Heteroskedastic Tensor Clustering

Yuchen Zhou*   Yuxin Chen*†

November 3, 2023

Abstract

Tensor clustering, which seeks to extract underlying cluster structures from noisy tensor observations, has gained increasing attention. One extensively studied model for tensor clustering is the tensor block model, which postulates the existence of clustering structures along each mode and has found broad applications in areas like multi-tissue gene expression analysis and multilayer network analysis. However, currently available computationally feasible methods for tensor clustering either are limited to handling i.i.d. sub-Gaussian noise or suffer from suboptimal statistical performance, which restrains their utility in applications that have to deal with heteroskedastic data and/or low signal-to-noise-ratio (SNR).

To overcome these challenges, we propose a two-stage method, named High-order HeteroClustering (HHC), which starts by performing tensor subspace estimation via a novel spectral algorithm called Thresholded Deflated-HeteroPCA, followed by approximate k-means to obtain cluster nodes. Encouragingly, our algorithm provably achieves exact clustering as long as the SNR exceeds the computational limit (ignoring logarithmic factors); here, the SNR refers to the ratio of the pairwise disparity between nodes to the noise level, and the computational limit indicates the lowest SNR that enables exact clustering with polynomial runtime. Comprehensive simulation and real-data experiments suggest that our algorithm outperforms existing algorithms across various settings, delivering more reliable clustering performance.

Keywords: tensor clustering, heteroskedastic noise, tensor block model, spectral clustering

Contents

---

*Department of Statistics and Data Science, Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA.
†Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA.

1   Introduction

The past few years have witnessed a surge of interest in tensor data analysis across various domains, including recommendation systems (Bi et al., 2018; Nasiri et al., 2014), neuroimaging (Wozniak et al., 2007; Zhou et al., 2013), computational imaging (Li and Li, 2010; Zhang et al., 2020), medical imaging (Fu and Dong, 2016), signal processing (Cichocki et al., 2015; Sidiropoulos et al., 2017), among other things. Compared to vectors and matrices, tensors (or multiway data arrays) offer the ability to characterize complex interrelations and interactions across multiple dimensions, allowing one to simultaneously capture the effects brought about by multiple factors. The growing prevalence of tensor data has sparked in-depth statistical research — from both methodological and theoretical perspectives — into various tensor estimation and learning problems (see, e.g., Zhou et al. (2013); Richard and Montanari (2014); Yuan and Zhang (2016); Xia et al. (2021); Cai et al. (2022a); Liu and Moitra (2020); Bi et al. (2021); Han et al. (2022b); Deng et al. (2023)).

    Within this body of research, one important problem that has garnered increasing attention is tensor clustering, which aims to extract the underlying cluster structures inherent in the observed tensor data. A model of this kind that has received widespread adoption is the *tensor block model* (Wang and Zeng, 2019; Chi et al., 2020; Han et al., 2022a). Concretely, suppose that the observed data takes the form of an order-three tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, drawn from the following data generating mechanism.

- *Cluster structure.* In each mode $1 \leq i \leq 3$, the n_i indices (or nodes) are divided into k_i clusters, and the cluster memberships of these nodes are encoded by a *cluster assignment vector* $\boldsymbol{z}_i^\star = [z_{i,j}^\star]_{1 \leq j \leq n_i} \in [k_i]^{n_i}$ such that

$$z_{i,j}^\star = \ell \ \text{ if the } j\text{th node falls within cluster } \ell \qquad (1 \leq j \leq n_i). \tag{1}$$

  Here and throughout, we denote by $[d] = \{1, \ldots, d\}$ for any positive integer d.

- *Observations.* The entries of the observed tensor $\boldsymbol{\mathcal{Y}} = [Y_{i,j,\ell}]$ obey

$$Y_{i,j,\ell} = S_{z_{1,i}^\star, z_{2,j}^\star, z_{3,\ell}^\star}^\star + E_{i,j,\ell}, \qquad \forall (i,j,\ell) \in [n_1] \times [n_2] \times [n_3]. \tag{2}$$

  Here, $S_{j_1,j_2,j_3}^\star$ is the (j_1, j_2, j_3)-th entry of an underlying core tensor $\boldsymbol{\mathcal{S}}^\star \in \mathbb{R}^{k_1 \times k_2 \times k_3}$ (which often has much lower dimension than $\boldsymbol{\mathcal{Y}}$), whereas the $E_{i,j,\ell}$'s represent independent zero-mean noise contaminating the measurements. Alternatively, we can rewrite this model in the tensor form as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}^\star + \boldsymbol{\mathcal{E}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \tag{3}$$

where $\boldsymbol{\mathcal{E}} = [E_{i,j,\ell}]_{(i,j,\ell)\in[n_1]\times[n_2]\times[n_3]}$ stands for a noise tensor, and the underlying tensor

$$\boldsymbol{\mathcal{X}}^\star = \left[S^\star_{z^\star_{1,i}, z^\star_{2,j}, z^\star_{3,\ell}}\right]_{(i,j,\ell)\in[n_1]\times[n_2]\times[n_3]} \tag{4}$$

exhibits block — and hence low-rank — structures. Crucially, for any set of indices $(i,j,\ell) \in [n_1] \times [n_2] \times [n_3]$, the mean of the observed entry $Y_{i,j,\ell}$ is determined by the cluster membership of (i,j,ℓ), making it possible to retrieve the cluster assignment information from the observed tensor as long as the noise level is not overly large.

- *Goal.* The aim is to reconstruct the underlying cluster structure along each mode — namely, recovering each $z^\star_i$ $(1 \le i \le 3)$ — on the basis of the observation $\boldsymbol{\mathcal{Y}}$.

Notably, this tensor block model finds a diverse range of applications. For instance, in multi-tissue gene expression analysis (Wang et al., 2019; Wang and Zeng, 2019; Han et al., 2022a), the expression levels of numerous genes are measured from various tissues across multiple individuals, and there could be natural group structures for genes, tissues, and individuals, respectively, which can be captured by the above model. Another instance arises from multilayer network analysis (Lei et al., 2020), wherein multiple (directed or undirected) graphs with identical vertices are gathered from various scenarios or experiments, inherently forming a tensor. A task stemming from this kind of data is identifying the clustering structures among the vertices and across different layers based on their connectivity patterns.

While numerous clustering algorithms have been studied in the literature, directly applying traditional clustering methods, such as k-means, to (the unfoldings of) the tensor data $\boldsymbol{\mathcal{Y}}$ may fail to capture the inherent tensor structures and therefore lead to unsatisfactory results. To overcome this issue, Han et al. (2022a) proposed a spectral clustering method called High-order Spectral Clustering (HSC), which starts by projecting the tensor data onto their estimated top singular subspaces along each mode, followed by an approximate k-means procedure to cluster nodes. Informally speaking, the singular subspace estimation procedure adopted by HSC directly calculates the left singular subspaces of the unfoldings of $\boldsymbol{\mathcal{Y}}$ along each mode — which we shall refer to as a vanilla SVD-based approach in the sequel. To further improve the spectral estimates, Han et al. (2022a) also came up with an algorithm called High-order Lloyd Algorithm (HLloyd) to iteratively refine the block membership estimates. When the noise tensor $\boldsymbol{\mathcal{E}}$ has i.i.d. sub-Gaussian noise entries, HSC (resp. HSC followed by HLloyd) provably achieves consistent (resp. exact) clustering results while accommodating a near-optimal range of signal-to-noise ratio (SNR) conditions (among polynomial-time algorithms) (Han et al., 2022a).

However, the HSC algorithm and the intriguing theory developed by Han et al. (2022a) fall short of accommodating heteroskedastic data, a common scenario in practice where variances of noise entries vary across locations. It has now been widely recognized that the vanilla SVD-based approach mentioned above could generate highly sub-optimal subspace estimates in the face of heteroskedastic noise (Zhang et al., 2022; Cai et al., 2021; Zhou and Chen, 2023); as a consequence, HSC, which is initialized based on this approach, becomes statistically sub-optimal. This issue severely hinders the performance of HSC in, say, a broad array of applications with discrete-valued observations — including multi-tissue gene expression data analysis and multilayer network data analysis — which often have to deal with heterogeneous data. To the best of our knowledge, no computationally efficient algorithm has been shown to achieve consistent estimation — not to mention exact recovery — of the underlying cluster structure under the widest possible SNR conditions.

1.1 Main contributions

Aimed at addressing the challenges resulting from heteroskedastic data, this paper proposes a new tensor clustering algorithm called High-order HeteroClustering (HHC). The key innovation compared to HSC lies in the development of a new paradigm for estimating the top singular subspaces of the unfolded tensor, in the hope of tackling heteroskedasticity. In a nutshell, the proposed HHC algorithm encompasses two stages:

1. *Subspace estimation.* This stage seeks to estimate the column subspaces of the unfoldings of $\boldsymbol{\mathcal{X}}^\star$ along each mode. Inspired by a spectral algorithm Deflated-HeteroPCA that proves effective in the face of heteroskedastic noise (Zhou and Chen, 2023), we propose a new variant, called Thresholded Deflated-HeteroPCA, that combines Deflated-HeteroPCA with a *data-driven* thresholding procedure. Our procedure only attempts to estimate the "useful" part of the column subspaces — that is, the subspace associated with

reasonably large singular values — of the unfolded tensors, which plays a crucial role in achieving statistical guarantees that are independent of the magnitude of the smallest singular value of $\boldsymbol{\mathcal{S}}^\star$.

2. *Approximate k-means.* Armed with the above subspace estimates, the second stage projects the unfolding of $\boldsymbol{\mathcal{Y}}$ onto the estimated subspace for denoising purposes, followed by an approximate k-means algorithm to cluster nodes.

Encouragingly, the proposed HHC algorithm allows for exact clustering as long as a certain "necessary" SNR condition holds (up to logarithmic factors), where the SNR is captured by the ratio of certain "pairwise" difference between nodes to the noise level. Here, a "necessary" SNR condition refers to a condition that is essential to ensure that the cluster assignment vectors $\boldsymbol{z}_i^\star$ can be exactly recovered in polynomial time. Empirically, we conduct simulation experiments and find that HHC can reliably estimate the cluster structures, and that HHC combined with HLloyd (Han et al., 2022a) enables enhanced numerical performance. We also apply our method (HHC + HLloyd) and HSC + HLloyd to the flight route network data, in which our method leads to better clustering results. It is noteworthy that: while the current paper focuses on three-way tensor for simplicity of presentation, both our algorithm and the proof can be straightforwardly extended to accommodate general higher-order tensors.

Paper organization. The rest of this article is organized as follows. Section 2 formulates the mathematical model and introduces the key assumptions, while Section 3 presents the proposed algorithm. The theoretical guarantees for our algorithm are provided in Section 4, with the analysis deferred to the appendix. Numerical performance on both synthetic and real data is reported in Section 5.

1.2   Notation

Throughout the paper, we denote $[n] := \{1, \ldots, n\}$ for any integer $n > 0$. We often use bold capital letters (e.g., $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$) and bold lowercase letters (e.g., $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$) to denote matrices and vectors, respectively, and employ boldface calligraphic letters (e.g., $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{Z}}$) to represent tensors. For any matrix $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$, we let $\lambda_i(\boldsymbol{X})$ and $\sigma_i(\boldsymbol{X})$ denote the i-th largest eigenvalue (in magnitude) and the i-th largest singular value of $\boldsymbol{A}$, respectively. Define $\|\cdot\|_{\mathrm{F}}$ for Frobenious norm and $\|\cdot\|$ for spectral norm. We denote by $\boldsymbol{A}_{i,:}$ and $\boldsymbol{A}_{:,j}$ the i-th column and the j-th row of a matrix $\boldsymbol{A}$, respectively, and define its $\ell_{2,\infty}$ norm as $\|\boldsymbol{A}\|_{2,\infty} := \max_{i \in [n_1]} \|\boldsymbol{A}_{i,:}\|_2$. Let $\mathcal{O}^{n,r} := \{\boldsymbol{U} \in \mathbb{R}^{n \times r} : \boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_r\}$ denote the set containing all n-by-r matrices with orthonormal columns. We use $\mathcal{P}_{\mathsf{diag}}$ to represent the projection that keeps all diagonal entries and zeros out all non-diagonal entries, and define $\mathcal{P}_{\mathsf{off\text{-}diag}}(\boldsymbol{M}) := \boldsymbol{M} - \mathcal{P}_{\mathsf{diag}}(\boldsymbol{M})$ for any $\boldsymbol{M} \in \mathbb{R}^{n \times n}$. For any vector $\boldsymbol{a} = [a_i]_{1 \le i \le n}$, we denote by $\mathsf{diag}(\boldsymbol{a})$ the diagonal matrix whose (i,i)-th entry is a_i. We let $C, c, C_0, c_0, \ldots$ denote absolute constants whose values may change from line to line.

For any two matrices $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{B} \in \mathbb{R}^{p \times q}$, we define the Kronecker product of them as

$$\boldsymbol{A} \otimes \boldsymbol{B} := \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}.$$

For any tensor $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ and any matrix $\boldsymbol{V}_1 \in \mathbb{R}^{n_1 \times r_1}$, the multi-linear product $\times_1$ is defined as

$$\boldsymbol{\mathcal{G}} \times_1 \boldsymbol{V}_1 = \left(\sum_{j=1}^{r_1} G_{j,i_2,i_3} V_{i_1,j} \right)_{i_1 \in [n_1], i_2 \in [r_2], i_3 \in [r_3]}.$$

We can also define $\times_2$ and $\times_3$ analogously. For any tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $1 \le j \le 3$, let $\mathcal{M}_j(\boldsymbol{\mathcal{X}}) \in \mathbb{R}^{n_j \times (n_1 n_2 n_3 / n_j)}$ denote the j-th matricization of $\boldsymbol{\mathcal{X}}$ such that

$$\left[\mathcal{M}_1\left(\boldsymbol{\mathcal{X}}\right)\right]_{i_1, i_2 + n_2(i_3 - 1)} = \left[\mathcal{M}_2\left(\boldsymbol{\mathcal{X}}\right)\right]_{i_2, i_3 + n_3(i_1 - 1)} = \left[\mathcal{M}_3\left(\boldsymbol{\mathcal{X}}\right)\right]_{i_3, i_1 + n_1(i_2 - 1)} = X_{i_1, i_2, i_3}$$

for all $(i_1, i_2, i_3) \in [n_1] \times [n_2] \times [n_3]$. We further define the Frobenious norm of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as

$$\|\boldsymbol{\mathcal{X}}\|_{\mathrm{F}} = \left(\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} X_{i,j,k}^2 \right)^{1/2}.$$

4

In addition, we say that $f(x) \lesssim g(x)$ or $f(x) = O(g(x))$ if $|f(x)| \leq Cg(x)$ for some constant $C > 0$; we let $f(x) \gtrsim g(x)$ denote $f(x) \geq C|g(x)|$ for some constant $C > 0$; we say $f(x) \asymp g(x)$ or $f(x) = \Omega(g(x))$ if $f(x) \lesssim g(x)$ and $f(x) \gtrsim g(x)$ hold; we use the notation $f(x) \ll g(x)$ to represent that $f(n_1, n_2) \leq cg(n_1, n_2)$ holds for some sufficiently small constant $c > 0$, and we say $f(n_1, n_2) \gg g(n_1, n_2)$ if $g(n_1, n_2) \ll f(n_1, n_2)$. In addition, we use $f(n_1, n_2) = o(g(n_1, n_2))$ to indicate that $f(n_1, n_2)/g(n_1, n_2) \to 0$ as $\min\{n_1, n_2\} \to \infty$. For any $a, b \in \mathbb{R}$, let $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$. Moreover, denote by Φ the set of all permutations $\phi : [k] \to [k]$. For any $\boldsymbol{z} = [z_j]_{1 \leq j \leq n}, \overline{\boldsymbol{z}} = [\widehat{z}_j]_{1 \leq j \leq n} \in [k]^n$, we define the misclassification rate as follows:

$$\mathsf{MCR}\,(\boldsymbol{z}, \widehat{\boldsymbol{z}}) := \inf_{\phi \in \Phi} \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}\big\{\widehat{z}_j \neq \phi(z_j)\big\}. \tag{5}$$

Also, for any $\phi \in \Phi$, we use the notation $\widehat{\boldsymbol{z}} = \phi(\boldsymbol{z})$ to mean that

$$\widehat{z}_j = \phi(z_j), \qquad \forall j \in [n].$$

Clearly, $\mathsf{MCR}\,(\boldsymbol{z}, \widehat{\boldsymbol{z}}) = 0$ holds if and only if $\widehat{\boldsymbol{z}} = \phi(\boldsymbol{z})$ for some $\phi \in \Phi$.

2 Problem formulation

2.1 Models

Recall that the vectors $\boldsymbol{z}_i^\star$'s encode the cluster assignment. The tensor block model (2) can be equivalently written as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}^\star + \boldsymbol{\mathcal{E}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \tag{6}$$

where the low-rank tensor $\boldsymbol{\mathcal{X}}^\star$ can be decomposed as $\boldsymbol{\mathcal{X}}^\star = \boldsymbol{\mathcal{S}}^\star \times_1 \boldsymbol{M}_1^\star \times_2 \boldsymbol{M}_2^\star \times_3 \boldsymbol{M}_3^\star$. Here, $\boldsymbol{\mathcal{S}}^\star \in \mathbb{R}^{k_1 \times k_2 \times k_3}$ stands for the core tensor, and $\boldsymbol{M}_i^\star \in \{0, 1\}^{n_i \times k_i}$ $(1 \leq i \leq 3)$ represents a membership matrix satisfying

$$(\boldsymbol{M}_i^\star)_{j,\ell} = \begin{cases} 1, & \text{if } z_{i,j}^\star = \ell, \\ 0, & \text{else.} \end{cases} \tag{7}$$

The goal is to recover the cluster assignment vectors $\{\boldsymbol{z}_i^\star\}$, or equivalently, the membership matrices $\{\boldsymbol{M}_i^\star\}$, based on the observation $\boldsymbol{\mathcal{Y}}$.

We would like to immediately single out two special cases of the above model, which represent two distinctive types of noise models of important practical value (see more discussions in Han et al. (2022a)).

1. *Sub-Gaussian tensor block models.* In this scenario, the entries of the noise tensor $\boldsymbol{\mathcal{E}}$ are zero-mean sub-Gaussian random variables generated independently, which capture, say, random contamination during the data collection process for measuring $\boldsymbol{\mathcal{X}}^\star$. Importantly, we allow the noise to be heteroskedastic — namely, the variance of the noise components can be location-varying — a remarkable extension of the one studied in Han et al. (2022a) (recall that the noise is assumed to be i.i.d. therein).

2. *Stochastic tensor block models.* Consider another scenario where each entry of the core tensor $\boldsymbol{\mathcal{S}}^\star$ falls within $[0, 1]$, and the observed entries $\{Y_{i,j,\ell}\}$ are independent Bernoulli random variables satisfying

$$Y_{i,j,\ell} = \begin{cases} 1, & \text{with probability } S_{z_{1,i}^\star, z_{2,j}^\star, z_{3,\ell}^\star}^\star, \\ 0, & \text{with probability } 1 - S_{z_{1,i}^\star, z_{2,j}^\star, z_{3,\ell}^\star}^\star. \end{cases} \tag{8}$$

This scenario can be understood as a generalization of the classical bipartite stochastic block model (e.g., Florescu and Perkins (2016)). Informally, each binary variable $Y_{i,j,\ell}$ encodes whether there is a hyper-edge connecting the vertices $(i, j, \ell) \in [n_1] \times [n_2] \times [n_3]$, and the probability that such a hyper-edge is present is determined by the clusters they belong to. Clearly, this noise model is, in general, heteroskedastic.

We seek to develop a suite of theory and algorithms that can readily accommodate these two important scenarios.

2.2 Assumptions and definitions

Next, let us introduce a couple of definitions and assumptions that will be used throughout this paper. Before proceeding, we find it helpful to define

$$n = \max_{1 \leq i \leq 3} n_i, \qquad \text{and} \qquad k = \max_{1 \leq i \leq 3} k_i, \tag{9}$$

and denote

$$\omega^2_{i_1,i_2,i_3} \coloneqq \mathbb{E}\left[E^2_{i_1,i_2,i_3}\right] \qquad \text{and} \qquad \omega^2_{\mathsf{max}} \coloneqq \max_{i_1 \in [n_1], i_2 \in [n_2], i_3 \in [n_3]} \omega^2_{i,j,k}. \tag{10}$$

We start by imposing the following assumption on the noise tensor $\boldsymbol{\mathcal{E}}$.

Assumption 1. *Suppose that the noise components satisfy the following conditions:*

1. *The E_{i_1,i_2,i_3}'s are independent and zero-mean;*

2. *For every (i_1, i_2, i_3), one has $\mathbb{P}(|E_{i_1,i_2,i_3}| > B) \leq n^{-24}$ for some quantity B satisfying*

$$B \leq C_{\mathsf{b}} \omega_{\mathsf{max}} \frac{(n_1 n_2 n_3)^{1/4}}{\log n},$$

where $C_{\mathsf{b}} > 0$ is some universal constant.

Remark 1. *Here, the tail condition $\mathbb{P}(|E_{i_1,i_2,i_3}| > B) \leq n^{-24}$ can be relaxed to $\mathbb{P}(|E_{i_1,i_2,i_3}| > B) \leq n^{-c}$ for any constant $c \geq 4$. We choose the exponent 24 to streamline the presentation of the proof a little bit.*

Notably, Assumption 1 is very mild and accommodates a broad range of scenarios of interest. For example, all ω_{max}-sub-Gaussian random variables (see, e.g., Vershynin (2018)) satisfy Condition 2 of Assumption 1 with $B \asymp \omega_{\mathsf{max}} \sqrt{\log n}$; centered Poisson random variables also easily satisfy this condition (Boucheron et al., 2013; Zhang and Zhou, 2020). In addition, the aforementioned stochastic tensor block model (8) obeys Assumption 1 as long as the following conditions hold:

$$\frac{2 \log^2 n}{C_{\mathsf{b}}^2 (n_1 n_2 n_3)^{1/2}} \leq S^{\star}_{i_1,i_2,i_3} \leq 1 - \frac{2 \log^2 n}{C_{\mathsf{b}}^2 (n_1 n_2 n_3)^{1/2}}, \qquad \forall (i_1, i_2, i_3) \in [k_1] \times [k_2] \times [k_3]. \tag{11}$$

Recognizing that $(n_1 n_2 n_3)^{1/2} \gg \log^2 n$, we see that the validity of Condition (11) is guaranteed as long as the entries of the center tensor $\boldsymbol{\mathcal{S}}^{\star}$ are not extremely close to 0 or 1.

In addition, we introduce the following parameter that reveals cluster size information.

Definition 1 (Balance of cluster sizes). *Let $\beta \leq 1$ denote the largest quantity such that*

$$\left| \left\{ j \in [n_i] : (\boldsymbol{z}^{\star}_i)_j = \ell \right\} \right| \geq \beta n_i / k_i, \qquad 1 \leq i \leq 3, \quad 1 \leq \ell \leq k_i. \tag{12}$$

In words, the parameter β measures how balanced these cluster sizes are, with a larger β indicating more balanced cluster sizes; for instance, $\beta = 1$ corresponds to the scenario where all clusters are of the same size.

Another quantity that plays an important role in our theory is concerned with the separation condition.

Definition 2 (Separation). *For any $1 \leq i \leq 3$, define*

$$\Delta_i^2 \coloneqq \min_{1 \leq j_1 \neq j_2 \leq k_i} \left\| \left[\mathcal{M}_i(\mathcal{S}^{\star}) \right]_{j_1,:} - \left[\mathcal{M}_i(\mathcal{S}^{\star}) \right]_{j_2,:} \right\|_2^2 \tag{13}$$

to be the minimum distance between the rows of $\mathcal{M}_i(\mathcal{S})$, the i-th matrization of the core tensor. We can also define

$$\Delta_{\mathsf{min}}^2 \coloneqq \min \left\{ \Delta_1^2, \Delta_2^2, \Delta_3^2 \right\}. \tag{14}$$

In a nutshell, the above separation between clusters captures the "signal strength," which determines the degree of noise variability that can be tolerated without compromising the feasibility of exact clustering. In contrast to generic low-rank tensor estimation problems where the signal strength is typically represented by the least singular value of $\boldsymbol{\mathcal{S}}^\star$, the above separation condition is more natural in capturing the differentiability between two different clusters. Noteworthily, having a desirable separation condition does not necessarily imply that the least singular value of $\boldsymbol{\mathcal{S}}^\star$ is sufficiently large.

Armed with the above separation metrics, we can readily introduce the following quantity to quantify the "signal-to-noise ratio":

$$\mathsf{SNR} \coloneqq \Delta_{\mathsf{min}}/\omega_{\mathsf{max}}. \tag{15}$$

An ideal tensor clustering algorithm would allow for exact clustering for the widest possible range of SNRs.

3   Algorithm: High-order HeteroClustering

In this section, we introduce the proposed procedure for tensor clustering. Akin to other spectral-method-based tensor clustering schemes, our algorithm begins by performing subspace estimation with the aid of matricization, followed by an application of the (approximate) k-means algorithm to estimate clustering assignment.

3.1   Stage 1: subspace estimation via Thresholded Deflated-HeteroPCA

First of all, we would like to estimate the "important" column subspaces of $\boldsymbol{\mathcal{X}}$ after matricization along each dimension, namely, the important column subspace of $\boldsymbol{X}_i = \mathcal{M}_i(\boldsymbol{\mathcal{X}})$ for each $1 \leq i \leq 3$. It is noteworthy that: it might not be necessary to estimate the entire rank-k_i column subspace for $\mathcal{M}_i(\boldsymbol{\mathcal{X}})$, given that those singular vectors corresponding to overly small singular values might only exert a negligible impact on the final clustering outcome. Instead, for each $1 \leq i \leq 3$, it often suffices to find a suitable estimator $\boldsymbol{U}_i \in \mathcal{O}^{n_i, r_i}$, for some $r_i \leq k_i$, that can reliably estimate the subspace formed by the singular vectors associated with large enough singular values of $\boldsymbol{X}_i$. As alluded to previously, however, the vanilla SVD-based approach (i.e., directly computing the SVD of $\boldsymbol{X}_i$) might result in unsatisfactory subspace estimation results when the noise is heteroskedastic. This motivates us to develop a more sophisticated algorithm, inspired by our recent work Zhou and Chen (2023).

Review: Deflated-HeteroPCA.   The recently proposed Deflated-HeteroPCA algorithm is particularly effective in subspace estimation in the face of heteroskedastic noise (Zhou and Chen, 2023), which we briefly review here. Let $\boldsymbol{Y}_i = \mathcal{M}_i(\boldsymbol{\mathcal{Y}})$ denote the i-th matricization of $\boldsymbol{Y}_i$. Starting from a diagonal-deleted gram matrix $\boldsymbol{G}_0 = \mathcal{P}_{\mathsf{off\text{-}diag}}(\boldsymbol{Y}_i \boldsymbol{Y}_i^\top)$, the main idea of Deflated-HeteroPCA is to sequentially choose ranks $0 < r_1 < \cdots < r_{k_{\mathsf{max}}} = r$ that divide the eigenvalues of $\boldsymbol{X}_i^\star \boldsymbol{X}_i^{\star\top}$ into "well-conditioned" and sufficiently separated subblocks, and progressively improve the estimation accuracy. Informally, we sequentially incorporate new subblocks into consideration and invoke the HeteroPCA algorithm (Zhang et al., 2022) to gradually improve the estimation accuracy of both the column subspace and the diagonal entries of $\boldsymbol{X}_i^\star \boldsymbol{X}_i^{\star\top}$. As proven in Zhou and Chen (2023), Deflated-HeteroPCA enjoys theoretical guarantees that are condition-number-free and accommodate the widest possible range of SNRs, all of which are appealing for the tensor clustering application. However, existing theory of Deflated-HeteroPCA requires the least singular value of $\boldsymbol{X}_i^\star$ to exceed a certain level (depending on the noise variance), which might oftentimes be unnecessary for clustering applications. In fact, even in the presence of a large separation metric (14), we cannot preclude the possibility of $\boldsymbol{X}_i^\star$ having a (nearly) zero singular value, thus limiting the utility of Deflated-HeteroPCA.

Proposed procedure: Thresholded Deflated-HeteroPCA.   To address the aforementioned issue, we incorporate a thresholding procedure into Deflated-HeteroPCA in order to make sure we only include important subblocks. More specifically, suppose that in the k-th round, we choose rank r_k and perform HeteroPCA with rank r_k and initialization $\boldsymbol{G}_{k-1}$ to obtain $\boldsymbol{G}_k$, where both $\boldsymbol{G}_{k-1}$ and $\boldsymbol{G}_k$ are intermediate estimates of the gram matrix $\boldsymbol{X}_i^\star \boldsymbol{X}_i^{\star\top}$. We then decide whether to proceed to the $(k+1)$-th round based on whether the condition $\sigma_{r_k+1}(\boldsymbol{G}_k) > \tau$ is met for some pre-determined threshold τ. With a properly chosen τ, we can

---

Algorithm 1: Thresholded Deflated-HeteroPCA$(\boldsymbol{Y}, r, \tau, \{t_j\}_{j \geq 1})$

---

1 input: data matrix $\boldsymbol{Y}$, rank r, threshold τ, maximum numbers of iterations $\{t_j\}_{j \geq 1}$.

2 initialization: $j = 0, r_0 = 0, \boldsymbol{G}_0 = \mathcal{P}_{\text{off-diag}}\left(\boldsymbol{Y}\boldsymbol{Y}^\top\right)$.

    /* sequentially invoke HeteroPCA until eigenvalues fall below the threshold.   */

3 while $r_j < r$ and $\sigma_{r_j+1}(\boldsymbol{G}_j) > \tau$ **do**

4 $j \leftarrow j + 1$.

5   compute $r_j = \mathsf{RankSelection}(\boldsymbol{G}_{j-1}, r, r_{j-1})$.

6 $(\boldsymbol{G}_j, \boldsymbol{U}_j) = \mathsf{HeteroPCA}(\boldsymbol{G}_{j-1}, r_j, t_j)$.

7 output: subspace estimate $\boldsymbol{U} = \boldsymbol{U}_j$.

Algorithm 2: RankSelection$(\boldsymbol{G}_{j-1}, r, r_{j-1})$

---

1 input: rank r, selected rank r_{j-1}, matrix $\boldsymbol{G}_{j-1}$.

    /* identify a subblock of eigenvalues that are well-conditioned and well-separated from the remaining eigenvalues.   */

2 output rank

$$r_j = \begin{cases} \max \mathcal{R}_j, & \text{if } \mathcal{R}_j \neq \emptyset, \\ r, & \text{otherwise,} \end{cases} \tag{16}$$

where

$$\mathcal{R}_j := \left\{ r' : r_{j-1} < r' \leq r, \ \frac{\sigma_{r_{j-1}+1}(\boldsymbol{G}_{j-1})}{\sigma_{r'}(\boldsymbol{G}_{j-1})} \leq 4 \text{ and } \sigma_{r'}(\boldsymbol{G}_{j-1}) - \sigma_{r'+1}(\boldsymbol{G}_{j-1}) \geq \frac{1}{r}\sigma_{r'}(\boldsymbol{G}_{j-1}) \right\}.$$

---

extract sufficient information needed for clustering. The details of Thresholded Deflated-HeteroPCA can be found in Algorithm 1. A theoretically-guided procedure for selecting the tuning parameter τ is deferred to Section 4.2.

3.2 Stage 2: approximate k-means

Having obtained the subspace estimates $\boldsymbol{U}_j, 1 \leq j \leq 3$, we would like to extract clustering information based on these subspace estimates as well as the observation $\boldsymbol{\mathcal{Y}}$. More specifically, let us construct the following matrix

$$\widehat{\boldsymbol{B}}_i = \boldsymbol{U}_i\boldsymbol{U}_i^\top \mathcal{M}_i(\boldsymbol{\mathcal{Y}})\left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1}\right) \in \mathbb{R}^{n_i \times (r_1 r_2 r_3 / r_i)} \tag{17}$$

for each $1 \leq i \leq 3$, and we propose to apply the (approximate) k-means algorithm on the rows of $\widehat{\boldsymbol{B}}_i$ to estimate the cluster assignment vectors. Here, the indices $i + 1$ and $i + 2$ are computed module 3. This procedure is equivalent to applying (approximate) k-means on the rows of the i-th matricization of the tensor estimate $\widehat{\boldsymbol{\mathcal{Y}}} = \boldsymbol{\mathcal{Y}} \times_1 \boldsymbol{U}_1\boldsymbol{U}_1^\top \times_2 \boldsymbol{U}_2\boldsymbol{U}_2^\top \times_3 \boldsymbol{U}_3\boldsymbol{U}_3^\top$, but operates upon matrices with significantly reduced sizes. Recognizing that performing exact k-means can be computationally intractable, we instead employ an M-approximate k-means approach for some $M > 1$. To be more specific, we find a cluster assignment vector estimate $\widehat{z}_i$ and centroids $\{\widehat{\boldsymbol{b}}_j^{(i)}\}$ that satisfy (18a) for each $1 \leq i \leq 3$. This relaxed version of k-means can be efficiently solved by a number of algorithms. For example, for $M = O(\log k)$, this problem can be solved with running time $O(nk^3)$ by k-means++ (Bahmani et al., 2012; Arthur and Vassilvitskii, 2007)[1]. The requirement (18b) ensures the "optimality" of the cluster estimates given the centroids.

---

[1]More generally, the time complexity of k-means++ for n points in $\mathbb{R}^d$ is $O(nkd)$. Here, the dimension $d = r_1 r_2 r_3 / r_i \leq k^2$ and thus the time complexity does not exceed $O(nk^3)$.

Algorithm 3: HeteroPCA($\boldsymbol{G}_{\mathsf{in}}$, r, t_{max}) (Zhang et al., 2022)

---

1 **input:** symmetric matrix $\boldsymbol{G}_{\mathsf{in}}$, rank r, number of iterations t_{max}.

2 **initialization:** $\boldsymbol{G}^0 = \boldsymbol{G}_{\mathsf{in}}$.

3 **for** $t = 0, 1, \ldots, t_{\mathsf{max}}$ **do**

4 $\quad$ $\boldsymbol{U}^t \boldsymbol{\Lambda}^t \boldsymbol{U}^{t\top} \leftarrow$ rank-r leading eigendecompostion of $\boldsymbol{G}^t$.

5 $\quad$ $\boldsymbol{G}^{t+1} = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^t\right) + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{U}^t \boldsymbol{\Lambda}^t \boldsymbol{U}^{t\top}\right)$.

6 **output:** matrix estimate $\boldsymbol{G} = \boldsymbol{G}^{t_{\mathsf{max}}}$ and subspace estimate $\boldsymbol{U} = \boldsymbol{U}^{t_{\mathsf{max}}}$.

Algorithm 4: High-order HeteroClustering (HHC)

---

1 **input:** observed tensor $\boldsymbol{\mathcal{Y}}$, numbers of clusters k_1, k_2, k_3, numbers of iterations $\{t_{i,j}\}_{1 \le i \le 3, j \ge 1}$, thresholds τ_1, τ_2, τ_3, relaxation factor $M > 1$.

$\quad$ /* Stage 1: subspace estimation $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ */

2 **subspace estimation:** for each $1 \le i \le 3$, compute $\boldsymbol{U}_i \in \mathcal{O}^{n_i, k_i}$ as follows

$$\boldsymbol{U}_i = \begin{cases} \mathsf{Thresholded\ Deflated\text{-}HeteroPCA}\big(\mathcal{M}_i(\boldsymbol{\mathcal{Y}}), k_i, \tau_i, \{t_{i,j}\}_{j \ge 1}\big), & k_i \ge 2, \\ (1, \ldots, 1)^\top / \sqrt{n_i}, & k_i = 1, \end{cases}$$

$\quad$ and set $\boldsymbol{U}_4 = \boldsymbol{U}_1$ and $\boldsymbol{U}_5 = \boldsymbol{U}_2$ for convenience.

$\quad$ /* Stage 2: approximate k-means $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ */

3 **for** $i = 1, 2, 3$ **do**

4 $\quad$ compute $\widehat{\boldsymbol{B}}_i = \boldsymbol{U}_i \boldsymbol{U}_i^\top \mathcal{M}_i(\boldsymbol{\mathcal{Y}})\big(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1}\big) \in \mathbb{R}^{n_i \times (r_1 r_2 r_3 / r_i)}$.

5 $\quad$ perform approximate k-means on the rows of $\widehat{\boldsymbol{B}}_i$, i.e., find a cluster assignment vector estimate $\widehat{\boldsymbol{z}}_i \in [k_i]^{n_i}$ and center estimates $\widehat{\boldsymbol{b}}_l^{(i)} \in \mathbb{R}^{r_1 r_2 r_3 / r_i}, l \in [k_i]$ such that

$$\sum_{j=1}^{n_i} \left\| \big(\widehat{\boldsymbol{B}}_i\big)_{j,:}^\top - \widehat{\boldsymbol{b}}_{\widehat{z}_{i,j}}^{(i)} \right\|_2^2 \le M \min_{\substack{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{k_i} \in \mathbb{R}^{r_1 r_2 r_3 / r_i} \\ \boldsymbol{z}_i \in [k_i]^{n_i}}} \sum_{j=1}^{n_i} \left\| \big(\widehat{\boldsymbol{B}}_i\big)_{j,:}^\top - \boldsymbol{b}_{z_{i,j}} \right\|_2^2, \qquad (18\mathrm{a})$$

$$\widehat{z}_{i,j} \in \arg\min_{\ell \in [k_i]} \left\| \big(\widehat{\boldsymbol{B}}_i\big)_{j,:}^\top - \widehat{\boldsymbol{b}}_\ell^{(i)} \right\|_2, \qquad \forall j \in [n_i]. \qquad (18\mathrm{b})$$

6 **output:** estimates $\widehat{\boldsymbol{z}}_1, \widehat{\boldsymbol{z}}_2, \widehat{\boldsymbol{z}}_3$ of cluster assignment vectors.

3.3 Full procedure

The full procedure of the proposed High-order HeteroClustering (HHC) is summarized in Algorithm 4. Since both Thresholded Deflated-HeteroPCA and approximate k-means are polynomial-time algorithms, our proposed method is computationally efficient. Empirically, we recommend using the high-order Lloyd Algorithm (HLloyd) (Han et al., 2022a) to further refine the clustering results obtained by HHC, where the description of this refinement procedure is deferred to Algorithm 5 in Appendix A. As will be demonstrated in Section 5, the combined application of our algorithm with HLloyd can yield superior empirical performance when compared to other methods.

4 Main theory

In this section, we develop theoretical performance guarantees for our algorithm proposed in Section 3. Before proceeding, we find it convenient to introduce the following additional notation for each $1 \le i \le 3$:

- $\sigma_{i,j}^\star$: the j-th largest singular value of $\mathcal{M}_i(\boldsymbol{\mathcal{S}}^\star)$ (i.e., the i-th matricization of $\boldsymbol{\mathcal{S}}^\star$);

- $\{r_{i,j}\}_{j \ge 1}$: the ranks selected in Algorithm 1 with the input matrix $\boldsymbol{Y} = \mathcal{M}_i(\boldsymbol{\mathcal{Y}})$, the rank $r = k_i$, the

threshold $\tau = \tau_i$, and the numbers of iterations $\{t_{i,j}\}_{j \geq 1}$;

- $r_{i,j_{\max}^i}$: the largest rank selected by Algorithm 1 with the above inputs.

4.1 Theoretical guarantees for exact clustering

The first theorem below demonstrates that HHC achieves intriguing exact recovery guarantees, as long as the tuning parameters $\{\tau_i\}_{1 \leq i \leq 3}$ are suitably selected.

Theorem 1. *Assume that $k_i \lesssim 1$ for each $1 \leq i \leq 3$, $\beta \asymp 1$, and*

$$n_1 n_2 n_3 \geq c_1 n^2, \tag{19a}$$

$$c_\tau (n_1 n_2 n_3)^{1/2} \log^2 n \leq \tau_i / \omega_{\mathsf{max}}^2 \leq C_\tau (n_1 n_2 n_3)^{1/2} \log^2 n, \qquad \forall 1 \leq i \leq 3, \tag{19b}$$

$$\mathsf{SNR} \ = \Delta_{\mathsf{min}} / \omega_{\mathsf{max}} \geq C_1 \sqrt{M} \, (n_1 n_2 n_3)^{-1/4} \log n, \tag{19c}$$

where C_1, c_1, C_τ and c_τ are some large enough positive constants. Suppose Assumption 1 holds. If the numbers of iterations satisfy

$$t_{i,j} \geq \log \left(C \frac{\sigma_{i,r_{i,j-1}+1}^{\star 2}}{\sigma_{i,r_{i,j}+1}^{\star 2}} \right), \qquad 1 \leq j \leq j_{\mathsf{max}}^i - 1, \tag{20a}$$

$$t_{i,j_{\mathsf{max}}^i} \geq \log \left(n^3 \frac{\sigma_{i,r_{i,j_{\mathsf{max}}^i - 1}+1}^{\star 2}}{\omega_{\mathsf{max}}^2} \right), \tag{20b}$$

for all $1 \leq i \leq 3$ with $C > 0$ some large enough constant, then with probability exceeding $1 - O(n^{-10})$, the misclassification rate (cf. (5)) of the outputs $\{\widehat{z}_i\}$ returned by Algorithm 4 satisfy

$$\mathsf{MCR}\big(\widehat{z}_i, z_i^\star\big) = 0, \qquad \forall 1 \leq i \leq 3.$$

In words, this theorem asserts that Algorithm 4 enables exact clustering under the assumptions imposed above. A more general version of Theorem 1 — which allows k_i and β grows with n — as well as its proof can be found in Section B.

Let us now take a moment to discuss the conditions assumed in Theorem 1. Condition (19a) assumes that the dimensions of the observed tensor are not extremely unbalanced; for instance, it holds in the scenario where $n_1 \lesssim n_2 n_3$, $n_2 \lesssim n_1 n_3$ and $n_3 \lesssim n_1 n_2$. In the regime where $n_1 \asymp n_2 \asymp n_3 \asymp n$ and $M \asymp \log k \asymp 1$, the signal-to-noise ratio condition (19c) simplifies to

$$\mathsf{SNR} \gtrsim n^{-3/4} \log n. \tag{21}$$

Interestingly, this condition (21) is almost necessary among polynomial-time algorithms; as shown in Han et al. (2022a, Theorem 7), if $\mathsf{SNR} = n^\gamma$ for any $\gamma < -3/4$, then there exists no polynomial-time algorithm that can exactly recover the cluster assignment vectors. Furthermore, combining Theorem 1 with Han et al. (2022a, Theorem 2), we know that HHC+ HLloyd and HHC can achieve the same theoretical guarantees in terms of the exact cluster recovery. In other words, applying HLloyd to further refinement would not degrade the theoretical performance at all. As we will illustrate in Section 5, HHC combined with HLloyd might sometimes achieve improved empirical results compared to HHC on its own.

Comparisons with HSC and HSC + HLloyd. To highlight the advantages of our algorithm and theory, we make comparisons with the state-of-the-art prior work Han et al. (2022a), which proposed the HSC algorithm and its combination with a follow-up HLloyd procedure. Firstly, in stark contrast to HSC and HSC + HLloyd — which assumes *identical variances* of the noise entries in order to guarantee their desired theoretical results Han et al. (2022a, Theorems 3 and 4) — our algorithm HHC is able to handle heteroskedastic noise efficiently without compromising the applicable range of SNRs. Secondly, in comparison with Han et al. (2022a, Theorems 3 and 4) that assume sub-Gaussian noise, our assumption is more mild and can accommodate a wider range of applications, including those with binary or count data outcomes. In addition, HHC can exactly recover the cluster assignment vectors without the aid of HLloyd, a feature that stands in contrast to HSC.

Other prior results. In addition to Han et al. (2022a), the tensor block model has been studied in several other past works. Focusing on sub-Gaussian noise, Wang and Zeng (2019) characterized the misclassification rate and the tensor estimation error for the the least-square estimator; this estimator, however, is computationally intractable. Chi et al. (2020) proposed a convex method and provided theoretical guarantees for the tensor estimation error, but they did not establish misclassification-rate-based theory for their proposed method. Agterberg and Zhang (2022) further investigated a more general model, called the tensor mixed-membership block model, in the presence of sub-Gaussian noise. When applied to Model (2) with $k \asymp 1$ and $\beta \asymp 1$, their signal-to-noise ratio condition becomes

$$\Delta/\omega_{\mathsf{max}} \gtrsim \kappa^2 \frac{n\sqrt{\log n}}{(n_1 n_2 n_3)^{1/2}(\min_i n_i)^{1/4}},$$

where κ is the condition number of the tensor $\mathcal{S}^\star$ satisfying $\kappa \lesssim (\min_i n_i)^{1/8}$. In comparison, Theorem 1 does not require any assumptions on κ and our signal-to-noise ratio condition is less stringent.

A glimpse of proof highlights. To prove that HHC alone is enough to achieve exact clustering, a crucial step lies in carefully controlling the magnitude of $\|(\boldsymbol{I} - \boldsymbol{U}_i\boldsymbol{U}_i^\top)\boldsymbol{X}_i^\star\|_{2,\infty}$. Here, $\boldsymbol{X}_i^\star = \mathcal{M}_i(\boldsymbol{\mathcal{X}}^\star)$ is the i-th matricization of $\boldsymbol{\mathcal{X}}^\star$, and $\boldsymbol{U}_i$ is the subspace estimator. Unlike the subspace/matrix estimation problems considered in the literature, we aim to derive sharp and condition-number-free $\ell_{2,\infty}$ guarantees for $(\boldsymbol{I} - \boldsymbol{U}_i\boldsymbol{U}_i^\top)\boldsymbol{X}_i^\star$ without imposing any restrictions on the condition number and the least singular value of $\boldsymbol{X}_i^\star$. In this context, existing techniques for deriving $\ell_{2,\infty}$ guarantees are inadequate for reaching our target bound; for instance, existing leave-one-out analyses (Zhong and Boumal, 2018; Chen et al., 2021a; Ma et al., 2020) require the condition number to not be overly large, whereas the subspace representation theorem (Xia, 2021; Zhou and Chen, 2023) relies on assumptions on the least singular value.

