Tensor Clustering with Heteroskedastic Noise

Yuchen Zhou[*] Yuxin Chen[*†]

October 6, 2023

Abstract

Tensor clustering, which seeks to uncover the underlying cluster structures from tensor observations, has garnered increasing attention recently. A commonly studied model in the literature is the tensor block model, assuming clustering structures along each mode. This model has been widely applied in various fields, including multi-tissue gene expression analysis and multilayer network analysis. However, existing methods on tensor clustering are limited to the i.i.d. noise setting, and hence can be highly unreliable for the count data or binary data arising in the aforementioned main application scenarios.

To better cope with heteroskedastic noise, we propose a novel method, named High-order HeteroClustering (HHC), which first tailors Deflated-HeteroPCA with a thresholding procedure to obtain subspace estimates and then performs approximate k-means to obtain clustering results. Theoretically, we demonstrate that our algorithm can achieve exact clustering, provided that the signal-to-noise ratio – the ratio of the minimum of the distances between center matrices for each mode and the noise level – exceeds the computational limit (ignoring logarithmic factors). Here, the computational limit is the minimum threshold that ensures exact clustering is computable within polynomial time. Through comprehensive simulation and real data experiments, we find our algorithm outperforms popular methods in the literature across various settings, delivering more reliable and reasonable clustering results.

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 the success of tensor data analysis in many applications, 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), signal processing (Cichocki et al., 2015), etc. Compared to vectors and matrices, tensors (or multidimensional arrays) can characterize the complex interrelations across multiple dimensions and simultaneously capture effects brought by multiple factors. In the literature, various tensor estimation problems and methodologies have been extensively studied.

In addition to tensor estimation problems, tensor clustering, which aims to reveal the underlying cluster structures from tensor observations, has garnered increasing attention recently. A commonly studied model is the tensor block model (Wang and Zeng, 2019; Chi et al., 2020; Han et al., 2022a): suppose that the data is an order-3 tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, for each mode $i \in [3]$, there is a cluster assignment vector $\boldsymbol{z}_i^\star \in [k_i]^{n_i}$ that divides n_i indices to k_i clusters. Here, for any positive integer d, we define $[d] = \{1, \dots, d\}$. The observation satisfies

$$Y_{i_1,i_2,i_3} = S^\star_{(\boldsymbol{z}_1^\star)_{i_1},(\boldsymbol{z}_2^\star)_{i_2},(\boldsymbol{z}_3^\star)_{i_3}} + E_{i_1,i_2,i_3}, \qquad \forall (i_1,i_2,i_3) \in [n_1] \times [n_2] \times [n_3]. \tag{1}$$

Here, $S^\star_{j_1,j_2,j_3}$ is the (j_1, j_2, j_3)-th entry of an underlying core tensor $\boldsymbol{\mathcal{S}}^\star \in \mathbb{R}^{k_1 \times k_2 \times k_3}$ and E_{i_1,i_2,i_3}'s are independent zero-mean noise. We aim to find out the clustering structures along each mode. Equivalently, we can rewrite Model (1) as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{E}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}, \tag{2}$$

where the underlying tensor $\boldsymbol{\mathcal{X}} = [S^\star_{(\boldsymbol{z}_1^\star)_{i_1},(\boldsymbol{z}_2^\star)_{i_2},(\boldsymbol{z}_3^\star)_{i_3}}]_{1 \le i_j \le n_j, 1 \le j \le 3}$ has high-order block structures, $\boldsymbol{\mathcal{E}}$ is a noise tensor, and the goal is to recover the community structures, or the $\boldsymbol{z}_i^{\star\prime}$s, based on $\boldsymbol{\mathcal{Y}}$. This model has a wide 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, different tissues, and individuals, which can be captured by the model above. Another instance is 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 many clustering algorithms have been extensively 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 tensor structures and thus the results can be highly inaccurate. To address this issue, Han et al. (2022a) proposed a spectral clustering method called High-order Spectral Clustering (HSC), which projects the tensor data on their estimated top singular spaces along each mode, and then performs approximate k-means. In particular, their singular subspace estimation procedure are built on top of a vanilla SVD-based approach (i.e., directly calculating the left singular spaces of the unfoldings of $\boldsymbol{\mathcal{Y}}$). In addition, they introduced an algorithm, High-order Lloyd Algorithm (HLloyd), to further update the block membership estimates. When the noise tensor $\boldsymbol{\mathcal{E}}$ has i.i.d. sub-Gaussian noise entries, Han et al. (2022a) proved that under a computationally near-optimal signal-to-noise ratio (SNR) condition, HSC can provide us with consistent block membership estimators and it combined with HLloyd can even achieve exact community recovery. Here, the SNR is defined as quotient of the minimum of the distances between center matrices for each mode, and the noise level (see Section 2 for a more detailed description).

However, in many applications, especially when the entries of the observations are count data or binary, it is more reasonable to assume that the noise can be heteroskedastic (i.e., the variances of the entries of $\boldsymbol{E}$ can vary drastically across different locations). In the presence of heteroskedastic noise, recent works (Zhang et al., 2022; Cai et al., 2021; Zhou and Chen, 2023) show that the vanilla SVD-based approach may return sup-optimal estimators. As a consequence, HSC, which heavily relies on SVD initializations, becomes less reliable. This obstacle severely hinders the performance of HSC in a wide array of applications with discrete observations, including multi-tissue gene expression data analysis and multilayer network data analysis, which are two main motivations for the tensor block model. To the best of our knowledge, under Model (1) with heteroskedastic noise, no known computationally-efficient algorithm can consistently estimate (not to mention exactly recover) the community structures as long as the SNR reaches the computational limit proved in Han et al. (2022a).

Our contributions. In this paper, we fill this void by proposing a novel spectral clustering method for tensors, called High-order HeteroClustering (HHC). Different from HSC, we devise a new strategy for estimating the top singular subspaces that performs well even in the presence of heteroskedastic noise. As a consequence, the subsequent clustering result is also robust to the heteroskedasity of noise. More specifically, HHC consists of two steps:

1. *Subspace estimation.* We leverage an idea from a recently proposed subspace estimation algorithm, Deflated-HeteroPCA (Zhou and Chen, 2023), which has been shown to accurately estimate the subspace even when the noise is heteroskedastic, provided that the noise does not mask the least singular value of $\boldsymbol{\mathcal{X}}^\star$. Nevertheless, for the clustering problem, there is no restriction on the singular values of $\boldsymbol{\mathcal{X}}^\star$ or $\boldsymbol{\mathcal{S}}^\star$, and they can even be degenerate. Applying Deflated-HeteroPCA directly to estimate the entire subspace of the unfoldings of $\boldsymbol{\mathcal{X}}^\star$ could result in inaccurate estimations of the subspace corresponding to small singular values. This inaccuracy may further affect the clustering results.

   To resolve this problem, we introduce a new method called Thresholded Deflated-HeteroPCA, which is similar to Deflated-HeteroPCA but includes a thresholding procedure. This procedure ensures that we only estimate the "useful" column subspaces – that is, the subspace associated with large singular values – of the unfoldings. Our method is adaptive, as the tuning parameter can be chosen in a theory-guided data-driven manner.

2. *Clustering.* Once we have these subspace estimates, we perform approximate k-means on the rows of a "denoised version" of the i-th unfolding of $\boldsymbol{\mathcal{Y}}$. This 'denoised version' is derived from $\boldsymbol{\mathcal{Y}}$ and the subspace estimates, the $\boldsymbol{U}_i'$s. This process yields the label vector estimate $\widehat{z}_i$ for each $i \in \{1, 2, 3\}$.

Theoretically, we prove that HHC can exactly recover the community structures as long as some "necessary" SNR condition holds (up to logarithmic factors). Here, a "necessary" SNR condition means that the condition is essential to ensure that the label 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 community structures, and that HHC combined with HLloyd (Han et al., 2022a) can achieve better numerical performance. We also apply our method (HHC + HLloyd) and HSC + HLloyd to the flight route network data, and our method outputs more reasonable clustering results. It is worth noting that both our algorithm and the proof can be easily extended to general order-d tensors for $d > 3$.

Paper organization. The rest of the article is organized as follows. In Sections 2 and 3, we introduce the model setting and our algorithm, respectively. We provide theoretical guarantees for our algorithm in Section 4. Numerical experiments and real data analysis are provided in Section 5.

1.1 Notation

Throughout the paper, the bold capital letters, e.g., $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$, and the bold lowercase letters, e.g., $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, are used to denote matrices and vectors, respectively. Let $[n] := \{1, \ldots, n\}$. 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 $\boldsymbol{A}$, respectively. We define the $\ell_{2,\infty}$ norm of $\boldsymbol{A}$ 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 sets all non-diagonal entries to zero, and we 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 $\boldsymbol{a} = (a_1, \ldots, a_n)$, 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.

We use the boldface calligraphic letters, e.g., $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{Y}}$, $\boldsymbol{\mathcal{Z}}$, to represent tensors. For tensor $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1\times\cdots\times r_d}$ and matrix $\boldsymbol{V}_1 \in \mathbb{R}^{n_i\times r_i}$, we define the multi-linear product $\times_1$ as follows:

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

We can also define $\times_\ell$ for all $\ell \in [d]$ similarly. For any tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1\times\cdots\times n_d}$, let $\mathcal{M}_j(\boldsymbol{\mathcal{X}}) \in \mathbb{R}^{n_j\times(\prod_{i=1}^d n_i/n_j)}$ denote the j-th matricization of $\boldsymbol{\mathcal{X}}$ such that for any $(i_1, ..., i_d) \in [n_1]\times\cdots\times[n_d]$,

$$[\mathcal{M}_j(\boldsymbol{\mathcal{X}})]_{i_j, i_{j+1}+\sum_{\ell=2}^{d-1}(\prod_{h=2}^{\ell} n_{j+h-1})(i_{j+\ell}-1)} = X_{i_1,\ldots,i_d}.$$

For example, when $d = 3$,

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

We define the Frobenious norm of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ as follows:

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

We say $f(x) \lesssim g(x)$ or $f(x) = O(g(x))$ if $|f(x)| \le Cg(x)$ for some constant $C > 0$; we let $f(x) \gtrsim g(x)$ denote $f(x) \ge C|g(x)|$ 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 $f(n_1, n_2) \le cg(n_1, n_2)$ 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 $\min\{n_1, n_2\} \to 0$. For any $a, b \in \mathbb{R}$, let $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$. For notational convenience, for any $\phi \in \Phi$, we say $\overline{\boldsymbol{z}} = \phi(\boldsymbol{z})$ if

$$\overline{\boldsymbol{z}}_j = \phi(\boldsymbol{z}_j), \qquad \forall j \in [n],$$

where Φ is set consisting of all permutations $\phi : [k] \to [k]$.

2 Problem formulation

Models. For $d = 3$, Model (1) can be equivalently written as

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}}^\star + \boldsymbol{\mathcal{E}}, \tag{3}$$

where the j-th mode of the true tensor $\boldsymbol{\mathcal{X}}^\star = \boldsymbol{\mathcal{S}}^\star \times_1 \boldsymbol{M}_1^\star \times_2 \boldsymbol{M}_2^\star \times_3 \boldsymbol{M}_3^\star$ with the center tensor $\boldsymbol{\mathcal{S}}^\star \in \mathbb{R}^{k_1 \times k_2 \times k_3}$ and membership matrices $\boldsymbol{M}_i^\star \in \{0,1\}^{n_i \times k_i}$ satisfying

$$(\boldsymbol{M}_i^\star)_{j,\ell} = \begin{cases} 1, & \text{if } (\boldsymbol{z}_i^\star)_j = \ell, \\ 0, & \text{else.} \end{cases} \tag{4}$$

Our goal is to recover the label vectors $\boldsymbol{z}_i^\star$ or equivalently the membership matrices $\boldsymbol{M}_i^\star$, $i \in [3]$. As revealed by Han et al. (2022a), the following two kinds of tensor clustering models, which find important applications, are special cases of Model (3):

1. *Sub-Gaussian tensor block models.* The entries of the noise tensor $\boldsymbol{\mathcal{E}}$ are independent, zero-mean, and sub-Gaussian random variables. In addition to the models studied in Han et al. (2022a), here we allow the noise to be heteroskedastic.
2. *Stochastic tensor block models.* The entries of the core tensor $\boldsymbol{\mathcal{S}}^\star$ fall in the interval $[0,1]$ and the Y_{i_1,i_2,i_3}'s are independent Bernoulli random variables satisfying

$$Y_{i_1,i_2,i_3} = \begin{cases} 1, & \text{with probability } S^\star_{(\boldsymbol{z}_1^\star)_{i_1},(\boldsymbol{z}_2^\star)_{i_2},(\boldsymbol{z}_3^\star)_{i_3}}, \\ 0, & \text{with probability } 1 - S^\star_{(\boldsymbol{z}_1^\star)_{i_1},(\boldsymbol{z}_2^\star)_{i_2},(\boldsymbol{z}_3^\star)_{i_3}}. \end{cases} \tag{5}$$

   Under this model, the noise is heteroskedastic unless there exists a quantity $p \in [0,1]$ such that the entries of $\boldsymbol{\mathcal{S}}^\star$ are either p or $1-p$, which is fairly restricted.

Definitions and assumptions. Now, we introduce the definitions and assumptions that will be used in the paper. We define

$$n = \max_{1 \le i \le 3} n_i, \qquad \text{and} \qquad k = \max_{1 \le i \le 3} k_i. \tag{6}$$

We start with the assumption on the noise tensor $\boldsymbol{\mathcal{E}}$. We denote

$$\omega_{i_1,i_2,i_3}^2 := \mathbb{E}\left[E_{i_1,i_2,i_3}^2\right] \qquad \text{and} \qquad \omega_{\mathsf{max}}^2 := \max_{i_1 \in [n_1], i_2 \in [n_2], i_3 \in [n_3]} \omega_{i,j,k}^2. \tag{7}$$

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. $\mathbb{P}(|E_{i_1,i_2,i_3}| > B) \le n^{-24}$, *where B satisfies*

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

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

Remark 1. *Assumption 1 is very mild and can be easily satisfied. For example, all ω_{max}-sub-Gaussian random variables satisfy Condition 2 of Assumption 1 with $B \asymp \omega_{\mathsf{max}} \sqrt{\log n}$. Moreover, centered Poisson random variables also satisfy this condition (Boucheron et al., 2013; Zhang and Zhou, 2020). In addition, the aforementioned stochastic tensor block models meet the conditions of Assumption 1 as long as the following conditions holds:*

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

Recognizing that $(n_1 n_2 n_3)^{1/2} \gg \log^2 n$, we know that this assumption holds for stochastic tensor block models as long as the entries of the center tensor $\boldsymbol{\mathcal{S}}^\star$ are not extremely close to 0 and 1.

In addition, we define the following cluster size parameter β.

Definition 1. *Let $\beta \leq 1$ denote the largest quantity such that for all $i \in [3], \ell \in [k_i]$,*

$$\left|\left\{j \in [n_i] : \left(\boldsymbol{z}_i^\star\right)_j = \ell\right\}\right| \geq \beta n_i / k_i. \tag{8}$$

The parameter β measures how balanced these cluster sizes are, with a larger β indicating more balanced sizes. We also let

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

denote the minimum distance between the rows of $\mathcal{M}_i(\mathcal{S})$, the i-th matrizication of the core tensor and the signal strength is defined as

$$\Delta_{\mathsf{min}}^2 := \min\left\{\Delta_1^2, \Delta_2^2, \Delta_3^2\right\}.$$

We define the signal-to-noise ratio as follows:

$$\mathrm{SNR} := \Delta_{\mathsf{min}} / \omega_{\mathsf{max}}. \tag{9}$$

Different from the low-rank tensor estimation, we aim for the clustering performance results to depend solely on the SNR defined in (9), without being affected by the least singular value of $\boldsymbol{\mathcal{S}}^\star$. More precisely, we want to design an polynomial-time algorithm that can exactly recover the $\boldsymbol{z}_i^{\star'}$s under the weakest possible assumption on the SNR.

3 Algorithm

In this section, we introduce the procedure of the proposed algorithm, aimed at achieving the above goal.

High-order HeteroClustering : Thresholded Deflated-HeteroPCA + approximate k-means. Similar to other spectral clustering methods, our algorithm begins with subspace estimation, followed by the application of the (approximate) k-means algorithm. However, as we discussed earlier, the vanilla SVD-based approach may provide unsatisfying subspace estimation results when the noise is heteroskedastic. Therefore, we use a more sophisticated algorithm instead to accomplish this task. A detailed description of our algorithm is given as follows:

Step 1: Thresholded Deflated-HeteroPCA. First, we need to estimate the "important" column subspaces of the unfoldings of $\boldsymbol{\mathcal{X}}$. Here, it is not necessary to estimate the entire rank-k_i column spaces for these unfoldings as the singular vectors corresponding to small singular values have a negligible impact on the final clustering results. Instead, for each $i \in [3]$, we want to find a suitable estimator $\boldsymbol{U}_i \in \mathcal{O}^{n_i, r_i}$ for some $r_i \leq k_i$ such that it can accurately estimate the subspace formed by the singular vectors associated with large enough singular values of $\boldsymbol{X}_i = \mathcal{M}_i\left(\boldsymbol{\mathcal{X}}\right)$.

Review: Deflated-HeteroPCA. Although heteroskedastic noise makes this task more challenging, fortunately, the Deflated-HeteroPCA algorithm (Zhou and Chen, 2023) provides us a ray of hope. We let $\boldsymbol{Y}_i = \mathcal{M}_i\left(\boldsymbol{\mathcal{Y}}\right)$ denote the i-th matricization of $\boldsymbol{Y}_i$. Starting from $\boldsymbol{G}_0 = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{Y}_i \boldsymbol{Y}_i^\top\right)$, 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 add new blocks and apply HeteroPCA (Zhang et al., 2022) with rank r_k and initialization $\boldsymbol{G}_k$ to impute the diagonal entries, thus obtaining $\boldsymbol{G}_{k+1}$. The subspace estimator can then be calculated by performing SVD on the final matrix estimator $\boldsymbol{G}_{k_{\mathsf{max}}}$. As shown in Zhou and Chen (2023), Deflated-HeteroPCA enjoys promising theoretical guarantees and does not make any assumptions on the condition number of the signal matrix. This is suitable for tensor clustering, as we hope the clustering result depends only on the SNR and we do not want to make unnecessary assumptions on the singular values of the signal part. However, it requires some conditions on the ratio between the least singular value of the signal matrix and the noise level. Under our setting, where some singular values of $\boldsymbol{X}_i$ can even be zero, Deflated-HeteroPCA may fail to estimate the whole column subspace.

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

input: data matrix $\boldsymbol{Y}$, rank r, threshold τ, maximum number of iterations t_i, $i = 1, 2, \ldots$
initialization: $k = 0, r_0 = 0, \boldsymbol{G}_0 = \mathcal{P}_{\text{off-diag}}\left(\boldsymbol{Y}\boldsymbol{Y}^\top\right)$.
if $k = 0$ **then**
  $k = k + 1$.
  select

$$r_k = \begin{cases} \max \mathcal{R}_k, & \text{if } \mathcal{R}_k \neq \emptyset, \\ r, & \text{otherwise.} \end{cases} \tag{10}$$

  Here,

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

  $(\boldsymbol{G}_k, \boldsymbol{U}_k) =$HeteroPCA$(\boldsymbol{G}_{k-1}, r_k, t_k)$.
while $r_k < r$ and $\sigma_{r_k+1}(\boldsymbol{G}_k) > \tau$ **do**
  $k = k + 1$.
  select r_k via (10).
  $(\boldsymbol{G}_k, \boldsymbol{U}_k) =$HeteroPCA$(\boldsymbol{G}_{k-1}, r_k, t_k)$.
output: subspace estimate $\boldsymbol{U} = \boldsymbol{U}_k$.

Thresholding procedure. To address this issue, we introduce a thresholding procedure to make sure that we only choose 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 $\boldsymbol{G}_k$. Then we proceed to the $(k+1)$-th round only if the condition $\sigma_{r_k+1}(\boldsymbol{G}_k) > \tau$ holds for some pre-determined threshold τ. With a properly chosen τ, we can extract all essential information needed for clustering. The complete details of Thresholded Deflated-HeteroPCA are summarized in Algorithm 1, and a theoretical-guided selection procedure of the tuning parameter τ will be provided in Section 4.2.

Step 2: *Approximate k-means.* Having obtained these subspace estimates $\boldsymbol{U}_j, j \in [3]$, we apply k-means on the rows of $\widehat{\boldsymbol{B}}_i = \boldsymbol{U}_i\boldsymbol{U}_i^\top \mathcal{M}_i(\boldsymbol{\mathcal{Y}})\big(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1}\big) \in \mathbb{R}^{n_i \times (r_1 r_2 r_3 / r_i)}$ for each $i \in [3]$ to obtain the label vector estimates. Here, the indices $i+1$ and $i+2$ are computed module 3. This procedure is equivalent to performing 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 offers a significant reduction in matrix size. While solving the exact k-means is usually computationally hard, we instead employ an M-approximate k-means for some $M > 1$. More specifically, we find a label vector estimate $\widehat{z}_i$ and a centroid $\widehat{\boldsymbol{b}}_j^{(i)}$ that satisfy (11a) for each $i \in [3]$. This relaxed version of k-means can be efficiently solved by many algorithms. For example, for $M = 1 + \varepsilon$ with any fixed $\varepsilon > 0$, Kumar et al. (2004) showed that the problem can be solved in linear time; for $M = O(\log k)$, this problem can be efficiently solved by k-means++ (Arthur and Vassilvitskii, 2007). The requirement (11b) ensures the "optimality" of the label estimates given the center estimates.

The full procedure of High-order HeteroClustering (HHC) is summarized in Algorithm 3. Since both Thresholded Deflated-HeteroPCA and Approximate k-means are polynomial-time algorithms, our method is computationally efficient. Empirically, we recommend using the high-order Lloyd Algorithm (HLloyd) (Han et al., 2022a, see also Algorithm 4) to further refine the clustering results obtained by HHC. 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.

Algorithm 2: HeteroPCA($\boldsymbol{G}_{\mathsf{in}}$, r, t_{max}) (Zhang et al., 2022)

input: symmetric matrix $\boldsymbol{G}_{\mathsf{in}}$, rank r, number of iterations t_{max}.
initialization: $\boldsymbol{G}^0 = \boldsymbol{G}_{\mathsf{in}}$.
for $t = 0, 1, \ldots, t_{\mathsf{max}}$ **do**
  $\boldsymbol{U}^t \boldsymbol{\Lambda}^t \boldsymbol{U}^{t\top} \leftarrow$ rank-r leading eigendecompostion of $\boldsymbol{G}^t$.
  $\boldsymbol{G}^{t+1} = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{G}^t\right) + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{U}^t \boldsymbol{\Lambda}^t \boldsymbol{U}^{t\top}\right)$.
output: matrix estimate $\boldsymbol{G} = \boldsymbol{G}^{t_{\mathsf{max}}}$ and subspace estimate $\boldsymbol{U} = \boldsymbol{U}^{t_{\mathsf{max}}}$.

Algorithm 3: High-order HeteroClustering (HHC)

Input: observed tensor $\boldsymbol{\mathcal{Y}}$, numbers of clusters k_1, k_2, k_3, numbers of iterations $\{t_{i,j}\}_{i\in[3], j\geq 1}$, thresholds τ_1, τ_2, τ_3, constant $M > 1$.
initialization:

$$\boldsymbol{U}_i = \begin{cases} \text{Thresholded Deflated-HeteroPCA}(\mathcal{M}_i(\boldsymbol{\mathcal{Y}}), k_i, \tau_i, \{t_{i,j}\}_{j\geq 1}), & k_i \geq 2, \\ (1, \ldots, 1)^\top / \sqrt{n_i}, & k_i = 1 \end{cases} \in \mathcal{O}^{n_i, k_i}, \qquad i \in [3].$$

for $i = 1, 2, 3$ **do**
  $\widehat{\boldsymbol{B}}_i = \boldsymbol{U}_i \boldsymbol{U}_i^\top \mathcal{M}_i(\boldsymbol{\mathcal{Y}})\big(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1}\big) \in \mathbb{R}^{n_i \times (r_1 r_2 r_3 / r_i)}$.
  perform approximate k-means on the rows of $\widehat{\boldsymbol{B}}_i$, i.e., find a label vector estimate $\widehat{\boldsymbol{z}}_i \in [k_i]^{n_i}$ and center estimates $\widehat{\boldsymbol{b}}_j^{(i)} \in \mathbb{R}^{r_1 r_2 r_3 / r_i}$, $j \in [k_i]$ such that

$$\sum_{j=1}^{n_i} \left\| \big(\widehat{\boldsymbol{B}}_i\big)_{j,:}^\top - \widehat{\boldsymbol{b}}_{(\widehat{\boldsymbol{z}}_i)_j}^{(i)} \right\|_2^2 \leq M \min_{\substack{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{k_i} \in \mathbb{R}^{r_1 r_2 r_3 / r_i} \\ \boldsymbol{z}_i \in [k_i]^{n_i}}} \sum_{j=1}^{n_i} \left\| \big(\widehat{\boldsymbol{B}}_i\big)_{j,:}^\top - \boldsymbol{b}_{(\boldsymbol{z}_i)_j} \right\|_2^2, \tag{11a}$$

$$\left(\widehat{\boldsymbol{z}}_i\right)_j \in \arg\min_{\ell \in [k_i]} \left\| \big(\widehat{\boldsymbol{B}}_i\big)_{j,:}^\top - \widehat{\boldsymbol{b}}_\ell^{(i)} \right\|_2. \tag{11b}$$

Output: label vector estimates $\widehat{\boldsymbol{z}}_1, \widehat{\boldsymbol{z}}_2, \widehat{\boldsymbol{z}}_3$.

4 Main theory

In this section, we provide theoretical guarantees for the proposed algorithm. We let $\sigma_{i,j}^\star$ denote the j-th largest singular value of $\mathcal{M}_i(\mathcal{S}^\star)$, the i-th matricization of $\mathcal{S}^\star$. For $i \in [3]$ and denote by $\{r_{i,j}\}_{j\geq 1}$ the ranks selected in Algorithm 2 with the input matrix $\boldsymbol{Y} = \mathcal{M}_i(\mathcal{Y})$, the rank $r = k_i$, the threshold τ_i and the number of iterations $t_{i,j}$, and $r_{i, j_{\mathsf{max}}^i}$ is the largest rank selected by Algorithm 1.

4.1 Exact clustering guarantees for High-order HeteroClustering

The first theorem shows that HHC achieves promising exact recovery guarantees if the tuning parameter τ is selected appropriately.

Theorem 1. *Suppose that Assumption 1 holds, $k_i = O(1), i \in [3]$, $\beta = \Omega(1)$, and*

$$n_1 n_2 n_3 \geq c_1 n^2, \tag{12a}$$

$$c_\tau \left(n_1 n_2 n_3\right)^{1/2} \log^2 n \leq \tau_i / \omega_{\mathsf{max}}^2 \leq C_\tau \left(n_1 n_2 n_3\right)^{1/2} \log^2 n, \qquad \forall i \in [3], \tag{12b}$$

$$\mathrm{SNR} \ = \Delta_{\mathsf{min}} / \sigma_{\mathsf{max}} \geq C_1 \left(n_1 n_2 n_3\right)^{-1/4} \log n, \tag{12c}$$

where C_1, c_1, C_τ and c_τ are some large enough constants. If the numbers of iterations satisfy

$$t_{i,j} \geq \log\left(C \frac{\sigma_{i, r_{i,j-1}+1}^{\star 2}}{\sigma_{i, r_{i,j}+1}^{\star 2}} \right), \qquad 1 \leq j \leq j_{\mathsf{max}}^i - 1, \tag{13a}$$

$$t_{i,j^i_{\max}} \geq \log\left(n^3 \frac{\sigma^{\star 2}_{i,r_{i,j^i_{\max}-1}+1}}{\omega^2_{\max}}\right), \tag{13b}$$

for all $i \in [3]$ with some large constant $C > 0$, then with probability exceeding $1 - O(n^{-10})$, the outputs $\widehat{\boldsymbol{z}}_i, i \in [3]$ returned by Algorithm 3 satisfy

$$\widehat{\boldsymbol{z}}_i = \phi_i\left(\boldsymbol{z}^\star_i\right)$$

for some permutation $\phi_i : [k_i] \to [k_i]$. In other words, Algorithm 3 can exactly recover the label vectors.

A more general version of the theorem, which allows k_i and β grows with n, along with its proof, can be found in Section B. Condition (12a) assumes that the dimensions of the observed tensor is not extremely unbalanced and it holds if $n_i \lesssim n_{i+1}n_{i+2}$ for all $i \in [3]$. Under the regime $n_1 \asymp n_2 \asymp n_3 \asymp n$, the signal-to-noise ratio condition (12c) becomes SNR $\gtrsim n^{-3/4}\log n$. This condition is computationally near-optimal – as shown in (Han et al., 2022a, Theorem 7), for any $\gamma < -3/4$, there is no polynomial-time algorithm that can exactly recover the label vectors if the SNR $= n^\gamma$. 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 label recovery. In other words, applying HLloyd to further update the results would not degrade the theoretical performance. In Section 5, we will demonstrate that empirically, HHC combined with HLloyd achieves more accurate results than HHC on its own.

Comparison with HSC and HSC + HLloyd. To highlight the advantages of both our algorithm and our theory, we compare our results with those from prior works. First, unlike HSC and HSC + HLloyd, which assumes equal variances of the noise entries in order to guarantee the desired theoretical results (Han et al., 2022a, Theorems 3 and 4), our algorithm HHC can effectively handle heteroskedastic noise without compromising the range of SNRs. Second, (Han et al., 2022a, Theorems 3 and 4) assume sub-Gaussian noise assumptions, our assumption is more mild and can accommodate a wider range of applications, including those with binary or count data outcomes. In addition, different from HSC, HHC can exactly recover the label vector without the aid of HLloyd.

A glimpse of proof highlights. To prove that HHC alone is enough to achieve the exact recovery result without any further updates, it turns out to be crucial to carefully control the magnitude of $\|(\boldsymbol{I}_{n_i} - \boldsymbol{U}_i\boldsymbol{U}_i^\top)\boldsymbol{X}^\star_i\|_{2,\infty}$. Here, $\boldsymbol{X}^\star_i = \mathcal{M}_i\left(\boldsymbol{\mathcal{X}}^\star\right)$ 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 $\left(\boldsymbol{I}_{n_i} - \boldsymbol{U}_i\boldsymbol{U}_i^\top\right)\boldsymbol{X}^\star_i$ without imposing any restrictions on the condition number and the least singular value of $\boldsymbol{X}^\star_i$. In this context, existing techniques for deriving $\ell_{2,\infty}$ guarantees, including the leave-one-out analysis (Zhong and Boumal, 2018; Chen et al., 2021a; Ma et al., 2020), which requires the condition number not too large, and the subspace representation theorem (Xia, 2021; Zhou and Chen, 2023), which is contingent on the least singular value assumption, are inadequate for reaching our target bound.

To establish such a result, we develop a new technical tool (Lemma 1) that enables us to control $\|(\boldsymbol{I}_{n_i} - \boldsymbol{U}_i\boldsymbol{U}_i^\top)\boldsymbol{X}^\star_i\|_{2,\infty}$ 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 that in Zhou and Chen (2023), we are able to deal with those $\ell_{2,\infty}$ norms of the error polynomials and thus achieve the desired guarantees.

4.2 Selecting tuning parameter τ_i

As shown in Theorem 1, HHC can successfully recover the label vectors if the tuning parameters $\tau_i, i \in [3]$ satisfy (12b). However, the maximum variance $\omega^2_{\max}$ is usually unknown and thus we need to carefully choose τ_i. In this subsection, we discuss how to select these tuning parameters. Without loss of generality, we assume $n_1 \leq n_2 \leq n_3$ and we let

$$\widehat{\omega} = \sigma_{k_1+1}\left(\mathcal{M}_1\left(\boldsymbol{\mathcal{Y}}\right)\right)/\sqrt{n_2 n_3}, \tag{14}$$

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

$$\tau_i = \tau = \overline{C}_\tau \left(n_1 n_2 n_3\right)^{1/2} \widehat{\omega}^2 \log^2 n = \overline{C}_\tau \sqrt{\frac{n_1}{n_2 n_3}} \sigma_{k_1+1}^2 \left(\mathcal{M}_1\left(\boldsymbol{\mathcal{Y}}\right)\right) \log^2 n, \qquad \forall i \in [3]. \tag{15}$$

Here, $\overline{C}_\tau$ is some sufficiently large constant. The following theorem asserts that τ_i, $i \in [3]$ defined in (15) match the desired property (12b).

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

1. *For all* $i \in [n_1]$,

$$\sum_{j=1}^{n_2} \sum_{\ell=1}^{n_3} \omega_{i,j,\ell}^2 \geq c n_2 n_3 \omega_{\mathsf{max}}^2 \qquad \text{where} \qquad \omega_{i,j,\ell}^2 = \mathsf{Var}\left[E_{i,j,\ell}\right] \tag{16}$$

   for some constant $c > 0$;

2. *Under the stochastic tensor block model (5), the numbers of clusters* $k_i = O(1), i \in [3]$, *and the cluster size parameter* $\beta = \Omega(1)$.

Then with probability exceeding $1 - O(n^{-10})$, *the thresholds* τ_i *defined in (15) satisfy (12b).*

Remark 2. *Here, Condition (16) posits that the average variance for each row of* $\mathcal{M}_1(\boldsymbol{\mathcal{E}})$ *has the same order of* ω_{max}^2. *This condition is met when the noise isn't 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,* $k_i = O(1), i \in [3]$, $\beta = \Omega(1)$, *and either the following condition is satisfied:*

1. *Condition (16) holds;*

2. *The model we are considering is the stochastic tensor block model (5).*

We further assume that

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

for some large enough constants $C_1, c_1 > 0$. *If we choose the tuning parameter* τ *as in (15) and the numbers of iterations satisfy (13a) and (13b), then with probability exceeding* $1 - O(n^{-10})$, HHC *can exactly recover the label vectors, i.e.,*

$$\widehat{\boldsymbol{z}}_i = \phi_i\left(\boldsymbol{z}_i^\star\right)$$

for some permutation $\phi_i : [k_i] \to [k_i]$.

