Supplementary notes to "Multiway Spherical Clustering via Degree-Corrected Tensor Block Models"

Jiaxin Hu and Miaoyan Wang
University of Wisconsin – Madison

Software

We have developed an R-package `dTBM` basd on our algorithm in the main text. The software and the cleaned datasets used in Section VII are available at `https://cran.r-project.org/package=dTBM`.

Appendices

Appendices include the additional numerical experiment results and the technical proofs for all theorems and lemmas in the main text.

A Additional numerical experiments

Bernoulli phase transition. The first additional experiment verifies the statistical-computational gap in Section III under the Bernoulli model. Consider the Bernoulli model with $p = \{80, 100\}$, $r = 5$. We vary γ in $[-1.2, -0.4]$ and $[-2.1, -1.4]$ for matrix ($K = 2$) and tensor ($K = 3$) clustering, respectively. We approximate MLE using an oracle estimator, i.e., the output of Sub-algorithm 2 initialized from the true assignment. Figure 1 shows a similar pattern as Figure 4. The algorithm and oracle estimators have no gap in the matrix case, while an error gap emerges between the critical values $\gamma_{\text{stat}} = -2$ and $\gamma_{\text{comp}} = -1.5$ in the tensor case. Figure 4 suggests the statistical-computational gap in Bernoulli models.

Sparsity. The second additional experiment evaluates the algorithm performances under the sparse binary dTBM (18). We fix the signal exponent $\gamma = -1.2$ and vary the sparsity parameter $\alpha_p \in [0.05, 0.9]$. A smaller α_p leads to a higher probability of zero entries in the observation. In addition to the three algorithms mentioned in Section VI-B (denoted **Initialization**, **dTBM**, and **SCORE**), we consider other three algorithms based on the discussion in Section IV-C:

- **D-HOSVD**, the diagonal-deleted HOSVD in Ke et al. (2019);
- **D-HOSVD + Angle**, the combined algorithm of our angle-based iteration with initialization from **D-HOSVD**;
- **SCORE + Angle**, the combined algorithms of our angle-based iteration with initialization from **SCORE**.

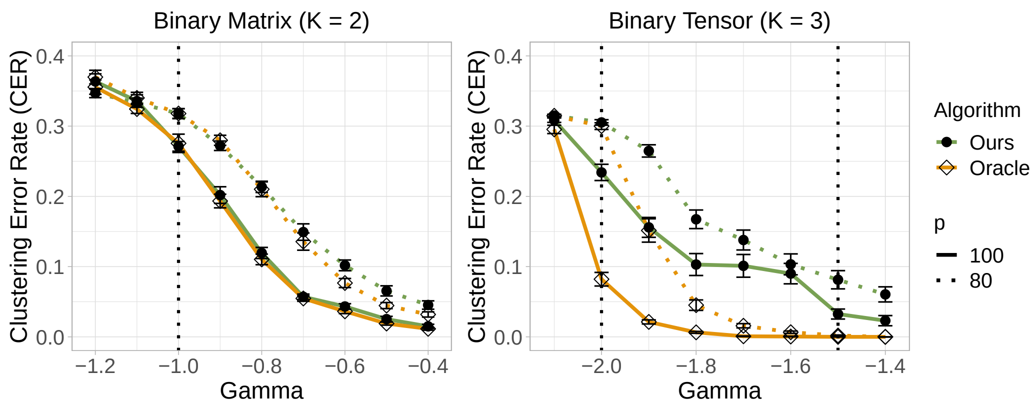

Figure 1: SNR phase transitions for Bernoulli dTBM with $p = \{80, 100\}, r = 5$ under (a) matrix case with $\gamma \in [-1.2, -0.4]$ and (b) tensor case with $\gamma \in [-2.1, -1.4]$.

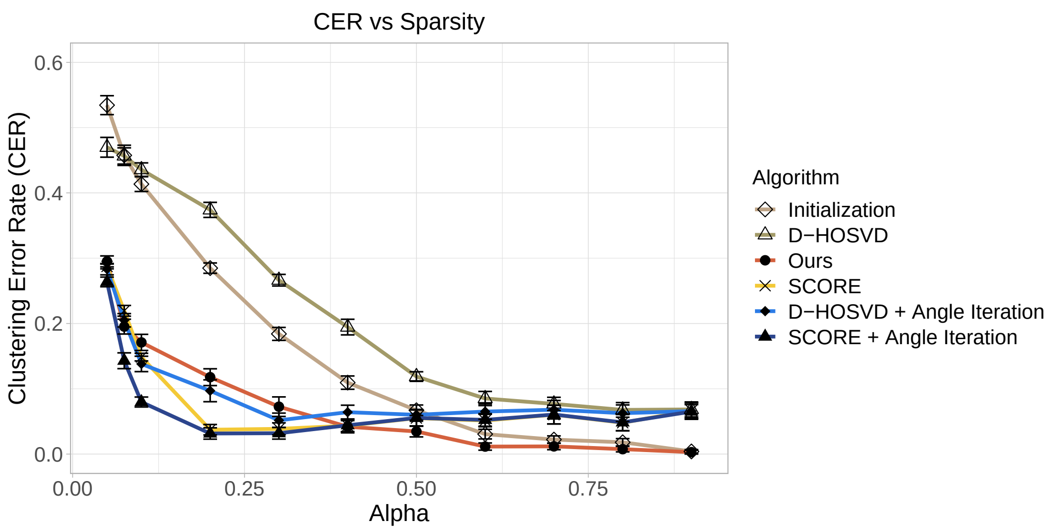

Figure 2: CER comparison versus sparsity parameter α_p in $[0.05, 0.9]$. We set $p = 100, r = 5$ and $\gamma = -1.2$ under sparse binary dTBM.

Figure 2 shows a slightly larger error in **dTBM** than that in **SCORE**, **D-HOSVD + Angle**, and **SCORE + Angle** under the sparse setting with $\alpha_p < 0.3$. The small gap between **dTBM** and other sparse-specific methods implies the robustness of our algorithm. In addition, comparing **SCORE** versus **SCORE + Angle** (or **D-HOSVD** versus **D-HOSVD + Angle**) indicates the benefit of our angle iterations under the sparse dTBM. In the intermediate and dense cases with $\alpha_p \geq 0.3$, our proposed **dTBM** has a clear improvement over others, which again verifies the success of our algorithm in dense settings.

B Proofs

We provide the proofs for all the theorems in our main paper. In each sub-section, we first show the proof of main theorem and then collect the useful lemmas in the end. We combine the proofs of MLE achievement in Theorem 2 and polynomial-time achievement in Theorem 5 in the last section due to the similar idea.

B.1 Notation

Before the proofs, we first introduce the notation used throughout the appendix and the general dTBM without symmetric assumptions. The parameter space and minimal gap assumption are also extended for the general asymmetric dTBM.

Preliminaries.

1. For mode $k \in [K]$, denote mode-k tensor matricizations by

$$\boldsymbol{Y}_k = \mathrm{Mat}_k(\mathcal{Y}), \quad \boldsymbol{S}_k = \mathrm{Mat}_k(\mathcal{S}), \quad \boldsymbol{E}_k = \mathrm{Mat}_k(\mathcal{E}), \quad \boldsymbol{X}_k = \mathrm{Mat}_k(\mathcal{X}).$$

2. For a vector $\boldsymbol{a}$, let $\boldsymbol{a}^s := \boldsymbol{a}/\|\boldsymbol{a}\|$ denote the normalized vector. We make the convention that $\boldsymbol{a}^s = \boldsymbol{0}$ if $\boldsymbol{a} = \boldsymbol{0}$.

3. For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$, let $\boldsymbol{A}^{\otimes K} := \boldsymbol{A} \otimes \cdots \otimes \boldsymbol{A} \in \mathbb{R}^{n^K \times m^K}$ denote the Kronecker product of K copies of matrices $\boldsymbol{A}$.