To establish the desired performance guarantees, we develop a new technique (i.e., Lemma 1) that allows for effective control of $\|(\boldsymbol{I} - \boldsymbol{U}_i\boldsymbol{U}_i^\top)\boldsymbol{X}_i^\star\|_{2,\infty}$ via bounding an infinite sum of $\ell_{2,\infty}$ norms of polynomials of the error matrix $\boldsymbol{E}_i = \mathcal{M}_i(\boldsymbol{\mathcal{E}})$, alongside other terms that can be easily bounded. Using a strategy akin to, but more intricate than, the one used in Zhou and Chen (2023), we are able to control those $\ell_{2,\infty}$ norms of the error polynomials and, in turn, achieve the desired guarantees.

4.2 Data-driven selection of the thresholds $\{\tau_i\}$

As shown in Theorem 1, HHC can successfully recover the cluster assignment vectors of interest if the tuning parameters $\{\tau_i\}$ satisfy (19b). However, the maximum variance ω_{max}^2 is usually unknown *a priori* and, therefore, we need to carefully choose τ_i. In what follows, we discuss how to select these tuning parameters.

Without loss of generality, we assume $n_1 \leq n_2 \leq n_3$ and let

$$\widehat{\omega} = \sigma_{k_1+1}\big(\mathcal{M}_1(\boldsymbol{\mathcal{Y}})\big)/\sqrt{n_2 n_3}, \tag{22}$$

which can be used to estimate the order of the noise level. Then the threshold τ_i can be chosen as follows:

$$\tau_i = \tau = \overline{C}_\tau \, (n_1 n_2 n_3)^{1/2} \, \widehat{\omega}^2 \log^2 n = \overline{C}_\tau \sqrt{\frac{n_1}{n_2 n_3}} \sigma_{k_1+1}^2\big(\mathcal{M}_1(\boldsymbol{\mathcal{Y}})\big) \log^2 n, \qquad \forall 1 \leq i \leq 3, \tag{23}$$

where $\overline{C}_\tau > 0$ is some sufficiently large constant. The following theorem asserts that the τ_i's computed in (23) satisfy the desired property (19b).

Theorem 2. *Suppose that Assumption 1 holds, and $\min\{n_2/k_2, n_3/k_3\} \geq C\sqrt{\log n}$ holds for some sufficiently large constant $C > 0$. Assume that either of the following conditions is satisfied:*

1. *For all $i \in [n_1]$, there exists some numerical constant $c > 0$ such that*

$$\sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3} \omega_{i,j,\ell}^2 \geq c n_2 n_3 \omega_{\mathsf{max}}^2 \qquad where \qquad \omega_{i,j,\ell}^2 = \mathsf{Var}\,[E_{i,j,\ell}]; \tag{24}$$

2. *The observation model is the stochastic tensor block model (8), with the numbers of clusters satisfying $k_i \lesssim 1$ for each $1 \leq i \leq 3$ and the balance parameter obeying $\beta \asymp 1$.*

Then with probability exceeding $1 - O(n^{-10})$, the thresholds $\{\tau_i\}$ defined in (23) satisfy (19b).

Remark 2. *Here, Condition (24) posits that the average variance for each row of $\mathcal{M}_1(\mathcal{E})$ is on the same order as ω_{max}^2. This condition is met when the noise is not excessively spiky. For example, any noise tensor $\mathcal{E}$ with variances $\omega_{i,j,k}^2 \asymp \omega_{\mathsf{max}}^2$ satisfies this condition.*

The proof of Theorem 2 can be found in Section E. Putting Theorem 1 and Theorem 2 together, we arrive at the following result:

Theorem 3. *Suppose that Assumption 1 holds, $k_i \lesssim 1$ for every $1 \le i \le 3$, and $\beta \asymp 1$. Assume that either the following conditions is satisfied:*

1. *Condition (24) holds;*

2. *The observation model is the stochastic tensor block model (8).*

We further assume that

$$n_1 n_2 n_3 \ge c_1 n^2,$$
$$\mathsf{SNR} \ = \Delta_{\mathsf{min}}/\sigma_{\mathsf{max}} \ge C_1 \sqrt{M} \left(n_1 n_2 n_3\right)^{-1/4} \log n$$

for some large enough constants $C_1, c_1 > 0$. If we choose the tuning parameter τ as in (23) and the numbers of iterations satisfy (20a) and (20b), then with probability exceeding $1 - O(n^{-10})$, HHC achieves exact clustering, i.e., the misclassification rate (cf. (5)) obeys

$$\mathsf{MCR}\big(\widehat{z}_i, z_i^\star\big) = 0, \qquad \forall 1 \le i \le 3.$$

Theorem 3 shows that our data-driven procedure can still achieve the exact clustering under the same signal-to-noise ratio condition as in Theorem 1, provided that the noise condition (24) is satisfied. If the model of interest is the stochastic tensor block model, no extra assumptions on the noise are needed to justify the validity of the data-driven choices of $\{\tau_i\}$.

5 Empirical studies

In this section, we conduct a series of numerical experiments to evaluate the practical effectiveness of the proposed algorithms: HHC, and HHC + HLloyd. Throughout this section, the thresholds are chosen in a data-driven manner as

$$\tau_i \equiv \tau = 1.1 \left(n_1 n_2 n_3\right)^{1/2} \widehat{\omega}^2 = 1.1 \sqrt{\frac{n_1}{n_2 n_3}} \sigma_{k_1+1}^2 \big(\mathcal{M}_1\left(\mathcal{Y}\right)\big), \tag{25}$$

where $\widehat{\omega}$ is defined in (22).

5.1 Experiments on synthetic data

First, we carry out numerical experiments on synthetic data to corroborate the efficacy of HHC and HHC + HLloyd. Following the settings in Han et al. (2022a), we set the dimensions to be $n_1 = n_2 = n_3 = n$ and the numbers of clusters as $k_1 = k_2 = k_3 = k$, and let the cluster sizes be balanced. The following four methods are considered: (1) HSC: the high-order spectral clustering algorithm proposed in Han et al. (2022a); (2) HSC + HLloyd: the procedure that uses HSC to obtain initial clustering results, followed by a 10-iteration high-order Lloyd algorithm (Han et al., 2022a) for refinement; (3) HHC: the method proposed in Algorithm 4 with the numbers of iterations $t_{i,j} = 10$; (4) HHC + HLloyd: the procedure that employs HHC (where $t_{i,j} = 10$) as initial cluster assignment vector estimators and then applies HLloyd with the iteration number $t = 10$ to compute the final clustering results. To evaluate the clustering performance, we calculate, for each method, the empirical clustering error rate (CER), which is one minus the adjusted random index (Milligan and Cooper, 1986). A lower CER indicates a better clustering result. Specifically, an exact recovery of clustering is achieved when CER equals 0. All results are averaged over 100 independent replicates.

(a) $k = 3$, averaged CER

(b) $k = 3$, percentage of exact recovery

(c) $k = 5$, averaged CER

(d) $k = 5$, percentage of exact recovery

Figure 1: Averaged CER and percentage of exact community recovery for HSC, HSC + HLloyd, HHC and HHC + HLloyd under the sub-Gaussian tensor block models with $n = 100$. Here, $\Delta_{\min} = 40n^{-\delta}$.

Sub-Gaussian tensor block models. Let us begin by considering Model (6) with Gaussian noise. We fix the dimensions $n_1 = n_2 = n_3 = n \in \{100, 150\}$, generate a random tensor $\overline{\mathcal{S}} \in \mathcal{R}^{k,k,k}$ with independent entries $\overline{S}_{i_1, i_2, i_3} \sim \mathcal{N}(0, 1)$ and the core tensor $\mathcal{S}^\star$ is obtained by rescaling $\overline{\mathcal{S}}$ such that $\Delta_{\min} = 40n^{-\delta}$ (so that SNR decreases as δ increases). We randomly generate the cluster assignment vectors $\boldsymbol{z}_i \in [k]^n$, $1 \leq i \leq 3$. We generate three vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ such that $\{\alpha_i\}, \{\beta_j\}, \{\gamma_k\}$ are independently and uniformly drawn from $[0, 2]$. The entries of the noise tensor $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{n \times n \times n}$ are generated independently with $E_{i,j,k} \sim \mathcal{N}(0, \alpha_i^2 \beta_j^2 \gamma_k^2)$. For each method, we report the averaged CER and the percentage of exact recovery for cluster assignment vectors. The results for $n = 100$ and $n = 150$ are illustrated in Figures 1 and 2, respectively. As can be seen, HHC and HHC + HLloyd achieve much smaller CER compared with HSC and HSC + HLloyd. In terms of the percentage of exact recovery, HHC + HLloyd achieves the best performance among all these four methods.

Stochastic tensor block models. Next, we study the stochastic block model (8). We choose the core tensor $\mathcal{S}^\star \in \mathbb{R}^{k \times k \times k}$ satisfying

$$S^\star_{i_1, i_2, i_3} = \begin{cases} 10a \cdot n^{-3/2} \left(1 - \frac{i_1 - 1}{2(k-1)}\right), & i_1 = i_2 = i_3, \\ 0.1a \cdot n^{-3/2}, & \text{otherwise.} \end{cases} \tag{26}$$

Here, a is a scalar. When a is not too large, SNR increases with a. The empirical results of the above four methods for $n = 100$ and $n = 150$ are displayed in Figures 3 and 4. From these plots, one sees that HHC and HHC + HLloyd achieve more accurate clustering results, and HHC + HLloyd outperforms all other methods in achieving the highest percentage of the exact recovery for the cluster assignment vectors.

5.2 Real data analysis and real-data-inspired simulation studies

Real data example: the flight route network. We now turn attention to the flight route network data studied in Han et al. (2022a). In adherence to their setup, we also take into account the top 50 airports

(a) $k = 3$, averaged CER

(b) $k = 3$, percentage of exact recovery

(c) $k = 5$, averaged CER

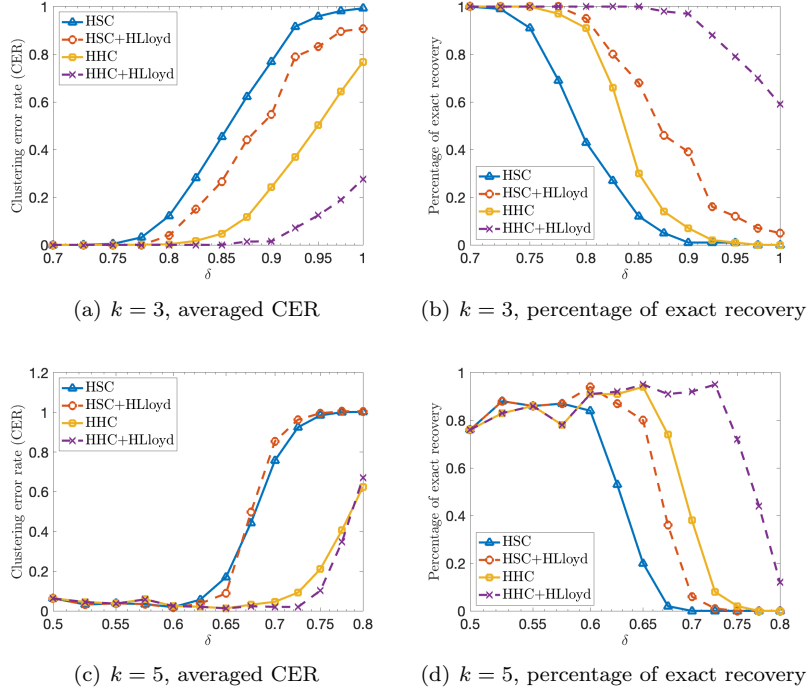(d) $k = 5$, percentage of exact recovery

Figure 2: Averaged CER and percentage of exact community recovery for HSC, HSC + HLloyd, HHC and HHC + HLloyd under the sub-Gaussian tensor block models with $n = 150$. Here, $\Delta_{\min} = 40n^{-\delta}$.

based on the number of flight routes.[2] This results in a $39 \times 50 \times 50$ tensor $\mathcal{Y}$ with binary entries, where the first mode represent airlines, and the remaining two modes represent airports. The entries of the tensor $\mathcal{Y}$ satisfies

$$Y_{i,j,k} = \begin{cases} 1, & \text{if airline } i \text{ operates a flight route from airport } j \text{ to airport } k, \\ 0, & \text{otherwise.} \end{cases} \qquad (27)$$

We select the clustering sizes based on the Bayesian information criterion (BIC) as described in Wang and Zeng (2019); Han et al. (2022a). This criterion suggests the numbers of clusters $(k_1, k_2, k_3) = (5, 5, 5)$. We apply HHC + HLloyd and HSC + HLloyd to the data $\mathcal{Y}$, with results summarized in Tables 1-4.[3] Tables 1 and 3 reveal that HHC + HLloyd produces reasonable clustering results, effectively grouping airlines/airports from China, Europe, and the United States. A comparison of Tables 1 and 2 indicates that HHC + HLloyd outperforms HSC + HLloyd in clustering European and US airlines. For instance, Cluster 2 in both tables shows HHC + HLloyd grouping three US airlines together, whereas HSC + HLloyd includes only two (AA and US); in Cluster 3, HHC + HLloyd groups three European airlines, but HSC + HLloyd gives a mixture of US and European airlines. For airport clustering, our results in Table 3 appear more reasonable than those for HSC + HLloyd in Table 4. Notably, with regards to Cluster 3 in both tables, HHC + HLloyd identifies a cluster of airports from four major European cities along with ATL (a hub). In contrast, HSC + HLloyd groups only CDG (France) and ATL (USA) together.

Real-data-inspired numerical studies. While HHC + HLloyd appears to yield more reasonable real data results, a challenge arises due to the absence of a known ground truth for validation. To draw a more convincing conclusion, we adopt real-data-inspired numerical studies to establish a quantitative comparison between HHC + HLloyd and HSC + HLloyd. Recall that in the real data example, HHC + HLloyd gives

---

[2]The original database at https://openflights.org/data.html#route. Here, we use the processed data provided at https://github.com/Rungang/HLloyd/blob/master/experiment/flight_route.RData.

[3]For each method, we run 100 independent replicates and choose the result that occurs most frequently.

(a) $k = 3$, averaged CER

(b) $k = 3$, percentage of exact recovery

(c) $k = 4$, averaged CER

(d) $k = 4$, percentage of exact recovery

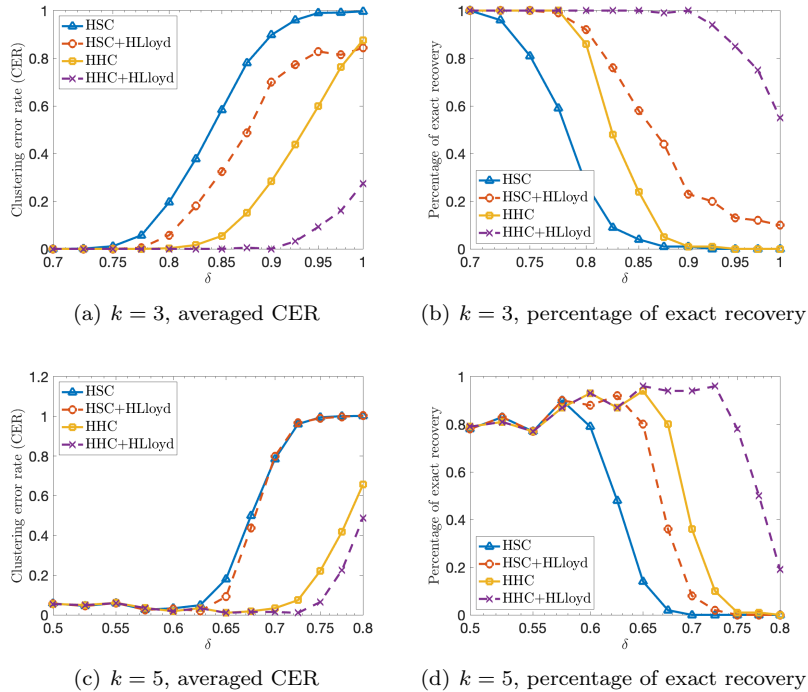Figure 3: Averaged CER and percentage of exact community recovery for HSC, HSC + HLloyd, HHC and HHC + HLloyd under the Stochastic tensor block models with $n = 100$. Here, the quantity a satisfies (26).

	Airlines
Cluster 1	CA, MU, CZ, HU, 3U, ZH (China)
Cluster 2	AA, UA, US (USA)
Cluster 3	AF, AZ, KL (Europe)
Cluster 4	BA, AY, IB (Europe), DL (USA)
Cluster 5	SU, AB, AI, AM, NH, AC, AS, FL, DE, ET, etc. (Mixture)

Table 1: Airline clustering results using HHC + HLloyd.

us the following estimates: the centroid tensor $\widehat{\boldsymbol{\mathcal{S}}}^{\mathsf{HHC+HLloyd}} \in [0,1]^{5 \times 5 \times 5}$ and the cluster assignment vector estimates $\widehat{\boldsymbol{z}}_i^{\mathsf{HHC+HLloyd}}$. In contrast, HSC + HLloyd provides $\widehat{\boldsymbol{\mathcal{S}}}^{\mathsf{HSC+HLloyd}} \in [0,1]^{5 \times 5 \times 5}$ and $\widehat{\boldsymbol{z}}_i^{\mathsf{HSC+HLloyd}}$. We then generate stochastic tensor block models, setting the truth $\boldsymbol{\mathcal{S}}^\star = \widehat{\boldsymbol{\mathcal{S}}}^{\mathsf{HHC+HLloyd}}$ and $\boldsymbol{z}_i^\star = \widehat{\boldsymbol{z}}_i^{\mathsf{HHC+HLloyd}}$ for the first scenario, and $\boldsymbol{\mathcal{S}}^\star = \widehat{\boldsymbol{\mathcal{S}}}^{\mathsf{HSC+HLloyd}}$ and $\boldsymbol{z}_i^\star = \widehat{\boldsymbol{z}}_i^{\mathsf{HSC+HLloyd}})$ for the second. We apply the four methods — HHC, HHC + HLloyd, HSC and HSC + HLloyd — to the generated data. The results are averaged over 100 Monte Carlo runs and are reported in Table 5 and Table 6, respectively. From these results, it becomes evident that under the model with $\boldsymbol{\mathcal{S}}^\star = \widehat{\boldsymbol{\mathcal{S}}}^{\mathsf{HHC+HLloyd}}$ and $\boldsymbol{z}_i^\star = \widehat{\boldsymbol{z}}_i^{\mathsf{HHC+HLloyd}}$, HHC + HLloyd outperforms in terms of estimation error and recovery rate. For the second model, while all four methods demonstrate comparable exact recovery percentages, HHC and HHC + HLloyd have noticeably smaller estimation errors. This means that even for data that best fits HSC + HLloyd, our methods can achieve better clustering performance.

15

(a) $k = 3$, averaged CER

(b) $k = 3$, percentage of exact recovery

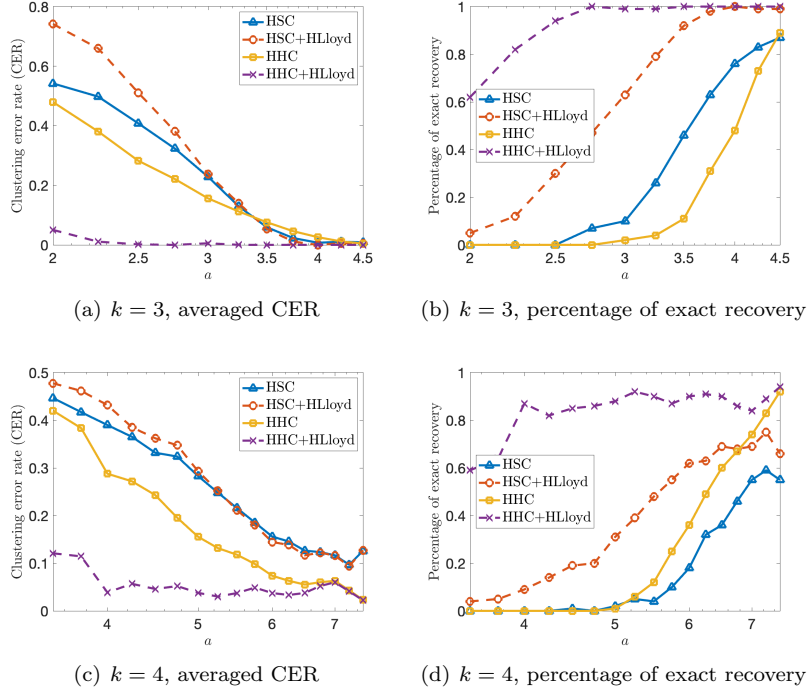(c) $k = 4$, averaged CER

(d) $k = 4$, percentage of exact recovery
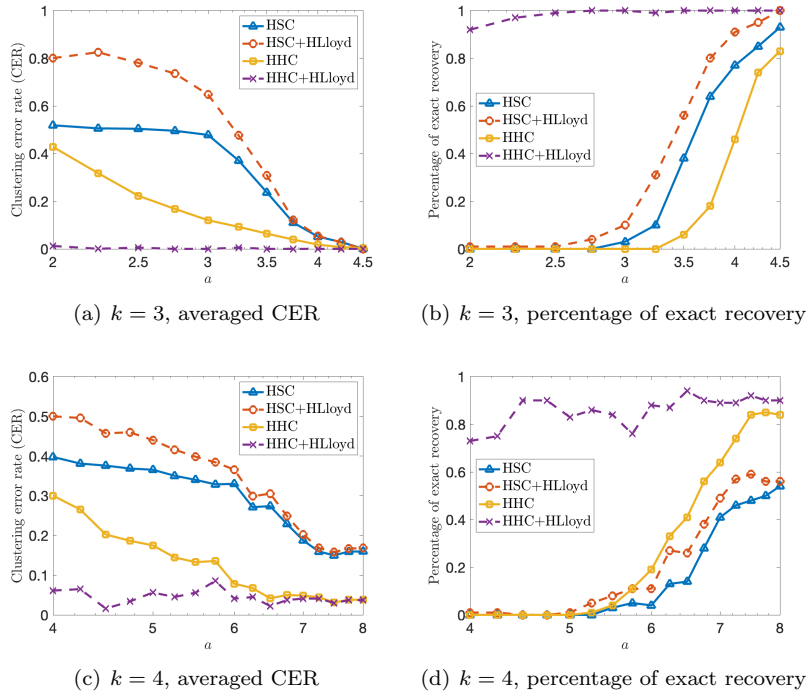
Figure 4: Averaged CER and percentage of exact community recovery for HSC, HSC + HLloyd, HHC and HHC + HLloyd under the Stochastic tensor block models with $n = 150$. Here, the quantity a satisfies (26).

	Airlines
Cluster 1	CA, MU, CZ, HU, 3U, ZH (China)
Cluster 2	AA, US (USA)
Cluster 3	AF, AZ, KL (Europe), DL (USA)
Cluster 4	BA, AY, IB (Europe), UA (USA)
Cluster 5	SU, AB, AI, AM, NH, AC, AS, FL, DE, ET, etc. (Mixture)

Table 2: Airline clustering results for HSC + HLloyd.

6 Related work

The tensor clustering problem considered in the current paper is closely related to several classical clustering problems, which we briefly review here. Among the most commonly studied clustering models are stochastic and censored block models (Holland et al., 1983; Rohe et al., 2011; Mossel et al., 2014; Lei and Rinaldo, 2015; Abbe et al., 2015; Mossel et al., 2015; Hajek et al., 2016a,b; Cai and Li, 2015; Abbe and Sandon, 2015; Chin et al., 2015; Zhang and Zhou, 2016; Florescu and Perkins, 2016; Chen et al., 2016; Guédon and Vershynin, 2016; Abbe, 2017; Gao et al., 2017; Deshpande et al., 2017; Amini and Levina, 2018; Li et al., 2021; Cai et al., 2021), synchronization (Singer, 2011; Javanmard et al., 2016; Bandeira et al., 2017; Chen and Candès, 2018; Zhong and Boumal, 2018; Gao and Zhang, 2021; Li and Wei, 2022; Li et al., 2023; Celentano et al., 2023), and (sub-)Gaussian mixture models (Lu and Zhou, 2016; Cai and Zhang, 2018; Li et al., 2020; Chen and Yang, 2021; Ndaoud, 2022; Abbe et al., 2022; Han et al., 2023). For all these models, spectral clustering algorithms have emerged as a powerful paradigm and have achieved both theoretical and empirical success (Von Luxburg, 2007; Kannan and Vempala, 2009; Abbe, 2017; Chen et al., 2021a). Sharp and intriguing statistical guarantees have recently been derived for spectral clustering (Lei and Rinaldo, 2015; Chin et al., 2015; Abbe et al., 2020; Löffler et al., 2021; Zhang and Zhou, 2022; Zhang, 2023). However, direct applications of these methods and theories to the tensor block model might yield highly sub-optimal signal-to-noise ratio

	Airlines
Cluster 1	BRU, DUS, MUC, MAN, LGW, AMS, BCN, VIE, etc. (Mixture)
Cluster 2	LAX, MIA, DFW, PHL, JFK, ORD, CLT (USA)
Cluster 3	Europe: LHR (London), MAD (Madrid), CDG (Paris), FCO (Rome) USA: ATL (Atlanta)
Cluster 4	PEK, CAN, XIY, KMG, HGH, CKG, CTU, PVG (China)
Cluster 5	PHX, SFO, EWR, IAH, DEN, LAS (USA) YYZ (Canada), FRA (Germany), MEX (Mexico)

Table 3: Airport clustering results using HHC + HLloyd.

	Airlines
Cluster 1	BRU, MUC, LGW, AMS, BCN, VIE, ZRH, DXB, etc. (Mixture)
Cluster 2	LHR (UK), MIA, DFW, PHL, JFK, ORD, CLT (USA)
Cluster 3	CDG (France), ATL (USA)
Cluster 4	PEK, CAN, XIY, KMG, HGH, CKG, CTU, PVG (China)
Cluster 5	YYZ, FRA, DUS, MAN, MAD, FCO (Europe), MEX (Mexico) PHX, SFO, LAX, EWR, IAH, DEN, LAS (USA)

Table 4: Airport clustering results for HSC + HLloyd.

conditions, while in the meantime introducing unnecessary assumptions on the condition number.

Turning to the tensor block model, Wang and Zeng (2019) investigated the theoretical properties for the MLE estimator, which is, however, computationally intractable. Chi et al. (2020) considered a convex procedure and derived its theoretical guarantees concerning the tensor estimation error. Under the i.i.d. noise setting, Han et al. (2022a) proved the existence of a statistical-computational gap for this problem and proposed a polynomial-time algorithm that can achieve exact clustering if the signal-to-noise ratio exceeds the computational limit (ignoring logarithmic factors). However, in the presence of heteroskedastic noise, these methods fall short of statistical efficiency. Going beyond this model, Hu and Wang (2023) considered degree-corrected tensor block models and Agterberg and Zhang (2022) studied mixed-membership tensor block models. However, the methods and theoretical results in these two papers either lean on i.i.d. noise assumptions or require the underlying tensor to be well-conditioned, both of which can be relaxed using our approach. In addition to the model considered in this paper, a couple of other tensor clustering problems have been proposed and studied in the literature; see, e.g., Jegelka et al. (2009); Sun and Li (2019); Wu et al. (2019); Lyu and Xia (2022); Mai et al. (2022).

Our work is also closely related to the tensor PCA models (Richard and Montanari, 2014; Hopkins et al., 2015; Anandkumar et al., 2017; Zhang and Xia, 2018; Arous et al., 2019; Han et al., 2022b; Cai et al., 2022a,b; Xia et al., 2022; Zhou et al., 2022), which aim to estimate the true tensor or the associated subspaces based on noisy observations. To accomplish this task, a commonly used strategy is to apply spectral methods (Chen et al., 2021a) to obtain initial subspace estimates, followed by further refinement steps (De Lathauwer et al., 2000; Zhang and Xia, 2018; Han et al., 2022b; Tong et al., 2022; Cai et al., 2022a). Some popular initialization methods include the vanilla SVD-based approach (Cai and Zhang, 2018; Zhang and Xia, 2018), diagonal-deleted/reweighted PCA (Lounici, 2014; Florescu and Perkins, 2016; Montanari and Sun, 2018; Cai et al., 2021, 2022a) and HeteroPCA (Zhang et al., 2022; Yan et al., 2021; Han et al., 2022b). However, in contrast to the tensor PCA models, the tensor block models studied in the present paper do not impose any assumptions on the least singular value or on the singular gaps of the true tensor. Therefore, directly applying these tensor PCA methods may not yield subspace estimates with the desired statistical accuracy.

Recently, it has been shown that sharp $\ell_{2,\infty}$ or ℓ_∞ guarantees for singular subspaces play a pivotal role for proving that spectral clustering (with or without the help of k-means) can achieve exact recovery or optimal mis-clustering rates for many clustering problems (Abbe et al., 2020; Cai et al., 2021; Abbe et al., 2022; Zhang, 2023). To derive such subspace estimation guarantees, a powerful and perhaps the most popular tool is the leave-one-out analysis (Zhong and Boumal, 2018; Ma et al., 2020; Chen et al., 2019a; Abbe et al.,

	error mean	standard deviation	recovery rate
HSC	0.0225	0.0269	0.36
HSC + HLloyd	0.0129	0.0275	0.69
HHC	0.0181	0.0472	0.6
HHC + HLloyd	0.0115	0.0453	0.83

Table 5: Real-data-inspired numerical experiments: $\boldsymbol{\mathcal{S}}^{\star} = \widehat{\boldsymbol{\mathcal{S}}}^{\mathsf{HHC+HLloyd}}$ and $\boldsymbol{z}_i^{\star} = \widehat{\boldsymbol{z}}_i^{\mathsf{HHC+HLloyd}}$.

	error mean	standard deviation	recovery rate
HSC	0.0273	0.0942	0.89
HSC + HLloyd	0.0311	0.0959	0.84
HHC	0.0120	0.0386	0.85
HHC + HLloyd	0.0124	0.0419	0.88

Table 6: Real-data-inspired nemerical experiments: $\boldsymbol{\mathcal{S}}^{\star} = \widehat{\boldsymbol{\mathcal{S}}}^{\mathsf{HSC+HLloyd}}$ and $\boldsymbol{z}_i^{\star} = \widehat{\boldsymbol{z}}_i^{\mathsf{HSC+HLloyd}}$.

2020; Lei, 2019; Chen et al., 2020, 2019b, 2021b; Cai et al., 2021; Chen et al., 2023; Cai et al., 2022a; Abbe et al., 2022; Yan et al., 2021; Ling, 2022; Ke and Wang, 2022; Zhang and Zhou, 2022; Yang and Ma, 2022). However, the results obtained using the leave-one-out analysis are often sub-optimal with respect to the condition number of the truth. This can lead to unsatisfactory results under the tensor block models, especially since there is no assumption made on the singular value of the underlying tensor $\boldsymbol{\mathcal{S}}^{\star}$.

7 Discussion

In this paper, we have studied the tensor clustering problem in the presence of heteroskedastic noise. To better deal with heteroskedastic noise and improve statistical performance, we have proposed a novel method called High-order HeteroClustering (HHC), which first employs Thresholded Deflated-HeteroPCA to obtain subspace estimates and then applies approximate k-means for clustering. The proposed method provably achieves exact clustering for a wide range of signal-to-noise ratio conditions that are essentially unimprovable among polynomial-time algorithms. Empirically, we have evaluated the numerical performance of HHC, and HHC followed by the high-order Lloyd algorithm (HLloyd, Han et al. (2022a)), on both synthetic and real data. Both of these two methods achieve low empirical mis-clustering rates, with HHC + HLloyd outperforming other existing methods proposed in the literature.

Moving beyond, there are numerous future directions that are worth investigating. For instance, thus far our theory has focused primarily on exact clustering; it remains unclear whether our algorithm can achieve optimal mis-classification rates when only partial recovery is feasible. In addition, our signal-to-noise ratio condition might be sub-optimal if the clusters are highly-unbalanced, a scenario where the balance parameter β is exceedingly small. It would be interesting to investigate the plausibility of further improvement under such imbalanced settings. Furthermore, it would be worthwhile to explore the feasibility of extending our paradigm to tackle the tensor mixed-membership block model (Agterberg and Zhang, 2022), with the aim of achieving optimal statistical performance without being affected by the condition number of the true tensor.

Acknowledgements

A Procedure of High-order Lloyd Algorithm (HLloyd)

This section provides a formal description of the procedure of High-order Lloyd Algorithm (HLloyd) proposed by (Han et al., 2022a); see Algorithm 5.

Algorithm 5: High-order Lloyd Algorithm (HLloyd) (Han et al., 2022a)

---

1 input: observed tensor $\boldsymbol{\mathcal{Y}}$, numbers of clusters k_1, k_2, k_3, initial cluster assignment vector estimates $\{\widehat{\boldsymbol{z}}_\ell^{(0)}\}_{1 \leq \ell \leq 3}$, number of iterations T.

2 for $t = 0, \ldots, T-1$ **do**

3   **block mean update:** calculate $\widehat{\boldsymbol{\mathcal{S}}}^{(t)} \in \mathbb{R}^{k_1 \times k_2 \times k_3}$ such that

$$\widehat{\boldsymbol{S}}_{i_1,i_2,i_3}^{(t)} = \text{Average}\left(\left\{\boldsymbol{\mathcal{Y}}_{j_1,j_2,j_3} : \widehat{z}_{\ell,j_\ell}^{(t)} = i_\ell, \forall \ell \in [3]\right\}\right), \quad \forall i_\ell \in [k_\ell], \ell \in [3].$$

4   calculate $\widehat{\boldsymbol{\mathcal{B}}}_1^{(t)} \in \mathbb{R}^{n_1 \times k_2 \times k_3}, \widehat{\boldsymbol{\mathcal{B}}}_2^{(t)} \in \mathbb{R}^{k_1 \times n_2 \times k_3}, \widehat{\boldsymbol{\mathcal{B}}}_3^{(t)} \in \mathbb{R}^{k_1 \times k_2 \times n_3}$ such that

$$\left(\widehat{\boldsymbol{\mathcal{B}}}_1^{(t)}\right)_{j_1,i_2,i_3} = \text{Average}\left(\left\{\boldsymbol{\mathcal{Y}}_{j_1,j_2,j_3} : \widehat{z}_{\ell,j_\ell}^{(t)} = i_\ell, \ell = 2,3\right\}\right), \quad \forall j_1 \in [n_1], i_2 \in [k_2], i_3 \in [k_3],$$

$$\left(\widehat{\boldsymbol{\mathcal{B}}}_2^{(t)}\right)_{i_1,j_2,i_3} = \text{Average}\left(\left\{\boldsymbol{\mathcal{Y}}_{j_1,j_2,j_3} : \widehat{z}_{\ell,j_\ell}^{(t)} = i_\ell, \ell = 1,3\right\}\right), \quad \forall i_1 \in [k_1], j_2 \in [n_2], i_3 \in [k_3],$$

$$\left(\widehat{\boldsymbol{\mathcal{B}}}_3^{(t)}\right)_{i_1,i_2,j_3} = \text{Average}\left(\left\{\boldsymbol{\mathcal{Y}}_{j_1,j_2,j_3} : \widehat{z}_{\ell,j_\ell}^{(t)} = i_\ell, \ell = 1,2\right\}\right), \quad \forall i_1 \in [k_1], i_2 \in [k_2], j_3 \in [n_3].$$

   cluster update: calculate cluster assignment vector estimates $\{\widehat{\boldsymbol{z}}_i^{(t+1)}\}_{i \in [3]}$:

$$\widehat{z}_{i,j}^{(t+1)} \in \underset{\ell \in [k_i]}{\arg\min} \left\|\left(\mathcal{M}_i(\widehat{\boldsymbol{\mathcal{B}}}_i^{(t)})\right)_{j,:} - \left(\mathcal{M}_i(\widehat{\boldsymbol{\mathcal{S}}}^{(t)})\right)_{\ell,:}\right\|_2, \quad \forall i \in [3], j \in [n_i]. \tag{28}$$

5 output: cluster assignment vector estimates $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{S}}^{(T-1)}, \widehat{\boldsymbol{z}}_1 = \widehat{\boldsymbol{z}}_1^{(T)}, \widehat{\boldsymbol{z}}_2 = \widehat{\boldsymbol{z}}_2^{(T)}, \widehat{\boldsymbol{z}}_3 = \widehat{\boldsymbol{z}}_2^{(T)}$.

B Proof of Theorem 1

In this section, we present the proof of our main result Theorem 1, by establishing a more general version as follows. Here and throughout, we define $k_{-i} = k_1 k_2 k_3 / k_i$ and $n_{-i} = n_1 n_2 n_3 / n_i$ for $i \in [3]$.

Theorem 4. *Suppose that Assumption (1) holds, and assume that for all $1 \leq i \leq 3$,*

$$n_1 n_2 n_3 \gtrsim k^4 n^2, \tag{29a}$$

$$n_i \geq \frac{c_1 k^4}{\beta^2}, \tag{29b}$$

$$k_i \gtrsim k_{-i}, \tag{29c}$$

$$c_\tau k_i^2 (n_1 n_2 n_3)^{1/2} \log^2 n \leq \tau_i / \omega_{\text{max}}^2 \leq C_\tau k_i^2 (n_1 n_2 n_3)^{1/2} \log^2 n, \tag{29d}$$

$$\frac{\Delta_{\text{min}}}{\omega_{\text{max}}} \geq C_1 \sqrt{M} \left(\frac{k^{9/2}}{\beta^{5/2}} (n_1 n_2 n_3)^{-1/4} \log n + \frac{k^9}{\beta^5} (n_1 n_2 n_3/n)^{-1/2} \sqrt{\log n}\right), \tag{29e}$$

where C_1, c_1, C_τ and c_τ are some large enough constants. If we choose the numbers of iterations to obey

$$t_{i,j} \geq \log\left(C \frac{k^3}{\beta^3} \frac{\sigma_{i,r_{i,j-1}+1}^{\star 2}}{\sigma_{i,r_{i,j}+1}^{\star 2}}\right), \quad 1 \leq j \leq j_{\text{max}}^i - 1 \tag{30a}$$

$$t_{i,j_{\max}^i} \geq \log \left(C n^3 \frac{\sigma_{i,r_{i,j_{\max}^i-1}+1}^{\star 2}}{\omega_{\max}^2} \right) \tag{30b}$$

for all $1 \leq i \leq 3$, then with probability exceeding $1 - O(n^{-10})$, the misclassification rate (cf. (5)) of the outputs $\{\widehat{z}_i\}$ returned by Algorithm 4 satisfy

$$\mathsf{MCR}(\widehat{z}_i, z_i^\star) = 0, \qquad \forall 1 \leq i \leq 3.$$

The rest of this section is dedicated to proving Theorem 4.

B.1   Several key results under the matrix setting

To begin with, we first consider the following model: suppose we observe

$$\boldsymbol{Y} = \boldsymbol{X}^\star + \boldsymbol{E} \in \mathbb{R}^{m_1 \times m_2}, \tag{31}$$

where the noise matrix $\boldsymbol{E}$ has independent and zero-mean entries, and $\boldsymbol{X}^\star$ is a matrix with rank not exceeding r and admits the following SVD decomposition:

$$\boldsymbol{X}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{V}^{\star\top} = \sum_{i=1}^r \sigma_i^\star \boldsymbol{u}_i^\star \boldsymbol{v}_i^{\star\top}, \tag{32}$$

where $\sigma_1^\star \geq \cdots \geq \sigma_r^\star \geq 0$ are the singular values of $\boldsymbol{X}$, $\boldsymbol{U}^\star = [\boldsymbol{u}_1^\star, \ldots, \boldsymbol{u}_r^\star] \in \mathcal{O}^{m_1,r}$ (resp. $\boldsymbol{V} = [\boldsymbol{v}_1^\star, \ldots, \boldsymbol{v}_r^\star] \in \mathcal{O}^{m_2,r}$) is the column (resp. row) subspace of $\boldsymbol{X}^\star$, and $\boldsymbol{\Sigma}^\star = \mathsf{diag}(\sigma_1^\star, \ldots, \sigma_r^\star)$. In addition, we define the incoherence parameter

$$\text{(Incoherence parameter)} \qquad \mu = \mu(\boldsymbol{X}^\star) := \max \left\{ \frac{m_1}{r} \max_{i \in [m_1]} \left\| \boldsymbol{U}_{i,:}^\star \right\|_2^2, \frac{m_2}{r} \max_{j \in [m_2]} \left\| \boldsymbol{V}_{j,:}^\star \right\|_2^2 \right\}. \tag{33}$$

For notational convenience, we also define

$$m := \max\{m_1, m_2\} \qquad \text{and} \qquad \sigma_{r+1}^\star = 0. \tag{34}$$

Furthermore, we impose the noise assumption on the noise matrix $\boldsymbol{E}$:

Assumption 2. *Suppose that the following conditions on the noise matrix $\boldsymbol{E}$ hold:*

1. *The $E_{i,j}$'s, the entries of $\boldsymbol{E}$, are independently generated and satisfy $\mathbb{E}[E_{i,j}] = 0$;*

2. *$\mathbb{P}(|E_{i,j}| > B) \leq m^{-12}$, where B is some quantity satisfying*

$$B \leq C_{\mathsf{b}} \omega_{\max} \frac{\min\left\{ (m_1 m_2)^{1/4}, \sqrt{m_2} \right\}}{\log m}.$$

One can immediately find that $\mathcal{M}_i(\boldsymbol{\mathcal{E}})$ obeys the conditions in Assumption 2 with dimension $m_1 = n_i$ and $m_2 = n_{-i}$. Moreover, we define

$$\boldsymbol{M} = (\boldsymbol{U}^\star \boldsymbol{\Sigma}^\star + \boldsymbol{E} \boldsymbol{V}^\star)(\boldsymbol{U}^\star \boldsymbol{\Sigma}^\star + \boldsymbol{E} \boldsymbol{V}^\star)^\top \tag{35}$$

and

$$\boldsymbol{M}^{\mathsf{oracle}} = \boldsymbol{M} + \mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{E} \boldsymbol{E}^\top - \boldsymbol{E} \boldsymbol{V}^\star \boldsymbol{V}^{\star\top} \boldsymbol{E}^\top \right) = \mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{Y} \boldsymbol{Y}^\top \right) + \mathcal{P}_{\mathsf{diag}}(\boldsymbol{M}). \tag{36}$$

Let $\boldsymbol{U}^{\mathsf{oracle}} \in \mathcal{O}^{m_1,r}$ denote the leading-r eigenvector of $\boldsymbol{M}^{\mathsf{oracle}}$. The following theorem shows that $\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}}$ and $\boldsymbol{U}_{:,1:r'}^\star$ are reasonably close if there is a sufficiently large gap between $\sigma_{r'}^\star$ and $\sigma_{r'+1}^\star$; the proof is deferred to Section C.

Theorem 5. *Suppose that $r \geq 2$, Assumption 2 holds and*

$$\sigma_1^\star / \omega_{\mathsf{max}} \geq 2 C_0 r \big[(m_1 m_2)^{1/4} + r m_1^{1/2} \big] \log m \tag{37a}$$

$$\mu \leq c_0 \frac{m_1}{r^3} \tag{37b}$$

hold for some sufficiently large (resp. small) constant $C_0 > 0$ (resp. $c_0 > 0$).
  (a) The set defined below

$$\mathcal{A} = \left\{ j : 1 \leq j \leq r, \sigma_j^\star \geq \frac{4r}{4r-1} \sigma_{j+1}^\star \vee C_0 r \big[(m_1 m_2)^{1/4} + r m_1^{1/2} \big] \omega_{\mathsf{max}} \log m \right\} \tag{38}$$

is non-empty.
  (b) With probability exceeding $1 - O(n^{-10})$, for all $r' \in \mathcal{A}$, we have

$$\left\| \boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}} \boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{:,1:r'} \widetilde{\boldsymbol{U}}_{:,1:r'}^\top \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{r^2 \sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma_{r'}^\star} + \frac{r^2 \sqrt{m_1 m_2} \omega_{\mathsf{max}}^2 \log^2 m}{\sigma_{r'}^{\star 2}} \right), \tag{39a}$$

$$\left\| \boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}} \boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top} - \boldsymbol{U}_{:,1:r'}^\star \boldsymbol{U}_{:,1:r'}^{\star\top} \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{r^2 \sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma_{r'}^\star} + \frac{r^2 \sqrt{m_1 m_2} \omega_{\mathsf{max}}^2 \log^2 m}{\sigma_{r'}^{\star 2}} \right). \tag{39b}$$

Here, $\widetilde{\boldsymbol{U}}_{:,1:r'}$ is the leading rank-r' left singular space of $\boldsymbol{U}_{:,1:\overline{r}}^\star \boldsymbol{\Sigma}_{1:\overline{r},1:\overline{r}}^\star + \boldsymbol{E} \boldsymbol{V}_{:,1:\overline{r}}^\star$ with $\overline{r} = \max \mathcal{A}$.

With the aid of Theorem 5, we are able to develop an upper bound on $\left\| (\boldsymbol{I}_{m_1} - \boldsymbol{U} \boldsymbol{U}^\top) \boldsymbol{X}^\star \right\|_{2,\infty}$ if the threshold τ is properly chosen, as asserted by the following theorem.