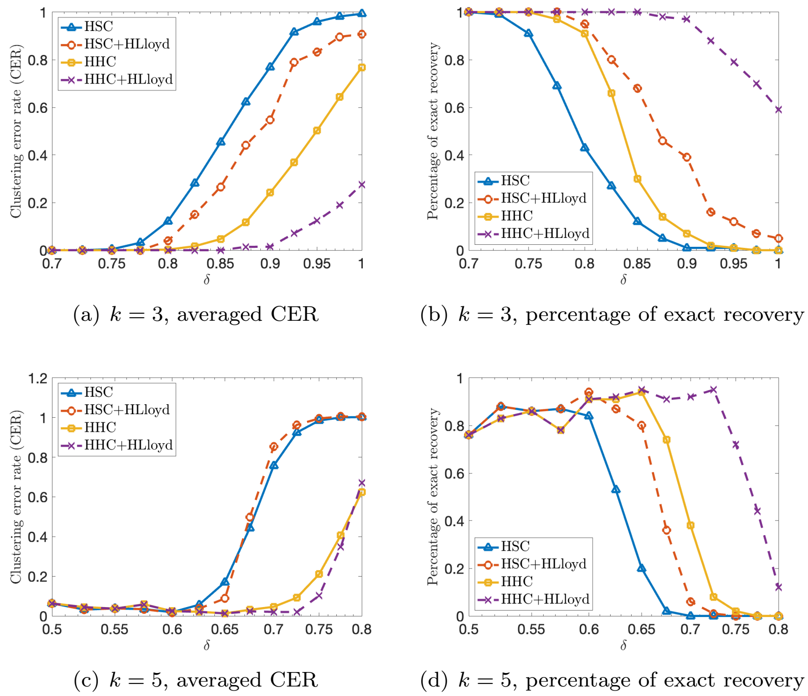
Theorem 3 shows that our data-driven procedure can still achieve the exact clustering under the same signal-to-noise ratio condition as in Theorem 1, provided that the noise condition 16. Specifically, for the stochastic tensor block model, no extra assumptions on the noise are needed.

5 Empirical studies

In this section, we study the empirical performance of the proposed algorithm, HHC, and HHC + HLloyd. Here, we choose the tuning parameter as

$$\tau = 1.1 \left(n_1 n_2 n_3\right)^{1/2} \widehat{\omega}^2 = 1.1 \sqrt{\frac{n_1}{n_2 n_3}} \sigma_{k_1+1}^2 \left(\mathcal{M}_1\left(\boldsymbol{\mathcal{Y}}\right)\right), \tag{17}$$

where $\widehat{\omega}$ is defined in (14).



(a) $k = 3$, averaged CER

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

(c) $k = 5$, averaged CER

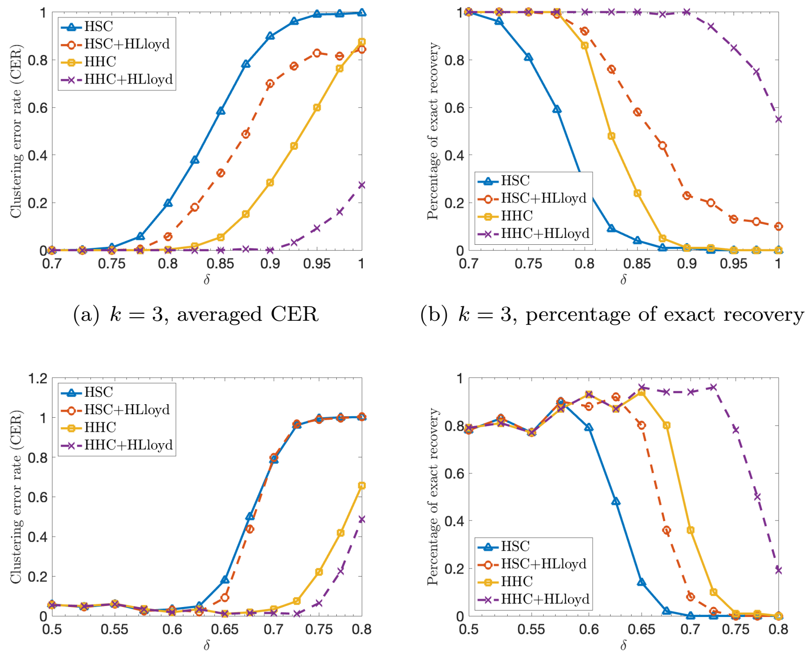
(d) $k = 5$, percentage of exact recovery

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, $\Delta_{\mathsf{min}} = 40n^{-\delta}$.

5.1 Synthetic data analysis

First, we conduct numerical experiments to show the usefulness of HHC and HHC + HLloyd. Following the setting in Han et al. (2022a), we set the dimensions $n_1 = n_2 = n_3 = n$, numbers of clusters $k_1 = k_2 = k_3 = k$, and the cluster sizes are balanced. We consider the following four method: (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) to refine the label estimators; (3) HHC: the method proposed in Algorithm 3 with the numbers of iterations $t_{i,j} = 10$; (4) HHC + HLloyd: the procedure that uses HHC (where $t_{i,j} = 10$) as initial label estimators and then applies HLloyd with the iteration number $t = 10$ to derive the final label estimates. To evaluate the clustering performance, for each method, we calculate the 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.

Sub-Gaussian tensor block models. First, we consider Model (3) with Gaussian noise. We fix the dimensions $n_1 = n_2 = n_3 = n \in \{100, 150\}$, generate a random tensor $\overline{\boldsymbol{\mathcal{S}}} \in \mathcal{R}^{k,k,k}$ with independent entries $\overline{S}_{i_1,i_2,i_3} \sim \mathcal{N}(0,1)$ and the core tensor $\boldsymbol{\mathcal{S}}^{\star}$ is obtained by rescaling $\overline{\boldsymbol{\mathcal{S}}}$ such that $\Delta_{\mathsf{min}} = 40n^{-\delta}$ (so the SNR decreases as δ increases). We randomly generate the labels $\boldsymbol{z}_i \in [k]^n$, $i \in [3]$. 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 $E_{i,j,k} \sim \mathcal{N}(0, \alpha_i^2 \beta_j^2 \gamma_k^2)$. We report the averaged CER and the percentage of exact label recovery for each method. We present the results for $n = 100$ and $n = 150$ in Figures 1 and 2, respectively. We can see that HHC and HHC + HLloyd achieves much smaller CER than HSC and HSC + HLloyd. In terms of the percentage of exact recovery, HHC + HLloyd has the best performance among all these four methods.



(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, $\Delta_{\mathsf{min}} = 40n^{-\delta}$.

Stochastic tensor block models. Next, we study the stochastic block model (5). We choose the core tensor $\boldsymbol{\mathcal{S}}^\star \in \mathbb{R}^{k\times k\times k}$ satisfying

$$S^\star_{i_1,i_2,i_3} = \begin{cases} 10a \cdot n^{-3/2}\left(1 - \frac{i_1-1}{2(k-1)}\right), & i_1 = i_2 = i_3, \\ 0.1a \cdot n^{-3/2}, & \text{otherwise.} \end{cases} \tag{18}$$

Here, a is a scalar. When a is not too large, the SNR increases with a. The results of the above four methods for $n = 100$ and $n = 150$ are presented in Figures 3 and 4, respectively. From these figures, we know 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 label vectors.

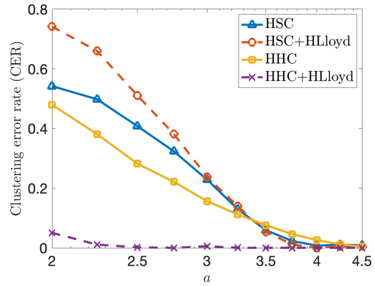
5.2 Real data analysis and real-data-inspired simulation studies.

Real data example: flight route network. Here, we consider 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 based on the number of flight routes.[1] This results in a $39 \times 50 \times 50$ tensor $\boldsymbol{\mathcal{Y}}$ with binary entries, where the first mode represent airlines, and the remaining two modes represent airports. The entries of the tensor $\boldsymbol{\mathcal{Y}}$ satisfies

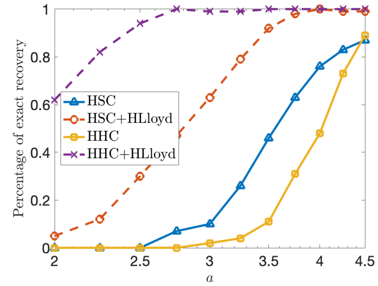
$$Y_{i,j,k} = \begin{cases} 1, & \text{if airline } i \text{ operates a flight route from airport } j \text{ to airport } k, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

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 $\boldsymbol{\mathcal{Y}}$, with results summarized in Tables 1 through

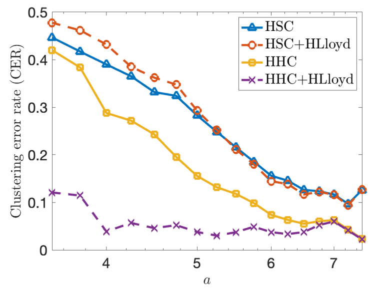
[1]The original database at `https://openflights.org/data.html#route`. Here, we use the processed data provided at `https://github.com/Rungang/HLloyd/blob/master/experiment/flight_route.RData`.



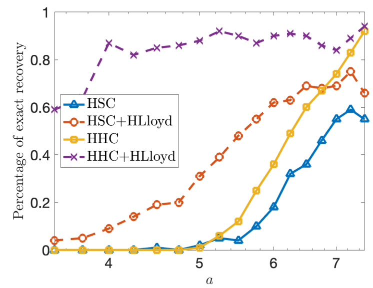
(a) $k = 3$, averaged CER



(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 (18).

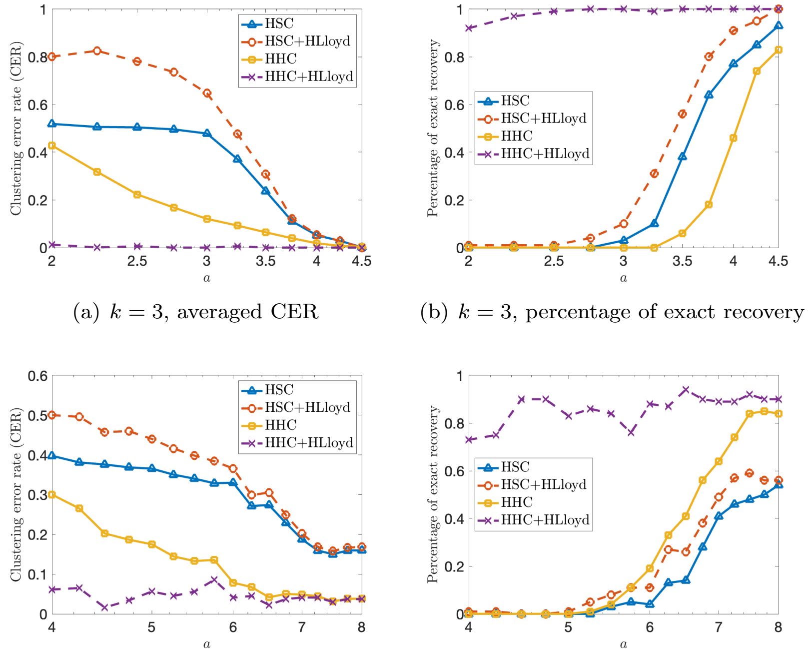
	Airlines
Cluster 1	CA, MU, CZ, HU, 3U, ZH (China)
Cluster 2	AA, UA, US (USA)
Cluster 3	AF, AZ, KL (Europe)
Cluster 4	BA, AY, IB (Europe), DL (USA)
Cluster 5	SU, AB, AI, AM, NH, AC, AS, FL, DE, ET, etc. (Mixture)

Table 1: Airline clustering results using HHC + HLloyd.

4[2]. 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 yields

[2]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 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 (18).

	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.

the centroid tensor $\widehat{\boldsymbol{\mathcal{S}}}^{\mathsf{HHC+HLloyd}} \in [0,1]^{5\times5\times5}$ (resp. HSC + HLloyd) gives us the following estimates: the centroid tensor $\widehat{\boldsymbol{\mathcal{S}}}^{\mathsf{HHC+HLloyd}} \in [0,1]^{5\times5\times5}$ and the label vector estimates $\widehat{\boldsymbol{z}}_i^{\mathsf{HHC+HLloyd}}$. In contrast, HSC + HLloyd provides $\widehat{\boldsymbol{\mathcal{S}}}^{\mathsf{HSC+HLloyd}} \in [0,1]^{5\times5\times5}$ 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 presented in Table 5 and Table 6, respectively.From the results, it's evident that under the model with 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.

	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.

6 Related work

This paper has close ties to various clustering problems. Among the commonly studied clustering models are stochastic block models (Holland et al., 1983; Rohe et al., 2011; Abbe et al., 2015; Lei and Rinaldo, 2015; Zhang and Zhou, 2016; Florescu and Perkins, 2016; Abbe, 2017; Gao et al., 2017; Amini and Levina, 2018; Cai et al., 2021) and (sub-)Gaussian mixture models (Lu and Zhou, 2016; Cai and Zhang, 2018; Chen and Yang, 2021; Ndaoud, 2022; Han et al., 2023). Spectral clustering has emerged as a powerful tool for obtaining clusters specifically under these two models, and has achieved significant success. Recently, it has been shown that spectral clustering can achieve optimal statistical guarantees (Abbe et al., 2020; Löffler et al., 2021; Abbe et al., 2022; Zhang and Zhou, 2022; Zhang, 2023). However, direct application of these methods and theories to the tensor block model might yield sub-optimal signal-to-noise ratio conditions or introduce unnecessary assumptions regarding the tensor's condition number.

Turning to the tensor block model, Wang and Zeng (2019) investigated the theoretical properties for the MLE estimator, which is computationally intractable. Chi et al. (2020) considered a convex procedure and derived its theoretical guarantees concerning the tensor estimation error. Moreover, under the i.i.d. noise setting Han et al. (2022a) proved 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 fail to provide computationally efficient and accurate estimators. Going beyond this model, Hu and Wang (2023) considered degree corrected tensor block models and Agterberg and Zhang (2022) studied tensor mixed membership blockmodels. However, the methods and theoretical results in these two papers either lean on i.i.d. noise assumptions or or necessitate the underlying tensor to be well-conditioned, both of which are superfluous in our setting. Besides the model considered in this paper, many tensor clustering problems have been proposed and extensively studied in the literature (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;

	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 $\boldsymbol{z}_i^{\star} = \widehat{\boldsymbol{z}}_i^{\mathsf{HSC+HLloyd}}$.

Xia et al., 2022; Zhou et al., 2022), which aims to estimate the true tensor or associated subspaces based on a noisy observation. 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 PCA (Cai et al., 2021, 2022a) and HeteroPCA (Zhang et al., 2022; Yan et al., 2021; Han et al., 2022b). However, different from the tensor PCA models, the tensor block models do not assume any conditions on the least singular value or on the singular gaps of the true tensor. Therefore, directly using these 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 exactly recover or achieve optimal mis-clustering rate 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., 2020; Chen et al., 2020, 2019b, 2021b; Cai et al., 2021; Chen et al., 2021c; Cai et al., 2022a; Abbe et al., 2022; Yan et al., 2021; Ling, 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 restriction on the singular value of the underlying tensor $\boldsymbol{\mathcal{S}}^{\star}$. Here, to demonstrate exact clustering results, we employ a novel strategy leveraging a new tool (Lemma 1).

7 Discussion

In this paper, we have studied the tensor clustering problem in the presence of heteroskedastic noise. To better deal with heteroskedastic noise, we have proposed a novel method, High-order HeteroClustering (HHC), which uses Thresholded Deflated-HeteroPCA to obtain subspace estimates and then performs approximate k-means to obtain clustering results. The proposed method can achieve exact clustering results under near-optimal signal-to-noise ratio conditions for polynomial-time algorithms. Empirically, we have studied the performance of HHC and HHC followed by the high-order Lloyd algorithm (HLloyd, Han et al. (2022a)). Both these two methods can achieve low clustering error rate, and HHC + HLloyd have been shown to achieve more accurate clustering results than existing methods in the literature.

Going beyond the problem considered in this paper, there are many future directions that are worth

Algorithm 4: High-order Lloyd Algorithm (HLloyd) (Han et al., 2022a)

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

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

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

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

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

$$\begin{aligned}
\left(\widehat{\boldsymbol{\mathcal{B}}}_1^{(t)}\right)_{j_1,i_2,i_3} &= \text{Average}\left(\left\{\boldsymbol{\mathcal{Y}}_{j_1,j_2,j_3} : \left(\widehat{\boldsymbol{z}}_\ell^{(t)}\right)_{j_\ell} = i_\ell, \ell = 2,3\right\}\right), &&\forall j_1\in[n_1], i_2\in[k_2], i_3\in[k_3],\\
\left(\widehat{\boldsymbol{\mathcal{B}}}_2^{(t)}\right)_{i_1,j_2,i_3} &= \text{Average}\left(\left\{\boldsymbol{\mathcal{Y}}_{j_1,j_2,j_3} : \left(\widehat{\boldsymbol{z}}_\ell^{(t)}\right)_{j_\ell} = i_\ell, \ell = 1,3\right\}\right), &&\forall i_1\in[k_1], j_2\in[n_2], i_3\in[k_3],\\
\left(\widehat{\boldsymbol{\mathcal{B}}}_3^{(t)}\right)_{i_1,i_2,j_3} &= \text{Average}\left(\left\{\boldsymbol{\mathcal{Y}}_{j_1,j_2,j_3} : \left(\widehat{\boldsymbol{z}}_\ell^{(t)}\right)_{j_\ell} = i_\ell, \ell = 1,2\right\}\right), &&\forall i_1\in[k_1], i_2\in[k_2], j_3\in[n_3].
\end{aligned}$$

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

$$\left(\widehat{\boldsymbol{z}}_i^{(t+1)}\right)_j \in \arg\min_{\ell\in[k_i]} \left\| \left(\mathcal{M}_i\left(\widehat{\boldsymbol{\mathcal{B}}}_i^{(t)}\right)\right)_{j,:} - \left(\mathcal{M}_i\left(\widehat{\boldsymbol{\mathcal{S}}}^{(t)}\right)\right)_{\ell,:}\right\|_2, \qquad \forall i\in[3], j\in[n_i]. \tag{20}$$

Output: label vector estimates $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{S}}^{(T-1)}, \widehat{\boldsymbol{z}}_1 = \widehat{\boldsymbol{z}}_1^{(T)}, \widehat{\boldsymbol{z}}_2 = \widehat{\boldsymbol{z}}_2^{(T)}, \widehat{\boldsymbol{z}}_3 = \widehat{\boldsymbol{z}}_2^{(T)}$.

investigating. For instance, we mainly focus on the exact community recovery in this paper. It remains unknown if our algorithm can achieve optimal mean squared error. In addition, our signal-to-noise ratio condition might be sub-optimal if the clusters are highly-unbalanced, i.e., the cluster size parameter β is very small. It would be interesting to design a more reliable algorithm that can achieve satisfying theoretical guarantees under a weaker signal-to-noise ratio condition in this setting.

Acknowledgements

Y. Chen is supported in part by the Alfred P. Sloan Research Fellowship, and the NSF grants CCF-1907661, DMS-2014279, IIS-2218713 and IIS-2218773.

A Procedure of High-order Lloyd Algorithm (HLloyd)

We introduce the procedure of High-order Lloyd Algorithm (HLloyd) (Han et al., 2022a) in Algorithm 4.

B Proof of Theorem 1

We will prove the following theorem, which is a more general version of Theorem 5.

Theorem 4. *Suppose that Assumption (1) holds and for all $i\in[3]$,*

$$n_1 n_2 n_3 \gtrsim k^4 n^2, \tag{21a}$$

$$n_i \ge \frac{c_1 k^4}{\beta^2}, \tag{21b}$$

$$k_i \gtrsim k_{-i}, \tag{21c}$$

$$c_\tau k_i^2 \left(n_1 n_2 n_3\right)^{1/2} \log^2 n \le \tau_i / \omega_{\mathsf{max}}^2 \le C_\tau k_i^2 \left(n_1 n_2 n_3\right)^{1/2} \log^2 n \tag{21d}$$

$$\Delta_{\mathsf{min}} / \sigma_{\mathsf{max}} \ge C_1 \sqrt{M} \left(\frac{k^{9/2}}{\beta^{5/2}} \left(n_1 n_2 n_3\right)^{-1/4} \log n + \frac{k^9}{\beta^5} \left(n_1 n_2 n_3 / n\right)^{-1/2} \sqrt{\log n} \right), \tag{21e}$$

where C_1, c_1, C_τ and c_τ are some large enough constants. Then by choosing the numbers of iterations

$$t_{i,j} \ge \log \left(C \frac{k^3}{\beta^3} \frac{\sigma_{i, r_{i,j-1}+1}^{\star 2}}{\sigma_{i, r_{i,j}+1}^{\star 2}} \right), \qquad 1 \le j \le j_{\mathsf{max}}^i - 1, \tag{22a}$$

$$t_{i, j_{\mathsf{max}}^i} \ge \log \left(C n^3 \frac{\sigma_{i, r_{i, j_{\mathsf{max}}^i - 1}+1}^{\star 2}}{\omega_{\mathsf{max}}^2} \right), \tag{22b}$$

for all $i \in [3]$, with probability exceeding $1 - O(n^{-10})$, one has

$$\widehat{\boldsymbol{z}}_i = \phi_i \left(\boldsymbol{z}_i^\star \right), \qquad \forall i \in [3]$$

for some permutation $\phi_i : [k_i] \to [k_i]$.

B.1 Several key results under the matrix setting

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

$$\boldsymbol{Y} = \boldsymbol{X}^\star + \boldsymbol{E} \in \mathbb{R}^{m_1 \times m_2}, \tag{23}$$

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 has the following SVD decomposition:

$$\boldsymbol{X}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{V}^{\star\top} = \sum_{i=1}^r \sigma_i^\star \boldsymbol{u}_i^\star \boldsymbol{v}_i^{\star\top} \tag{24}$$

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

$$\text{(Incoherence parameter)} \qquad \mu = \mu\left(\boldsymbol{X}^\star\right) := \max\left\{ \frac{m_1}{r} \max_{i \in [m_1]} \left\| \boldsymbol{U}_{i,:}^\star \right\|_2^2, \frac{m_2}{r} \max_{j \in [m_2]} \left\| \boldsymbol{V}_{j,:}^\star \right\|_2^2 \right\}. \tag{25}$$

For notational convenience, we also define

$$m = \max\{m_1, m_2\} \qquad \text{and} \qquad \sigma_{r+1}^\star = 0. \tag{26}$$

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 independent and satisfy $\mathbb{E}[E_{i,j}] = 0$;*
2. *$\mathbb{P}\left(|E_{i,j}| > B\right) \le m^{-12}$, where B is some quantity satisfying*

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

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

$$\boldsymbol{M} = \left(\boldsymbol{U}^\star \boldsymbol{\Sigma}^\star + \boldsymbol{E} \boldsymbol{V}^\star\right) \left(\boldsymbol{U}^\star \boldsymbol{\Sigma}^\star + \boldsymbol{E} \boldsymbol{V}^\star\right)^\top \tag{27}$$

and

$$\boldsymbol{M}^{\mathsf{oracle}} = \boldsymbol{M} + \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star \boldsymbol{V}^{\star\top} \boldsymbol{E}^\top\right) = \mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{Y}\boldsymbol{Y}^\top\right) + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{M}\right). \tag{28}$$

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 close if there is a significant gap between $\sigma_{r'}^\star$ and $\sigma_{r'+1}^\star$, and its proof is deferred to Section C.

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

$$\sigma_1^\star/\omega_{\mathsf{max}} \geq 2C_0 r\left[(m_1 m_2)^{1/4} + r m_1^{1/2}\right]\log m, \tag{29a}$$

$$\mu \leq c_0 \frac{m_1}{r^3}, \tag{29b}$$

for some sufficiently large (resp. small) constant $C_0 > 0$ (resp. $c_0 > 0$).

(a) Let

$$\mathcal{A} = \left\{ j : 1 \leq j \leq r, \sigma_j^\star \geq \frac{4r}{4r-1}\sigma_{j+1}^\star \vee C_0 r\left[(m_1 m_2)^{1/4} + r m_1^{1/2}\right]\omega_{\mathsf{max}} \log m \right\}. \tag{30}$$

Then $\mathcal{A}$ is non-empty.

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

$$\left\| \boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}} \boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{:,1:r'} \widetilde{\boldsymbol{U}}_{:,1:r'}^{\top} \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma_{r'}^\star} + \frac{r^2\sqrt{m_1 m_2}\omega_{\mathsf{max}}^2 \log^2 m}{\sigma_{r'}^{\star 2}} \right), \tag{31a}$$

$$\left\| \boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}} \boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top} - \boldsymbol{U}_{:,1:r'}^{\star} \boldsymbol{U}_{:,1:r'}^{\star\top} \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(\frac{r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma_{r'}^\star} + \frac{r^2\sqrt{m_1 m_2}\omega_{\mathsf{max}}^2 \log^2 m}{\sigma_{r'}^{\star 2}} \right). \tag{31b}$$

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

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

$$c_\tau r^2\left[(m_1 m_2)^{1/2} + r^2 m_1\right]\log^2 m \leq \tau/\omega_{\mathsf{max}}^2 \leq C_\tau r^2\left[(m_1 m_2)^{1/2} + r^2 m_1\right]\log^2 m, \tag{32a}$$

$$\sigma_1^\star/\omega_{\mathsf{max}} \geq C_1 r\left[(m_1 m_2)^{1/4} + r m_1^{1/2}\right]\log m, \tag{32b}$$

$$\mu \leq c_1 \frac{m_1}{r^3}, \tag{32c}$$

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

$$t_k > \log\left(C \frac{\sigma_{r_{k-1}}^{\star 2}}{\sigma_{r_k}^{\star 2}} \right), \qquad 1 \leq k < k_{\mathsf{max}}, \tag{33a}$$

$$t_{k_{\mathsf{max}}} > \log\left(C \frac{\sigma_{r_{k_{\mathsf{max}}-1}+1}^{\star 2}}{\omega_{\mathsf{max}}^2} \right) \tag{33b}$$

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

$$\left\| \boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}_{:,1:r_{k_{\mathsf{max}}}}^\star \boldsymbol{U}_{:,1:r_{k_{\mathsf{max}}}}^{\star\top} \right\| \lesssim \sqrt{\frac{\mu r^3}{m_1}}, \tag{34a}$$

$$\left\| \left(\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}_{:,1:r_{k_{\mathsf{max}}}}^\star \boldsymbol{U}_{:,1:r_{k_{\mathsf{max}}}}^{\star\top}\right) \boldsymbol{X}^\star \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m + r(m_1 m_2)^{1/4}\omega_{\mathsf{max}}\log m \right), \tag{34b}$$

$$\left\| \left(\boldsymbol{I}_{m_1} - \boldsymbol{U}\boldsymbol{U}^\top\right) \boldsymbol{X}^\star \right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}} \left(r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m + r(m_1 m_2)^{1/4}\omega_{\mathsf{max}}\log m \right),. \tag{34c}$$

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 in the position to prove Theorem 4. The proof consists of four steps.

B.2 Main steps for proving Theorem 4

Step 1: verifying (32a) - (32c). To apply Theorem 6, one needs to verify the conditions (32a) - (32c) for $\mathcal{M}_1(\boldsymbol{\mathcal{Y}}) = \mathcal{M}_1(\boldsymbol{\mathcal{X}}) + \mathcal{M}_i(\boldsymbol{\mathcal{E}})$ with dimensions $m_1 = n_i$, $m_2 = n_{-i}$ and the rank $r = k_i$.

Step 1.1: verifying (32a). Noting that $n_1n_2n_3 \geq k^4n^2$, we have $(n_1n_2n_3)^{1/2} + k_i^2 n_i \asymp (n_1n_2n_3)^{1/2}$. Then we know from (21d) that (32a) is valid.

Step 1.2: verifying (32c). For notational convenience, we let

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

and

$$\overline{\boldsymbol{\mathcal{S}}}^{\star} = \boldsymbol{\mathcal{S}}^{\star} \times_1 \left(\boldsymbol{M}_1^{\star\top}\boldsymbol{M}_1^{\star}\right)^{1/2} \times_2 \left(\boldsymbol{M}_2^{\star\top}\boldsymbol{M}_2^{\star}\right)^{1/2} \times_3 \left(\boldsymbol{M}_3^{\star\top}\boldsymbol{M}_3^{\star}\right)^{1/2}.$$

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

$$\mathcal{P}_{\boldsymbol{U}} = \boldsymbol{U}\boldsymbol{U}^{\top}. \tag{35}$$

Recognizing that

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

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

$$\boldsymbol{U}_{\boldsymbol{X}_i^{\star}} = \overline{\boldsymbol{M}}_i^{\star}\boldsymbol{A}_i \qquad (\text{resp. } \boldsymbol{V}_{\boldsymbol{X}_i^{\star}} = \big(\overline{\boldsymbol{M}}_{i+2}^{\star}\otimes\overline{\boldsymbol{M}}_{i+1}^{\star}\big)\boldsymbol{B}_i) \tag{36}$$

for some $\boldsymbol{A}_i \in \mathcal{O}^{k_i,k_i}$ and $\boldsymbol{B} \in \mathcal{O}^{k_{-i},k_i}$. Then we immediately know that

$$\left\|\boldsymbol{U}_{\boldsymbol{X}_i^{\star}}\right\|_{2,\infty} \leq \left\|\overline{\boldsymbol{M}}_i^{\star}\right\|_{2,\infty} \leq \left\|\boldsymbol{M}_i^{\star}\right\|_{2,\infty}\left\|\sigma_{k_i}\left(\boldsymbol{M}_i^{\star\top}\boldsymbol{M}_i^{\star}\right)\right\|^{-1/2} \leq 1\cdot\sqrt{\frac{k_i}{\beta n_i}}. \tag{37}$$

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

$$\left(\boldsymbol{M}_i^{\star\top}\boldsymbol{M}_i^{\star}\right)_{\ell,\ell} = \left|\left\{j \in [n_i] : \left(\boldsymbol{z}_i^{\star}\right)_j = \ell\right\}\right| \geq \beta n_i/k_i. \tag{38}$$

Similarly, one can bound $\|\boldsymbol{V}_{\boldsymbol{X}_i^{\star}}\|_{2,\infty}$ as follows:

$$\left\|\boldsymbol{V}_{\boldsymbol{X}_i^{\star}}\right\|_{2,\infty} \leq \left\|\overline{\boldsymbol{M}}_{i+1}^{\star}\right\|_{2,\infty}\left\|\overline{\boldsymbol{M}}_{i+2}^{\star}\right\|_{2,\infty} \leq \sqrt{\frac{k_{i+1}}{\beta n_{i+1}}}\sqrt{\frac{k_{i+2}}{\beta n_{i+2}}} = \sqrt{\frac{\left(\frac{k_{-i}}{\beta^2 k_i}\right)k_i}{n_{-i}}}. \tag{39}$$

Then we have

$$\mu_i = \max\left\{\frac{1}{\beta}, \frac{k_{-i}}{\beta^2 k_i}\right\} \overset{(21b)}{\leq} c_1\frac{n_i}{k_i^3}. \tag{40}$$

Step 1.3: verifying (32b). Next, we validate Condition (32b). Note that for any $1 \leq j_1 \neq j_2 \leq k_i$,

$$\begin{aligned}
\Delta_i^2 &\leq \left\|\mathcal{M}_i\left(\mathcal{S}^{\star}\right)_{j_1,:} - \mathcal{M}_i\left(\mathcal{S}^{\star}\right)_{j_2,:}\right\|_2^2 \\
&= \left\|\left(\boldsymbol{e}_{j_1} - \boldsymbol{e}_{j_2}\right)^{\top}\mathcal{M}_i\left(\mathcal{S}^{\star}\right)\right\|_2^2 \\
&\leq \left\|\mathcal{M}_i\left(\mathcal{S}^{\star}\right)\right\|^2\left\|\boldsymbol{e}_{j_1} - \boldsymbol{e}_{j_2}\right\|_2^2
\end{aligned}$$

$$= 2\left\|\mathcal{M}_i\left(\mathcal{S}^\star\right)\right\|^2, \tag{41}$$

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 (38), we know that

$$\sigma_{k_i}\left(\boldsymbol{M}_i\right) \geq \sqrt{\frac{\beta n_i}{k_i}} \tag{42}$$

and consequently

$$\left\|\boldsymbol{X}_i^\star\right\| = \left\|\boldsymbol{M}_i^\star\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}^\star\right)\left(\boldsymbol{M}_{i+2}^\star\otimes\boldsymbol{M}_{i+1}^\star\right)^\top\right\| \geq \left\|\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}^\star\right)\right\|\prod_{i=1}^3\sigma_{k_i}\left(\boldsymbol{M}_i^\star\right) \geq \sqrt{\frac{\beta^3 n_1n_2n_3}{k_1k_2k_3}}\left\|\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}^\star\right)\right\|. \tag{43}$$

Combining (41) and (43) and the assumption $\Delta_i/\sigma_{\mathsf{max}} \gg k^{5/2}\left(n_1n_2n_3\right)^{-1/4}\log m/\beta^{3/2}$, one has

$$\left\|\boldsymbol{X}_i^\star\right\| \geq \sqrt{\frac{\beta^3 n_1n_2n_3}{2k_1k_2k_3}}\Delta_i \gg k_i\left(n_1n_2n_3\right)^{1/4}\log n \asymp k_i\left[\left(n_1n_2n_3\right)^{1/4}+k_in_i^{1/2}\right]\log n, \tag{44}$$

and thus condition (32b) holds. Here, the last inequality holds since $n_1n_2n_3 \geq k^4n^2$.