4. For a matrix $\boldsymbol{A}$, let $\|\boldsymbol{A}\|_\sigma$ denote the spectral norm of matrix $\boldsymbol{A}$, which is equal to the maximal singular value of $\boldsymbol{A}$; let $\lambda_k(\boldsymbol{A})$ denote the k-th largest singular value of $\boldsymbol{A}$; let $\|\boldsymbol{A}\|_F$ denote the Frobenius norm of matrix $\boldsymbol{A}$.

Extension to general asymmetric dTBM.

The general order-K $(p_1, \ldots, p_K)$-dimensional dTBM with r_k communities and degree heterogeneity $\boldsymbol{\theta}_k = [\![\theta_k(i)]\!] \in \mathbb{R}_+^{p_k}$ is represented by

$$\mathcal{Y} = \mathcal{X} + \mathcal{E}, \text{ where } \mathcal{X} = \mathcal{S} \times_1 \boldsymbol{\Theta}_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{\Theta}_K \boldsymbol{M}_K, \tag{26}$$

where $\mathcal{Y} \in \mathbb{R}^{p_1 \times \cdots \times p_K}$ is the data tensor, $\mathcal{X} \in \mathbb{R}^{p_1 \times \cdots \times p_K}$ is the mean tensor, $\mathcal{S} \in \mathbb{R}^{r_1 \times \cdots \times r_K}$ is the core tensor, $\mathcal{E} \in \mathbb{R}^{p_1 \times \cdots \times p_K}$ is the noise tensor consisting of independent zero-mean sub-Gaussian entries with variance bounded by σ^2, $\boldsymbol{\Theta}_k = \mathrm{diag}(\boldsymbol{\theta}_k)$, and $\boldsymbol{M}_k \in \{0,1\}^{p_k \times r_k}$ is the membership matrix corresponding to the assignment $z_k : [p_k] \mapsto [r_k]$, for all $k \in [K]$.

For ease of notation, we use $\{z_k\}$ to denote the collection $\{z_k\}_{k=1}^K$, and $\{\boldsymbol{\theta}_k\}$ to denote the collection $\{\boldsymbol{\theta}_k\}_{k=1}^K$. Correspondingly, we consider the parameter space for the triplet $(\{z_k\}, \mathcal{S}, \{\boldsymbol{\theta}_k\})$,

$$\begin{aligned}\mathcal{P}(\{r_k\}) = \Big\{ &(\{z_k\}, \mathcal{S}, \{\boldsymbol{\theta}_k\}) : \boldsymbol{\theta}_k \in \mathbb{R}_+^p, \frac{c_1 p_k}{r_k} |z_k^{-1}(a)| \leq \frac{c_2 p_k}{r_k}, \\ &c_3 \leq \|\boldsymbol{S}_{k,a:}\| \leq c_4, \|\boldsymbol{\theta}_{k,z_k^{-1}(a)}\|_1 = |z_k^{-1}(a)|, \text{for all } a \in [r_k], k \in [K] \Big\}.\end{aligned} \tag{27}$$

We call the degree heterogeneity $\{\boldsymbol{\theta}_k\}$ is balanced if for all $k \in [K]$,

$$\min_{a \in [r]} \|\boldsymbol{\theta}_{k,z_k^{-1}(a)}\| = (1 + o(1)) \max_{a \in [r]} \|\boldsymbol{\theta}_{k,z_k^{-1}(a)}\|.$$

We also consider the generalized Assumption 1 on angle gap.

Assumption 2 (Generalized angle gap)**.** Recall $\boldsymbol{S}_k = \text{Mat}_k(\mathcal{S})$. We assume the minimal gap between normalized rows of $\boldsymbol{S}_k$ is bounded away from zero for all $k \in [K]$; i.e.,

$$\Delta_{\min} \coloneqq \min_{k\in[K]} \min_{a\neq b\in[r_k]} \left\| \boldsymbol{S}^s_{k,a:} - \boldsymbol{S}^s_{k,b:} \right\| > 0.$$

Similarly, let $\text{SNR} = \Delta^2_{\min}/\sigma^2$ with the generalized minimal gap $\Delta^2_{\min}$ defined in Assumption 2. We define the regime

$$\mathcal{P}(\gamma) = \mathcal{P}(\{r_k\}) \cap \{\mathcal{S} \text{ satisfies SNR} = p^\gamma \text{ and } p_k \asymp p, k \in [K]\}.$$

B.2 Proof of Theorem 1

Proof of Theorem 1. To study the identifiability, we consider the noiseless model with $\mathcal{E} = 0$. Assume that there exist two parameterizations satisfying

$$\mathcal{X} = \mathcal{S} \times_1 \Theta_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \Theta_K \boldsymbol{M}'_K = \mathcal{S}' \times_1 \Theta'_1 \boldsymbol{M}'_1 \times_2 \cdots \times_K \Theta'_K \boldsymbol{M}'_K, \tag{28}$$

where $(\{z_k\}, \mathcal{S}, \{\boldsymbol{\theta}_k\}) \in \mathcal{P}(\{r_k\})$ and $(\{z'_k\}, \mathcal{S}', \{\boldsymbol{\theta}'_k\}) \in \mathcal{P}(\{r'_k\})$ are two sets of parameters. We prove the sufficient and necessary conditions separately.

($\Leftarrow$) For the necessity, it suffices to construct two distinct parameters up to cluster label permutation, if the model (26) violates Assumption 2. Note that $\Delta^2_{\min} = 1$ when there exists $k \in [K]$ such that $r_k = 1$. Hence, we consider the case that $r_k \geq 2$ for all $k \in [K]$. Without loss of generality, we assume $\left\| \boldsymbol{S}^s_{1,1:} - \boldsymbol{S}^s_{1,2:} \right\| = 0$.

By constraints in parameter space (27), neither $\boldsymbol{S}_{1,1:}$ nor $\boldsymbol{S}_{1,2:}$ is a zero vector. There exists a positive constant c such that $\boldsymbol{S}_{1,1:} = c\boldsymbol{S}_{1,2:}$. Thus, there exists a core tensor $\mathcal{S}_0 \in \mathbb{R}^{r_1-1\times\cdots\times r_K}$ such that

$$\mathcal{S} = \mathcal{S}_0 \times_1 \boldsymbol{C}\boldsymbol{R},$$

where $\boldsymbol{C} = \text{diag}(1, c, 1, ..., 1) \in \mathbb{R}^{r_1\times r_1}$ and

$$\boldsymbol{R} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & \mathbf{1}_{r_1-2} \end{pmatrix} \in \mathbb{R}^{r_1\times(r_1-1)}.$$

Let $\boldsymbol{D} = \text{diag}(1+c, 1, ..., 1) \in \mathbb{R}^{r_1-1\times r_1-1}$. Consider the parameterization $\boldsymbol{M}'_1 = \boldsymbol{M}_1\boldsymbol{R}, \mathcal{S}' = \mathcal{S}_0 \times_1 \boldsymbol{D}$, and

$$\theta'_1(i) = \begin{cases} \frac{1}{1+c}\theta_1(i) & i \in z_1^{-1}(1), \\ \frac{c}{1+c}\theta_1(i) & i \in z_1^{-1}(2), \\ \theta_1(i) & \text{otherwise}, \end{cases}$$

and $\boldsymbol{M}'_k = \boldsymbol{M}_k, \boldsymbol{\theta}'_k = \boldsymbol{\theta}_k$ for all $k = 2, \ldots, K$. Then we have constructed a triplet $(\{z'_k\}, \mathcal{S}', \{\boldsymbol{\theta}'_k\})$ that is distinct from $(\{z_k\}, \mathcal{S}, \{\boldsymbol{\theta}_k\})$ up to label permutation.