Theorem 6. *Suppose that $r \geq 2$, Assumption 2 holds, and*

$$c_\tau r^2 \big[(m_1 m_2)^{1/2} + r^2 m_1 \big] \log^2 m \leq \tau / \omega_{\mathsf{max}}^2 \leq C_\tau r^2 \big[(m_1 m_2)^{1/2} + r^2 m_1 \big] \log^2 m \tag{40a}$$

$$\sigma_1^\star / \omega_{\mathsf{max}} \geq C_1 r \big[(m_1 m_2)^{1/4} + r m_1^{1/2} \big] \log m \tag{40b}$$

$$\mu \leq c_1 \frac{m_1}{r^3} \tag{40c}$$

hold for some sufficiently large (resp. small) constant $C_1, C_\tau, c_\tau > 0$ satisfying $C_1^2 / 2 > C_\tau > c_\tau$ (resp. $c_1 > 0$). If the numbers of iterations obey

$$t_k > \log \left(C \frac{\sigma_{r_{k-1}}^{\star 2}}{\sigma_{r_k}^{\star 2}} \right), \qquad 1 \leq k < k_{\mathsf{max}} \tag{41a}$$

$$t_{k_{\mathsf{max}}} > \log \left(C \frac{\sigma_{r_{k_{\mathsf{max}}-1}+1}^{\star 2}}{\omega_{\mathsf{max}}^2} \right) \tag{41b}$$

for some sufficiently large constants $C > 0$, then with probability exceeding $1 - O(m^{-10})$, the output of Algorithm 1 satisfies

$$\left\| \boldsymbol{U} \boldsymbol{U}^\top - \boldsymbol{U}_{:,1:r_{k_{\mathsf{max}}}}^\star \boldsymbol{U}_{:,1:r_{k_{\mathsf{max}}}}^{\star\top} \right\| \lesssim \sqrt{\frac{\mu r^3}{m_1}}, \tag{42a}$$

$$\left\| \left(\boldsymbol{U} \boldsymbol{U}^\top - \boldsymbol{U}_{:,1:r_{k_{\mathsf{max}}}}^\star \boldsymbol{U}_{:,1:r_{k_{\mathsf{max}}}}^{\star\top} \right) \boldsymbol{X}^\star \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(r^2 \sqrt{m_1} \omega_{\mathsf{max}} \log m + r (m_1 m_2)^{1/4} \omega_{\mathsf{max}} \log m \right), \tag{42b}$$

$$\left\| \left(\boldsymbol{I}_{m_1} - \boldsymbol{U} \boldsymbol{U}^\top \right) \boldsymbol{X}^\star \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(r^2 \sqrt{m_1} \omega_{\mathsf{max}} \log m + r (m_1 m_2)^{1/4} \omega_{\mathsf{max}} \log m \right). \tag{42c}$$

Here, $r_0 = 0.r_1, \ldots, r_{k_{\mathsf{max}}}$ are the ranks selected in Algorithm 1 and k_{max} satisfies $r_{k_{\mathsf{max}}} = r$ or $\sigma_{r_{k_{\mathsf{max}}}+1}(\boldsymbol{G}_{k_{\mathsf{max}}}) \leq \tau$.

The proof of Theorem 6 can be found in Section D. With Theorem 6 in hand, we are now positioned to prove Theorem 4. The proof consists of four steps, to be detailed next.

B.2 Main steps for proving Theorem 4

Step 1: verifying (40a) **-** (40c). To apply Theorem 6, one needs to verify the conditions (40a) - (40c) for $\mathcal{M}_1(\mathcal{Y}) = \mathcal{M}_1(\mathcal{X}) + \mathcal{M}_i(\mathcal{E})$ with dimensions $m_1 = n_i$, $m_2 = n_{-i}$ and the rank $r = k_i$.

Step 1.1: verifying (40a). Noting that $n_1 n_2 n_3 \geq k^4 n^2$, we have $(n_1 n_2 n_3)^{1/2} + k_i^2 n_i \asymp (n_1 n_2 n_3)^{1/2}$. Then we know from (29d) that (40a) is valid.

Step 1.2: verifying (40c). For notational convenience, we let

$$\overline{M}_i^\star = M_i^\star \left(M_i^{\star\top} M_i^\star \right)^{-1/2} \in \mathcal{O}^{n_i, k_i}, \qquad \forall i \in [3]$$

and

$$\overline{\mathcal{S}}^\star = \mathcal{S}^\star \times_1 \left(M_1^{\star\top} M_1^\star \right)^{1/2} \times_2 \left(M_2^{\star\top} M_2^\star \right)^{1/2} \times_3 \left(M_3^{\star\top} M_3^\star \right)^{1/2}.$$

In addition, for any $U \in \mathcal{O}^{n,r}$, we define the projection matrix

$$\mathcal{P}_U = UU^\top. \tag{43}$$

Recognizing that

$$\mathcal{M}_i(\mathcal{X}^\star) = M_i^\star \mathcal{M}_i\left(\mathcal{S}^\star\right) \left(M_{i+2}^\star \otimes M_{i+1}^\star \right)^\top = \overline{M}_i^\star \mathcal{M}_i(\overline{\mathcal{S}}^\star)\left(\overline{M}_{i+2}^\star \otimes \overline{M}_{i+1}^\star \right)^\top,$$

We let $U_{X_i^\star} \Sigma_{X_i^\star} V_{X_i^\star}^\top$ denote the SVD of $X_i^\star := \mathcal{M}_i(\mathcal{X}^\star)$, where $\Sigma_{X_i^\star} = \mathsf{diag}(\sigma_1(X_i^\star), \ldots, \sigma_{k_i}(X_i^\star))$ is a diagonal matrix containing all singular values of $X_i^\star$ (in decreasing order), and $U_{X_i^\star} \in \mathcal{O}^{n_i, k_i}$ (resp. $V_{X_i^\star} \in \mathcal{O}^{n_{-i}, k_i}$) denotes the left (resp. right) singular subspace of $X_i^\star$ and satisfies

$$U_{X_i^\star} = \overline{M}_i^\star A_i \qquad (\text{resp. } V_{X_i^\star} = \left(\overline{M}_{i+2}^\star \otimes \overline{M}_{i+1}^\star \right) B_i) \tag{44}$$

for some $A_i \in \mathcal{O}^{k_i, k_i}$ and $B \in \mathcal{O}^{k_{-i}, k_i}$. Then it follows immediately that

$$\left\| U_{X_i^\star} \right\|_{2,\infty} \leq \left\| \overline{M}_i^\star \right\|_{2,\infty} \leq \left\| M_i^\star \right\|_{2,\infty} \left\| \sigma_{k_i}\left(M_i^{\star\top} M_i^\star \right) \right\|^{-1/2} \leq 1 \cdot \sqrt{\frac{k_i}{\beta n_i}}. \tag{45}$$

Here, the second inequality comes from the fact that $M_i^{\star\top} M_i^\star$ is a diagonal matrix and its diagonal entries

$$\left(M_i^{\star\top} M_i^\star \right)_{\ell,\ell} = \left| \left\{ j \in [n_i] : z_{i,j}^\star = \ell \right\} \right| \geq \beta n_i / k_i. \tag{46}$$

Similarly, one can bound $\|V_{X_i^\star}\|_{2,\infty}$ as follows:

$$\left\| V_{X_i^\star} \right\|_{2,\infty} \leq \left\| \overline{M}_{i+1}^\star \right\|_{2,\infty} \left\| \overline{M}_{i+2}^\star \right\|_{2,\infty} \leq \sqrt{\frac{k_{i+1}}{\beta n_{i+1}}} \sqrt{\frac{k_{i+2}}{\beta n_{i+2}}} = \sqrt{\frac{\left(\frac{k_{-i}}{\beta^2 k_i} \right) k_i}{n_{-i}}}. \tag{47}$$

We then arrive at

$$\mu_i = \max\left\{ \frac{1}{\beta}, \frac{k_{-i}}{\beta^2 k_i} \right\} \overset{(29\mathrm{b})}{\leq} c_1 \frac{n_i}{k_i^3}. \tag{48}$$

Step 1.3: verifying (40b). Next, let us validate Condition (40b). Note that for any $1 \leq j_1 \neq j_2 \leq k_i$,

$$\Delta_i^2 \leq \left\| \mathcal{M}_i\left(\mathcal{S}^\star\right)_{j_1,:} - \mathcal{M}_i\left(\mathcal{S}^\star\right)_{j_2,:} \right\|_2^2$$
$$= \left\| (e_{j_1} - e_{j_2})^\top \mathcal{M}_i\left(\mathcal{S}^\star\right) \right\|_2^2$$
$$\leq \left\| \mathcal{M}_i\left(\mathcal{S}^\star\right) \right\|^2 \left\| e_{j_1} - e_{j_2} \right\|_2^2$$

22

$$= 2 \left\| \mathcal{M}_i \left(\mathcal{S}^\star \right) \right\|^2, \tag{49}$$

where e_j is the j-th canonical basis of $\mathbb{R}^{k_i}$. Furthermore, recognizing that $\boldsymbol{M}_i^{\star\top} \boldsymbol{M}_i^\star$ is a diagonal matrix with diagonal entries satisfying (46), we know that

$$\sigma_{k_i} \left(\boldsymbol{M}_i \right) \geq \sqrt{\frac{\beta n_i}{k_i}}, \tag{50}$$

and consequently

$$\left\| \boldsymbol{X}_i^\star \right\| = \left\| \boldsymbol{M}_i^\star \mathcal{M}_i \left(\mathcal{S}^\star \right) \left(\boldsymbol{M}_{i+2}^\star \otimes \boldsymbol{M}_{i+1}^\star \right)^\top \right\| \geq \left\| \mathcal{M}_i \left(\mathcal{S}^\star \right) \right\| \prod_{i=1}^3 \sigma_{k_i} \left(\boldsymbol{M}_i^\star \right) \geq \sqrt{\frac{\beta^3 n_1 n_2 n_3}{k_1 k_2 k_3}} \left\| \mathcal{M}_i \left(\mathcal{S}^\star \right) \right\|. \tag{51}$$

Combining (49), (51) and the assumption $\Delta_i / \sigma_{\mathsf{max}} \gg k^{5/2} \left(n_1 n_2 n_3 \right)^{-1/4} \log m / \beta^{3/2}$, one has

$$\left\| \boldsymbol{X}_i^\star \right\| \geq \sqrt{\frac{\beta^3 n_1 n_2 n_3}{2 k_1 k_2 k_3}} \Delta_i \gg k_i \left(n_1 n_2 n_3 \right)^{1/4} \log n \asymp k_i \left[\left(n_1 n_2 n_3 \right)^{1/4} + k_i n_i^{1/2} \right] \log n, \tag{52}$$

and thus condition (40b) holds. Here, the last inequality holds since $n_1 n_2 n_3 \geq k^4 n^2$.

Step 2: bounding $\left\| \mathcal{Y} \times_1 \mathcal{P}_{\boldsymbol{U}_1} \times_2 \mathcal{P}_{\boldsymbol{U}_2} \times_3 \mathcal{P}_{\boldsymbol{U}_3} - \mathcal{X}^\star \right\|_{\mathrm{F}}^2$. We define

$$\widehat{\mathcal{X}} = \mathcal{Y} \times_1 \mathcal{P}_{\boldsymbol{U}_1} \times_2 \mathcal{P}_{\boldsymbol{U}_2} \times_3 \mathcal{P}_{\boldsymbol{U}_3}.$$

Then the triangle inequality and the fact $\left\| \mathcal{P}_{\boldsymbol{U}_i} \right\| = 1$ lead to the following upper bound:

$$
\begin{aligned}
\left\| \widehat{\mathcal{X}} - \mathcal{X}^\star \right\|_{\mathrm{F}}^2 &= \left\| \mathcal{Y} \times_1 \mathcal{P}_{\boldsymbol{U}_1} \times_2 \mathcal{P}_{\boldsymbol{U}_2} \times_3 \mathcal{P}_{\boldsymbol{U}_3} - \mathcal{X}^\star \right\|_{\mathrm{F}}^2 \\
&\leq 2 \left(\left\| \mathcal{X}^\star \times_1 \mathcal{P}_{\boldsymbol{U}_1} \times_2 \mathcal{P}_{\boldsymbol{U}_2} \times_3 \mathcal{P}_{\boldsymbol{U}_3} - \mathcal{X}^\star \right\|_{\mathrm{F}}^2 + \left\| \mathcal{E} \times_1 \mathcal{P}_{\boldsymbol{U}_1} \times_2 \mathcal{P}_{\boldsymbol{U}_2} \times_3 \mathcal{P}_{\boldsymbol{U}_3} \right\|_{\mathrm{F}}^2 \right) \\
&\leq 6 \big(\left\| \mathcal{X}^\star \times_1 \left(\boldsymbol{I}_{n_1} - \mathcal{P}_{\boldsymbol{U}_1} \right) \times_2 \mathcal{P}_{\boldsymbol{U}_2} \times_3 \mathcal{P}_{\boldsymbol{U}_3} \right\|_{\mathrm{F}}^2 + \left\| \mathcal{X}^\star \times_2 \left(\boldsymbol{I}_{n_2} - \mathcal{P}_{\boldsymbol{U}_2} \right) \times_3 \mathcal{P}_{\boldsymbol{U}_3} \right\|_{\mathrm{F}}^2 \\
&\quad + \left\| \mathcal{X}^\star \times_3 \left(\boldsymbol{I}_{n_3} - \mathcal{P}_{\boldsymbol{U}_3} \right) \right\|_{\mathrm{F}}^2 \big) + 2 \left\| \mathcal{E} \times_1 \mathcal{P}_{\boldsymbol{U}_1} \times_2 \mathcal{P}_{\boldsymbol{U}_2} \times_3 \mathcal{P}_{\boldsymbol{U}_3} \right\|_{\mathrm{F}}^2 \\
&\leq 6 \left(\left\| \left(\boldsymbol{I}_{n_1} - \boldsymbol{U}_1 \boldsymbol{U}_1^\top \right) \boldsymbol{X}_1^\star \right\|_{\mathrm{F}}^2 + \left\| \left(\boldsymbol{I}_{n_2} - \boldsymbol{U}_2 \boldsymbol{U}_2^\top \right) \boldsymbol{X}_2^\star \right\|_{\mathrm{F}}^2 + \left\| \left(\boldsymbol{I}_{n_3} - \boldsymbol{U}_3 \boldsymbol{U}_3^\top \right) \boldsymbol{X}_3^\star \right\|_{\mathrm{F}}^2 \right) \\
&\quad + 2 \left\| \mathcal{E} \times_1 \mathcal{P}_{\boldsymbol{U}_1} \times_2 \mathcal{P}_{\boldsymbol{U}_2} \times_3 \mathcal{P}_{\boldsymbol{U}_3} \right\|_{\mathrm{F}}^2.
\end{aligned}
\tag{53}
$$

Recognizing that for any $1 \leq \ell \leq k_i$, we have

$$\sigma_\ell \left(\mathcal{M}_i \left(\mathcal{X}^\star \right) \right) = \sigma_\ell \left(\boldsymbol{M}_i^\star \mathcal{M}_i \left(\mathcal{S}^\star \right) \left(\boldsymbol{M}_{i+2}^\star \otimes \boldsymbol{M}_{i+1}^\star \right)^\top \right) \geq \prod_{i=1}^3 \sigma_{k_i} \left(\boldsymbol{M}_i^\star \right) \sigma_\ell \left(\mathcal{M}_i \left(\mathcal{S}^\star \right) \right) \overset{(50)}{\geq} \sqrt{\frac{\beta^3 n_1 n_2 n_3}{k_1 k_2 k_3}} \sigma_\ell \left(\mathcal{M}_i \left(\mathcal{S}^\star \right) \right)$$

and

$$\sigma_\ell \left(\mathcal{M}_i \left(\mathcal{X}^\star \right) \right) \leq \prod_{i=1}^3 \left\| \boldsymbol{M}_i^\star \right\| \sigma_\ell \left(\mathcal{M}_i \left(\mathcal{S}^\star \right) \right) \leq \sqrt{n_1 n_2 n_3} \sigma_\ell \left(\mathcal{M}_i \left(\mathcal{S}^\star \right) \right).$$

In view of Theorem 6, by choosing the numbers of iterations as in (30a) and (30b), we have

$$
\begin{aligned}
\left\| \left(\boldsymbol{I}_{n_1} - \boldsymbol{U}_1 \boldsymbol{U}_1^\top \right) \boldsymbol{X}_1^\star \right\|_{\mathrm{F}}^2 &\leq n_1 \left\| \left(\boldsymbol{I}_{n_1} - \boldsymbol{U}_1 \boldsymbol{U}_1^\top \right) \boldsymbol{X}_1^\star \right\|_{2,\infty}^2 \\
&\lesssim n_1 \cdot \frac{\mu_1 k_1^3}{n_1} \left(k_1^2 \sqrt{n_1} \omega_{\mathsf{max}} \log n + k_1 \left(n_1 n_2 n_3 \right)^{1/4} \omega_{\mathsf{max}} \log n \right)^2 \\
&\overset{\text{(29a) and (48)}}{\lesssim} \frac{k^6}{\beta^2} \left(n_1 n_2 n_3 \right)^{1/2} \omega_{\mathsf{max}}^2 \log^2 n
\end{aligned}
\tag{54}
$$

with probability at least $1 - O(n^{-10})$. Similarly, one has, with probability exceeding $1 - O(n^{-10})$,

$$\left\|\left(\boldsymbol{I}_{n_2} - \boldsymbol{U}_2\boldsymbol{U}_2^\top\right)\boldsymbol{X}_2^\star\right\|_{\mathrm{F}}^2 \lesssim \frac{k^6}{\beta^2}\left(n_1 n_2 n_3\right)^{1/2}\omega_{\mathsf{max}}^2 \log^2 n, \tag{55a}$$

$$\left\|\left(\boldsymbol{I}_{n_2} - \boldsymbol{U}_2\boldsymbol{U}_2^\top\right)\boldsymbol{X}_2^\star\right\|_{\mathrm{F}}^2 \lesssim \frac{k^6}{\beta^2}\left(n_1 n_2 n_3\right)^{1/2}\omega_{\mathsf{max}}^2 \log^2 n. \tag{55b}$$

Moreover, we learn from Theorem 6 that with probability at least $1 - O(n^{-10})$,

$$\left\|\boldsymbol{U}_i\right\|_{2,\infty} = \left\|\boldsymbol{U}_i\boldsymbol{U}_i^\top\right\|_{2,\infty} \le \left\|\boldsymbol{U}_i\boldsymbol{U}_i^\top - \boldsymbol{U}^\star\boldsymbol{U}^\star\right\|_{2,\infty} + \left\|\boldsymbol{U}^\star\right\|_{2,\infty} \le 2\sqrt{\frac{\mu_i k_i^3}{n_i}}, \qquad \forall i \in [3]. \tag{56}$$

Applying Lemma 7 yields that with probability at least $1 - O(n^{-10})$,

$$
\begin{aligned}
\left\|\boldsymbol{\mathcal{E}} \times_1 \mathcal{P}_{\boldsymbol{U}_1} \times_2 \mathcal{P}_{\boldsymbol{U}_2} \times_3 \mathcal{P}_{\boldsymbol{U}_3}\right\|_{\mathrm{F}}^2 &\le k_1 \left\|\boldsymbol{U}_1^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2\right)\right\|^2 \\
&\lesssim kn\left(\mu_1 k_1^2\right)\left(\mu_2 k_2^2\right)\left(\mu_3 k_3^2\right)k^3\omega_{\mathsf{max}}^2 \log n \\
&\overset{(48)}{\lesssim} \frac{k^{13}}{\beta^6}n\omega_{\mathsf{max}}^2 \log n.
\end{aligned} \tag{57}
$$

Putting (53) - (57) together, we obtain

$$\left\|\widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^\star\right\|_{\mathrm{F}}^2 \lesssim \frac{k^6}{\beta^2}\left(n_1 n_2 n_3\right)^{1/2}\omega_{\mathsf{max}}^2 \log^2 n + \frac{k^{13}}{\beta^6}n\omega_{\mathsf{max}}^2 \log n \tag{58}$$

with probability exceeding $1 - O(n^{-10})$.

Step 3: deriving estimation accuracy of the center estimates. We let

$$\widehat{\boldsymbol{\theta}}_\ell^{(i)} = \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1}\right)\widehat{\boldsymbol{b}}_\ell^{(i)} \in \mathbb{R}^{n_{-i}}, \qquad \forall i \in [3], \ell \in [k_i], \tag{59}$$

and also define

$$\boldsymbol{\theta}_\ell^{(i)\star} = \left(\mathcal{M}_i\left(\boldsymbol{\mathcal{X}}^\star\right)\right)_{j,:}^\top \in \mathbb{R}^{n_{-i}}, \qquad \forall i \in [3], \ell \in [k_i]. \tag{60}$$

Here, $j \in [n_i]$ is any index satisfying $z_{i,j}^\star = \ell$ and the $\widehat{\boldsymbol{b}}_\ell^{(i)}$'s are the center estimates satisfying (18a). Recalling that $\widehat{\boldsymbol{B}}_i = \boldsymbol{U}_i\boldsymbol{U}_i^\top \mathcal{M}_i(\boldsymbol{\mathcal{Y}})(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1})$, we have

$$\mathcal{M}_i\left(\widehat{\boldsymbol{\mathcal{X}}}\right) = \widehat{\boldsymbol{B}}_i\left(\boldsymbol{U}_{i+2}^\top \otimes \boldsymbol{U}_{i+1}^\top\right), \qquad \forall i \in [3].$$

This allows one to show that

$$
\begin{aligned}
\sum_{j=1}^{n_i}\left\|\left(\mathcal{M}_i(\widehat{\boldsymbol{\mathcal{X}}})\right)_{j,:}^\top - \widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j}}^{(i)}\right\|_2^2 &= \sum_{j=1}^{n_i}\left\|\left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1}\right)\left(\left(\widehat{\boldsymbol{B}}_i\right)_{j,:}^\top - \widehat{\boldsymbol{b}}_{\widehat{z}_{i,j}}^{(i)}\right)\right\|_2^2 \\
&= \sum_{j=1}^{n_i}\left\|\left(\widehat{\boldsymbol{B}}_i\right)_{j,:}^\top - \widehat{\boldsymbol{b}}_{\widehat{z}_{i,j}}^{(i)}\right\|_2^2 \\
&\overset{(18a)}{\le} M \min_{\substack{\boldsymbol{b}_1,\dots,\boldsymbol{b}_{k_i} \in \mathbb{R}^{r_1 r_2 r_3/r_i} \\ \boldsymbol{z}_i \in [k_i]^{n_i}}} \sum_{j=1}^{n_i}\left\|\left(\widehat{\boldsymbol{B}}_i\right)_{j,:}^\top - \boldsymbol{b}_{z_{i,j}}\right\|_2^2 \\
&= M \min_{\substack{\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_{k_i} \in \mathbb{R}^{n_{-i}} \\ \boldsymbol{z}_i \in [k_i]^{n_i}}} \sum_{j=1}^{n_i}\left\|\left(\mathcal{M}_i(\widehat{\boldsymbol{\mathcal{X}}})\right)_{j,:}^\top - \boldsymbol{\theta}_{z_{i,j}}\right\|_2^2 \\
&\le M \sum_{j=1}^{n_i}\left\|\left(\mathcal{M}_i(\widehat{\boldsymbol{\mathcal{X}}})\right)_{j,:}^\top - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star}\right\|_2^2
\end{aligned}
$$

24

$$= M \sum_{j=1}^{n_i} \left\| \left(\mathcal{M}_i(\widehat{\boldsymbol{\mathcal{X}}})\right)_{j:}^\top - \left(\mathcal{M}_i(\boldsymbol{\mathcal{X}}^\star)\right)_{j:}^\top \right\|_2^2$$

$$= M \left\| \widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^\star \right\|_{\mathrm{F}}^2. \tag{61}$$

Here, the fourth line makes use of Han et al. (2022a, Eqn. (38)). As a result, we have

$$\sum_{j=1}^{n_i} \left\| \widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j}}^{(i)} - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star} \right\|_2^2 \le 2 \left(\sum_{j=1}^{n_i} \left\| \left(\mathcal{M}_i(\widehat{\boldsymbol{\mathcal{X}}})\right)_{j:}^\top - \widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j}}^{(i)} \right\|_2^2 + \sum_{j=1}^{n_i} \left\| \left(\mathcal{M}_i(\widehat{\boldsymbol{\mathcal{X}}})\right)_{j:}^\top - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star} \right\|_2^2 \right)$$

$$\le 2 \left(M \left\| \widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^\star \right\|_{\mathrm{F}}^2 + \left\| \widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^\star \right\|_{\mathrm{F}}^2 \right)$$

$$\le 4M \left\| \widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^\star \right\|_{\mathrm{F}}^2. \tag{62}$$

Noting that $\boldsymbol{M}_i^{\star\top} \boldsymbol{M}_i^\star$ is a diagonal matrix and the diagonal entries

$$\left(\boldsymbol{M}_i^{\star\top} \boldsymbol{M}_i^\star\right)_{\ell,\ell} = \sum_{j=1}^{n_i} (\boldsymbol{M}_i^\star)_{j,\ell}^2 = \left\{ j \in [n_i] : z_{i,j}^\star = \ell \right\} \ge \beta n_i / k_i,$$

we have

$$\sigma_{k_i}(\boldsymbol{M}_i^\star) \ge \sqrt{\beta n_i / k_i}, \qquad \forall i \in [3]. \tag{63}$$

As a consequence, for all $\ell_1 \ne \ell_2 \in [k_i]$, we can derive

$$\left\| \boldsymbol{\theta}_{\ell_1}^{(i)\star} - \boldsymbol{\theta}_{\ell_2}^{(i)\star} \right\| = \left\| \left(\mathcal{M}_i(\boldsymbol{\mathcal{X}})\right)_{j_1,:} - \left(\mathcal{M}_i(\boldsymbol{\mathcal{X}})\right)_{j_2,:} \right\|$$

$$= \left\| \left(\left(\mathcal{M}_i(\boldsymbol{\mathcal{S}})\right)_{\ell_1,:} - \left(\mathcal{M}_i(\boldsymbol{\mathcal{S}})\right)_{\ell_2,:} \right) \left(\boldsymbol{M}_{i+2}^\star \otimes \boldsymbol{M}_{i+1}^\star \right) \right\|$$

$$\ge \left\| \left(\mathcal{M}_i(\boldsymbol{\mathcal{S}})\right)_{\ell_1,:} - \left(\mathcal{M}_i(\boldsymbol{\mathcal{S}})\right)_{\ell_2,:} \right\| \sigma_{k_{i+1}}(\boldsymbol{M}_{i+1}^\star) \sigma_{k_{i+2}}(\boldsymbol{M}_{i+2}^\star)$$

$$\ge \beta \sqrt{\frac{n_{-i}}{k_{-i}}} \Delta_i, \tag{64}$$

where j_1 (resp. j_2) is any index such that $z_{i,j_1}^\star = \ell_1$ (resp. $z_{i,j_2}^\star = \ell_2$). We define

$$\mathcal{S}_i := \left\{ j \in [n_i] : \left\| \widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j}}^{(i)} - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star} \right\|_2 \ge \frac{\beta}{2} \sqrt{\frac{n_{-i}}{k_{-i}}} \Delta_i \right\}. \tag{65}$$

In view of (58) and (62), with probability exceeding $1 - O(n^{-10})$,

$$|\mathcal{S}_i| \le \frac{\sum_{j=1}^{n_i} \left\| \widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j}}^{(i)} - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star} \right\|_2^2}{\left(\frac{\beta}{2} \sqrt{\frac{n_{-i}}{k_{-i}}} \Delta_i \right)^2}$$

$$\le \frac{4M \cdot C_6 \left(\frac{k^6}{\beta^2} (n_1 n_2 n_3)^{1/2} \omega_{\mathsf{max}}^2 \log^2 n + \frac{k^{13}}{\beta^6} n \omega_{\mathsf{max}}^2 \log n \right)}{\left(\frac{\beta}{2} \sqrt{\frac{n_{-i}}{k_{-i}}} \Delta_i \right)^2}$$

$$\le \frac{\beta}{2} \frac{n_i}{k_i}, \tag{66}$$

provided that

$$\Delta_i \ge C_1 \sqrt{M} \left(\frac{k^{9/2}}{\beta^{5/2}} (n_1 n_2 n_3)^{-1/4} \omega_{\mathsf{max}} \log n + \frac{k^8}{\beta^{9/2}} (n_1 n_2 n_3 / n)^{-1/2} \omega_{\mathsf{max}} \sqrt{\log n} \right).$$

For each $1 \le i \le 3$ and $\ell \in [k_i]$, denote by $\mathcal{N}_{i,\ell}$ the following set:

$$\mathcal{N}_{i,\ell} := \left\{ j \in [n_i] : z_{i,j}^\star = \ell, j \in \mathcal{S}^c \right\}. \tag{67}$$

Then we can verify that with probability exceeding $1 - O(n^{-10})$, the following two important properties of the $\mathcal{N}_{i,\ell}$'s hold:

1 . The $\mathcal{N}_{i,\ell}$'s are nonempty:

$$|\mathcal{N}_{i,\ell}| \geq \left|\left\{j \in [n_i] : z_{i,j}^\star = \ell\right\}\right| - |\mathcal{S}_i| \overset{(12) \text{ and } (66)}{\geq} \beta \frac{n_i}{k_i} - \frac{\beta}{2} \frac{n_i}{k_i} = \frac{\beta}{2} \frac{n_i}{k_i} > 0. \tag{68}$$

2. For any $i \in [3]$, the sets $\{\widehat{z}_{i,j} : j \in \mathcal{N}_{i,\ell}\}, \ell \in [k_i]$ are disjoint: for all $\ell_1 \neq \ell_2 \in [k_i], j_1 \in \mathcal{N}_{i,\ell_1}, j_2 \in \mathcal{N}_{i,\ell_2}$,

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j_1}}^{(i)} - \widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j_2}}^{(i)}\right\|_2 &\geq \left\|\boldsymbol{\theta}_{z_{i,j_1}^\star}^{(i)\star} - \boldsymbol{\theta}_{z_{i,j_2}^\star}^{(i)\star}\right\|_2 - \left\|\widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j_1}}^{(i)} - \boldsymbol{\theta}_{z_{i,j_1}^\star}^{(i)\star}\right\|_2 - \left\|\widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j_2}}^{(i)} - \boldsymbol{\theta}_{z_{i,j_2}^\star}^{(i)\star}\right\|_2 \\
&= \left\|\boldsymbol{\theta}_{\ell_1}^{(i)\star} - \boldsymbol{\theta}_{\ell_2}^{(i)\star}\right\|_2 - \left\|\widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j_1}}^{(i)} - \boldsymbol{\theta}_{z_{i,j_1}^\star}^{(i)\star}\right\|_2 - \left\|\widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j_2}}^{(i)} - \boldsymbol{\theta}_{z_{i,j_2}^\star}^{(i)\star}\right\|_2 \\
&\overset{(64)}{>} \beta\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i - \frac{\beta}{2}\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i - \frac{\beta}{2}\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i = 0,
\end{aligned}$$

which implies $\widehat{z}_{i,j_1} \neq \widehat{z}_{i,j_2}$ and further tells us $\{\widehat{z}_{i,j} : j \in \mathcal{N}_{i,\ell_1}\} \cap \{\widehat{z}_{i,j} : j \in \mathcal{N}_{i,\ell_2}\} = \emptyset$.

Therefore, with probability at least $1 - O(n^{-10})$, for any $i \in [3]$, there exists a permutation $\phi_i : [k_i] \to [k_i]$ such that

$$\widehat{z}_{i,j} = \phi_i\left(z_{i,j}^\star\right), \qquad \forall j \in \mathcal{S}_i^c, i \in [3]. \tag{69}$$

In view of (66), with probability at least $1 - O(n^{-10})$, one has

$$\begin{aligned}
\mathsf{MCR}\left(\widehat{\boldsymbol{z}}_i, \boldsymbol{z}_i^\star\right) &\leq \frac{1}{n_i}\left|\left\{j \in [n_i] : \widehat{z}_{i,j} \neq \phi_i\left(z_{i,j}^\star\right)\right\}\right| \leq \frac{1}{n_i}|\mathcal{S}_i| \\
&\leq \frac{4M \cdot C_6\left(\frac{k^6}{\beta^2}\left(n_1 n_2 n_3\right)^{1/2}\omega_{\mathsf{max}}^2 \log^2 n + \frac{k^{13}}{\beta^6}n\omega_{\mathsf{max}}^2 \log n\right)}{n_i\left(\frac{\beta}{2}\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i\right)^2}
\end{aligned} \tag{70}$$

for all $i \in [3]$, and

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\theta}}_{\phi_i(\ell)}^{(i)} - \boldsymbol{\theta}_\ell^{(i)\star}\right\|_2^2 &\leq \frac{\sum_{j=1}^{n_i}\left\|\widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j}}^{(i)} - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star}\right\|_2^2}{\left|\left\{j \in [n_i] : \widehat{z}_{i,j} = \phi_i\left(\ell\right), z_{i,j}^\star = \ell\right\}\right|} \\
&\leq \frac{\sum_{j=1}^{n_i}\left\|\widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j}}^{(i)} - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star}\right\|_2^2}{\left|\left\{j \in [n_i] : z_{i,j}^\star = \ell, j \in \mathcal{S}^c\right\}\right|} \\
&= \frac{\sum_{j=1}^{n_i}\left\|\widehat{\boldsymbol{\theta}}_{\widehat{z}_{i,j}}^{(i)} - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star}\right\|_2^2}{|\mathcal{N}_{i,\ell}|} \\
&\leq \frac{4M \cdot C\left(\frac{k^6}{\beta^2}\left(n_1 n_2 n_3\right)^{1/2}\omega_{\mathsf{max}}^2 \log^2 n + \frac{k^{13}}{\beta^6}n\omega_{\mathsf{max}}^2 \log n\right)}{\frac{\beta n_i}{2k_i}} \\
&\leq \frac{C_7 M\left(\frac{k^7}{\beta^3}\left(n_1 n_2 n_3\right)^{1/2}\omega_{\mathsf{max}}^2 \log^2 n + \frac{k^{14}}{\beta^7}n\omega_{\mathsf{max}}^2 \log n\right)}{n_i}
\end{aligned} \tag{71}$$

for all $\ell \in [k_i]$, provided that

$$\Delta_i \geq C_1\sqrt{M}\left(\frac{k^{9/2}}{\beta^{5/2}}\left(n_1 n_2 n_3\right)^{-1/4}\omega_{\mathsf{max}} \log n + \frac{k^8}{\beta^{9/2}}\left(n_1 n_2 n_3/n\right)^{-1/2}\omega_{\mathsf{max}}\sqrt{\log n}\right).$$

Here, the fourth line of (71) makes use of (58), (62) and (68).

Step 4: proving $\mathsf{MCR}(\widehat{z}_i, z_i^\star) = 0$. Finally, we would like to show that with probability exceeding $1 - O(n^{-10})$, $\widehat{z}_i = \phi_i(z_i^\star)$ for all $i \in [3]$. In view of Theorem 6, Lemma 7, (48) and (56), with probability at least $1 - O(n^{-10})$,

$$
\left\| \left(\boldsymbol{I}_{n_i} - \boldsymbol{U}_i \boldsymbol{U}_i^\top \right) \boldsymbol{X}_i^\star \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu_i k_i^3}{n_i}} \left(k_i^2 \sqrt{n_1} \omega_{\mathsf{max}} \log n + k_i \left(n_1 n_2 n_3 \right)^{1/4} \omega_{\mathsf{max}} \log n \right)
$$

$$
\overset{(48)}{\lesssim} \frac{k^3}{\beta} \frac{\left(n_1 n_2 n_3 \right)^{1/4}}{n_i^{1/2}} \omega_{\mathsf{max}} \log n \tag{72}
$$

and

$$
\left\| \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{E}_i \left(\boldsymbol{U}_{i+2} \boldsymbol{U}_{i+2}^\top \otimes \boldsymbol{U}_{i+1} \boldsymbol{U}_{i+1}^\top \right) \right\|_{2,\infty} \leq \left\| \boldsymbol{U}_i \right\|_{2,\infty} \left\| \boldsymbol{U}_i^\top \boldsymbol{E}_i \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \right\|
$$

$$
\overset{(56) \text{ and Lemma } 7}{\lesssim} \sqrt{\frac{\mu_i k_i^3}{n_i}} \omega_{\mathsf{max}} \sqrt{n \left(\mu_1 k_1^2 \right) \left(\mu_2 k_2^2 \right) \left(\mu_3 k_3^2 \right) k^3 \log n}
$$

$$
\overset{(48)}{\lesssim} \sqrt{\frac{k^4}{\beta^2 n_i}} \frac{k^6}{\beta^3} \omega_{\mathsf{max}} \sqrt{n \log n}
$$

$$
\leq \frac{k^8}{\beta^4} \omega_{\mathsf{max}} \sqrt{\frac{n \log n}{n_i}} \tag{73}
$$

holds for all $i \in [3]$. Here, $\boldsymbol{X}_i^\star = \mathcal{M}_i(\boldsymbol{\mathcal{X}}^\star)$ and $\boldsymbol{E}_i = \mathcal{M}_i(\boldsymbol{\mathcal{E}})$. By virtue of (18b), we know that for any $i \in [3]$,

$$
\left\{ j \in [n_1] : \widehat{z}_{i,j} \neq \phi_i\left(z_{i,j}^\star \right) \right\} \subseteq \left\{ j \in [n_1] : \exists \ell \in [k_i] \text{ s.t. } \left\| \left(\widehat{\boldsymbol{B}}_i \right)_{j:}^\top - \widehat{\boldsymbol{b}}_\ell^{(i)} \right\|_2 \leq \left\| \left(\widehat{\boldsymbol{B}}_i \right)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)} \right\|_2 \right\}. \tag{74}
$$

For any fixed $\ell \neq \phi_i(z_{i,j}^\star) \in [k_i]$, recalling that $\widehat{\boldsymbol{\theta}}_\ell^{(i)} = \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \widehat{\boldsymbol{b}}_\ell^{(i)}$, one has

$$
\mathbb{1}\left\{ \left\| \left(\widehat{\boldsymbol{B}}_i \right)_{j:}^\top - \widehat{\boldsymbol{b}}_\ell^{(i)} \right\|_2 \leq \left\| \left(\widehat{\boldsymbol{B}}_i \right)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)} \right\|_2 \right\}
$$

$$
= \mathbb{1}\left\{ \left\| \left(\widehat{\boldsymbol{B}}_i \right)_{j:}^\top - \widehat{\boldsymbol{b}}_\ell^{(i)} \right\|_2 + \left\| \left(\widehat{\boldsymbol{B}}_i \right)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)} \right\|_2 \leq 2 \left\| \left(\widehat{\boldsymbol{B}}_i \right)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)} \right\|_2 \right\}
$$

$$
\leq \mathbb{1}\left\{ \left\| \widehat{\boldsymbol{b}}_\ell^{(i)} - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)} \right\|_2 \leq 2 \left\| \left(\widehat{\boldsymbol{B}}_i \right)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)} \right\|_2 \right\}
$$

$$
\leq \mathbb{1}\left\{ \left\| \widehat{\boldsymbol{\theta}}_\ell^{(i)} - \widehat{\boldsymbol{\theta}}_{\phi_i(z_{i,j}^\star)}^{(i)} \right\|_2 \leq 2 \left\| \left(\widehat{\boldsymbol{B}}_i \right)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)} \right\|_2 \right\}
$$

$$
= \mathbb{1}\left\{ \left\| \widehat{\boldsymbol{\theta}}_\ell^{(i)} - \widehat{\boldsymbol{\theta}}_{\phi_i(z_{i,j}^\star)}^{(i)} \right\|_2 \leq 2 \left\| \left(\boldsymbol{U}_i \right)_{j,:} \boldsymbol{U}_i^\top \boldsymbol{Y}_i \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)\top} \right\|_2 \right\}, \tag{75}
$$

where $\boldsymbol{Y}_i = \mathcal{M}_i(\boldsymbol{\mathcal{Y}}) = \boldsymbol{X}_i^\star + \boldsymbol{E}_i$. Note that

$$
\left\| \left(\boldsymbol{U}_i \right)_{j,:} \boldsymbol{U}_i^\top \boldsymbol{Y}_i \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)\top} \right\|_2
$$

$$
\leq \left\| \left(\boldsymbol{U}_i \right)_{j,:} \boldsymbol{U}_i^\top \boldsymbol{E}_i \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \right\|_2 + \left\| \left(\boldsymbol{U}_i \right)_{j,:} \boldsymbol{U}_i^\top \boldsymbol{X}_i^\star \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)\top} \right\|_2
$$

$$
\leq \left\| \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{E}_i \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \right\|_{2,\infty} + \left\| \left(\boldsymbol{U}_i \right)_{j,:} \boldsymbol{U}_i^\top \boldsymbol{X}_i^\star \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) - \widehat{\boldsymbol{\theta}}_{\phi_i(z_{i,j}^\star)}^{(i)\top} \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \right\|_2
$$

$$
\leq \left\| \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{E}_i \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \right\|_{2,\infty} + \left\| \left(\boldsymbol{U}_i \right)_{j,:} \boldsymbol{U}_i^\top \boldsymbol{X}_i^\star - \widehat{\boldsymbol{\theta}}_{\phi_i(z_{i,j}^\star)}^{(i)\top} \right\|_2
$$

$$
\leq \left\| \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{E}_i \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \right\|_{2,\infty} + \left\| \left(\boldsymbol{U}_i \right)_{j,:} \boldsymbol{U}_i^\top \boldsymbol{X}_i^\star - \left(\boldsymbol{X}_i^\star \right)_{j,:} \right\|_2 + \left\| \left(\boldsymbol{X}_i^\star \right)_{j,:} - \widehat{\boldsymbol{\theta}}_{\phi_i(z_{i,j}^\star)}^{(i)\top} \right\|_2
$$

$$
\leq \left\| \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{E}_i \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \right\|_{2,\infty} + \left\| \left(\boldsymbol{I}_{n_i} - \boldsymbol{U}_i \boldsymbol{U}_i^\top \right) \boldsymbol{X}_i^\star \right\|_{2,\infty} + \left\| \left(\boldsymbol{X}_i^\star \right)_{j,:} - \widehat{\boldsymbol{\theta}}_{\phi_i(z_{i,j}^\star)}^{(i)\top} \right\|_2
$$

$$
= \left\| \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{E}_i \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \right\|_{2,\infty} + \left\| \left(\boldsymbol{I}_{n_i} - \boldsymbol{U}_i \boldsymbol{U}_i^\top \right) \boldsymbol{X}_i^\star \right\|_{2,\infty} + \left\| \widehat{\boldsymbol{\theta}}_{\phi_i(z_{i,j}^\star)}^{(i)} - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star} \right\|_2
$$

$$
\leq \left\| \boldsymbol{U}_i \boldsymbol{U}_i^\top \boldsymbol{E}_i \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \right\|_{2,\infty} + \left\| \left(\boldsymbol{I}_{n_i} - \boldsymbol{U}_i \boldsymbol{U}_i^\top \right) \boldsymbol{X}_i^\star \right\|_{2,\infty} + \sup_{a \in [k_i]} \left\| \widehat{\boldsymbol{\theta}}_{\phi_i(a)}^{(i)} - \boldsymbol{\theta}_a^{(i)\star} \right\|_2
$$

and

$$\left\|\widehat{\boldsymbol{\theta}}_\ell^{(i)} - \widehat{\boldsymbol{\theta}}_{\phi_i(z_{i,j}^\star)}^{(i)}\right\|_2 \geq \left\|\boldsymbol{\theta}_{\phi_i^{-1}(\ell)}^{(i)\star} - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star}\right\|_2 - \left\|\widehat{\boldsymbol{\theta}}_\ell^{(i)} - \boldsymbol{\theta}_{\phi_i^{-1}(\ell)}^{(i)\star}\right\|_2 - \left\|\widehat{\boldsymbol{\theta}}_{\phi_i(z_{i,j}^\star)}^{(i)} - \boldsymbol{\theta}_{z_{i,j}^\star}^{(i)\star}\right\|_2$$

$$\overset{(64)}{\geq} \beta\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i - 2\sup_{a\in[k_i]}\left\|\widehat{\boldsymbol{\theta}}_{\phi_i(a)}^{(i)} - \boldsymbol{\theta}_a^{(i)\star}\right\|_2.$$

Putting the previous two inequalities, (71), (72) and (73) together yields: with probability at least $1 - O(n^{-10})$,

$$\left\|\widehat{\boldsymbol{\theta}}_\ell^{(i)} - \widehat{\boldsymbol{\theta}}_{\phi_i(z_{i,j}^\star)}^{(i)}\right\|_2 - 2\left\|(\boldsymbol{U}_i)_{j,:}\,\boldsymbol{U}_i^\top\boldsymbol{Y}_i\,(\boldsymbol{U}_{i+2}\otimes\boldsymbol{U}_{i+1}) - \widehat{\boldsymbol{b}}_{\phi_i(z_{i,j}^\star)}^{(i)\top}\right\|_2$$

$$\geq \beta\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i - 2\sup_{a\in[k_i]}\left\|\widehat{\boldsymbol{\theta}}_{\phi_i(a)}^{(i)} - \boldsymbol{\theta}_a^{(i)\star}\right\|_2$$

$$- 2\left(\left\|\boldsymbol{U}_i\boldsymbol{U}_i^\top\boldsymbol{E}_i\,(\boldsymbol{U}_{i+2}\otimes\boldsymbol{U}_{i+1})\right\|_{2,\infty} + \left\|\left(\boldsymbol{I}_{n_i} - \boldsymbol{U}_i\boldsymbol{U}_i^\top\right)\boldsymbol{X}_i^\star\right\|_{2,\infty} + \sup_{a\in[k_i]}\left\|\widehat{\boldsymbol{\theta}}_{\phi_i(a)}^{(i)} - \boldsymbol{\theta}_a^{(i)\star}\right\|_2\right)$$

$$\geq \beta\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i - 4\sqrt{\frac{C_7 M\left(\frac{k^7}{\beta^3}(n_1 n_2 n_3)^{1/2}\omega_{\mathsf{max}}^2\log^2 n + \frac{k^{14}}{\beta^7}n\omega_{\mathsf{max}}^2\log n\right)}{n_i}}$$

$$- C_8\frac{k^3}{\beta}\frac{(n_1 n_2 n_3)^{1/4}}{n_i^{1/2}}\omega_{\mathsf{max}}\log n - C_8\frac{k^8}{\beta^4}\omega_{\mathsf{max}}\sqrt{\frac{n\log n}{n_i}}$$

$$> 0 \tag{76}$$

holds for all $\ell\in[k_i]$, provided that

$$\Delta_i/\omega_{\mathsf{max}} \geq C_1\sqrt{M}\left(\frac{k^{9/2}}{\beta^{5/2}}(n_1 n_2 n_3)^{-1/4}\log n + \frac{k^9}{\beta^5}(n_1 n_2 n_3/n)^{-1/2}\sqrt{\log n}\right).$$

Combining (74), (75) and (76), we arrive at

$$\mathsf{MCR}\big(\widehat{\boldsymbol{z}}_i, \boldsymbol{z}_i^\star\big) = 0, \qquad \forall i\in[3]$$

with probability exceeding $1 - O(n^{-10})$.