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

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

Then the triangle inequality and the fact $\|\mathcal{P}_{\boldsymbol{U}_i}\| = 1$ give us the following upper bound:

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\mathcal{X}}}-\boldsymbol{\mathcal{X}}^\star\right\|_{\mathrm{F}}^2 &= \|\boldsymbol{\mathcal{Y}}\times_1\mathcal{P}_{\boldsymbol{U}_1}\times_2\mathcal{P}_{\boldsymbol{U}_2}\times_3\mathcal{P}_{\boldsymbol{U}_3}-\boldsymbol{\mathcal{X}}^\star\|_{\mathrm{F}}^2\\
&\leq 2\left(\|\boldsymbol{\mathcal{X}}^\star\times_1\mathcal{P}_{\boldsymbol{U}_1}\times_2\mathcal{P}_{\boldsymbol{U}_2}\times_3\mathcal{P}_{\boldsymbol{U}_3}-\boldsymbol{\mathcal{X}}^\star\|_{\mathrm{F}}^2+\|\boldsymbol{\mathcal{E}}\times_1\mathcal{P}_{\boldsymbol{U}_1}\times_2\mathcal{P}_{\boldsymbol{U}_2}\times_3\mathcal{P}_{\boldsymbol{U}_3}\|_{\mathrm{F}}^2\right)\\
&\leq 6\big(\|\boldsymbol{\mathcal{X}}^\star\times_1\left(\boldsymbol{I}_{n_1}-\mathcal{P}_{\boldsymbol{U}_1}\right)\times_2\mathcal{P}_{\boldsymbol{U}_2}\times_3\mathcal{P}_{\boldsymbol{U}_3}\|_{\mathrm{F}}^2+\|\boldsymbol{\mathcal{X}}^\star\times_2\left(\boldsymbol{I}_{n_2}-\mathcal{P}_{\boldsymbol{U}_2}\right)\times_3\mathcal{P}_{\boldsymbol{U}_3}\|_{\mathrm{F}}^2\\
&\qquad+\|\boldsymbol{\mathcal{X}}^\star\times_3\left(\boldsymbol{I}_{n_3}-\mathcal{P}_{\boldsymbol{U}_3}\right)\|_{\mathrm{F}}^2\big)+2\|\boldsymbol{\mathcal{E}}\times_1\mathcal{P}_{\boldsymbol{U}_1}\times_2\mathcal{P}_{\boldsymbol{U}_2}\times_3\mathcal{P}_{\boldsymbol{U}_3}\|_{\mathrm{F}}^2\\
&\leq 6\left(\left\|\left(\boldsymbol{I}_{n_1}-\boldsymbol{U}_1\boldsymbol{U}_1^\top\right)\boldsymbol{X}_1^\star\right\|_{\mathrm{F}}^2+\left\|\left(\boldsymbol{I}_{n_2}-\boldsymbol{U}_2\boldsymbol{U}_2^\top\right)\boldsymbol{X}_2^\star\right\|_{\mathrm{F}}^2+\left\|\left(\boldsymbol{I}_{n_3}-\boldsymbol{U}_3\boldsymbol{U}_3^\top\right)\boldsymbol{X}_3^\star\right\|_{\mathrm{F}}^2\right)\\
&\quad+2\|\boldsymbol{\mathcal{E}}\times_1\mathcal{P}_{\boldsymbol{U}_1}\times_2\mathcal{P}_{\boldsymbol{U}_2}\times_3\mathcal{P}_{\boldsymbol{U}_3}\|_{\mathrm{F}}^2.
\end{aligned} \tag{45}$$

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

$$\sigma_\ell\left(\mathcal{M}_i\left(\boldsymbol{\mathcal{X}}^\star\right)\right) = \sigma_\ell\left(\boldsymbol{M}_i^\star\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}^\star\right)\left(\boldsymbol{M}_{i+2}^\star\otimes\boldsymbol{M}_{i+1}^\star\right)^\top\right) \geq \prod_{i=1}^3\sigma_{k_i}\left(\boldsymbol{M}_i^\star\right)\sigma_\ell\left(\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}^\star\right)\right) \overset{(42)}{\geq} \sqrt{\frac{\beta^3n_1n_2n_3}{k_1k_2k_3}}\sigma_\ell\left(\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}^\star\right)\right)$$

and

$$\sigma_\ell\left(\mathcal{M}_i\left(\boldsymbol{\mathcal{X}}^\star\right)\right) \leq \prod_{i=1}^3\left\|\boldsymbol{M}_i^\star\right\|\sigma_\ell\left(\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}^\star\right)\right) \leq \sqrt{n_1n_2n_3}\sigma_\ell\left(\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}^\star\right)\right).$$

In view of Theorem 6, we know that by choosing the numbers of iterations as in (22a) and (22b), with probability at least $1-O(n^{-10})$,

$$\begin{aligned}
\left\|\left(\boldsymbol{I}_{n_1}-\boldsymbol{U}_1\boldsymbol{U}_1^\top\right)\boldsymbol{X}_1^\star\right\|_{\mathrm{F}}^2 &\leq n_1\left\|\left(\boldsymbol{I}_{n_1}-\boldsymbol{U}_1\boldsymbol{U}_1^\top\right)\boldsymbol{X}_1^\star\right\|_{2,\infty}^2\\
&\lesssim n_1\cdot\frac{\mu_1k_1^3}{n_1}\left(k_1^2\sqrt{n_1}\omega_{\mathsf{max}}\log n+k_1\left(n_1n_2n_3\right)^{1/4}\omega_{\mathsf{max}}\log n\right)^2\\
&\overset{\text{(21a) and (40)}}{\lesssim} \frac{k^6}{\beta^2}\left(n_1n_2n_3\right)^{1/2}\omega_{\mathsf{max}}^2\log^2 n.
\end{aligned} \tag{46}$$

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

$$\left\| \left(\boldsymbol{I}_{n_2} - \boldsymbol{U}_2 \boldsymbol{U}_2^\top \right) \boldsymbol{X}_2^\star \right\|_{\mathrm{F}}^2 \lesssim \frac{k^6}{\beta^2} \left(n_1 n_2 n_3 \right)^{1/2} \omega_{\mathsf{max}}^2 \log^2 n, \tag{47a}$$

$$\left\| \left(\boldsymbol{I}_{n_2} - \boldsymbol{U}_2 \boldsymbol{U}_2^\top \right) \boldsymbol{X}_2^\star \right\|_{\mathrm{F}}^2 \lesssim \frac{k^6}{\beta^2} \left(n_1 n_2 n_3 \right)^{1/2} \omega_{\mathsf{max}}^2 \log^2 n. \tag{47b}$$

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

$$\|\boldsymbol{U}_i\|_{2,\infty} = \left\| \boldsymbol{U}_i \boldsymbol{U}_i^\top \right\|_{2,\infty} \le \left\| \boldsymbol{U}_i \boldsymbol{U}_i^\top - \boldsymbol{U}^\star \boldsymbol{U}^\star \right\|_{2,\infty} + \|\boldsymbol{U}^\star\|_{2,\infty} \le 2\sqrt{\frac{\mu_i k_i^3}{n_i}}, \qquad \forall i \in [3]. \tag{48}$$

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

$$\begin{aligned} \|\boldsymbol{\mathcal{E}} \times_1 \mathcal{P}_{\boldsymbol{U}_1} \times_2 \mathcal{P}_{\boldsymbol{U}_2} \times_3 \mathcal{P}_{\boldsymbol{U}_3}\|_{\mathrm{F}}^2 &\le k_1 \left\| \boldsymbol{U}_1^\top \mathcal{M}_1 \left(\boldsymbol{\mathcal{E}} \right) \left(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2 \right) \right\|^2 \\ &\lesssim kn \left(\mu_1 k_1^2 \right) \left(\mu_2 k_2^2 \right) \left(\mu_3 k_3^2 \right) k^3 \omega_{\mathsf{max}}^2 \log n \\ &\overset{(40)}{\lesssim} \frac{k^{13}}{\beta^6} n \omega_{\mathsf{max}}^2 \log n. \end{aligned} \tag{49}$$

Putting (45) - (49) together, we obtain

$$\left\| \widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^\star \right\|_{\mathrm{F}}^2 \lesssim \frac{k^6}{\beta^2} \left(n_1 n_2 n_3 \right)^{1/2} \omega_{\mathsf{max}}^2 \log^2 n + \frac{k^{13}}{\beta^6} n \omega_{\mathsf{max}}^2 \log n \tag{50}$$

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

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

$$\widehat{\boldsymbol{\theta}}_\ell^{(i)} = \left(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \right) \widehat{\boldsymbol{b}}_\ell^{(i)} \in \mathbb{R}^{n_{-i}}, \qquad \forall i \in [3], \ell \in [k_i], \tag{51}$$

and also define

$$\boldsymbol{\theta}_\ell^{(i)\star} = \left(\mathcal{M}_i \left(\boldsymbol{\mathcal{X}}^\star \right) \right)_{j,:}^\top \in \mathbb{R}^{n_{-i}}, \qquad \forall i \in [3], \ell \in [k_i]. \tag{52}$$

Here, $j \in [n_i]$ is any index satisfying $\left(\boldsymbol{z}_i^\star \right)_j = \ell$ and the $\widehat{\boldsymbol{b}}_\ell^{(i)}$'s are the center estimates satisfying (11a). 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

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

Then one can show that

$$\begin{aligned} \sum_{j=1}^{n_i} \left\| \big(\mathcal{M}_i \big(\widehat{\boldsymbol{\mathcal{X}}} \big) \big)_{j:}^\top - \widehat{\boldsymbol{\theta}}_{(\widehat{\boldsymbol{z}}_i)_j}^{(i)} \right\|_2^2 &= \sum_{j=1}^{n_i} \left\| \big(\boldsymbol{U}_{i+2} \otimes \boldsymbol{U}_{i+1} \big) \left(\big(\widehat{\boldsymbol{B}}_i \big)_{j:}^\top - \widehat{\boldsymbol{b}}_{(\widehat{\boldsymbol{z}}_i)_j}^{(i)} \right) \right\|_2^2 \\ &= \sum_{j=1}^{n_i} \left\| \big(\widehat{\boldsymbol{B}}_i \big)_{j:}^\top - \widehat{\boldsymbol{b}}_{(\widehat{\boldsymbol{z}}_i)_j}^{(i)} \right\|_2^2 \\ &\overset{(11a)}{\le} M \min_{\substack{\boldsymbol{b}_1, \dots, \boldsymbol{b}_{k_i} \in \mathbb{R}^{r_1 r_2 r_3 / r_i} \\ \boldsymbol{z}_i \in [k_i]^{n_i}}} \sum_{j=1}^{n_i} \left\| \big(\widehat{\boldsymbol{B}}_i \big)_{j:}^\top - \boldsymbol{b}_{(\boldsymbol{z}_i)_j} \right\|_2^2 \\ &= M \min_{\substack{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{k_i} \in \mathbb{R}^{n_{-i}} \\ \boldsymbol{z}_i \in [k_i]^{n_i}}} \sum_{j=1}^{n_i} \left\| \big(\mathcal{M}_i \big(\widehat{\boldsymbol{\mathcal{X}}} \big) \big)_{j:}^\top - \boldsymbol{\theta}_{(\boldsymbol{z}_i)_j} \right\|_2^2 \\ &\le M \sum_{j=1}^{n_i} \left\| \big(\mathcal{M}_i \big(\widehat{\boldsymbol{\mathcal{X}}} \big) \big)_{j:}^\top - \boldsymbol{\theta}_{(\boldsymbol{z}_i^\star)_j}^{(i)\star} \right\|_2^2 \end{aligned}$$

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

$$= M \left\| \widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^{\star} \right\|_{\mathrm{F}}^2. \tag{53}$$

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

$$\begin{aligned} \sum_{j=1}^{n_i} \left\| \widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_j} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^{\star})_j} \right\|_2^2 &\le 2 \left(\sum_{j=1}^{n_i} \left\| \left(\mathcal{M}_i(\widehat{\boldsymbol{\mathcal{X}}})\right)_{j:}^{\top} - \widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_j} \right\|_2^2 + \sum_{j=1}^{n_i} \left\| \left(\mathcal{M}_i(\widehat{\boldsymbol{\mathcal{X}}})\right)_{j:}^{\top} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^{\star})_j} \right\|_2^2 \right) \\ &\le 2 \left(M \left\| \widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^{\star} \right\|_{\mathrm{F}}^2 + \left\| \widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^{\star} \right\|_{\mathrm{F}}^2 \right) \\ &\le 4M \left\| \widehat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}^{\star} \right\|_{\mathrm{F}}^2. \end{aligned} \tag{54}$$

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

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

we have

$$\sigma_{k_i}\left(\boldsymbol{M}_i^{\star}\right) \ge \sqrt{\beta n_i / k_i}, \qquad \forall i \in [3]. \tag{55}$$

As a consequence, for all $\ell_1 \neq \ell_2 \in [k_i]$, we have

$$\begin{aligned} \left\| \boldsymbol{\theta}^{(i)\star}_{\ell_1} - \boldsymbol{\theta}^{(i)\star}_{\ell_2} \right\| &= \left\| \left(\mathcal{M}_i\left(\boldsymbol{\mathcal{X}}\right)\right)_{j_1,:} - \left(\mathcal{M}_i\left(\boldsymbol{\mathcal{X}}\right)\right)_{j_2,:} \right\| \\ &= \left\| \left(\left(\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}\right)\right)_{\ell_1,:} - \left(\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}\right)\right)_{\ell_2,:} \right) \left(\boldsymbol{M}_{i+2}^{\star} \otimes \boldsymbol{M}_{i+1}^{\star} \right) \right\| \\ &\ge \left\| \left(\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}\right)\right)_{\ell_1,:} - \left(\mathcal{M}_i\left(\boldsymbol{\mathcal{S}}\right)\right)_{\ell_2,:} \right\| \sigma_{k_{i+1}}\left(\boldsymbol{M}_{i+1}^{\star}\right) \sigma_{k_{i+2}}\left(\boldsymbol{M}_{i+2}^{\star}\right) \\ &\ge \beta \sqrt{\frac{n_{-i}}{k_{-i}}} \Delta_i, \end{aligned} \tag{56}$$

where j_1 (resp. j_2) is any index such that $(\boldsymbol{z}_i^{\star})_{j_1} = \ell_1$ (resp. $(\boldsymbol{z}_i^{\star})_{j_2} = \ell_2$). We define

$$\mathcal{S}_i := \left\{ j \in [n_i] : \left\| \widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_j} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^{\star})_j} \right\|_2 \ge \frac{\beta}{2} \sqrt{\frac{n_{-i}}{k_{-i}}} \Delta_i \right\}. \tag{57}$$

In view of (50) and (54), with probability exceeding $1 - O(n^{-10})$,

$$\begin{aligned} |\mathcal{S}_i| &\le \frac{\sum_{j=1}^{n_i} \left\| \widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_j} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^{\star})_j} \right\|_2^2}{\left(\frac{\beta}{2} \sqrt{\frac{n_{-i}}{k_{-i}}} \Delta_i \right)^2} \\ &\le \frac{4M \cdot C_6 \left(\frac{k^6}{\beta^2} \left(n_1 n_2 n_3\right)^{1/2} \omega_{\mathsf{max}}^2 \log^2 n + \frac{k^{13}}{\beta^6} n \omega_{\mathsf{max}}^2 \log n \right)}{\left(\frac{\beta}{2} \sqrt{\frac{n_{-i}}{k_{-i}}} \Delta_i \right)^2} \\ &\le \frac{\beta}{2} \frac{n_i}{k_i}, \end{aligned} \tag{58}$$

provided that $\Delta_i \ge C_1 \sqrt{M} \left(\frac{k^{9/2}}{\beta^{5/2}} \left(n_1 n_2 n_3\right)^{-1/4} \omega_{\mathsf{max}} \log n + \frac{k^8}{\beta^{9/2}} \left(n_1 n_2 n_3 / n\right)^{-1/2} \omega_{\mathsf{max}} \sqrt{\log n} \right)$. For each $i \in [3], \ell \in [k_i]$, let $\mathcal{N}_{i,\ell}$ denote the following set:

$$\mathcal{N}_{i,\ell} := \left\{ j \in [n_i] : \left(\boldsymbol{z}_i^{\star}\right)_j = \ell, j \in \mathcal{S}^c \right\}. \tag{59}$$

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:

$$|\mathcal{N}_{i,\ell}| \geq \left|\left\{j \in [n_i] : (\boldsymbol{z}_i^\star)_j = \ell\right\}\right| - |\mathcal{S}_i| \overset{(8) \text{ and } (58)}{\geq} \beta \frac{n_i}{k_i} - \frac{\beta}{2}\frac{n_i}{k_i} = \frac{\beta}{2}\frac{n_i}{k_i} > 0. \tag{60}$$

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

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_{j_1}} - \widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_{j_2}}\right\|_2 &\geq \left\|\boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^\star)_{j_1}} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^\star)_{j_2}}\right\|_2 - \left\|\widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_{j_1}} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^\star)_{j_1}}\right\|_2 - \left\|\widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_{j_2}} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^\star)_{j_2}}\right\|_2 \\
&= \left\|\boldsymbol{\theta}^{(i)\star}_{\ell_1} - \boldsymbol{\theta}^{(i)\star}_{\ell_2}\right\|_2 - \left\|\widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_{j_1}} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^\star)_{j_1}}\right\|_2 - \left\|\widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_{j_2}} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^\star)_{j_2}}\right\|_2 \\
&\overset{(56)}{>} \beta\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i - \frac{\beta}{2}\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i - \frac{\beta}{2}\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i = 0,
\end{aligned}$$

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

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

$$(\widehat{\boldsymbol{z}}_i)_j = \phi_i\big((\boldsymbol{z}_i^\star)_j\big), \qquad \forall j \in \mathcal{S}_i^c, i \in [3]. \tag{61}$$

We define the misclassification rate

$$\ell\left(\boldsymbol{z}_i^\star, \widehat{\boldsymbol{z}}_i\right) = \inf_{\phi \in \Phi_i} \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\big\{(\widehat{\boldsymbol{z}}_i)_j \neq \phi_i\big((\boldsymbol{z}_i^\star)_j\big)\big\},$$

where Φ_i is set consisting of all permutations $\phi : [k_i] \to [k_i]$. In view of (58), with probability at least $1 - O(n^{-10})$, for all $i \in [3]$, one has

$$\begin{aligned}
\ell\left(\widehat{\boldsymbol{z}}_i, \boldsymbol{z}_i^\star\right) &\leq \frac{1}{n_i}\left|\left\{j \in [n_i] : (\widehat{\boldsymbol{z}}_i)_j \neq \phi_i\big((\boldsymbol{z}_i^\star)_j\big)\right\}\right| \leq \frac{1}{n_i}|\mathcal{S}_i| \\
&\leq \frac{4M \cdot C_6\left(\frac{k^6}{\beta^2}(n_1n_2n_3)^{1/2}\omega_{\mathsf{max}}^2 \log^2 n + \frac{k^{13}}{\beta^6} n\omega_{\mathsf{max}}^2 \log n\right)}{n_i\left(\frac{\beta}{2}\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i\right)^2}
\end{aligned} \tag{62}$$

and for all $\ell \in [k_i]$,

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\theta}}^{(i)}_{\phi_i(\ell)} - \boldsymbol{\theta}^{(i)\star}_\ell\right\|_2^2 &\leq \frac{\sum_{j=1}^{n_i}\left\|\widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_j} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^\star)_j}\right\|_2^2}{\left|\left\{j \in [n_i] : (\widehat{\boldsymbol{z}}_i)_j = \phi_i(\ell), (\boldsymbol{z}_i^\star)_j = \ell\right\}\right|} \\
&\leq \frac{\sum_{j=1}^{n_i}\left\|\widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_j} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^\star)_j}\right\|_2^2}{\left|\left\{j \in [n_i] : (\boldsymbol{z}_i^\star)_j = \ell, j \in \mathcal{S}^c\right\}\right|} \\
&= \frac{\sum_{j=1}^{n_i}\left\|\widehat{\boldsymbol{\theta}}^{(i)}_{(\widehat{\boldsymbol{z}}_i)_j} - \boldsymbol{\theta}^{(i)\star}_{(\boldsymbol{z}_i^\star)_j}\right\|_2^2}{|\mathcal{N}_{i,\ell}|} \\
&\leq \frac{4M \cdot C\left(\frac{k^6}{\beta^2}(n_1n_2n_3)^{1/2}\omega_{\mathsf{max}}^2 \log^2 n + \frac{k^{13}}{\beta^6} n\omega_{\mathsf{max}}^2 \log n\right)}{\frac{\beta n_i}{2k_i}} \\
&\leq \frac{C_7 M\left(\frac{k^7}{\beta^3}(n_1n_2n_3)^{1/2}\omega_{\mathsf{max}}^2 \log^2 n + \frac{k^{14}}{\beta^7} n\omega_{\mathsf{max}}^2 \log n\right)}{n_i},
\end{aligned} \tag{63}$$

provided that $\Delta_i \geq C_1\sqrt{M}\left(\frac{k^{9/2}}{\beta^{5/2}}(n_1n_2n_3)^{-1/4}\omega_{\mathsf{max}} \log n + \frac{k^8}{\beta^{9/2}}(n_1n_2n_3/n)^{-1/2}\omega_{\mathsf{max}}\sqrt{\log n}\right)$. Here, the fourth line of (63) makes use of (50), (54) and (60).

Step 4: proving $\widehat{\boldsymbol{z}}_i = \phi_i(\boldsymbol{z}_i^\star)$. Finally, we will show that with probability exceeding $1-O(n^{-10})$, $\widehat{\boldsymbol{z}}_i = \phi_i(\boldsymbol{z}_i^\star)$ for all $i \in [3]$. In view of Theorem 6, Lemma 7, (40) and (48), with probability at least $1-O(n^{-10})$, for all $i\in[3]$,

$$\begin{aligned}\left\|\left(\boldsymbol{I}_{n_i}-\boldsymbol{U}_i\boldsymbol{U}_i^\top\right)\boldsymbol{X}_i^\star\right\|_{2,\infty} &\lesssim \sqrt{\frac{\mu_i k_i^3}{n_i}}\left(k_i^2\sqrt{n_1}\omega_{\mathsf{max}}\log n + k_i\left(n_1n_2n_3\right)^{1/4}\omega_{\mathsf{max}}\log n\right)\\ &\overset{(40)}{\lesssim} \frac{k^3}{\beta}\frac{\left(n_1n_2n_3\right)^{1/4}}{n_i^{1/2}}\omega_{\mathsf{max}}\log n, \end{aligned} \tag{64}$$

and

$$\begin{aligned}\left\|\boldsymbol{U}_i\boldsymbol{U}_i^\top\boldsymbol{E}_i\left(\boldsymbol{U}_{i+2}\boldsymbol{U}_{i+2}^\top\otimes\boldsymbol{U}_{i+1}\boldsymbol{U}_{i+1}^\top\right)\right\|_{2,\infty} &\leq \left\|\boldsymbol{U}_i\right\|_{2,\infty}\left\|\boldsymbol{U}_i^\top\boldsymbol{E}_i\left(\boldsymbol{U}_{i+2}\otimes\boldsymbol{U}_{i+1}\right)\right\|\\ &\overset{(48)\text{ and Lemma 7}}{\lesssim} \sqrt{\frac{\mu_ik_i^3}{n_i}}\omega_{\mathsf{max}}\sqrt{n\left(\mu_1k_1^2\right)\left(\mu_2k_2^2\right)\left(\mu_3k_3^2\right)k^3\log n}\\ &\overset{(40)}{\lesssim}\sqrt{\frac{k^4}{\beta^2 n_i}}\frac{k^6}{\beta^3}\omega_{\mathsf{max}}\sqrt{n\log n}\\ &\leq \frac{k^8}{\beta^4}\omega_{\mathsf{max}}\sqrt{\frac{n\log n}{n_i}}. \end{aligned}\tag{65}$$

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 (11b), we know that for any $i\in[3]$,

$$\left\{j\in[n_1] : (\widehat{\boldsymbol{z}}_i)_j \neq \phi_i\big((\boldsymbol{z}_i^\star)_j\big)\right\} \subseteq \left\{j\in[n_1] : \exists \ell\in[k_i] \text{ s.t. } \left\|\big(\widehat{\boldsymbol{B}}_i\big)_{j:}^\top - \widehat{\boldsymbol{b}}_\ell^{(i)}\right\|_2 \leq \left\|\big(\widehat{\boldsymbol{B}}_i\big)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2\right\}. \tag{66}$$

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

$$\begin{aligned}&\mathbb{1}\left\{\left\|\big(\widehat{\boldsymbol{B}}_i\big)_{j:}^\top - \widehat{\boldsymbol{b}}_\ell^{(i)}\right\|_2 \leq \left\|\big(\widehat{\boldsymbol{B}}_i\big)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2\right\}\\ &= \mathbb{1}\left\{\left\|\big(\widehat{\boldsymbol{B}}_i\big)_{j:}^\top - \widehat{\boldsymbol{b}}_\ell^{(i)}\right\|_2 + \left\|\big(\widehat{\boldsymbol{B}}_i\big)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2 \leq 2\left\|\big(\widehat{\boldsymbol{B}}_i\big)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2\right\}\\ &\leq \mathbb{1}\left\{\left\|\widehat{\boldsymbol{b}}_\ell^{(i)} - \widehat{\boldsymbol{b}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2 \leq 2\left\|\big(\widehat{\boldsymbol{B}}_i\big)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2\right\}\\ &\leq \mathbb{1}\left\{\left\|\widehat{\boldsymbol{\theta}}_\ell^{(i)} - \widehat{\boldsymbol{\theta}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2 \leq 2\left\|\big(\widehat{\boldsymbol{B}}_i\big)_{j:}^\top - \widehat{\boldsymbol{b}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2\right\}\\ &= \mathbb{1}\left\{\left\|\widehat{\boldsymbol{\theta}}_\ell^{(i)} - \widehat{\boldsymbol{\theta}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2 \leq 2\left\|\left(\boldsymbol{U}_i\right)_{j,:}\boldsymbol{U}_i^\top\boldsymbol{Y}_i\left(\boldsymbol{U}_{i+2}\otimes\boldsymbol{U}_{i+1}\right) - \widehat{\boldsymbol{b}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)\top}\right\|_2\right\}, \end{aligned}\tag{67}$$

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

$$\begin{aligned}&\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((\boldsymbol{z}_i^\star)_j)}^{(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((\boldsymbol{z}_i^\star)_j)}^{(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((\boldsymbol{z}_i^\star)_j)}^{(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((\boldsymbol{z}_i^\star)_j)}^{(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((\boldsymbol{z}_i^\star)_j)}^{(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((\boldsymbol{z}_i^\star)_j)}^{(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((\boldsymbol{z}_i^\star)_j)}^{(i)} - \boldsymbol{\theta}_{(\boldsymbol{z}_i^\star)_j}^{(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 \end{aligned}$$

and

$$\left\|\widehat{\boldsymbol{\theta}}_{\ell}^{(i)}-\widehat{\boldsymbol{\theta}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2 \geq \left\|\boldsymbol{\theta}_{\phi_i^{-1}(\ell)}^{(i)\star}-\boldsymbol{\theta}_{(\boldsymbol{z}_i^\star)_j}^{(i)\star}\right\|_2-\left\|\widehat{\boldsymbol{\theta}}_{\ell}^{(i)}-\boldsymbol{\theta}_{\phi_i^{-1}(\ell)}^{(i)\star}\right\|_2-\left\|\widehat{\boldsymbol{\theta}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}-\boldsymbol{\theta}_{(\boldsymbol{z}_i^\star)_j}^{(i)\star}\right\|_2$$
$$\overset{(56)}{\geq} \beta\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i-2\sup_{a\in[k_i]}\left\|\widehat{\boldsymbol{\theta}}_{\phi_i(a)}^{(i)}-\boldsymbol{\theta}_{a}^{(i)\star}\right\|_2.$$

Putting the previous two inequalities, (63), (64) and (65) together, we know that, with probability at least $1-O(n^{-10})$, for all $\ell\in[k_i]$,

$$\begin{aligned}
&\left\|\widehat{\boldsymbol{\theta}}_{\ell}^{(i)}-\widehat{\boldsymbol{\theta}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)}\right\|_2-2\left\|(\boldsymbol{U}_i)_{j,:}\boldsymbol{U}_i^\top\boldsymbol{Y}_i\left(\boldsymbol{U}_{i+2}\otimes\boldsymbol{U}_{i+1}\right)-\widehat{\boldsymbol{b}}_{\phi_i((\boldsymbol{z}_i^\star)_j)}^{(i)\top}\right\|_2\\
&\geq \beta\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i-2\sup_{a\in[k_i]}\left\|\widehat{\boldsymbol{\theta}}_{\phi_i(a)}^{(i)}-\boldsymbol{\theta}_{a}^{(i)\star}\right\|_2\\
&\quad-2\left(\left\|\boldsymbol{U}_i\boldsymbol{U}_i^\top\boldsymbol{E}_i\left(\boldsymbol{U}_{i+2}\otimes\boldsymbol{U}_{i+1}\right)\right\|_{2,\infty}+\left\|\left(\boldsymbol{I}_{n_i}-\boldsymbol{U}_i\boldsymbol{U}_i^\top\right)\boldsymbol{X}_i^\star\right\|_{2,\infty}+\sup_{a\in[k_i]}\left\|\widehat{\boldsymbol{\theta}}_{\phi_i(a)}^{(i)}-\boldsymbol{\theta}_{a}^{(i)\star}\right\|_2\right)\\
&\geq \beta\sqrt{\frac{n_{-i}}{k_{-i}}}\Delta_i-4\sqrt{\frac{C_7M\left(\frac{k^7}{\beta^3}\left(n_1n_2n_3\right)^{1/2}\omega_{\mathsf{max}}^2\log^2 n+\frac{k^{14}}{\beta^7}n\omega_{\mathsf{max}}^2\log n\right)}{n_i}}\\
&\quad-C_8\frac{k^3}{\beta}\frac{\left(n_1n_2n_3\right)^{1/4}}{n_i^{1/2}}\omega_{\mathsf{max}}\log n-C_8\frac{k^8}{\beta^4}\omega_{\mathsf{max}}\sqrt{\frac{n\log n}{n_i}}\\
&>0,
\end{aligned}\tag{68}$$

provided that $\Delta_i/\omega_{\mathsf{max}}\geq C_1\sqrt{M}\left(\frac{k^{9/2}}{\beta^{5/2}}\left(n_1n_2n_3\right)^{-1/4}\log n+\frac{k^9}{\beta^5}\left(n_1n_2n_3/n\right)^{-1/2}\sqrt{\log n}\right)$.

Combining (66), (67) and (68), we arrive at

$$\widehat{\boldsymbol{z}}_i=\phi_i\left(\boldsymbol{z}_i^\star\right),\qquad \forall i\in[3]$$

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

C Proof of Theorem 5

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

$$\overline{r}=\begin{cases}\max\mathcal{A}, & \text{if } \mathcal{A}\neq\emptyset;\\ 0, & \text{otherwise.}\end{cases},\tag{69}$$

We claim that

$$\sigma_{\overline{r}+1}^\star\leq 2C_0r\left[(m_1m_2)^{1/4}+rm_1^{1/2}\right]\omega_{\mathsf{max}}\log m.\tag{70}$$

In fact, if $\sigma_{\overline{r}+1}^\star=0$, then (70) 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 $\overline{r}$ immediately tells us that $\sigma_i^\star\leq C_0r[(m_1m_2)^{1/4}+rm_1^{1/2}]\log m$ and consequently one has

$$\sigma_{\overline{r}+1}^\star=\sigma_i^\star\prod_{j=\overline{r}+1}^{i-1}\frac{\sigma_j^\star}{\sigma_{j+1}^\star}\leq\sigma_i^\star\left(\frac{4r}{4r-1}\right)^{i-\overline{r}-1}\leq\sigma_i^\star\cdot\left(1+\frac{1}{3r}\right)^r\leq 2C_0r\left[(m_1m_2)^{1/4}+rm_1^{1/2}\right]\omega_{\mathsf{max}}\log m.$$

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

$$\overline{r}=\max\mathcal{A}.\tag{71}$$

Part (b): proving (31b). The rest of the proof is devoted to prove (31b). We let

$$\boldsymbol{U}^{\star(1)} = [\boldsymbol{u}_1^\star, \ldots, \boldsymbol{u}_{\overline{r}}^\star], \qquad \boldsymbol{\Sigma}^{\star(1)} = \mathsf{diag}\left(\sigma_1^\star, \ldots, \sigma_{\overline{r}}^\star\right), \qquad \boldsymbol{V}^{\star(1)} = [\boldsymbol{v}_1^\star, \ldots, \boldsymbol{v}_{\overline{r}}^\star], \tag{72}$$

$$\boldsymbol{U}^{\star(2)} = \left[\boldsymbol{u}_{\overline{r}+1}^\star, \ldots, \boldsymbol{u}_r^\star\right], \qquad \boldsymbol{\Sigma}^{\star(2)} = \mathsf{diag}\left(\sigma_{\overline{r}+1}^\star, \ldots, \sigma_r^\star\right), \qquad \boldsymbol{V}^{\star(2)} = \left[\boldsymbol{v}_{\overline{r}+1}^\star, \ldots, \boldsymbol{v}_r^\star\right]. \tag{73}$$

Then we have

$$\boldsymbol{U}^\star = \left[\boldsymbol{U}^{\star(1)}\ \boldsymbol{U}^{\star(2)}\right], \qquad \boldsymbol{V}^\star = \left[\boldsymbol{V}^{\star(1)}\ \boldsymbol{V}^{\star(2)}\right], \qquad \boldsymbol{\Sigma}^\star = \begin{bmatrix}\boldsymbol{\Sigma}^{\star(1)} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}^{\star(2)}\end{bmatrix}. \tag{74}$$

We denote the SVD of $\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}$ by

$$\widetilde{\boldsymbol{U}}^{(1)}\widetilde{\boldsymbol{\Sigma}}^{(1)}\widetilde{\boldsymbol{W}}^{(1)\top} = \boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}. \tag{75}$$

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

$$\left(\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}\right)\left(\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}\right)^\top = \widetilde{\boldsymbol{U}}^{(1)}\left(\widetilde{\boldsymbol{\Sigma}}^{(1)}\right)^2\widetilde{\boldsymbol{U}}^{(1)\top}. \tag{76}$$

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

$$\begin{aligned}
\boldsymbol{M}^{\mathsf{oracle}} &= \left(\boldsymbol{X}^\star + \boldsymbol{E}\right)\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\left(\boldsymbol{X}^\star + \boldsymbol{E}\right)^\top + \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)\\
&= \left(\boldsymbol{X}^\star + \boldsymbol{E}\right)\boldsymbol{V}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\left(\boldsymbol{X}^\star + \boldsymbol{E}\right)^\top + \left(\boldsymbol{X}^\star + \boldsymbol{E}\right)\boldsymbol{V}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\left(\boldsymbol{X}^\star + \boldsymbol{E}\right)^\top\\
&\quad + \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)\\
&= \underbrace{\left(\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}\right)\left(\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}\right)^\top + \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}}_{=:\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)}\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)}}}_{=:\boldsymbol{Z}_2}\\
&\quad + \underbrace{\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)}_{=:\boldsymbol{Z}_3}.
\end{aligned} \tag{77}$$

For convenience, we also let

$$\boldsymbol{Z} = \boldsymbol{Z}_1 + \boldsymbol{Z}_2 + \boldsymbol{Z}_3. \tag{78}$$

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

$$\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1} > 2\|\boldsymbol{Z}\|, \tag{79}$$

then it holds that

$$\left\|\overline{\boldsymbol{U}}_1\overline{\boldsymbol{U}}_1^\top - \boldsymbol{U}_1\boldsymbol{U}_1^\top\right\|_{2,\infty} \leq \frac{8}{\pi}\sum_{k\geq 1}\left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^k \sum_{\substack{0\leq j_1,\ldots,j_{k+1}\leq r\\ (j_1,\ldots,j_{k+1})\neq \boldsymbol{0}}}\left\|\overline{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}, \tag{80a}$$

$$\left\|(\overline{\boldsymbol{U}}_1\overline{\boldsymbol{U}}_1^\top - \boldsymbol{U}_1\boldsymbol{U}_1^\top)\overline{\boldsymbol{M}}\right\|_{2,\infty} \le \frac{40}{\pi}\sum_{k\ge 1}\overline{\lambda}_{r_1}\left(\frac{2}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\right)^k \sum_{\substack{0\le j_1,\dots,j_{k+1}\le r\\(j_1,\dots,j_{k+1})\neq\boldsymbol{0}}}\left\|\overline{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}. \tag{80b}$$