($\Rightarrow$) For the sufficiency, it suffices to show that all possible triplets $(\{z_k'\}, \mathcal{S}', \{\boldsymbol{\theta}_k'\})$ are identical to $(\{z_k\}, \mathcal{S}, \{\boldsymbol{\theta}_k\})$ up to label permutation if the model (26) satisfies Assumption (2). We show the uniqueness of the three parameters, $\{\boldsymbol{M}_k\}, \{\mathcal{S}\}, \{\boldsymbol{\theta}_k\}$ separately.

First, we show the uniqueness of $\boldsymbol{M}_k$ for all $k \in [K]$. When $r_k = 1$, all possible $\boldsymbol{M}_k$'s are equal to the vector $\mathbf{1}_{p_k}$, and the uniqueness holds trivially. Hence, we consider the case that $r_k \geq 2$. Without loss of generality, we consider $k = 1$ with $r_1 \geq 2$ and show the uniqueness of the first mode membership matrix; i.e., $\boldsymbol{M}_1' = \boldsymbol{M}_1 \boldsymbol{P}_1$ where $\boldsymbol{P}_1$ is a permutation matrix. The conclusion for $k \geq 2$ can be showed similarly and thus omitted.

Consider an arbitrary node pair (i, j). If $z_1(i) = z_1(j)$, then we have $\|\boldsymbol{X}^s_{1,z_1(i):} - \boldsymbol{X}^s_{1,z_1(j):}\| = 0$ and thus $\|(\boldsymbol{S}')^s_{1,z_1'(i):} - (\boldsymbol{S}')^s_{1,z_1'(j):}\| = 0$ by Lemma 3. Then, by Assumption (2), we have $z_1'(i) = z_1'(j)$. Conversely, if $z_1(i) \neq z_1(j)$, then we have $\left\|\boldsymbol{X}^s_{1,i:} - \boldsymbol{X}^s_{1,j:}\right\| \neq 0$ and thus $\left\|(\boldsymbol{S}')^s_{1,z_1'(i):} - (\boldsymbol{S}')^s_{1,z_1'(j):}\right\| \neq 0$ by Lemma 3. Hence, we have $z_1'(i) \neq z_1'(j)$. Therefore, we have proven that z_1' is identical z_i up to label permutation.

Next, we show the uniqueness of $\boldsymbol{\theta}_k$ for all $k \in [K]$ provided that $z_k = z_k'$. Similarly, consider $k = 1$ only, and omit the procedure for $k \geq 2$.

Consider an arbitrary $j \in [p_1]$ such that $z_1(j) = a$. Then for all the nodes $i \in z_1^{-1}(a)$ in the same cluster of j, we have

$$\frac{\boldsymbol{X}_{1,z_1(i):}}{\boldsymbol{X}_{1,z_1(j):}} = \frac{\boldsymbol{X}'_{1,z_1(i):}}{\boldsymbol{X}'_{1,z_1(j):}}, \text{ which implies } \frac{\theta_1(j)}{\theta_1(i)} = \frac{\theta_1'(j)}{\theta_1'(i)}. \tag{29}$$

Let $\theta_1'(j) = c\theta_1(j)$ for some positive constant c. By equation (29), we have $\theta_1'(i) = c\theta_1(i)$ for all $i \in z_1^{-1}(a)$. By the constraint $(\{z_k\}, \mathcal{S}', \{\boldsymbol{\theta}_k'\}) \in \mathcal{P}(\{r_k\})$, we have

$$\sum_{j \in z_1^{-1}(a)} \theta_1'(j) = c \sum_{j \in z_1^{-1}(a)} \theta_1(j) = 1,$$

which implies $c = 1$. Hence, we have proven $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1'$ provided that $z_1 = z_1'$.

Last, we show the uniqueness of $\mathcal{S}$; i.e., $\mathcal{S}' = \mathcal{S} \times_1 \boldsymbol{P}_1^{-1} \times_2 \cdots \times_K \boldsymbol{P}_K^{-1}$, where $\boldsymbol{P}_k$'s are permutation matrices for all $k \in [K]$. Provided $z_k' = z_k, \boldsymbol{\theta}_k' = \boldsymbol{\theta}_k$, we have $\boldsymbol{M}_k' = \boldsymbol{M}_k \boldsymbol{P}_k$ and $\boldsymbol{\Theta}_k' = \boldsymbol{\Theta}_k$ for all $k \in [K]$.

Let $\boldsymbol{D}_k = \left[(\boldsymbol{\Theta}_k'\boldsymbol{M}_k')^T(\boldsymbol{\Theta}_k'\boldsymbol{M}_k')\right]^{-1}(\boldsymbol{\Theta}_k'\boldsymbol{M}_k')^T, k \in [K]$. By the parameterization (28), we have

$$\begin{aligned}
\mathcal{S}' &= \mathcal{X} \times_1 \boldsymbol{D}_1 \times_2 \cdots \times_K \boldsymbol{D}_K \\
&= \mathcal{S} \times_1 \boldsymbol{D}_1 \boldsymbol{\Theta}_1 \boldsymbol{M}_1 \times_1 \cdots \times_K \boldsymbol{D}_K \boldsymbol{\Theta}_K \boldsymbol{M}_K \\
&= \mathcal{S} \times_1 \boldsymbol{P}_1^{-1} \times_2 \cdots \times_K \boldsymbol{P}_K^{-1}.
\end{aligned}$$

Therefore, we finish the proof of Theorem 1. $\square$

Useful Lemma for the Proof of Theorem 1

Lemma 3 (Motivation of angle-based clustering)**.** Consider the signal tensor $\mathcal{X}$ in the general asymmetric dTBM (26) with $(\{z_k\}, \mathcal{S}, \{\boldsymbol{\theta}_k\}) \in \mathcal{P}(\{r_k\})$ and $r_k \geq 2, k \in [K]$. Then, for any $k \in [K]$ and index pair $(i,j) \in [p_k]^2$, we have

$$\left\| \boldsymbol{S}^s_{k,z_k(i):} - \boldsymbol{S}^s_{k,z_k(j):} \right\| = 0 \quad \text{if and only if} \quad \left\| \boldsymbol{X}^s_{k,z_k(i):} - \boldsymbol{X}^s_{k,z_k(j):} \right\| = 0.$$

Proof of Lemma 3. Without loss of generality, we prove $k=1$ only and drop the subscript k in $\boldsymbol{X}_k, \boldsymbol{S}_k$ for notational convenience. By tensor matricization, we have

$$\boldsymbol{X}_{j:} = \theta_1(j) \boldsymbol{S}_{z_1(j):} \left[\boldsymbol{\Theta}_2 \boldsymbol{M}_2 \otimes \cdots \otimes \boldsymbol{\Theta}_K \boldsymbol{M}_K \right]^T.$$

Let $\tilde{\boldsymbol{M}} = \boldsymbol{\Theta}_2 \boldsymbol{M}_2 \otimes \cdots \otimes \boldsymbol{\Theta}_K \boldsymbol{M}_K$. Notice that for two vectors $\boldsymbol{a}, \boldsymbol{b}$ and two positive constants $c_1, c_2 > 0$, we have

$$\|\boldsymbol{a}^s - \boldsymbol{b}^s\| = \|(c_1 \boldsymbol{a})^s - (c_2 \boldsymbol{b})^s\|.$$

Thus it suffices to show the following statement holds for any index pair $(i,j) \in [p_1]^2$,

$$\left\| \boldsymbol{S}^s_{z_1(i):} - \boldsymbol{S}^s_{z_1(j):} \right\| = 0 \quad \text{if and only if} \quad \left\| \left[\boldsymbol{S}_{z_1(i):} \tilde{\boldsymbol{M}}^T \right]^s - \left[\boldsymbol{S}_{z_1(j):} \tilde{\boldsymbol{M}}^T \right]^s \right\| = 0.$$