C Proof of Theorem 5

Part (a): proving $\mathcal{A}\neq\emptyset$. Let

$$\overline{r} = \begin{cases} \max\mathcal{A}, & \text{if } \mathcal{A}\neq\emptyset; \\ 0, & \text{otherwise.} \end{cases} \tag{77}$$

We claim that

$$\sigma_{\overline{r}+1}^\star \leq 2C_0 r\big[(m_1 m_2)^{1/4} + r m_1^{1/2}\big]\omega_{\mathsf{max}}\log m. \tag{78}$$

In fact, if $\sigma_{\overline{r}+1}^\star = 0$, then (78) clearly holds. If $\sigma_{\overline{r}+1}^\star > 0$, then we must have $\overline{r} < r$. Let

$$i = \min\left\{j : \overline{r}+1 \leq j \leq r, \sigma_j^\star \geq \frac{4r}{4r-1}\sigma_{j+1}^\star\right\}.$$

Note that such i does exist as the largest $j \leq r$ satisfying $\sigma_j^\star > 0$ must obey $\sigma_j^\star \geq \frac{4r}{4r-1}\sigma_{j+1}^\star$. The definition of $\bar{r}$ immediately tells us that $\sigma_i^\star \leq C_0 r[(m_1 m_2)^{1/4} + rm_1^{1/2}]\log m$ and consequently one has

$$\sigma_{\bar{r}+1}^\star = \sigma_i^\star \prod_{j=\bar{r}+1}^{i-1} \frac{\sigma_j^\star}{\sigma_{j+1}^\star} \leq \sigma_i^\star \left(\frac{4r}{4r-1}\right)^{i-\bar{r}-1} \leq \sigma_i^\star \cdot \left(1 + \frac{1}{3r}\right)^r \leq 2C_0 r\big[(m_1 m_2)^{1/4} + rm_1^{1/2}\big]\omega_{\mathsf{max}}\log m.$$

The first inequality holds due to the definition of i. This combined with the assumption on $\sigma_1^\star$ reveals that $\bar{r} \neq 0$, i.e., $\mathcal{A} \neq \emptyset$ and

$$\bar{r} = \max \mathcal{A}. \tag{79}$$

Part (b): proving (39b). The rest of the proof is devoted to proving (39b). Letting

$$\boldsymbol{U}^{\star(1)} = [\boldsymbol{u}_1^\star, \ldots, \boldsymbol{u}_{\bar{r}}^\star], \qquad \boldsymbol{\Sigma}^{\star(1)} = \mathsf{diag}\,(\sigma_1^\star, \ldots, \sigma_{\bar{r}}^\star), \qquad \boldsymbol{V}^{\star(1)} = [\boldsymbol{v}_1^\star, \ldots, \boldsymbol{v}_{\bar{r}}^\star], \tag{80a}$$
$$\boldsymbol{U}^{\star(2)} = [\boldsymbol{u}_{\bar{r}+1}^\star, \ldots, \boldsymbol{u}_r^\star], \qquad \boldsymbol{\Sigma}^{\star(2)} = \mathsf{diag}\,(\sigma_{\bar{r}+1}^\star, \ldots, \sigma_r^\star), \qquad \boldsymbol{V}^{\star(2)} = [\boldsymbol{v}_{\bar{r}+1}^\star, \ldots, \boldsymbol{v}_r^\star], \tag{80b}$$

we can derive

$$\boldsymbol{U}^\star = \big[\boldsymbol{U}^{\star(1)}\, \boldsymbol{U}^{\star(2)}\big], \qquad \boldsymbol{V}^\star = \big[\boldsymbol{V}^{\star(1)}\, \boldsymbol{V}^{\star(2)}\big], \qquad \boldsymbol{\Sigma}^\star = \begin{bmatrix} \boldsymbol{\Sigma}^{\star(1)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}^{\star(2)} \end{bmatrix}. \tag{81}$$

Let the SVD of $\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}$ be denoted by

$$\widetilde{\boldsymbol{U}}^{(1)}\widetilde{\boldsymbol{\Sigma}}^{(1)}\widetilde{\boldsymbol{W}}^{(1)\top} = \boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}. \tag{82}$$

Here, $\widetilde{\boldsymbol{U}}^{(1)} \in \mathcal{O}^{n_1, \bar{r}}$, $\widetilde{\boldsymbol{\Sigma}}^{(1)} = \mathsf{diag}(\widetilde{\sigma}_1, \ldots, \widetilde{\sigma}_{\bar{r}})$ where $\widetilde{\sigma}_1 \geq \cdots \geq \widetilde{\sigma}_{\bar{r}} \geq 0$, $\widetilde{\boldsymbol{W}}^{(1)} \in \mathcal{O}^{\bar{r}, \bar{r}}$. Then one has

$$\big(\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}\big)\big(\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}\big)^\top = \widetilde{\boldsymbol{U}}^{(1)}\big(\widetilde{\boldsymbol{\Sigma}}^{(1)}\big)^2\widetilde{\boldsymbol{U}}^{(1)\top}. \tag{83}$$

We can then write $\boldsymbol{M}^{\mathsf{oracle}}$ as follows:

$$\begin{aligned}
\boldsymbol{M}^{\mathsf{oracle}} &= (\boldsymbol{X}^\star + \boldsymbol{E})\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}(\boldsymbol{X}^\star + \boldsymbol{E})^\top + \mathcal{P}_{\mathsf{off\text{-}diag}}\big(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\big) \\
&= (\boldsymbol{X}^\star + \boldsymbol{E})\boldsymbol{V}^{\star(1)}\boldsymbol{V}^{\star(1)\top}(\boldsymbol{X}^\star + \boldsymbol{E})^\top + (\boldsymbol{X}^\star + \boldsymbol{E})\boldsymbol{V}^{\star(2)}\boldsymbol{V}^{\star(2)\top}(\boldsymbol{X}^\star + \boldsymbol{E})^\top \\
&\quad + \mathcal{P}_{\mathsf{off\text{-}diag}}\big(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\big) \\
&= \underbrace{\big(\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}\big)\big(\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}\big)^\top + \mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp}\boldsymbol{U}^{\star(2)}\big(\boldsymbol{\Sigma}^{\star(2)}\big)^2\boldsymbol{U}^{\star(2)\top}\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp}}_{=:\widetilde{\boldsymbol{M}}} \\
&\quad + \underbrace{\boldsymbol{U}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\boldsymbol{E}^\top + \boldsymbol{E}\boldsymbol{V}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{U}^{\star(2)\top} + \boldsymbol{E}\boldsymbol{V}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\boldsymbol{E}^\top}_{=:\boldsymbol{Z}_1} \\
&\quad + \underbrace{\mathcal{P}_{\widetilde{\boldsymbol{U}}^{(1)}}\boldsymbol{U}^{\star(2)}\big(\boldsymbol{\Sigma}^{\star(2)}\big)^2\boldsymbol{U}^{\star(2)\top}\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp} + \boldsymbol{U}^{\star(2)}\big(\boldsymbol{\Sigma}^{\star(2)}\big)^2\boldsymbol{U}^{\star(2)\top}\mathcal{P}_{\widetilde{\boldsymbol{U}}^{(1)}}}_{=:\boldsymbol{Z}_2} \\
&\quad + \underbrace{\mathcal{P}_{\mathsf{off\text{-}diag}}\big(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\big)}_{=:\boldsymbol{Z}_3}.
\end{aligned} \tag{84}$$

For convenience, we shall also let

$$\boldsymbol{Z} = \boldsymbol{Z}_1 + \boldsymbol{Z}_2 + \boldsymbol{Z}_3. \tag{85}$$

C.1   Several key lemmas

Before proceeding, we first introduce the following lemma, which allows us to bound an infinite sum of $\ell_{2,\infty}$ norms of perturbation matrix polynomials instead of bounding $\big\|\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}}\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^\top\big\|_{2,\infty}$ directly.

Lemma 1. *Suppose that $\boldsymbol{M} = \overline{\boldsymbol{M}} + \boldsymbol{Z} \in \mathbb{R}^{n \times n}$, where $\overline{\boldsymbol{M}}$ and $\boldsymbol{Z}$ are both symmetric matrices. Assume that $\overline{\boldsymbol{M}}$ is a matrix with rank not exceeding r and has eigenvalues $\overline{\lambda}_1 \geq \cdots \geq \overline{\lambda}_r \geq 0$ and rank-r leading eigenspace $\overline{\boldsymbol{U}} = [\overline{\boldsymbol{u}}_1, \ldots, \overline{\boldsymbol{u}}_r]$ (so that $\overline{\boldsymbol{u}}_i$ is the eigenvector associated with $\overline{\lambda}_i$). If there exists some r_1 obeying $1 \leq r_1 \leq r$ and*

$$\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1} > 2\|\boldsymbol{Z}\|, \tag{86}$$

then it holds that

$$\left\| \overline{\boldsymbol{U}}_1 \overline{\boldsymbol{U}}_1^\top - \boldsymbol{U}_1 \boldsymbol{U}_1^\top \right\|_{2,\infty} \leq \frac{8}{\pi} \sum_{k \geq 1} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \sum_{\substack{0 \leq j_1,\ldots,j_{k+1} \leq r \\ (j_1,\ldots,j_{k+1}) \neq \boldsymbol{0}}} \left\| \overline{\boldsymbol{P}}_{j_1} \boldsymbol{Z} \overline{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \overline{\boldsymbol{P}}_{j_{k+1}} \right\|_{2,\infty}, \tag{87a}$$

$$\left\| (\overline{\boldsymbol{U}}_1 \overline{\boldsymbol{U}}_1^\top - \boldsymbol{U}_1 \boldsymbol{U}_1^\top) \overline{\boldsymbol{M}} \right\|_{2,\infty} \leq \frac{40}{\pi} \sum_{k \geq 1} \overline{\lambda}_{r_1} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \sum_{\substack{0 \leq j_1,\ldots,j_{k+1} \leq r \\ (j_1,\ldots,j_{k+1}) \neq \boldsymbol{0}}} \left\| \overline{\boldsymbol{P}}_{j_1} \boldsymbol{Z} \overline{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \overline{\boldsymbol{P}}_{j_{k+1}} \right\|_{2,\infty}. \tag{87b}$$

Here, $\overline{\boldsymbol{U}}_1$ and $\boldsymbol{U}_1$ denote the rank-r_1 leading eigen-subspace of $\overline{\boldsymbol{M}}$ and $\boldsymbol{M}$, respectively; and we denote $\overline{\boldsymbol{P}}_j = \overline{\boldsymbol{u}}_j \overline{\boldsymbol{u}}_j^\top$ for any $1 \leq j \leq r$ and $\overline{\boldsymbol{P}}_0 = \overline{\boldsymbol{U}}_\perp \overline{\boldsymbol{U}}_\perp^\top$.

The proof of Lemma 1 can be found in Section C.3. In addition, the following lemmas deliver sharp $\ell_{2,\infty}$ guarantees for some polynomials of the noise matrix.

Lemma 2 (Zhou and Chen (2023), Lemma 2)**.** *Suppose that Assumption 2 holds. Then we have, with probability at least $1 - O(m^{-10})$,*

$$\left\| \left[\mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{E} \boldsymbol{E}^\top \right) \right]^k \boldsymbol{E} \boldsymbol{V}^\star \right\|_{2,\infty} \leq C_3 \sqrt{\mu r} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^k \omega_{\mathsf{max}} \log m \tag{88}$$

for all $0 \leq k \leq \log n$, where C_3 is some sufficiently large constant.

Lemma 3 (Zhou and Chen (2023), Lemma 3)**.** *Suppose that Assumption 2 holds. Then we have, with probability at least $1 - O(m^{-10})$,*

$$\left\| \left[\mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{E} \boldsymbol{E}^\top \right) \right]^k \boldsymbol{U}^\star \right\|_{2,\infty} \leq C_3 \sqrt{\frac{\mu r}{n_1}} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^k \tag{89}$$

for all $0 \leq k \leq \log n$, where C_3 is some sufficiently large constant.

The following lemma, which was also established in Zhou and Chen (2023), provides some helpful consequences on the eigenvalue perturbation, the size of some perturbation matrix, and some incoherence properties of $\widetilde{\boldsymbol{U}}^{(1)}$.

Lemma 4 (Zhou and Chen (2023), Lemma 4)**.** *Instate the assumptions in Theorem 5. Then there exist some large enough constant $C_5 > 0$ such that with probability exceeding $1 - O(m^{-10})$,*

$$\left| \widetilde{\sigma}_i - \sigma_i^\star \right| \leq \left\| \boldsymbol{E} \boldsymbol{V}^{\star(1)} \right\| \leq \left\| \boldsymbol{E} \boldsymbol{V}^\star \right\| \leq \sqrt{C_5} \sqrt{m_1} \omega_{\mathsf{max}} \log m, \qquad \forall i \in [\overline{r}], \tag{90a}$$

$$\left\| \mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{E} \boldsymbol{E}^\top - \boldsymbol{E} \boldsymbol{V}^\star \boldsymbol{V}^{\star\top} \boldsymbol{E}^\top \right) \right\| \leq 3 C_5 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \tag{90b}$$

$$\left\| \boldsymbol{U}^{\star(1)} \boldsymbol{U}^{\star(1)\top} \widetilde{\boldsymbol{U}}^{(1)} - \widetilde{\boldsymbol{U}}^{(1)} \right\|_{2,\infty} \leq \frac{4 C_5 \sqrt{\mu r} \omega_{\mathsf{max}} \log m}{\sigma_r^\star} \leq \sqrt{\frac{\mu r}{m_1}}, \tag{90c}$$

$$\left\| \widetilde{\boldsymbol{U}}^{(1)} \right\|_{2,\infty} \leq 2 \sqrt{\frac{\mu r}{m_1}}. \tag{90d}$$

Finally, the following lemma develops $\ell_{2,\infty}$ bounds on the polynomials of the perturbation matrix $\boldsymbol{Z}_3$, which will play a key role in the subsequent proof.

Lemma 5. *Suppose that Assumption 2 holds. Let*

$$\mathcal{E} = \{(88) \text{ and } (89) \text{ hold for } 0 \le k \le \log n\} \cap \{(90a), (90b), (90c) \text{ and } (90d) \text{ hold}\}. \tag{91}$$

Then there exists some large enough constant $C_2, C_3 > 0$ (independent of C_0) such that under $\mathcal{E}$, for any $0 \le i \le \log n$, one has

$$\left\| \boldsymbol{Z}_3^i \boldsymbol{U}^\star \right\|_{2,\infty} \le 3C_3 \sqrt{\frac{\mu r}{m_1}} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i, \tag{92a}$$

$$\left\| \boldsymbol{Z}_3^i \boldsymbol{E} \boldsymbol{V}^\star \right\|_{2,\infty} \le 3C_3 \sqrt{\mu r} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i \omega_{\mathsf{max}} \log m, \tag{92b}$$

$$\left\| \boldsymbol{Z}_3^i \widetilde{\boldsymbol{U}}^{(1)} \right\|_{2,\infty} \le 4C_3 \sqrt{\frac{\mu r}{m_1}} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i, \tag{92c}$$

$$\left\| \boldsymbol{Z}_3^i \boldsymbol{Z}_1 \right\|_{2,\infty} \le C_2 \sqrt{\mu r} \left(\sigma_{\bar{r}+1}^\star + \sqrt{m_1} \omega_{\mathsf{max}} \log m \right) \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i \omega_{\mathsf{max}} \log m, \tag{92d}$$

$$\left\| \boldsymbol{Z}_3^i \boldsymbol{Z}_2 \right\|_{2,\infty} \le C_2 \sqrt{\mu r} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i \omega_{\mathsf{max}} \sigma_{\bar{r}+1}^\star \log m. \tag{92e}$$

The proof of Lemma 5 is postponed to Section C.4. The union bound taken together with Lemma 2, Lemma 3 and Lemma 4 shows that

$$\mathbb{P}\left(\mathcal{E} \right) \ge 1 - O\left(n^{-10} \right). \tag{93}$$

In the rest of the proof, we assume that $\mathcal{E}$ occurs unless otherwise noted.

C.2 Main steps for proving (39b)

Step 1: bounding $\|\widetilde{\boldsymbol{U}}^{(2)}\|_2$. We start with controlling $\|\widetilde{\boldsymbol{U}}^{(2)}\|_2$. Combining (90a), (Chen et al., 2021a, Lemma 2.5) and Wedin's $\sin \boldsymbol{\Theta}$ theorem, one has

$$\left\| \widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)} \right\| \le \left\| \widetilde{\boldsymbol{U}}^{(1)\top} \left(\boldsymbol{U}^{\star(1)} \right)_{\perp} \right\| = \left\| \widetilde{\boldsymbol{U}}^{(1)} \widetilde{\boldsymbol{U}}^{(1)\top} - \boldsymbol{U}^{\star(1)} \boldsymbol{U}^{\star(1)\top} \right\|$$
$$\le \frac{2 \left\| \boldsymbol{E} \boldsymbol{V}^{\star(1)} \right\|}{\sigma_{\bar{r}}^\star} \le \frac{2\sqrt{C_5}\sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma_{\bar{r}}^\star} \ll \frac{1}{2}, \tag{94}$$

where the first inequality makes use of the fact that $\boldsymbol{U}_1^{\star\top} \boldsymbol{U}_2^\star = \boldsymbol{0}$. Note that the rank of $\widetilde{\boldsymbol{M}}$ is at most r. We also denote the eigendecomposition of $\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_{\perp}} \boldsymbol{U}^{\star(2)} \left(\boldsymbol{\Sigma}^{\star(2)} \right)^2 \boldsymbol{U}^{\star(2)\top} \mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_{\perp}}$ by

$$\widetilde{\boldsymbol{U}}^{(2)} \left(\widetilde{\boldsymbol{\Sigma}}^{(2)} \right)^2 \widetilde{\boldsymbol{U}}^{(2)} = \mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_{\perp}} \boldsymbol{U}^{\star(2)} \left(\boldsymbol{\Sigma}^{\star(2)} \right)^2 \boldsymbol{U}^{\star(2)\top} \mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_{\perp}}, \tag{95}$$

where $\widetilde{\boldsymbol{U}}^{(2)} \in \mathcal{O}^{m_1, r-\bar{r}}$ and $\widetilde{\boldsymbol{\Sigma}}^{(1)} = \mathsf{diag}(\widetilde{\sigma}_{\bar{r}+1}, \dots, \widetilde{\sigma}_r)$ with $\widetilde{\sigma}_{\bar{r}+1} \ge \cdots \ge \widetilde{\sigma}_r \ge 0$. Recognizing that $\widetilde{\boldsymbol{U}}^{(1)\top} \widetilde{\boldsymbol{U}}^{(2)} = \boldsymbol{0}$, we see that the eigendecomposition of $\widetilde{\boldsymbol{M}}$ can be written as

$$\widetilde{\boldsymbol{M}} = \widetilde{\boldsymbol{U}} \widetilde{\boldsymbol{\Lambda}} \widetilde{\boldsymbol{U}}^\top, \tag{96}$$

where

$$\widetilde{\boldsymbol{U}} = \left[\widetilde{\boldsymbol{U}}^{(1)} \ \widetilde{\boldsymbol{U}}^{(2)} \right], \qquad \text{and} \qquad \widetilde{\boldsymbol{\Lambda}} = \mathsf{diag}\left(\widetilde{\sigma}_1^2, \dots, \widetilde{\sigma}_r^2 \right) = \begin{bmatrix} \left(\widetilde{\boldsymbol{\Sigma}}^{(1)} \right)^2 & \boldsymbol{0} \\ \boldsymbol{0} & \left(\widetilde{\boldsymbol{\Sigma}}^{(2)} \right)^2 \end{bmatrix}. \tag{97}$$

In addition, one observes that

$$\sigma_{r-\bar{r}}\left(\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_{\perp}} \boldsymbol{U}^{\star(2)} \right) = \sigma_{r-\bar{r}}\left(\left(\widetilde{\boldsymbol{U}}^{(1)} \right)_{\perp}^\top \boldsymbol{U}^{\star(2)} \right) = \min_{\boldsymbol{a} \in \mathbb{R}^{r-\bar{r}} : \|\boldsymbol{a}\|_2 = 1} \left\| \left(\widetilde{\boldsymbol{U}}^{(1)} \right)_{\perp}^\top \boldsymbol{U}^{\star(2)} \boldsymbol{a} \right\|_2$$
$$= \sqrt{\min_{\boldsymbol{a} \in \mathbb{R}^{r-\bar{r}} : \|\boldsymbol{a}\|_2 = 1} \left(\left\| \boldsymbol{U}^{\star(2)} \boldsymbol{a} \right\|_2^2 - \left\| \widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)} \boldsymbol{a} \right\|_2^2 \right)}$$

$$= \sqrt{1 - \max_{\boldsymbol{a} \in \mathbb{R}^{r-\overline{r}}: \|\boldsymbol{a}\|_2 = 1} \left\| \widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)} \boldsymbol{a} \right\|_2^2}$$

$$= \sqrt{1 - \left\| \widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)} \right\|^2}$$

$$\geq \sqrt{1 - \frac{1}{Cr^2}}$$

$$\geq \sqrt{1 - \left(\frac{1}{2}\right)^2} = \frac{\sqrt{3}}{2}. \tag{98}$$

The last line holds due to (94). Noting that $\widetilde{\boldsymbol{U}}^{(2)}$ is also the column subspace of $\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp} \boldsymbol{U}^{\star(2)} \in \mathbb{R}^{m_1, r-\overline{r}}$ and combining (90d), (94) and (98), one reaches

$$\left\| \widetilde{\boldsymbol{U}}^{(2)} \right\|_{2,\infty} \leq \left\| \mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp} \boldsymbol{U}^{\star(2)} \right\|_{2,\infty} \sigma_{r-\overline{r}}^{-1} \left(\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp} \boldsymbol{U}^{\star(2)} \right)$$

$$\leq \frac{2}{\sqrt{3}} \left(\left\| \boldsymbol{U}^{\star(2)} \right\|_{2,\infty} + \left\| \mathcal{P}_{\widetilde{\boldsymbol{U}}^{(1)}} \boldsymbol{U}^{\star(2)} \right\|_{2,\infty} \right)$$

$$\leq \frac{2}{\sqrt{3}} \left(\sqrt{\frac{\mu r}{m_1}} + \left\| \widetilde{\boldsymbol{U}}^{(1)} \right\|_{2,\infty} \left\| \widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)} \right\| \right)$$

$$\leq \frac{2}{\sqrt{3}} \left(\sqrt{\frac{\mu r}{m_1}} + 2\sqrt{\frac{\mu r}{m_1}} \cdot \frac{1}{2} \right)$$

$$\leq 2\sqrt{\frac{\mu r}{m_1}}. \tag{99}$$

Step 2: bounding $\widetilde{\sigma}_{r'}^2 - \widetilde{\sigma}_{r'+1}^2$ and $\|\boldsymbol{Z}\|$. Recall that $\lambda_i(\widetilde{M}) = \widetilde{\sigma}_i^2$ for $i \in [r]$. To apply Lemma 1, one needs to check the condition

$$\widetilde{\sigma}_{r'}^2 - \widetilde{\sigma}_{r'+1}^2 > 2\|\boldsymbol{Z}\|, \qquad \forall r' \in \mathcal{A}. \tag{100}$$

It is seen from the definition of $\widetilde{\sigma}_{\overline{r}+1}$ that

$$\widetilde{\sigma}_{\overline{r}+1} \leq \left\| \boldsymbol{\Sigma}^{\star(2)} \right\| = \sigma_{\overline{r}+1}^\star. \tag{101}$$

Further, (95) and (98) taken together imply that

$$\widetilde{\sigma}_{\overline{r}+1}^2 \geq \sigma_{r-\overline{r}}^2 \left(\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp} \boldsymbol{U}^{\star(2)} \right) \sigma_{\overline{r}+1}^{\star 2} \geq \left(1 - \frac{2\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}} \log m}{\sigma_{\overline{r}}^\star} \right)^2 \sigma_{\overline{r}+1}^{\star 2} \geq \left(1 - \frac{1}{Cr^2} \right)^2 \sigma_{\overline{r}+1}^{\star 2} \tag{102}$$

for some large constant $C > 0$. By virtue of (101) and (102), one has

$$\max \left\{ \left(1 - \frac{1}{Cr^2} \right) \sigma_{\overline{r}+1}^\star, \sigma_{\overline{r}+1}^\star - 2\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}} \log m \right\} \leq \widetilde{\sigma}_{\overline{r}+1} \leq \sigma_{\overline{r}+1}^\star. \tag{103}$$

Putting (90a), (101) and the fact $\sigma_{r'}^\star \geq \sigma_{\overline{r}}^\star \geq C_0 r[(m_1 m_2)^{1/4} + m_1^{1/2}] \log m$ together yields

$$\widetilde{\sigma}_{r'} - \widetilde{\sigma}_{r'+1} \geq \sigma_{r'}^\star - \sigma_{r'+1}^\star - 2\left\| \boldsymbol{E}\boldsymbol{V}^\star \right\|$$

$$\geq \sigma_{r'}^\star - \sigma_{r'+1}^\star - 2\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}} \log m$$

$$\geq \sigma_{r'}^\star - \sigma_{r'+1}^\star - \frac{\sigma_{r'}^\star}{Cr}$$

$$\geq \frac{1}{2}\left(\sigma_{r'}^\star - \sigma_{r'+1}^\star \right) + \frac{1}{2}\left(\sigma_{r'}^\star - \frac{4r-1}{4r}\sigma_{r'}^\star \right) - \frac{\sigma_{r'}^\star}{Cr}$$

$$\geq \frac{1}{2}\left(\sigma_{r'}^\star - \sigma_{r'+1}^\star \right) \geq \frac{\sigma_{r'}^\star}{8r}.$$

Here, the penultimate and the last lines hold due to the fact $r' \in \mathcal{A}$. In addition, we observe that

$$\widetilde{\sigma}_{r'} + \widetilde{\sigma}_{r'+1} \geq \sigma_{r'}^\star + \sigma_{r'+1}^\star - 2\|\boldsymbol{EV}^\star\| \geq \sigma_{r'}^\star + \sigma_{r'+1}^\star - \frac{\sigma_{r'}^\star}{Cr} \geq \frac{1}{2}\left(\sigma_{r'}^\star + \sigma_{r'+1}^\star\right).$$

Combining the previous two inequalities leads to

$$\widetilde{\sigma}_{r'}^2 - \widetilde{\sigma}_{r'+1}^2 \geq \frac{1}{4}\left(\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}\right) \geq \frac{\sigma_{r'}^{\star 2}}{16r}. \tag{104}$$

Now, we move on to control $\|\boldsymbol{Z}\|$. In view of (90a) and (94), we have

$$\|\boldsymbol{Z}_1\| \leq 2\|\boldsymbol{\Sigma}^{\star(2)}\|\|\boldsymbol{EV}^\star\| + \|\boldsymbol{EV}^\star\|^2 \leq 2\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\widetilde{r}+1}^\star + C_5 m_1 \omega_{\mathsf{max}}^2 \log^2 m \tag{105}$$

and

$$\|\boldsymbol{Z}_2\| \leq 2\|\widetilde{\boldsymbol{U}}_1^\top \boldsymbol{U}_2^{\star(2)}\|\|\boldsymbol{\Sigma}^{\star(2)}\|^2 \lesssim \frac{\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma_{\widetilde{r}}^\star}\sigma_{\widetilde{r}+1}^{\star 2} \leq \sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\widetilde{r}+1}^\star. \tag{106}$$

Combining (105), (106) and (90b), we arrive at

$$\|\boldsymbol{Z}\| \lesssim \sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\widetilde{r}+1}^\star + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m \ll \frac{\sigma_{r'}^{\star 2}}{16r} \leq \widetilde{\sigma}_{r'}^2 - \widetilde{\sigma}_{r'+1}^2, \tag{107}$$

which validates (100). Here, the second inequality uses the facts $\sigma_{r'}^\star \geq \sigma_{\widetilde{r}}^\star \geq C_0 r[(m_1 m_2)^{1/4} + r m_1^{1/2}]\log m$. By virtue of Lemma 1 and (104), we have

$$\left\|\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}}\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^\top\right\|_{2,\infty}$$

$$\leq \frac{8}{\pi}\sum_{k\geq 1}\frac{2^k}{\left(\widetilde{\sigma}_{r'}^2 - \widetilde{\sigma}_{r'+1}^2\right)^k}\sum_{\substack{0 \leq j_1,\ldots,j_{k+1} \leq r \\ (j_1,\ldots,j_{k+1})^\top \neq \boldsymbol{0}_{k+1}}}\left\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}$$

$$\leq \frac{8}{\pi}\sum_{k\geq 1}\left(\frac{8}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}\right)^k\sum_{\substack{0 \leq j_1,\ldots,j_{k+1} \leq r \\ (j_1,\ldots,j_{k+1})^\top \neq \boldsymbol{0}_{k+1}}}\left\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}, \tag{108}$$

Here, for any $1 \leq j \leq r$, $\widetilde{\boldsymbol{P}}_j = \widetilde{\boldsymbol{u}}_j\widetilde{\boldsymbol{u}}_j^\top$ and $\widetilde{\boldsymbol{P}}_0 = \widetilde{\boldsymbol{U}}_\perp\widetilde{\boldsymbol{U}}_\perp^\top$.

To bound $\|\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}}\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^\top\|_{2,\infty}$, we will bound each single term $\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\|_{2,\infty}$ for $1 \leq k \leq \log n$, and show that the total contribution of the remaining terms is small.

Step 3: bounding $\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\|_{2,\infty}$ for small k. For any $1 \leq k \leq \log n$ and $(j_1,\ldots,j_{k+1}) \in \{0,1,\ldots,r\}^{k+1}\backslash\boldsymbol{0}$, let ℓ denote the the smallest i such that $j_i \neq 0$.

Step 3.1: bounding $\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\|_{2,\infty}$ when $\ell = 1$. If $\ell = 1$, then (90d) and (99) taken collectively show that

$$\|\widetilde{\boldsymbol{u}}_{j_1}\|_\infty \leq \max\left\{\|\widetilde{\boldsymbol{U}}_1\|_{2,\infty}, \|\widetilde{\boldsymbol{U}}_2\|_{2,\infty}\right\} \leq 2\sqrt{\frac{\mu r}{m_1}}. \tag{109}$$

Inequality (107) taken together with (109) leads to

$$\begin{aligned}
\left\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty} &= \left\|\widetilde{\boldsymbol{u}}_{j_1}\widetilde{\boldsymbol{u}}_{j_1}^\top\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty} \\
&\leq \|\widetilde{\boldsymbol{u}}_{j_1}\|_\infty\left\|\widetilde{\boldsymbol{u}}_{j_1}^\top\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\| \\
&\leq 2\sqrt{\frac{\mu r}{m_1}}\|\boldsymbol{Z}\|^k \\
&\leq 2\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\widetilde{r}+1}^\star + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\right)^k. \tag{110}
\end{aligned}$$

33

Step 3.2: bounding $\|Z^i\widetilde{U}\|_{2,\infty}$. Turning to $\ell \geq 2$, we see from the triangle inequality that

$$\left\|\widetilde{P}_{j_1}Z\widetilde{P}_{j_2}Z\cdots Z\widetilde{P}_{j_{k+1}}\right\|_{2,\infty} \leq \left\|Z^{\ell-1}\widetilde{P}_{j_\ell}Z\cdots Z\widetilde{P}_{j_{k+1}}\right\|_{2,\infty} + \sum_{i=1}^{\ell-1}\left\|Z^{i-1}P_{\widetilde{U}}Z\widetilde{P}_{j_{i+1}}Z\cdots Z\widetilde{P}_{j_{k+1}}\right\|_{2,\infty}. \tag{111}$$

To bound the right-hand side of (111), it is helpful to bound $\|Z^i\widetilde{U}\|_{2,\infty}$ first.

Step 3.2.1: bounding $\|Z^i\widetilde{U}^{(1)}\|_{2,\infty}$. Recognizing that for any matrices $A, B \in \mathbb{R}^{m_1 \times m_1}$, we see that the following equation holds:

$$(A + B)^i = B^i + \sum_{j=0}^{i-1}B^j A (A + B)^{i-j-1}.$$

This allows one to derive

$$
\begin{aligned}
Z^i\widetilde{U}^{(1)} &= (Z_1 + Z_2 + Z_3)^i\,\widetilde{U}^{(1)} \\
&= Z_3^i\widetilde{U}^{(1)} + \sum_{j=0}^{i-1}Z_3^j Z_1 Z^{i-j-1}\widetilde{U}^{(1)} + \sum_{j=0}^{i-1}Z_3^j Z_2 Z^{i-j-1}\widetilde{U}^{(1)}.
\end{aligned}
\tag{112}
$$

By virtue of Lemma 5 and (107), one has

$$
\begin{aligned}
\left\|Z^i\widetilde{U}^{(1)}\right\|_{2,\infty} &\leq \left\|Z_3^i\widetilde{U}^{(1)}\right\|_{2,\infty} + \sum_{j=0}^{i-1}\left\|Z_3^j(Z_1 + Z_2)Z^{i-j-1}\widetilde{U}^{(1)}\right\|_{2,\infty} \\
&\leq \left\|Z_3^i\widetilde{U}^{(1)}\right\|_{2,\infty} + \sum_{j=0}^{i-1}\left(\left\|Z_3^j Z_1\right\|_{2,\infty} + \left\|Z_3^j Z_2\right\|_{2,\infty}\right)\left\|Z^{i-j-1}\widetilde{U}^{(1)}\right\| \\
&\leq \left\|Z_3^i\widetilde{U}^{(1)}\right\|_{2,\infty} + \sum_{j=0}^{i-1}\left(\left\|Z_3^j Z_1\right\|_{2,\infty} + \left\|Z_3^j Z_2\right\|_{2,\infty}\right)\left\|Z\right\|^{i-j-1} \\
&\leq 4C_3\sqrt{\frac{\mu r}{m_1}}\left(C_3\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^i \\
&\quad + \sum_{j=0}^{i-1}C_2\sqrt{\mu r}\left(\sigma_{\bar{r}+1}^\star + \sqrt{m_1}\omega_{\mathsf{max}}\log m\right)\left(C_3\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^j \omega_{\mathsf{max}}\log m\cdot \\
&\qquad\quad \left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\right)^{i-j-1} \\
&\leq 4C_3\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\right)^i \\
&\quad + \sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\right)^i\sum_{j=0}^{i-1}\frac{1}{2^j} \\
&\leq C_2\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\right)^i,
\end{aligned}
\tag{113}
$$

provided that $C_2 \geq 4C_3 + 1$.

Step 3.2.2: bounding $\|Z^i\widetilde{U}^{(2)}\|_{2,\infty}$. Note that $\widetilde{U}^{(2)}$ is also the column subspace of $\mathcal{P}_{(\widetilde{U}^{(1)})_\perp}U^{\star(2)} \in \mathbb{R}^{m_1, r-\bar{r}}$. In view of Lemma 5, (94), (98) and (113), we arrive at

$$
\begin{aligned}
\left\|Z^i\widetilde{U}^{(2)}\right\|_{2,\infty} &\leq \left\|Z^i\mathcal{P}_{(\widetilde{U}^{(1)})_\perp}U^{\star(2)}\right\|_{2,\infty}\sigma_{r-\bar{r}}^{-1}\left(\mathcal{P}_{(\widetilde{U}^{(1)})_\perp}U^{\star(2)}\right) \\
&\leq \frac{2}{\sqrt{3}}\left(\left\|Z^i U^{\star(2)}\right\|_{2,\infty} + \left\|Z^i\mathcal{P}_{\widetilde{U}^{(1)}}U^{\star(2)}\right\|_{2,\infty}\right)
\end{aligned}
$$

$$\leq \frac{2}{\sqrt{3}} \left(\big\| \boldsymbol{Z}^i \boldsymbol{U}^{\star(2)} \big\|_{2,\infty} + \big\| \boldsymbol{Z}^i \widetilde{\boldsymbol{U}}^{(1)} \big\|_{2,\infty} \big\| \widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)} \big\| \right)$$

$$\leq \frac{2}{\sqrt{3}} \left(3 C_3 \sqrt{\frac{\mu r}{m_1}} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i \right.$$

$$\left. + C_2 \sqrt{\frac{\mu r}{m_1}} \left(C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right) \right)^i \cdot \frac{1}{2} \right)$$

$$\leq C_2 \sqrt{\frac{\mu r}{m_1}} \left(C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right) \right)^i. \tag{114}$$

Putting (113) and (114) together and recognizing that $\widetilde{\boldsymbol{U}} = [\widetilde{\boldsymbol{U}}^{(1)} \ \widetilde{\boldsymbol{U}}^{(2)}]$, we conclude that

$$\big\| \boldsymbol{Z}^i \widetilde{\boldsymbol{U}} \big\|_{2,\infty} \leq \big\| \boldsymbol{Z}^i \widetilde{\boldsymbol{U}}^{(1)} \big\|_{2,\infty} + \big\| \boldsymbol{Z}^i \widetilde{\boldsymbol{U}}^{(2)} \big\|_{2,\infty}$$

$$\leq 2 C_2 \sqrt{\frac{\mu r}{m_1}} \left(C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right) \right)^i. \tag{115}$$

Step 4: bounding $\big\| \widetilde{\boldsymbol{P}}_{j_1} \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \big\|_{2,\infty}$ when $\ell > 1$. Plugging (107) and (115) into (111) yields that, for $\ell \geq 2$,

$$\big\| \widetilde{\boldsymbol{P}}_{j_1} \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \big\|_{2,\infty}$$

$$\leq \big\| \boldsymbol{Z}^{\ell-1} \widetilde{\boldsymbol{u}}_{j_\ell} \big\|_{2,\infty} \big\| \widetilde{\boldsymbol{u}}_{j_\ell}^\top \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \big\|_{2,\infty} + \sum_{i=1}^{\ell-1} \big\| \boldsymbol{Z}^{i-1} \widetilde{\boldsymbol{U}} \big\|_{2,\infty} \big\| \widetilde{\boldsymbol{U}}^\top \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{i+1}} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \big\|_{2,\infty}$$

$$\leq \big\| \boldsymbol{Z}^{\ell-1} \widetilde{\boldsymbol{U}} \big\|_{2,\infty} \|\boldsymbol{Z}\|^{k-\ell+1} + \sum_{i=1}^{\ell-1} \big\| \boldsymbol{Z}^{i-1} \widetilde{\boldsymbol{U}} \big\|_{2,\infty} \|\boldsymbol{Z}\|^{k-i+1}$$

$$\leq 2 C_2 \sqrt{\frac{\mu r}{m_1}} \left(C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right) \right)^{\ell-1} \cdot$$

$$\left(C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right) \right)^{k-\ell+1}$$

$$+ \sum_{i=1}^{\ell-1} 2 C_2 \sqrt{\frac{\mu r}{m_1}} \left(C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right) \right)^{i-1} \cdot$$

$$\left(C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right) \right)^{k-i+1}$$

$$= 2 C_2 \sqrt{\frac{\mu r}{m_1}} \left(C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right) \right)^k \cdot \ell$$

$$\leq 2 C_2 \sqrt{\frac{\mu r}{m_1}} \left(2 C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right) \right)^k.$$

The last inequality comes from $\ell \leq k+1 \leq 2^k$. Combining the previous inequality and (110) reveals that: for any $1 \leq k \leq \log n$ and $(j_1, \ldots, j_{k+1}) \in \{0, 1, \ldots, r\}^{k+1} \backslash \boldsymbol{0}$, it holds that

$$\big\| \widetilde{\boldsymbol{P}}_{j_1} \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \big\|_{2,\infty}$$

$$\leq 2 C_2 \sqrt{\frac{\mu r}{m_1}} \left(2 C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right) \right)^k. \tag{116}$$

Step 5: bounding $\|\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}} \boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{:,1:r'} \widetilde{\boldsymbol{U}}_{:,1:r'}^\top \|_{2,\infty}$. Now, we are ready to bound $\|\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}} \boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{:,1:r'} \widetilde{\boldsymbol{U}}_{:,1:r'}^\top \|_{2,\infty}$. As a consequence of (116), for any $1 \leq k \leq \log n$, one has

$$\left(\frac{8}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}} \right)^k \sum_{\substack{0 \leq j_1, \ldots, j_{k+1} \leq r \\ (j_1, \ldots, j_{k+1})^\top \neq \boldsymbol{0}_{k+1}}} \big\| \widetilde{\boldsymbol{P}}_{j_1} \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \big\|_{2,\infty}$$

$$\leq \left(\frac{8}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}\right)^k \cdot (r+1)^{k+1} \cdot 2C_2 \sqrt{\frac{\mu r}{m_1}} \left(2C_2 \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)\right)^k$$

$$\leq 4C_2 \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{32C_2 r \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}\right)^k. \tag{117}$$

Recalling that $\sigma_{r'}^{\star} \geq \sigma_{\bar{r}}^{\star} \geq C_0 r[(m_1 m_2)^{1/4} + r m_1^{1/2}]\log m$, we know from (104) that there exists some large constant $C > 0$ such that

$$\left(\frac{32C_2 r \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}\right)^{k-1}$$

$$\leq \left(\frac{32C_2 r \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)}{\sigma_{r'}^{\star 2}/(4r)}\right)^{k-1}$$

$$\leq \left(\frac{1}{C^2}\right)^{k-1} \leq \frac{1}{C^k}. \tag{118}$$

For any $k \geq \lfloor \log m \rfloor + 1$, in view of (107) and the the previous inequality, we have

$$\left(\frac{8}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}\right)^k \sum_{\substack{0 \leq j_1, \ldots, j_{k+1} \leq r \\ (j_1, \ldots, j_{k+1})^\top \neq \mathbf{0}_{k+1}}} \left\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}$$

$$\leq \left(\frac{8}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}\right)^k \cdot (r+1)^{k+1} \|\boldsymbol{Z}\|^k$$

$$\leq 2r \cdot \left(\frac{16rC_2 \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}\right)^k$$

$$\leq \frac{2r}{C^k} \cdot \frac{16rC_2 \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}. \tag{119}$$

Combining (108), (117), (118) and (119), one can obtain

$$\left\|\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}}\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^{\top}\right\|_{2,\infty}$$

$$\leq \sum_{1 \leq k \leq \log m} 4C_2 \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{32C_2 r \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}\right)^k$$

$$+ \sum_{k \geq \lfloor \log m \rfloor + 1} \frac{2r}{C^k} \cdot \frac{16C_2 r \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}$$

$$\lesssim \sqrt{\frac{\mu r^3}{m_1}} \frac{r \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma_{\bar{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)}{\sigma_{r'}^{\star 2} - \sigma_{r'+1}^{\star 2}}$$

$$\lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{r^2 \sqrt{m_1}\omega_{\mathsf{max}} \log m}{\sigma_{r'}^{\star}} + \frac{r^2 \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m}{\sigma_{r'}^{\star 2}}\right)$$

$$\asymp \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{r^2 \sqrt{m_1}\omega_{\mathsf{max}} \log m}{\sigma_{r'}^{\star}} + \frac{r^2 \sqrt{m_1 m_2}\omega_{\mathsf{max}}^2 \log^2 m}{\sigma_{r'}^{\star 2}}\right). \tag{120}$$

Here, the second last line holds due to (104) and the the last line makes use of the inequality

$$\frac{r^2 m_1 \omega_{\mathsf{max}}^2 \log^2 m}{\sigma_{r'}^{\star 2}} = r^2 \left(\frac{\sqrt{m_1}\omega_{\mathsf{max}} \log m}{\sigma_{r'}^{\star}}\right)^2 \lesssim r^2 \frac{\sqrt{m_1}\omega_{\mathsf{max}} \log m}{\sigma_{r'}^{\star}},$$

provided that $\sigma_{r'}^{\star} \gtrsim \sqrt{m_1}\omega_{\mathsf{max}} \log m$.