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 $1\le j\le r$ 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 $1-O(m^{-10})$,*

$$\left\|\left[\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^k\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty} \le C_3\sqrt{\mu r}\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^k\omega_{\textsf{max}}\log m \tag{81}$$

for all $0\le k\le \log n$, 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 $1-O(m^{-10})$,*

$$\left\|\left[\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^k\boldsymbol{U}^\star\right\|_{2,\infty} \le C_3\sqrt{\frac{\mu r}{n_1}}\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^k \tag{82}$$

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

The following lemma, which was 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 $1-O(m^{-10})$,*

$$\left|\widetilde{\sigma}_i-\sigma_i^\star\right| \le \left\|\boldsymbol{E}\boldsymbol{V}^{\star(1)}\right\| \le \left\|\boldsymbol{E}\boldsymbol{V}^\star\right\| \le \sqrt{C}_5\sqrt{m_1}\omega_{\textsf{max}}\log m, \qquad \forall i\in[\overline{r}], \tag{83a}$$

$$\left\|\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top - \boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\right\| \le 3C_5\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m \tag{83b}$$

$$\left\|\boldsymbol{U}^{\star(1)}\boldsymbol{U}^{\star(1)\top}\widetilde{\boldsymbol{U}}^{(1)}-\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty} \le \frac{4C_5\sqrt{\mu r}\omega_{\textsf{max}}\log m}{\sigma_r^\star} \le \sqrt{\frac{\mu r}{m_1}}, \tag{83c}$$

$$\left\|\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty} \le 2\sqrt{\frac{\mu r}{m_1}}. \tag{83d}$$

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

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

$$\mathcal{E} = \{(81)\text{ and }(82)\text{ hold for }0\le k\le\log n\}\cap\{(83a),(83b),(83c)\text{ and }(83d)\text{ hold}\}. \tag{84}$$

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

$$\left\|\boldsymbol{Z}_3^i\boldsymbol{U}^\star\right\|_{2,\infty} \le 3C_3\sqrt{\frac{\mu r}{m_1}}\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^i, \tag{85a}$$

$$\left\|\boldsymbol{Z}_3^i\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty} \le 3C_3\sqrt{\mu r}\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^i\omega_{\textsf{max}}\log m, \tag{85b}$$

$$\left\|\boldsymbol{Z}_3^i\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty} \le 4C_3\sqrt{\frac{\mu r}{m_1}}\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^i, \tag{85c}$$

$$\left\|\boldsymbol{Z}_3^i\boldsymbol{Z}_1\right\|_{2,\infty} \le C_2\sqrt{\mu r}\left(\sigma_{\overline{r}+1}^\star+\sqrt{m_1}\omega_{\textsf{max}}\log m\right)\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^i\omega_{\textsf{max}}\log m, \tag{85d}$$

$$\left\|\boldsymbol{Z}_3^i\boldsymbol{Z}_2\right\|_{2,\infty} \le C_2\sqrt{\mu r}\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^i\omega_{\textsf{max}}\sigma_{\overline{r}+1}^\star\log m. \tag{85e}$$

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

$$\mathbb{P}(\mathcal{E}) \geq 1 - O\left(n^{-10}\right). \tag{86}$$

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

C.2 Main steps for proving (31b)

Step 1: bounding $\|\widetilde{\boldsymbol{U}}^{(2)}\|_2$. We start with controlling $\|\widetilde{\boldsymbol{U}}^{(2)}\|_2$. Combining (83a), (Chen et al., 2021a, Lemma 2.5) and the Wedin's $\sin\boldsymbol{\Theta}$ theorem, one has

$$\begin{aligned}\left\|\widetilde{\boldsymbol{U}}^{(1)\top}\boldsymbol{U}^{\star(2)}\right\| &\leq \left\|\widetilde{\boldsymbol{U}}^{(1)\top}\left(\boldsymbol{U}^{\star(1)}\right)_{\perp}\right\| = \left\|\widetilde{\boldsymbol{U}}^{(1)}\widetilde{\boldsymbol{U}}^{(1)\top} - \boldsymbol{U}^{\star(1)}\boldsymbol{U}^{\star(1)\top}\right\| \\ &\leq \frac{2\left\|\boldsymbol{E}\boldsymbol{V}^{\star(1)}\right\|}{\sigma_{\overline{r}}^{\star}} \leq \frac{2\sqrt{C_5}\sqrt{m}_1\omega_{\mathsf{max}}\log m}{\sigma_{\overline{r}}^{\star}} \ll \frac{1}{2}.\end{aligned} \tag{87}$$

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

$$\widetilde{\boldsymbol{U}}^{(2)}\left(\widetilde{\boldsymbol{\Sigma}}^{(2)}\right)^2\widetilde{\boldsymbol{U}}^{(2)} = \mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_{\perp}}\boldsymbol{U}^{\star(2)}\left(\boldsymbol{\Sigma}^{\star(2)}\right)^2\boldsymbol{U}^{\star(2)\top}\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_{\perp}} \tag{88}$$

where $\widetilde{\boldsymbol{U}}^{(2)} \in \mathcal{O}^{m_1, r-\overline{r}}$ and $\widetilde{\boldsymbol{\Sigma}}^{(1)} = \mathsf{diag}(\widetilde{\sigma}_{\overline{r}+1}, \dots, \widetilde{\sigma}_r)$ with $\widetilde{\sigma}_{\overline{r}+1} \geq \cdots \geq \widetilde{\sigma}_r \geq 0$. Recognizing that $\widetilde{\boldsymbol{U}}^{(1)\top}\widetilde{\boldsymbol{U}}^{(2)} = \boldsymbol{0}$, we know that the eigendecomposition of $\widetilde{\boldsymbol{M}}$ is

$$\widetilde{\boldsymbol{M}} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^{\top}, \tag{89}$$

where

$$\widetilde{\boldsymbol{U}} = \left[\widetilde{\boldsymbol{U}}^{(1)}\ \widetilde{\boldsymbol{U}}^{(2)}\right], \qquad \textit{and} \qquad \widetilde{\boldsymbol{\Lambda}} = \mathsf{diag}\left(\widetilde{\sigma}_1^2, \dots, \widetilde{\sigma}_r^2\right) = \begin{bmatrix}\left(\widetilde{\boldsymbol{\Sigma}}^{(1)}\right)^2 & \boldsymbol{0} \\ \boldsymbol{0} & \left(\widetilde{\boldsymbol{\Sigma}}^{(2)}\right)^2\end{bmatrix}. \tag{90}$$

In addition, one has

$$\begin{aligned}\sigma_{r-\overline{r}}\left(\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_{\perp}}\boldsymbol{U}^{\star(2)}\right) &= \sigma_{r-\overline{r}}\left(\left(\widetilde{\boldsymbol{U}}^{(1)}\right)_{\perp}^{\top}\boldsymbol{U}^{\star(2)}\right) = \min_{\boldsymbol{a}\in\mathbb{R}^{r-\overline{r}}:\|\boldsymbol{a}\|_2=1}\left\|\left(\widetilde{\boldsymbol{U}}^{(1)}\right)_{\perp}^{\top}\boldsymbol{U}^{\star(2)}\boldsymbol{a}\right\|_2 \\ &= \sqrt{\min_{\boldsymbol{a}\in\mathbb{R}^{r-\overline{r}}:\|\boldsymbol{a}\|_2=1}\left(\left\|\boldsymbol{U}^{\star(2)}\boldsymbol{a}\right\|_2^2 - \left\|\widetilde{\boldsymbol{U}}^{(1)\top}\boldsymbol{U}^{\star(2)}\boldsymbol{a}\right\|_2^2\right)} \\ &= \sqrt{1 - \max_{\boldsymbol{a}\in\mathbb{R}^{r-\overline{r}}:\|\boldsymbol{a}\|_2=1}\left\|\widetilde{\boldsymbol{U}}^{(1)\top}\boldsymbol{U}^{\star(2)}\boldsymbol{a}\right\|_2^2} \\ &= \sqrt{1 - \left\|\widetilde{\boldsymbol{U}}^{(1)\top}\boldsymbol{U}^{\star(2)}\right\|^2} \\ &\geq \sqrt{1 - \frac{1}{Cr^2}} \\ &\geq \sqrt{1 - \left(\frac{1}{2}\right)^2} = \frac{\sqrt{3}}{2}.\end{aligned} \tag{91}$$

The last line holds due to (87). 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 (83d), (87) and (91), one has

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

$$\le \frac{2}{\sqrt{3}}\left(\sqrt{\frac{\mu r}{m_1}}+\|\widetilde{\boldsymbol{U}}^{(1)}\|_{2,\infty}\|\widetilde{\boldsymbol{U}}^{(1)\top}\boldsymbol{U}^{\star(2)}\|\right)$$
$$\le \frac{2}{\sqrt{3}}\left(\sqrt{\frac{\mu r}{m_1}}+2\sqrt{\frac{\mu r}{m_1}}\cdot\frac{1}{2}\right)$$
$$\le 2\sqrt{\frac{\mu r}{m_1}}. \tag{92}$$

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

$$\widetilde{\sigma}_{r'}^2-\widetilde{\sigma}_{r'+1}^2>2\|\boldsymbol{Z}\|, \qquad \forall r'\in\mathcal{A}. \tag{93}$$

We know from the definition of $\widetilde{\sigma}_{\overline{r}+1}$ that

$$\widetilde{\sigma}_{\overline{r}+1}\le\left\|\boldsymbol{\Sigma}^{\star(2)}\right\|=\sigma_{\overline{r}+1}^{\star}. \tag{94}$$

Further, (88) and (91) together imply that

$$\widetilde{\sigma}_{\overline{r}+1}^2\ge\sigma_{r-\overline{r}}^2\big(\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp}\boldsymbol{U}^{\star(2)}\big)\sigma_{\overline{r}+1}^{\star 2}\ge\left(1-\frac{2\sqrt{C_5}\sqrt{m}_1\omega_{\mathsf{max}}\log m}{\sigma_{\overline{r}}^{\star}}\right)^2\sigma_{\overline{r}+1}^{\star 2}\ge\left(1-\frac{1}{Cr^2}\right)^2\sigma_{\overline{r}+1}^{\star 2} \tag{95}$$

for some large constant $C>0$. By virtue of (94) and (95), one has

$$\max\left\{\left(1-\frac{1}{Cr^2}\right)\sigma_{\overline{r}+1}^{\star},\sigma_{\overline{r}+1}^{\star}-2\sqrt{C_5}\sqrt{m}_1\omega_{\mathsf{max}}\log m\right\}\le\widetilde{\sigma}_{\overline{r}+1}\le\sigma_{\overline{r}+1}^{\star}. \tag{96}$$

Putting (83a), (94) and the fact $\sigma_{r'}^{\star}\ge\sigma_{\overline{r}}^{\star}\ge C_0r[(m_1m_2)^{1/4}+m_1^{1/2}]\log m$ together, one has

$$\begin{aligned}
\widetilde{\sigma}_{r'}-\widetilde{\sigma}_{r'+1}&\ge\sigma_{r'}^{\star}-\sigma_{r'+1}^{\star}-2\left\|\boldsymbol{E}\boldsymbol{V}^{\star}\right\|\\
&\ge\sigma_{r'}^{\star}-\sigma_{r'+1}^{\star}-2\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m\\
&\ge\sigma_{r'}^{\star}-\sigma_{r'+1}^{\star}-\frac{\sigma_{r'}^{\star}}{Cr}\\
&\ge\frac{1}{2}\left(\sigma_{r'}^{\star}-\sigma_{r'+1}^{\star}\right)+\frac{1}{2}\left(\sigma_{r'}^{\star}-\frac{4r-1}{4r}\sigma_{r'}^{\star}\right)-\frac{\sigma_{r'}^{\star}}{Cr}\\
&\ge\frac{1}{2}\left(\sigma_{r'}^{\star}-\sigma_{r'+1}^{\star}\right)\ge\frac{\sigma_{r'}^{\star}}{8r}.
\end{aligned}$$

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

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

Combining the previous two inequalities, we arrive at

$$\widetilde{\sigma}_{r'}^2-\widetilde{\sigma}_{r'+1}^2\ge\frac{1}{4}\left(\sigma_{r'}^{\star 2}-\sigma_{r'+1}^{\star 2}\right)\ge\frac{\sigma_{r'}^{\star 2}}{16r}. \tag{97}$$

Now, we move on to control $\|\boldsymbol{Z}\|$. In view of (83a) and (87), we have

$$\|\boldsymbol{Z}_1\|\le2\left\|\boldsymbol{\Sigma}^{\star(2)}\right\|\left\|\boldsymbol{E}\boldsymbol{V}^{\star}\right\|+\left\|\boldsymbol{E}\boldsymbol{V}^{\star}\right\|^2\le2\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\overline{r}+1}^{\star}+C_5m_1\omega_{\mathsf{max}}^2\log^2 m \tag{98}$$

and

$$\|\boldsymbol{Z}_2\|\le2\left\|\widetilde{\boldsymbol{U}}_1^{\top}\boldsymbol{U}_2^{\star(2)}\right\|\left\|\boldsymbol{\Sigma}^{\star(2)}\right\|^2\lesssim\frac{\sqrt{m}_1\omega_{\mathsf{max}}\log m}{\sigma_{\overline{r}}^{\star}}\sigma_{\overline{r}+1}^{\star 2}\le\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\overline{r}+1}^{\star}. \tag{99}$$

Combining (98), (99) and (83b), we arrive at

$$\|\boldsymbol{Z}\| \lesssim \sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma^\star_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right) \omega^2_{\mathsf{max}} \log^2 m \ll \frac{\sigma^{\star 2}_{r'}}{16r} \le \widetilde{\sigma}^2_{r'} - \widetilde{\sigma}^2_{r'+1}, \tag{100}$$

which validates (93). Here, the second inequality uses the facts $\sigma^\star_{r'} \ge \sigma^\star_{\bar{r}} \ge C_0 r[(m_1 m_2)^{1/4} + r m_1^{1/2}] \log m$. By virtue of Lemma 1 and (97), we have

$$\begin{aligned} &\left\| \boldsymbol{U}^{\mathsf{oracle}}_{:,1:r'} \boldsymbol{U}^{\mathsf{oracle}\top}_{:,1:r'} - \widetilde{\boldsymbol{U}}_{:,1:r'} \widetilde{\boldsymbol{U}}^\top_{:,1:r'} \right\|_{2,\infty} \\ &\quad \le \frac{8}{\pi} \sum_{k\ge 1} \frac{2^k}{\left(\widetilde{\sigma}^2_{r'} - \widetilde{\sigma}^2_{r'+1}\right)^k} \sum_{\substack{0\le j_1,\dots,j_{k+1}\le r \\ (j_1,\dots,j_{k+1})^\top \ne \mathbf{0}_{k+1}}} \left\| \widetilde{\boldsymbol{P}}_{j_1} \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \right\|_{2,\infty} \\ &\quad \le \frac{8}{\pi} \sum_{k\ge 1} \left(\frac{8}{\sigma^{\star 2}_{r'} - \sigma^{\star 2}_{r'+1}} \right)^k \sum_{\substack{0\le j_1,\dots,j_{k+1}\le r \\ (j_1,\dots,j_{k+1})^\top \ne \mathbf{0}_{k+1}}} \left\| \widetilde{\boldsymbol{P}}_{j_1} \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \right\|_{2,\infty}, \end{aligned} \tag{101}$$

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

To bound $\|\boldsymbol{U}^{\mathsf{oracle}}_{:,1:r'} \boldsymbol{U}^{\mathsf{oracle}\top}_{:,1:r'} - \widetilde{\boldsymbol{U}}_{:,1:r'} \widetilde{\boldsymbol{U}}^\top_{:,1:r'}\|_{2,\infty}$, 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 $1 \le k \le \log n$, and show that the total contribution of the remaining terms are 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 $1 \le k \le \log n$ and $(j_1, \dots, j_{k+1}) \in \{0, 1, \dots, r\}^{k+1} \backslash \mathbf{0}$, let ℓ denote the the smallest i such that $j_i \ne 0$.

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 (83d) and (92) together show that

$$\|\widetilde{\boldsymbol{u}}_{j_1}\|_\infty \le \max\left\{ \|\widetilde{\boldsymbol{U}}_1\|_{2,\infty}, \|\widetilde{\boldsymbol{U}}_2\|_{2,\infty} \right\} \le 2\sqrt{\frac{\mu r}{m_1}}. \tag{102}$$

Inequality (100) taken together with (102) leads us to

$$\begin{aligned} \left\| \widetilde{\boldsymbol{P}}_{j_1} \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \right\|_{2,\infty} &= \left\| \widetilde{\boldsymbol{u}}_{j_1} \widetilde{\boldsymbol{u}}_{j_1}^\top \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \right\|_{2,\infty} \\ &\le \|\widetilde{\boldsymbol{u}}_{j_1}\|_\infty \left\| \widetilde{\boldsymbol{u}}_{j_1}^\top \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \right\| \\ &\le 2\sqrt{\frac{\mu r}{m_1}} \|\boldsymbol{Z}\|^k \\ &\le 2\sqrt{\frac{\mu r}{m_1}} \left(C_2 \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma^\star_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right) \omega^2_{\mathsf{max}} \log^2 m \right) \right)^k. \end{aligned} \tag{103}$$

Step 3.2: bounding $\|\boldsymbol{Z}^i \widetilde{\boldsymbol{U}}\|_{2,\infty}$. Turning to $\ell \ge 2$, the triangle inequality implies that

$$\left\| \widetilde{\boldsymbol{P}}_{j_1} \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_2} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \right\|_{2,\infty} \le \left\| \boldsymbol{Z}^{\ell-1} \widetilde{\boldsymbol{P}}_{j_\ell} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \right\|_{2,\infty} + \sum_{i=1}^{\ell-1} \left\| \boldsymbol{Z}^{i-1} \boldsymbol{P}_{\widetilde{\boldsymbol{U}}} \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{i+1}} \boldsymbol{Z} \cdots \boldsymbol{Z} \widetilde{\boldsymbol{P}}_{j_{k+1}} \right\|_{2,\infty}. \tag{104}$$

To bound the right-hand side of (104), it is helpful to bound $\|\boldsymbol{Z}^i \widetilde{\boldsymbol{U}}\|_{2,\infty}$ first.

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

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

we know that

$$\begin{aligned}\boldsymbol{Z}^i\widetilde{\boldsymbol{U}}^{(1)} &= \left(\boldsymbol{Z}_1+\boldsymbol{Z}_2+\boldsymbol{Z}_3\right)^i\widetilde{\boldsymbol{U}}^{(1)}\\ &= \boldsymbol{Z}_3^i\widetilde{\boldsymbol{U}}^{(1)}+\sum_{j=0}^{i-1}\boldsymbol{Z}_3^j\boldsymbol{Z}_1\boldsymbol{Z}^{i-j-1}\widetilde{\boldsymbol{U}}^{(1)}+\sum_{j=0}^{i-1}\boldsymbol{Z}_3^j\boldsymbol{Z}_2\boldsymbol{Z}^{i-j-1}\widetilde{\boldsymbol{U}}^{(1)}. \end{aligned} \tag{105}$$

By virtue of Lemma 5 and (100), one has

$$\begin{aligned}\left\|\boldsymbol{Z}^i\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty} &\le \left\|\boldsymbol{Z}_3^i\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty}+\sum_{j=0}^{i-1}\left\|\boldsymbol{Z}_3^j\left(\boldsymbol{Z}_1+\boldsymbol{Z}_2\right)\boldsymbol{Z}^{i-j-1}\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty}\\
&\le \left\|\boldsymbol{Z}_3^i\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty}+\sum_{j=0}^{i-1}\left(\left\|\boldsymbol{Z}_3^j\boldsymbol{Z}_1\right\|_{2,\infty}+\left\|\boldsymbol{Z}_3^j\boldsymbol{Z}_2\right\|_{2,\infty}\right)\left\|\boldsymbol{Z}^{i-j-1}\widetilde{\boldsymbol{U}}^{(1)}\right\|\\
&\le \left\|\boldsymbol{Z}_3^i\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty}+\sum_{j=0}^{i-1}\left(\left\|\boldsymbol{Z}_3^j\boldsymbol{Z}_1\right\|_{2,\infty}+\left\|\boldsymbol{Z}_3^j\boldsymbol{Z}_2\right\|_{2,\infty}\right)\left\|\boldsymbol{Z}\right\|^{i-j-1}\\
&\le 4C_3\sqrt{\frac{\mu r}{m_1}}\left(C_3\left(\sqrt{m_1m_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_{\overline{r}+1}^{\star}+\sqrt{m}_1\omega_{\mathsf{max}}\log m\right)\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^j\omega_{\mathsf{max}}\log m\cdot\\
&\qquad\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\overline{r}+1}^{\star}+\left(\sqrt{m_1m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\right)^{i-j-1}\\
&\le 4C_3\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\overline{r}+1}^{\star}+\left(\sqrt{m_1m_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_{\overline{r}+1}^{\star}+\left(\sqrt{m_1m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\right)^i\sum_{j=0}^{i-1}\frac{1}{2^j}\\
&\le C_2\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\overline{r}+1}^{\star}+\left(\sqrt{m_1m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\right)^i, \end{aligned}\tag{106}$$

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

Step 3.2.2: bounding $\|\boldsymbol{Z}^i\widetilde{\boldsymbol{U}}^{(2)}\|_{2,\infty}$. Note that $\widetilde{\boldsymbol{U}}^{(2)}$ is also the column subspace of $\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp}\boldsymbol{U}^{\star(2)}\in\mathbb{R}^{m_1,r-\overline{r}}$. In view of Lemma 5, (87), (91) and (106), we arrive at

$$\begin{aligned}\left\|\boldsymbol{Z}^i\widetilde{\boldsymbol{U}}^{(2)}\right\|_{2,\infty} &\le \left\|\boldsymbol{Z}^i\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp}\boldsymbol{U}^{\star(2)}\right\|_{2,\infty}\sigma_{r-\overline{r}}^{-1}\big(\mathcal{P}_{(\widetilde{\boldsymbol{U}}^{(1)})_\perp}\boldsymbol{U}^{\star(2)}\big)\\
&\le \frac{2}{\sqrt{3}}\left(\left\|\boldsymbol{Z}^i\boldsymbol{U}^{\star(2)}\right\|_{2,\infty}+\left\|\boldsymbol{Z}^i\mathcal{P}_{\widetilde{\boldsymbol{U}}^{(1)}}\boldsymbol{U}^{\star(2)}\right\|_{2,\infty}\right)\\
&\le \frac{2}{\sqrt{3}}\left(\left\|\boldsymbol{Z}^i\boldsymbol{U}^{\star(2)}\right\|_{2,\infty}+\left\|\boldsymbol{Z}^i\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty}\left\|\widetilde{\boldsymbol{U}}^{(1)\top}\boldsymbol{U}^{\star(2)}\right\|\right)\\
&\le \frac{2}{\sqrt{3}}\Bigg(3C_3\sqrt{\frac{\mu r}{m_1}}\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)^i\\
&\qquad+C_2\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\overline{r}+1}^{\star}+\left(\sqrt{m_1m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\right)^i\cdot\frac{1}{2}\Bigg)\\
&\le C_2\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\overline{r}+1}^{\star}+\left(\sqrt{m_1m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\right)^i. \end{aligned}\tag{107}$$

Putting (106) and (107) together and recognizing that $\widetilde{\boldsymbol{U}}=[\widetilde{\boldsymbol{U}}^{(1)}\ \widetilde{\boldsymbol{U}}^{(2)}]$, we know that

$$\left\|\boldsymbol{Z}^i\widetilde{\boldsymbol{U}}\right\|_{2,\infty}\le\left\|\boldsymbol{Z}^i\widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty}+\left\|\boldsymbol{Z}^i\widetilde{\boldsymbol{U}}^{(2)}\right\|_{2,\infty}$$

$$\le 2C_2\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)\right)^i. \tag{108}$$

Step 4: 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 $\ell>1$. Plugging (100) and (108) into (104) yields that, for $\ell\ge 2$,

$$\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}\\
&\quad\le\left\|\boldsymbol{Z}^{\ell-1}\widetilde{\boldsymbol{u}}_{j_\ell}\right\|_{2,\infty}\left\|\widetilde{\boldsymbol{u}}_{j_\ell}^\top\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}+\sum_{i=1}^{\ell-1}\left\|\boldsymbol{Z}^{i-1}\widetilde{\boldsymbol{U}}\right\|_{2,\infty}\left\|\widetilde{\boldsymbol{U}}^\top\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{i+1}}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}\\
&\quad\le\left\|\boldsymbol{Z}^{\ell-1}\widetilde{\boldsymbol{U}}\right\|_{2,\infty}\left\|\boldsymbol{Z}\right\|^{k-\ell+1}+\sum_{i=1}^{\ell-1}\left\|\boldsymbol{Z}^{i-1}\widetilde{\boldsymbol{U}}\right\|_{2,\infty}\left\|\boldsymbol{Z}\right\|^{k-i+1}\\
&\quad\le 2C_2\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)\right)^{\ell-1}\cdot\\
&\qquad\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)\right)^{k-\ell+1}\\
&\qquad+\sum_{i=1}^{\ell-1}2C_2\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)\right)^{i-1}\cdot\\
&\qquad\qquad\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)\right)^{k-i+1}\\
&\quad=2C_2\sqrt{\frac{\mu r}{m_1}}\left(C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)\right)^{k}\cdot\ell\\
&\quad\le 2C_2\sqrt{\frac{\mu r}{m_1}}\left(2C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)\right)^{k}.
\end{aligned}$$

The last inequality comes from $\ell\le k+1\le 2^k$. Combining the previous inequality and (103), for any $1\le k\le\log n$ and $(j_1,\ldots,j_{k+1})\in\{0,1,\ldots,r\}^{k+1}\backslash\boldsymbol{0}$, we have

$$\begin{aligned}
&\left\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}\\
&\qquad\le 2C_2\sqrt{\frac{\mu r}{m_1}}\left(2C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)\right)^{k}.
\end{aligned}\tag{109}$$

Step 5: bounding $\|\boldsymbol{U}^{\mathsf{oracle}}_{:,1:r'}\boldsymbol{U}^{\mathsf{oracle}\top}_{:,1:r'}-\widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}^\top_{:,1:r'}\|_{2,\infty}$. Now, we are ready to bound $\|\boldsymbol{U}^{\mathsf{oracle}}_{:,1:r'}\boldsymbol{U}^{\mathsf{oracle}\top}_{:,1:r'}-\widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}^\top_{:,1:r'}\|_{2,\infty}$. As a consequence of (109), for any $1\le k\le\log n$, one has

$$\begin{aligned}
&\left(\frac{8}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}\right)^k\sum_{\substack{0\le j_1,\ldots,j_{k+1}\le r\\(j_1,\ldots,j_{k+1})^\top\neq\boldsymbol{0}_{k+1}}}\left\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}\\
&\le\left(\frac{8}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}\right)^k\cdot(r+1)^{k+1}\cdot 2C_2\sqrt{\frac{\mu r}{m_1}}\left(2C_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)\right)^{k}\\
&\le 4C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\frac{32C_2r\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}\right)^k.
\end{aligned}\tag{110}$$

Recalling that $\sigma^\star_{r'}\ge\sigma^\star_{\overline{r}}\ge C_0 r[(m_1m_2)^{1/4}+rm_1^{1/2}]\log m$, we know from (97) that there exists some large constant $C>0$ such that

$$\left(\frac{32C_2r\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}\right)^{k-1}$$

$$
\begin{aligned}
&\le \left(\frac{32C_2 r\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)}{\sigma^{\star 2}_{r'}/(4r)}\right)^{k-1}\\
&\le \left(\frac{1}{C^2}\right)^{k-1}\le\frac{1}{C^k}.
\end{aligned}\tag{111}
$$

For any $k\ge\lfloor\log m\rfloor+1$, in view of (100) and the the previous inequality, we have

$$
\begin{aligned}
&\left(\frac{8}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}\right)^k\sum_{\substack{0\le j_1,\dots,j_{k+1}\le r\\ (j_1,\dots,j_{k+1})^\top\neq\mathbf{0}_{k+1}}}\left\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}\\
&\le\left(\frac{8}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}\right)^k\cdot(r+1)^{k+1}\left\|\boldsymbol{Z}\right\|^k\\
&\le 2r\cdot\left(\frac{16rC_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}\right)^k\\
&\le\frac{2r}{C^k}\cdot\frac{16rC_2\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}.
\end{aligned}\tag{112}
$$

Combining (101), (110), (111) and (112), one can obtain

$$
\begin{aligned}
&\left\|\boldsymbol{U}^{\mathsf{oracle}}_{:,1:r'}\boldsymbol{U}^{\mathsf{oracle}\top}_{:,1:r'}-\widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}^\top_{:,1:r'}\right\|_{2,\infty}\\
&\le\sum_{1\le k\le\log m}4C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\frac{32C_2 r\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}\right)^k\\
&\quad+\sum_{k\ge\lfloor\log m\rfloor+1}\frac{2r}{C^k}\cdot\frac{16C_2 r\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}\\
&\lesssim\sqrt{\frac{\mu r^3}{m_1}}\frac{r\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right)}{\sigma^{\star 2}_{r'}-\sigma^{\star 2}_{r'+1}}\\
&\lesssim\sqrt{\frac{\mu r^3}{m_1}}\left(\frac{r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma^\star_{r'}}+\frac{r^2\left(\sqrt{m_1m_2}+m_1\right)\omega^2_{\mathsf{max}}\log^2 m}{\sigma^{\star 2}_{r'}}\right)\\
&\asymp\sqrt{\frac{\mu r^3}{m_1}}\left(\frac{r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma^\star_{r'}}+\frac{r^2\sqrt{m_1m_2}\omega^2_{\mathsf{max}}\log^2 m}{\sigma^{\star 2}_{r'}}\right).
\end{aligned}\tag{113}
$$

Here, the second last line holds due to (97) and the the last line makes use of the inequality

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

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

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

$$
\begin{aligned}
&\mathcal{P}_{\left(\widetilde{\boldsymbol{U}}_{:,1:r'}\right)_\perp}\boldsymbol{U}^\star_{:,1:r'}\\
&\quad=\mathcal{P}_{\left(\widetilde{\boldsymbol{U}}_{:,1:r'}\right)_\perp}\left(\boldsymbol{U}^\star_{:,1:r'}\boldsymbol{\Sigma}^\star_{1:r',1:\overline{r}}\right)\left(\boldsymbol{I}_{r'}\ \ \mathbf{0}_{r'\times(\overline{r}-r')}\right)^\top\left(\boldsymbol{\Sigma}^\star_{1:r',1:r'}\right)^{-1}
\end{aligned}
$$

$$
\begin{aligned}
&= \mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} \left(\widetilde{U}^{(1)} \widetilde{\Sigma}^{(1)} \widetilde{W}^{(1)\top} - EV^{\star(1)} - U^\star_{:,r'+1:\overline{r}} \Sigma^\star_{r'+1:\overline{r},1:\overline{r}} \right) \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top \left(\Sigma^\star_{1:r',1:r'} \right)^{-1} \\
&= \mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} \left(\widetilde{U}^{(1)}_{:,1:r'} \widetilde{\Sigma}^{(1)}_{1:r',1:r'} \widetilde{W}^{(1)\top}_{:,1:r'} + \widetilde{U}^{(1)}_{:,r'+1:\overline{r}} \widetilde{\Sigma}^{(1)}_{r'+1:\overline{r},r'+1:\overline{r}} \widetilde{W}^{(1)\top}_{:,r'+1:\overline{r}} - EV^{\star(1)} \right) \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top \\
&\quad \cdot \left(\Sigma^\star_{1:r',1:r'} \right)^{-1} \\
&= \widetilde{U}^{(1)}_{:,r'+1:\overline{r}} \widetilde{\Sigma}^{(1)}_{r'+1:\overline{r},r'+1:\overline{r}} \widetilde{W}^{(1)\top}_{:,r'+1:\overline{r}} \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top \left(\Sigma^\star_{1:r',1:r'} \right)^{-1} \\
&\quad - \mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} EV^{\star(1)} \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top \left(\Sigma^\star_{1:r',1:r'} \right)^{-1}, \qquad (114)
\end{aligned}
$$

where the second identity is valid since

$$
\Sigma^\star_{r'+1:\overline{r},1:\overline{r}} \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top = \left(\mathbf{0}_{(\overline{r}-r')\times r'} \ \Sigma^\star_{r'+1:\overline{r},r'+1:\overline{r}} \right) \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top = \mathbf{0}_{(\overline{r}-r')\times r'}
$$

and the last line comes from $\mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} \widetilde{U}^{(1)}_{:,1:r'} = \mathbf{0}$ and $\mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} \widetilde{U}^{(1)}_{:,r'+1:\overline{r}} = \widetilde{U}^{(1)}_{:,r'+1:\overline{r}}$. Note that $\widetilde{W}^{(1)}_{:,1:r'}$ (resp. $(I_{r'} \ \mathbf{0})^\top$) 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

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

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

$$
\left\| \widetilde{W}^{(1)\top}_{:,r'+1:\overline{r}} \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top \right\| = \left\| \left(\widetilde{W}_{:,1:r'} \right)^{(1)\top}_\perp \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top \right\| \lesssim \frac{\| EV^{\star(1)} \|}{\sigma^\star_{r'} - \sigma^\star_{r'+1}} \lesssim \frac{r \sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma^\star_{r'}}, \qquad (115a)
$$

$$
\left\| \left(U^\star_{:,1:r'} \right)^\top_\perp \widetilde{U}_{:,1:r'} \right\| \lesssim \frac{\| EV^{\star(1)} \|}{\sigma^\star_{r'} - \sigma^\star_{r'+1}} \lesssim \frac{r \sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma^\star_{r'}}. \qquad (115b)
$$

Moreover, combining (83a) and the assumption $\sigma^\star_{\overline{r}} \geq C_0 r [(m_1 m_2)^{1/4} + r m_1^{1/2}] \omega_{\mathsf{max}} \log m$, one has

$$
\widetilde{\sigma}_i \leq \sigma^\star_i + \| EV^\star \| \leq 2 \sigma^\star_i, \qquad \forall i \in [\overline{r}]. \qquad (116)
$$