$(\Leftarrow)$ Suppose $\left\| \left[\boldsymbol{S}_{z_1(i):} \tilde{\boldsymbol{M}}^T \right]^s - \left[\boldsymbol{S}_{z_1(j):} \tilde{\boldsymbol{M}}^T \right]^s \right\| = 0$. There exists a positive constant c such that $\boldsymbol{S}_{z_1(i):} \tilde{\boldsymbol{M}}^T = c \boldsymbol{S}_{z_1(j):} \tilde{\boldsymbol{M}}^T$. Note that

$$\boldsymbol{S}_{z_1(i):} = \boldsymbol{S}_{z_1(i):} \tilde{\boldsymbol{M}}^T \left[\tilde{\boldsymbol{M}} \left(\tilde{\boldsymbol{M}}^T \tilde{\boldsymbol{M}} \right)^{-1} \right],$$

where $\tilde{\boldsymbol{M}}^T \tilde{\boldsymbol{M}}$ is an invertiable diagonal matrix with positive diagonal elements. Thus, we have $\boldsymbol{S}_{z_1(i):} = c \boldsymbol{S}_{z_1(j):}$, which implies $\left\| \boldsymbol{S}^s_{z_1(i):} - \boldsymbol{S}^s_{z_1(j):} \right\| = 0$.

$(\Rightarrow)$ Suppose $\left\| \boldsymbol{S}^s_{z_1(i):} - \boldsymbol{S}^s_{z_1(j):} \right\| = 0$. There exists a positive constant c such that $\boldsymbol{S}_{z_1(i):} = c \boldsymbol{S}_{z_1(j):}$, and thus $\boldsymbol{S}_{z_1(i):} \tilde{\boldsymbol{M}}^T = c \boldsymbol{S}_{z_1(j):} \tilde{\boldsymbol{M}}^T$, which implies $\left\| \left[\boldsymbol{S}_{z_1(i):} \tilde{\boldsymbol{M}}^T \right]^s - \left[\boldsymbol{S}_{z_1(j):} \tilde{\boldsymbol{M}}^T \right]^s \right\| = 0$.

Therefore, we finish the proof of Lemma 3. $\square$

B.3 Proof of Lemma 1 and Lemma 2

Proof of Lemma 1. Note that the vector $\boldsymbol{S}_{z(i):}$ can be folded to a tensor $\mathcal{S}' = [\![\mathcal{S}'_{a_2,\ldots,a_K}]\!] \in \mathbb{R}^{r^{K-1}}$; i.e., $\text{vec}(\mathcal{S}') = \boldsymbol{S}_{z(i):}$. Define weight vectors $\boldsymbol{w}_{a_2,\cdots,a_K}$ corresponding to the elements in $\mathcal{S}'_{a_2,\ldots,a_K}$ by

$$\boldsymbol{w}_{a_2 \cdots a_K} = [\boldsymbol{\theta}^T_{z^{-1}(a_2)} \otimes \cdots \otimes \boldsymbol{\theta}^T_{z^{-1}(a_K)}] \in \mathbb{R}^{|z^{-1}(a_2)| \times \cdots \times |z^{-1}(a_K)|},$$

for all $a_k \in [r], k = 2, \ldots, K$, where $\otimes$ denotes the Kronecker product. Therefore, we have $\boldsymbol{X}_{i:} = \theta(i) \text{Pad}_{\boldsymbol{w}}(\boldsymbol{S}_{z(i):})$ where $\boldsymbol{w} = \{\boldsymbol{w}_{a_2,\cdots,a_K}\}_{a_k \in [r], k \in [K]/\{1\}}$. Specifically, we have $\|\boldsymbol{w}_{a_2,\ldots,a_K}\|^2 = \prod_{k=2}^K \|\boldsymbol{\theta}_{z^{-1}(a_k)}\|^2$, and by the balanced assumption (6) we have

$$\max_{(a_2,\ldots,a_K)} \|\boldsymbol{w}_{a_2,\ldots,a_K}\|^2 = (1+o(1)) \min_{(a_2,\ldots,a_K)} \|\boldsymbol{w}_{a_2,\ldots,a_K}\|^2. \tag{30}$$

Consider the inner product of $\boldsymbol{X}_{i:}$ and $\boldsymbol{X}_{j:}$ for $z(i) \neq z(j)$. By the definition of weighted padding operator (56) and the balanced assumption (30), we have

$$\begin{aligned}\langle \boldsymbol{X}_{i:}, \boldsymbol{X}_{j:}\rangle &= \theta(i)\theta(j)\left\langle \mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{S}_{z(i):}), \mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{S}_{z(j):})\right\rangle \\ &= \theta(i)\theta(j) \min_{(a_2,\ldots,a_K)} \|\boldsymbol{w}_{a_2,\ldots,a_K}\|^2 \left\langle \boldsymbol{S}_{z(i):}, \boldsymbol{S}_{z(j):}\right\rangle (1+o(1)).\end{aligned}$$

Therefore, when p large enough, the inner product $\langle \boldsymbol{X}_{i:}, \boldsymbol{X}_{j:}\rangle$ has the same sign as $\left\langle \boldsymbol{S}_{z(i):}, \boldsymbol{S}_{z(j):}\right\rangle$.

Then, we have

$$\cos(\boldsymbol{S}_{z_1(i):}, \boldsymbol{S}_{z_1(j):}) = \frac{\left\langle \boldsymbol{S}_{z_1(i):}, \boldsymbol{S}_{z_1(j):}\right\rangle}{\|\boldsymbol{S}_{z_1(i):}\|\|\boldsymbol{S}_{z_1(j):}\|} = (1+o(1))\frac{\langle \boldsymbol{X}_{i:}, \boldsymbol{X}_{j:}\rangle}{\|\boldsymbol{X}_{i:}\|\|\boldsymbol{X}_{j:}\|} = (1+o(1))\cos(\boldsymbol{X}_{i:}, \boldsymbol{X}_{j:}),$$

where the second inequality follows by the balance assumption on $\boldsymbol{\theta}$.

Further, notice that $\|\boldsymbol{v}_1^s - \boldsymbol{v}_2^s\|^2 = 2(1-\cos(\boldsymbol{v}_1, \boldsymbol{v}_2))$. For all i,j such that $z(i) \neq z(j)$, when $p \to \infty$, we have

$$\|\boldsymbol{X}_{i:}^s - \boldsymbol{X}_{j:}^s\| \asymp \|\boldsymbol{S}_{z_1(i):}^s - \boldsymbol{S}_{z_1(j):}^s\| \gtrsim \Delta_{\min}.$$

□

Proof of Lemma 2. By the definition of minimal gap in Assumption 1, we have

$$\begin{aligned}L^{(t)} &= \frac{1}{p}\sum_{i\in[p]} \theta(i) \sum_{b\in[r]} \mathbb{1}\left\{z^{(t)}(i) = b\right\} \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \\ &\geq \frac{1}{p}\sum_{i\in[p]} \theta(i) \sum_{b\in[r]} \mathbb{1}\left\{z^{(t)}(i) = b\right\} \Delta_{\min}^2 \\ &\geq c\ell^{(t)}\Delta_{\min}^2,\end{aligned}$$

where the last inequality follows from the assumption $\min_{i\in[p]}\theta(i) \geq c > 0$. □

B.4 Proof of Theorem 2 (Impossibility)

Proof of Theorem 2 (Impossibility). Consider the general asymmetric dTBM (26) in the special case that $p_k = p$ and $r_k = r$ for all $k \in [K]$ with $K \geq 2$, $2 \leq r \lesssim p^{1/3}$ as $p \to \infty$. For simplicity, we show the minimax rate for the estimation on the first mode $\hat{z}_1$; the proof for other modes are essentially the same.