Step 6: bounding $\|\widetilde{U}_{:,1:r'}\widetilde{U}_{:,1:r'}^{\top} - U_{:,1:r'}^{\star}U_{:,1:r'}^{\star\top}\|_{2,\infty}$. To control $\|U_{:,1:r'}^{\mathsf{oracle}}U_{:,1:r'}^{\mathsf{oracle}\top} - U_{:,1:r'}^{\star}U_{:,1:r'}^{\star\top}\|_{2,\infty}$, one still needs to bound $\|\widetilde{U}_{:,1:r'}\widetilde{U}_{:,1:r'}^{\top} - U_{:,1:r'}^{\star}U_{:,1:r'}^{\star\top}\|_{2,\infty}$. Recall that $\widetilde{U}^{(1)}\widetilde{\Sigma}^{(1)}\widetilde{W}^{(1)\top}$ is the SVD of $U^{\star(1)}\Sigma^{\star(1)} + EV^{\star(1)} = U_{:,1:\overline{r}}^{\star}\Sigma_{1:\overline{r},1:\overline{r}}^{\star} + EV_{:,1:\overline{r}}^{\star}$ and $\widetilde{U}_{:,1:r'}$ (resp. $U_{:,1:r'}^{\star}$) is the matrix containing the first r' columns of $\widetilde{U}^{(1)}$ (resp. $U^{\star(1)}$). We make the observation that

$$
\begin{aligned}
&\mathcal{P}_{\left(\widetilde{U}_{:,1:r'}\right)_{\perp}} U_{:,1:r'}^{\star} \\
&= \mathcal{P}_{\left(\widetilde{U}_{:,1:r'}\right)_{\perp}} \left(U_{:,1:r'}^{\star}\Sigma_{1:r',1:\overline{r}}^{\star}\right)\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top}\left(\Sigma_{1:r',1:r'}^{\star}\right)^{-1} \\
&= \mathcal{P}_{\left(\widetilde{U}_{:,1:r'}\right)_{\perp}} \left(\widetilde{U}^{(1)}\widetilde{\Sigma}^{(1)}\widetilde{W}^{(1)\top} - EV^{\star(1)} - U_{:,r'+1:\overline{r}}^{\star}\Sigma_{r'+1:\overline{r},1:\overline{r}}^{\star}\right)\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top}\left(\Sigma_{1:r',1:r'}^{\star}\right)^{-1} \\
&= \mathcal{P}_{\left(\widetilde{U}_{:,1:r'}\right)_{\perp}} \left(\widetilde{U}_{:,1:r'}^{(1)}\widetilde{\Sigma}_{1:r',1:r'}^{(1)}\widetilde{W}_{:,1:r'}^{(1)\top} + \widetilde{U}_{:,r'+1:\overline{r}}^{(1)}\widetilde{\Sigma}_{r'+1:\overline{r},r'+1:\overline{r}}^{(1)}\widetilde{W}_{:,r'+1:\overline{r}}^{(1)\top} - EV^{\star(1)}\right)\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top} \\
&\quad \cdot \left(\Sigma_{1:r',1:r'}^{\star}\right)^{-1} \\
&= \widetilde{U}_{:,r'+1:\overline{r}}^{(1)}\widetilde{\Sigma}_{r'+1:\overline{r},r'+1:\overline{r}}^{(1)}\widetilde{W}_{:,r'+1:\overline{r}}^{(1)\top}\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top}\left(\Sigma_{1:r',1:r'}^{\star}\right)^{-1} \\
&\quad - \mathcal{P}_{\left(\widetilde{U}_{:,1:r'}\right)_{\perp}} EV^{\star(1)}\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top}\left(\Sigma_{1:r',1:r'}^{\star}\right)^{-1},
\end{aligned}
\tag{121}
$$

where the second identity is valid since

$$
\Sigma_{r'+1:\overline{r},1:\overline{r}}^{\star}\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top} = \left(0_{(\overline{r}-r')\times r'}\; \Sigma_{r'+1:\overline{r},r'+1:\overline{r}}^{\star}\right)\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top} = 0_{(\overline{r}-r')\times r'}
$$

and the last line comes from $\mathcal{P}_{\left(\widetilde{U}_{:,1:r'}\right)_{\perp}}\widetilde{U}_{:,1:r'}^{(1)} = 0$ and $\mathcal{P}_{\left(\widetilde{U}_{:,1:r'}\right)_{\perp}}\widetilde{U}_{:,r'+1:\overline{r}}^{(1)} = \widetilde{U}_{:,r'+1:\overline{r}}^{(1)}$. Note that $\widetilde{W}_{:,1:r'}^{(1)}$ (resp. $\left(I_{r'}\; 0\right)^{\top}$) is the leading r' right singular space of $U^{\star(1)}\Sigma^{\star(1)} + EV^{\star(1)}$ (resp. $U^{\star(1)}\Sigma^{\star(1)}$) and

$$
\sigma_{r'}^{\star} - \sigma_{r'+1}^{\star} \geq \frac{1}{4r}\sigma_{r'}^{\star} \geq \frac{C_0}{4}\left[(m_1 m_2)^{1/4} + r m_1^{1/2}\right]\omega_{\max}\log m \gg r\sqrt{m_1}\omega_{\max}\log m.
$$

Chen et al. (2021a, Lemma 2.6, Eqn. (2.26a)) and Lemma 4 taken together imply that

$$
\left\|\widetilde{W}_{:,r'+1:\overline{r}}^{(1)\top}\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top}\right\| = \left\|\left(\widetilde{W}_{:,1:r'}\right)_{\perp}^{(1)\top}\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top}\right\| \lesssim \frac{\|EV^{\star(1)}\|}{\sigma_{r'}^{\star} - \sigma_{r'+1}^{\star}} \lesssim \frac{r\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r'}^{\star}},
\tag{122a}
$$

$$
\left\|\left(U_{:,1:r'}^{\star}\right)_{\perp}^{\top}\widetilde{U}_{:,1:r'}\right\| \lesssim \frac{\|EV^{\star(1)}\|}{\sigma_{r'}^{\star} - \sigma_{r'+1}^{\star}} \lesssim \frac{r\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r'}^{\star}}.
\tag{122b}
$$

Moreover, combining (90a) and the assumption $\sigma_{\overline{r}}^{\star} \geq C_0 r[(m_1 m_2)^{1/4} + r m_1^{1/2}]\omega_{\max}\log m$ gives

$$
\widetilde{\sigma}_i \leq \sigma_i^{\star} + \|EV^{\star}\| \leq 2\sigma_i^{\star}, \qquad \forall i \in [\overline{r}].
\tag{123}
$$

Inequality (90d) combined with (122a) and (123) gives

$$
\begin{aligned}
&\left\|\widetilde{U}_{:,r'+1:\overline{r}}^{(1)}\widetilde{\Sigma}_{r'+1:\overline{r},r'+1:\overline{r}}^{(1)}\widetilde{W}_{:,r'+1:\overline{r}}^{(1)\top}\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top}\left(\Sigma_{1:r',1:r'}^{\star}\right)^{-1}\right\|_{2,\infty} \\
&\qquad \leq \left\|\widetilde{U}^{(1)}\right\|_{2,\infty}\left\|\widetilde{\Sigma}_{r'+1:\overline{r},r'+1:\overline{r}}^{(1)}\right\|\left\|\widetilde{W}_{:,r'+1:\overline{r}}^{(1)\top}\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top}\right\|\left\|\left(\Sigma_{1:r',1:r'}^{\star}\right)^{-1}\right\| \\
&\qquad \lesssim 2\sqrt{\frac{\mu r}{m_1}}\cdot\widetilde{\sigma}_{r'+1}\cdot\frac{r\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r'}^{\star}}\cdot\frac{1}{\sigma_{r'}^{\star}} \\
&\qquad \lesssim 2\sqrt{\frac{\mu r}{m_1}}\cdot 2\sigma_{r'+1}^{\star}\cdot\frac{r\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r'}^{\star}}\cdot\frac{1}{\sigma_{r'}^{\star}} \\
&\qquad \lesssim \sqrt{\frac{\mu r}{m_1}}\frac{r\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r'}^{\star}}.
\end{aligned}
\tag{124}
$$

In addition, Lemma 2 and (90d) taken together imply that

$$
\left\|\mathcal{P}_{\left(\widetilde{U}_{:,1:r'}\right)_{\perp}} EV^{\star(1)}\left(I_{r'}\; 0_{r'\times(\overline{r}-r')}\right)^{\top}\left(\Sigma_{1:r',1:r'}^{\star}\right)^{-1}\right\|_{2,\infty}
$$

$$\leq \frac{1}{\sigma_{r'}^\star} \left\| \mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} EV^{\star(1)} \right\|_{2,\infty}$$

$$\leq \frac{1}{\sigma_{r'}^\star} \left(\left\| EV^{\star(1)} \right\|_{2,\infty} + \left\| \mathcal{P}_{\widetilde{U}_{:,1:r'}} EV^{\star(1)} \right\|_{2,\infty} \right)$$

$$\leq \frac{1}{\sigma_{r'}^\star} \left(\left\| EV^{\star(1)} \right\|_{2,\infty} + \left\| \widetilde{U}_{:,1:r'} \right\|_{2,\infty} \left\| \widetilde{U}_{:,1:r'} \right\| \left\| EV^{\star(1)} \right\| \right)$$

$$\lesssim \frac{1}{\sigma_{r'}^\star} \left(\sqrt{\mu r}\, \omega_{\mathsf{max}} \log n + 2\sqrt{\frac{\mu r}{m_1}} \cdot \sqrt{m_1}\, \omega_{\mathsf{max}} \log n \right)$$

$$\asymp \sqrt{\frac{\mu r}{m_1}} \frac{\sqrt{m_1}\, \omega_{\mathsf{max}} \log m}{\sigma_{r'}^\star}. \tag{125}$$

Taking (121), (124) and (125) together gives

$$\left\| \mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} U_{:,1:r'}^\star \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r}{m_1}} \frac{r\sqrt{m_1}\, \omega_{\mathsf{max}} \log m}{\sigma_{r'}^\star}.$$

The previous inequality taken together with (122b) and (90d) reveals that

$$\left\| \widetilde{U}_{:,1:r'} \widetilde{U}_{:,1:r'}^\top - U_{:,1:r'}^\star U_{:,1:r'}^{\star\top} \right\|_{2,\infty}$$

$$\leq \left\| (U_{:,1:r'}^\star - \widetilde{U}_{:,1:r'} \widetilde{U}_{:,1:r'}^\top U_{:,1:r'}^\star) U_{:,1:r'}^{\star\top} \right\|_{2,\infty} + \left\| \widetilde{U}_{:,1:r'} \widetilde{U}_{:,1:r'}^\top U_{:,1:r'}^\star U_{:,1:r'}^{\star\top} - \widetilde{U}_{:,1:r'} \widetilde{U}_{:,1:r'}^\top \right\|_{2,\infty}$$

$$\leq \left\| \left(\mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} U_{:,1:r'}^\star \right) U_{:,1:r'}^{\star\top} \right\|_{2,\infty} + \left\| \widetilde{U}_{:,1:r'} \widetilde{U}_{:,1:r'}^\top (U_{:,1:r'}^\star)_\perp (U_{:,1:r'}^\star)_\perp^\top \right\|_{2,\infty}$$

$$\leq \left\| \mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} U_{:,1:r'}^\star \right\|_{2,\infty} + \left\| \widetilde{U}_{:,1:r'} \right\|_{2,\infty} \left\| (U_{:,1:r'}^\star)_\perp^\top \widetilde{U}_{:,1:r'} \right\|$$

$$\lesssim \sqrt{\frac{\mu r}{m_1}} \frac{r\sqrt{m_1}\, \omega_{\mathsf{max}} \log m}{\sigma_{r'}^\star} + 2\sqrt{\frac{\mu r}{m_1}} \frac{r\sqrt{m_1}\, \omega_{\mathsf{max}} \log m}{\sigma_{r'}^\star}$$

$$\asymp \sqrt{\frac{\mu r}{m_1}} \frac{r\sqrt{m_1}\, \omega_{\mathsf{max}} \log m}{\sigma_{r'}^\star}. \tag{126}$$

By virtue of (120) and (126), we arrive at

$$\left\| U_{:,1:r'}^{\mathsf{oracle}} U_{:,1:r'}^{\mathsf{oracle}\top} - U_{:,1:r'}^\star U_{:,1:r'}^{\star\top} \right\|_{2,\infty}$$

$$\leq \left\| U_{:,1:r'}^{\mathsf{oracle}} U_{:,1:r'}^{\mathsf{oracle}\top} - \widetilde{U}_{:,1:r'} \widetilde{U}_{:,1:r'}^\top \right\|_{2,\infty} + \left\| \widetilde{U}_{:,1:r'} \widetilde{U}_{:,1:r'}^\top - U_{:,1:r'}^\star U_{:,1:r'}^{\star\top} \right\|_{2,\infty}$$

$$\lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{r^2 \sqrt{m_1}\, \omega_{\mathsf{max}} \log m}{\sigma_{r'}^\star} + \frac{r^2 \sqrt{m_1 m_2}\, \omega_{\mathsf{max}}^2 \log^2 m}{\sigma_{r'}^{\star 2}} \right). \tag{127}$$

$\square$

C.3   Proof of Lemma 1

Denote by γ_1 the following counterclockwise contour on the complex plane:

$$\gamma_1 = \Bigg\{ x + y\mathrm{i} : x = \frac{\overline{\lambda}_{r_1} + \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{r_1} + \overline{\lambda}_{r_1+1}}{2},$$

$$\text{or } x = \overline{\lambda}_1 + \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{r_1} + \overline{\lambda}_{r_1+1}}{2},$$

$$\text{or } y = \pm \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2}, \frac{\overline{\lambda}_{r_1} + \overline{\lambda}_{r_1+1}}{2} \leq x \leq \overline{\lambda}_1 + \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2} \Bigg\}.$$

Then $\{\overline{\lambda}_i\}_{i=1}^{r_1}$ lie inside the contour γ_1 and $\{\overline{\lambda}_i\}_{i=r_1+1}^n$ (where $\overline{\lambda}_i = 0$ for $i \geq r+1$) reside outside γ_1. Moreover, for any $\eta \in \gamma_1$ and $1 \leq i \leq n$, one has

$$\left| \eta - \overline{\lambda}_i \right| \leq \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2}. \tag{128}$$

Step 1: decompose $U_1 U_1^\top - \overline{U}_1 \overline{U}_1^\top$. First, we invoke a similar argument as in Xia (2021, Theorem 1) to express $U_1 U_1^\top - \overline{U}_1 \overline{U}_1^\top$ as an infinite sum. Denote by $\lambda_1 \geq \cdots \geq \lambda_n$ the eigenvalues of M. Apply Weyl's inequality to obtain

$$\max_{1 \leq i \leq r} \left| \lambda_i - \overline{\lambda}_i \right| \leq \|Z\| < \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2}.$$

As a result, we know that $\{\lambda_i\}_{i=1}^{r_1}$ are inside the contour γ_1, and $\{\lambda_i\}_{i=r_1+1}^{n}$ are outside the contour. Similar to Eqn. (10) in Xia (2021), one has

$$U_1 U_1^\top = \frac{1}{2\pi i} \oint_{\gamma_1} (\eta I - M)\, d\eta. \tag{129}$$

We define

$$\mathcal{R}_{\overline{M}}(\eta) := \left(\eta I - \overline{M}\right)^{-1} = \sum_{i=1}^{n} \frac{1}{\eta - \overline{\lambda}_j} \overline{u}_j \overline{u}_j^\top.$$

In view of (128), for any $\eta \in \gamma_1$, we have

$$\left\| \mathcal{R}_{\overline{M}}(\eta) \right\| \leq \frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}},$$

and consequently,

$$\left\| \mathcal{R}_{\overline{M}}(\eta) Z \right\| \leq \left\| \mathcal{R}_{\overline{M}}(\eta) \right\| \|Z\| \leq \frac{2\|Z\|}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} < 1.$$

A similar argument in Xia (2021, Eqn. (13)) yields

$$
\begin{aligned}
U_1 U_1^\top - \overline{U}_1 \overline{U}_1^\top &= \sum_{k \geq 1} \frac{1}{2\pi i} \oint_{\gamma_1} \left[\mathcal{R}_{\overline{M}}(\eta) Z \right]^k \mathcal{R}_{\overline{M}}(\eta)\, d\eta \\
&= \sum_{k \geq 1} \sum_{1 \leq j_1, \ldots, j_{k+1} \leq n} \frac{1}{2\pi i} \oint_{\gamma_1} \frac{d\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} P_{\overline{u}_{j_1}} Z P_{\overline{u}_{j_2}} Z \cdots P_{\overline{u}_{j_k}} Z P_{\overline{u}_{j_{k+1}}} \\
&= \sum_{k \geq 1} \sum_{0 \leq j_1, \ldots, j_{k+1} \leq r} \frac{1}{2\pi i} \oint_{\gamma_1} \frac{d\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \overline{P}_{j_1} Z \overline{P}_{j_2} Z \cdots \overline{P}_{j_k} Z \overline{P}_{j_{k+1}}. \tag{130}
\end{aligned}
$$

Here, we define $\overline{\lambda}_0 = 0$ and the last line holds since $\overline{\lambda}_i = 0$ for all $i \geq r+1$ and $\overline{P}_0 = \overline{U}_\perp \overline{U}_\perp^\top = \sum_{i=r+1}^{n} \overline{u}_i \overline{u}_i^\top$.

Step 2: bounding $\|U_1 U_1^\top - \overline{U}_1 \overline{U}_1^\top\|_{2,\infty}$. By virtue of (130) and the triangle inequality, we see that: to prove (87a), it suffices to bound $\left| \frac{1}{2\pi i} \oint_{\gamma_1} \frac{d\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \right|$. We consider two scenarios: all of $j_1, \ldots, j_{k+1}$ lie in the set $\{0\} \cup \{r_1 + 1, \ldots, r_1\}$, and at least one of $j_1, \ldots, j_{k+1}$ is in the set $\{1, \ldots, r_1\}$.

Case 1: all of $j_1, \ldots, j_{k+1}$ are either 0 or larger than r_1. In this case, none of $\overline{\lambda}_{j_1}, \ldots, \overline{\lambda}_{j_{k+1}}$ is inside γ_1 and as a result, $f(\eta) = \frac{1}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}$ is analytic within and on γ_1. Cauchy's integral theorem tells us that

$$\frac{1}{2\pi i} \oint_{\gamma_1} \frac{d\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} = 0, \tag{131}$$

and thus we have

$$\frac{1}{2\pi i} \oint_{\gamma_1} \frac{d\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \overline{P}_{j_1} Z \overline{P}_{j_2} Z \cdots \overline{P}_{j_k} Z \overline{P}_{j_{k+1}} = 0. \tag{132}$$

Case 2: at least one of $j_1, \ldots, j_{k+1}$ is between 1 and r_1. Let

$$\mathcal{J} = \{j : 1 \le j \le r_1, \exists 1 \le \ell \le k+1 \text{ s.t. } j_\ell = j\}, \tag{133}$$

and let j_{max} and j_{min} denote the largest and smallest elements in $\mathcal{J}$, respectively. We define the following counterclockwise rectangular contour:

$$
\gamma_2 = \left\{ x + y\mathrm{i} : x = \overline{\lambda}_{j_{\mathsf{max}}} - \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2}, \; -\frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2} \le y \le \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2}, \right.
$$
$$
\text{or } x = \overline{\lambda}_{j_{\mathsf{min}}} + \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2}, \; -\frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2} \le y \le \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2},
$$
$$
\left. \text{or } y = \pm \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2}, \; \overline{\lambda}_{j_{\mathsf{max}}} - \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2} \le x \le \overline{\lambda}_{j_{\mathsf{min}}} + \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2} \right\}.
$$

It is easy to verify that

$$|\eta - \lambda_{j_\ell}| \ge \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2}, \qquad \forall 1 \le \ell \le k+1, \eta \in \gamma_2 \tag{134}$$

and

$$\frac{1}{2\pi\mathrm{i}} \oint_{\gamma_1} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} = \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_2} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}. \tag{135}$$

Moreover, the length of γ_2 is

$$L(\gamma_2) = 2\left(\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{max}}}\right) + 4\left(\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}\right).$$

If $j_{\mathsf{max}} = j_{\mathsf{min}}$, i.e., there is only one element in $\mathcal{I}$, then applying the triangle inequality for contour integrals yields

$$
\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_2} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \right| \le \frac{1}{2\pi} \sup_{\eta \in \gamma_2} \left| \frac{1}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \right| \cdot L(\gamma_2)
$$
$$
\overset{(134)}{\le} \frac{1}{2\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^{k+1} \cdot 4\left(\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}\right)
$$
$$
\le \frac{4}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^{k}.
$$

If $j_{\mathsf{max}} \ne j_{\mathsf{min}}$, then the triangle inequality tells us that for any $\eta \in \gamma_2$,

$$\max\left\{ \left| \eta - \overline{\lambda}_{j_{\mathsf{max}}} \right|, \left| \eta - \overline{\lambda}_{j_{\mathsf{min}}} \right| \right\} \ge \frac{\left| (\eta - \overline{\lambda}_{j_{\mathsf{max}}}) - (\eta - \overline{\lambda}_{j_{\mathsf{min}}}) \right|}{2} = \frac{\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{max}}}}{2}. \tag{136}$$

In view of (134), (136) and the basic inequality $\min\{\frac{a}{b}, \frac{c}{d}\} \le \frac{a+c}{b+d}$ for $a, b, c, d > 0$, one has

$$
\frac{1}{\left| (\eta - \overline{\lambda}_{j_{\mathsf{max}}})(\eta - \overline{\lambda}_{j_{\mathsf{min}}}) \right|} \le \min\left\{ \frac{4}{(\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{max}}})(\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1})}, \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^2 \right\}
$$
$$
\le \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right) \cdot \left(\frac{4}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1} + \overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{j_{\mathsf{min}}}} \right) \tag{137}
$$

for all $\eta \in \gamma_2$, and consequently, one has

$$\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_2} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \right| \le \frac{1}{2\pi} \sup_{\eta \in \gamma_2} \left| \frac{1}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \right| \cdot L(\gamma_2)$$

40

$$\leq \frac{1}{2\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right) \cdot \left(\frac{4}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1} + \overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\min}}} \right)$$

$$\cdot \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^{k-1} \cdot \left(2 \left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}} \right) + 4 \left(\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1} \right) \right)$$

$$\leq \frac{8}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^{k}.$$

Therefore, it is guaranteed that

$$\left| \frac{1}{2\pi \mathrm{i}} \oint_{\gamma_2} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \right| \leq \frac{8}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^{k}. \tag{138}$$

Combining (130), (132), (138) and the triangle inequality finishes the proof of (87a).

Step 3: bounding $\|(\overline{U}_1 \overline{U}_1^\top - U_1 U_1^\top) \overline{M}\|_{2,\infty}$. Next, we move on to control $\|(\overline{U}_1 \overline{U}_1^\top - U_1 U_1^\top) \overline{M}\|_{2,\infty}$. By virtue of (130), we have

$$\left(U_1 U_1^\top - \overline{U}_1 \overline{U}_1^\top \right) \overline{M}$$

$$= \sum_{k \geq 1} \sum_{0 \leq j_1, \ldots, j_k \leq r} \sum_{j_{k+1}=1}^{r} \frac{1}{2\pi \mathrm{i}} \oint_{\gamma_1} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \overline{P}_{j_1} Z \overline{P}_{j_2} Z \cdots \overline{P}_{j_k} Z \overline{P}_{j_{k+1}} \overline{M}$$

$$= \sum_{k \geq 1} \sum_{0 \leq j_1, \ldots, j_k \leq r} \sum_{j_{k+1}=1}^{r} \frac{1}{2\pi \mathrm{i}} \oint_{\gamma_1} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \overline{P}_{j_1} Z \overline{P}_{j_2} Z \cdots \overline{P}_{j_k} Z \overline{P}_{j_{k+1}}. \tag{139}$$

The second line and the third line make use of $\overline{U}_\perp \overline{M} = 0$ and $\overline{P}_j \overline{M} = \overline{u}_j \overline{u}_j^\top (\sum_{i=1}^{r} \overline{\lambda}_i \overline{u}_i \overline{u}_i^\top) = \overline{\lambda}_j$, respectively. In the rest of the proof, we will establish upper bounds for $|\frac{1}{2\pi \mathrm{i}} \oint_{\gamma_1} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}|$ and justify the validity of (87b).

Step 3.1: bounding $|\frac{1}{2\pi \mathrm{i}} \oint_{\gamma_1} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}|$ **for** $k = 1$. We consider two scenarios: (1) $1 \leq j_1, j_2 \leq r_1$ or $j_1, j_2 \in \{r_1+1, \ldots, r\} \cup \{0\}$ and (2) only one of j_1 and j_2 falls into the set $\{1, \ldots, r_1\}$.

Case 1: $1 \leq j_1, j_2 \leq r_1$ **or** $j_1, j_2 \in \{r_1+1, \ldots, r\} \cup \{0\}$. In this case, Cauchy's integral formula asserts that

$$\frac{1}{2\pi \mathrm{i}} \oint_{\gamma_1} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1})(\eta - \overline{\lambda}_{j_2})} = 0,$$

and consequently,

$$\left| \frac{1}{2\pi \mathrm{i}} \oint_{\gamma_1} \frac{\overline{\lambda}_{j_2} \mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1})(\eta - \overline{\lambda}_{j_2})} \right| = 0. \tag{140}$$

Case 2: exactly one of $j_i \in \{1, \ldots, r_1\}$. Apply Cauchy's integral formula to yield

$$\left| \frac{1}{2\pi \mathrm{i}} \oint_{\gamma_1} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1})(\eta - \overline{\lambda}_{j_2})} \right| = \left| \frac{1}{2\pi \mathrm{i}} \oint_{\gamma_1} \frac{\frac{1}{\eta - \overline{\lambda}_{j_1}}}{\eta - \overline{\lambda}_{j_2}} \mathrm{d}\eta \right| = \frac{1}{|\overline{\lambda}_{j_1} - \overline{\lambda}_{j_2}|}.$$

Observing that the function $f(x) = \frac{1}{|x-1|}$ is increasing on $(-\infty, 1)$ and decreasing on $(1, \infty)$, one has

$$\left| \frac{1}{2\pi \mathrm{i}} \oint_{\gamma_1} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1})(\eta - \overline{\lambda}_{j_2})} \right| = \frac{\overline{\lambda}_{j_2}}{|\overline{\lambda}_{j_1} - \overline{\lambda}_{j_2}|} = \frac{1}{\left| \frac{\overline{\lambda}_{j_1}}{\overline{\lambda}_{j_2}} - 1 \right|} \leq \frac{1}{\max \left\{ \frac{\overline{\lambda}_{r_1}}{\overline{\lambda}_{r_1+1}} - 1, 1 - \frac{\overline{\lambda}_{r_1+1}}{\overline{\lambda}_{r_1}} \right\}} = \frac{\overline{\lambda}_{r_1}}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}. \tag{141}$$

41

Step 3.2: bounding $\left|\frac{1}{2\pi i}\oint_{\gamma_1}\frac{\overline{\lambda}_{j_{k+1}}d\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\right|$ **for** $k>1$**.** If (1) all $j_1,\ldots,j_{k+1}$ lie in the set $\{1,\ldots,r_1\}$ or (2) all $j_1,\ldots,j_{k+1}$ are all in the set $\{0\}\cup\{r_1+1,\ldots,r\}$, then the function

$$g(x)=\frac{\overline{\lambda}_{j_{k+1}}}{\left(\eta-\overline{\lambda}_{j_1}\right)\cdots\left(\eta-\overline{\lambda}_{j_{k+1}}\right)}$$

is analytic in and on γ_1 or outside on on γ_1. As a result, Cauchy's integral formula tells us that

$$\left|\frac{1}{2\pi i}\oint_{\gamma_1}\frac{\overline{\lambda}_{j_{k+1}}d\eta}{\left(\eta-\overline{\lambda}_{j_1}\right)\cdots\left(\eta-\overline{\lambda}_{j_{k+1}}\right)}\right|=0. \tag{142}$$

In the following proof, we assume that these two cases would not happen, i.e.,

$$1\leq|\{j_1,\ldots,j_{k+1}\}\cap\{1,\ldots,r_1\}|\leq k. \tag{143}$$

Let γ_3 denote the following counterclockwise rectangular contour:

$$\gamma_3=\left\{x+y i:x=\frac{\overline{\lambda}_{j_{\max}}+\overline{\lambda}_{r_1+1}}{2},-\frac{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}{2}\leq y\leq\frac{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}{2},\right.$$

$$\text{or }x=\overline{\lambda}_{j_{\min}}+\frac{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}{2},-\frac{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}{2}\leq y\leq\frac{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}{2},$$

$$\left.\text{or }y=\pm\frac{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}{2},\frac{\overline{\lambda}_{j_{\max}}+\overline{\lambda}_{r_1+1}}{2}\leq x\leq\overline{\lambda}_{j_{\min}}+\frac{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}{2}\right\},$$

where we recall that $j_{\min}$ (resp. $j_{\max}$) is the smallest (resp. largest) element in the set $\mathcal{J}$ defined in (133). Then one can check that

$$\left|\eta-\overline{\lambda}_{j_\ell}\right|\geq\frac{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}{2},\qquad\forall 1\leq\ell\leq k+1,\eta\in\gamma_3, \tag{144}$$

and the length of γ_3 satisfies

$$L(\gamma_3)=2\left(\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_{\max}}\right)+4\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)=2\overline{\lambda}_{j_{\min}}+2\overline{\lambda}_{j_{\max}}-4\overline{\lambda}_{r_1+1}. \tag{145}$$

In addition, we have

$$\frac{1}{2\pi i}\oint_{\gamma_1}\frac{d\eta}{\left(\eta-\overline{\lambda}_{j_1}\right)\cdots\left(\eta-\overline{\lambda}_{j_{k+1}}\right)}=\frac{1}{2\pi i}\oint_{\gamma_3}\frac{d\eta}{\left(\eta-\overline{\lambda}_{j_1}\right)\cdots\left(\eta-\overline{\lambda}_{j_{k+1}}\right)}. \tag{146}$$

Case 1: $\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_{\max}}\leq 3(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1})$**.** In this scenario, one has

$$\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}=\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_{\max}}+\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\leq 4\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right), \tag{147}$$

which further leads to

$$L(\gamma_3)\leq 10\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right).$$

In view of (144), (146), (147) and the previous inequality, one has

$$\left|\frac{1}{2\pi i}\oint_{\gamma_1}\frac{\overline{\lambda}_{j_{k+1}}d\eta}{\left(\eta-\overline{\lambda}_{j_1}\right)\cdots\left(\eta-\overline{\lambda}_{j_{k+1}}\right)}\right|\leq\frac{1}{2\pi}\overline{\lambda}_{j_{k+1}}\left(\frac{2}{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}\right)^{k+1}L(\gamma_3)$$

$$\leq\frac{1}{2\pi}\overline{\lambda}_{j_{\min}}\left(\frac{2}{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}\right)^{k+1}\cdot 10\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)$$

$$\leq\frac{80}{\pi}\frac{\overline{\lambda}_{j_{\min}}}{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}}\left(\frac{2}{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}\right)^{k-1}$$

$$\leq\frac{80}{\pi}\frac{\overline{\lambda}_{r_1}}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\left(\frac{2}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\right)^{k-1}=\frac{40}{\pi}\overline{\lambda}_{r_1}\left(\frac{2}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\right)^k. \tag{148}$$

The third inequality holds because of (147), whereas the last one applies the monotonicity of the function $f(x)=\frac{x}{x-\overline{\lambda}_{r_1+1}}$ for $x>\overline{\lambda}_{r_1+1}$ and the inequality $\overline{\lambda}_{j_{\min}}\geq\overline{\lambda}_{j_{\max}}\geq\overline{\lambda}_{r_1}$.

Case 2: $\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}} > 3(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1})$. Denote the following two disjoint sets

$$\mathcal{I}_1 = \left\{ i : 1 \le i \le k+1, \overline{\lambda}_{j_i} \ge \frac{2}{3}\overline{\lambda}_{j_{\min}} + \frac{1}{3}\overline{\lambda}_{j_{\max}} \right\} \tag{149}$$

and

$$\mathcal{I}_2 = \left\{ i : 1 \le i \le k+1, \overline{\lambda}_{j_i} \le \frac{1}{3}\overline{\lambda}_{j_{\min}} + \frac{2}{3}\overline{\lambda}_{j_{\max}} \right\}. \tag{150}$$

By the definition of $j_{\min}$ and $j_{\max}$, we know that $\mathcal{I}_1$ and $\mathcal{I}_2$ are nonempty. We consider the following three scenarios: (1) $\min\{|\mathcal{I}_1|, |\mathcal{I}_2|\} \ge 2$; (2) $|\mathcal{I}_1| = 1$ and (3) $|\mathcal{I}_2| = 1$.

Case 2.1: $\min\{|\mathcal{I}_1|, |\mathcal{I}_2|\} \ge 2$. When it comes to this case, one can find four different indices i_1, i_2, i_3, i_4 such that $i_1, i_3 \in \mathcal{I}_1$, $i_2, i_4 \in \mathcal{I}_2$, $\overline{\lambda}_{j_{i_1}} = \overline{\lambda}_{j_{\min}}$ and $\overline{\lambda}_{j_{i_2}} = \overline{\lambda}_{j_{\max}}$. Then the triangle inequality tells us that

$$\max\left\{ \left|\eta - \overline{\lambda}_{j_{i_1}}\right|, \left|\eta - \overline{\lambda}_{j_{i_2}}\right| \right\} \ge \frac{1}{2}\left(\overline{\lambda}_{j_{i_1}} - \overline{\lambda}_{j_{i_2}}\right) = \frac{1}{2}\left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}}\right)$$

and

$$\max\left\{ \left|\eta - \overline{\lambda}_{j_{i_3}}\right|, \left|\eta - \overline{\lambda}_{j_{i_4}}\right| \right\} \ge \frac{1}{2}\left(\overline{\lambda}_{j_{i_3}} - \overline{\lambda}_{j_{i_4}}\right) \ge \frac{1}{6}\left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}}\right).$$

Similar to (137), one can derive

$$\left| \frac{1}{\left(\eta - \overline{\lambda}_{j_{i_1}}\right)\left(\eta - \overline{\lambda}_{j_{i_2}}\right)\left(\eta - \overline{\lambda}_{j_{i_3}}\right)\left(\eta - \overline{\lambda}_{j_{i_4}}\right)} \right|$$

$$= \frac{1}{\left|\eta - \overline{\lambda}_{j_{i_1}}\right|\left|\eta - \overline{\lambda}_{j_{i_2}}\right|} \frac{1}{\left|\eta - \overline{\lambda}_{j_{i_3}}\right|\left|\eta - \overline{\lambda}_{j_{i_4}}\right|}$$

$$= \min\left\{ \frac{4}{\left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}}\right)\left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right)}, \frac{4}{\left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right)^2} \right\}$$

$$\quad \cdot \min\left\{ \frac{12}{\left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}}\right)\left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right)}, \frac{4}{\left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right)^2} \right\}$$

$$\le \frac{8}{\left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right)\left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}\right)} \cdot \frac{16}{\left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right)\left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}\right)}$$

$$= \frac{128}{\left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right)^2\left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}\right)^2}.$$

The penultimate line uses the basic inequality $\min\{a/b, c/d\} \le (a+c)/(b+d)$ for $a, b, c, d > 0$. Putting the previous inequality, (144), (145) and (123) together, we reach

$$\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_1} \frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1}\right)\cdots\left(\eta - \overline{\lambda}_{j_{k+1}}\right)} \right|$$

$$\le \frac{1}{2\pi} \sup_{\eta \in \gamma_3} \left| \frac{\overline{\lambda}_{j_{k+1}}}{\left(\eta - \overline{\lambda}_{j_1}\right)\cdots\left(\eta - \overline{\lambda}_{j_{k+1}}\right)} \right| \cdot L(\gamma_3)$$

$$\le \frac{1}{2\pi}\overline{\lambda}_{j_{k+1}} \frac{128}{\left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right)^2\left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}\right)^2}\left(\frac{2}{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}} \right)^{k+1-4} \cdot 4\left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}\right)$$

$$\le \frac{64}{\pi} \frac{\overline{\lambda}_{j_{\min}}}{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}}\left(\frac{2}{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}} \right)^{k-1}$$

43

$$\leq \frac{64}{\pi} \frac{\overline{\lambda}_{r_1}}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \left(\frac{2}{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}} \right)^{k-1}$$

$$\leq \frac{32}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^{k} \overline{\lambda}_{r_1}. \tag{151}$$

The fourth line is due to $\overline{\lambda}_{j_{k+1}} \leq \overline{\lambda}_{j_{\min}}$ and the fifth line holds since $f(x) = \frac{x}{x - \overline{\lambda}_{r_1+1}}$ is a decreasing function on $(\overline{\lambda}_{r_1+1}, \infty)$.

Case 2.2: $|\mathcal{I}_1| = 1$. Let us choose

$$\ell \in \underset{i:1 \leq i \leq k+1, i \notin \mathcal{I}_1}{\arg \max} \overline{\lambda}_{j_i}. \tag{152}$$

We can see from the definition of $\mathcal{I}_1$ that

$$\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_\ell} \geq \overline{\lambda}_{j_{\min}} - \left(\frac{2}{3} \overline{\lambda}_{j_{\min}} + \frac{1}{3} \overline{\lambda}_{j_{\max}} \right) = \frac{1}{3} \left(\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}} \right) > \overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}.$$

The last inequality is valid since $\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}} > 3(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1})$. We let γ_4 and γ_5 denote the following counterclockwise contours:

$$\gamma_4 = \left\{ x + y\mathrm{i} : x = \frac{\overline{\lambda}_{j_{\max}} + \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, \right.$$

$$\text{or } x = \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2},$$

$$\left. \text{or } y = \pm \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, \frac{\overline{\lambda}_{j_{\max}} + \overline{\lambda}_{r_1+1}}{2} \leq x \leq \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \right\}$$

and

$$\gamma_5 = \left\{ x + y\mathrm{i} : x = \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, \right.$$

$$\text{or } x = \overline{\lambda}_{j_{\min}} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2},$$

$$\left. \text{or } y = \pm \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq x \leq \overline{\lambda}_{j_{\min}} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \right\}.$$

For any complex number $\eta = x + y\mathrm{i}$ with $x = \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}$ and $-\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}$, we know that $g(\eta) = \frac{1}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \neq 0$, and thus $g(\eta)$ is analytic on $\{\eta = x + y\mathrm{i} : x = \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}\}$. Applying Cauchy's integral formula yields

$$\frac{1}{2\pi\mathrm{i}} \oint_{\gamma_3} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}$$

$$= \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_4} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} + \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_5} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}. \tag{153}$$

Repeating a similar argument as for (138) reveals that

$$\frac{1}{2\pi} \sup_{\eta \in \gamma_4} \left| \frac{1}{\prod_{1 \leq i \leq k+1, i \notin \mathcal{I}_1} (\eta - \overline{\lambda}_{j_i})} \right| \cdot L(\gamma_4) \leq \frac{8}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^{k-1}. \tag{154}$$

In addition, for any $\eta \in \gamma_4$, we know that its real part

$$\mathrm{Re}(\eta) \leq \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}$$

$$\leq \frac{2}{3}\overline{\lambda}_{j_{\min}} + \frac{1}{3}\overline{\lambda}_{j_{\max}} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}$$

$$= \overline{\lambda}_{j_{\min}} - \frac{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}}}{3} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}$$

$$\leq \overline{\lambda}_{j_{\min}} - \frac{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}}}{8} - \frac{5}{8}\left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right) + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}$$

$$= \overline{\lambda}_{j_{\min}} - \frac{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}}{8},$$

which further tells us that

$$\left|\eta - \overline{\lambda}_{j_{\min}}\right| \geq \frac{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}}{8}, \qquad \forall \eta \in \gamma_4.$$

Putting (154) and the previous inequality together, one has

$$\left|\frac{1}{2\pi i}\oint_{\gamma_4} \frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1}\right)\cdots\left(\eta - \overline{\lambda}_{j_{k+1}}\right)}\right|$$

$$\leq \frac{1}{2\pi}\sup_{\eta\in\gamma_4}\left|\frac{1}{\prod_{1\leq i\leq k, i\notin\mathcal{I}_1}\left(\eta - \overline{\lambda}_{j_i}\right)}\right| \cdot \sup_{\eta\in\gamma_4}\frac{\overline{\lambda}_{j_{k+1}}}{\left|\eta - \overline{\lambda}_{j_{\min}}\right|} \cdot L\left(\gamma_4\right)$$

$$\leq \frac{8}{\pi}\left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^{k-1} \cdot \frac{8\overline{\lambda}_{j_{\min}}}{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}}$$

$$\leq \frac{32}{\pi}\left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^{k}\overline{\lambda}_{r_1}. \tag{155}$$

Here, the second and the third lines also make use of $|\mathcal{I}_1| = 1$. Note that for all $i \in \{1,\ldots,k+1\}\backslash\mathcal{I}_1$, $\overline{\lambda}_{j_i}$ is not in or on γ_5. By virtue of Cauchy's integral formula, one has

$$\frac{1}{2\pi i}\oint_{\gamma_5} \frac{\mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1}\right)\cdots\left(\eta - \overline{\lambda}_{j_{k+1}}\right)} = \frac{1}{2\pi i}\oint_{\gamma_5} \frac{\prod_{1\leq i\leq k+1, i\notin\mathcal{I}_1}\frac{1}{\eta-\overline{\lambda}_{j_i}}}{\eta - \overline{\lambda}_{j_{\min}}}\mathrm{d}\eta = \prod_{1\leq i\leq k+1, i\notin\mathcal{I}_1}\frac{1}{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_i}}. \tag{156}$$

Moreover, the definition of $\mathcal{I}_1$ tells us that

$$\min_{i:1\leq i\leq k+1, i\notin\mathcal{I}_1}\left|\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_i}\right| \geq \frac{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}}}{3}$$

$$= \frac{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}}}{4} + \frac{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}}}{12}$$

$$\geq \left(\frac{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_{\max}}}{4} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{4}\right) \vee \left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right)$$

$$= \frac{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}}{4} \vee \left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right).$$

Combining (156) and the previous inequality, one has

$$\left|\frac{1}{2\pi i}\oint_{\gamma_5} \frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1}\right)\cdots\left(\eta - \overline{\lambda}_{j_{k+1}}\right)}\right| \leq \prod_{1\leq i\leq k+1, i\notin\mathcal{I}_1}\frac{1}{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{j_i}}\overline{\lambda}_{j_{\min}}$$

$$\leq \frac{1}{\left(\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}\right)^{k-1}}\frac{4\overline{\lambda}_{j_{\min}}}{\overline{\lambda}_{j_{\min}} - \overline{\lambda}_{r_1+1}}$$

$$\leq \frac{1}{\left(\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}\right)^{k-1}}\frac{4\overline{\lambda}_{r_1}}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}$$

45

$$\leq \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \overline{\lambda}_{r_1}. \tag{157}$$

Eqn. (146) together with (153), (155) and (157) implies that

$$\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_1} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1} \right) \cdots \left(\eta - \overline{\lambda}_{j_{k+1}} \right)} \right| \leq \frac{36}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \overline{\lambda}_{r_1}. \tag{158}$$

Case 2.3: $|\mathcal{I}_2| = 1$. In this case, we define

$$\ell' \in \operatorname*{arg\,min}_{i:1 \leq i \leq k+1, i \notin \mathcal{I}_2} \overline{\lambda}_{j_i}.$$

Denote by γ_6 and γ_7 the following counterclockwise contours:

$$\gamma_4 = \left\{ x + y\mathrm{i} : x = \frac{\overline{\lambda}_{j_{\max}} + \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, \right.$$

$$\text{or } x = \overline{\lambda}_{j_{\ell'}} - \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2},$$

$$\left. \text{or } y = \pm\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, \frac{\overline{\lambda}_{j_{\max}} + \overline{\lambda}_{r_1+1}}{2} \leq x \leq \overline{\lambda}_{j_{\ell'}} - \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \right\}$$

and

$$\gamma_5 = \left\{ x + y\mathrm{i} : x = \overline{\lambda}_{j_{\ell'}} - \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, \right.$$

$$\text{or } x = \overline{\lambda}_{j_{\min}} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2},$$

$$\left. \text{or } y = \pm\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, \overline{\lambda}_{j_{\ell'}} - \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \leq x \leq \overline{\lambda}_{j_{\min}} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \right\}.$$

Similar to (153), one has

$$\frac{1}{2\pi\mathrm{i}} \oint_{\gamma_3} \frac{\mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1} \right) \cdots \left(\eta - \overline{\lambda}_{j_{k+1}} \right)}$$

$$= \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_6} \frac{\mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1} \right) \cdots \left(\eta - \overline{\lambda}_{j_{k+1}} \right)} + \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_7} \frac{\mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1} \right) \cdots \left(\eta - \overline{\lambda}_{j_{k+1}} \right)}. \tag{159}$$

Repeating similar arguments as in (155) and (157) yields

$$\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_6} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1} \right) \cdots \left(\eta - \overline{\lambda}_{j_{k+1}} \right)} \right| \leq \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \overline{\lambda}_{r_1},, \tag{160a}$$

$$\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_7} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1} \right) \cdots \left(\eta - \overline{\lambda}_{j_{k+1}} \right)} \right| \leq \frac{32}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \overline{\lambda}_{r_1}. \tag{160b}$$

Putting (146), (159), (160a) and (160b) together, one has

$$\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_1} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1} \right) \cdots \left(\eta - \overline{\lambda}_{j_{k+1}} \right)} \right| \leq \frac{36}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \overline{\lambda}_{r_1}. \tag{161}$$

In summary, we are guaranteed to have

$$\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_1} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1} \right) \cdots \left(\eta - \overline{\lambda}_{j_{k+1}} \right)} \right| \leq \frac{40}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \overline{\lambda}_{r_1}.$$

This together with (139) finishes the proof of (87b). $\qquad \square$

C.4 Proof of Lemma 5

Throughout the subsection, we assume that $\mathcal{E}$ holds.

Proof of (92a). (92a) clearly holds for $i = 0$ due to the definition of μ. Now we consider the case $i \geq 1$. It is easy to verify that for any matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m_1 \times m_1}$,

$$(\boldsymbol{A} + \boldsymbol{B})^i = \boldsymbol{B}^i + \sum_{j=0}^{i-1} \boldsymbol{B}^j \boldsymbol{A} \left(\boldsymbol{A} + \boldsymbol{B}\right)^{i-j-1}. \tag{162}$$

This allows us to decompose $\boldsymbol{Z}_3^i \boldsymbol{U}^\star$ as follows:

$$
\begin{aligned}
\boldsymbol{Z}_3^i \boldsymbol{U}^\star &= \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right]^i \boldsymbol{U}^\star \\
&= -\sum_{j=0}^{i-1} \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j \mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right) \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right]^{i-j-1} \boldsymbol{U}^\star \\
&\quad + \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^i \boldsymbol{U}^\star \\
&= -\sum_{j=0}^{i-1} \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right]^{i-j-1} \boldsymbol{U}^\star \\
&\quad + \sum_{j=0}^{i-1} \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j \mathcal{P}_{\text{diag}}\left(\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right) \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right]^{i-j-1} \boldsymbol{U}^\star \\
&\quad + \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^i \boldsymbol{U}^\star.
\end{aligned}
$$

In view of (88), (89), (90a) and (90b), one can obtain the following upper bound for $\|\boldsymbol{Z}^i \boldsymbol{U}^\star\|_{2,\infty}$:

$$
\begin{aligned}
&\left\|\boldsymbol{Z}_3^i \boldsymbol{U}^\star\right\|_{2,\infty} \\
&\leq \sum_{j=0}^{i-1} \left\|\left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j \boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty} \left\|\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right]^{i-j-1} \boldsymbol{U}^\star\right\| \\
&\quad + \sum_{j=0}^{i-1} \left\|\left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j \mathcal{P}_{\text{diag}}\left(\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right) \left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right]^{i-j-1} \boldsymbol{U}^\star\right\| \\
&\quad + \left\|\left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^i \boldsymbol{U}^\star\right\|_{2,\infty} \\
&\leq \sum_{j=0}^{i-1} \left\|\left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j \boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty} \|\boldsymbol{E}\boldsymbol{V}^\star\| \left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right\|^{i-j-1} \\
&\quad + \sum_{j=0}^{i-1} \left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right\|^j \|\boldsymbol{E}\boldsymbol{V}^\star\|_{2,\infty}^2 \left\|\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right\|^{i-j-1} \\
&\quad + \left\|\left[\mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^i \boldsymbol{U}^\star\right\|_{2,\infty} \\
&\leq \sum_{j=0}^{i-1} \left(C_3 \sqrt{\mu r}\left(C_3\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\max}^2 \log^2 m\right)^j \omega_{\max} \log m\right) \cdot C_5 \sqrt{m_1}\omega_{\max} \log m \\
&\qquad\qquad \cdot \left(3C_5\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\max}^2 \log^2 m\right)^{i-j-1} \\
&\quad + \sum_{j=0}^{i-1} \left(C_5\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\max}^2 \log^2 m\right)^j \left(C_3 \sqrt{\mu r}\omega_{\max} \log m\right)^2 \cdot \left(3C_5\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\max}^2 \log^2 m\right)^{i-j-1} \\
&\quad + C_3 \sqrt{\frac{\mu r}{m_1}} \left(C_3\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\max}^2 \log^2 m\right)^i
\end{aligned}
$$

$$\leq 3C_3\sqrt{\frac{\mu r}{m_1}}\left(C_3\left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^i, \tag{163}$$

provided that $C_3 \geq 6C_5$.