Inequality (83d) combined with (115a) and (116) gives

$$
\begin{aligned}
&\left\| \widetilde{U}^{(1)}_{:,r'+1:\overline{r}} \widetilde{\Sigma}^{(1)}_{r'+1:\overline{r},r'+1:\overline{r}} \widetilde{W}^{(1)\top}_{:,r'+1:\overline{r}} \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top \left(\Sigma^\star_{1:r',1:r'} \right)^{-1} \right\|_{2,\infty} \\
&\quad \leq \left\| \widetilde{U}^{(1)} \right\|_{2,\infty} \left\| \widetilde{\Sigma}^{(1)}_{r'+1:\overline{r},r'+1:\overline{r}} \right\| \left\| \widetilde{W}^{(1)\top}_{:,r'+1:\overline{r}} \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top \right\| \left\| \left(\Sigma^\star_{1:r',1:r'} \right)^{-1} \right\| \\
&\quad \lesssim 2 \sqrt{\frac{\mu r}{m_1}} \cdot \widetilde{\sigma}_{r'+1} \cdot \frac{r \sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma^\star_{r'}} \cdot \frac{1}{\sigma^\star_{r'}} \\
&\quad \lesssim 2 \sqrt{\frac{\mu r}{m_1}} \cdot 2 \sigma^\star_{r'+1} \cdot \frac{r \sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma^\star_{r'}} \cdot \frac{1}{\sigma^\star_{r'}} \\
&\quad \lesssim \sqrt{\frac{\mu r}{m_1}} \frac{r \sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma^\star_{r'}}. \qquad (117)
\end{aligned}
$$

In addition, Lemma 2 and (83d) together imply that

$$
\begin{aligned}
&\left\| \mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} EV^{\star(1)} \left(I_{r'} \ \mathbf{0}_{r'\times(\overline{r}-r')} \right)^\top \left(\Sigma^\star_{1:r',1:r'} \right)^{-1} \right\|_{2,\infty} \\
&\quad \leq \frac{1}{\sigma^\star_{r'}} \left\| \mathcal{P}_{(\widetilde{U}_{:,1:r'})_\perp} EV^{\star(1)} \right\|_{2,\infty} \\
&\quad \leq \frac{1}{\sigma^\star_{r'}} \left(\left\| EV^{\star(1)} \right\|_{2,\infty} + \left\| \mathcal{P}_{\widetilde{U}_{:,1:r'}} EV^{\star(1)} \right\|_{2,\infty} \right) \\
&\quad \leq \frac{1}{\sigma^\star_{r'}} \left(\left\| EV^{\star(1)} \right\|_{2,\infty} + \left\| \widetilde{U}_{:,1:r'} \right\|_{2,\infty} \left\| \widetilde{U}_{:,1:r'} \right\| \left\| EV^{\star(1)} \right\| \right)
\end{aligned}
$$

$$\lesssim \frac{1}{\sigma_{r'}^{\star}}\left(\sqrt{\mu r}\omega_{\max}\log n+2\sqrt{\frac{\mu r}{m_1}}\cdot\sqrt{m_1}\omega_{\max}\log n\right)$$
$$\asymp \sqrt{\frac{\mu r}{m_1}}\frac{\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r'}^{\star}}. \tag{118}$$

Combing (114), (117) and (118), we know that

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

The previous inequality taken together with (115b) and (83d) reveal that

$$\begin{aligned}
&\|\widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^{\top}-\boldsymbol{U}_{:,1:r'}^{\star}\boldsymbol{U}_{:,1:r'}^{\star\top}\|_{2,\infty}\\
&\quad\le\|(\boldsymbol{U}_{:,1:r'}^{\star}-\widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^{\top}\boldsymbol{U}_{:,1:r'}^{\star})\boldsymbol{U}_{:,1:r'}^{\star\top}\|_{2,\infty}+\|\widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^{\top}\boldsymbol{U}_{:,1:r'}^{\star}\boldsymbol{U}_{:,1:r'}^{\star\top}-\widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^{\top}\|_{2,\infty}\\
&\quad\le\left\|\left(\mathcal{P}_{(\widetilde{\boldsymbol{U}}_{:,1:r'})_{\perp}}\boldsymbol{U}_{:,1:r'}^{\star}\right)\boldsymbol{U}_{:,1:r'}^{\star\top}\right\|_{2,\infty}+\|\widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^{\top}\left(\boldsymbol{U}_{:,1:r'}^{\star}\right)_{\perp}\left(\boldsymbol{U}_{:,1:r'}^{\star}\right)_{\perp}^{\top}\|_{2,\infty}\\
&\quad\le\left\|\mathcal{P}_{(\widetilde{\boldsymbol{U}}_{:,1:r'})_{\perp}}\boldsymbol{U}_{:,1:r'}^{\star}\right\|_{2,\infty}+\|\widetilde{\boldsymbol{U}}_{:,1:r'}\|_{2,\infty}\left\|\left(\boldsymbol{U}_{:,1:r'}^{\star}\right)_{\perp}^{\top}\widetilde{\boldsymbol{U}}_{:,1:r'}\right\|\\
&\quad\lesssim\sqrt{\frac{\mu r}{m_1}}\frac{r\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r'}^{\star}}+2\sqrt{\frac{\mu r}{m_1}}\frac{r\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r'}^{\star}}\\
&\quad\asymp\sqrt{\frac{\mu r}{m_1}}\frac{r\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r'}^{\star}}.
\end{aligned}\tag{119}$$

By virtue of (113) and (119), we arrive at

$$\begin{aligned}
&\|\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}}\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top}-\boldsymbol{U}_{:,1:r'}^{\star}\boldsymbol{U}_{:,1:r'}^{\star\top}\|_{2,\infty}\\
&\quad\le\|\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}}\boldsymbol{U}_{:,1:r'}^{\mathsf{oracle}\top}-\widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^{\top}\|_{2,\infty}+\|\widetilde{\boldsymbol{U}}_{:,1:r'}\widetilde{\boldsymbol{U}}_{:,1:r'}^{\top}-\boldsymbol{U}_{:,1:r'}^{\star}\boldsymbol{U}_{:,1:r'}^{\star\top}\|_{2,\infty}\\
&\quad\lesssim\sqrt{\frac{\mu r^3}{m_1}}\left(\frac{r^2\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r'}^{\star}}+\frac{r^2\sqrt{m_1m_2}\omega_{\max}^2\log^2 m}{\sigma_{r'}^{\star2}}\right).
\end{aligned}\tag{120}$$

□

C.3 Proof of Lemma 1

Let γ_1 denote the following counterclockwise contour on the complex plane:

$$\begin{aligned}
\gamma_1=\Bigg\{x+y\mathrm{i}:\ &x=\frac{\overline{\lambda}_{r_1}+\overline{\lambda}_{r_1+1}}{2},-\frac{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}{2}\le y\le\frac{\overline{\lambda}_{r_1}+\overline{\lambda}_{r_1+1}}{2},\\
&\text{or } x=\overline{\lambda}_1+\frac{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}{2},-\frac{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}{2}\le y\le\frac{\overline{\lambda}_{r_1}+\overline{\lambda}_{r_1+1}}{2},\\
&\text{or } y=\pm\frac{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}{2},\frac{\overline{\lambda}_{r_1}+\overline{\lambda}_{r_1+1}}{2}\le x\le\overline{\lambda}_1+\frac{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}{2}\Bigg\}.
\end{aligned}$$

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

$$\left|\eta-\overline{\lambda}_i\right|\le\frac{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}{2}. \tag{121}$$

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

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

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

$$\boldsymbol{U}_1\boldsymbol{U}_1^\top = \frac{1}{2\pi \mathrm{i}}\oint_{\gamma_1} (\eta \boldsymbol{I} - \boldsymbol{M})\,\mathrm{d}\eta. \tag{122}$$

We define

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

In view of (121), for any $\eta \in \gamma_1$, we have

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

and consequently

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

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

$$\begin{aligned}
\boldsymbol{U}_1\boldsymbol{U}_1^\top - \overline{\boldsymbol{U}}_1\overline{\boldsymbol{U}}_1^\top &= \sum_{k\ge 1}\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\left[\mathcal{R}_{\overline{\boldsymbol{M}}}(\eta)\boldsymbol{Z}\right]^k\mathcal{R}_{\overline{\boldsymbol{M}}}(\eta)\mathrm{d}\eta \\
&= \sum_{k\ge 1}\sum_{1\le j_1,\dots,j_{k+1}\le n}\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\boldsymbol{P}_{\overline{\boldsymbol{u}}_{j_1}}\boldsymbol{Z}\boldsymbol{P}_{\overline{\boldsymbol{u}}_{j_2}}\boldsymbol{Z}\cdots\boldsymbol{P}_{\overline{\boldsymbol{u}}_{j_k}}\boldsymbol{Z}\boldsymbol{P}_{\overline{\boldsymbol{u}}_{j_{k+1}}} \\
&= \sum_{k\ge 1}\sum_{0\le j_1,\dots,j_{k+1}\le r}\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\overline{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\overline{\boldsymbol{P}}_{j_k}\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_{k+1}}.
\end{aligned} \tag{123}$$

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

Step 2: bounding $\|\boldsymbol{U}_1\boldsymbol{U}_1^\top - \overline{\boldsymbol{U}}_1\overline{\boldsymbol{U}}_1^\top\|_{2,\infty}$. By virtue of (123) and the triangle inequality, to prove (80a), one only needs to bound $|\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}|$. We consider two scenarios: all of $j_1,\dots,j_{k+1}$ are in the set $\{0\}\cup\{r_1+1,\dots,r_1\}$ and at least one of $j_1,\dots,j_{k+1}$ is in the set $\{1,\dots,r_1\}$.

Case 1: all of $j_1,\dots,j_{k+1}$ are 0 or larger than r_1. In this case, none of $\overline{\lambda}_{j_1},\dots,\overline{\lambda}_{j_{k+1}}$ 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

$$\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})} = 0, \tag{124}$$

and thus we have

$$\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\overline{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\overline{\boldsymbol{P}}_{j_k}\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_{k+1}} = \boldsymbol{0}. \tag{125}$$

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

$$\mathcal{J} = \{j : 1\le j\le r_1, \exists 1\le \ell\le k+1 \text{ s.t. } j_\ell = j\}, \tag{126}$$

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

$$|\eta - \lambda_{j_\ell}| \ge \frac{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}{2}, \qquad \forall 1 \le \ell \le k+1, \eta \in \gamma_2 \tag{127}$$

and

$$\frac{1}{2\pi \mathrm{i}} \oint_{\gamma_1} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} = \frac{1}{2\pi \mathrm{i}} \oint_{\gamma_2} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}. \tag{128}$$

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

$$\begin{aligned}
\left|\frac{1}{2\pi \mathrm{i}} \oint_{\gamma_2} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}\right| &\le \frac{1}{2\pi} \sup_{\eta \in \gamma_2} \left|\frac{1}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}\right| \cdot L(\gamma_2) \\
&\overset{(127)}{\le} \frac{1}{2\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^{k+1} \cdot 4\left(\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}\right) \\
&\le \frac{4}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^{k}.
\end{aligned}$$

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

$$\max\left\{\left|\eta - \overline{\lambda}_{j_{\mathsf{max}}}\right|, \left|\eta - \overline{\lambda}_{j_{\mathsf{min}}}\right|\right\} \ge \frac{\left|(\eta - \overline{\lambda}_{j_{\mathsf{max}}}) - (\eta - \overline{\lambda}_{j_{\mathsf{min}}})\right|}{2} = \frac{\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{max}}}}{2}. \tag{129}$$

In view of (127), (129) 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

$$\begin{aligned}
\frac{1}{\left|(\eta - \overline{\lambda}_{j_{\mathsf{max}}})(\eta - \overline{\lambda}_{j_{\mathsf{min}}})\right|} &\le \min\left\{\frac{4}{\left(\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{max}}}\right)\left(\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}\right)}, \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^2\right\} \\
&\le \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right) \cdot \left(\frac{4}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1} + \overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{j_{\mathsf{min}}}}\right)
\end{aligned} \tag{130}$$

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

$$\begin{aligned}
\left|\frac{1}{2\pi \mathrm{i}} \oint_{\gamma_2} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}\right| &\le \frac{1}{2\pi} \sup_{\eta \in \gamma_2} \left|\frac{1}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}\right| \cdot L(\gamma_2) \\
&\le \frac{1}{2\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right) \cdot \left(\frac{4}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1} + \overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{min}}}}\right) \\
&\quad \cdot \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^{k-1} \cdot \left(2\left(\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{max}}}\right) + 4\left(\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}\right)\right) \\
&\le \frac{8}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^{k}.
\end{aligned}$$

Therefore, we always have

$$\left|\frac{1}{2\pi \mathrm{i}} \oint_{\gamma_2} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}\right| \le \frac{8}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^{k}. \tag{131}$$

Combining (123), (125), (131) and the triangle inequality, we have finished the proof of (80a).

Step 3: bounding $\|(\overline{\boldsymbol{U}}_1\overline{\boldsymbol{U}}_1^\top - \boldsymbol{U}_1\boldsymbol{U}_1^\top)\overline{\boldsymbol{M}}\|_{2,\infty}$. Next, we move on to control $\|(\overline{\boldsymbol{U}}_1\overline{\boldsymbol{U}}_1^\top - \boldsymbol{U}_1\boldsymbol{U}_1^\top)\overline{\boldsymbol{M}}\|_{2,\infty}$. By virtue of (123), we have

$$
\begin{aligned}
&\left(\boldsymbol{U}_1\boldsymbol{U}_1^\top - \overline{\boldsymbol{U}}_1\overline{\boldsymbol{U}}_1^\top\right)\overline{\boldsymbol{M}} \\
&= \sum_{k\geq 1}\sum_{0\leq j_1,\dots,j_k\leq r}\sum_{j_{k+1}=1}^{r}\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\overline{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\overline{\boldsymbol{P}}_{j_k}\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_{k+1}}\overline{\boldsymbol{M}} \\
&= \sum_{k\geq 1}\sum_{0\leq j_1,\dots,j_k\leq r}\sum_{j_{k+1}=1}^{r}\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\overline{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\overline{\boldsymbol{P}}_{j_k}\boldsymbol{Z}\overline{\boldsymbol{P}}_{j_{k+1}}.
\end{aligned}
\tag{132}
$$

The second line and the third line make use of $\overline{\boldsymbol{U}}_\perp\overline{\boldsymbol{M}} = 0$ and $\overline{\boldsymbol{P}}_j\overline{\boldsymbol{M}} = \overline{\boldsymbol{u}}_j\overline{\boldsymbol{u}}_j^\top(\sum_{i=1}^r\overline{\lambda}_i\overline{\boldsymbol{u}}_i\overline{\boldsymbol{u}}_i^\top) = \overline{\lambda}_j$, respectively. For the rest of the proof, we prove 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 show that (80b) is valid.

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 $j_1,j_2\in\{r_1+1,\dots,r\}\cup\{0\}$ and (2) only one of j_1 and j_2 falls into the set $\{1,\dots,r_1\}$.

Case 1: $1\leq j_1,j_2\leq r_1$ or $j_1,j_2\in\{r_1+1,\dots,r\}\cup\{0\}$. 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 one has

$$
\left|\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\overline{\lambda}_{j_2}\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})(\eta-\overline{\lambda}_{j_2})}\right| = 0. \tag{133}
$$

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

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

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

$$
\left|\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})(\eta-\overline{\lambda}_{j_2})}\right| = \frac{\overline{\lambda}_{j_2}}{\left|\overline{\lambda}_{j_1}-\overline{\lambda}_{j_2}\right|} = \frac{1}{\left|\frac{\overline{\lambda}_{j_1}}{\overline{\lambda}_{j_2}}-1\right|} \leq \frac{1}{\max\left\{\frac{\overline{\lambda}_{r_1}}{\overline{\lambda}_{r_1+1}}-1, 1-\frac{\overline{\lambda}_{r_1+1}}{\overline{\lambda}_{r_1}}\right\}} = \frac{\overline{\lambda}_{r_1}}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}. \tag{134}
$$

Step 3.2: 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 $k>1$. If (1) all $j_1,\dots,j_{k+1}$ are all in the set $\{1,\dots,r_1\}$ or (2) all $j_1,\dots,j_{k+1}$ are all in the set $\{0\}\cup\{r_1+1,\dots,r\}$, then the function

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

is analytic in and on γ_1 or outside on on γ_1. As a result, one can use Cauchy's integral formula to obtain

$$
\left|\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\right| = 0. \tag{135}
$$

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

$$1 \leq |\{j_1, \ldots, j_{k+1}\} \cap \{1, \ldots, r_1\}| \leq k. \tag{136}$$

Let γ_3 denote the following counterclockwise rectangular contour:

$$\gamma_3 = \Big\{x + y\mathrm{i} : x = \frac{\overline{\lambda}_{j_{\mathsf{max}}} + \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2},$$
$$\text{or } x = \overline{\lambda}_{j_{\mathsf{min}}} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2},$$
$$\text{or } y = \pm\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}, \frac{\overline{\lambda}_{j_{\mathsf{max}}} + \overline{\lambda}_{r_1+1}}{2} \leq x \leq \overline{\lambda}_{j_{\mathsf{min}}} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}\Big\},$$

where we recall that j_{min} (resp. j_{max}) is the smallest (resp. largest) element in the set $\mathcal{J}$ defined in (126). Then one can check that

$$\left|\eta - \overline{\lambda}_{j_\ell}\right| \geq \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}, \qquad \forall 1 \leq \ell \leq k+1, \eta \in \gamma_3 \tag{137}$$

and the length of γ_3 satisfies

$$L(\gamma_3) = 2\left(\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{max}}}\right) + 4\left(\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}\right) = 2\overline{\lambda}_{j_{\mathsf{min}}} + 2\overline{\lambda}_{j_{\mathsf{max}}} - 4\overline{\lambda}_{r_1+1}. \tag{138}$$

In addition, we have

$$\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1} \frac{\mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1}\right)\cdots\left(\eta - \overline{\lambda}_{j_{k+1}}\right)} = \frac{1}{2\pi\mathrm{i}}\oint_{\gamma_3} \frac{\mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1}\right)\cdots\left(\eta - \overline{\lambda}_{j_{k+1}}\right)}. \tag{139}$$

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

$$\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{r_1+1} = \overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{max}}} + \overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1} \leq 4\left(\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}\right), \tag{140}$$

which further leads to

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

In view of (137), (139), (140) and the previous inequality, one has

$$\begin{aligned}
\left|\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1} \frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{\left(\eta - \overline{\lambda}_{j_1}\right)\cdots\left(\eta - \overline{\lambda}_{j_{k+1}}\right)}\right| &\leq \frac{1}{2\pi}\overline{\lambda}_{j_{k+1}}\left(\frac{2}{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}\right)^{k+1} L(\gamma_3) \\
&\leq \frac{1}{2\pi}\overline{\lambda}_{j_{\mathsf{min}}}\left(\frac{2}{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}\right)^{k+1} \cdot 10\left(\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}\right) \\
&\leq \frac{80}{\pi}\frac{\overline{\lambda}_{j_{\mathsf{min}}}}{\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{r_1+1}}\left(\frac{2}{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}\right)^{k-1} \\
&\leq \frac{80}{\pi}\frac{\overline{\lambda}_{r_1}}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^{k-1} = \frac{40}{\pi}\overline{\lambda}_{r_1}\left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}}\right)^{k}.
\end{aligned} \tag{141}$$

The third inequality holds because of (140) and the last one uses 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 $\overline{\lambda}_{j_{\mathsf{min}}} \geq \overline{\lambda}_{j_{\mathsf{max}}} \geq \overline{\lambda}_{r_1}$.

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

$$\mathcal{I}_1 = \left\{i : 1 \leq i \leq k+1, \overline{\lambda}_{j_i} \geq \frac{2}{3}\overline{\lambda}_{j_{\mathsf{min}}} + \frac{1}{3}\overline{\lambda}_{j_{\mathsf{max}}}\right\} \tag{142}$$

and

$$\mathcal{I}_2 = \left\{i : 1 \leq i \leq k+1, \overline{\lambda}_{j_i} \leq \frac{1}{3}\overline{\lambda}_{j_{\mathsf{min}}} + \frac{2}{3}\overline{\lambda}_{j_{\mathsf{max}}}\right\}. \tag{143}$$

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) $\min\{|\mathcal{I}_1|, |\mathcal{I}_2|\} \geq 2$; (2) $|\mathcal{I}_1| = 1$ and (3) $|\mathcal{I}_2| = 1$.

Case 2.1: $\min\{|\mathcal{I}_1|,|\mathcal{I}_2|\}\geq 2$. 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

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

and

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

Similar to (130), one can derive

$$\begin{aligned}
&\left|\frac{1}{\left(\eta-\overline{\lambda}_{j_{i_1}}\right)\left(\eta-\overline{\lambda}_{j_{i_2}}\right)\left(\eta-\overline{\lambda}_{j_{i_3}}\right)\left(\eta-\overline{\lambda}_{j_{i_4}}\right)}\right|\\
&\quad=\frac{1}{\left|\eta-\overline{\lambda}_{j_{i_1}}\right|\left|\eta-\overline{\lambda}_{j_{i_2}}\right|}\frac{1}{\left|\eta-\overline{\lambda}_{j_{i_3}}\right|\left|\eta-\overline{\lambda}_{j_{i_4}}\right|}\\
&\quad=\min\left\{\frac{4}{\left(\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_{\max}}\right)\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)},\frac{4}{\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)^2}\right\}\\
&\qquad\cdot\min\left\{\frac{12}{\left(\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_{\max}}\right)\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)},\frac{4}{\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)^2}\right\}\\
&\quad\leq\frac{8}{\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)\left(\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}\right)}\cdot\frac{16}{\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)\left(\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}\right)}\\
&\quad=\frac{128}{\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)^2\left(\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}\right)^2}.
\end{aligned}$$

The penultimate line uses the basic inequality $\min\{a/b,c/d\}\leq(a+c)/(b+d)$ for $a,b,c,d>0$. Putting the previous inequality, (137), (138) and (116) together, we have

$$\begin{aligned}
&\left|\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{\left(\eta-\overline{\lambda}_{j_1}\right)\cdots\left(\eta-\overline{\lambda}_{j_{k+1}}\right)}\right|\\
&\qquad\leq\frac{1}{2\pi}\sup_{\eta\in\gamma_3}\left|\frac{\overline{\lambda}_{j_{k+1}}}{\left(\eta-\overline{\lambda}_{j_1}\right)\cdots\left(\eta-\overline{\lambda}_{j_{k+1}}\right)}\right|\cdot L(\gamma_3)\\
&\qquad\leq\frac{1}{2\pi}\overline{\lambda}_{j_{k+1}}\frac{128}{\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)^2\left(\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}\right)^2}\left(\frac{2}{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}\right)^{k+1-4}\cdot 4\left(\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}\right)\\
&\qquad\leq\frac{64}{\pi}\frac{\overline{\lambda}_{j_{\min}}}{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}}\left(\frac{2}{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}\right)^{k-1}\\
&\qquad\leq\frac{64}{\pi}\frac{\overline{\lambda}_{r_1}}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\left(\frac{2}{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}\right)^{k-1}\\
&\qquad\leq\frac{32}{\pi}\left(\frac{2}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\right)^{k}\overline{\lambda}_{r_1}.
\end{aligned}\tag{144}$$

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: $|\mathcal{I}_1|=1$. We choose

$$\ell\in\underset{i:1\leq i\leq k+1,i\notin\mathcal{I}_1}{\arg\max}\ \overline{\lambda}_{j_i}.\tag{145}$$

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

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

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

$$\begin{aligned}\gamma_4 = \Big\{x + y\mathrm{i} : x = \frac{\overline{\lambda}_{j_{\mathsf{max}}} + \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2},\\ \text{or } x = \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2},\\ \text{or } y = \pm\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}, \frac{\overline{\lambda}_{j_{\mathsf{max}}} + \overline{\lambda}_{r_1+1}}{2} \leq x \leq \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}\Big\}\end{aligned}$$

and

$$\begin{aligned}\gamma_5 = \Big\{x + y\mathrm{i} : x = \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2},\\ \text{or } x = \overline{\lambda}_{j_{\mathsf{min}}} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2},\\ \text{or } y = \pm\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}, \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2} \leq x \leq \overline{\lambda}_{j_{\mathsf{min}}} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}\Big\}.\end{aligned}$$

For any complex number $\eta = x + y\mathrm{i}$ with $x = \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}$ and $-\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}$, 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 $\{\eta = x + y\mathrm{i} : x = \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2} \leq y \leq \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}\}$. Applying Cauchy's integral formula yields

$$\begin{aligned}&\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_3}\frac{\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\\ &\quad= \frac{1}{2\pi\mathrm{i}}\oint_{\gamma_4}\frac{\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})} + \frac{1}{2\pi\mathrm{i}}\oint_{\gamma_5}\frac{\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}.\end{aligned} \tag{146}$$

Repeating a similar argument as for (131) reveals that

$$\frac{1}{2\pi}\sup_{\eta\in\gamma_4}\left|\frac{1}{\prod_{1\leq i\leq k+1, i\notin\mathcal{I}_1}\left(\eta-\overline{\lambda}_{j_i}\right)}\right|\cdot L\left(\gamma_4\right) \leq \frac{8}{\pi}\left(\frac{2}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\right)^{k-1}. \tag{147}$$

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

$$\begin{aligned}\mathrm{Re}(\eta) &\leq \overline{\lambda}_{j_\ell} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}\\ &\leq \frac{2}{3}\overline{\lambda}_{j_{\mathsf{min}}} + \frac{1}{3}\overline{\lambda}_{j_{\mathsf{max}}} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}\\ &= \overline{\lambda}_{j_{\mathsf{min}}} - \frac{\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{max}}}}{3} + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}\\ &\leq \overline{\lambda}_{j_{\mathsf{min}}} - \frac{\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{j_{\mathsf{max}}}}{8} - \frac{5}{8}\left(\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}\right) + \frac{\overline{\lambda}_{j_{\mathsf{max}}} - \overline{\lambda}_{r_1+1}}{2}\\ &= \overline{\lambda}_{j_{\mathsf{min}}} - \frac{\overline{\lambda}_{j_{\mathsf{min}}} - \overline{\lambda}_{r_1+1}}{8},\end{aligned}$$

which further tells us that

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

Putting (147) and the previous inequality together, one has

$$\begin{aligned}
&\left|\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_4}\frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\right| \\
&\qquad\leq \frac{1}{2\pi}\sup_{\eta\in\gamma_4}\left|\frac{1}{\prod_{1\leq i\leq k,i\notin\mathcal{I}_1}(\eta-\overline{\lambda}_{j_i})}\right|\cdot\sup_{\eta\in\gamma_4}\frac{\overline{\lambda}_{j_{k+1}}}{|\eta-\overline{\lambda}_{j_{\min}}|}\cdot L(\gamma_4) \\
&\qquad\leq \frac{8}{\pi}\left(\frac{2}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\right)^{k-1}\cdot\frac{8\overline{\lambda}_{j_{\min}}}{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}} \\
&\qquad\leq \frac{32}{\pi}\left(\frac{2}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\right)^{k}\overline{\lambda}_{r_1}. \qquad (148)
\end{aligned}$$

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

$$\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_5}\frac{\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}=\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_5}\frac{\prod_{1\leq i\leq k+1,i\notin\mathcal{I}_1}\frac{1}{\eta-\overline{\lambda}_{j_i}}}{\eta-\overline{\lambda}_{j_{\min}}}\mathrm{d}\eta=\prod_{1\leq i\leq k+1,i\notin\mathcal{I}_1}\frac{1}{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_i}}. \qquad (149)$$

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

$$\begin{aligned}
\min_{i:1\leq i\leq k+1,i\notin\mathcal{I}_1}\left|\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_i}\right| &\geq \frac{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_{\max}}}{3} \\
&= \frac{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_{\max}}}{4}+\frac{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_{\max}}}{12} \\
&\geq \left(\frac{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_{\max}}}{4}+\frac{\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}}{4}\right)\vee\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right) \\
&= \frac{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}}{4}\vee\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right).
\end{aligned}$$

Combining (149) and the previous inequality, one has

$$\begin{aligned}
\left|\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_5}\frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\right| &\leq \prod_{1\leq i\leq k+1,i\notin\mathcal{I}_1}\frac{1}{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{j_i}}\overline{\lambda}_{j_{\min}} \\
&\leq \frac{1}{\left(\overline{\lambda}_{j_{\max}}-\overline{\lambda}_{r_1+1}\right)^{k-1}}\frac{4\overline{\lambda}_{j_{\min}}}{\overline{\lambda}_{j_{\min}}-\overline{\lambda}_{r_1+1}} \\
&\leq \frac{1}{\left(\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}\right)^{k-1}}\frac{4\overline{\lambda}_{r_1}}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}} \\
&\leq \left(\frac{2}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\right)^{k}\overline{\lambda}_{r_1}. \qquad (150)
\end{aligned}$$

Eqn. (139) together with (146), (148) and (150) implies that

$$\left|\frac{1}{2\pi\mathrm{i}}\oint_{\gamma_1}\frac{\overline{\lambda}_{j_{k+1}}\mathrm{d}\eta}{(\eta-\overline{\lambda}_{j_1})\cdots(\eta-\overline{\lambda}_{j_{k+1}})}\right|\leq\frac{36}{\pi}\left(\frac{2}{\overline{\lambda}_{r_1}-\overline{\lambda}_{r_1+1}}\right)^{k}\overline{\lambda}_{r_1}. \qquad (151)$$

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:

$$\gamma_4 = \Big\{ x + y\mathrm{i} : x = \frac{\overline{\lambda}_{j_{\max}} + \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \le y \le \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2},$$
$$\text{or } x = \overline{\lambda}_{j_{\ell'}} - \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \le y \le \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2},$$
$$\text{or } y = \pm\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, \frac{\overline{\lambda}_{j_{\max}} + \overline{\lambda}_{r_1+1}}{2} \le x \le \overline{\lambda}_{j_{\ell'}} - \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \Big\}$$

and

$$\gamma_5 = \Big\{ x + y\mathrm{i} : x = \overline{\lambda}_{j_{\ell'}} - \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \le y \le \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2},$$
$$\text{or } x = \overline{\lambda}_{j_{\min}} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, -\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \le y \le \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2},$$
$$\text{or } y = \pm\frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2}, \overline{\lambda}_{j_{\ell'}} - \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \le x \le \overline{\lambda}_{j_{\min}} + \frac{\overline{\lambda}_{j_{\max}} - \overline{\lambda}_{r_1+1}}{2} \Big\}.$$

Similar to (146), one has

$$\begin{aligned} &\frac{1}{2\pi\mathrm{i}} \oint_{\gamma_3} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \\ &= \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_6} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} + \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_7} \frac{\mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})}. \end{aligned} \tag{152}$$

Repeating similar arguments as in (148) and (150) yields

$$\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_6} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \right| \le \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \overline{\lambda}_{r_1},, \tag{153a}$$

$$\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_7} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \right| \le \frac{32}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \overline{\lambda}_{r_1}. \tag{153b}$$

Putting (139), (152), (153a) and (153b) together, one has

$$\left| \frac{1}{2\pi\mathrm{i}} \oint_{\gamma_1} \frac{\overline{\lambda}_{j_{k+1}} \mathrm{d}\eta}{(\eta - \overline{\lambda}_{j_1}) \cdots (\eta - \overline{\lambda}_{j_{k+1}})} \right| \le \frac{36}{\pi} \left(\frac{2}{\overline{\lambda}_{r_1} - \overline{\lambda}_{r_1+1}} \right)^k \overline{\lambda}_{r_1}. \tag{154}$$

In summary, we always have

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

This together with (132) finished the proof of (80b). □

C.4 Proof of Lemma 5

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

Proof of (85a). (85a) clearly holds for $i = 0$ due to the definition of μ. Now we consider the case $i \ge 1$. It is easy to verify that for any matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m_1 \times m_1}$,

$$(\boldsymbol{A} + \boldsymbol{B})^i = \boldsymbol{B}^i + \sum_{j=0}^{i-1} \boldsymbol{B}^j \boldsymbol{A} (\boldsymbol{A} + \boldsymbol{B})^{i-j-1}. \tag{155}$$

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

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

In view of (81), (82), (83a) and (83b), 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} \\
&\quad\le \sum_{j=0}^{i-1}\left\|\left[\mathcal{P}_{\textsf{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}_{\textsf{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\| \\
&\qquad + \sum_{j=0}^{i-1}\left\|\left[\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j\mathcal{P}_{\textsf{diag}}\left(\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right)\left[\mathcal{P}_{\textsf{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\| \\
&\qquad + \left\|\left[\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^i\boldsymbol{U}^\star\right\|_{2,\infty} \\
&\quad\le \sum_{j=0}^{i-1}\left\|\left[\mathcal{P}_{\textsf{off-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\|\left\|\mathcal{P}_{\textsf{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} \\
&\qquad + \sum_{j=0}^{i-1}\left\|\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right\|^j\left\|\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}^2\left\|\mathcal{P}_{\textsf{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} \\
&\qquad + \left\|\left[\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^i\boldsymbol{U}^\star\right\|_{2,\infty} \\
&\quad\le \sum_{j=0}^{i-1}\left(C_3\sqrt{\mu r}\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^j\omega_{\textsf{max}}\log m\right)\cdot C_5\sqrt{m_1}\omega_{\textsf{max}}\log m \\
&\qquad\qquad\cdot\left(3C_5\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^{i-j-1} \\
&\qquad + \sum_{j=0}^{i-1}\left(C_5\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^j\left(C_3\sqrt{\mu r}\omega_{\textsf{max}}\log m\right)^2\cdot\left(3C_5\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^{i-j-1} \\
&\qquad + C_3\sqrt{\frac{\mu r}{m_1}}\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^i \\
&\quad\le 3C_3\sqrt{\frac{\mu r}{m_1}}\left(C_3\left(\sqrt{m_1m_2}+m_1\right)\omega_{\textsf{max}}^2\log^2 m\right)^i, \qquad (156)
\end{aligned}$$

provided that $C_3 \ge 6C_5$.