To prove the minimax rate (10), it suffices to take an arbitrary $\mathcal{S}^* \in \mathcal{P}_{\mathcal{S}}(\gamma)$ wih $\gamma < -(K-1)$ and construct $(z_k^*, \boldsymbol{\theta}_k^*)$ such that

$$\inf_{\hat{z}_1} \mathbb{E}\left[p\ell(\hat{z}_1, z_1^*)|(z_k^*, \mathcal{S}^*, \boldsymbol{\theta}_k^*)\right] \geq 1.$$

We first define a subset of indices $T_k \subset [p_k], k \in [K]$ in order to avoid the complication of label permutation. Based on Han et al. (2022a, Proof of Theorem 6), we consider the restricted family of $\hat{z}_k$'s for which the following three conditions are satisfied:

$$\text{(a) } \hat{z}_k(i) = z_k(i) \text{ for all } i \in T_k; \quad \text{(b) } |T_k^c| \asymp \frac{p}{r};$$

$$\text{(c)} \min_{\pi\in\Pi}\sum_{i\in[p]}\mathbb{1}\{\hat{z}_k(i)\neq \pi\circ z_k(i)\} = \sum_{i\in[p]}\mathbb{1}\{\hat{z}_k(i)\neq z_k(i)\},$$

for all $k \in [K]$. Now, we consider the construction:

(i) $\{z_k^*\}$ satisfies properties (a)-(c) with misclassification sets T_k^c for all $k \in [K]$;

(ii) $\{\boldsymbol{\theta}_k^*\}$ such that $\boldsymbol{\theta}_k^*(i) \leq \sigma r^{(K-1)/2} p^{-(K-1)/2}$ for all $i \in T_k^c, k \in [K]$ and $\max_{k\in[K],a\in[r]}\|\boldsymbol{\theta}_{k,z_k^{*,-1}(a)}\|_2^2 \asymp p/r$.

Combining the inequalities (39) and (40) in the proof of Theorem 2 in Gao et al. (2018), we have

$$\begin{aligned}
&\inf_{\hat{z}_1}\mathbb{E}\left[\ell(\hat{z}_1, z_1^*)|(z_k^*, \mathcal{S}^*, \boldsymbol{\theta}_k^*)\right] \geq \\
&\quad \frac{C}{r^3|T_1^c|}\sum_{i\in T_1^c}\inf_{\hat{z}_1(i)}\{\mathbb{P}[\hat{z}_1(i)=1|z_1^*(i)=2, z_k^*, \mathcal{S}^*, \boldsymbol{\theta}_k^*] + \mathbb{P}[\hat{z}_1(i)=2|z_1^*(i)=1, z_k^*, \mathcal{S}^*, \boldsymbol{\theta}_k^*]\}, \quad (31)
\end{aligned}$$

where C is some positive constant, $\hat{z}_1$ on the left hand side denote the generic assignment functions in $\mathcal{P}(\gamma)$, and the infimum on the right hand side is taken over the generic assignment function family of $\hat{z}_1(i)$ for all nodes $i \in T_1^c$. Here, the factor $r^3 = r \cdot r^2$ in (31) comes from two sources: $r^2 \asymp \binom{r}{2}$ comes from the multiple testing burden for all pairwise comparisons among r clusters; and another r comes from the number of elements $|T_k^c| \asymp p/r$ to be clustered.

Next, we need to find the lower bound of the rightmost side in (31). We consider the hypothesis test based on model (26). First, we reparameterize the model under the construction (i)-(ii).

$$\boldsymbol{x}_a^* = \left[\mathrm{Mat}_1\left(\mathcal{S}^* \times_2 \boldsymbol{\Theta}_2^*\boldsymbol{M}_2^* \times_3 \cdots \times_K \boldsymbol{\Theta}_K^*\boldsymbol{M}_K^*\right)\right]_{a:},$$

for all $a \in [r]$, where $\boldsymbol{x}_a^*$'s are centroids in $\mathbb{R}^{p^{K-1}}$. Without loss of generality, we consider the lower bound for the summand in (31) for $i = 1$. The analysis for other $i \in T_1^c$ are similar. For notational simplicity, we suppress the subscript i and write $\boldsymbol{y}, \theta^*, z$ in place of $\boldsymbol{y}_1, \boldsymbol{\theta}_1^*(1)$ and $z_1(1)$, respectively. The equivalent vector problem for assessing the summand in (31) is

$$\boldsymbol{y} = \theta^*\boldsymbol{x}_z^* + \boldsymbol{e}, \tag{32}$$

where $z \in \{1, 2\}$ is an unknown parameter, $\theta^* \in \mathbb{R}_+$ is the given heterogeneity degree, $\boldsymbol{x}_1^*, \boldsymbol{x}_2^* \in \mathbb{R}^{p^{K-1}}$ are given centroids, and $\boldsymbol{e} \in \mathbb{R}^{p^{K-1}}$ consists of i.i.d. $N(0, \sigma^2)$ entries. Then, we consider the hypothesis testing under the model (32):

$$H_0: z = 1, \boldsymbol{y} = \theta^*\boldsymbol{x}_1^* + \boldsymbol{e} \ \leftrightarrow\ H_1: z = 2, \boldsymbol{y} = \theta^*\boldsymbol{x}_2^* + \boldsymbol{e}, \tag{33}$$

The hypothesis testing (33) is a simple versus simple testing, since the assignment z is the only unknown parameter in the test. By Neyman-Pearson lemma, the likelihood ratio test is optimal with minimal Type I + II error. Under Gaussian model, the likelihood ratio test of (33) is equivalent to the least square estimator $\hat{z}_{LS} = \arg\min_{a=\{1,2\}}\|\boldsymbol{y} - \theta^*\boldsymbol{x}_a^*\|_F^2$.

Let $\boldsymbol{S} = \mathrm{Mat}_1(\mathcal{S})$. Note that

$$\|\theta^*\boldsymbol{x}_1^* - \theta^*\boldsymbol{x}_2^*\|_F \leq \theta^*\|\boldsymbol{S}_{1:}^* - \boldsymbol{S}_{2:}^*\|_F\prod_{k=2}^K\lambda_{\max}(\boldsymbol{\Theta}_k^*\boldsymbol{M}_k^*)$$

$$
\begin{aligned}
&\leq \theta^* \|\boldsymbol{S}^*_{1:} - \boldsymbol{S}^*_{2:}\|_F \max_{k\in[K]/\{1\}, a\in[r]} \|\boldsymbol{\theta}_{k, z_k^{*,-1}(a)}\|_2^{K-1} \\
&\leq \sigma r^{(K-1)/2} p^{-(K-1)/2} 2c_4 p^{(K-1)/2} r^{-(K-1)/2} \\
&\leq 2c_4\sigma,
\end{aligned}
$$

where $\lambda_{\max}(\cdot)$ denotes the maximal singular value, the second inequality follows from Lemma 6, and the third inequality follows from property (ii) and the boundedness constraint in $\mathcal{P}_{\mathcal{S}}(\gamma)$ such that $\|\boldsymbol{S}^*_{1:} - \boldsymbol{S}^*_{2:}\|_F \leq \|\boldsymbol{S}^*_{1:}\|_F + \|\boldsymbol{S}^*_{2:}\|_F \leq 2c_4$.