Proof of (92b). When $i=0$, (92b) is a direct consequence of Lemma 2. For $i \geq 1$, similar to (163), one has

$$\left\|\boldsymbol{Z}_3^i\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}$$
$$\leq \sum_{j=0}^{i-1}\left\|\left[\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}\left\|\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\left[\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top-\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right]^{i-j-1}\boldsymbol{E}\boldsymbol{V}^\star\right\|$$
$$+\sum_{j=0}^{i-1}\left\|\left[\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\left[\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top-\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right]^{i-j-1}\boldsymbol{E}\boldsymbol{V}^\star\right\|$$
$$+\left\|\left[\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^i\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}$$
$$\leq \sum_{j=0}^{i-1}\left\|\left[\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}\left\|\boldsymbol{E}\boldsymbol{V}^\star\right\|^2\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top-\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right\|^{i-j-1}$$
$$+\sum_{j=0}^{i-1}\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right\|^j\left\|\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}^2\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top-\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right\|^{i-j-1}\left\|\boldsymbol{E}\boldsymbol{V}^\star\right\|$$
$$+\left\|\left[\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^i\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}$$
$$\leq \sum_{j=0}^{i-1}\left(C_3\sqrt{\mu r}\left(C_3\left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^j\omega_{\mathsf{max}}\log m\right)\cdot\left(\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m\right)^2$$
$$\qquad\cdot\left(3C_5\left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^{i-j-1}$$
$$+\sum_{j=0}^{i-1}\left(C_5\left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^j\left(C_3\sqrt{\mu r}\omega_{\mathsf{max}}\log m\right)^2\cdot\left(3C_5\left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^{i-j-1}$$
$$\qquad\cdot\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m$$
$$+C_3\sqrt{\mu r}\left(C_3\left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^i\omega_{\mathsf{max}}\log m$$
$$\leq 3C_3\sqrt{\mu r}\left(C_3\left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^i\omega_{\mathsf{max}}\log m, \tag{164}$$

provided that $C_3 \geq 6C_5$. In the last inequality, we used the assumption that $m_1 \gg \mu r$.

Proof of (92c). We can directly use the same argument of Zhou and Chen (2023, Eqn. (121)) to prove (92c). We omit the details here for the sake of brevity.

Proof of (92d). Putting (78), (90a), (92a) and (92b) together shows that: for all $0 \leq i \leq \log m$, we have

$$\left\|\boldsymbol{Z}_3^i\boldsymbol{Z}_1\right\|_{2,\infty}=\left\|\boldsymbol{Z}_3^i\left(\boldsymbol{U}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\boldsymbol{E}+\boldsymbol{E}\boldsymbol{V}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{U}^{\star(2)\top}+\boldsymbol{E}\boldsymbol{V}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\boldsymbol{E}\right)\right\|_{2,\infty}$$
$$\leq\left\|\boldsymbol{Z}_3^i\boldsymbol{U}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\boldsymbol{E}^\top\right\|_{2,\infty}+\left\|\boldsymbol{Z}_3^i\boldsymbol{E}\boldsymbol{V}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{U}^{\star(2)\top}\right\|_{2,\infty}+\left\|\boldsymbol{Z}_3^i\boldsymbol{E}\boldsymbol{V}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\boldsymbol{E}^\top\right\|_{2,\infty}$$
$$\leq\left\|\boldsymbol{Z}_3^i\boldsymbol{U}^{\star(2)}\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}^{\star(2)}\right\|\left\|\boldsymbol{E}\boldsymbol{V}^{\star(2)}\right\|+\left\|\boldsymbol{Z}_3^i\boldsymbol{E}\boldsymbol{V}^{\star(2)}\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}^{\star(2)}\right\|+\left\|\boldsymbol{Z}_3^i\boldsymbol{E}\boldsymbol{V}^{\star(2)}\right\|_{2,\infty}\left\|\boldsymbol{E}\boldsymbol{V}^{\star(2)}\right\|$$
$$\leq\sigma_{\widetilde{r}+1}^\star\left\|\boldsymbol{Z}_3^i\boldsymbol{U}^\star\right\|_{2,\infty}\left\|\boldsymbol{E}\boldsymbol{V}^\star\right\|+\sigma_{\widetilde{r}+1}^\star\left\|\boldsymbol{Z}_3^i\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}+\left\|\boldsymbol{Z}_3^i\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}\left\|\boldsymbol{E}\boldsymbol{V}^\star\right\|$$
$$\leq\sigma_{\widetilde{r}+1}^\star\cdot 3C_3\sqrt{\frac{\mu r}{m_1}}\left(C_3\left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^i\cdot\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m$$

$$+ \sigma_{\bar{r}+1}^{\star} \omega_{\mathsf{max}} \log m \cdot 3 C_3 \sqrt{\mu r} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i \omega_{\mathsf{max}} \log m$$

$$+ 3 C_3 \sqrt{\mu r} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i \omega_{\mathsf{max}} \log m \cdot \sqrt{C_5} \sqrt{m_1} \omega_{\mathsf{max}} \log m$$

$$\leq C_2 \sqrt{\mu r} \left(\sigma_{\bar{r}+1}^{\star} + \sqrt{m_1} \omega_{\mathsf{max}} \log m \right) \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i \omega_{\mathsf{max}} \log m, \tag{165}$$

provided that $C_2 \geq 9 C_3 \sqrt{C_5}$. The fourth line makes use of the fact that $\|\boldsymbol{B}\| \leq \|\boldsymbol{A}\|$ and $\|\boldsymbol{B}\|_{2,\infty} \leq \|\boldsymbol{A}\|_{2,\infty}$ for any $\boldsymbol{A}$ and its submatrix $\boldsymbol{B}$.

Proof of (92e). By virtue of (78), (92a), (92c) and (94), we know that for all $0 \leq i \leq \log m$,

$$\begin{aligned}
\left\| \boldsymbol{Z}_3^i \boldsymbol{Z}_2 \right\|_{2,\infty} &\leq \left\| \boldsymbol{Z}_3^i \left(\mathcal{P}_{\widetilde{\boldsymbol{U}}^{(1)}} \boldsymbol{U}^{\star(2)} \left(\boldsymbol{\Sigma}^{\star(2)} \right)^2 \boldsymbol{U}^{\star(2)\top} \mathcal{P}_{\left(\widetilde{\boldsymbol{U}}^{(1)} \right)_\perp} + \boldsymbol{U}^{\star(2)} \left(\boldsymbol{\Sigma}^{\star(2)} \right)^2 \boldsymbol{U}^{\star(2)\top} \mathcal{P}_{\widetilde{\boldsymbol{U}}^{(1)}} \right) \right\|_{2,\infty} \\
&\leq \left\| \boldsymbol{Z}_3^i \widetilde{\boldsymbol{U}}^{(1)} \right\|_{2,\infty} \left\| \widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)} \right\| \left\| \boldsymbol{\Sigma}^{\star(2)} \right\|^2 + \left\| \boldsymbol{Z}_3^i \boldsymbol{U}^{\star(2)} \right\|_{2,\infty} \left\| \boldsymbol{\Sigma}^{\star(2)} \right\|^2 \left\| \widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)} \right\| \\
&\leq \left\| \boldsymbol{Z}_3^i \widetilde{\boldsymbol{U}}^{(1)} \right\|_{2,\infty} \left\| \widetilde{\boldsymbol{U}}^{(1)\top} \left(\boldsymbol{U}_1^{\star} \right)_\perp \right\| \sigma_{\bar{r}+1}^{\star 2} + \left\| \boldsymbol{Z}_3^i \boldsymbol{U}^{\star(2)} \right\|_{2,\infty} \left\| \widetilde{\boldsymbol{U}}^{(1)\top} \left(\boldsymbol{U}_1^{\star} \right)_\perp \right\| \sigma_{\bar{r}+1}^{\star 2} \\
&\lesssim \sqrt{\frac{\mu r}{m_1}} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i \cdot \frac{\sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma_{\bar{r}}^{\star}} \sigma_{\bar{r}+1}^{\star 2} \\
&\leq \sqrt{\frac{\mu r}{m_1}} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i \cdot \sqrt{m_1} \omega_{\mathsf{max}} \sigma_{\bar{r}+1}^{\star} \log m \\
&= \sqrt{\mu r} \left(C_3 \left(\sqrt{m_1 m_2} + m_1 \right) \omega_{\mathsf{max}}^2 \log^2 m \right)^i \omega_{\mathsf{max}} \sigma_{\bar{r}+1}^{\star} \log m. \tag{166}
\end{aligned}$$

The third line holds since $\boldsymbol{U}^{\star(1)\top} \boldsymbol{U}^{\star(2)} = \boldsymbol{0}$, whereas the fifth line is due to the basic fact that $\sigma_{\bar{r}}^{\star} \geq \sigma_{\bar{r}+1}^{\star}$. $\quad\square$

D Proof of Theorem 6

D.1 Several notation

First, we introduce some notation that will be useful throughout the proof. We let

$$\boldsymbol{G}_{k+1}^0 := \boldsymbol{G}_k, \qquad \forall 0 \leq k \leq k_{\mathsf{max}}, \tag{167}$$

For any $0 \leq t \leq t_{k+1}$ and $0 \leq k \leq k_{\mathsf{max}}$, we define

$$\boldsymbol{U}_{k+1}^t \boldsymbol{\Lambda}_{k+1}^t \boldsymbol{U}_{k+1}^{t\top} := \text{ the leading } r_k \text{ eigendecomposition of } \boldsymbol{G}_{k+1}^t, \tag{168}$$

and denote

$$\boldsymbol{G}_{k+1}^{t+1} := \mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{G}_{k+1}^t \right) + \mathcal{P}_{\mathsf{diag}} \left(\boldsymbol{U}_{k+1}^t \boldsymbol{\Lambda}_{k+1}^t \boldsymbol{U}_{k+1}^{t\top} \right). \tag{169}$$

Recall that we can decompose $\boldsymbol{M}^{\mathsf{oracle}}$ into four terms:

$$\boldsymbol{M}^{\mathsf{oracle}} = \widetilde{\boldsymbol{M}} + \boldsymbol{Z}_1 + \boldsymbol{Z}_2 + \boldsymbol{Z}_3 = \widetilde{\boldsymbol{M}} + \boldsymbol{Z}, \tag{170}$$

where $\widetilde{\boldsymbol{M}}, \boldsymbol{Z}_1, \boldsymbol{Z}_2$ and $\boldsymbol{Z}_3$ are defined in (84).

For notational convenience, we let

$$D_k^t = \left\| \boldsymbol{G}_k^t - \boldsymbol{M}^{\mathsf{oracle}} \right\|, \qquad F_k^t = \left\| \mathcal{P}_{\mathsf{diag}} \left(\boldsymbol{G}_k^t - \widetilde{\boldsymbol{M}} \right) \right\|, \qquad \text{and} \qquad L_k^t = \left\| \boldsymbol{G}_k^t - \widetilde{\boldsymbol{M}} \right\|. \tag{171}$$

In addition, we let

$$\widetilde{\boldsymbol{U}}_k = \widetilde{\boldsymbol{U}}_{:,1:r_k}, \qquad \boldsymbol{U}_k^{\mathsf{oracle}} = \boldsymbol{U}_{:,1:r_k}^{\mathsf{oracle}}, \qquad \text{and} \qquad \boldsymbol{U}_k^{\star} = \boldsymbol{U}_{:,1:r_k}^{\star}, \tag{172}$$

where $\widetilde{\boldsymbol{U}}$ (resp. $\boldsymbol{U}^{\mathsf{oracle}}$) is the rank-$r$ leading eigenspace of $\widetilde{\boldsymbol{M}}$ (resp. $\boldsymbol{M}^{\mathsf{oracle}}$). We also let $\mathcal{E}$ denote the following event:

$$\mathcal{E} = \{(88) \text{ and } (89) \text{ hold for } 0 \leq k \leq \log n\} \cap \{(90a), (90b), (90c) \text{ and } (90d) \text{ hold}\}$$

$$\cap \{(39a) \text{ and } (39b) \text{ for all } r' \in \mathcal{A}\}. \tag{173}$$

We know from Lemmas 2, 3, 4, Theorem 5 and the union bound that

$$\mathbb{P}(\mathcal{E}) \geq 1 - O\left(m^{-10}\right). \tag{174}$$

Throughout the rest of the proof, we assume that $\mathcal{E}$ occurs.

D.2 Main steps for proving Theorem 6

Step 1: a key property of r_1 selected in Algorithm 1. First, we show that

$$r_1 \in \mathcal{R}_1 \cap \mathcal{A}, \tag{175}$$

where

$$\mathcal{R}_1 := \left\{ r' \leq r : \frac{\sigma_1(\boldsymbol{G}_0)}{\sigma_{r'}(\boldsymbol{G}_0)} \leq 4 \quad \text{and} \quad \sigma_{r'}(\boldsymbol{G}_0) - \sigma_{r'+1}(\boldsymbol{G}_0) \geq \frac{1}{r}\sigma_{r'}(\boldsymbol{G}_0) \right\}. \tag{176}$$

and $\mathcal{A}$ is defined in (38). Noting that $\widetilde{\boldsymbol{U}} = [\widetilde{\boldsymbol{U}}^{(1)} \ \widetilde{\boldsymbol{U}}^{(2)}]$ and putting (90d) and (99) together, one has

$$\left\|\widetilde{\boldsymbol{U}}\right\|_{2,\infty} = \left\|[\widetilde{\boldsymbol{U}}^{(1)} \ \widetilde{\boldsymbol{U}}^{(2)}]\right\|_{2,\infty} \leq \left\|\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty} + \left\|\widetilde{\boldsymbol{U}}^{(2)}\right\|_{2,\infty} \leq 4\sqrt{\frac{\mu r}{m_1}}. \tag{177}$$

In view of (107), (177) and the definition $\boldsymbol{G}_1^0 = \boldsymbol{G}_0 = \mathcal{P}_{\text{off-diag}}(\boldsymbol{M}) = \mathcal{P}_{\text{off-diag}}(\boldsymbol{M}^{\text{oracle}}) = \mathcal{P}_{\text{off-diag}}(\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^{\top}+\boldsymbol{Z})$, we can derive

$$\begin{aligned}
L_1^0 &= \left\|\boldsymbol{G}_0 - \widetilde{\boldsymbol{M}}\right\| \\
&= \left\|\mathcal{P}_{\text{diag}}(\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^{\top}) - \mathcal{P}_{\text{off-diag}}(\boldsymbol{Z})\right\| \\
&\leq \left\|\mathcal{P}_{\text{diag}}(\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^{\top})\right\| + \left\|\mathcal{P}_{\text{off-diag}}(\boldsymbol{Z})\right\| \\
&\leq \left\|\widetilde{\boldsymbol{U}}\right\|_{2,\infty}^2 \left\|\widetilde{\boldsymbol{\Lambda}}\right\| + 2\left\|\boldsymbol{Z}\right\| \\
&\leq 16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2 + C_2\left(\sqrt{m_1}\omega_{\text{max}}\log m \cdot \sigma_{\bar{r}+1}^{\star} + (\sqrt{m_1 m_2} + m_1)\omega_{\text{max}}^2\log^2 m\right).
\end{aligned} \tag{178}$$

This together with Weyl's inequality shows that, for all $i \in [m_1]$,

$$\begin{aligned}
\left|\sigma_i(\boldsymbol{G}_0) - \widetilde{\sigma}_i^2\right| &= \left|\sigma_i(\boldsymbol{G}_0) - \sigma_i(\widetilde{\boldsymbol{M}})\right| \\
&\leq 16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2 + C_2\left(\sqrt{m_1}\omega_{\text{max}}\log m \cdot \sigma_{\bar{r}+1}^{\star} + (\sqrt{m_1 m_2} + m_1)\omega_{\text{max}}^2\log^2 m\right).
\end{aligned} \tag{179}$$

Moreover, (90a) tells us that

$$\widetilde{\sigma}_i \leq \sigma_i^{\star} + \sqrt{C_5}\sqrt{m_1}\omega_{\text{max}}\log m \leq \left(1 + \frac{1}{Cr^2}\right)\sigma_i^{\star}, \qquad \forall i \in [\bar{r}]$$

for some large constant $C > 0$, where $\bar{r} = \max\mathcal{A}$ and the last inequality holds since $\sigma_{\bar{r}}^{\star} \geq C_0 r\left[(m_1 m_2)^{1/4} + rm_1^{1/2}\right]\omega_{\text{max}}\log m$. Similarly, one can show that

$$\left(1 - \frac{1}{Cr^2}\right)\sigma_i^{\star} \leq \widetilde{\sigma}_i \leq \left(1 + \frac{1}{Cr^2}\right)\sigma_i^{\star}, \qquad \forall i \in [\bar{r}]. \tag{180}$$

By virtue of (180) and the assumption $\sigma_{\bar{r}}^{\star} \geq C_0 r\left[(m_1 m_2)^{1/4} + rm_1^{1/2}\right]\omega_{\text{max}}\log m$, one has

$$16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2 + C_2\left(\sqrt{m_1}\omega_{\text{max}}\log m \cdot \sigma_{\bar{r}+1}^{\star} + (\sqrt{m_1 m_2} + m_1)\omega_{\text{max}}^2\log^2 m\right)$$

$$\leq 16 \frac{\mu r}{m_1} \widetilde{\sigma}_1^2 + C_2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \sigma_{\overline{r}+1}^\star + \frac{1}{C_0^2 r^2} \sigma_{\overline{r}}^{\star 2} \right)$$

$$\leq 16 \frac{\mu r}{m_1} \widetilde{\sigma}_1^2 + C_2 \left(2\sqrt{m_1} \omega_{\mathsf{max}} \log m \cdot \widetilde{\sigma}_{\overline{r}+1} + \frac{4}{C_0^2 r^2} \widetilde{\sigma}_{\overline{r}}^2 \right)$$

$$\leq 16 \frac{\mu r}{m_1} \widetilde{\sigma}_1^2 + C_2 \left(\frac{2}{C_0 r^2} \widetilde{\sigma}_{\overline{r}}^2 + \frac{4}{C_0^2 r^2} \widetilde{\sigma}_{\overline{r}}^2 \right)$$

$$\leq \frac{16 c_1}{r^2} \widetilde{\sigma}_1^2 + \frac{1}{2 C r^2} \widetilde{\sigma}_{\overline{r}}^2$$

$$\leq \frac{1}{C r^2} \widetilde{\sigma}_1^2, \tag{181}$$

provided that $C_0 \geq 8C \cdot C_2$ and $c_1 \leq \frac{1}{32C}$. Here, the fourth line comes from $\sigma_{\overline{r}}^\star \geq C_0 r \big[(m_1 m_2)^{1/4} + r m_1^{1/2} \big] \omega_{\mathsf{max}} \log m$ and (180), whereas the penultimate line uses the assumption $\mu \leq c_0 m_1 / r^3$. Combining (179) and (181) gives

$$\sigma_1 (\boldsymbol{G}_0) \geq \widetilde{\sigma}_1^2 - \frac{1}{C r^2} \widetilde{\sigma}_1^2 > \max \left\{ \frac{1}{2} \widetilde{\sigma}_1^2, \frac{1}{2} \sigma_1^{\star 2}, \tau \right\}, \tag{182}$$

where the last inequality holds due to (180) and $C_1^2 / 2 \geq C_\tau$.

Equipped with (179), (180), (181) and (182), we can establish the property $r_1 \in \mathcal{R}_1$ using a similar argument as in the proof of Zhou and Chen (2023, Eqn. (62)). Therefore, we only need to prove $r_1 \in \mathcal{A}$. In view of (180), (182) and the definition of $\mathcal{R}_1$, we know that

$$\sigma_{r_1} (\boldsymbol{G}_0) \geq \frac{1}{4} \sigma_1 (\boldsymbol{G}_0) \geq \frac{1}{8} \sigma_1^{\star 2}, \tag{183}$$

and consequently,

$$\sigma_{r_1}^\star \geq \left(1 - \frac{C}{r^2} \right) \widetilde{\sigma}_{r_1} \geq \left(1 - \frac{C}{r^2} \right) \left[\sigma_{r_1} (\boldsymbol{G}_0) - \frac{1}{C r^2} \widetilde{\sigma}_1^2 \right]^{1/2} \geq \left(1 - \frac{C}{r^2} \right) \left(\frac{1}{8} \sigma_1^{\star 2} - \frac{4}{C r^2} \sigma_1^{\star 2} \right)^{1/2}$$

$$\geq \frac{1}{3} \sigma_1^\star > 2 C_0 r \big[(m_1 m_2)^{1/4} + r m_1^{1/2} \big] \omega_{\mathsf{max}} \log m. \tag{184}$$

Furthermore, inequality (179), (181) and (183) combined imply that

$$\widetilde{\sigma}_{r_1}^2 \geq \sigma_{r_1} (\boldsymbol{G}_0) - \frac{1}{C r^2} \widetilde{\sigma}_1^2 \geq \frac{1}{4} \sigma_1 (\boldsymbol{G}_0) - \frac{1}{C r^2} \widetilde{\sigma}_1^2 \geq \frac{1}{4} \left(\widetilde{\sigma}_1^2 - \frac{1}{C r^2} \widetilde{\sigma}_1^2 \right) - \frac{1}{C r^2} \widetilde{\sigma}_1^2 \geq \frac{1}{5} \widetilde{\sigma}_1^2. \tag{185}$$

Inequalities (179), (181), (185) and the triangle inequality taken together show that

$$\widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2 \geq \sigma_{r_1} (\boldsymbol{G}_0) - \sigma_{r_1+1} (\boldsymbol{G}_0) - \left| \sigma_{r_1} (\boldsymbol{G}_0) - \widetilde{\sigma}_{r_1}^2 \right| - \left| \sigma_{r_1+1} (\boldsymbol{G}_0) - \widetilde{\sigma}_{r_1+1}^2 \right|$$

$$\geq \frac{1}{r} \sigma_{r_1} (\boldsymbol{G}_0) - \frac{2}{C r^2} \widetilde{\sigma}_1^2$$

$$\geq \frac{1}{r} \widetilde{\sigma}_{r_1}^2 - \frac{1}{r} \left| \sigma_{r_1} (\boldsymbol{G}_0) - \widetilde{\sigma}_{r_1}^2 \right| - \frac{2}{C r^2} \widetilde{\sigma}_1^2$$

$$\geq \frac{1}{r} \widetilde{\sigma}_{r_1}^2 - \frac{3}{C r^2} \widetilde{\sigma}_1^2$$

$$\geq \frac{1}{r} \widetilde{\sigma}_{r_1}^2 - \frac{15}{C r^2} \widetilde{\sigma}_{r_1}^2$$

$$\geq \frac{9}{10 r} \widetilde{\sigma}_{r_1}^2. \tag{186}$$

Note that (78) together with (184) reveals that $r_1 \leq \overline{r}$, where $\overline{r}$ is the largest element in $\mathcal{A}$. Putting (90a), (103), (180) and the previous inequality together, we arrive at

$$\sigma_{r_1}^\star - \sigma_{r_1+1}^\star \geq \widetilde{\sigma}_{r_1} - \widetilde{\sigma}_{r_1+1} - \left| \widetilde{\sigma}_{r_1} - \sigma_{r_1}^\star \right| - \left| \widetilde{\sigma}_{r_1+1} - \sigma_{r_1+1}^\star \right|$$

$$\geq \frac{1}{\widetilde{\sigma}_{r_1} + \widetilde{\sigma}_{r_1+1}} \left(\widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2 \right) - \sqrt{C_5} \sqrt{m_1} \omega_{\max} \log m - 2\sqrt{C_5} \sqrt{m_1} \omega_{\max} \log m$$

$$\geq \frac{1}{2\widetilde{\sigma}_{r_1}} \cdot \frac{9}{10r} \widetilde{\sigma}_{r_1}^2 - \frac{1}{20r^2} \sigma_{r_1}^\star$$

$$\geq \frac{9}{20r} \left(1 - \frac{1}{Cr} \right) \sigma_{r_1}^\star - \frac{1}{20r^2} \sigma_{r_1}^\star$$

$$\geq \frac{1}{4r} \sigma_{r_1}^\star. \tag{187}$$

Inequality (184) taken together with (187) validates $r_1 \in \mathcal{A}$, thus finishing the proof of (175).

Step 2: bounding $D_1^t = \|\boldsymbol{G}_1^t - \boldsymbol{M}^{\mathsf{oracle}}\|$. Now, we would like to deal with the quantities $\{D_1^t\}$. Recognizing that for all t and k,

$$\mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{G}_k^t \right) = \mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{M}^{\mathsf{oracle}} \right) = \mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{Y} \boldsymbol{Y}^\top \right),$$

we can deduce that

$$D_k^t = \left\| \mathcal{P}_{\mathsf{diag}} \left(\boldsymbol{G}_k^t - \boldsymbol{M}^{\mathsf{oracle}} \right) \right\|. \tag{188}$$

We would like to prove the following inequalities by induction:

$$F_1^t - 40\sqrt{\frac{\mu r^3}{m_1}} \|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}} \widetilde{\sigma}_{r_1+1}^2 \leq \frac{1}{e^t} \left(F_1^0 - 40\sqrt{\frac{\mu r^3}{m_1}} \|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}} \widetilde{\sigma}_{r_1+1}^2 \right), \tag{189a}$$

$$D_1^t \leq F_1^t + 6C_3 \sqrt{\frac{\mu r}{m_1}} \sqrt{m_1} \omega_{\max} \log m \cdot \sigma_{\overline{r}+1}^\star + C_3^2 \mu r \omega_{\max}^2 \log^2 m, \tag{189b}$$

$$\left\| \boldsymbol{U}_1^t \boldsymbol{U}_1^{t\top} - \boldsymbol{U}_1^{\mathsf{oracle}} \boldsymbol{U}_1^{\mathsf{oracle}\top} \right\| \leq 2 \frac{D_1^t}{\lambda_{r_1} \left(\boldsymbol{M}^{\mathsf{oracle}} \right) - \lambda_{r_1+1} \left(\boldsymbol{M}^{\mathsf{oracle}} \right)} \leq \frac{1}{8}, \tag{189c}$$

$$\left\| \boldsymbol{U}_1^t \right\|_{2,\infty} \leq \left\| \boldsymbol{U}_1^t \boldsymbol{U}_1^{t\top} - \boldsymbol{U}_1^{\mathsf{oracle}} \boldsymbol{U}_1^{\mathsf{oracle}\top} \right\| + \left\| \boldsymbol{U}_1^{\mathsf{oracle}} \right\|_{2,\infty} \leq \frac{1}{4e}. \tag{189d}$$

Step 2.1: the base case ($t = 0$) for (189a)-(189d). Note that (189a) automatically holds when $t = 0$. Recalling that $\mathcal{P}_{\mathsf{diag}}(\boldsymbol{G}_1^0) = 0$, we can invoke Zhang et al. (2022, Lemma 1) together with (177) to obtain

$$F_1^0 = \left\| \mathcal{P}_{\mathsf{diag}}(\widetilde{\boldsymbol{M}}) \right\| = \left\| \mathcal{P}_{\mathsf{diag}}(\widetilde{\boldsymbol{U}} \widetilde{\boldsymbol{\Lambda}} \widetilde{\boldsymbol{U}}^\top) \right\| \leq \|\widetilde{\boldsymbol{U}}\|_{2,\infty}^2 \|\widetilde{\boldsymbol{\Lambda}}\| \leq 16 \frac{\mu r}{m_1} \widetilde{\sigma}_1^2. \tag{190}$$

Furthermore, putting Lemma 2, (84), (90d), (94) and (190) together, we have

$$D_1^0 = \left\| \mathcal{P}_{\mathsf{diag}} \left(\boldsymbol{M}^{\mathsf{oracle}} \right) \right\| = \left\| \mathcal{P}_{\mathsf{diag}} \left(\widetilde{\boldsymbol{M}} + \boldsymbol{Z}_1 + \boldsymbol{Z}_2 + \boldsymbol{Z}_3 \right) \right\|$$

$$\leq \left\| \mathcal{P}_{\mathsf{diag}}(\widetilde{\boldsymbol{M}}) \right\| + \|\mathcal{P}_{\mathsf{diag}} (\boldsymbol{Z}_1)\| + \|\mathcal{P}_{\mathsf{diag}} (\boldsymbol{Z}_2)\| + \|\mathcal{P}_{\mathsf{diag}} (\boldsymbol{Z}_3)\|$$

$$\leq F_1^0 + \left\| \mathcal{P}_{\mathsf{off\text{-}diag}} \left(\boldsymbol{U}^{\star(2)} \boldsymbol{\Sigma}^{\star(2)} \boldsymbol{V}^{\star(2)\top} \boldsymbol{E}^\top + \boldsymbol{E} \boldsymbol{V}^{\star(2)} \boldsymbol{\Sigma}^{\star(2)} \boldsymbol{U}^{\star(2)\top} + \boldsymbol{E} \boldsymbol{V}^{\star(2)} \boldsymbol{V}^{\star(2)\top} \boldsymbol{E}^\top \right) \right\|$$

$$\quad + 2 \left\| \mathcal{P}_{\mathsf{diag}} \left(\mathcal{P}_{\widetilde{\boldsymbol{U}}^{(1)}} \boldsymbol{U}^{\star(2)} \left(\boldsymbol{\Sigma}^{\star(2)} \right)^2 \boldsymbol{U}^{\star(2)\top} \mathcal{P}_{\left(\widetilde{\boldsymbol{U}}^{(1)} \right)_\perp} \right) \right\| + 0$$

$$\leq F_1^0 + 2\|\boldsymbol{U}^{\star(2)}\|_{2,\infty} \|\boldsymbol{E} \boldsymbol{V}^{\star(2)}\|_{2,\infty} \|\boldsymbol{\Sigma}^{\star(2)}\| + \|\boldsymbol{E} \boldsymbol{V}^{\star(2)}\|_{2,\infty}^2 + 2\|\widetilde{\boldsymbol{U}}^{(1)}\|_{2,\infty} \|\widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)}\| \|\boldsymbol{\Sigma}^{\star(2)}\|^2$$

$$\leq F_1^0 + 2\sqrt{\frac{\mu r}{m_1}} \cdot C_3 \sqrt{\mu r} \omega_{\max} \log m \cdot \sigma_{\overline{r}+1}^\star + \left(C_3 \sqrt{\mu r} \omega_{\max} \log m \right)^2 + 4\sqrt{\frac{\mu r}{m_1}} \cdot C_3 \frac{\sqrt{m_1} \omega_{\max} \log m}{\sigma_{\overline{r}}^\star} \sigma_{\overline{r}+1}^{\star 2}$$

$$\leq F_1^0 + 6C_3 \sqrt{\frac{\mu r}{m_1}} \sqrt{m_1} \omega_{\max} \log m \cdot \sigma_{\overline{r}+1}^\star + C_3^2 \mu r \omega_{\max}^2 \log^2 m, \tag{191}$$

which validates (189b) for $t = 0$. Here, the last line holds since $\mu \le c_0 m_1 / r^3$ and $\sigma_{\widetilde{r}}^\star \ge \sigma_{\widetilde{r}+1}^\star$. Combining (184), (185), (186) and (187), one has

$$\max\left\{\frac{\sigma_1^\star}{\sigma_{r_1}^\star}, \frac{\widetilde{\sigma}_1}{\widetilde{\sigma}_{r_1}}\right\} \le 3 \qquad \text{and} \qquad \min\left\{\frac{\widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2}{\widetilde{\sigma}_{r_1}^2}, \frac{\sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2}}{\sigma_{r_1}^{\star 2}}\right\} \ge \frac{1}{4r}. \tag{192}$$

Moreover, by virtue of (90a), (103) and the fact $r_1 \in \mathcal{A}$, we know that

$$\left|\left(\widetilde{\sigma}_{r_1} - \widetilde{\sigma}_{r_1+1}\right) - \left(\sigma_{r_1}^\star - \sigma_{r_1+1}^\star\right)\right| \le \left|\widetilde{\sigma}_{r_1} - \sigma_{r_1}^\star\right| + \left|\widetilde{\sigma}_{r_1+1} - \sigma_{r_1+1}^\star\right| \le \sqrt{C_5}\sqrt{m_1}\omega_{\max}\log m + 2\sqrt{C_5}\sqrt{m_1}\omega_{\max}\log m$$

$$\ll \frac{1}{r^2}\sigma_{r_1}^\star \lesssim \frac{1}{r}\left(\sigma_{r_1}^\star - \sigma_{r_1+1}^\star\right),$$

where the last inequality comes from (187). This implies that

$$\left(1 - \frac{1}{Cr}\right)\left(\sigma_{r_1}^\star - \sigma_{r_1+1}^\star\right) \le \widetilde{\sigma}_{r_1} - \widetilde{\sigma}_{r_1+1} \le \left(1 + \frac{1}{Cr}\right)\left(\sigma_{r_1}^\star - \sigma_{r_1+1}^\star\right).$$

The previous inequality together with (180), (107) and (192) gives

$$\widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2 \asymp \sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2} \gg \|\boldsymbol{Z}\|.$$

Recalling that $\boldsymbol{M}^{\text{oracle}} = \widetilde{\boldsymbol{M}} + \boldsymbol{Z}$, one can invoke Weyl's inequality to obtain

$$\lambda_{r_1}\left(\boldsymbol{M}^{\text{oracle}}\right) - \lambda_{r_1+1}\left(\boldsymbol{M}^{\text{oracle}}\right) \asymp \widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2 \asymp \sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2} \gg \|\boldsymbol{Z}\|. \tag{193}$$

Note that $\boldsymbol{U}_1^0$ (resp. $\boldsymbol{U}_1^{\text{oracle}}$) is the rank-$r$ leading eigenspace of $\boldsymbol{G}_0$ (resp. $\boldsymbol{M}^{\text{oracle}}$). We know from the Davis-Kahan Theorem (Chen et al., 2021a, Theorem 2.7), (191), (192) and (193) that

$$\left\|\boldsymbol{U}_1^0\boldsymbol{U}_1^{0\top} - \boldsymbol{U}_1^{\text{oracle}}\boldsymbol{U}_1^{\text{oracle}\top}\right\| \le 2\frac{\left\|\boldsymbol{G}_0 - \boldsymbol{M}^{\text{oracle}}\right\|}{\lambda_{r_1}\left(\boldsymbol{M}^{\text{oracle}}\right) - \lambda_{r_1+1}\left(\boldsymbol{M}^{\text{oracle}}\right)} = 2\frac{D_1^0}{\lambda_{r_1}\left(\boldsymbol{M}^{\text{oracle}}\right) - \lambda_{r_1+1}\left(\boldsymbol{M}^{\text{oracle}}\right)}$$

$$\overset{(193)}{\lesssim} \frac{\frac{\mu r}{m_1}\widetilde{\sigma}_1^2}{\widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2} + \frac{\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\max}\log m \cdot \sigma_{\widetilde{r}+1}^\star}{\sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2}} + \frac{\mu r\omega_{\max}^2\log^2 m}{\sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2}}$$

$$\overset{(192)}{\lesssim} \frac{\frac{\mu r}{m_1}\widetilde{\sigma}_1^2}{\widetilde{\sigma}_{r_1}^2/r} + \frac{\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\max}\log m \cdot \sigma_{\widetilde{r}+1}^\star}{\sigma_{r_1}^{\star 2}/r} + \frac{\mu r\omega_{\max}^2\log^2 m}{\sigma_{r_1}^{\star 2}/r}$$

$$\overset{(192)}{\lesssim} \frac{\mu r^2}{m_1} + \frac{\sqrt{\frac{\mu r}{m_1}}r\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r_1}^\star} + \frac{\mu r\omega_{\max}^2\log^2 m}{\sigma_{r_1}^{\star 2}/r}$$

$$\ll \sqrt{\frac{\mu r}{m_1}} \le \frac{1}{8}, \tag{194}$$

which proves (189c) for $t = 0$. Here, the last inequality is due to $\mu \le c_0 m_1/r^3$ and $\sigma_{r_1}^\star \ge \sigma_{\widetilde{r}}^\star \ge C_0 r\left[(m_1 m_2)^{1/4} + r m_1^{1/2}\right]\omega_{\max}\log m$. Inequality (39b) and the fact $r_1 \in \mathcal{A}$ together imply that

$$\left\|\boldsymbol{U}_1^{\text{oracle}}\right\|_{2,\infty} = \left\|\boldsymbol{U}_1^{\text{oracle}}\boldsymbol{U}_1^{\text{oracle}\top}\right\|_{2,\infty} \le \left\|\boldsymbol{U}_1^{\text{oracle}}\boldsymbol{U}_1^{\text{oracle}\top} - \boldsymbol{U}_1^\star\boldsymbol{U}_1^{\star\top}\right\|_{2,\infty} + \left\|\boldsymbol{U}_1^\star\right\|_{2,\infty} \le 2\sqrt{\frac{\mu r^3}{m_1}}. \tag{195}$$

Combining (194) and (195), one can further obtain that

$$\left\|\boldsymbol{U}_1^0\right\|_{2,\infty} = \left\|\boldsymbol{U}_1^0\boldsymbol{U}_1^{0\top}\right\|_{2,\infty} \le \left\|\boldsymbol{U}_1^0\boldsymbol{U}_1^{0\top} - \boldsymbol{U}_1^{\text{oracle}}\boldsymbol{U}_1^{\text{oracle}\top}\right\| + \left\|\boldsymbol{U}_1^{\text{oracle}}\right\|_{2,\infty} \le 3\sqrt{\frac{\mu r^3}{m_1}} \le \frac{1}{4e}, \tag{196}$$

i.e., (189d) holds for $t = 0$.

Step 2.2: induction step $(t > 0)$ for (189a)-(189d). Suppose that (189a)-(189d) hold for $t = t'$. We aim to show that they continue to hold for $t = t' + 1$.

Recognizing that $\boldsymbol{U}_1^{t'}$ is the top-r_1 singular space of

$$\boldsymbol{G}_1^{t'} = \mathcal{P}_{\widetilde{\boldsymbol{U}}_1}\widetilde{\boldsymbol{M}} + \left(\boldsymbol{G}_1^{t'} - \mathcal{P}_{\widetilde{\boldsymbol{U}}_1}\widetilde{\boldsymbol{M}}\right),$$

we have

$$
\begin{aligned}
F_1^{t'+1} &= \left\|\mathcal{P}_{\mathsf{diag}}\big(\boldsymbol{G}_1^{t'+1} - \widetilde{\boldsymbol{M}}\big)\right\| \\
&= \left\|\mathcal{P}_{\mathsf{diag}}\big(\mathcal{P}_{\boldsymbol{U}_1^{t'}}\boldsymbol{G}_1^{t'} - \widetilde{\boldsymbol{M}}\big)\right\| \\
&\le \left\|\mathcal{P}_{\mathsf{diag}}\left(\mathcal{P}_{\boldsymbol{U}_1^{t'}}(\boldsymbol{G}_1^{t'} - \widetilde{\boldsymbol{M}})\right)\right\| + \left\|\mathcal{P}_{\mathsf{diag}}\left(\mathcal{P}_{(\boldsymbol{U}_1^{t'})_\perp}\widetilde{\boldsymbol{M}}\mathcal{P}_{\widetilde{\boldsymbol{U}}}\right)\right\| \\
&\le \left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty}\left\|\boldsymbol{G}_1^{t'} - \widetilde{\boldsymbol{M}}\right\| + \left\|\widetilde{\boldsymbol{U}}\right\|_{2,\infty}\left\|(\boldsymbol{U}_1^{t'})_\perp\widetilde{\boldsymbol{M}}\right\| \\
&\le \left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty}L_1^{t'} + 4\sqrt{\frac{\mu r}{m_1}}\left(\left\|(\boldsymbol{U}_1^{t'})_\perp\mathcal{P}_{\widetilde{\boldsymbol{U}}_1}\widetilde{\boldsymbol{M}}\right\| + \left\|\mathcal{P}_{\widetilde{\boldsymbol{U}}_{:,r_1+1:r}}\widetilde{\boldsymbol{M}}\right\|\right) \\
&\le \left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty}L_1^{t'} + 4\sqrt{\frac{\mu r}{m_1}}\left(2\left\|\boldsymbol{G}_1^{t'} - \mathcal{P}_{\widetilde{\boldsymbol{U}}_1}\widetilde{\boldsymbol{M}}\right\| + \left\|\mathcal{P}_{\widetilde{\boldsymbol{U}}_{:,r_1+1:r}}\widetilde{\boldsymbol{M}}\right\|\right) \\
&\le \left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty}L_1^{t'} + 4\sqrt{\frac{\mu r}{m_1}}\left(2\left\|\boldsymbol{G}_1^{t'} - \widetilde{\boldsymbol{M}}\right\| + 3\left\|\mathcal{P}_{\widetilde{\boldsymbol{U}}_{:,r_1+1:r}}\widetilde{\boldsymbol{M}}\right\|\right) \\
&\le \left(\left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty} + 8\sqrt{\frac{\mu r}{m_1}}\right)L_1^{t'} + 12\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2.
\end{aligned}
$$
(197)

Here, the second line holds since $\mathcal{P}_{\mathsf{diag}}(\boldsymbol{G}_1^{t'+1}) = \mathcal{P}_{\mathsf{diag}}(\boldsymbol{U}_1^{t'}\boldsymbol{\Lambda}_1^{t'}\boldsymbol{U}_1^{t'\top}) = \mathcal{P}_{\mathsf{diag}}(\mathcal{P}_{\boldsymbol{U}_1^{t'}}\boldsymbol{G}_1^{t'})$; the fourth line comes from Zhang et al. (2022, Lemma 1); the fifth line makes use of (177); the sixth line applies Zhou and Chen (2023, Lemma 8); and the penultimate line invokes the triangle inequality. Note that

$$
\begin{aligned}
L_1^{t'} &\le F_1^{t'} + \left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\big(\boldsymbol{G}_1^{t'+1} - \widetilde{\boldsymbol{M}}\big)\right\| = F_1^{t'} + \left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\big(\boldsymbol{M}^{\mathsf{oracle}} - \widetilde{\boldsymbol{M}}\big)\right\| \\
&\le F_1^{t'} + \left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{Z}\right)\right\| \le F_1^{t'} + \left\|\boldsymbol{Z}\right\| + \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{Z}\right)\right\| \le F_1^{t'} + 2\left\|\boldsymbol{Z}\right\|.
\end{aligned}
$$

Inequality (197) taken together with the previous inequality leads us to

$$
\begin{aligned}
F_1^{t'+1} &\le \left(\left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty} + 8\sqrt{\frac{\mu r}{m_1}}\right)F_1^{t'} + 2\left(\left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty} + 8\sqrt{\frac{\mu r}{m_1}}\right)\|\boldsymbol{Z}\| + 12\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2 \\
&\overset{\text{(189d)}}{\le} \left(\frac{1}{4e} + \frac{1}{4e}\right)F_1^{t'} + 2\left(\left\|\boldsymbol{U}_1^{t}\boldsymbol{U}_1^{t\top} - \boldsymbol{U}_1^{\mathsf{oracle}}\boldsymbol{U}_1^{\mathsf{oracle}\top}\right\| + \left\|\boldsymbol{U}_1^{\mathsf{oracle}}\right\|_{2,\infty} + 8\sqrt{\frac{\mu r}{m_1}}\right)\|\boldsymbol{Z}\| + 12\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2 \\
&\overset{\text{(189c) and (195)}}{\le} \frac{1}{2e}F_1^{t'} + 2\left(2\frac{D_1^{t'}}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)} + 2\sqrt{\frac{\mu r^3}{m_1}} + 8\sqrt{\frac{\mu r}{m_1}}\right)\|\boldsymbol{Z}\| + 12\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2 \\
&\overset{\text{(189b)}}{\le} \frac{1}{2e}F_1^{t'} + 2\left(2\frac{F_1^{t'}}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)} + 10\sqrt{\frac{\mu r^3}{m_1}}\right)\|\boldsymbol{Z}\| + 12\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2 \\
&\quad + 4\frac{6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\widetilde{r}+1}^\star + C_3^2\mu r\omega_{\mathsf{max}}^2\log^2 m}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}\|\boldsymbol{Z}\| \\
&\overset{\text{(192) and (193)}}{\le} \frac{1}{e}F_1^{t'} + 20\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| + 12\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2 + C_3^3\frac{\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\widetilde{r}+1}^\star + \mu r\omega_{\mathsf{max}}^2\log^2 m}{\sigma_{r_1}^{\star 2}/r}\|\boldsymbol{Z}\| \\
&\le \frac{1}{e}F_1^{t'} + 21\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| + 12\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2,
\end{aligned}
$$
(198)

where the last inequality is a consequence of $\sigma_{r_1}^\star \geq \sigma_{\bar r}^\star \geq C_0 r[(m_1 m_2)^{1/4} + r m_1^{1/2}]\omega_{\max}\log m$. Then one immediately has

$$F_1^{t'+1} - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde\sigma_{r_1+1}^2 \leq \frac{1}{e}\left(F_1^{t'} - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde\sigma_{r_1+1}^2\right)$$

$$\leq \frac{1}{e^{t'+1}}\left(F_1^0 - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde\sigma_{r_1+1}^2\right),$$

which confirms that (189a) holds for $t = t'+1$.