Proof of (85b). When $i=0$, (85b) is a direct consequence of Lemma 2. For $i \ge 1$, similar to (156), one has

$$\left\|\boldsymbol{Z}_3^i\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}$$

$$
\begin{aligned}
&\le \sum_{j=0}^{i-1} \left\| \left[\mathcal{P}_{\textsf{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}_{\textsf{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{E}\boldsymbol{V}^\star \right\| \\
&\quad + \sum_{j=0}^{i-1} \left\| \left[\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^j \mathcal{P}_{\textsf{diag}}\left(\boldsymbol{E}\boldsymbol{V}^\star\boldsymbol{V}^{\star\top}\boldsymbol{E}^\top\right) \left[\mathcal{P}_{\textsf{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{E}\boldsymbol{V}^\star \right\| \\
&\quad + \left\| \left[\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^i \boldsymbol{E}\boldsymbol{V}^\star \right\|_{2,\infty} \\
&\le \sum_{j=0}^{i-1} \left\| \left[\mathcal{P}_{\textsf{off-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}_{\textsf{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}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right\|^j \left\|\boldsymbol{E}\boldsymbol{V}^\star\right\|_{2,\infty}^2 \left\|\mathcal{P}_{\textsf{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} \left\|\boldsymbol{E}\boldsymbol{V}^\star\right\| \\
&\quad + \left\| \left[\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{E}\boldsymbol{E}^\top\right)\right]^i \boldsymbol{E}\boldsymbol{V}^\star \right\|_{2,\infty} \\
&\le \sum_{j=0}^{i-1} \left(C_3\sqrt{\mu r} \left(C_3\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{max}}^2 \log^2 m\right)^j \omega_{\textsf{max}} \log m \right) \cdot \left(\sqrt{C_5}\sqrt{m_1}\omega_{\textsf{max}}\log m\right)^2 \\
&\qquad \cdot \left(3C_5\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{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_{\textsf{max}}^2 \log^2 m\right)^j \left(C_3\sqrt{\mu r}\omega_{\textsf{max}}\log m\right)^2 \cdot \left(3C_5\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{max}}^2 \log^2 m\right)^{i-j-1} \\
&\qquad \cdot \sqrt{C_5}\sqrt{m_1}\omega_{\textsf{max}}\log m \\
&\quad + C_3\sqrt{\mu r}\left(C_3\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{max}}^2 \log^2 m\right)^i \omega_{\textsf{max}}\log m \\
&\le 3C_3\sqrt{\mu r}\left(C_3\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{max}}^2 \log^2 m\right)^i \omega_{\textsf{max}}\log m, \qquad (157)
\end{aligned}
$$

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

Proof of (85c). We can directly use the same argument of (Zhou and Chen, 2023, Eqn. (121)) to prove (85c). We omit the details here for the sake of brevity.

Proof of (85d). Putting (70), (83a), (85a) and (85b) together, for all $0 \le i \le \log m$, we have

$$
\begin{aligned}
\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} \\
&\le \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} \\
&\le \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\| \\
&\le \sigma_{\bar r+1}^\star \left\|\boldsymbol{Z}_3^i \boldsymbol{U}^\star\right\|_{2,\infty}\left\|\boldsymbol{E}\boldsymbol{V}^\star\right\| + \sigma_{\bar 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\| \\
&\le \sigma_{\bar 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_{\textsf{max}}^2 \log^2 m\right)^i \cdot \sqrt{C_5}\sqrt{m_1}\omega_{\textsf{max}}\log m \\
&\quad + \sigma_{\bar r+1}^\star \omega_{\textsf{max}}\log m \cdot 3C_3\sqrt{\mu r}\left(C_3\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{max}}^2 \log^2 m\right)^i \omega_{\textsf{max}}\log m \\
&\quad + 3C_3\sqrt{\mu r}\left(C_3\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{max}}^2 \log^2 m\right)^i \omega_{\textsf{max}}\log m \cdot \sqrt{C_5}\sqrt{m_1}\omega_{\textsf{max}}\log m \\
&\le C_2\sqrt{\mu r}\left(\sigma_{\bar r+1}^\star + \sqrt{m_1}\omega_{\textsf{max}}\log m\right)\left(C_3\left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{max}}^2 \log^2 m\right)^i \omega_{\textsf{max}}\log m, \qquad (158)
\end{aligned}
$$

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

Proof of (85e)**.** By virtue of (70), (85a), (85c) and (87), we know that for all $0 \le i \le \log m$,

$$
\begin{aligned}
\left\|\boldsymbol{Z}_3^i \boldsymbol{Z}_2\right\|_{2,\infty} &\le \left\|\boldsymbol{Z}_3^i \left(\mathcal{P}_{\widetilde{\boldsymbol{U}}^{(1)}} \boldsymbol{U}^{\star(2)} \left(\boldsymbol{\Sigma}^{\star(2)}\right)^2 \boldsymbol{U}^{\star(2)\top} \mathcal{P}_{\left(\widetilde{\boldsymbol{U}}^{(1)}\right)_\perp} + \boldsymbol{U}^{\star(2)} \left(\boldsymbol{\Sigma}^{\star(2)}\right)^2 \boldsymbol{U}^{\star(2)\top} \mathcal{P}_{\widetilde{\boldsymbol{U}}^{(1)}}\right)\right\|_{2,\infty} \\
&\le \left\|\boldsymbol{Z}_3^i \widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty} \left\|\widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)}\right\| \left\|\boldsymbol{\Sigma}^{\star(2)}\right\|^2 + \left\|\boldsymbol{Z}_3^i \boldsymbol{U}^{\star(2)}\right\|_{2,\infty} \left\|\boldsymbol{\Sigma}^{\star(2)}\right\|^2 \left\|\widetilde{\boldsymbol{U}}^{(1)\top} \boldsymbol{U}^{\star(2)}\right\| \\
&\le \left\|\boldsymbol{Z}_3^i \widetilde{\boldsymbol{U}}^{(1)}\right\|_{2,\infty} \left\|\widetilde{\boldsymbol{U}}^{(1)\top} \left(\boldsymbol{U}_1^\star\right)_\perp\right\| \sigma_{\overline{r}+1}^{\star 2} + \left\|\boldsymbol{Z}_3^i \boldsymbol{U}^{\star(2)}\right\|_{2,\infty} \left\|\widetilde{\boldsymbol{U}}^{(1)\top} \left(\boldsymbol{U}_1^\star\right)_\perp\right\| \sigma_{\overline{r}+1}^{\star 2} \\
&\lesssim \sqrt{\frac{\mu r}{m_1}} \left(C_3 \left(\sqrt{m_1 m_2} + m_1\right) \omega_{\mathsf{max}}^2 \log^2 m\right)^i \cdot \frac{\sqrt{m_1} \omega_{\mathsf{max}} \log m}{\sigma_{\overline{r}}^\star} \sigma_{\overline{r}+1}^{\star 2} \\
&\le \sqrt{\frac{\mu r}{m_1}} \left(C_3 \left(\sqrt{m_1 m_2} + m_1\right) \omega_{\mathsf{max}}^2 \log^2 m\right)^i \cdot \sqrt{m_1} \omega_{\mathsf{max}} \sigma_{\overline{r}+1}^\star \log m \\
&= \sqrt{\mu r} \left(C_3 \left(\sqrt{m_1 m_2} + m_1\right) \omega_{\mathsf{max}}^2 \log^2 m\right)^i \omega_{\mathsf{max}} \sigma_{\overline{r}+1}^\star \log m.
\end{aligned} \tag{159}
$$

The third line holds since $\boldsymbol{U}^{\star(1)\top}\boldsymbol{U}^{\star(2)} = \boldsymbol{0}$ and the fifth line is due to the basic fact $\sigma_{\overline{r}}^\star \ge \sigma_{\overline{r}+1}^\star$. $\square$

D Proof of Theorem 6

D.1 Several notation

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

$$\boldsymbol{G}_{k+1}^0 := \boldsymbol{G}_k, \qquad \forall 0 \le k \le k_{\mathsf{max}}, \tag{160}$$

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

$$\boldsymbol{U}_{k+1}^t \boldsymbol{\Lambda}_{k+1}^t \boldsymbol{U}_{k+1}^{t\top} := \text{ the leading } r_k \text{ eigendecomposition of } \boldsymbol{G}_{k+1}^t, \tag{161}$$

and denote

$$\boldsymbol{G}_{k+1}^{t+1} := \mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{G}_{k+1}^t\right) + \mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{U}_{k+1}^t \boldsymbol{\Lambda}_{k+1}^t \boldsymbol{U}_{k+1}^{t\top}\right). \tag{162}$$

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

$$\boldsymbol{M}^{\mathsf{oracle}} = \widetilde{\boldsymbol{M}} + \boldsymbol{Z}_1 + \boldsymbol{Z}_2 + \boldsymbol{Z}_3 = \widetilde{\boldsymbol{M}} + \boldsymbol{Z}, \tag{163}$$

where $\widetilde{\boldsymbol{M}}, \boldsymbol{Z}_1, \boldsymbol{Z}_2$ and $\boldsymbol{Z}_3$ are defined in (77).

For notational convenience, we let

$$D_k^t = \left\|\boldsymbol{G}_k^t - \boldsymbol{M}^{\mathsf{oracle}}\right\|, \qquad F_k^t = \left\|\mathcal{P}_{\mathsf{diag}}\big(\boldsymbol{G}_k^t - \widetilde{\boldsymbol{M}}\big)\right\|, \qquad \text{and} \qquad L_k^t = \left\|\boldsymbol{G}_k^t - \widetilde{\boldsymbol{M}}\right\|. \tag{164}$$

In addition, we let

$$\widetilde{\boldsymbol{U}}_k = \widetilde{\boldsymbol{U}}_{:,1:r_k}, \qquad \boldsymbol{U}_k^{\mathsf{oracle}} = \boldsymbol{U}_{:,1:r_k}^{\mathsf{oracle}}, \qquad \text{and} \qquad \boldsymbol{U}_k^\star = \boldsymbol{U}_{:,1:r_k}^\star, \tag{165}$$

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:

$$
\begin{aligned}
\mathcal{E} = &\{(81) \text{ and } (82) \text{ hold for } 0 \le k \le \log n\} \cap \{(83a), (83b), (83c) \text{ and } (83d) \text{ hold}\} \\
&\cap \{(31a) \text{ and } (31b) \text{ for all } \ r' \in \mathcal{A}\}.
\end{aligned} \tag{166}
$$

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

$$\mathbb{P}\left(\mathcal{E}\right) \ge 1 - O\left(m^{-10}\right). \tag{167}$$

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

$$r_1 \in \mathcal{R}_1 \cap \mathcal{A}, \tag{168}$$

where

$$\mathcal{R}_1 := \left\{ r' \le r : \frac{\sigma_1\left(\boldsymbol{G}_0\right)}{\sigma_{r'}\left(\boldsymbol{G}_0\right)} \le 4 \quad \text{and} \quad \sigma_{r'}\left(\boldsymbol{G}_0\right) - \sigma_{r'+1}\left(\boldsymbol{G}_0\right) \ge \frac{1}{r}\sigma_{r'}\left(\boldsymbol{G}_0\right) \right\}. \tag{169}$$

and $\mathcal{A}$ is defined in (30). Noting that $\widetilde{\boldsymbol{U}} = [\widetilde{\boldsymbol{U}}^{(1)}\ \widetilde{\boldsymbol{U}}^{(2)}]$ and putting (83d) and (92) together, one has

$$\big\|\widetilde{\boldsymbol{U}}\big\|_{2,\infty} = \big\|\big[\widetilde{\boldsymbol{U}}^{(1)}\ \widetilde{\boldsymbol{U}}^{(2)}\big]\big\|_{2,\infty} \le \big\|\widetilde{\boldsymbol{U}}^{(1)}\big\|_{2,\infty} + \big\|\widetilde{\boldsymbol{U}}^{(2)}\big\|_{2,\infty} \le 4\sqrt{\frac{\mu r}{m_1}}. \tag{170}$$

In view of (100), (170) and the definition $\boldsymbol{G}_1^0 = \boldsymbol{G}_0 = \mathcal{P}_{\textsf{off-diag}}(\boldsymbol{M}) = \mathcal{P}_{\textsf{off-diag}}(\boldsymbol{M}^{\textsf{oracle}}) = \mathcal{P}_{\textsf{off-diag}}(\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^{\top} + \boldsymbol{Z})$.

$$\begin{aligned}
L_1^0 &= \big\|\boldsymbol{G}_0 - \widetilde{\boldsymbol{M}}\big\| \\
&= \big\|\mathcal{P}_{\textsf{diag}}\big(\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^{\top}\big) - \mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{Z}\right)\big\| \\
&\le \big\|\mathcal{P}_{\textsf{diag}}\big(\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^{\top}\big)\big\| + \left\|\mathcal{P}_{\textsf{off-diag}}\left(\boldsymbol{Z}\right)\right\| \\
&\le \big\|\widetilde{\boldsymbol{U}}\big\|_{2,\infty}^2 \big\|\widetilde{\boldsymbol{\Lambda}}\big\| + 2\left\|\boldsymbol{Z}\right\| \\
&\le 16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2 + C_2\left(\sqrt{m_1}\omega_{\textsf{max}}\log m \cdot \sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{max}}^2 \log^2 m\right).
\end{aligned} \tag{171}$$

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

$$\begin{aligned}
\left|\sigma_i\left(\boldsymbol{G}_0\right) - \widetilde{\sigma}_i^2\right| &= \big|\sigma_i\left(\boldsymbol{G}_0\right) - \sigma_i\big(\widetilde{\boldsymbol{M}}\big)\big| \\
&\le 16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2 + C_2\left(\sqrt{m_1}\omega_{\textsf{max}}\log m \cdot \sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{max}}^2 \log^2 m\right).
\end{aligned} \tag{172}$$

Moreover, (83a) tells us that

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

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

$$\left(1 - \frac{1}{Cr^2}\right)\sigma_i^{\star} \le \widetilde{\sigma}_i \le \left(1 + \frac{1}{Cr^2}\right)\sigma_i^{\star}, \qquad \forall i \in [\overline{r}]. \tag{173}$$

By virtue of (173) and the assumption $\sigma_{\overline{r}}^{\star} \ge C_0 r\big[(m_1 m_2)^{1/4} + r m_1^{1/2}\big]\omega_{\textsf{max}}\log m$, one has

$$\begin{aligned}
&16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2 + C_2\left(\sqrt{m_1}\omega_{\textsf{max}}\log m \cdot \sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\textsf{max}}^2 \log^2 m\right) \\
&\quad \le 16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2 + C_2\left(\sqrt{m_1}\omega_{\textsf{max}}\log m \cdot \sigma_{\overline{r}+1}^{\star} + \frac{1}{C_0^2 r^2}\sigma_{\overline{r}}^{\star 2}\right) \\
&\quad \le 16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2 + C_2\left(2\sqrt{m_1}\omega_{\textsf{max}}\log m \cdot \widetilde{\sigma}_{\overline{r}+1} + \frac{4}{C_0^2 r^2}\widetilde{\sigma}_{\overline{r}}^2\right) \\
&\quad \le 16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2 + C_2\left(\frac{2}{C_0 r^2}\widetilde{\sigma}_{\overline{r}}^2 + \frac{4}{C_0^2 r^2}\widetilde{\sigma}_{\overline{r}}^2\right) \\
&\quad \le \frac{16 c_1}{r^2}\widetilde{\sigma}_1^2 + \frac{1}{2Cr^2}\widetilde{\sigma}_{\overline{r}}^2
\end{aligned}$$

$$\leq \frac{1}{Cr^2}\widetilde{\sigma}_1^2, \tag{174}$$

provided that $C_0 \geq 8C \cdot C_2$ and $c_1 \leq \frac{1}{32C}$. Here, the fourth line comes from $\sigma_{\overline{r}}^{\star} \geq C_0 r\big[(m_1m_2)^{1/4} + rm_1^{1/2}\big]\omega_{\mathsf{max}}\log m$ and (173), the penultimate line uses the assumption $\mu \leq c_0 m_1/r^3$. Combining (172) and (174), we have

$$\sigma_1\left(\boldsymbol{G}_0\right) \geq \widetilde{\sigma}_1^2 - \frac{1}{Cr^2}\widetilde{\sigma}_1^2 > \max\left\{\frac{1}{2}\widetilde{\sigma}_1^2, \frac{1}{2}\sigma_1^{\star 2}, \tau\right\}, \tag{175}$$

where the last inequality holds due to (173) and $C_1^2/2 \geq C_\tau$.

Equipped with (172), (173), (174) and (175), $r_1 \in \mathcal{R}_1$ comes from 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 (173), (175) and the definition of $\mathcal{R}_1$, we know that

$$\sigma_{r_1}\left(\boldsymbol{G}_0\right) \geq \frac{1}{4}\sigma_1\left(\boldsymbol{G}_0\right) \geq \frac{1}{8}\sigma_1^{\star 2} \tag{176}$$

and consequently

$$\begin{aligned}
\sigma_{r_1}^{\star} &\geq \left(1 - \frac{C}{r^2}\right)\widetilde{\sigma}_{r_1} \geq \left(1 - \frac{C}{r^2}\right)\left[\sigma_{r_1}\left(\boldsymbol{G}_0\right) - \frac{1}{Cr^2}\widetilde{\sigma}_1^2\right]^{1/2} \geq \left(1 - \frac{C}{r^2}\right)\left(\frac{1}{8}\sigma_1^{\star 2} - \frac{4}{Cr^2}\sigma_1^{\star 2}\right)^{1/2} \\
&\geq \frac{1}{3}\sigma_1^{\star} > 2C_0 r\big[(m_1m_2)^{1/4} + rm_1^{1/2}\big]\omega_{\mathsf{max}}\log m.
\end{aligned} \tag{177}$$

Furthermore, inequality (172), (174) and (176) together imply that

$$\widetilde{\sigma}_{r_1}^2 \geq \sigma_{r_1}\left(\boldsymbol{G}_0\right) - \frac{1}{Cr^2}\widetilde{\sigma}_1^2 \geq \frac{1}{4}\sigma_1\left(\boldsymbol{G}_0\right) - \frac{1}{Cr^2}\widetilde{\sigma}_1^2 \geq \frac{1}{4}\left(\widetilde{\sigma}_1^2 - \frac{1}{Cr^2}\widetilde{\sigma}_1^2\right) - \frac{1}{Cr^2}\widetilde{\sigma}_1^2 \geq \frac{1}{5}\widetilde{\sigma}_1^2. \tag{178}$$

Inequalities (172), (174), (178) and the triangle inequality together show that

$$\begin{aligned}
\widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2 &\geq \sigma_{r_1}\left(\boldsymbol{G}_0\right) - \sigma_{r_1+1}\left(\boldsymbol{G}_0\right) - \left|\sigma_{r_1}\left(\boldsymbol{G}_0\right) - \widetilde{\sigma}_{r_1}^2\right| - \left|\sigma_{r_1+1}\left(\boldsymbol{G}_0\right) - \widetilde{\sigma}_{r_1+1}^2\right| \\
&\geq \frac{1}{r}\sigma_{r_1}\left(\boldsymbol{G}_0\right) - \frac{2}{Cr^2}\widetilde{\sigma}_1^2 \\
&\geq \frac{1}{r}\widetilde{\sigma}_{r_1}^2 - \frac{1}{r}\left|\sigma_{r_1}\left(\boldsymbol{G}_0\right) - \widetilde{\sigma}_{r_1}^2\right| - \frac{2}{Cr^2}\widetilde{\sigma}_1^2 \\
&\geq \frac{1}{r}\widetilde{\sigma}_{r_1}^2 - \frac{3}{Cr^2}\widetilde{\sigma}_1^2 \\
&\geq \frac{1}{r}\widetilde{\sigma}_{r_1}^2 - \frac{15}{Cr^2}\widetilde{\sigma}_{r_1}^2 \\
&\geq \frac{9}{10r}\widetilde{\sigma}_{r_1}^2.
\end{aligned} \tag{179}$$

Note that (70) together with (177) reveals that $r_1 \leq \overline{r}$, where $\overline{r}$ is the largest element in $\mathcal{A}$. Putting (83a), (96), (173) and the previous inequality together, we arrive that

$$\begin{aligned}
\sigma_{r_1}^{\star} - \sigma_{r_1+1}^{\star} &\geq \widetilde{\sigma}_{r_1} - \widetilde{\sigma}_{r_1+1} - \left|\widetilde{\sigma}_{r_1} - \sigma_{r_1}^{\star}\right| - \left|\widetilde{\sigma}_{r_1+1} - \sigma_{r_1+1}^{\star}\right| \\
&\geq \frac{1}{\widetilde{\sigma}_{r_1} + \widetilde{\sigma}_{r_1+1}}\left(\widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2\right) - \sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m - 2\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m \\
&\geq \frac{1}{2\widetilde{\sigma}_{r_1}} \cdot \frac{9}{10r}\widetilde{\sigma}_{r_1}^2 - \frac{1}{20r^2}\sigma_{r_1}^{\star} \\
&\geq \frac{9}{20r}\left(1 - \frac{1}{Cr}\right)\sigma_{r_1}^{\star} - \frac{1}{20r^2}\sigma_{r_1}^{\star} \\
&\geq \frac{1}{4r}\sigma_{r_1}^{\star}.
\end{aligned} \tag{180}$$

(177) taken together with (180) validates $r_1 \in \mathcal{A}$. Therefore, we have finished the proof of (168).

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 have

$$D_k^t = \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{G}_k^t - \boldsymbol{M}^{\mathsf{oracle}}\right)\right\|. \tag{181}$$

We will prove the following inequalities by induction:

$$F_1^t - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2 \le \frac{1}{e^t}\left(F_1^0 - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2\right), \tag{182a}$$

$$D_1^t \le F_1^t + 6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\overline{r}+1}^\star + C_3^2\mu r\omega_{\mathsf{max}}^2\log^2 m, \tag{182b}$$

$$\left\|\boldsymbol{U}_1^t\boldsymbol{U}_1^{t\top} - \boldsymbol{U}_1^{\mathsf{oracle}}\boldsymbol{U}_1^{\mathsf{oracle}\top}\right\| \le 2\frac{D_1^t}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)} \le \frac{1}{8}, \tag{182c}$$

$$\left\|\boldsymbol{U}_1^t\right\|_{2,\infty} \le \left\|\boldsymbol{U}_1^t\boldsymbol{U}_1^{t\top} - \boldsymbol{U}_1^{\mathsf{oracle}}\boldsymbol{U}_1^{\mathsf{oracle}\top}\right\| + \left\|\boldsymbol{U}_1^{\mathsf{oracle}}\right\|_{2,\infty} \le \frac{1}{4e}. \tag{182d}$$

Step 2.1: the base case ($t=0$) for (182a)-(182d). Note that (182a) automatically holds when $t=0$. Recalling that $\mathcal{P}_{\mathsf{diag}}(\boldsymbol{G}_1^0)=0$, Zhang et al. (2022, Lemma 1) taken together with (170) tells us that

$$F_1^0 = \left\|\mathcal{P}_{\mathsf{diag}}(\widetilde{\boldsymbol{M}})\right\| = \left\|\mathcal{P}_{\mathsf{diag}}(\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{U}}^\top)\right\| \le \left\|\widetilde{\boldsymbol{U}}\right\|_{2,\infty}^2\left\|\widetilde{\boldsymbol{\Lambda}}\right\| \le 16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2. \tag{183}$$

Furthermore, putting Lemma 2, (77), (83d), (87) and (183) together, we have

$$\begin{aligned}
D_1^0 &= \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)\right\| = \left\|\mathcal{P}_{\mathsf{diag}}(\widetilde{\boldsymbol{M}} + \boldsymbol{Z}_1 + \boldsymbol{Z}_2 + \boldsymbol{Z}_3)\right\| \\
&\le \left\|\mathcal{P}_{\mathsf{diag}}(\widetilde{\boldsymbol{M}})\right\| + \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{Z}_1\right)\right\| + \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{Z}_2\right)\right\| + \left\|\mathcal{P}_{\mathsf{diag}}\left(\boldsymbol{Z}_3\right)\right\| \\
&\le F_1^0 + \left\|\mathcal{P}_{\mathsf{off-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 \\
&\le F_1^0 + 2\left\|\boldsymbol{U}^{\star(2)}\right\|_{2,\infty}\left\|\boldsymbol{E}\boldsymbol{V}^{\star(2)}\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}^{\star(2)}\right\| + \left\|\boldsymbol{E}\boldsymbol{V}^{\star(2)}\right\|_{2,\infty}^2 + 2\left\|\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 \\
&\le F_1^0 + 2\sqrt{\frac{\mu r}{m_1}} \cdot C_3\sqrt{\mu r}\omega_{\mathsf{max}}\log m \cdot \sigma_{\overline{r}+1}^\star + \left(C_3\sqrt{\mu r}\omega_{\mathsf{max}}\log m\right)^2 + 4\sqrt{\frac{\mu r}{m_1}} \cdot C_3\frac{\sqrt{m}_1\omega_{\mathsf{max}}\log m}{\sigma_{\overline{r}}^\star}\sigma_{\overline{r}+1}^{\star 2} \\
&\le F_1^0 + 6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m}_1\omega_{\mathsf{max}}\log m \cdot \sigma_{\overline{r}+1}^\star + C_3^2\mu r\omega_{\mathsf{max}}^2\log^2 m,
\end{aligned} \tag{184}$$

which validates (182b) for $t=0$. Here, the last line holds since $\mu \le c_0 m_1/r^3$ and $\sigma_{\overline{r}}^\star \ge \sigma_{\overline{r}+1}^\star$. Combining (177), (178), (179) and (180), one has

$$\max\left\{\frac{\sigma_1^\star}{\sigma_{r_1}^\star}, \frac{\widetilde{\sigma}_1}{\widetilde{\sigma}_{r_1}}\right\} \le 3 \qquad \text{and} \qquad \min\left\{\frac{\widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2}{\widetilde{\sigma}_{r_1}^2}, \frac{\sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2}}{\sigma_{r_1}^{\star 2}}\right\} \ge \frac{1}{4r}. \tag{185}$$

Moreover, by virtue of (83a), (96) and the fact $r_1 \in \mathcal{A}$, we know that

$$\begin{aligned}
\left|\left(\widetilde{\sigma}_{r_1} - \widetilde{\sigma}_{r_1+1}\right) - \left(\sigma_{r_1}^\star - \sigma_{r_1+1}^\star\right)\right| &\le \left|\widetilde{\sigma}_{r_1} - \sigma_{r_1}^\star\right| + \left|\widetilde{\sigma}_{r_1+1} - \sigma_{r_1+1}^\star\right| \le \sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m + 2\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m \\
&\ll \frac{1}{r^2}\sigma_{r_1}^\star \lesssim \frac{1}{r}\left(\sigma_{r_1}^\star - \sigma_{r_1+1}^\star\right),
\end{aligned}$$

where the last inequality comes from (180). 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 (173), (100) and (185) gives us

$$\widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2 \asymp \sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2} \gg \|\boldsymbol{Z}\|.$$

Recalling that $\boldsymbol{M}^{\mathsf{oracle}} = \widetilde{\boldsymbol{M}} + \boldsymbol{Z}$, Weyl's inequality further tells us

$$\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) \asymp \widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2 \asymp \sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2} \gg \|\boldsymbol{Z}\|. \tag{186}$$

Note that $\boldsymbol{U}_1^0$ (resp. $\boldsymbol{U}_1^{\mathsf{oracle}}$) is the rank-$r$ leading eigenspace of $\boldsymbol{G}_0$ (resp. $\boldsymbol{M}^{\mathsf{oracle}}$). We know from the Davis-Kahan Theorem (Chen et al., 2021a, Theorem 2.7), (184), (185) and (186) that

$$\begin{aligned}
\left\|\boldsymbol{U}_1^0\boldsymbol{U}_1^{0\top} - \boldsymbol{U}_1^{\mathsf{oracle}}\boldsymbol{U}_1^{\mathsf{oracle}\top}\right\| &\le 2\frac{\left\|\boldsymbol{G}_0 - \boldsymbol{M}^{\mathsf{oracle}}\right\|}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)} = 2\frac{D_1^0}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)} \\
&\overset{(186)}{\lesssim} \frac{\frac{\mu r}{m_1}\widetilde{\sigma}_1^2}{\widetilde{\sigma}_{r_1}^2 - \widetilde{\sigma}_{r_1+1}^2} + \frac{\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star}{\sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2}} + \frac{\mu r\omega_{\mathsf{max}}^2\log^2 m}{\sigma_{r_1}^{\star 2} - \sigma_{r_1+1}^{\star 2}} \\
&\overset{(185)}{\lesssim} \frac{\frac{\mu r}{m_1}\widetilde{\sigma}_1^2}{\widetilde{\sigma}_{r_1}^2/r} + \frac{\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma_{\bar{r}+1}^\star}{\sigma_{r_1}^{\star 2}/r} + \frac{\mu r\omega_{\mathsf{max}}^2\log^2 m}{\sigma_{r_1}^{\star 2}/r} \\
&\overset{(185)}{\lesssim} \frac{\mu r^2}{m_1} + \frac{\sqrt{\frac{\mu r}{m_1}}r\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma_{r_1}^\star} + \frac{\mu r\omega_{\mathsf{max}}^2\log^2 m}{\sigma_{r_1}^{\star 2}/r} \\
&\ll \sqrt{\frac{\mu r}{m_1}} \le \frac{1}{8},
\end{aligned} \tag{187}$$

which proves (182c) for $t = 0$. Here, the last inequality is due to $\mu \le c_0 m_1/r^3$ and $\sigma_{r_1}^\star \ge \sigma_{\bar{r}}^\star \ge C_0 r\big[(m_1m_2)^{1/4} + rm_1^{1/2}\big]\omega_{\mathsf{max}}\log m$. Inequality (31b) and the fact $r_1 \in \mathcal{A}$ together imply that

$$\left\|\boldsymbol{U}_1^{\mathsf{oracle}}\right\|_{2,\infty} = \left\|\boldsymbol{U}_1^{\mathsf{oracle}}\boldsymbol{U}_1^{\mathsf{oracle}\top}\right\|_{2,\infty} \le \left\|\boldsymbol{U}_1^{\mathsf{oracle}}\boldsymbol{U}_1^{\mathsf{oracle}\top} - \boldsymbol{U}_1^\star\boldsymbol{U}_1^{\star\top}\right\|_{2,\infty} + \|\boldsymbol{U}_1^\star\|_{2,\infty} \le 2\sqrt{\frac{\mu r^3}{m_1}}. \tag{188}$$

Combining (187) and (188), one can further obtain that

$$\left\|\boldsymbol{U}_1^0\right\|_{2,\infty} = \left\|\boldsymbol{U}_1^0\boldsymbol{U}_1^{0\top}\right\|_{2,\infty} \le \left\|\boldsymbol{U}_1^0\boldsymbol{U}_1^{0\top} - \boldsymbol{U}_1^{\mathsf{oracle}}\boldsymbol{U}_1^{\mathsf{oracle}\top}\right\| + \left\|\boldsymbol{U}_1^{\mathsf{oracle}}\right\|_{2,\infty} \le 3\sqrt{\frac{\mu r^3}{m_1}} \le \frac{1}{4e}, \tag{189}$$

i.e., (182d) holds for $t = 0$.

Step 2.2: induction step ($t > 0$) for (182a)-(182d). Suppose that (182a)-(182d) hold for $t = t'$. We aim to show that they also 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

$$\begin{aligned}
F_1^{t'+1} &= \left\|\mathcal{P}_{\mathsf{diag}}\big(\boldsymbol{G}_1^{t'+1} - \widetilde{\boldsymbol{M}}\big)\right\| \\
&= \left\|\mathcal{P}_{\mathsf{diag}}\big(\mathcal{P}_{\boldsymbol{U}_1^{t'}}\boldsymbol{G}_1^{t'} - \widetilde{\boldsymbol{M}}\big)\right\| \\
&\le \left\|\mathcal{P}_{\mathsf{diag}}\left(\mathcal{P}_{\boldsymbol{U}_1^{t'}}\big(\boldsymbol{G}_1^{t'} - \widetilde{\boldsymbol{M}}\big)\right)\right\| + \left\|\mathcal{P}_{\mathsf{diag}}\left(\mathcal{P}_{(\boldsymbol{U}_1^{t'})_\perp}\widetilde{\boldsymbol{M}}\mathcal{P}_{\widetilde{\boldsymbol{U}}}\right)\right\| \\
&\le \left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty}\left\|\boldsymbol{G}_1^{t'} - \widetilde{\boldsymbol{M}}\right\| + \|\widetilde{\boldsymbol{U}}\|_{2,\infty}\left\|\big(\boldsymbol{U}_1^{t'}\big)_\perp\widetilde{\boldsymbol{M}}\right\| \\
&\le \left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty} L_1^{t'} + 4\sqrt{\frac{\mu r}{m_1}}\left(\left\|\big(\boldsymbol{U}_1^{t'}\big)_\perp\mathcal{P}_{\widetilde{\boldsymbol{U}}_1}\widetilde{\boldsymbol{M}}\right\| + \left\|\mathcal{P}_{\widetilde{\boldsymbol{U}}_{:,r_1+1:r}}\widetilde{\boldsymbol{M}}\right\|\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty} L_1^{t'} + 4\sqrt{\frac{\mu r}{m_1}}\left(2\left\|\boldsymbol{G}_1^{t'} - \mathcal{P}_{\widetilde{\boldsymbol{U}}_1}\widetilde{\boldsymbol{M}}\right\| + \left\|\mathcal{P}_{\widetilde{\boldsymbol{U}}_{:,r_1+1:r}}\widetilde{\boldsymbol{M}}\right\|\right) \\
&\leq \left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty} L_1^{t'} + 4\sqrt{\frac{\mu r}{m_1}}\left(2\left\|\boldsymbol{G}_1^{t'} - \widetilde{\boldsymbol{M}}\right\| + 3\left\|\mathcal{P}_{\widetilde{\boldsymbol{U}}_{:,r_1+1:r}}\widetilde{\boldsymbol{M}}\right\|\right) \\
&\leq \left(\left\|\boldsymbol{U}_1^{t'}\right\|_{2,\infty} + 8\sqrt{\frac{\mu r}{m_1}}\right) L_1^{t'} + 12\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2. \qquad (190)
\end{aligned}
$$

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 (170); the sixth line applies Zhou and Chen (2023, Lemma 8); the penultimate line invokes the triangle inequality. Note that

$$
\begin{aligned}
L_1^{t'} &\leq F_1^{t'} + \left\|\mathcal{P}_{\mathsf{off\text{-}diag}}(\boldsymbol{G}_1^{t'+1} - \widetilde{\boldsymbol{M}})\right\| = F_1^{t'} + \left\|\mathcal{P}_{\mathsf{off\text{-}diag}}(\boldsymbol{M}^{\mathsf{oracle}} - \widetilde{\boldsymbol{M}})\right\| \\
&\leq F_1^{t'} + \left\|\mathcal{P}_{\mathsf{off\text{-}diag}}(\boldsymbol{Z})\right\| \leq F_1^{t'} + \|\boldsymbol{Z}\| + \left\|\mathcal{P}_{\mathsf{diag}}(\boldsymbol{Z})\right\| \leq F_1^{t'} + 2\|\boldsymbol{Z}\|.
\end{aligned}
$$

Inequality (190) taken together with the previous inequality leads us to

$$
\begin{aligned}
F_1^{t'+1} &\leq \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{(182d)}{\leq} \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{(182c)\text{ and }(188)}{\leq} \frac{1}{2e}F_1^{t'} + 2\left(2\frac{D_1^{t'}}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})} + 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{(182b)}{\leq} \frac{1}{2e}F_1^{t'} + 2\left(2\frac{F_1^{t'}}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})} + 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_{\bar{r}+1}^{\star} + C_3^2\mu r\omega_{\mathsf{max}}^2\log^2 m}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})}\|\boldsymbol{Z}\| \\
&\overset{(185)\text{ and }(186)}{\leq} \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_{\bar{r}+1}^{\star} + \mu r\omega_{\mathsf{max}}^2\log^2 m}{\sigma_{r_1}^{\star 2}/r}\|\boldsymbol{Z}\| \\
&\leq \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, \qquad (191)
\end{aligned}
$$

where the last inequality is a consequence of $\sigma_{r_1}^{\star} \geq \sigma_{\bar{r}}^{\star} \geq C_0 r\big[(m_1m_2)^{1/4} + rm_1^{1/2}\big]\omega_{\mathsf{max}}\log m$. Then one immediately has

$$
\begin{aligned}
F_1^{t'+1} - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2 &\leq \frac{1}{e}\left(F_1^{t'} - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2\right) \\
&\leq \frac{1}{e^{t'+1}}\left(F_1^{0} - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2\right),
\end{aligned}
$$

which confirms that (182a) holds for $t = t' + 1$.