Hence, we have

$$
\begin{aligned}
&\inf_{\hat{z}_1(1)} \{\mathbb{P}[\hat{z}_1(1) = 1 | z_1^*(1) = 2, z_k^*, \mathcal{S}^*, \boldsymbol{\theta}_k^*] + \mathbb{P}[\hat{z}_1(1) = 2 | z_1^*(1) = 1, z_k^*, \mathcal{S}^*, \boldsymbol{\theta}_k^*]\} \\
&\quad = 2\mathbb{P}[\hat{z}_{LS} = 1 | z_1^*(1) = 2, z_k^*, \mathcal{S}^*, \boldsymbol{\theta}_k^*] \\
&\quad = 2\mathbb{P}[\|\boldsymbol{y} - \theta^* \boldsymbol{x}_1^*\|_F^2 \leq \|\boldsymbol{y} - \theta^* \boldsymbol{x}_2^*\|_F^2 | z_1^*(1) = 2, z_k^*, \mathcal{S}^*, \boldsymbol{\theta}_k^*] \\
&\quad = 2\mathbb{P}[2\langle \boldsymbol{e}, \theta^* \boldsymbol{x}_1^* - \theta^* \boldsymbol{x}_2^*\rangle \geq \|\theta^* \boldsymbol{x}_1^* - \theta^* \boldsymbol{x}_2^*\|_F^2] \\
&\quad = 2\mathbb{P}[N(0,1) \geq \theta^* \|\boldsymbol{x}_1^* - \boldsymbol{x}_2^*\|_F / (2\sigma)] \\
&\quad \geq 2\mathbb{P}[N(0,1) \geq c_4] \geq c,
\end{aligned} \tag{34}
$$

where the first equation holds by symmetry, the third equation holds by rearrangement, the fourth equation holds from the fact that $\langle \boldsymbol{e}, \theta^* \boldsymbol{x}_1^* - \theta^* \boldsymbol{x}_2^*\rangle \sim N(0, \sigma\|\theta^* \boldsymbol{x}_1^* - \theta^* \boldsymbol{x}_2^*\|_F)$, and c is some positive constant in the last inequality.

Plugging the inequality (34) into the inequality (31) for all $i \in T_1^c$, then, we have

$$
\liminf_{p\to\infty} \inf_{\hat{z}_1} \mathbb{E}\left[p\ell(\hat{z}_1, z_1^*) | z_k^*, \boldsymbol{\theta}_k^*, \mathcal{S}^*\right] \geq \liminf_{p\to\infty} \frac{Ccp}{r^3} \geq Cc,
$$

where the last inequality follows by the condition $r = o(p^{1/3})$. By the discrete nature of the misclustering error, we obtain our conclusion

$$
\liminf_{p\to\infty} \inf_{\mathcal{S}^* \in \mathcal{P}_{\mathcal{S}}(\gamma)} \inf_{\hat{z}_{\text{stat}}} \sup_{(z^*, \boldsymbol{\theta}^*) \in \mathcal{P}_{z,\boldsymbol{\theta}}} \mathbb{E}\left[p\ell(\hat{z}_{\text{stat}}, z)\right] \geq 1.
$$

Last, with constructed $z_k^*, \boldsymbol{\theta}_k^*$ satisfying properties (i) and (ii) and $\gamma' < -(K-1)$, we construct a core tensor $\mathcal{S}^*$ such that $\Delta^2_{\boldsymbol{X}^*} \leq p^{-(K-1)}$. Based on the property (ii) and the boundedness constraint of $\mathcal{S}^*$ in $\mathcal{P}$, we still have $\|\theta^* \boldsymbol{x}_1^* - \theta^* \boldsymbol{x}_2^*\|_F \leq 2c_4\sigma$. Hence, we obtain the desired result

$$
\liminf_{p\to\infty} \inf_{\hat{z}_1} \sup_{(z,\mathcal{S},\boldsymbol{\theta}) \in \mathcal{P}'(\gamma')} \mathbb{E}\left[p\ell(\hat{z}_1, z_1)\right] \geq \liminf_{p\to\infty} \inf_{\hat{z}_{\text{stat}}} \mathbb{E}\left[p\ell(\hat{z}_1, z_1^*) | z_k^*, \mathcal{S}^*, \boldsymbol{\theta}_k^*\right] \geq 1.
$$

□

B.5 Proof of Theorem 3 (Impossibility)

Proof of Theorem 3 (Impossibility). The idea of proving computational hardness is to show the computational lower bound for a special class of degree-corrected tensor clustering model with $K \geq 2$ and $r \geq 2$. We construct the following special class of higher-order degree-corrected tensor clustering model. For a given signal level $\gamma \in \mathbb{R}$ and noise variance σ, define a rank-2 symmetric tensor $\mathcal{S} \in \mathbb{R}^{3\times\cdots\times 3}$ subject to

$$\mathcal{S} = \mathcal{S}(\gamma) = \begin{bmatrix}1\\1\\1\end{bmatrix}^{\otimes K} + \sigma p^{-\gamma/2}\begin{bmatrix}1\\-1\\0\end{bmatrix}^{\otimes K}. \tag{35}$$

Then, we consider the signal tensor family

$$\mathcal{P}_{\text{shifted}}(\gamma) = \{\mathcal{X}\colon \mathcal{X} = \mathcal{S} \times_1 \boldsymbol{M}_1 \times_2 \cdots \times_K \boldsymbol{M}_K,\ \boldsymbol{M}_k \in \{0,1\}^{p\times 3} \text{ is a membership matrix that satisfies } |\boldsymbol{M}_k(:\,,i)| \asymp p \text{ for all } i \in [3] \text{ and } k \in [K]\}.$$

We claim that the constructed family satisfies the following two properties:

(i) For every $\gamma \in \mathbb{R}$, $\mathcal{P}_{\text{shifted}}(\gamma) \subset \mathcal{P}(\gamma)$, where $\mathcal{P}(\gamma)$ is the degree-corrected cluster tensor family (5).

(ii) For every $\gamma \in \mathbb{R}$, $\{\mathcal{X} - 1\colon \mathcal{X} \in \mathcal{P}_{\text{shifted}}(\gamma)\} \subset \mathcal{P}_{\text{non-degree}}(\gamma)$, where $\mathcal{P}_{\text{non-degree}}(\gamma)$ denotes the sub-family of rank-one tensor block model constructed in proof of Han et al. (2022a, Theorem 7).

The verification of the above two properties is provided in the end of this proof.

Now, following the proof of Han et al. (2022a, Theorem 7), when $\gamma < -K/2$, every polynomial-time algorithm estimator $(\hat{\boldsymbol{M}}_k)_{k\in[K]}$ obeys

$$\liminf_{p\to\infty} \sup_{\mathcal{X}\in\mathcal{P}_{\text{non-degree}}(\gamma)} \mathbb{P}(\exists k \in [K],\ \hat{\boldsymbol{M}}_k \neq \boldsymbol{M}_k) \geq 1/2, \tag{36}$$

under the HPC Conjecture 1. The inequality (36) implies

$$\liminf_{p\to\infty} \sup_{\mathcal{X}\in\mathcal{P}_{\text{non-degree}}(\gamma)} \max_{k\in[K]} \mathbb{E}[p\ell(z_k, \hat{z}_k)] \geq 1.$$

Based on properties (i)-(ii), we conclude that

$$\liminf_{p\to\infty} \sup_{\mathcal{X}\in\mathcal{P}(\gamma)} \max_{k\in[K]} \mathbb{E}[p\ell(z_k, \hat{z}_k)] \geq 1.$$

We complete the proof by verifying the properties (i)-(ii). For (i), we verify that the angle gap for the core tensor $\mathcal{S}$ in (35) is on the order of $\sigma p^{-\gamma/2}$. Specifically, write $\mathbf{1} = (1,1,1)$ and $\boldsymbol{e} = (1,-1,0)$. We have

$$\text{Mat}(\mathcal{S}) = \begin{bmatrix}\text{Vec}(\mathbf{1}^{\otimes K-1}) + \sigma p^{-\gamma/2}\text{Vec}\left(\boldsymbol{e}^{\otimes(K-1)}\right)\\ \text{Vec}(\mathbf{1}^{\otimes K-1}) - \sigma p^{-\gamma/2}\text{Vec}\left(\boldsymbol{e}^{\otimes(K-1)}\right)\\ \text{Vec}(\mathbf{1}^{\otimes K-1})\end{bmatrix}.$$