In addition, we can prove (189b) for $t = t'+1$ by using the same argument as in (191). Combining (189a), (189b) for $t = t'+1$ and Weyl's inequality, we further have

$$\left\|\boldsymbol{U}_1^{t'+1}\boldsymbol{U}_1^{t'+1\top} - \boldsymbol{U}_1^{\mathsf{oracle}}\boldsymbol{U}_1^{\mathsf{oracle}\top}\right\|$$

$$\leq 2\frac{D_1^{t'+1}}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})}$$

$$\leq 2\frac{F_1^{t'+1}}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})} + 2\frac{6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\bar r+1}^\star + C_3^2\mu r\omega_{\max}^2\log^2 m}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})}$$

$$= 2\frac{F_1^{t'+1} - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde\sigma_{r_1+1}^2}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})} + 2\frac{6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\bar r+1}^\star + C_3^2\mu r\omega_{\max}^2\log^2 m}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})}$$

$$+ \frac{80\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| + 40\sqrt{\frac{\mu r}{m_1}}\widetilde\sigma_{r_1+1}^2}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})}$$

$$\overset{(193)}{\lesssim} \frac{1}{e^{t'+1}}\frac{F_1^0}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})} + \frac{\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| + \sqrt{\frac{\mu r}{m_1}}\widetilde\sigma_{r_1+1}^2}{\widetilde\sigma_{r_1}^2 - \widetilde\sigma_{r_1+1}^2}$$

$$2\frac{6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\bar r+1}^\star + C_3^2\mu r\omega_{\max}^2\log^2 m}{\sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2}}$$

$$\overset{(190),(192)\text{ and }(193)}{\ll} \sqrt{\frac{\mu r^3}{m_1}} \ll \frac{1}{8},$$

which validates (189c) for $t = t'+1$.

Putting the previous inequality and (195) together, one can prove that (189d) also holds for $t = t'+1$:

$$\left\|\boldsymbol{U}_1^{t'+1}\right\|_{2,\infty} = \left\|\boldsymbol{U}_1^{t'+1}\boldsymbol{U}_1^{t'+1\top}\right\|_{2,\infty} \leq \left\|\boldsymbol{U}_1^{t'+1}\boldsymbol{U}_1^{t'+1\top} - \boldsymbol{U}_1^{\mathsf{oracle}}\boldsymbol{U}_1^{\mathsf{oracle}\top}\right\| + \left\|\boldsymbol{U}_1^{\mathsf{oracle}}\right\|_{2,\infty} \leq 3\sqrt{\frac{\mu r^3}{m_1}} \leq \frac{1}{4e}.$$

Therefore, we have completed the proof of the induction step for (189a) - (189d).

Step 3: bounding D_k^t for $k > 1$ After establishing upper bounds on $\{D_1^t\}$, we now turn attention to the quantities $\{D_k^t\}_{k>1}$. By setting

$$t_1 \geq \log\left(C\frac{\sigma_1^{\star 2}}{\sigma_{r_1+1}^{\star 2} + \omega_{\max}^2}\right),$$

we can show that

$$F_2^0 = F_1^{t_1} \overset{(189a)}{\leq} 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| + 20\sqrt{\frac{\mu r}{m_1}}\widetilde\sigma_{r_1+1}^2 + \frac{1}{e^{t_1}}F_1^0$$

$$\overset{(191)}{\leq} 40\sqrt{\frac{\mu r^3}{m_1}}\,\|\boldsymbol{Z}\| + 20\sqrt{\frac{\mu r}{m_1}}\,\widetilde{\sigma}_{r_1+1}^2 + \frac{\sigma_{r_1+1}^{\star 2} + \omega_{\max}^2}{C\sigma_1^{\star 2}}\cdot 16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2$$

$$\overset{(107) \text{ and } (192)}{\leq} 40C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\bar{r}+1}^\star + (\sqrt{m_1 m_2}+m_1)\,\omega_{\max}^2\log^2 m\right)$$

$$+ 20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2 + \frac{1}{C}\sqrt{\frac{\mu r}{m_1}}\left(\omega_{\max}^2 + \sigma_{r_1+1}^{\star 2}\right)$$

$$\leq 41C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\bar{r}+1}^\star + (\sqrt{m_1 m_2}+m_1)\,\omega_{\max}^2\log^2 m\right) + 21\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2. \quad (199)$$

Here, the last line makes use of the following inequality:

$$\sigma_{r_1+1}^{\star 2} \overset{(90a) \text{ and } (103)}{\leq} \left(\widetilde{\sigma}_{r_1+1} + 2\sqrt{C_5}\sqrt{m_1}\omega_{\max}\log m\right)^2 \overset{\text{Cauchy-Schwarz}}{\leq} 2\widetilde{\sigma}_{r_1+1}^2 + 8C_5 m_1\omega_{\max}^2\log^2 m.$$

We define

$$\mathcal{R}_k := \left\{ r' : \frac{\sigma_{r_{k-1}+1}(\boldsymbol{G}_0)}{\sigma_{r'}(\boldsymbol{G}_0)} \leq 4 \quad\text{and}\quad \sigma_{r'}(\boldsymbol{G}_0) - \sigma_{r'+1}(\boldsymbol{G}_0) \geq \frac{1}{r}\sigma_{r'}(\boldsymbol{G}_0) \right\}, \quad (200)$$

Choosing the numbers of iterations $\{t_i\}$ as in (41a) and (41b) and repeating similar arguments as in (175), (189a) - (189d), (192), (193) and (199), we know that for all $1 \leq k \leq k_{\max}$ and $1 \leq t \leq t_k$,

$$r_k \in \mathcal{R}_k \cap \mathcal{A}, \quad (201\text{a})$$

$$\max\left\{\frac{\sigma_{r_{k-1}+1}^\star}{\sigma_{r_k}^\star}, \frac{\widetilde{\sigma}_{r_{k-1}+1}}{\widetilde{\sigma}_{r_k}}\right\} \leq 3 \quad\text{and}\quad \min\left\{\frac{\widetilde{\sigma}_{r_k}^2 - \widetilde{\sigma}_{r_k+1}^2}{\widetilde{\sigma}_{r_k}^2}, \frac{\sigma_{r_k}^{\star 2} - \sigma_{r_k+1}^{\star 2}}{\sigma_{r_k}^{\star 2}}\right\} \geq \frac{1}{4r}, \quad (201\text{b})$$

$$\lambda_{r_k}\left(\boldsymbol{M}^{\text{oracle}}\right) - \lambda_{r_k+1}\left(\boldsymbol{M}^{\text{oracle}}\right) \asymp \widetilde{\sigma}_{r_k}^2 - \widetilde{\sigma}_{r_k+1}^2 \asymp \sigma_{r_k}^{\star 2} - \sigma_{r_k+1}^{\star 2} \gg \|\boldsymbol{Z}\|, \quad (201\text{c})$$

$$F_k^t - 40\sqrt{\frac{\mu r^3}{m_1}}\,\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_k+1}^2 \leq \frac{1}{e^t}\left(F_k^0 - 40\sqrt{\frac{\mu r^3}{m_1}}\,\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_k+1}^2\right), \quad (201\text{d})$$

$$D_k^t \leq F_k^t + 6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\bar{r}+1}^\star + C_3^2\mu r\omega_{\max}^2\log^2 m, \quad (201\text{e})$$

$$\left\|\boldsymbol{U}_k^t\boldsymbol{U}_k^{t\top} - \boldsymbol{U}_k^{\text{oracle}}\boldsymbol{U}_k^{\text{oracle}\top}\right\| \leq 2\frac{D_k^t}{\lambda_{r_k}\left(\boldsymbol{M}^{\text{oracle}}\right) - \lambda_{r_k+1}\left(\boldsymbol{M}^{\text{oracle}}\right)} \leq \frac{1}{8}, \quad (201\text{f})$$

$$\left\|\boldsymbol{U}_k^t\right\|_{2,\infty} \leq \left\|\boldsymbol{U}_k^t\boldsymbol{U}_k^{t\top} - \boldsymbol{U}_k^{\text{oracle}}\boldsymbol{U}_k^{\text{oracle}\top}\right\| + \left\|\boldsymbol{U}_k^{\text{oracle}}\right\|_{2,\infty} \leq \frac{1}{4e}. \quad (201\text{g})$$

$$F_{k+1}^0 = F_k^{t_k} \leq 41C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\bar{r}+1}^\star + (\sqrt{m_1 m_2}+m_1)\,\omega_{\max}^2\log^2 m\right) + 21\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_k+1}^2. \quad (201\text{h})$$

Taking $k = k_{\max}$ in (201h) yields that

$$F_{k_{\max}}^{t_{k_{\max}}} \leq 41C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\bar{r}+1}^\star + (\sqrt{m_1 m_2}+m_1)\,\omega_{\max}^2\log^2 m\right) + 21\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_{k_{\max}}+1}^2. \quad (202)$$

This together with (201e) implies that

$$D_{k_{\max}}^{t_{k_{\max}}} \leq 42C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\bar{r}+1}^\star + (\sqrt{m_1 m_2}+m_1)\,\omega_{\max}^2\log^2 m\right) + 21\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_{k_{\max}}+1}^2. \quad (203)$$

Recall that $r_{k_{\max}}$ satisfies $r_{k_{\max}} = r$ or $\sigma_{r_{k_{\max}}+1}(\boldsymbol{G}_{k_{\max}}) = \sigma_{r_{k_{\max}}+1}(\boldsymbol{G}_{k_{\max}}^{t_{k_{\max}}}) \leq \tau$.

1. If $r_{k_{\max}} = r$, then it follows that

$$\widetilde{\sigma}_{r_{k_{\max}}+1}^2 = \widetilde{\sigma}_{r+1}^2 = 0.$$

2. If $\sigma_{r_{k_{\max}}+1}(\boldsymbol{G}_{k_{\max}}) \leq \tau$, then Weyl's inequality and (202) together show that

$$
\begin{aligned}
\widetilde{\sigma}^2_{r_{k_{\max}}+1} = \sigma_{r_{k_{\max}}+1}(\widetilde{\boldsymbol{M}}) &\leq \sigma_{r_{k_{\max}}+1}\left(\boldsymbol{G}^{t_{k_{\max}}}_{k_{\max}}\right) + \left\|\widetilde{\boldsymbol{M}} - \boldsymbol{G}^{t_{k_{\max}}}_{k_{\max}}\right\| \\
&= \sigma_{r_{k_{\max}}+1}\left(\boldsymbol{G}^{t_{k_{\max}}}_{k_{\max}}\right) + F^{t_{k_{\max}}}_{k_{\max}} \\
&\leq \tau + 41 C_2 \sqrt{\frac{\mu r^3}{m_1}} \left(\sqrt{m_1}\omega_{\max}\log m \cdot \sigma^{\star}_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\max}\log^2 m\right) + 21\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}^2_{r_{k_{\max}}+1} \\
&\leq \tau + 41 C_2 \sqrt{\frac{\mu r^3}{m_1}} \left(\sqrt{m_1}\omega_{\max}\log m \cdot \sigma^{\star}_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\max}\log^2 m\right) + \frac{1}{2}\widetilde{\sigma}^2_{r_{k_{\max}}+1},
\end{aligned}
$$

which further gives

$$
\widetilde{\sigma}^2_{r_{k_{\max}}+1} \leq 2\tau + 82 C_2 \sqrt{\frac{\mu r^3}{m_1}} \left(\sqrt{m_1}\omega_{\max}\log m \cdot \sigma^{\star}_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\max}\log^2 m\right) \leq 3\tau. \tag{204}
$$

Therefore, inequality (204) is guaranteed to hold.

Step 4: proving (42a). We know from (201a) that $r_{k_{\max}} \in \mathcal{A}$. Also, (39b) tell us that

$$
\left\|\boldsymbol{U}^{\mathsf{oracle}}_{k_{\max}}\boldsymbol{U}^{\mathsf{oracle}\top}_{k_{\max}} - \boldsymbol{U}^{\star}_{k_{\max}}\boldsymbol{U}^{\star\top}_{k_{\max}}\right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{r^2\sqrt{m_1}\omega_{\max}\log m}{\sigma^{\star}_{r_{k_{\max}}}} + \frac{r^2\sqrt{m_1 m_2}\omega^2_{\max}\log^2 m}{\sigma^{\star 2}_{r_{k_{\max}}}}\right) \leq \sqrt{\frac{\mu r^3}{m_1}}.
$$

In view of (203), (201b), (201c) and (201g), we can demonstrate that

$$
\begin{aligned}
&\left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}^{\top}_{k_{\max}} - \boldsymbol{U}^{\mathsf{oracle}}_{k_{\max}}\boldsymbol{U}^{\mathsf{oracle}\top}_{k_{\max}}\right\| \\
&\lesssim \frac{D^{t_{k_{\max}}}_{k_{\max}}}{\lambda_{r_{k_{\max}}}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_{k_{\max}}+1}(\boldsymbol{M}^{\mathsf{oracle}})} \\
&\lesssim \frac{\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m \cdot \sigma^{\star}_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\max}\log^2 m\right)}{\sigma^{\star 2}_{r_{k_{\max}}} - \sigma^{\star 2}_{r_{k_{\max}}+1}} + \frac{\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}^2_{r_{k_{\max}}+1}}{\widetilde{\sigma}^2_{r_{k_{\max}}} - \widetilde{\sigma}^2_{r_{k_{\max}}+1}} \\
&\lesssim \frac{\sqrt{\frac{\mu r^3}{m_1}}r\left(\sqrt{m_1}\omega_{\max}\log m \cdot \sigma^{\star}_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\max}\log^2 m\right)}{\sigma^{\star 2}_{r_{k_{\max}}}} + \frac{\sqrt{\frac{\mu r^3}{m_1}}\widetilde{\sigma}^2_{r_{k_{\max}}+1}}{\widetilde{\sigma}^2_{r_{k_{\max}}}} \\
&\lesssim \sqrt{\frac{\mu r^3}{m_1}}. \tag{205}
\end{aligned}
$$

Here, the last inequality holds since $\widetilde{\sigma}_{r_{k_{\max}}} \geq \widetilde{\sigma}_{r_{k_{\max}}+1}$ and $\sigma^{\star}_{r_{k_{\max}}} \geq \sigma^{\star}_{\bar{r}} \geq C_0 r\left[(m_1 m_2)^{1/4} + r m_1^{1/2}\right]\omega_{\max}\log m$. Combining the previous two inequalities yields

$$
\begin{aligned}
\left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}^{\top}_{k_{\max}} - \boldsymbol{U}^{\star}_{k_{\max}}\boldsymbol{U}^{\star\top}_{k_{\max}}\right\|_{2,\infty} &\leq \left\|\boldsymbol{U}^{\mathsf{oracle}}_{k_{\max}}\boldsymbol{U}^{\mathsf{oracle}\top}_{k_{\max}} - \boldsymbol{U}^{\star}_{k_{\max}}\boldsymbol{U}^{\star\top}_{k_{\max}}\right\|_{2,\infty} + \left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}^{\top}_{k_{\max}} - \boldsymbol{U}^{\mathsf{oracle}}_{k_{\max}}\boldsymbol{U}^{\mathsf{oracle}\top}_{k_{\max}}\right\| \\
&\lesssim \sqrt{\frac{\mu r^3}{m_1}}, \tag{206}
\end{aligned}
$$

which validates (42a).

Step 5: proving (42b) **and** (42c). Note that

$$
\begin{aligned}
&\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}^{\top}_{k_{\max}} - \boldsymbol{U}^{\star}_{k_{\max}}\boldsymbol{U}^{\star\top}_{k_{\max}}\right)\boldsymbol{X}^{\star}\right\|_{2,\infty} \\
&= \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}^{\top}_{k_{\max}} - \boldsymbol{U}^{\star}_{k_{\max}}\boldsymbol{U}^{\star\top}_{k_{\max}}\right)\left(\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top} + \boldsymbol{U}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\right)\right\|_{2,\infty}
\end{aligned}
$$

$$
\begin{aligned}
&= \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty} \\
&\quad + \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\boldsymbol{U}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\right\|_{2,\infty} \\
&\leq \underbrace{\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^{\top}\right)\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty}}_{=:\alpha_1} \\
&\quad + \underbrace{\left\|\left(\widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty}}_{=:\alpha_2} \\
&\quad + \underbrace{\left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right\|_{2,\infty}\sigma_{\widetilde{r}+1}^{\star}}_{=:\alpha_3},
\end{aligned}
\tag{207}
$$

where the last line holds due to $\|\boldsymbol{U}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\| = \|\boldsymbol{\Sigma}^{\star(2)}\| = \sigma_{\widetilde{r}+1}^{\star}$. To control $\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\boldsymbol{X}^{\star}\right\|_{2,\infty}$, one only needs to bound α_1, α_2 and α_3, respectively.

Step 5.1: bounding α_1. We first bound $\alpha_1 = \|(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^{\top})\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\|_{2,\infty}$.

Step 5.1.1: bounding $\|(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top})\widetilde{\boldsymbol{M}}\|_{2,\infty}$. By virtue of Weyl's inequality and (107), one has

$$
\lambda_{r_{k_{\max}}}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) \leq \widetilde{\sigma}_{r_{k_{\max}}}^2 + \|\boldsymbol{Z}\| \asymp \widetilde{\sigma}_{r_{k_{\max}}}^2 \asymp \sigma_{r_{k_{\max}}}^{\star 2}.
\tag{208}
$$

Recognizing that $\boldsymbol{G}_{k_{\max}}^{t_{k_{\max}}} = \boldsymbol{M}^{\mathsf{oracle}} + (\boldsymbol{G}_{k_{\max}}^{t_{k_{\max}}} - \boldsymbol{M}^{\mathsf{oracle}})$ and $\boldsymbol{U}_{k_{\max}}$ (resp. $\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}$) is the rank-$r_{k_{\max}}$ leading singular subspace of $\boldsymbol{G}_{k_{\max}}^{t_{k_{\max}}}$ (resp. $\boldsymbol{M}^{\mathsf{oracle}}$), we invoke Lemma 6 to yield

$$
\begin{aligned}
&\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\boldsymbol{M}^{\mathsf{oracle}}\right\|_{2,\infty} \\
&\leq \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\boldsymbol{M}^{\mathsf{oracle}}\right\| \\
&\lesssim \frac{\lambda_{r_{k_{\max}}}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}{\lambda_{r_{k_{\max}}}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_{k_{\max}}+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}\left\|\boldsymbol{G}_{k_{\max}}^{t_{k_{\max}}} - \boldsymbol{M}^{\mathsf{oracle}}\right\| \\
&\overset{\text{(201c) and (208)}}{\lesssim} \frac{\sigma_{r_{k_{\max}}}^{\star 2}}{\sigma_{r_{k_{\max}}}^{\star 2} - \sigma_{r_{k_{\max}}+1}^{\star 2}}D_{k_{\max}}^{t_{k_{\max}}} \\
&\overset{\text{(201b) and (203)}}{\lesssim} r\left[\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m \cdot \sigma_{\widetilde{r}+1}^{\star} + (\sqrt{m_1 m_2} + m_1)\omega_{\max}^2\log^2 m\right) + \sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_{k_{\max}}+1}^2\right].
\end{aligned}
\tag{209}
$$

Combining (107), (205) and (209) leads to

$$
\begin{aligned}
&\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\widetilde{\boldsymbol{M}}\right\|_{2,\infty} \\
&\leq \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\boldsymbol{M}^{\mathsf{oracle}}\right\|_{2,\infty} + \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\boldsymbol{Z}\right\|_{2,\infty} \\
&\leq \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\boldsymbol{M}^{\mathsf{oracle}}\right\|_{2,\infty} + \left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right\|_{2,\infty}\|\boldsymbol{Z}\| \\
&\lesssim r\left[\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m \cdot \sigma_{\widetilde{r}+1}^{\star} + (\sqrt{m_1 m_2} + m_1)\omega_{\max}^2\log^2 m\right) + \sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_{k_{\max}}+1}^2\right] \\
&\quad + \sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m \cdot \sigma_{\widetilde{r}+1}^{\star} + (\sqrt{m_1 m_2} + m_1)\omega_{\max}^2\log^2 m\right) \\
&\lesssim \sqrt{\frac{\mu r^5}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m \cdot \sigma_{\widetilde{r}+1}^{\star} + (\sqrt{m_1 m_2} + m_1)\omega_{\max}^2\log^2 m\right) + \sqrt{\frac{\mu r^3}{m_1}}\widetilde{\sigma}_{r_{k_{\max}}+1}^2.
\end{aligned}
\tag{210}
$$

58

Step 5.1.2: bounding $\left\|(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top)\widetilde{\boldsymbol{M}}\right\|_{2,\infty}$. Recalling that (107) holds, we can invoke Lemma 1 to obtain

$$
\left\|\left(\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right)\widetilde{\boldsymbol{M}}\right\|_{2,\infty}
$$
$$
\leq \frac{40}{\pi}\widetilde{\sigma}_{r_{k_{\max}}}^2 \sum_{k\geq 1}\frac{2^k}{\left(\widetilde{\sigma}_{r_{k_{\max}}}^2 - \widetilde{\sigma}_{r_{k_{\max}+1}}^2\right)^k}\sum_{\substack{0\leq j_1,\ldots,j_{k+1}\leq r\\ (j_1,\ldots,j_{k+1})^\top\neq \mathbf{0}_{k+1}}}\left\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}. \tag{211}
$$

Here, we recall that $\widetilde{\boldsymbol{P}}_j = \widetilde{\boldsymbol{u}}_j\widetilde{\boldsymbol{u}}_j^\top$ for $1\leq j\leq r$ and $\widetilde{\boldsymbol{P}}_0 = \widetilde{\boldsymbol{U}}_\perp\widetilde{\boldsymbol{U}}_\perp^\top$. Repeat similar arguments as in (120) to deduce that

$$
\left\|\left(\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right)\widetilde{\boldsymbol{M}}\right\|_{2,\infty}
$$
$$
\lesssim \sqrt{\frac{\mu r^3}{m_1}}\frac{r\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star + \left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)}{\sigma_{r_{k_{\max}}}^{\star 2} - \sigma_{r_{k_{\max}+1}}^{\star 2}}\cdot\widetilde{\sigma}_{r_{k_{\max}}}^2
$$
$$
\overset{\text{(201b) and (180)}}{\lesssim}\sqrt{\frac{\mu r^3}{m_1}}\frac{r\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star + \left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)}{\sigma_{r_{k_{\max}}}^{\star 2}/r}\cdot\sigma_{r_{k_{\max}}}^{\star 2}
$$
$$
= \sqrt{\frac{\mu r^3}{m_1}}r^2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star + \left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right). \tag{212}
$$

Inequality (210) taken together with (212) and the triangle inequality shows that

$$
\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right)\widetilde{\boldsymbol{M}}\right\|_{2,\infty}
$$
$$
\lesssim \sqrt{\frac{\mu r^3}{m_1}}r^2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star + \left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right) + \sqrt{\frac{\mu r^3}{m_1}}\widetilde{\sigma}_{r_{k_{\max}+1}}^2. \tag{213}
$$

Step 5.1.3: bounding α_1. Equipped with (213), we are now ready to bound α_1. Recall that

$$
\widetilde{\boldsymbol{M}} = \widetilde{\boldsymbol{U}}^{(1)}\left(\widetilde{\boldsymbol{\Sigma}}^{(1)}\right)^2\widetilde{\boldsymbol{U}}^{(1)\top} + \widetilde{\boldsymbol{U}}^{(2)}\left(\widetilde{\boldsymbol{\Sigma}}^{(2)}\right)^2\widetilde{\boldsymbol{U}}^{(2)\top},
$$

where $\widetilde{\boldsymbol{U}}^{(1)}$ and $\widetilde{\boldsymbol{\Sigma}}^{(1)}$ (resp. $\widetilde{\boldsymbol{U}}^{(2)}$ and $\widetilde{\boldsymbol{\Sigma}}^{(2)}$) are defined in (82) (resp. (95)). In view of (205) and (213), one can obtain

$$
\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right)\widetilde{\boldsymbol{U}}^{(1)}\widetilde{\boldsymbol{\Sigma}}^{(1)}\right\|_{2,\infty}
$$
$$
\leq \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right)\widetilde{\boldsymbol{U}}^{(1)}\left(\widetilde{\boldsymbol{\Sigma}}^{(1)}\right)^2\widetilde{\boldsymbol{U}}^{(1)\top}\right\|_{2,\infty}\left\|\left(\widetilde{\boldsymbol{\Sigma}}^{(1)}\right)^{-1}\right\|
$$
$$
\overset{(i)}{\lesssim}\left(\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right)\widetilde{\boldsymbol{M}}\right\|_{2,\infty} + \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right)\widetilde{\boldsymbol{U}}^{(2)}\left(\widetilde{\boldsymbol{\Sigma}}^{(2)}\right)^2\widetilde{\boldsymbol{U}}^{(2)\top}\right\|_{2,\infty}\right)\frac{1}{\sigma_{\bar{r}}^\star}
$$
$$
\leq \left(\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right)\widetilde{\boldsymbol{M}}\right\|_{2,\infty} + \left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right\|_{2,\infty}\left\|\widetilde{\boldsymbol{\Sigma}}^{(2)}\right\|^2\right)\frac{1}{\sigma_{\bar{r}}^\star}
$$
$$
\lesssim \left(\sqrt{\frac{\mu r^3}{m_1}}r^2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star + \left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right) + \sqrt{\frac{\mu r^3}{m_1}}\widetilde{\sigma}_{r_{k_{\max}+1}}^2\right.
$$
$$
\left. + \sqrt{\frac{\mu r^3}{m_1}}\left(\frac{r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma_{r_{k_{\max}}}^\star} + \frac{r^2\sqrt{m_1 m_2}\omega_{\mathsf{max}}^2\log^2 m}{\sigma_{r_{k_{\max}}}^{\star 2}}\right)\widetilde{\sigma}_{\bar{r}+1}^2\right)\frac{1}{\sigma_{\bar{r}}^\star}
$$
$$
\overset{(ii)}{\lesssim}\left(\sqrt{\frac{\mu r^3}{m_1}}r^2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star + \left(\sqrt{m_1 m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right) + \sqrt{\frac{\mu r^3}{m_1}}\widetilde{\sigma}_{r_{k_{\max}+1}}^2\right.
$$
$$
\left. + \sqrt{\frac{\mu r^3}{m_1}}\left(\frac{r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma_{r_{k_{\max}}}^\star} + \frac{r^2\sqrt{m_1 m_2}\omega_{\mathsf{max}}^2\log^2 m}{\sigma_{r_{k_{\max}}}^{\star 2}}\right)\sigma_{\bar{r}}^{\star 2}\right)\frac{1}{\sigma_{\bar{r}}^\star}
$$

$$\overset{(iii)}{\lesssim} \sqrt{\frac{\mu r^3}{m_1}} r^2 \left(\sqrt{m_1} \omega_{\mathsf{max}} \log m + \frac{(\sqrt{m_1 m_2} + m_1) \omega_{\mathsf{max}}^2 \log^2 m}{\sigma_{\overline{r}}^\star} \right) + \sqrt{\frac{\mu r^3}{m_1}} \frac{\widetilde{\sigma}_{r_{k_{\mathsf{max}}}+1}^2}{\sigma_{\overline{r}}^\star}$$

$$\overset{(iv)}{\lesssim} \sqrt{\frac{\mu r^3}{m_1}} \left(r^2 \sqrt{m_1} \omega_{\mathsf{max}} \log m + r (m_1 m_2)^{1/4} \omega_{\mathsf{max}} \log m \right) + \sqrt{\frac{\mu r^3}{m_1}} \frac{\tau}{\sigma_{\overline{r}}^\star}$$

$$\asymp \sqrt{\frac{\mu r^3}{m_1}} \left(r^2 \sqrt{m_1} \omega_{\mathsf{max}} \log m + r (m_1 m_2)^{1/4} \omega_{\mathsf{max}} \log m \right), \tag{214}$$

where (i), (ii) and (iii) are consequences of $\widetilde{\sigma}_{r_{k_{\mathsf{max}}}} \overset{(180)}{\asymp} \sigma_{r_{k_{\mathsf{max}}}}^\star \geq \sigma_{\overline{r}}^\star \asymp \widetilde{\sigma}_{\overline{r}} \gtrsim \max\{\sigma_{\overline{r}+1}^\star, \widetilde{\sigma}_{\overline{r}+1}\}$, and (iv) makes use of the fact $\sigma_{\overline{r}}^\star \geq C_0 r [(m_1 m_2)^{1/4} + r m_1^{1/2}] \omega_{\mathsf{max}} \log m$ and (204). Note that $\widetilde{U}^{(1)} \widetilde{\Sigma}^{(1)} \widetilde{W}^{(1)\top}$ is the SVD of $U^{\star(1)} \Sigma^{\star(1)} + E V^{\star(1)}$. By virtue of the previous inequality, (90a), Theorem 5 and the triangle inequality, we arrive at

$$\alpha_1 = \left\| (U_{k_{\mathsf{max}}} U_{k_{\mathsf{max}}}^\top - \widetilde{U}_{k_{\mathsf{max}}} \widetilde{U}_{k_{\mathsf{max}}}^\top) U^{\star(1)} \Sigma^{\star(1)} V^{\star(1)\top} \right\|_{2,\infty}$$

$$= \left\| (U_{k_{\mathsf{max}}} U_{k_{\mathsf{max}}}^\top - \widetilde{U}_{k_{\mathsf{max}}} \widetilde{U}_{k_{\mathsf{max}}}^\top) U^{\star(1)} \Sigma^{\star(1)} \right\|_{2,\infty}$$

$$\leq \left\| (U_{k_{\mathsf{max}}} U_{k_{\mathsf{max}}}^\top - \widetilde{U}_{k_{\mathsf{max}}} \widetilde{U}_{k_{\mathsf{max}}}^\top) \widetilde{U}^{(1)} \widetilde{\Sigma}^{(1)} \widetilde{W}^\top \right\|_{2,\infty} + \left\| (U_{k_{\mathsf{max}}} U_{k_{\mathsf{max}}}^\top - \widetilde{U}_{k_{\mathsf{max}}} \widetilde{U}_{k_{\mathsf{max}}}^\top) E V^{\star(1)} \right\|_{2,\infty}$$

$$\leq \left\| (U_{k_{\mathsf{max}}} U_{k_{\mathsf{max}}}^\top - \widetilde{U}_{k_{\mathsf{max}}} \widetilde{U}_{k_{\mathsf{max}}}^\top) \widetilde{U}^{(1)} \widetilde{\Sigma}^{(1)} \right\|_{2,\infty} + \left\| U_{k_{\mathsf{max}}} U_{k_{\mathsf{max}}}^\top - \widetilde{U}_{k_{\mathsf{max}}} \widetilde{U}_{k_{\mathsf{max}}}^\top \right\|_{2,\infty} \left\| E V^{\star(1)} \right\|$$

$$\leq \sqrt{\frac{\mu r^3}{m_1}} \left(r^2 \sqrt{m_1} \omega_{\mathsf{max}} \log m + r (m_1 m_2)^{1/4} \omega_{\mathsf{max}} \log m \right)$$

$$+ \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{r^2 \sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma_{r_{k_{\mathsf{max}}}}^\star} + \frac{r^2 \sqrt{m_1 m_2} \omega_{\mathsf{max}}^2 \log^2 m}{\sigma_{r_{k_{\mathsf{max}}}}^{\star 2}} \right) \cdot \sqrt{m_1} \omega_{\mathsf{max}} \log m$$

$$\leq \sqrt{\frac{\mu r^3}{m_1}} \left(r^2 \sqrt{m_1} \omega_{\mathsf{max}} \log m + r (m_1 m_2)^{1/4} \omega_{\mathsf{max}} \log m \right) + \sqrt{\frac{\mu r^3}{m_1}} \sqrt{m_1} \omega_{\mathsf{max}} \log m$$

$$\asymp \sqrt{\frac{\mu r^3}{m_1}} \left(r^2 \sqrt{m_1} \omega_{\mathsf{max}} \log m + r (m_1 m_2)^{1/4} \omega_{\mathsf{max}} \log m \right), \tag{215}$$

where the penultimate line follows since $\sigma_{\overline{r}}^\star \geq C_0 r [(m_1 m_2)^{1/4} + r m_1^{1/2}] \omega_{\mathsf{max}} \log m$.

Step 5.2: bounding α_2. In view of (121), we have

$$\mathcal{P}_{(\widetilde{U}_{k_{\mathsf{max}}})_\perp} \mathcal{P}_{U_{k_{\mathsf{max}}}^\star} U^{\star(1)} \Sigma^{\star(1)} V^{\star(1)\top}$$

$$= \left(\mathcal{P}_{(\widetilde{U}_{k_{\mathsf{max}}})_\perp} U_{k_{\mathsf{max}}}^\star \right) U_{k_{\mathsf{max}}}^{\star\top} U^{\star(1)} \Sigma^{\star(1)} V^{\star(1)\top}$$

$$= \left[\widetilde{U}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}^{(1)} \widetilde{\Sigma}_{r_{k_{\mathsf{max}}}+1:\overline{r},r_{k_{\mathsf{max}}}+1:\overline{r}}^{(1)} \widetilde{W}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}^{(1)\top} \left(I_{r_{k_{\mathsf{max}}}} \; \mathbf{0} \right)^\top \left(\Sigma_{1:r_{k_{\mathsf{max}}},1:r_{k_{\mathsf{max}}}}^\star \right)^{-1} \right.$$

$$\left. - \mathcal{P}_{(\widetilde{U}_{:,1:r_{k_{\mathsf{max}}}})_\perp} E V^{\star(1)} \left(I_{r_{k_{\mathsf{max}}}} \; \mathbf{0} \right)^\top \left(\Sigma_{1:r_{k_{\mathsf{max}}},1:r_{k_{\mathsf{max}}}}^\star \right)^{-1} \right]$$

$$\cdot \Sigma_{1:r_{k_{\mathsf{max}}},1:r_{k_{\mathsf{max}}}}^\star V_{:,1:r_{k_{\mathsf{max}}}}^{\star\top}$$

$$= \widetilde{U}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}^{(1)} \widetilde{\Sigma}_{r_{k_{\mathsf{max}}}+1:\overline{r},r_{k_{\mathsf{max}}}+1:\overline{r}}^{(1)} \widetilde{W}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}^{(1)\top} \left(I_{r_{k_{\mathsf{max}}}} \; \mathbf{0} \right)^\top V_{:,1:r_{k_{\mathsf{max}}}}^{\star\top}$$

$$- \mathcal{P}_{(\widetilde{U}_{:,1:r_{k_{\mathsf{max}}}})_\perp} E V^{\star(1)} \left(I_{r_{k_{\mathsf{max}}}} \; \mathbf{0} \right)^\top V_{:,1:r_{k_{\mathsf{max}}}}^{\star\top}. \tag{216}$$

Here, the third line holds since $r_{k_{\mathsf{max}}} \leq \overline{r}$ and

$$U_{k_{\mathsf{max}}}^{\star\top} U^{\star(1)} \Sigma^{\star(1)} V^{\star(1)\top} = U_{:,1:r_{k_{\mathsf{max}}}}^{\star\top} U_{:,1:\overline{r}}^\star \Sigma_{1:\overline{r},1:\overline{r}}^\star V_{:,1:\overline{r}}^{\star\top}$$

$$= U_{:,1:r_{k_{\mathsf{max}}}}^{\star\top} \left(U_{:,1:r_{k_{\mathsf{max}}}}^\star \Sigma_{1:r_{k_{\mathsf{max}}},1:r_{k_{\mathsf{max}}}}^\star V_{:,1:r_{k_{\mathsf{max}}}}^{\star\top} + U_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}^\star \Sigma_{r_{k_{\mathsf{max}}}+1:\overline{r},r_{k_{\mathsf{max}}}+1:\overline{r}}^\star V_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}^{\star\top} \right)$$

$$= \mathbf{\Sigma}_{1:r_{k_{\max}}, 1:r_{k_{\max}}}^{\star} \mathbf{V}_{:,1:r_{k_{\max}}}^{\star\top}. \tag{217}$$

Repeating similar arguments as in (124) and (125), one has

$$\left\| \widetilde{\mathbf{U}}_{:,r_{k_{\max}}+1:\overline{r}}^{(1)} \widetilde{\mathbf{\Sigma}}_{r_{k_{\max}}+1:\overline{r},r_{k_{\max}}+1:\overline{r}}^{(1)} \widetilde{\mathbf{W}}_{:,r_{k_{\max}}+1:\overline{r}}^{(1)\top} \left(\mathbf{I}_{r_{k_{\max}}} \quad \mathbf{0} \right)^{\top} \mathbf{V}_{:,1:r_{k_{\max}}}^{\star\top} \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r}{m_1}} r \sqrt{m_1} \omega_{\max} \log m, \tag{218a}$$

$$\left\| \mathcal{P}_{\left(\widetilde{\mathbf{U}}_{:,1:r_{k_{\max}}}\right)_{\perp}} \mathbf{E} \mathbf{V}^{\star(1)} \left(\mathbf{I}_{r_{k_{\max}}} \quad \mathbf{0} \right)^{\top} \mathbf{V}_{:,1:r_{k_{\max}}}^{\star\top} \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r}{m_1}} \sqrt{m_1} \omega_{\max} \log m. \tag{218b}$$

Combining (216), (218a), (218b) and the triangle inequality yields

$$\left\| \mathcal{P}_{\left(\widetilde{\mathbf{U}}_{k_{\max}}\right)_{\perp}} \mathcal{P}_{\mathbf{U}_{k_{\max}}^{\star}} \mathbf{U}^{\star(1)} \mathbf{\Sigma}^{\star(1)} \mathbf{V}^{\star(1)\top} \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \sqrt{m_1} \omega_{\max} \log m. \tag{219}$$

Then we can bound α_2 as follows:

$$\begin{aligned}
\alpha_2 &= \left\| \left(\widetilde{\mathbf{U}}_{k_{\max}} \widetilde{\mathbf{U}}_{k_{\max}}^{\top} - \mathbf{U}_{k_{\max}}^{\star} \mathbf{U}_{k_{\max}}^{\star\top} \right) \mathbf{U}^{\star(1)} \mathbf{\Sigma}^{\star(1)} \mathbf{V}^{\star(1)\top} \right\|_{2,\infty} \\
&\leq \left\| \left(\widetilde{\mathbf{U}}_{k_{\max}} \widetilde{\mathbf{U}}_{k_{\max}}^{\top} - \mathbf{U}_{k_{\max}}^{\star} \mathbf{U}_{k_{\max}}^{\star\top} \right) \mathbf{U}_{k_{\max}}^{\star} \mathbf{U}_{k_{\max}}^{\star\top} \mathbf{U}^{\star(1)} \mathbf{\Sigma}^{\star(1)} \mathbf{V}^{\star(1)\top} \right\|_{2,\infty} \\
&\quad + \left\| \widetilde{\mathbf{U}}_{k_{\max}} \widetilde{\mathbf{U}}_{k_{\max}}^{\top} \left(\mathbf{U}_{k_{\max}}^{\star} \right)_{\perp} \left(\mathbf{U}_{k_{\max}}^{\star} \right)_{\perp}^{\star\top} \mathbf{U}^{\star(1)} \mathbf{\Sigma}^{\star(1)} \mathbf{V}^{\star(1)\top} \right\|_{2,\infty} \\
&= \left\| \mathcal{P}_{\left(\widetilde{\mathbf{U}}_{k_{\max}}\right)_{\perp}} \mathcal{P}_{\mathbf{U}_{k_{\max}}^{\star}} \mathbf{U}^{\star(1)} \mathbf{\Sigma}^{\star(1)} \mathbf{V}^{\star(1)\top} \right\|_{2,\infty} \\
&\quad + \left\| \widetilde{\mathbf{U}}_{k_{\max}} \widetilde{\mathbf{U}}_{k_{\max}}^{\top} \left(\mathbf{U}_{k_{\max}}^{\star} \right)_{\perp} \left(\mathbf{U}_{k_{\max}}^{\star} \right)_{\perp}^{\star\top} \mathbf{U}_{:,r_{k_{\max}}+1:\overline{r}}^{\star} \mathbf{\Sigma}_{r_{k_{\max}}+1:\overline{r},r_{k_{\max}}+1:\overline{r}}^{\star} \mathbf{V}_{:,r_{k_{\max}}+1:\overline{r}}^{\star\top} \right\|_{2,\infty} \\
&\overset{(219)}{\lesssim} \sqrt{\frac{\mu r^3}{m_1}} \sqrt{m_1} \omega_{\max} \log m + \left\| \widetilde{\mathbf{U}}_{k_{\max}} \right\|_{2,\infty} \left\| \widetilde{\mathbf{U}}_{k_{\max}}^{\top} \left(\mathbf{U}_{k_{\max}}^{\star} \right)_{\perp} \right\| \left\| \mathbf{\Sigma}_{r_{k_{\max}}+1:\overline{r},r_{k_{\max}}+1:\overline{r}}^{\star} \right\| \\
&\overset{\text{(90d) and (122b)}}{\lesssim} \sqrt{\frac{\mu r^3}{m_1}} \sqrt{m_1} \omega_{\max} \log m + \sqrt{\frac{\mu r}{m_1}} \frac{r \sqrt{m_1} \omega_{\max} \log m}{\sigma_{r_{k_{\max}}}^{\star}} \sigma_{r_{k_{\max}}+1}^{\star} \\
&\asymp \sqrt{\frac{\mu r^3}{m_1}} \sqrt{m_1} \omega_{\max} \log m. \tag{220}
\end{aligned}$$

Step 5.3: bounding α_3. By virtue of (78) and (206), one has

$$\alpha_3 = \left\| \mathbf{U}_{k_{\max}} \mathbf{U}_{k_{\max}}^{\top} - \mathbf{U}_{k_{\max}}^{\star} \mathbf{U}_{k_{\max}}^{\star\top} \right\|_{2,\infty} \sigma_{\overline{r}+1}^{\star} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(r \left[(m_1 m_2)^{1/4} + r m_1^{1/2} \right] \omega_{\max} \log m \right). \tag{221}$$

Step 5.4: bounding $\|(\mathbf{U}_{k_{\max}} \mathbf{U}_{k_{\max}}^{\top} - \mathbf{U}_{k_{\max}}^{\star} \mathbf{U}_{k_{\max}}^{\star\top}) \mathbf{X}^{\star}\|_{2,\infty}$ and $\|(\mathbf{I}_{n_1} - \mathbf{U}_{k_{\max}} \mathbf{U}_{k_{\max}}^{\top}) \mathbf{X}^{\star}\|_{2,\infty}$. Putting (207), (215), (220) and (221) together, we arrive at

$$\left\| \left(\mathbf{U}_{k_{\max}} \mathbf{U}_{k_{\max}}^{\top} - \mathbf{U}_{k_{\max}}^{\star} \mathbf{U}_{k_{\max}}^{\star\top} \right) \mathbf{X}^{\star} \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(r^2 \sqrt{m_1} \omega_{\max} \log m + r (m_1 m_2)^{1/4} \omega_{\max} \log m \right). \tag{222}$$

Furthermore, we have

$$\begin{aligned}
\left\| \left(\mathbf{I}_{m_1} - \mathbf{U}_{k_{\max}}^{\star} \mathbf{U}_{k_{\max}}^{\star\top} \right) \mathbf{X}^{\star} \right\|_{2,\infty} &= \left\| \mathbf{U}^{\star} \mathbf{\Sigma}^{\star} \mathbf{V}^{\star\top} - \mathbf{U}_{k_{\max}}^{\star} \mathbf{U}_{k_{\max}}^{\star\top} \mathbf{U}^{\star} \mathbf{\Sigma}^{\star} \mathbf{V}^{\star\top} \right\|_{2,\infty} \\
&= \left\| \mathbf{U}^{\star} \mathbf{\Sigma}^{\star} \mathbf{V}^{\star\top} - \mathbf{U}_{:,1:r_{k_{\max}}}^{\star} \mathbf{\Sigma}_{1:r_{k_{\max}},1:r_{k_{\max}}}^{\star} \mathbf{V}_{:,1:r_{k_{\max}}}^{\star\top} \right\|_{2,\infty} \\
&= \left\| \mathbf{U}_{:,r_{k_{\max}}+1:r}^{\star} \mathbf{\Sigma}_{r_{k_{\max}}+1:r,r_{k_{\max}}+1:r}^{\star} \mathbf{V}_{:,r_{k_{\max}}+1:r}^{\star\top} \right\|_{2,\infty}
\end{aligned}$$

$$\leq \|\boldsymbol{U}^\star\|_{2,\infty} \left\| \boldsymbol{\Sigma}^\star_{r_{k_{\max}}+1:r, r_{k_{\max}}+1:r} \right\|$$

$$\leq \sqrt{\frac{\mu r}{m_1}} \sigma^\star_{r_{k_{\max}}+1}$$

$$\overset{\text{(90a) and (103)}}{\leq} \sqrt{\frac{\mu r}{m_1}} \left(\widetilde{\sigma}_{r_{k_{\max}}+1} + 2\sqrt{C_5}\sqrt{m_1}\omega_{\max}\log m \right)$$

$$\overset{\text{(204)}}{\lesssim} \sqrt{\frac{\mu r}{m_1}} \left(\sqrt{\tau} + \sqrt{m_1}\omega_{\max}\log m \right)$$

$$\asymp \sqrt{\frac{\mu r}{m_1}} \left(r^2\sqrt{m_1}\omega_{\max}\log m + r(m_1 m_2)^{1/4}\omega_{\max}\log m \right).$$

This together with (222) gives

$$\|(\boldsymbol{I}_{n_1} - \boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top)\boldsymbol{X}^\star\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(r^2\sqrt{m_1}\omega_{\max}\log m + r(m_1 m_2)^{1/4}\omega_{\max}\log m \right). \tag{223}$$