In addition, we can prove (182b) for $t = t' + 1$ by using the same argument as in (184). Combining (182a), (182b) for $t = t' + 1$ and Weyl's inequality, we further have

$$
\begin{aligned}
&\left\|\boldsymbol{U}_1^{t'+1}\boldsymbol{U}_1^{t'+1\top} - \boldsymbol{U}_1^{\mathsf{oracle}}\boldsymbol{U}_1^{\mathsf{oracle}\top}\right\| \\
&\quad\leq 2\frac{D_1^{t'+1}}{\lambda_{r_1}(\boldsymbol{M}^{\mathsf{oracle}}) - \lambda_{r_1+1}(\boldsymbol{M}^{\mathsf{oracle}})}
\end{aligned}
$$

$$
\begin{aligned}
&\leq 2\frac{F_1^{t'+1}}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)-\lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}+2\frac{6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m}_1\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+C_3^2\mu r\omega_{\mathsf{max}}^2\log^2 m}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)-\lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}\\
&=2\frac{F_1^{t'+1}-40\sqrt{\frac{\mu r^3}{m_1}}\left\|\boldsymbol{Z}\right\|-20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)-\lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}+2\frac{6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m}_1\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+C_3^2\mu r\omega_{\mathsf{max}}^2\log^2 m}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)-\lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}\\
&\quad+\frac{80\sqrt{\frac{\mu r^3}{m_1}}\left\|\boldsymbol{Z}\right\|+40\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)-\lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}\\
&\overset{(186)}{\lesssim}\frac{1}{e^{t'+1}}\frac{F_1^0}{\lambda_{r_1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)-\lambda_{r_1+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}+\frac{\sqrt{\frac{\mu r^3}{m_1}}\left\|\boldsymbol{Z}\right\|+\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2}{\widetilde{\sigma}_{r_1}^2-\widetilde{\sigma}_{r_1+1}^2}\\
&\quad 2\frac{6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m}_1\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+C_3^2\mu r\omega_{\mathsf{max}}^2\log^2 m}{\sigma_{r_1}^{\star 2}-\sigma_{r_1+1}^{\star 2}}\\
&\overset{(183),(185)\text{ and }(186)}{\ll}\sqrt{\frac{\mu r^3}{m_1}}\ll\frac{1}{8},
\end{aligned}
$$

which validates (182c) for $t=t'+1$.

Putting the previous inequality and (188) together, one can prove that (182d) 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 finished the proof of the induction step for (182a) - (182d).

Step 3: bounding D_k^t for $k>1$ After providing upper bounds for $\{D_1^t\}$, we now move to the quantities $\{D_k^t\}_{k>1}$. By setting

$$
t_1\geq\log\left(C\frac{\sigma_1^{\star 2}}{\sigma_{r_1+1}^{\star 2}+\omega_{\mathsf{max}}^2}\right),
$$

we know that

$$
\begin{aligned}
F_2^0=F_1^{t_1}&\overset{(182a)}{\leq}40\sqrt{\frac{\mu r^3}{m_1}}\left\|\boldsymbol{Z}\right\|+20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2+\frac{1}{e^{t_1}}F_1^0\\
&\overset{(184)}{\leq}40\sqrt{\frac{\mu r^3}{m_1}}\left\|\boldsymbol{Z}\right\|+20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2+\frac{\sigma_{r_1+1}^{\star 2}+\omega_{\mathsf{max}}^2}{C\sigma_1^{\star 2}}\cdot 16\frac{\mu r}{m_1}\widetilde{\sigma}_1^2\\
&\overset{(100)\text{ and }(185)}{\leq}40C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)\\
&\qquad+20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2+\frac{1}{C}\sqrt{\frac{\mu r}{m_1}}\left(\omega_{\mathsf{max}}^2+\sigma_{r_1+1}^{\star 2}\right)\\
&\leq 41C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m\cdot\sigma^\star_{\overline{r}+1}+\left(\sqrt{m_1m_2}+m_1\right)\omega_{\mathsf{max}}^2\log^2 m\right)+21\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_1+1}^2.
\end{aligned}
\tag{192}
$$

Here, the last line makes use of the following inequality:

$$
\sigma_{r_1+1}^{\star 2}\overset{(83a)\text{ and }(96)}{\leq}\left(\widetilde{\sigma}_{r_1+1}+2\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m\right)^2\overset{\text{Cauchy-Schwarz}}{\leq}2\widetilde{\sigma}_{r_1+1}^2+8C_5m_1\omega_{\mathsf{max}}^2\log^2 m.
$$

We define

$$
\mathcal{R}_k:=\left\{r':\frac{\sigma_{r_{k-1}+1}\left(\boldsymbol{G}_0\right)}{\sigma_{r'}\left(\boldsymbol{G}_0\right)}\leq 4\quad\text{and}\quad\sigma_{r'}\left(\boldsymbol{G}_0\right)-\sigma_{r'+1}\left(\boldsymbol{G}_0\right)\geq\frac{1}{r}\sigma_{r'}\left(\boldsymbol{G}_0\right)\right\},
\tag{193}
$$

Choosing the numbers of iterations $\{t_i\}$ as in (33a) and (33b) and repeating similar arguments as in (168), (182a) - (182d), (185), (186) and (192), we know that for all $1 \le k \le k_{\mathsf{max}}$ and $1 \le t \le t_k$,

$$r_k \in \mathcal{R}_k \cap \mathcal{A}, \tag{194a}$$

$$\max\left\{\frac{\sigma^\star_{r_{k-1}+1}}{\sigma^\star_{r_k}}, \frac{\widetilde{\sigma}_{r_{k-1}+1}}{\widetilde{\sigma}_{r_k}}\right\} \le 3 \qquad \text{and} \qquad \min\left\{\frac{\widetilde{\sigma}^2_{r_k} - \widetilde{\sigma}^2_{r_k+1}}{\widetilde{\sigma}^2_{r_k}}, \frac{\sigma^{\star 2}_{r_k} - \sigma^{\star 2}_{r_k+1}}{\sigma^{\star 2}_{r_k}}\right\} \ge \frac{1}{4r}, \tag{194b}$$

$$\lambda_{r_k}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_k+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) \asymp \widetilde{\sigma}^2_{r_k} - \widetilde{\sigma}^2_{r_k+1} \asymp \sigma^{\star 2}_{r_k} - \sigma^{\star 2}_{r_k+1} \gg \|\boldsymbol{Z}\|, \tag{194c}$$

$$F_k^t - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}^2_{r_k+1} \le \frac{1}{e^t}\left(F_k^0 - 40\sqrt{\frac{\mu r^3}{m_1}}\|\boldsymbol{Z}\| - 20\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}^2_{r_k+1}\right), \tag{194d}$$

$$D_k^t \le F_k^t + 6C_3\sqrt{\frac{\mu r}{m_1}}\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma^\star_{\widetilde{r}+1} + C_3^2\mu r\omega^2_{\mathsf{max}}\log^2 m, \tag{194e}$$

$$\left\|\boldsymbol{U}_k^t\boldsymbol{U}_k^{t\top} - \boldsymbol{U}_k^{\mathsf{oracle}}\boldsymbol{U}_k^{\mathsf{oracle}\top}\right\| \le 2\frac{D_k^t}{\lambda_{r_k}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_k+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)} \le \frac{1}{8}, \tag{194f}$$

$$\left\|\boldsymbol{U}_k^t\right\|_{2,\infty} \le \left\|\boldsymbol{U}_k^t\boldsymbol{U}_k^{t\top} - \boldsymbol{U}_k^{\mathsf{oracle}}\boldsymbol{U}_k^{\mathsf{oracle}\top}\right\| + \left\|\boldsymbol{U}_k^{\mathsf{oracle}}\right\|_{2,\infty} \le \frac{1}{4e}. \tag{194g}$$

$$F_{k+1}^0 = F_k^{t_k} \le 41C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma^\star_{\widetilde{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right) + 21\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}^2_{r_k+1}. \tag{194h}$$

Taking $k = k_{\mathsf{max}}$ in (194h) yields that

$$F_{k_{\mathsf{max}}}^{t_{k_{\mathsf{max}}}} \le 41C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma^\star_{\widetilde{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right) + 21\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1}. \tag{195}$$

This together with (194e) implies that

$$D_{k_{\mathsf{max}}}^{t_{k_{\mathsf{max}}}} \le 42C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma^\star_{\widetilde{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right) + 21\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1}. \tag{196}$$

Recall that $r_{k_{\mathsf{max}}}$ satisfies $r_{k_{\mathsf{max}}} = r$ or $\sigma_{r_{k_{\mathsf{max}}}+1}(\boldsymbol{G}_{k_{\mathsf{max}}}) = \sigma_{r_{k_{\mathsf{max}}}+1}(\boldsymbol{G}_{k_{\mathsf{max}}}^{t_{k_{\mathsf{max}}}}) \le \tau$.

1. If $r_{k_{\mathsf{max}}} = r$, then we have
$$\widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1} = \widetilde{\sigma}^2_{r+1} = 0.$$

2. If $\sigma_{r_{k_{\mathsf{max}}}+1}(\boldsymbol{G}_{k_{\mathsf{max}}}) \le \tau$, then Weyl's inequality and (195) together show that

$$\begin{aligned}
\widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1} &= \sigma_{r_{k_{\mathsf{max}}}+1}(\widetilde{\boldsymbol{M}}) \le \sigma_{r_{k_{\mathsf{max}}}+1}\left(\boldsymbol{G}_{k_{\mathsf{max}}}^{t_{k_{\mathsf{max}}}}\right) + \left\|\widetilde{\boldsymbol{M}} - \boldsymbol{G}_{k_{\mathsf{max}}}^{t_{k_{\mathsf{max}}}}\right\| \\
&= \sigma_{r_{k_{\mathsf{max}}}+1}\left(\boldsymbol{G}_{k_{\mathsf{max}}}^{t_{k_{\mathsf{max}}}}\right) + F_{k_{\mathsf{max}}}^{t_{k_{\mathsf{max}}}} \\
&\le \tau + 41C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma^\star_{\widetilde{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right) + 21\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1} \\
&\le \tau + 41C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma^\star_{\widetilde{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right) + \frac{1}{2}\widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1},
\end{aligned}$$

   which further gives us

$$\widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1} \le 2\tau + 82C_2\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma^\star_{\widetilde{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\mathsf{max}}\log^2 m\right) \le 3\tau. \tag{197}$$

Therefore, inequality (197) always holds.

Step 4: proving (34a). We know from (194a) that $r_{k_{\max}} \in \mathcal{A}$. (31b) tell us that

$$\left\|\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}}\left(\frac{r^2\sqrt{m_1}\omega_{\max}\log m}{\sigma_{r_{k_{\max}}}^{\star}} + \frac{r^2\sqrt{m_1m_2}\omega_{\max}^2\log^2 m}{\sigma_{r_{k_{\max}}}^{\star 2}}\right) \leq \sqrt{\frac{\mu r^3}{m_1}}.$$

In view of (196), (194b), (194c) and (194g), we have

$$\begin{aligned}
&\left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right\| \\
&\quad\lesssim \frac{D_{k_{\max}}^{t_{k_{\max}}}}{\lambda_{r_{k_{\max}}}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_{k_{\max}}+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)} \\
&\quad\lesssim \frac{\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1m_2}+m_1\right)\omega_{\max}^2\log^2 m\right)}{\sigma_{r_{k_{\max}}}^{\star 2} - \sigma_{r_{k_{\max}}+1}^{\star 2}} + \frac{\sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_{k_{\max}}+1}^{2}}{\widetilde{\sigma}_{r_{k_{\max}}}^{2} - \widetilde{\sigma}_{r_{k_{\max}}+1}^{2}} \\
&\quad\lesssim \frac{\sqrt{\frac{\mu r^3}{m_1}}r\left(\sqrt{m_1}\omega_{\max}\log m\cdot\sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1m_2}+m_1\right)\omega_{\max}^2\log^2 m\right)}{\sigma_{r_{k_{\max}}}^{\star 2}} + \frac{\sqrt{\frac{\mu r^3}{m_1}}\widetilde{\sigma}_{r_{k_{\max}}+1}^{2}}{\widetilde{\sigma}_{r_{k_{\max}}}^{2}} \\
&\quad\lesssim \sqrt{\frac{\mu r^3}{m_1}}.
\end{aligned}\tag{198}$$

Here, the last inequality holds since $\widetilde{\sigma}_{r_{k_{\max}}} \geq \widetilde{\sigma}_{r_{k_{\max}}+1}$ and $\sigma_{r_{k_{\max}}}^{\star} \geq \sigma_{\overline{r}}^{\star} \geq C_0 r\left[(m_1m_2)^{1/4} + rm_1^{1/2}\right]\omega_{\max}\log m$. Combining the previous two inequalities, one has

$$\begin{aligned}
\left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right\|_{2,\infty} &\leq \left\|\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right\|_{2,\infty} + \left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right\| \\
&\lesssim \sqrt{\frac{\mu r^3}{m_1}},
\end{aligned}\tag{199}$$

which validates (34a).

Step 5: proving (34b) and (34c). Note that

$$\begin{aligned}
&\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\boldsymbol{X}^{\star}\right\|_{2,\infty} \\
&\quad= \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\left(\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top} + \boldsymbol{U}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\right)\right\|_{2,\infty} \\
&\quad= \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty} \\
&\qquad+ \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\boldsymbol{U}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\right\|_{2,\infty} \\
&\quad\leq \underbrace{\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^{\top}\right)\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty}}_{=:\alpha_1} \\
&\qquad+ \underbrace{\left\|\left(\widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty}}_{=:\alpha_2} \\
&\qquad+ \underbrace{\left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top} - \boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right\|_{2,\infty}\sigma_{\overline{r}+1}^{\star}}_{=:\alpha_3},
\end{aligned}\tag{200}$$

where the last line holds due to $\|\boldsymbol{U}^{\star(2)}\boldsymbol{\Sigma}^{\star(2)}\boldsymbol{V}^{\star(2)\top}\| = \|\boldsymbol{\Sigma}^{\star(2)}\| = \sigma_{\overline{r}+1}^{\star}$. 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 (100), one has

$$\lambda_{r_{k_{\max}}}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) \le \widetilde{\sigma}_{r_{k_{\max}}}^2 + \|\boldsymbol{Z}\| \asymp \widetilde{\sigma}_{r_{k_{\max}}}^2 \asymp \sigma_{r_{k_{\max}}}^{\star 2}. \tag{201}$$

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

$$\begin{aligned}
&\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\boldsymbol{M}^{\mathsf{oracle}}\right\|_{2,\infty} \\
&\quad \le \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\boldsymbol{M}^{\mathsf{oracle}}\right\| \\
&\quad \lesssim \frac{\lambda_{r_{k_{\max}}}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}{\lambda_{r_{k_{\max}}}\left(\boldsymbol{M}^{\mathsf{oracle}}\right) - \lambda_{r_{k_{\max}}+1}\left(\boldsymbol{M}^{\mathsf{oracle}}\right)}\left\|\boldsymbol{G}_{k_{\max}}^{t_{k_{\max}}} - \boldsymbol{M}^{\mathsf{oracle}}\right\| \\
&\quad \overset{(194c)\text{ and }(201)}{\lesssim} \frac{\sigma_{r_{k_{\max}}}^{\star 2}}{\sigma_{r_{k_{\max}}}^{\star 2} - \sigma_{r_{k_{\max}}+1}^{\star 2}} D_{k_{\max}}^{t_{k_{\max}}} \\
&\quad \overset{(194b)\text{ and }(196)}{\lesssim} r\left[\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right) + \sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_{k_{\max}}+1}^2\right].
\end{aligned} \tag{202}$$

Combining (100), (198) and (202) leads to

$$\begin{aligned}
&\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\widetilde{\boldsymbol{M}}\right\|_{2,\infty} \\
&\quad \le \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\boldsymbol{M}^{\mathsf{oracle}}\right\|_{2,\infty} + \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\boldsymbol{Z}\right\|_{2,\infty} \\
&\quad \le \left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right)\boldsymbol{M}^{\mathsf{oracle}}\right\|_{2,\infty} + \left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top}\right\|_{2,\infty}\|\boldsymbol{Z}\| \\
&\quad \lesssim r\left[\sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right) + \sqrt{\frac{\mu r}{m_1}}\widetilde{\sigma}_{r_{k_{\max}}+1}^2\right] \\
&\qquad + \sqrt{\frac{\mu r^3}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right) \\
&\quad \lesssim \sqrt{\frac{\mu r^5}{m_1}}\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right) + \sqrt{\frac{\mu r^3}{m_1}}\widetilde{\sigma}_{r_{k_{\max}}+1}^2.
\end{aligned} \tag{203}$$

Step 5.1.2: bounding $\|(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^\top - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top)\widetilde{\boldsymbol{M}}\|_{2,\infty}$. Recalling that (100) holds, We invoke Lemma 1 to obtain

$$\begin{aligned}
&\left\|\left(\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right)\widetilde{\boldsymbol{M}}\right\|_{2,\infty} \\
&\quad \le \frac{40}{\pi}\widetilde{\sigma}_{r_{k_{\max}}}^2 \sum_{k\ge 1}\frac{2^k}{\left(\widetilde{\sigma}_{r_{k_{\max}}}^2 - \widetilde{\sigma}_{r_{k_{\max}}+1}^2\right)^k}\sum_{\substack{0\le j_1,\dots,j_{k+1}\le r \\ (j_1,\dots,j_{k+1})^\top \ne \boldsymbol{0}_{k+1}}}\left\|\widetilde{\boldsymbol{P}}_{j_1}\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_2}\boldsymbol{Z}\cdots\boldsymbol{Z}\widetilde{\boldsymbol{P}}_{j_{k+1}}\right\|_{2,\infty}.
\end{aligned} \tag{204}$$

Here, we recall that $\widetilde{\boldsymbol{P}}_j = \widetilde{\boldsymbol{u}}_j\widetilde{\boldsymbol{u}}_j^\top$ for $1 \le j \le r$ and $\widetilde{\boldsymbol{P}}_0 = \widetilde{\boldsymbol{U}}_\perp\widetilde{\boldsymbol{U}}_\perp^\top$. Repeat similar arguments as in (113) to yield that

$$\begin{aligned}
&\left\|\left(\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}}\boldsymbol{U}_{k_{\max}}^{\mathsf{oracle}\top} - \widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^\top\right)\widetilde{\boldsymbol{M}}\right\|_{2,\infty} \\
&\quad \lesssim \sqrt{\frac{\mu r^3}{m_1}}\frac{r\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)}{\sigma_{r_{k_{\max}}}^{\star 2} - \sigma_{r_{k_{\max}}+1}^{\star 2}} \cdot \widetilde{\sigma}_{r_{k_{\max}}}^2 \\
&\quad \overset{(194b)\text{ and }(173)}{\lesssim} \sqrt{\frac{\mu r^3}{m_1}}\frac{r\left(\sqrt{m_1}\omega_{\mathsf{max}}\log m \cdot \sigma_{\overline{r}+1}^{\star} + \left(\sqrt{m_1 m_2} + m_1\right)\omega_{\mathsf{max}}^2 \log^2 m\right)}{\sigma_{r_{k_{\max}}}^{\star 2}/r} \cdot \sigma_{r_{k_{\max}}}^{\star 2}
\end{aligned}$$

$$= \sqrt{\frac{\mu r^3}{m_1}} r^2 \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma^\star_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right) \omega^2_{\mathsf{max}} \log^2 m\right). \tag{205}$$

(203) taken together with (205) and the triangle inequality shows that

$$\begin{aligned}
&\big\|\big(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top\big)\widetilde{\boldsymbol{M}}\big\|_{2,\infty} \\
&\quad\lesssim \sqrt{\frac{\mu r^3}{m_1}} r^2 \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma^\star_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right) \omega^2_{\mathsf{max}} \log^2 m\right) + \sqrt{\frac{\mu r^3}{m_1}} \widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1}.
\end{aligned} \tag{206}$$

Step 5.1.3: bounding α_1. Equipped with (206), we are now ready to bound α_1. Recall that

$$\widetilde{\boldsymbol{M}} = \widetilde{\boldsymbol{U}}^{(1)}\big(\widetilde{\boldsymbol{\Sigma}}^{(1)}\big)^2\widetilde{\boldsymbol{U}}^{(1)\top} + \widetilde{\boldsymbol{U}}^{(2)}\big(\widetilde{\boldsymbol{\Sigma}}^{(2)}\big)^2\widetilde{\boldsymbol{U}}^{(2)\top},$$

where $\widetilde{\boldsymbol{U}}^{(1)}$ and $\widetilde{\boldsymbol{\Sigma}}^{(1)}$ (resp. $\widetilde{\boldsymbol{U}}^{(2)}$ and $\widetilde{\boldsymbol{\Sigma}}^{(2)}$) are defined in (75) (resp. (88)). In view of (198) and (206), one can obtain that

$$\begin{aligned}
&\big\|\big(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top\big)\widetilde{\boldsymbol{U}}^{(1)}\widetilde{\boldsymbol{\Sigma}}^{(1)}\big\|_{2,\infty} \\
&\quad\le \Big\|\big(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top\big)\widetilde{\boldsymbol{U}}^{(1)}\big(\widetilde{\boldsymbol{\Sigma}}^{(1)}\big)^2\widetilde{\boldsymbol{U}}^{(1)\top}\Big\|_{2,\infty}\Big\|\big(\widetilde{\boldsymbol{\Sigma}}^{(1)}\big)^{-1}\Big\| \\
&\quad\overset{(i)}{\lesssim} \left(\big\|\big(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top\big)\widetilde{\boldsymbol{M}}\big\|_{2,\infty} + \Big\|\big(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top\big)\widetilde{\boldsymbol{U}}^{(2)}\big(\widetilde{\boldsymbol{\Sigma}}^{(2)}\big)^2\widetilde{\boldsymbol{U}}^{(2)\top}\Big\|_{2,\infty}\right)\frac{1}{\sigma^\star_{\bar{r}}} \\
&\quad\le \left(\big\|\big(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top\big)\widetilde{\boldsymbol{M}}\big\|_{2,\infty} + \Big\|\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top\Big\|_{2,\infty}\big\|\widetilde{\boldsymbol{\Sigma}}^{(2)}\big\|^2\right)\frac{1}{\sigma^\star_{\bar{r}}} \\
&\quad\lesssim \Bigg(\sqrt{\frac{\mu r^3}{m_1}} r^2 \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma^\star_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right) \omega^2_{\mathsf{max}} \log^2 m\right) + \sqrt{\frac{\mu r^3}{m_1}} \widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1} \\
&\qquad + \sqrt{\frac{\mu r^3}{m_1}}\left(\frac{r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma^\star_{r_{k_{\mathsf{max}}}}} + \frac{r^2\sqrt{m_1 m_2}\omega^2_{\mathsf{max}}\log^2 m}{\sigma^{\star 2}_{r_{k_{\mathsf{max}}}}}\right)\widetilde{\sigma}^2_{\bar{r}+1}\Bigg)\frac{1}{\sigma^\star_{\bar{r}}} \\
&\quad\overset{(ii)}{\lesssim} \Bigg(\sqrt{\frac{\mu r^3}{m_1}} r^2 \left(\sqrt{m_1}\omega_{\mathsf{max}} \log m \cdot \sigma^\star_{\bar{r}+1} + \left(\sqrt{m_1 m_2} + m_1\right) \omega^2_{\mathsf{max}} \log^2 m\right) + \sqrt{\frac{\mu r^3}{m_1}} \widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1} \\
&\qquad + \sqrt{\frac{\mu r^3}{m_1}}\left(\frac{r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma^\star_{r_{k_{\mathsf{max}}}}} + \frac{r^2\sqrt{m_1 m_2}\omega^2_{\mathsf{max}}\log^2 m}{\sigma^{\star 2}_{r_{k_{\mathsf{max}}}}}\right)\sigma^{\star 2}_{\bar{r}}\Bigg)\frac{1}{\sigma^\star_{\bar{r}}} \\
&\quad\overset{(iii)}{\lesssim} \sqrt{\frac{\mu r^3}{m_1}} r^2 \left(\sqrt{m_1}\omega_{\mathsf{max}}\log m + \frac{\left(\sqrt{m_1 m_2} + m_1\right)\omega^2_{\mathsf{max}}\log^2 m}{\sigma^\star_{\bar{r}}}\right) + \sqrt{\frac{\mu r^3}{m_1}}\frac{\widetilde{\sigma}^2_{r_{k_{\mathsf{max}}}+1}}{\sigma^\star_{\bar{r}}} \\
&\quad\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^\star_{\bar{r}}} \\
&\quad\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).
\end{aligned} \tag{207}$$

where (i), (ii) and (iii) are consequences of $\widetilde{\sigma}_{r_{k_{\mathsf{max}}}} \overset{(173)}{\asymp} \sigma^\star_{r_{k_{\mathsf{max}}}} \ge \sigma^\star_{\bar{r}} \asymp \widetilde{\sigma}_{\bar{r}} \gtrsim \max\{\sigma^\star_{\bar{r}+1}, \widetilde{\sigma}_{\bar{r}+1}\}$, (iv) makes use of the fact $\sigma^\star_{\bar{r}} \ge C_0 r[(m_1 m_2)^{1/4} + r m_1^{1/2}]\omega_{\mathsf{max}}\log m$ and (197). Note that $\widetilde{\boldsymbol{U}}^{(1)}\widetilde{\boldsymbol{\Sigma}}^{(1)}\widetilde{\boldsymbol{W}}^{(1)\top}$ is the SVD of $\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)} + \boldsymbol{E}\boldsymbol{V}^{\star(1)}$. By virtue of the previous inequality, (83a), Theorem 5 and the triangle inequality, we arrive at

$$\begin{aligned}
\alpha_1 &= \big\|\big(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top\big)\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\big\|_{2,\infty} \\
&= \big\|\big(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top\big)\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\big\|_{2,\infty}
\end{aligned}$$

$$
\begin{aligned}
&\le \left\|(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top)\widetilde{\boldsymbol{U}}^{(1)}\widetilde{\boldsymbol{\Sigma}}^{(1)}\widetilde{\boldsymbol{W}}^\top\right\|_{2,\infty} + \left\|(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top)\boldsymbol{E}\boldsymbol{V}^{\star(1)}\right\|_{2,\infty} \\
&\le \left\|(\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top)\widetilde{\boldsymbol{U}}^{(1)}\widetilde{\boldsymbol{\Sigma}}^{(1)}\right\|_{2,\infty} + \left\|\boldsymbol{U}_{k_{\mathsf{max}}}\boldsymbol{U}_{k_{\mathsf{max}}}^\top - \widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top\right\|_{2,\infty}\left\|\boldsymbol{E}\boldsymbol{V}^{\star(1)}\right\| \\
&\le \sqrt{\frac{\mu r^3}{m_1}}\left(r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m + r(m_1m_2)^{1/4}\omega_{\mathsf{max}}\log m\right) \\
&\quad + \sqrt{\frac{\mu r^3}{m_1}}\left(\frac{r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma^\star_{r_{k_{\mathsf{max}}}}} + \frac{r^2\sqrt{m_1m_2}\omega_{\mathsf{max}}^2\log^2 m}{\sigma^{\star 2}_{r_{k_{\mathsf{max}}}}}\right)\cdot\sqrt{m_1}\omega_{\mathsf{max}}\log m \\
&\le \sqrt{\frac{\mu r^3}{m_1}}\left(r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m + r(m_1m_2)^{1/4}\omega_{\mathsf{max}}\log m\right) + \sqrt{\frac{\mu r^3}{m_1}}\sqrt{m_1}\omega_{\mathsf{max}}\log m \\
&\asymp \sqrt{\frac{\mu r^3}{m_1}}\left(r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m + r(m_1m_2)^{1/4}\omega_{\mathsf{max}}\log m\right), \qquad (208)
\end{aligned}
$$

where the penultimate line comes from $\sigma^\star_{\overline{r}} \ge C_0 r[(m_1m_2)^{1/4} + rm_1^{1/2}]\omega_{\mathsf{max}}\log m$.

Step 5.2: bounding α_2. In view of (114), we have

$$
\begin{aligned}
&\mathcal{P}_{(\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}})_\perp}\mathcal{P}_{\boldsymbol{U}^\star_{k_{\mathsf{max}}}}\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top} \\
&\quad = \left(\mathcal{P}_{(\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}})_\perp}\boldsymbol{U}^\star_{k_{\mathsf{max}}}\right)\boldsymbol{U}^{\star\top}_{k_{\mathsf{max}}}\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top} \\
&\quad = \Big[\widetilde{\boldsymbol{U}}^{(1)}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}\widetilde{\boldsymbol{\Sigma}}^{(1)}_{r_{k_{\mathsf{max}}}+1:\overline{r},r_{k_{\mathsf{max}}}+1:\overline{r}}\widetilde{\boldsymbol{W}}^{(1)\top}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}\left(\boldsymbol{I}_{r_{k_{\mathsf{max}}}}\ \ \boldsymbol{0}\right)^\top\left(\boldsymbol{\Sigma}^\star_{1:r_{k_{\mathsf{max}}},1:r_{k_{\mathsf{max}}}}\right)^{-1} \\
&\qquad - \mathcal{P}_{(\widetilde{\boldsymbol{U}}_{:,1:r_{k_{\mathsf{max}}}})_\perp}\boldsymbol{E}\boldsymbol{V}^{\star(1)}\left(\boldsymbol{I}_{r_{k_{\mathsf{max}}}}\ \ \boldsymbol{0}\right)^\top\left(\boldsymbol{\Sigma}^\star_{1:r_{k_{\mathsf{max}}},1:r_{k_{\mathsf{max}}}}\right)^{-1}\Big] \\
&\qquad \cdot\boldsymbol{\Sigma}^\star_{1:r_{k_{\mathsf{max}}},1:r_{k_{\mathsf{max}}}}\boldsymbol{V}^{\star\top}_{:,1:r_{k_{\mathsf{max}}}} \\
&\quad = \widetilde{\boldsymbol{U}}^{(1)}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}\widetilde{\boldsymbol{\Sigma}}^{(1)}_{r_{k_{\mathsf{max}}}+1:\overline{r},r_{k_{\mathsf{max}}}+1:\overline{r}}\widetilde{\boldsymbol{W}}^{(1)\top}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}\left(\boldsymbol{I}_{r_{k_{\mathsf{max}}}}\ \ \boldsymbol{0}\right)^\top\boldsymbol{V}^{\star\top}_{:,1:r_{k_{\mathsf{max}}}} \\
&\qquad - \mathcal{P}_{(\widetilde{\boldsymbol{U}}_{:,1:r_{k_{\mathsf{max}}}})_\perp}\boldsymbol{E}\boldsymbol{V}^{\star(1)}\left(\boldsymbol{I}_{r_{k_{\mathsf{max}}}}\ \ \boldsymbol{0}\right)^\top\boldsymbol{V}^{\star\top}_{:,1:r_{k_{\mathsf{max}}}}. \qquad (209)
\end{aligned}
$$

Here, the third line holds since $r_{k_{\mathsf{max}}} \le \overline{r}$ and

$$
\begin{aligned}
\boldsymbol{U}^{\star\top}_{k_{\mathsf{max}}}\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top} &= \boldsymbol{U}^{\star\top}_{:,1:r_{k_{\mathsf{max}}}}\boldsymbol{U}^\star_{:,1:\overline{r}}\boldsymbol{\Sigma}^\star_{1:\overline{r},1:\overline{r}}\boldsymbol{V}^{\star\top}_{:,1:\overline{r}} \\
&= \boldsymbol{U}^{\star\top}_{:,1:r_{k_{\mathsf{max}}}}\left(\boldsymbol{U}^\star_{:,1:r_{k_{\mathsf{max}}}}\boldsymbol{\Sigma}^\star_{1:r_{k_{\mathsf{max}}},1:r_{k_{\mathsf{max}}}}\boldsymbol{V}^{\star\top}_{:,1:r_{k_{\mathsf{max}}}} + \boldsymbol{U}^\star_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}\boldsymbol{\Sigma}^\star_{r_{k_{\mathsf{max}}}+1:\overline{r},r_{k_{\mathsf{max}}}+1:\overline{r}}\boldsymbol{V}^{\star\top}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}\right) \\
&= \boldsymbol{\Sigma}^\star_{1:r_{k_{\mathsf{max}}},1:r_{k_{\mathsf{max}}}}\boldsymbol{V}^{\star\top}_{:,1:r_{k_{\mathsf{max}}}}. \qquad (210)
\end{aligned}
$$

Repeating similar arguments as in (117) and (118), one has

$$
\left\|\widetilde{\boldsymbol{U}}^{(1)}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}\widetilde{\boldsymbol{\Sigma}}^{(1)}_{r_{k_{\mathsf{max}}}+1:\overline{r},r_{k_{\mathsf{max}}}+1:\overline{r}}\widetilde{\boldsymbol{W}}^{(1)\top}_{:,r_{k_{\mathsf{max}}}+1:\overline{r}}\left(\boldsymbol{I}_{r_{k_{\mathsf{max}}}}\ \ \boldsymbol{0}\right)^\top\boldsymbol{V}^{\star\top}_{:,1:r_{k_{\mathsf{max}}}}\right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r}{m_1}}r\sqrt{m}_1\omega_{\mathsf{max}}\log m, \qquad (211a)
$$

$$
\left\|\mathcal{P}_{(\widetilde{\boldsymbol{U}}_{:,1:r_{k_{\mathsf{max}}}})_\perp}\boldsymbol{E}\boldsymbol{V}^{\star(1)}\left(\boldsymbol{I}_{r_{k_{\mathsf{max}}}}\ \ \boldsymbol{0}\right)^\top\boldsymbol{V}^{\star\top}_{:,1:r_{k_{\mathsf{max}}}}\right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r}{m_1}}\sqrt{m}_1\omega_{\mathsf{max}}\log m. \qquad (211b)
$$