Based on the orthogonality $\langle \mathbf{1}, \boldsymbol{e}\rangle = 0$, the minimal angle gap among rows of $\text{Mat}(\mathcal{S})$ is

$$\Delta^2_{\min}(\mathcal{S}) \asymp \tan^2(\text{Mat}(\mathcal{S})_{1:}, \text{Mat}(\mathcal{S})_{3:}) = \left(\frac{\|\boldsymbol{e}\|_2}{\|\mathbf{1}\|_2}\right)^{2(K-1)} \sigma^2 d^{-\gamma} \asymp \sigma^2 d^{-\gamma}.$$

Therefore, we have shown that $\mathcal{P}_{\text{shifited}}(\gamma) = \mathcal{P}(\gamma)$. Finally, the property (ii) follows directly by comparing the definition of $\mathcal{S}$ in (35) with that in the proof of Han et al. (2022a, Theorem 7). $\square$

B.6 Proof of Theorem 4 and Proposition 1

Proof of Theorem 4. We prove Theorem 4 under the dTBM (1) with symmetric mean tensor, parameters $(z, \mathcal{S}, \boldsymbol{\theta})$, fixed $r \geq 1, K \geq 2$, and i.i.d. noise. For the case $r = 1$, we have $L(z^{(0)}, z) = 0, \ell(z^{(0)}, z) = 0$ trivially. Hence, we focus on the proof of the first mode clustering $z_1^{(0)}$ with $r \geq 2$; the proofs for the other modes can be extended similarly. We drop the subscript k in the matricizations $\boldsymbol{M}_k, \boldsymbol{X}_k, \boldsymbol{S}_k$ and in the estimate $z_1^{(0)}$. We firstly show the proof with balanced $\boldsymbol{\theta}$.

We firstly show the upper bound for misclustering error $\ell(z^{(0)}, z)$.

First, by Lemma 1, there exists a positive constant such that $\min_{z(i) \neq z(j)} \left\| \boldsymbol{X}_{i:}^s - \boldsymbol{X}_{j:}^s \right\| \geq c_0 \Delta_{\min}$. By the balance assumption on $\boldsymbol{\theta}$ and Lemma 8, we have

$$\min_{\pi \in \Pi} \sum_{i: z^{(0)}(i) \neq \pi(z(i))} \theta(i)^2 \leq \sum_{i \in S_I} \theta(i)^2 + 4 \sum_{i \in S} \theta(i)^2, \tag{37}$$

where

$$S_0 = \{i : \|\hat{\boldsymbol{X}}_{i:}\| = 0\}, S = \{i \in S_0^c : \|\hat{\boldsymbol{x}}_{z^{(0)}(i)} - \boldsymbol{X}_{i:}^s\| \geq c_0 \Delta_{\min}/2\}.$$

On one hand, note that for any set $P \in [p]$,

$$\begin{aligned} \sum_{i \in P} \|\boldsymbol{X}_{i:}\|^2 &= \sum_{i \in P} \|\theta(i) \boldsymbol{S}_{z(i):} (\boldsymbol{\Theta} \boldsymbol{M})^{T, \otimes(K-1)}\|^2 \\ &\geq \sum_{i \in P} \theta(i)^2 \min_{a \in [r]} \|\boldsymbol{S}_{a:}\|^2 \lambda_r^{2(K-1)}(\boldsymbol{\Theta} \boldsymbol{M}) \\ &\gtrsim \sum_{i \in P} \theta(i)^2 p^{K-1} r^{-(K-1)}, \end{aligned}$$

where the last inequality follows Lemma 6, the assumption that $\min_{i \in [p]} \theta(i) \geq c$, and the constraint $\min_{a \in [r]} \|\boldsymbol{S}_{a:}\| \geq c_3$ in the parameter space (2). Thus, we have

$$\sum_{i \in P} \theta(i)^2 \lesssim \sum_{i \in P} \|\boldsymbol{X}_{i:}\|^2 p^{-(K-1)} r^{K-1}. \tag{38}$$

On the other hand, note that

$$\begin{aligned} \sum_{i \in S} \|\boldsymbol{X}_{i:}\|^2 &\leq 2 \sum_{i \in S} \|\hat{\boldsymbol{X}}_{i:}\|^2 + 2 \sum_{i \in S} \|\hat{\boldsymbol{X}}_{i:} - \boldsymbol{X}_{i:}\|^2 & (39) \\ &\leq \frac{8}{c_0^2 \Delta_{\min}^2} \sum_{i \in S} \|\hat{\boldsymbol{X}}_{i:}\|^2 \|\hat{\boldsymbol{x}}_{z^{(0)}(i)} - \boldsymbol{X}_{i:}^s\|^2 + 2 \|\hat{\mathcal{X}} - \mathcal{X}\|_F^2 & (40) \\ &\leq \frac{16}{c_0^2 \Delta_{\min}^2} \sum_{i \in S} \|\hat{\boldsymbol{X}}_{i:}\|^2 \left[\|\hat{\boldsymbol{x}}_{z^{(0)}(i)} - \hat{\boldsymbol{X}}_{i:}^s\|^2 + \|\hat{\boldsymbol{X}}_{i:}^s - \boldsymbol{X}_{i:}^s\|^2 \right] + 2 \|\hat{\mathcal{X}} - \mathcal{X}\|_F^2 & (41) \\ &\leq \frac{16(1+\eta)}{c_0^2 \Delta_{\min}^2} \sum_{i \in S} \|\hat{\boldsymbol{X}}_{i:}\|^2 \|\hat{\boldsymbol{X}}_{i:}^s - \boldsymbol{X}_{i:}^s\|^2 + 2 \|\hat{\mathcal{X}} - \mathcal{X}\|_F^2 & (42) \\ &\leq \left(\frac{16(1+\eta)}{c_0^2 \Delta_{\min}^2} + 2 \right) \|\hat{\mathcal{X}} - \mathcal{X}\|_F^2 & (43) \end{aligned}$$

$$\lesssim \left(\frac{16(1+\eta)}{c_0^2\Delta_{\min}^2}+2\right)\left(p^{K/2}r+pr^2+r^K\right)\sigma^2, \tag{44}$$

where inequalities (39) and (41) follow from the triangle inequality, (40) follows from the definition of S, (42) follows from the update rule of k-means in Step 6 of Sub-algorithm 1, (43) follows from Lemma 4, and the last inequality (44) follows from Lemma 7. Also, note that

$$\sum_{i\in S_0}\|\boldsymbol{X}_{i:}\|^2=\sum_{i\in S_0}\|\hat{\boldsymbol{X}}_{i:}-\boldsymbol{X}_{i:}\|^2\le\|\hat{\mathcal{X}}-\mathcal{X}\|_F^2\lesssim\left(p^{K/2}r+pr^2+r^K\right)\sigma^2, \tag{45}$$

where the equation follows from the definition of S_0. Therefore, combining the inequalities (37), (38), (44), and (45), we have

$$\begin{aligned}\min_{\pi\in\Pi}\sum_{i:z^{(0)}(i)\neq\pi(z(i))}\theta(i)^2&\lesssim\left(\sum_{i\in S}\|\boldsymbol{X}_{i:}\|^2+\sum_{i\in S_0}\|\boldsymbol{X}_{i:}\|^2\right)p^{-(K-1)}r^{K-1}\\&\lesssim\frac{\sigma^2r^{K-1}}{\Delta_{\min}^2p^{K-1}}\left(p^{K/2}r+pr^2+r^K\right).\end{aligned} \tag{46}$$

With the assumption that $\min_{i\in[p]}\theta(i)\ge c$, we finally obtain the result

$$\ell(z^{(0)},z)\lesssim\frac{1}{p}\min_{\pi\in\Pi}\sum_{i:z^{(0)}(i)\neq\pi(z(i))}\theta(i)^2\lesssim\frac{r^Kp^{-K/2}}{\text{SNR}},$$

where the last inequality follows from the definition $\text{SNR}=\Delta_{\min}^2/\sigma^2$.