E Proof of Theorem 2

For notational convenience, we let $\boldsymbol{U}_i^\star = \boldsymbol{U}_{\boldsymbol{X}_i^\star} \in \mathcal{O}^{n_i, k_i}$ denote the left singular subspace of $\boldsymbol{X}_i^\star$ for all $i \in [3]$. Then we know from (45) that

$$\|\boldsymbol{U}_i^\star\|_{2,\infty} \leq \sqrt{\frac{k_i}{\beta n_i}}, \qquad \forall i \in [3]. \tag{224}$$

In view of Zhou and Chen (2023, Lemma 7), with probability exceeding $1 - O(n^{-10})$,

$$\left\| \mathcal{P}_{\text{off-diag}}\left(\boldsymbol{E}_1\boldsymbol{E}_1^\top \right) \right\| \lesssim B^2\log^2 n + \sqrt{n_1 n_2 n_3}\omega_{\max}^2\log n \leq \sqrt{n_1 n_2 n_3}\omega_{\max}^2\log n \ll n_2 n_3\omega_{\max}^2. \tag{225}$$

For any $i \in [n_1]$, we know that

$$\left(\boldsymbol{E}_1\boldsymbol{E}_1^\top \right)_{i,i} = \sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3} E_{i,j,\ell}^2 = \sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3} \omega_{i,j,\ell}^2 + \sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3} \left(E_{i,j,\ell}^2 - \omega_{i,j,\ell}^2 \right). \tag{226}$$

If the noise is bounded, i.e., $E_{i,j,k} \leq B$ for all $(i,j,k) \in [n_1] \times [n_2] \times [n_3]$, then $\{E_{i,j,\ell}^2 - \omega_{i,j,\ell}^2\}_{i,j,\ell}$ are zero-mean, and

$$\left| E_{i,j,\ell}^2 - \omega_{i,j,\ell}^2 \right| \leq 2B^2,$$
$$\mathbb{E}\left[\left(E_{i,j,\ell}^2 - \omega_{i,j,\ell}^2 \right)^2 \right] = \mathbb{E}\left[E_{i,j,\ell}^4 \right] - \omega_{i,j,\ell}^4 \leq B^2\mathbb{E}\left[E_{i,j,\ell}^2 \right] - \omega_{i,j,\ell}^4 \leq B^2\omega_{\max}^2.$$

In view of Bernstein's inequality and the union bound, one has, with probability exceeding $1 - O(n^{-10})$, for all $i \in [n_1]$,

$$\left| \sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3} \left(E_{i,j,\ell}^2 - \omega_{i,j,\ell}^2 \right) \right| \lesssim \sqrt{n_2 n_3}B\omega_{\max}\sqrt{\log n} + B^2\log n \ll n_2 n_3\omega_{\max}^2. \tag{228}$$

For the general case where the noise satisfies Assumption 1, using the truncation trick as in Zhou and Chen (2023, Section B.4.2), one can show that (228) also holds with probability exceeding $1 - O(n^{-10})$. Putting (225), (226) and (228) together, we know that with probability exceeding $1 - O(n^{-10})$,

$$\left\| \boldsymbol{E}_1\boldsymbol{E}_1^\top - \text{diag}\left(\left[\sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3} \omega_{i,j,\ell}^2 \right]_{1 \leq i \leq n_1} \right) \right\| \ll n_2 n_3\omega_{\max}^2. \tag{229}$$

For all $i \in [3]$, let $\boldsymbol{U}_i^\star \in \mathcal{O}^{n_i, k_i}$ denote the left singular subspace of $\boldsymbol{X}_i^\star$. The min-max principle for singular values reveals that

$$
\begin{aligned}
\sigma_{k_1+1}\left(\boldsymbol{Y}_1\right) &\geq \sigma_{k_1+1}\left(\boldsymbol{Y}_1\left(\mathcal{P}_{\boldsymbol{U}_{3\perp}^\star} \otimes \mathcal{P}_{\boldsymbol{U}_{2\perp}^\star}\right)\right) \\
&= \sigma_{k_1+1}\left(\boldsymbol{X}_1^\star\left(\mathcal{P}_{\boldsymbol{U}_{3\perp}^\star} \otimes \mathcal{P}_{\boldsymbol{U}_{2\perp}^\star}\right) + \boldsymbol{E}_1\left(\mathcal{P}_{\boldsymbol{U}_{3\perp}^\star} \otimes \mathcal{P}_{\boldsymbol{U}_{2\perp}^\star}\right)\right) \\
&= \sigma_{k_1+1}\left(\boldsymbol{E}_1\left(\mathcal{P}_{\boldsymbol{U}_{3\perp}^\star} \otimes \mathcal{P}_{\boldsymbol{U}_{2\perp}^\star}\right)\right) \\
&\geq \sigma_{k_1+1}\left(\boldsymbol{E}_1\right) - \left\|\boldsymbol{E}_1\left(\mathcal{P}_{\boldsymbol{U}_3^\star} \otimes \mathcal{P}_{\boldsymbol{U}_2^\star}\right)\right\| - \left\|\boldsymbol{E}_1\left(\mathcal{P}_{\boldsymbol{U}_{3\perp}^\star} \otimes \mathcal{P}_{\boldsymbol{U}_2^\star}\right)\right\| - \left\|\boldsymbol{E}_1\left(\mathcal{P}_{\boldsymbol{U}_3^\star} \otimes \mathcal{P}_{\boldsymbol{U}_{2\perp}^\star}\right)\right\| \\
&= \sqrt{\sigma_{k_1+1}\left(\boldsymbol{E}_1 \boldsymbol{E}_1^\top\right)} - \left\|\boldsymbol{E}_1\left(\boldsymbol{U}_3^\star \otimes \boldsymbol{U}_2^\star\right)\right\| - \left\|\boldsymbol{E}_1\left(\boldsymbol{U}_{3\perp}^\star \otimes \boldsymbol{U}_2^\star\right)\right\| - \left\|\boldsymbol{E}_1\left(\boldsymbol{U}_3^\star \otimes \boldsymbol{U}_{2\perp}^\star\right)\right\|,
\end{aligned}
\tag{230}
$$

where the fourth line makes use of Weyl's inequality. We know from (224) that

$$
\left\|\boldsymbol{U}_3^\star \otimes \boldsymbol{U}_2^\star\right\|_{2,\infty} \leq \sqrt{\frac{k_2 k_3}{\beta^2 n_2 n_3}},
$$

$$
\left\|\boldsymbol{U}_{3\perp}^\star \otimes \boldsymbol{U}_2^\star\right\|_{2,\infty} \leq \left\|\boldsymbol{U}_2^\star\right\|_{2,\infty} \leq \sqrt{\frac{k_2}{\beta n_2}},
$$

$$
\left\|\boldsymbol{U}_3^\star \otimes \boldsymbol{U}_{2\perp}^\star\right\|_{2,\infty} \leq \left\|\boldsymbol{U}_3^\star\right\|_{2,\infty} \leq \sqrt{\frac{k_3}{\beta n_3}}.
$$

Applying Zhou and Chen (2023, Lemma 5) yields that with probability at least $1 - O(n^{-10})$,

$$
\left\|\boldsymbol{E}_1\left(\boldsymbol{U}_3^\star \otimes \boldsymbol{U}_2^\star\right)\right\| \lesssim B\sqrt{\frac{k_2 k_3}{\beta^2 n_2 n_3}} \log n + \sqrt{n_1}\omega_{\mathsf{max}}\sqrt{\log n} \ll \sqrt{n_2 n_3}\omega_{\mathsf{max}},
\tag{231a}
$$

$$
\left\|\boldsymbol{E}_1\left(\boldsymbol{U}_{3\perp}^\star \otimes \boldsymbol{U}_2^\star\right)\right\| \lesssim B\sqrt{\frac{k_2}{\beta n_2}} \log n + \sqrt{n_1 + k_2 n_3}\omega_{\mathsf{max}}\sqrt{\log n} \ll \sqrt{n_2 n_3}\omega_{\mathsf{max}},
\tag{231b}
$$

$$
\left\|\boldsymbol{E}_1\left(\boldsymbol{U}_3^\star \otimes \boldsymbol{U}_{2\perp}^\star\right)\right\| \lesssim B\sqrt{\frac{k_3}{\beta n_3}} \log n + \sqrt{n_1 + k_3 n_2}\omega_{\mathsf{max}}\sqrt{\log n} \ll \sqrt{n_2 n_3}\omega_{\mathsf{max}}.
\tag{231c}
$$

Combining (229), (231a) - (231c) and Weyl's inequality, we obtain that with probability exceeding $1 - O(n^{-10})$,

$$
\sigma_{k_1+1}\left(\boldsymbol{Y}_1\right) \leq \sigma_{k_1+1}\left(\boldsymbol{X}_1^\star\right) + \left\|\boldsymbol{E}_1\right\| = \left\|\boldsymbol{E}_1\right\| = \left(\left\|\boldsymbol{E}_1 \boldsymbol{E}_1^\top\right\|\right)^{1/2} \leq 2\sqrt{n_2 n_3}\omega_{\mathsf{max}}.
\tag{232}
$$

1. If (24) holds, then (229), (230) and (231a) - (231c) together show that with probability exceeding $1 - O(n^{-10})$,

$$
\sigma_{k_1+1}\left(\boldsymbol{Y}_1\right) \geq \sqrt{c n_2 n_3 \omega_{\mathsf{max}}^2} - \frac{\sqrt{c}}{2}\sqrt{n_2 n_3}\omega_{\mathsf{max}} \geq \frac{\sqrt{c}}{2}\sqrt{n_2 n_3}\omega_{\mathsf{max}}.
\tag{233}
$$

The previous inequality together with (232) shows that

$$
\sigma_{k_1+1}\left(\boldsymbol{Y}_1\right) \asymp \sqrt{n_2 n_3}\omega_{\mathsf{max}}
$$

with probability at least $1 - O(n^{-10})$. As a consequence, there exist two large enough constants $C_\tau > c_\tau > 0$ such that with probability at least $1 - O(n^{-10})$,

$$
c_\tau \left(n_1 n_2 n_3\right)^{1/2} \log^2 n \leq \tau / \omega_{\mathsf{max}}^2 \leq C_\tau \left(n_1 n_2 n_3\right)^{1/2} \log^2 n.
\tag{234}
$$

2. If Assumption 2 in Theorem 2 holds, we choose $(\ell_1, \ell_2, \ell_3) \in [k_1] \times [k_2] \times [k_3]$ such that

$$
(\ell_1, \ell_2, \ell_3) \in \underset{i_1 \in [k_1], i_2 \in [k_2], i_3 \in [k_3]}{\arg\max} S_{i_1, i_2, i_3}^\star \left(1 - S_{i_1, i_2, i_3}^\star\right) = \omega_{\mathsf{max}}^2.
$$

Then for any $(j_1, j_2, j_3) \in [n_1] \times [n_2] \times [n_3]$ with $z^\star_{i,j_i} = \ell_i$, $i \in [3]$,

$$\mathbb{E}\left[E^2_{j_1,j_2,j_3}\right] = \omega^2_{\mathsf{max}}.$$

For any $i \in \{j \in [n_1] : z^\star_{1,j} = \ell_1\}$,

$$\sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3} \omega^2_{i,j,\ell} \geq \omega^2_{\mathsf{max}} \left|j \in [n_2] : z^\star_{2,j} = \ell_2\right| \cdot \left|j \in [n_3] : z^\star_{3,j} = \ell_3\right| \geq \frac{\beta n_2}{k_2}\frac{\beta n_3}{k_3}\omega^2_{\mathsf{max}} \asymp n_2 n_3 \omega^2_{\mathsf{max}}.$$

Since $\left|\{j \in [n_1] : z^\star_{1,j} = \ell_1\}\right| \geq \beta n_1/k_1 > k_1 + 1$, we can still show that (233) holds with probability exceeding $1 - O(n^{-10})$. Repeating similar arguments as in (234) shows that Theorem 2 also holds.

F Technical lemmas

Lemma 6. *Suppose that*

$$\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{E} \in \mathbb{R}^{n_1 \times n_2} \tag{235}$$

and the SVDs of $\boldsymbol{M}^\star$ and $\boldsymbol{M}$ are given by

$$\boldsymbol{M}^\star = \sum_{i=1}^{n_1} \sigma_i^\star \boldsymbol{u}_i^\star \boldsymbol{v}_i^{\star\top}, \qquad and \qquad \boldsymbol{M}^\star = \sum_{i=1}^{n_1} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top.$$

Here, $\sigma_1^\star \geq \cdots \geq \sigma_{n_1}^\star \geq 0$ (resp. $\sigma_1 \geq \cdots \geq \sigma_{n_1} \geq 0$) represent the singular values of $\boldsymbol{M}^\star$ (resp. $\boldsymbol{M}$), $u_i^\star$ (resp. u_i) denotes the left singular vector associated with the singular value $\sigma_i^\star$ (resp. σ_i), and $v_i^\star$ (resp. v_i) denotes the left singular vector associated with $\sigma_i^\star$ (resp. σ_i). We let $\boldsymbol{U}^\star = [u_1^\star, \ldots, u_r^\star] \in \mathbb{R}^{n_1 \times r}$ (resp. $\boldsymbol{U} = [u_1, \ldots, u_r] \in \mathbb{R}^{n_1 \times r}$) denote the rank-r leading singular subspace of $\boldsymbol{M}^\star$ (resp. $\boldsymbol{M}$). If

$$\sigma_r^\star - \sigma_{r+1}^\star > 2\|\boldsymbol{E}\|,$$

then we have

$$\left\|\left(\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right)\boldsymbol{M}^\star\right\| \leq \frac{4\sigma_r^\star\|\boldsymbol{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star}.$$

Proof. We define

$$\boldsymbol{\Sigma} := \mathsf{diag}\left(\sigma_1, \ldots, \sigma_r\right), \qquad \boldsymbol{\Sigma}_\perp := \mathsf{diag}\left(\sigma_{r+1}, \ldots, \sigma_{n_1}\right),$$
$$\boldsymbol{V} := [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r] \in \mathbb{R}^{n_2 \times r}, \qquad \boldsymbol{V}_\perp := [\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_{n_2}] \in \mathbb{R}^{n_2 \times (n_2-r)},$$

and define $\boldsymbol{\Sigma}^\star, \boldsymbol{\Sigma}_\perp^\star, \boldsymbol{V}^\star, \boldsymbol{V}_\perp^\star$ similarly. In view of Chen et al. (2021a, Eqn. (2.27)), we have

$$\begin{aligned}
\mathcal{P}_{\boldsymbol{U}_\perp}\mathcal{P}_{\boldsymbol{U}^\star}\boldsymbol{M}^\star &= \boldsymbol{U}_\perp\left(\boldsymbol{U}_\perp\boldsymbol{U}^\star\right)\left(\boldsymbol{U}^{\star\top}\boldsymbol{M}^\star\right) \\
&= \boldsymbol{U}_\perp\left(\boldsymbol{\Sigma}_\perp\boldsymbol{V}_\perp^\top\boldsymbol{V}^\star\boldsymbol{\Sigma}^{\star-1} - \boldsymbol{U}_\perp^\top\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{\Sigma}^{\star-1}\right)\boldsymbol{\Sigma}^\star\boldsymbol{V}^{\star\top} \\
&= \boldsymbol{U}_\perp\boldsymbol{\Sigma}_\perp\boldsymbol{V}_\perp^\top\boldsymbol{V}^\star\boldsymbol{V}^{\star\top} - \boldsymbol{U}_\perp^\top\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}.
\end{aligned}$$

Recognizing that

$$\begin{aligned}
\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top} &= \left(\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\star - \boldsymbol{U}^\star\right)\boldsymbol{U}^{\star\top} + \left(\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right) \\
&= -\mathcal{P}_{\boldsymbol{U}_\perp}\mathcal{P}_{\boldsymbol{U}^\star} + \boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}_\perp^\star\boldsymbol{U}_\perp^{\star\top},
\end{aligned}$$

we have

$$\begin{aligned}
\left\|\left(\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\right)\boldsymbol{M}^\star\right\| &\leq \left\|\mathcal{P}_{\boldsymbol{U}_\perp}\mathcal{P}_{\boldsymbol{U}^\star}\boldsymbol{M}^\star\right\| + \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}_\perp^\star\boldsymbol{U}_\perp^{\star\top}\boldsymbol{M}^\star\right\| \\
&\leq \left\|\boldsymbol{U}_\perp\boldsymbol{\Sigma}_\perp\boldsymbol{V}_\perp^\top\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\right\| + \left\|\boldsymbol{U}_\perp^\top\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\right\| + \left\|\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{U}_\perp^\star\boldsymbol{\Sigma}_\perp^\star\boldsymbol{V}_\perp^{\star\top}\right\|
\end{aligned}$$

$$
\begin{aligned}
&\leq \|\mathbf{\Sigma}_\perp\| \left\|\mathbf{V}_\perp^\top \mathbf{V}^\star\right\| + \|\mathbf{E}\| + \left\|\mathbf{U}_\perp^{\star\top}\mathbf{U}\right\| \|\mathbf{\Sigma}_\perp^\star\| \\
&\overset{(a)}{\leq} \sigma_{r+1} \frac{2\|\mathbf{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star} + \|\mathbf{E}\| + \sigma_{r+1}^\star \frac{2\|\mathbf{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star} \\
&\overset{(b)}{\leq} \left(\sigma_{r+1}^\star + \|\mathbf{E}\|\right) \frac{2\|\mathbf{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star} + \|\mathbf{E}\| + \sigma_{r+1}^\star \frac{2\|\mathbf{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star} \\
&\overset{(c)}{\leq} \frac{4\sigma_r^\star \|\mathbf{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star}.
\end{aligned}
\tag{236}
$$

Here, (a) comes from (Chen et al., 2021a, Eqn. (2.26a)), (b) makes use of Weyl's inequality, and (c) holds due to the assumption $\sigma_r^\star - \sigma_{r+1}^\star > 2\|\mathbf{E}\|$. $\qquad\square$

Lemma 7. *Suppose that Assumption 1 holds. We let $\mathcal{A}$ denote the following set:*

$$
\mathcal{U} = \left\{ (\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) : \mathbf{U}_i \in \mathbb{R}^{n_i \times r_i}, \|\mathbf{U}_i\| \leq 1, \|\mathbf{U}_i\|_{2,\infty} \leq \sqrt{\frac{\mu_i r_i}{n_i}}, i \in [3] \right\}
\tag{237}
$$

and we define

$$
n = \max_{1 \leq i \leq 3} n_i, \qquad \text{and} \qquad r = \max_{1 \leq i \leq 3} r_i.
$$

If $n_1 n_2 n_3 \geq r^4 n^2$, then with probability exceeding $1 - O(n^{-10})$, the following inequality holds:

$$
\sup_{(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) \in \mathcal{U}} \left\|\mathbf{U}_1^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\mathbf{U}_3 \otimes \mathbf{U}_2)\right\| \lesssim \omega_{\mathsf{max}} \sqrt{n \mu_1 \mu_2 \mu_3 r^3 \log n},
\tag{238}
$$

If, furthermore, the $E_{i,j,k}$'s are ω_{max}-sub-Gaussian, then with probability exceeding $1 - e^{-Cn}$, one has

$$
\sup_{(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) \in \mathcal{U}} \left\|\mathbf{U}_1^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\mathbf{U}_3 \otimes \mathbf{U}_2)\right\| \leq \sqrt{nr} \omega_{\mathsf{max}}.
\tag{239}
$$

Proof of Lemma 7. (239) can be proved by simply combining Zhou et al. (2022, Lemma A.2) (or Lemma 8.2 presented in its arxiv version) and the standard epsilon-net technique used in the proof of Zhang and Xia (2018, Lemma 5) and we omit the details here for the sake of brevity.

Proving (238): the bounded noise case. To prove (238), we first consider the bounded noise case, i.e., $|E_{i,j,k}| \leq B$ holds for all $(i,j,k) \in [n_1] \times [n_2] \times [n_3]$. For any fixed $(\overline{\mathbf{U}}_1, \overline{\mathbf{U}}_2, \overline{\mathbf{U}}_3) \in \mathcal{U}$, note that

$$
\overline{\mathbf{U}}_1^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\overline{\mathbf{U}}_3 \otimes \overline{\mathbf{U}}_2) = \sum_{i \in [n_1], j \in [n_2 n_3]} (\mathcal{M}_1(\boldsymbol{\mathcal{E}}))_{i,j} (\overline{\mathbf{U}}_1)_{i,:}^\top (\overline{\mathbf{U}}_3 \otimes \overline{\mathbf{U}}_2)_{j,:}
$$

is a sum of independent zero-mean matrices. In addition, we have

$$
\begin{aligned}
L &:= \max_{i \in [n_1], j \in [n_2 n_3]} \left\|(\mathcal{M}_1(\boldsymbol{\mathcal{E}}))_{i,j} (\overline{\mathbf{U}}_1)_{i,:}^\top (\overline{\mathbf{U}}_3 \otimes \overline{\mathbf{U}}_2)_{j,:}\right\| \leq B \prod_{i=1}^3 \|\overline{\mathbf{U}}_i\|_{2,\infty} \leq B \sqrt{\frac{\mu_1 \mu_2 \mu_3 r_1 r_2 r_3}{n_1 n_2 n_3}}, \\
V &:= \max \left\{ \left\| \sum_{i \in [n_1], j \in [n_2 n_3]} \mathbb{E}\left[(\mathcal{M}_1(\boldsymbol{\mathcal{E}}))_{i,j}^2\right] \left\|(\overline{\mathbf{U}}_3 \otimes \overline{\mathbf{U}}_2)_{j,:}\right\|_2^2 (\overline{\mathbf{U}}_1)_{i,:}^\top (\overline{\mathbf{U}}_1)_{i,:} \right\|, \right. \\
&\qquad\qquad \left. \left\| \sum_{i \in [n_1], j \in [n_2 n_3]} \mathbb{E}\left[(\mathcal{M}_1(\boldsymbol{\mathcal{E}}))_{i,j}^2\right] \left\|(\overline{\mathbf{U}}_1)_{i,:}\right\|_2^2 (\overline{\mathbf{U}}_3 \otimes \overline{\mathbf{U}}_2)_{j,:}^\top (\overline{\mathbf{U}}_3 \otimes \overline{\mathbf{U}}_2)_{j,:} \right\| \right\} \\
&\leq \max \left\{ \omega_{\mathsf{max}}^2 \left\|(\overline{\mathbf{U}}_3 \otimes \overline{\mathbf{U}}_2)\right\|_{\mathrm{F}}^2 \left\|\overline{\mathbf{U}}_1 \overline{\mathbf{U}}_1^\top\right\|, \omega_{\mathsf{max}}^2 \|\overline{\mathbf{U}}_1\|_{\mathrm{F}}^2 \left\|(\overline{\mathbf{U}}_3 \otimes \overline{\mathbf{U}}_2)(\overline{\mathbf{U}}_3 \otimes \overline{\mathbf{U}}_2)^\top\right\| \right\} \\
&\lesssim \omega_{\mathsf{max}}^2 \max\{r_2 r_3, r_1\} \leq \omega_{\mathsf{max}}^2 r^2.
\end{aligned}
$$

In view of the matrix Bernstein inequality, with probability exceeding $1 - e^{-Cnr \log n}$,

$$
\begin{aligned}
\left\| \overline{\boldsymbol{U}}_1^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\overline{\boldsymbol{U}}_3 \otimes \overline{\boldsymbol{U}}_2\right) \right\| &\lesssim \sqrt{Vnr \log n} + Lnr \log n \\
&\lesssim \omega_{\max}\sqrt{nr^3 \log n} + \omega_{\max}\frac{(n_1 n_2 n_3)^{1/4}}{\log n}\sqrt{\frac{\mu_1 \mu_2 \mu_3 r_1 r_2 r_3}{n_1 n_2 n_3}}nr \log n \\
&\leq \omega_{\max}\sqrt{n\mu_1 \mu_2 \mu_3 r^3 \log n}.
\end{aligned}
\tag{240}
$$

Here, the last line makes use of $n_1 n_2 n_3 \geq r^4 n^2$.

Repeating a similar argument as in Vershynin (2010, Lemma 5.2) yields that: there exists a set $\mathcal{B}_i \subset \mathcal{D}_i = \{\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^{r_i}, \|\boldsymbol{x}\|_2 \leq \sqrt{\mu_i r_i / n_i}\}$ with cardinality at most $(1 + 8\sqrt{\mu_i r_i})^{r_i}$ such that for any $\boldsymbol{x} \in \mathcal{D}_i$, one can find $\boldsymbol{x}' \in \mathcal{B}_i$ such that

$$
\|\boldsymbol{x} - \boldsymbol{x}'\|_2 \leq \frac{1}{4}\sqrt{\frac{1}{n_i}}.
$$

$\|\boldsymbol{x} - \boldsymbol{x}'\|_2 \leq c\sqrt{1/n_i}$. As a direct consequence, for any $\boldsymbol{U}_i \in \mathcal{U}_i := \left\{ \boldsymbol{U} \in \mathbb{R}^{n_i \times r_i} : \|\boldsymbol{U}\| \leq 1, \|\boldsymbol{U}\|_{2,\infty} \leq \sqrt{\frac{\mu_i r_i}{n_i}} \right\}$, one can find $\boldsymbol{U}_i' \in \mathcal{F}_i := \{\boldsymbol{U} \in \mathbb{R}^{n_i \times r_i} : \boldsymbol{U}_{j,:}^\top \in \mathcal{B}_i, \forall j \in [n_i]\}$ such that

$$
\|\boldsymbol{U}_i - \boldsymbol{U}_i'\|_{2,\infty} \leq \frac{1}{4}\sqrt{\frac{1}{n_i}}.
$$

Let $\mathcal{U}_i'$ denote the following set:

$$
\mathcal{U}_i' := \left\{ \boldsymbol{U}' \in \mathbb{R}^{n_i \times r_i} : \boldsymbol{U}' \in \mathcal{F}_i, \inf_{\boldsymbol{U} \in \mathcal{U}_i} \|\boldsymbol{U} - \boldsymbol{U}'\|_{2,\infty} \leq \frac{1}{4}\sqrt{\frac{1}{n_i}} \right\}.
\tag{241}
$$

Then we can verify the following three properties:

$$
|\mathcal{U}_i'| \leq |\mathcal{F}_i| = |\mathcal{B}_i|^{n_i} \leq (1 + 8\sqrt{\mu_i r_i})^{n_i r_i} \leq n_i^{n_i r_i} \leq e^{n_i r_i \log n},
\tag{242a}
$$

$$
\forall \boldsymbol{U}_i \in \mathcal{U}_i, \quad \exists \boldsymbol{U}_i' \in \mathcal{U}_i' \quad \text{s.t.} \quad \|\boldsymbol{U}_i - \boldsymbol{U}_i'\|_{2,\infty} \leq \frac{1}{4}\sqrt{\frac{1}{n_i}} \quad \text{and} \quad \|\boldsymbol{U}_i - \boldsymbol{U}_i'\| \leq \frac{1}{4},
\tag{242b}
$$

$$
\forall \boldsymbol{U}_i' \in \mathcal{U}_i', \quad \exists \boldsymbol{U}_i \in \mathcal{U}_i \quad \text{s.t.} \quad \|\boldsymbol{U}_i - \boldsymbol{U}_i'\|_{2,\infty} \leq \frac{1}{4}\sqrt{\frac{1}{n_i}} \quad \text{and} \quad \|\boldsymbol{U}_i - \boldsymbol{U}_i'\| \leq \frac{1}{4}.
\tag{242c}
$$

Here, the last two inequalities make use of $\|\boldsymbol{U}_i - \boldsymbol{U}_i'\| \leq \sqrt{n_i}\|\boldsymbol{U}_i - \boldsymbol{U}_i'\|_{2,\infty}$.

We define

$$
A = \sup_{(\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3) \in \mathcal{U}} \left\| \boldsymbol{U}_1^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2\right) \right\| \quad \text{and} \quad B = \sup_{(\boldsymbol{U}_1', \boldsymbol{U}_2', \boldsymbol{U}_3') \in \mathcal{U}_1' \times \mathcal{U}_2' \times \mathcal{U}_3'} \left\| \boldsymbol{U}_1'^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\boldsymbol{U}_3' \otimes \boldsymbol{U}_2'\right) \right\|.
$$

For any $(\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3) \in \mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{U}_3$, we know from (242b) and (242c) that there exist $(\boldsymbol{U}_1', \boldsymbol{U}_2', \boldsymbol{U}_3') \in \mathcal{U}_1' \times \mathcal{U}_2' \times \mathcal{U}_3'$ such that

$$
\boldsymbol{U}_i - \boldsymbol{U}_i' \in \frac{1}{4}\mathcal{U}_i, \quad i \in [3],
\tag{243a}
$$

$$
\boldsymbol{U}_i' \in \frac{5}{4}\mathcal{U}_i, \quad i \in [3].
\tag{243b}
$$

The triangle inequality, (243a) and (243b) together show that

$$
\begin{aligned}
\left\| \boldsymbol{U}_1^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2\right) \right\| &\leq \left\| (\boldsymbol{U}_1 - \boldsymbol{U}_1')^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2\right) \right\| + \left\| \boldsymbol{U}_1'^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\boldsymbol{U}_3 \otimes (\boldsymbol{U}_2 - \boldsymbol{U}_2')\right) \right\| \\
&\quad + \left\| \boldsymbol{U}_1'^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left((\boldsymbol{U}_3 - \boldsymbol{U}_3') \otimes \boldsymbol{U}_2'\right) \right\| + \left\| \boldsymbol{U}_1'^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\boldsymbol{U}_3' \otimes \boldsymbol{U}_2'\right) \right\| \\
&\leq \frac{1}{4}A + \frac{1}{4}\cdot\frac{5}{4}A + \frac{1}{4}\cdot\frac{5}{4}\cdot\frac{5}{4}A + B
\end{aligned}
$$

66

$$\leq \frac{61}{64}A + B,$$

which implies

$$A \leq \frac{64}{3}B. \tag{244}$$

Furthermore, (240), (242a), (243b) and the union bound together imply that with probability exceeding $1 - e^{-Cnr \log n} \cdot \prod_i e^{n_i r_i \log n} \geq 1 - e^{-C'nr \log n}$,

$$B \lesssim \left(\frac{5}{4}\right)^3 \omega_{\mathsf{max}} \sqrt{n\mu_1\mu_2\mu_3 r^3 \log n} \asymp \omega_{\mathsf{max}} \sqrt{n\mu_1\mu_2\mu_3 r^3 \log n}. \tag{245}$$

Putting (244) and (245) together, we arrive at

$$\sup_{(\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3) \in \mathcal{U}} \left\| \boldsymbol{U}_1^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2) \right\| = A \lesssim \omega_{\mathsf{max}} \sqrt{n\mu_1\mu_2\mu_3 r^3 \log n}$$

with probability exceeding $1 - e^{-C'nr \log n}$.

Proving (238): the general case. For the general case where the noise matrix $\boldsymbol{E}$ satisfies Assumption 2, we can prove (238) by repeating a similar argument as in Zhou and Chen (2023, Section B.4.2). $\qquad \square$

References

Abbe, E. (2017). Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531.

Abbe, E., Bandeira, A. S., and Hall, G. (2015). Exact recovery in the stochastic block model. *IEEE Transactions on information theory*, 62(1):471–487.

Abbe, E., Fan, J., and Wang, K. (2022). An ℓ_p theory of PCA and spectral clustering. *The Annals of Statistics*, 50(4):2359–2385.

Abbe, E., Fan, J., Wang, K., and Zhong, Y. (2020). Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of statistics*, 48(3):1452.

Abbe, E. and Sandon, C. (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *Annual Symposium on Foundations of Computer Science*, pages 670–688. IEEE.

Agterberg, J. and Zhang, A. (2022). Estimating higher-order mixed memberships via the $\ell_{2,\infty}$ tensor perturbation bound. *arXiv preprint arXiv:2212.08642*.

Amini, A. A. and Levina, E. (2018). On semidefinite relaxations for the block model.

Anandkumar, A., Deng, Y., Ge, R., and Mobahi, H. (2017). Homotopy analysis for tensor pca. In *Conference on Learning Theory*, pages 79–104. PMLR.

Arous, G. B., Mei, S., Montanari, A., and Nica, M. (2019). The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330.

Arthur, D. and Vassilvitskii, S. (2007). K-means++: the advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035.

Bahmani, B., Moseley, B., Vattani, A., Kumar, R., and Vassilvitskii, S. (2012). Scalable k-means++. *Proceedings of the VLDB Endowment*, 5(7):622–633.

Bandeira, A. S., Boumal, N., and Singer, A. (2017). Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming*, 163:145–167.

Bi, X., Qu, A., and Shen, X. (2018). Multilayer tensor factorization with applications to recommender systems. *The Annals of Statistics*, 46(6B):3308–3333.

Bi, X., Tang, X., Yuan, Y., Zhang, Y., and Qu, A. (2021). Tensors in statistics. *Annual review of statistics and its application*, 8:345–368.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

Cai, C., Li, G., Chi, Y., Poor, H. V., and Chen, Y. (2021). Subspace estimation from unbalanced and incomplete data matrices: $\ell_{2,\infty}$ statistical guarantees. *The Annals of Statistics*, 49(2):944–967.

Cai, C., Li, G., Poor, H. V., and Chen, Y. (2022a). Nonconvex low-rank tensor completion from noisy data. *Operations Research*, 70(2):1219–1237.

Cai, C., Poor, H. V., and Chen, Y. (2022b). Uncertainty quantification for nonconvex tensor completion: Confidence intervals, heteroscedasticity and optimality. *IEEE Transactions on Information Theory*, 69(1):407–452.

Cai, T. T. and Li, X. (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *The Annals of Statistics*, 43(3):1027–1059.

Cai, T. T. and Zhang, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60–89.

Celentano, M., Fan, Z., and Mei, S. (2023). Local convexity of the TAP free energy and AMP convergence for z_2-synchronization. *The Annals of Statistics*, 51(2):519–546.

Chen, X. and Yang, Y. (2021). Cutoff for exact recovery of Gaussian mixture models. *IEEE Transactions on Information Theory*, 67(6):4223–4238.

Chen, Y. and Candès, E. J. (2018). The projected power method: An efficient algorithm for joint alignment from pairwise differences. *Communications on Pure and Applied Mathematics*, 71(8):1648–1714.

Chen, Y., Chi, Y., Fan, J., and Ma, C. (2021a). Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, 14(5):566–806.

Chen, Y., Chi, Y., Fan, J., Ma, C., and Yan, Y. (2020). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM journal on optimization*, 30(4):3098–3121.

Chen, Y., Fan, J., Ma, C., and Wang, K. (2019a). Spectral method and regularized MLE are both optimal for top-K ranking. *Annals of statistics*, 47(4):2204.

Chen, Y., Fan, J., Ma, C., and Yan, Y. (2019b). Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences*, 116(46):22931–22937.

Chen, Y., Fan, J., Ma, C., and Yan, Y. (2021b). Bridging convex and nonconvex optimization in robust PCA: Noise, outliers and missing data. *The Annals of Statistics*, 49(5):2948–2971.

Chen, Y., Fan, J., Wang, B., and Yan, Y. (2023). Convex and nonconvex optimization are both minimax-optimal for noisy blind deconvolution under random designs. *Journal of the American Statistical Association*, 118(542):858–868.

Chen, Y., Kamath, G., Suh, C., and Tse, D. (2016). Community recovery in graphs with locality. In *International conference on machine learning*, pages 689–698. PMLR.

Chi, E. C., Gaines, B. R., Sun, W. W., Zhou, H., and Yang, J. (2020). Provable convex co-clustering of tensors. *The Journal of Machine Learning Research*, 21(1):8792–8849.

Chin, P., Rao, A., and Vu, V. (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Conference on Learning Theory*, pages 391–423. PMLR.

Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C., and Phan, H. A. (2015). Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE signal processing magazine*, 32(2):145–163.

De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000). On the best rank-1 and rank-$(r_1, r_2, ..., r_n)$ approximation of higher-order tensors. *SIAM journal on Matrix Analysis and Applications*, 21(4):1324–1342.

Deng, Y., Tang, X., and Qu, A. (2023). Correlation tensor decomposition and its application in spatial imaging data. *Journal of the American Statistical Association*, 118(541):440–456.

Deshpande, Y., Abbe, E., and Montanari, A. (2017). Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference: A Journal of the IMA*, 6(2):125–170.

Florescu, L. and Perkins, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory*, pages 943–959. PMLR.

Fu, Y. and Dong, W. (2016). 3D magnetic resonance image denoising using low-rank tensor approximation. *Neurocomputing*, 195:30–39.

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2017). Achieving optimal misclassification proportion in stochastic block models. *The Journal of Machine Learning Research*, 18(1):1980–2024.

Gao, C. and Zhang, A. Y. (2021). Exact minimax estimation for phase synchronization. *IEEE Transactions on Information Theory*, 67(12):8236–8247.

Guédon, O. and Vershynin, R. (2016). Community detection in sparse networks via grothendieck's inequality. *Probability Theory and Related Fields*, 165(3-4):1025–1049.

Hajek, B., Wu, Y., and Xu, J. (2016a). Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797.

Hajek, B., Wu, Y., and Xu, J. (2016b). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Transactions on Information Theory*, 62(10):5918–5937.

Han, R., Luo, Y., Wang, M., and Zhang, A. R. (2022a). Exact clustering in tensor block model: Statistical optimality and computational limit. *Journal of the Royal Statistical Society Series B*, 84(5):1666–1698.

Han, R., Willett, R., and Zhang, A. R. (2022b). An optimal statistical and computational framework for generalized tensor estimation. *The Annals of Statistics*, 50(1):1–29.

Han, X., Tong, X., and Fan, Y. (2023). Eigen selection in spectral clustering: a theory-guided practice. *Journal of the American Statistical Association*, 118(541):109–121.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

Hopkins, S. B., Shi, J., and Steurer, D. (2015). Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory*, pages 956–1006. PMLR.

Hu, J. and Wang, M. (2023). Multiway spherical clustering via degree-corrected tensor block models. *IEEE Transactions on Information Theory*.

Javanmard, A., Montanari, A., and Ricci-Tersenghi, F. (2016). Phase transitions in semidefinite relaxations. *Proceedings of the National Academy of Sciences*, 113(16):E2218–E2223.

Jegelka, S., Sra, S., and Banerjee, A. (2009). Approximation algorithms for tensor clustering. In *International Conference on Algorithmic Learning Theory*, pages 368–383. Springer.

Kannan, R. and Vempala, S. (2009). Spectral algorithms. *Foundations and Trends® in Theoretical Computer Science*, 4(3–4):157–288.

Ke, Z. T. and Wang, J. (2022). Optimal network membership estimation under severe degree heterogeneity. *arXiv preprint arXiv:2204.12087*.

Lei, J., Chen, K., and Lynch, B. (2020). Consistent community detection in multi-layer network data. *Biometrika*, 107(1):61–73.

Lei, J. and Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237.

Lei, L. (2019). Unified $\ell_{2\to\infty}$ eigenspace perturbation theory for symmetric random matrices. *arXiv preprint arXiv:1909.04798*.

Li, G., Fan, W., and Wei, Y. (2023). Approximate message passing from random initialization with applications to z_2 synchronization. *Proceedings of the National Academy of Sciences (PNAS)*, 120(31).

Li, G. and Wei, Y. (2022). A non-asymptotic framework for approximate message passing in spiked models. *arXiv preprint arXiv:2208.03313*.

Li, N. and Li, B. (2010). Tensor completion for on-board compression of hyperspectral images. In *2010 IEEE International Conference on Image Processing*, pages 517–520. IEEE.

Li, X., Chen, Y., and Xu, J. (2021). Convex relaxation methods for community detection. *Statistical science*, 36(1):2–15.

Li, X., Li, Y., Ling, S., Strohmer, T., and Wei, K. (2020). When do birds of a feather flock together? k-means, proximity, and conic programming. *Mathematical Programming*, 179:295–341.

Ling, S. (2022). Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *Applied and Computational Harmonic Analysis*, 60:20–52.

Liu, A. and Moitra, A. (2020). Tensor completion made practical. *Advances in Neural Information Processing Systems*, 33:18905–18916.

Löffler, M., Zhang, A. Y., and Zhou, H. H. (2021). Optimality of spectral clustering in the Gaussian mixture model. *The Annals of Statistics*, 49(5):2506–2530.

Lounici, K. (2014). High-dimensional covariance matrix estimation with missing observations. *Bernoulli*, 20(3):1029–1058.

Lu, Y. and Zhou, H. H. (2016). Statistical and computational guarantees of Lloyd's algorithm and its variants. *arXiv preprint arXiv:1612.02099*.

Lyu, Z. and Xia, D. (2022). Optimal clustering by lloyd algorithm for low-rank mixture model. *arXiv preprint arXiv:2207.04600*.

Ma, C., Wang, K., Chi, Y., and Chen, Y. (2020). 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.

Mai, Q., Zhang, X., Pan, Y., and Deng, K. (2022). A doubly enhanced em algorithm for model-based tensor clustering. *Journal of the American Statistical Association*, 117(540):2120–2134.

Milligan, G. W. and Cooper, M. C. (1986). A study of the comparability of external criteria for hierarchical cluster analysis. *Multivariate behavioral research*, 21(4):441–458.

Montanari, A. and Sun, N. (2018). Spectral algorithms for tensor completion. *Communications on Pure and Applied Mathematics*, 71(11):2381–2425.

Mossel, E., Neeman, J., and Sly, A. (2014). Belief propagation, robust reconstruction and optimal recovery of block models. In *Conference on Learning Theory*, pages 356–370. PMLR.

Mossel, E., Neeman, J., and Sly, A. (2015). Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162:431–461.

Nasiri, M., Rezghi, M., and Minaei, B. (2014). Fuzzy dynamic tensor decomposition algorithm for recommender system. *UCT Journal of Research in Science, Engineering and Technology*, 2(2):52–55.

Ndaoud, M. (2022). Sharp optimal recovery in the two component Gaussian mixture model. *The Annals of Statistics*, 50(4):2096–2126.

Richard, E. and Montanari, A. (2014). A statistical model for tensor PCA. In *Advances in Neural Information Processing Systems*, pages 2897–2905.

Rohe, K., Chatterjee, S., and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic block-model. *The Annals of Statistics*, 39(4):1878–1915.

Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., and Faloutsos, C. (2017). Tensor decomposition for signal processing and machine learning. *IEEE Transactions on signal processing*, 65(13):3551–3582.

Singer, A. (2011). Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36.

Sun, W. W. and Li, L. (2019). Dynamic tensor clustering. *Journal of the American Statistical Association*, 114(528):1894–1907.

Tong, T., Ma, C., Prater-Bennette, A., Tripp, E., and Chi, Y. (2022). Scaling and scalability: Provable non-convex low-rank tensor estimation from incomplete measurements. *Journal of Machine Learning Research*, 23(163):1–77.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17:395–416.

Wang, M., Fischer, J., and Song, Y. S. (2019). Three-way clustering of multi-tissue multi-individual gene expression data using semi-nonnegative tensor decomposition. *The annals of applied statistics*, 13(2):1103.

Wang, M. and Zeng, Y. (2019). Multiway clustering via tensor block models. *Advances in neural information processing systems*, 32.

Wozniak, J. R., Krach, L., Ward, E., Mueller, B. A., Muetzel, R., Schnoebelen, S., Kiragu, A., and Lim, K. O. (2007). Neurocognitive and neuroimaging correlates of pediatric traumatic brain injury: a diffusion tensor imaging (dti) study. *Archives of Clinical Neuropsychology*, 22(5):555–568.

Wu, J., Lin, Z., and Zha, H. (2019). Essential tensor learning for multi-view spectral clustering. *IEEE Transactions on Image Processing*, 28(12):5910–5922.

Xia, D. (2021). Normal approximation and confidence region of singular subspaces. *Electronic Journal of Statistics*, 15(2):3798–3851.

Xia, D., Yuan, M., and Zhang, C.-H. (2021). Statistically optimal and computationally efficient low rank tensor completion from noisy entries. *The Annals of Statistics*, 49(1).

Xia, D., Zhang, A. R., and Zhou, Y. (2022). Inference for low-rank tensors—no need to debias. *The Annals of Statistics*, 50(2):1220–1245.

Yan, Y., Chen, Y., and Fan, J. (2021). Inference for heteroskedastic PCA with missing data. *arXiv preprint arXiv:2107.12365*.

Yang, Y. and Ma, C. (2022). Optimal tuning-free convex relaxation for noisy matrix completion. *arXiv preprint arXiv:2207.05802*.

Yuan, M. and Zhang, C.-H. (2016). On tensor completion via nuclear norm minimization. *Foundations of Computational Mathematics*, 16(4):1031–1068.

Zhang, A. and Xia, D. (2018). Tensor SVD: Statistical and computational limits. *IEEE Transactions on Information Theory*, 64(11):7311–7338.

Zhang, A. R., Cai, T. T., and Wu, Y. (2022). Heteroskedastic PCA: Algorithm, optimality, and applications. *The Annals of Statistics*, 50(1):53–80.

Zhang, A. R. and Zhou, Y. (2020). On the non-asymptotic and sharp lower tail bounds of random variables. *Stat*, 9(1):e314.

Zhang, A. Y. (2023). Fundamental limits of spectral clustering in stochastic block models. *arXiv preprint arXiv:2301.09289*.

Zhang, A. Y. and Zhou, H. H. (2016). Minimax rates of community detection in stochastic block models. *The Annals of Statistics*, 44(5):2252–2280.

Zhang, A. Y. and Zhou, H. H. (2022). Leave-one-out singular subspace perturbation analysis for spectral clustering. *arXiv preprint arXiv:2205.14855*.

Zhang, C., Han, R., Zhang, A. R., and Voyles, P. M. (2020). Denoising atomic resolution 4d scanning transmission electron microscopy data with tensor singular value decomposition. *Ultramicroscopy*, 219:113123.

Zhong, Y. and Boumal, N. (2018). Near-optimal bounds for phase synchronization. *SIAM Journal on Optimization*, 28(2):989–1016.

Zhou, H., Li, L., and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552.

Zhou, Y. and Chen, Y. (2023). Deflated HeteroPCA: Overcoming the curse of ill-conditioning in heteroskedastic PCA. *arXiv preprint arXiv:2303.06198*.

Zhou, Y., Zhang, A. R., Zheng, L., and Wang, Y. (2022). Optimal high-order tensor SVD via tensor-train orthogonal iteration. *IEEE Transactions on Information Theory*, 68(6):3991–4019.