Combining (209), (211a), (211b) and the triangle inequality yields

$$
\left\|\mathcal{P}_{(\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}})_\perp}\mathcal{P}_{\boldsymbol{U}^\star_{k_{\mathsf{max}}}}\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty} \lesssim \sqrt{\frac{\mu r^3}{m_1}}\sqrt{m}_1\omega_{\mathsf{max}}\log m. \qquad (212)
$$

Then we can bound α_2 as follows:

$$
\alpha_2 = \left\|(\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}\widetilde{\boldsymbol{U}}_{k_{\mathsf{max}}}^\top - \boldsymbol{U}^\star_{k_{\mathsf{max}}}\boldsymbol{U}^{\star\top}_{k_{\mathsf{max}}})\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty}
$$

$$
\begin{aligned}
&\le \left\|(\widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^{\top}-\boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top})\boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty}\\
&\quad+\left\|\widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^{\top}\left(\boldsymbol{U}_{k_{\max}}^{\star}\right)_{\perp}\left(\boldsymbol{U}_{k_{\max}}^{\star}\right)_{\perp}^{\star\top}\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty}\\
&=\left\|\mathcal{P}_{(\widetilde{\boldsymbol{U}}_{k_{\max}})_{\perp}}\mathcal{P}_{\boldsymbol{U}_{k_{\max}}^{\star}}\boldsymbol{U}^{\star(1)}\boldsymbol{\Sigma}^{\star(1)}\boldsymbol{V}^{\star(1)\top}\right\|_{2,\infty}\\
&\quad+\left\|\widetilde{\boldsymbol{U}}_{k_{\max}}\widetilde{\boldsymbol{U}}_{k_{\max}}^{\top}\left(\boldsymbol{U}_{k_{\max}}^{\star}\right)_{\perp}\left(\boldsymbol{U}_{k_{\max}}^{\star}\right)_{\perp}^{\star\top}\boldsymbol{U}^{\star}_{:,r_{k_{\max}}+1:\overline{r}}\boldsymbol{\Sigma}^{\star}_{r_{k_{\max}}+1:\overline{r},r_{k_{\max}}+1:\overline{r}}\boldsymbol{V}^{\star\top}_{:,r_{k_{\max}}+1:\overline{r}}\right\|_{2,\infty}\\
&\overset{(212)}{\lesssim}\sqrt{\frac{\mu r^3}{m_1}}\sqrt{m_1}\omega_{\mathsf{max}}\log m+\left\|\widetilde{\boldsymbol{U}}_{k_{\max}}\right\|_{2,\infty}\left\|\widetilde{\boldsymbol{U}}_{k_{\max}}^{\top}\left(\boldsymbol{U}_{k_{\max}}^{\star}\right)_{\perp}\right\|\left\|\boldsymbol{\Sigma}^{\star}_{r_{k_{\max}}+1:\overline{r},r_{k_{\max}}+1:\overline{r}}\right\|\\
&\overset{(83d)\text{ and }(115b)}{\lesssim}\sqrt{\frac{\mu r^3}{m_1}}\sqrt{m_1}\omega_{\mathsf{max}}\log m+\sqrt{\frac{\mu r}{m_1}}\frac{r\sqrt{m_1}\omega_{\mathsf{max}}\log m}{\sigma^{\star}_{r_{k_{\max}}}}\sigma^{\star}_{r_{k_{\max}}+1}\\
&\asymp\sqrt{\frac{\mu r^3}{m_1}}\sqrt{m_1}\omega_{\mathsf{max}}\log m. \qquad (213)
\end{aligned}
$$

Step 5.3: bounding α_3. By virtue of (70) and (199), one has

$$
\alpha_3=\left\|\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top}-\boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right\|_{2,\infty}\sigma^{\star}_{\overline{r}+1}\lesssim\sqrt{\frac{\mu r^3}{m_1}}\left(r\left[(m_1m_2)^{1/4}+rm_1^{1/2}\right]\omega_{\mathsf{max}}\log m\right). \qquad (214)
$$

Step 5.4: bounding $\|(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top}-\boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top})\boldsymbol{X}^{\star}\|_{2,\infty}$ and $\|(\boldsymbol{I}_{n_1}-\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top})\boldsymbol{X}^{\star}\|_{2,\infty}$. Putting (200), (208), (213) and (214) together, we arrive at

$$
\left\|\left(\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top}-\boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\boldsymbol{X}^{\star}\right\|_{2,\infty}\lesssim\sqrt{\frac{\mu r^3}{m_1}}\left(r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m+r(m_1m_2)^{1/4}\omega_{\mathsf{max}}\log m\right). \qquad (215)
$$

Furthermore, we have

$$
\begin{aligned}
\left\|\left(\boldsymbol{I}_{m_1}-\boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\right)\boldsymbol{X}^{\star}\right\|_{2,\infty}&=\left\|\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{V}^{\star\top}-\boldsymbol{U}_{k_{\max}}^{\star}\boldsymbol{U}_{k_{\max}}^{\star\top}\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{V}^{\star\top}\right\|_{2,\infty}\\
&=\left\|\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{V}^{\star\top}-\boldsymbol{U}^{\star}_{:,1:r_{k_{\max}}}\boldsymbol{\Sigma}^{\star}_{1:r_{k_{\max}},1:r_{k_{\max}}}\boldsymbol{V}^{\star\top}_{:,1:r_{k_{\max}}}\right\|_{2,\infty}\\
&=\left\|\boldsymbol{U}^{\star}_{:,r_{k_{\max}}+1:r}\boldsymbol{\Sigma}^{\star}_{r_{k_{\max}}+1:r,r_{k_{\max}}+1:r}\boldsymbol{V}^{\star\top}_{:,r_{k_{\max}}+1:r}\right\|_{2,\infty}\\
&\le\left\|\boldsymbol{U}^{\star}\right\|_{2,\infty}\left\|\boldsymbol{\Sigma}^{\star}_{r_{k_{\max}}+1:r,r_{k_{\max}}+1:r}\right\|\\
&\le\sqrt{\frac{\mu r}{m_1}}\sigma^{\star}_{r_{k_{\max}}+1}\\
&\overset{(83a)\text{ and }(96)}{\le}\sqrt{\frac{\mu r}{m_1}}\left(\widetilde{\sigma}_{r_{k_{\max}}+1}+2\sqrt{C_5}\sqrt{m_1}\omega_{\mathsf{max}}\log m\right)\\
&\overset{(197)}{\lesssim}\sqrt{\frac{\mu r}{m_1}}\left(\sqrt{\tau}+\sqrt{m_1}\omega_{\mathsf{max}}\log m\right)\\
&\asymp\sqrt{\frac{\mu r}{m_1}}\left(r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m+r(m_1m_2)^{1/4}\omega_{\mathsf{max}}\log m\right).
\end{aligned}
$$

This together with (215) gives

$$
\left\|(\boldsymbol{I}_{n_1}-\boldsymbol{U}_{k_{\max}}\boldsymbol{U}_{k_{\max}}^{\top})\boldsymbol{X}^{\star}\right\|_{2,\infty}\lesssim\sqrt{\frac{\mu r^3}{m_1}}\left(r^2\sqrt{m_1}\omega_{\mathsf{max}}\log m+r(m_1m_2)^{1/4}\omega_{\mathsf{max}}\log m\right). \qquad (216)
$$

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 (37) that

$$\|\boldsymbol{U}_i^\star\|_{2,\infty} \le \sqrt{\frac{k_i}{\beta n_i}}, \qquad \forall i \in [3]. \tag{217}$$

In view of Zhou and Chen (2023, Lemma 7), with probability exceeding $1 - O(n^{-10})$,

$$\left\|\mathcal{P}_{\mathsf{off\text{-}diag}}\left(\boldsymbol{E}_1\boldsymbol{E}_1^\top\right)\right\| \lesssim B^2 \log^2 n + \sqrt{n_1n_2n_3}\omega_{\mathsf{max}}^2 \log n \le \sqrt{n_1n_2n_3}\omega_{\mathsf{max}}^2 \log n \ll n_2n_3\omega_{\mathsf{max}}^2. \tag{218}$$

For any $i \in [n_1]$, we know that

$$\left(\boldsymbol{E}_1\boldsymbol{E}_1^\top\right)_{i,i} = \sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3} E_{i,j,\ell}^2 = \sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3} \omega_{i,j,\ell}^2 + \sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3}\left(E_{i,j,\ell}^2 - \omega_{i,j,\ell}^2\right). \tag{219}$$

If the noise is bounded, i.e., $E_{i,j,k} \le B$ 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

$$\begin{aligned}
\left|E_{i,j,\ell}^2 - \omega_{i,j,\ell}^2\right| &\le 2B^2,\\
\mathbb{E}\left[\left(E_{i,j,\ell}^2 - \omega_{i,j,\ell}^2\right)^2\right] &= \mathbb{E}\left[E_{i,j,\ell}^4\right] - \omega_{i,j,\ell}^4 \le B^2\mathbb{E}\left[E_{i,j,\ell}^2\right] - \omega_{i,j,\ell}^4 \le B^2\omega_{\mathsf{max}}^2.
\end{aligned}$$

In view of Bernstein's inequality and the union bound, one has, with probability exceeding $1 - O(n^{-10})$, for all $i \in [n_1]$,

$$\left|\sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3}\left(E_{i,j,\ell}^2 - \omega_{i,j,\ell}^2\right)\right| \lesssim \sqrt{n_2n_3}B\omega_{\mathsf{max}}\sqrt{\log n} + B^2\log n \ll n_2n_3\omega_{\mathsf{max}}^2. \tag{221}$$

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 (221) also holds with probability exceeding $1 - O(n^{-10})$. Putting (218), (219) and (221) together, we know that with probability exceeding $1 - O(n^{-10})$,

$$\left\|\boldsymbol{E}_1\boldsymbol{E}_1^\top - \mathsf{diag}\left(\left[\sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3}\omega_{i,j,\ell}^2\right]_{1\le i\le n_1}\right)\right\| \ll n_2n_3\omega_{\mathsf{max}}^2. \tag{222}$$

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

$$\begin{aligned}
\sigma_{k_1+1}\left(\boldsymbol{Y}_1\right) &\ge \sigma_{k_1+1}\left(\boldsymbol{Y}_1\left(\mathcal{P}_{\boldsymbol{U}_{3\perp}^\star}\otimes\mathcal{P}_{\boldsymbol{U}_{2\perp}^\star}\right)\right)\\
&= \sigma_{k_1+1}\left(\boldsymbol{X}_1^\star\left(\mathcal{P}_{\boldsymbol{U}_{3\perp}^\star}\otimes\mathcal{P}_{\boldsymbol{U}_{2\perp}^\star}\right) + \boldsymbol{E}_1\left(\mathcal{P}_{\boldsymbol{U}_{3\perp}^\star}\otimes\mathcal{P}_{\boldsymbol{U}_{2\perp}^\star}\right)\right)\\
&= \sigma_{k_1+1}\left(\boldsymbol{E}_1\left(\mathcal{P}_{\boldsymbol{U}_{3\perp}^\star}\otimes\mathcal{P}_{\boldsymbol{U}_{2\perp}^\star}\right)\right)\\
&\ge \sigma_{k_1+1}\left(\boldsymbol{E}_1\right) - \left\|\boldsymbol{E}_1\left(\mathcal{P}_{\boldsymbol{U}_3^\star}\otimes\mathcal{P}_{\boldsymbol{U}_2^\star}\right)\right\| - \left\|\boldsymbol{E}_1\left(\mathcal{P}_{\boldsymbol{U}_{3\perp}^\star}\otimes\mathcal{P}_{\boldsymbol{U}_2^\star}\right)\right\| - \left\|\boldsymbol{E}_1\left(\mathcal{P}_{\boldsymbol{U}_3^\star}\otimes\mathcal{P}_{\boldsymbol{U}_{2\perp}^\star}\right)\right\|\\
&= \sqrt{\sigma_{k_1+1}\left(\boldsymbol{E}_1\boldsymbol{E}_1^\top\right)} - \left\|\boldsymbol{E}_1\left(\boldsymbol{U}_3^\star\otimes\boldsymbol{U}_2^\star\right)\right\| - \left\|\boldsymbol{E}_1\left(\boldsymbol{U}_{3\perp}^\star\otimes\boldsymbol{U}_2^\star\right)\right\| - \left\|\boldsymbol{E}_1\left(\boldsymbol{U}_3^\star\otimes\boldsymbol{U}_{2\perp}^\star\right)\right\|,
\end{aligned} \tag{223}$$

where the fourth line makes use of Weyl's inequality. We know from (217) that

$$\begin{aligned}
\|\boldsymbol{U}_3^\star\otimes\boldsymbol{U}_2^\star\|_{2,\infty} &\le \sqrt{\frac{k_2k_3}{\beta^2n_2n_3}},\\
\|\boldsymbol{U}_{3\perp}^\star\otimes\boldsymbol{U}_2^\star\|_{2,\infty} &\le \|\boldsymbol{U}_2^\star\|_{2,\infty} \le \sqrt{\frac{k_2}{\beta n_2}},
\end{aligned}$$

$$\|\boldsymbol{U}_3^\star \otimes \boldsymbol{U}_{2\perp}^\star\|_{2,\infty} \le \|\boldsymbol{U}_3^\star\|_{2,\infty} \le \sqrt{\frac{k_3}{\beta n_3}}.$$

Applying Zhou and Chen (2023, Lemma 5) yields that with probability at least $1 - O(n^{-10})$,

$$\|\boldsymbol{E}_1 (\boldsymbol{U}_3^\star \otimes \boldsymbol{U}_2^\star)\| \lesssim B\sqrt{\frac{k_2 k_3}{\beta^2 n_2 n_3}} \log n + \sqrt{n_1}\omega_{\mathsf{max}}\sqrt{\log n} \ll \sqrt{n_2 n_3}\omega_{\mathsf{max}}, \tag{224a}$$

$$\|\boldsymbol{E}_1 (\boldsymbol{U}_{3\perp}^\star \otimes \boldsymbol{U}_2^\star)\| \lesssim B\sqrt{\frac{k_2}{\beta n_2}} \log n + \sqrt{n_1 + k_2 n_3}\omega_{\mathsf{max}}\sqrt{\log n} \ll \sqrt{n_2 n_3}\omega_{\mathsf{max}}, \tag{224b}$$

$$\|\boldsymbol{E}_1 (\boldsymbol{U}_3^\star \otimes \boldsymbol{U}_{2\perp}^\star)\| \lesssim B\sqrt{\frac{k_3}{\beta n_3}} \log n + \sqrt{n_1 + k_3 n_2}\omega_{\mathsf{max}}\sqrt{\log n} \ll \sqrt{n_2 n_3}\omega_{\mathsf{max}}. \tag{224c}$$

Combining (222), (224a) - (224c) and Weyl's inequality, we obtain that with probability exceeding $1 - O(n^{-10})$,

$$\sigma_{k_1+1} (\boldsymbol{Y}_1) \le \sigma_{k_1+1} (\boldsymbol{X}_1^\star) + \|\boldsymbol{E}_1\| = \|\boldsymbol{E}_1\| = \left(\left\|\boldsymbol{E}_1 \boldsymbol{E}_1^\top\right\|\right)^{1/2} \le 2\sqrt{n_2 n_3}\omega_{\mathsf{max}}. \tag{225}$$

1. If (16) holds, then (222), (223) and (224a) - (224c) together show that with probability exceeding $1 - O(n^{-10})$,

$$\sigma_{k_1+1} (\boldsymbol{Y}_1) \ge \sqrt{c n_2 n_3 \omega_{\mathsf{max}}^2} - \frac{\sqrt{c}}{2}\sqrt{n_2 n_3}\omega_{\mathsf{max}} \ge \frac{\sqrt{c}}{2}\sqrt{n_2 n_3}\omega_{\mathsf{max}}. \tag{226}$$

   The previous inequality together with (225) shows that

$$\sigma_{k_1+1} (\boldsymbol{Y}_1) \asymp \sqrt{n_2 n_3}\omega_{\mathsf{max}}$$

   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 $1 - O(n^{-10})$,

$$c_\tau (n_1 n_2 n_3)^{1/2} \log^2 n \le \tau/\omega_{\mathsf{max}}^2 \le C_\tau (n_1 n_2 n_3)^{1/2} \log^2 n \tag{227}$$

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

$$(\ell_1, \ell_2, \ell_3) \in \mathop{\arg\max}_{i_1 \in [k_1], i_2 \in [k_2], i_3 \in [k_3]} S_{i_1,i_2,i_3}^\star \left(1 - S_{i_1,i_2,i_3}^\star\right) = \omega_{\mathsf{max}}^2.$$

   Then for any $(j_1, j_2, j_3) \in [n_1] \times [n_2] \times [n_3]$ with $(\boldsymbol{z}_i^\star)_{j_i} = \ell_i$, $i \in [3]$,

$$\mathbb{E}\left[E_{j_1,j_2,j_3}^2\right] = \omega_{\mathsf{max}}^2.$$

   For any $i \in \{j \in [n_1] : (\boldsymbol{z}_1^\star)_j = \ell_1\}$,

$$\sum_{j=1}^{n_2}\sum_{\ell=1}^{n_3} \omega_{i,j,\ell}^2 \ge \omega_{\mathsf{max}}^2 \left|j \in [n_2] : (\boldsymbol{z}_2^\star)_j = \ell_2\right| \cdot \left|j \in [n_3] : (\boldsymbol{z}_3^\star)_j = \ell_3\right| \ge \frac{\beta n_2}{k_2}\frac{\beta n_3}{k_3}\omega_{\mathsf{max}}^2 \asymp n_2 n_3 \sigma_{\mathsf{max}}^2.$$

   Since $|\{j \in [n_1] : (\boldsymbol{z}_1^\star)_j = \ell_1\}| \ge \beta n_1/k_1 > k_1 + 1$, we can still show that (226) holds with probability exceeding $1 - O(n^{-10})$. Repeating similar arguments as in (227) shows that Theorem 2 also holds.

F Technical lemmas

Lemma 6. *Suppose that*

$$\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{E} \in \mathbb{R}^{n_1 \times n_2} \tag{228}$$

and the SVDs of $\boldsymbol{M}^\star$ and $\boldsymbol{M}$ are given by

$$\boldsymbol{M}^\star = \sum_{i=1}^{n_1} \sigma_i^\star \boldsymbol{u}_i^\star \boldsymbol{v}_i^{\star\top}, \qquad \text{and} \qquad \boldsymbol{M}^\star = \sum_{i=1}^{n_1} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\top}.$$

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

$$\begin{aligned} \boldsymbol{\Sigma} &:= \mathsf{diag}\left(\sigma_1, \ldots, \sigma_r\right), & \boldsymbol{\Sigma}_\perp &:= \mathsf{diag}\left(\sigma_{r+1}, \ldots, \sigma_{n_1}\right), \\ \boldsymbol{V} &:= [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r] \in \mathbb{R}^{n_2 \times r}, & \boldsymbol{V}_\perp &:= [\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_{n_2}] \in \mathbb{R}^{n_2 \times (n_2 - r)}, \end{aligned}$$

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

$$\begin{aligned} \left\| \left(\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top} \right) \boldsymbol{M}^\star \right\| &\leq \left\| \mathcal{P}_{\boldsymbol{U}_\perp} \mathcal{P}_{\boldsymbol{U}^\star} \boldsymbol{M}^\star \right\| + \left\| \boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{U}_\perp^\star \boldsymbol{U}_\perp^{\star\top} \boldsymbol{M}^\star \right\| \\ &\leq \left\| \boldsymbol{U}_\perp \boldsymbol{\Sigma}_\perp \boldsymbol{V}_\perp^\top \boldsymbol{V}^\star \boldsymbol{V}^{\star\top} \right\| + \left\| \boldsymbol{U}_\perp^\top \boldsymbol{E} \boldsymbol{V}^\star \boldsymbol{V}^{\star\top} \right\| + \left\| \boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{U}_\perp^\star \boldsymbol{\Sigma}_\perp^\star \boldsymbol{V}_\perp^{\star\top} \right\| \\ &\leq \left\| \boldsymbol{\Sigma}_\perp \right\| \left\| \boldsymbol{V}_\perp^\top \boldsymbol{V}^\star \right\| + \|\boldsymbol{E}\| + \left\| \boldsymbol{U}_\perp^{\star\top} \boldsymbol{U} \right\| \left\| \boldsymbol{\Sigma}_\perp^\star \right\| \\ &\overset{(a)}{\leq} \sigma_{r+1} \frac{2\|\boldsymbol{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star} + \|\boldsymbol{E}\| + \sigma_{r+1}^\star \frac{2\|\boldsymbol{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star} \\ &\overset{(b)}{\leq} \left(\sigma_{r+1}^\star + \|\boldsymbol{E}\| \right) \frac{2\|\boldsymbol{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star} + \|\boldsymbol{E}\| + \sigma_{r+1}^\star \frac{2\|\boldsymbol{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star} \\ &\overset{(c)}{\leq} \frac{4\sigma_r^\star \|\boldsymbol{E}\|}{\sigma_r^\star - \sigma_{r+1}^\star}. \end{aligned} \tag{229}$$

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 $\sigma_r^\star - \sigma_{r+1}^\star > 2\|\boldsymbol{E}\|$. ∎

Lemma 7. *Suppose that Assumption 1 holds. We let $\mathcal{A}$ denote the following set:*

$$\mathcal{U} = \left\{ (\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3) : \boldsymbol{U}_i \in \mathbb{R}^{n_i \times r_i}, \|\boldsymbol{U}_i\| \leq 1, \|\boldsymbol{U}_i\|_{2,\infty} \leq \sqrt{\frac{\mu_i r_i}{n_i}}, i \in [3] \right\} \tag{230}$$

and we define

$$n = \max_{1\le i\le 3} n_i, \qquad \textit{and} \qquad r = \max_{1\le i\le 3} r_i.$$

If $n_1 n_2 n_3 \ge r^4 n^2$, then with probability exceeding $1 - O(n^{-10})$, the following inequality holds:

$$\sup_{(\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3)\in\mathcal{U}} \left\|\boldsymbol{U}_1^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\boldsymbol{U}_3\otimes\boldsymbol{U}_2\right)\right\| \lesssim \omega_{\mathsf{max}}\sqrt{n\mu_1\mu_2\mu_3 r^3\log n}, \tag{231}$$

If, furthermore, the $E_{i,j,k}$'s are ω_{max}-sub-Gaussian, then with probability exceeding $1 - e^{-Cn}$, one has

$$\sup_{(\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3)\in\mathcal{U}} \left\|\boldsymbol{U}_1^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\boldsymbol{U}_3\otimes\boldsymbol{U}_2\right)\right\| \le \sqrt{nr}\,\omega_{\mathsf{max}}. \tag{232}$$

Proof of Lemma 7. (232) 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 (231): the bounded noise case. To prove (231), we first consider the bounded noise case, i.e., $|E_{i,j,k}| \le B$ holds for all $(i,j,k) \in [n_1]\times[n_2]\times[n_3]$. For any fixed $\left(\overline{\boldsymbol{U}}_1,\overline{\boldsymbol{U}}_2,\overline{\boldsymbol{U}}_3\right)\in\mathcal{U}$, note that

$$\overline{\boldsymbol{U}}_1^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\overline{\boldsymbol{U}}_3\otimes\overline{\boldsymbol{U}}_2\right) = \sum_{i\in[n_1],j\in[n_2n_3]} \left(\mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\right)_{i,j}\left(\overline{\boldsymbol{U}}_1\right)_{i,:}^\top\left(\overline{\boldsymbol{U}}_3\otimes\overline{\boldsymbol{U}}_2\right)_{j,:}$$

is a sum of independent zero-mean matrices. In addition, we have

$$\begin{aligned}
L &:= \max_{i\in[n_1],j\in[n_2n_3]} \left\|\left(\mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\right)_{i,j}\left(\overline{\boldsymbol{U}}_1\right)_{i,:}^\top\left(\overline{\boldsymbol{U}}_3\otimes\overline{\boldsymbol{U}}_2\right)_{j,:}\right\| \le B\prod_{i=1}^{3}\left\|\overline{\boldsymbol{U}}_i\right\|_{2,\infty} \le B\sqrt{\frac{\mu_1\mu_2\mu_3 r_1 r_2 r_3}{n_1n_2n_3}}, \\
V &:= \max\left\{\left\|\sum_{i\in[n_1],j\in[n_2n_3]}\mathbb{E}\left[\left(\mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\right)_{i,j}^2\right]\left\|\left(\overline{\boldsymbol{U}}_3\otimes\overline{\boldsymbol{U}}_2\right)_{j,:}\right\|_2^2\left(\overline{\boldsymbol{U}}_1\right)_{i,:}^\top\left(\overline{\boldsymbol{U}}_1\right)_{i,:}\right\|,\right. \\
&\qquad \left.\left\|\sum_{i\in[n_1],j\in[n_2n_3]}\mathbb{E}\left[\left(\mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\right)_{i,j}^2\right]\left\|\left(\overline{\boldsymbol{U}}_1\right)_{i,:}\right\|_2^2\left(\overline{\boldsymbol{U}}_3\otimes\overline{\boldsymbol{U}}_2\right)_{j,:}^\top\left(\overline{\boldsymbol{U}}_3\otimes\overline{\boldsymbol{U}}_2\right)_{j,:}\right\|\right\} \\
&\le \max\left\{\sigma_{\mathsf{max}}^2\left\|\left(\overline{\boldsymbol{U}}_3\otimes\overline{\boldsymbol{U}}_2\right)\right\|_{\mathrm{F}}^2\left\|\overline{\boldsymbol{U}}_1\overline{\boldsymbol{U}}_1^\top\right\|, \sigma_{\mathsf{max}}^2\left\|\overline{\boldsymbol{U}}_1\right\|_{\mathrm{F}}^2\left\|\left(\overline{\boldsymbol{U}}_3\otimes\overline{\boldsymbol{U}}_2\right)\left(\overline{\boldsymbol{U}}_3\otimes\overline{\boldsymbol{U}}_2\right)^\top\right\|\right\} \\
&\lesssim \omega_{\mathsf{max}}^2\max\left\{r_2r_3, r_1\right\} \le \omega_{\mathsf{max}}^2 r^2.
\end{aligned}$$

In view of the matrix Bernstein inequality, with probability exceeding $1 - e^{-Cnr\log n}$,

$$\begin{aligned}
\left\|\overline{\boldsymbol{U}}_1^\top \mathcal{M}_1\left(\boldsymbol{\mathcal{E}}\right)\left(\overline{\boldsymbol{U}}_3\otimes\overline{\boldsymbol{U}}_2\right)\right\| &\lesssim \sqrt{Vnr\log n} + Lnr\log n \\
&\lesssim \sigma_{\mathsf{max}}\sqrt{nr^3\log n} + \omega_{\mathsf{max}}\frac{\left(n_1n_2n_3\right)^{1/4}}{\log n}\sqrt{\frac{\mu_1\mu_2\mu_3 r_1r_2r_3}{n_1n_2n_3}}nr\log n \\
&\le \sigma_{\mathsf{max}}\sqrt{n\mu_1\mu_2\mu_3 r^3\log n}.
\end{aligned} \tag{233}$$

Here, the last line makes use of $n_1n_2n_3 \ge r^4n^2$.

Repeating a similar argument as in (Vershynin, 2010, Lemma 5.2) yields that: there exists a set $\mathcal{B}_i \subset \mathcal{D}_i = \{\boldsymbol{x} : \boldsymbol{x}\in\mathbb{R}^{r_i}, \|\boldsymbol{x}\|_2 \le \sqrt{\mu_i r_i/n_i}\}$ 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

$$\|\boldsymbol{x}-\boldsymbol{x}'\|_2 \le \frac{1}{4}\sqrt{\frac{1}{n_i}}.$$

$\|\boldsymbol{x}-\boldsymbol{x}'\|_2 \le c\sqrt{1/n_i}$. As a direct consequence, for any $\boldsymbol{U}_i \in \mathcal{U}_i := \left\{\boldsymbol{U}\in\mathbb{R}^{n_i\times r_i} : \|\boldsymbol{U}\| \le 1, \|\boldsymbol{U}\|_{2,\infty} \le \sqrt{\frac{\mu_i r_i}{n_i}}\right\}$, 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

$$\|\boldsymbol{U}_i - \boldsymbol{U}_i'\|_{2,\infty} \le \frac{1}{4}\sqrt{\frac{1}{n_i}}.$$

Let $\mathcal{U}_i'$ denote the following set:

$$\mathcal{U}_i' := \left\{ \boldsymbol{U}' \in \mathbb{R}^{n_i \times r_i} : \boldsymbol{U}' \in \mathcal{F}_i, \inf_{\boldsymbol{U} \in \mathcal{U}_i} \|\boldsymbol{U} - \boldsymbol{U}'\|_{2,\infty} \le \frac{1}{4}\sqrt{\frac{1}{n_i}} \right\}. \tag{234}$$

Then we can verify the following three properties:

$$|\mathcal{U}_i'| \le |\mathcal{F}_i| = |\mathcal{B}_i|^{n_i} \le (1 + 8\sqrt{\mu_i r_i})^{n_i r_i} \le n_i^{n_i r_i} \le e^{n_i r_i \log n}, \tag{235a}$$

$$\forall \boldsymbol{U}_i \in \mathcal{U}_i, \quad \exists \boldsymbol{U}_i' \in \mathcal{U}_i' \quad \text{s.t.} \quad \|\boldsymbol{U}_i - \boldsymbol{U}_i'\|_{2,\infty} \le \frac{1}{4}\sqrt{\frac{1}{n_i}} \quad \text{and} \quad \|\boldsymbol{U}_i - \boldsymbol{U}_i'\| \le \frac{1}{4}, \tag{235b}$$

$$\forall \boldsymbol{U}_i' \in \mathcal{U}_i', \quad \exists \boldsymbol{U}_i \in \mathcal{U}_i \quad \text{s.t.} \quad \|\boldsymbol{U}_i - \boldsymbol{U}_i'\|_{2,\infty} \le \frac{1}{4}\sqrt{\frac{1}{n_i}} \quad \text{and} \quad \|\boldsymbol{U}_i - \boldsymbol{U}_i'\| \le \frac{1}{4}. \tag{235c}$$

Here, the last two inequalities make use of $\|\boldsymbol{U}_i - \boldsymbol{U}_i'\| \le \sqrt{n_i}\|\boldsymbol{U}_i - \boldsymbol{U}_i'\|_{2,\infty}$.

We define

$$A = \sup_{(\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3) \in \mathcal{U}} \left\|\boldsymbol{U}_1^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2)\right\| \qquad \text{and} \qquad B = \sup_{(\boldsymbol{U}_1', \boldsymbol{U}_2', \boldsymbol{U}_3') \in \mathcal{U}_1' \times \mathcal{U}_2' \times \mathcal{U}_3'} \left\|\boldsymbol{U}_1'^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\boldsymbol{U}_3' \otimes \boldsymbol{U}_2')\right\|.$$

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 (235b) and (235c) 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

$$\boldsymbol{U}_i - \boldsymbol{U}_i' \in \frac{1}{4}\mathcal{U}_i, \qquad i \in [3], \tag{236a}$$

$$\boldsymbol{U}_i' \in \frac{5}{4}\mathcal{U}_i, \qquad i \in [3]. \tag{236b}$$

The triangle inequality, (236a) and (236b) together show that

$$\begin{aligned}
\left\|\boldsymbol{U}_1^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2)\right\| &\le \left\|(\boldsymbol{U}_1 - \boldsymbol{U}_1')^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2)\right\| + \left\|\boldsymbol{U}_1'^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\boldsymbol{U}_3 \otimes (\boldsymbol{U}_2 - \boldsymbol{U}_2'))\right\| \\
&\quad + \left\|\boldsymbol{U}_1'^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})((\boldsymbol{U}_3 - \boldsymbol{U}_3') \otimes \boldsymbol{U}_2')\right\| + \left\|\boldsymbol{U}_1'^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\boldsymbol{U}_3' \otimes \boldsymbol{U}_2')\right\| \\
&\le \frac{1}{4}A + \frac{1}{4}\cdot\frac{5}{4}A + \frac{1}{4}\cdot\frac{5}{4}\cdot\frac{5}{4}A + B \\
&\le \frac{61}{64}A + B,
\end{aligned}$$

which implies

$$A \le \frac{64}{3}B. \tag{237}$$

Furthermore, (233), (235a), (236b) and the union bound together imply that with probability exceeding $1 - e^{-Cnr\log n} \cdot \prod_i e^{n_i r_i \log n} \ge 1 - e^{-C'nr\log n}$,

$$B \lesssim \left(\frac{5}{4}\right)^3 \sigma_{\mathsf{max}}\sqrt{n\mu_1\mu_2\mu_3 r^3 \log n} \asymp \sigma_{\mathsf{max}}\sqrt{n\mu_1\mu_2\mu_3 r^3 \log n}. \tag{238}$$

Putting (237) and (238) together, we arrive at, with probability exceeding $1 - e^{-C'nr\log n}$,

$$\sup_{(\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3) \in \mathcal{U}} \left\|\boldsymbol{U}_1^\top \mathcal{M}_1(\boldsymbol{\mathcal{E}})(\boldsymbol{U}_3 \otimes \boldsymbol{U}_2)\right\| = A \lesssim \sigma_{\mathsf{max}}\sqrt{n\mu_1\mu_2\mu_3 r^3 \log n}.$$

Proving (231): the general case. For the general case where the noise matrix $\boldsymbol{E}$ satisfies Assumption 2, we can prove (231) by repeating a similar argument as in Zhou and Chen (2023, Section B.4.2).

□

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.

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.

Bi, X., Qu, A., and Shen, X. (2018). Multilayer tensor factorization with applications to recommender systems. *The Annals of Statistics*, 46(6B):3308–3333.

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 Zhang, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *The Annals of Statistics*, 46(1):60–89.

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., 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. (2021c). Convex and nonconvex optimization are both minimax-optimal for noisy blind deconvolution under random designs. *Journal of the American Statistical Association*, pages 1–11.

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.

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.

Florescu, L. and Perkins, W. (2016). Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory*, pages 943–959. PMLR.

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.

Han, R., Luo, Y., Wang, M., Zhang, A. R., et al. (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*.

Jegelka, S., Sra, S., and Banerjee, A. (2009). Approximation algorithms for tensor clustering. In *International Conference on Algorithmic Learning Theory*, pages 368–383. Springer.

Kumar, A., Sabharwal, Y., and Sen, S. (2004). A simple linear time $(1 + \varepsilon)$-approximation algorithm for k-means clustering in any dimensions. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 454–462. IEEE.

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.

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.

Ling, S. (2022). Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *Applied and Computational Harmonic Analysis*, 60:20–52.

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.

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.

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 blockmodel. *The Annals of Statistics*, 39(4):1878–1915.

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 nonconvex 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*.

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., 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*.

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.