Without the balanced $\boldsymbol{\theta}$, we have $\min_{z(i)\neq z(j)}\left\|\boldsymbol{X}_{i:}^s-\boldsymbol{X}_{j:}^s\right\|\ge c_0\Delta_{\boldsymbol{X}}$. Replacing the definition of S with $\Delta_{\boldsymbol{X}}$, we obtain the desired result.

Next, we show the bound for $L(z^{(0)},z)$.

Note that $\boldsymbol{X}_{i:}^s$ have only r different values. We let $\boldsymbol{X}_a^s=\boldsymbol{X}_{i:}^s$ for all i such that $z(i)=a, a\in[r]$. Notice that

$$\|\boldsymbol{X}_{i:}\|^2\gtrsim p^{K-1}r^{-(K-1)}$$

and

$$\|\boldsymbol{X}_{i:}-\hat{\boldsymbol{X}}_{i:}\|^2\le\|\hat{\mathcal{X}}-\mathcal{X}\|_F^2\lesssim p^{K/2}r+pr^2+r^K.$$

Therefore, when p is large enough, we have

$$\begin{aligned}\sum_{i\in[p]}\|\boldsymbol{X}_{i:}\|^2\|\hat{\boldsymbol{X}}_i^s-\hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2&\lesssim\sum_{i\in[p]}\left(\|\boldsymbol{X}_{i:}\|^2-\|\boldsymbol{X}_{i:}-\hat{\boldsymbol{X}}_{i:}\|^2\right)\|\hat{\boldsymbol{X}}_{i:}^s-\hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2\\&\lesssim\sum_{i\in[p]}\|\hat{\boldsymbol{X}}_{i:}\|^2\|\hat{\boldsymbol{X}}_{i:}^s-\hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2\\&\lesssim\eta\sum_{i\in[p]}\|\hat{\boldsymbol{X}}_{i:}\|^2\|\hat{\boldsymbol{X}}_{i:}^s-\boldsymbol{X}_{i:}^s\|^2\\&\lesssim\|\hat{\mathcal{X}}-\mathcal{X}\|_F^2\end{aligned}$$

$$\lesssim p^{K/2}r + pr^2 + r^K. \tag{47}$$

Hence, we have

$$\begin{aligned}
\sum_{i\in[p]} \|\hat{\boldsymbol{X}}_{i:}^s - \hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2 &\lesssim \sum_{i\in[p]} \theta(i)^2 \|\hat{\boldsymbol{X}}_i^s - \hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2 \\
&\lesssim \frac{r^{K-1}}{p^{K-1}} \sum_{i\in[p]} \|\boldsymbol{X}_{i:}\|^2 \|\hat{\boldsymbol{X}}_{i:}^s - \hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2 \\
&\lesssim \frac{r^{K-1}}{p^{K-1}} \left(p^{K/2}r + pr^2 + r^K\right), \qquad (48)
\end{aligned}$$

where the first inequality follows from the assumption $\min_{i\in[p]} \theta(i) \geq c > 0$, the second inequality follows from the inequality (38), and the last inequality comes from the inequality (47).

Next, we consider the following quantity,

$$\begin{aligned}
\sum_{i\in[p]} \theta(i) \|\boldsymbol{X}_{i:}^s - \hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2 &\lesssim \sum_{i\in[p]} \theta(i)^2 \|\boldsymbol{X}_{i:}^s - \hat{\boldsymbol{X}}_{i:}^s\|^2 + \sum_{i\in[p]} \theta(i)^2 \|\hat{\boldsymbol{X}}_{i:}^s - \hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2 \\
&\lesssim \sum_{i\in[p]} \frac{\theta(i)^2}{\|\boldsymbol{X}_{i:}\|^2} \|\boldsymbol{X}_{i:} - \hat{\boldsymbol{X}}_{i:}\|^2 + \sum_{i\in[p]} \theta(i)^2 \|\hat{\boldsymbol{X}}_{i:}^s - \hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2 \\
&\lesssim \frac{r^{K-1}}{p^{K-1}} \left(p^{K/2}r + pr^2 + r^K\right), \qquad (49)
\end{aligned}$$

where the first inequality follows from the assumption of $\theta(i)$ and triangle inequality, the second inequality follows from Lemma 4, and the last inequality follows from (48). In addition, with Theorem 4 and the condition $\text{SNR} \gtrsim p^{-K/2}\log p$, for all $a \in [r]$, we have

$$|z^{-1}(a) \cap (z^{(0)})^{-1}(a)| \geq |z^{-1}(a)| - p\ell(z^{(0)}, z) \gtrsim \frac{p}{r} - \frac{p}{\log p} \gtrsim \frac{p}{r},$$

when p is large enough. Therefore, for all $a \in [r]$, we have

$$\begin{aligned}
\|\hat{\boldsymbol{x}}_a - \boldsymbol{X}_a^s\|^2 &= \frac{\sum_{i\in z^{-1}(a)\cap(z^{(0)})^{-1}(a)} \left\|\boldsymbol{X}_{i:}^s - \hat{\boldsymbol{x}}_{z^{(0)}(i)}\right\|^2}{|z^{-1}(a) \cap (z^{(0)})^{-1}(a)|} \\
&\lesssim \frac{r}{p}\left(\sum_{i\in[p]} \|\boldsymbol{X}_{i:}^s - \hat{\boldsymbol{X}}_{i:}^s\|^2 + \sum_{i\in[p]} \|\hat{\boldsymbol{X}}_{i:}^s - \hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2\right) \\
&\lesssim \frac{r^K}{p^K}\left(p^{K/2}r + pr^2 + r^K\right), \qquad (50)
\end{aligned}$$

where the last inequality follows from the inequality (48).

Finally, we obtain

$$L^{(0)} = \frac{1}{p}\sum_{i\in[p]} \theta(i) \sum_{b\in[r]} \mathbb{1}\left\{z^{(0)}(i) = b\right\} \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2$$

$$
\begin{aligned}
&\lesssim \frac{1}{p} \sum_{i\in[p], z^{(0)}(i)\neq z(i)} \theta(i)\|\boldsymbol{X}^s_{i:} - \boldsymbol{X}^s_{z^{(0)}(i)}\|^2 \\
&\lesssim \frac{1}{p} \sum_{i\in[p], z^{(0)}(i)\neq z(i)} \theta(i)\Big(\|\boldsymbol{X}^s_{i:} - \hat{\boldsymbol{x}}_{z^{(0)}(i)}\|^2 + \|\hat{\boldsymbol{x}}_{z^{(0)}(i)} - \boldsymbol{X}^s_{z^{(0)}(i)}\|^2\Big) \\
&\leq \bar{C}\frac{r^K}{p^K}\left(p^{K/2}r + pr^2 + r^K\right), \\
&\leq \frac{\bar{C}\Delta^2_{\min}}{\tilde{C}r\log p}
\end{aligned}
$$

where the first inequality follows from Lemma 1, the third inequality follows from inequalities (49) and (50), and the last inequality follows from the assumption that $\text{SNR} \geq \tilde{C}p^{-K/2}\log p$. $\square$

Proof of Proposition 1. Sub-algorithm 3 shares the same algorithm strategy as Sub-algorithm 1 but with a different estimation of the mean tensor, $\hat{\mathcal{X}}'$. Hence, the proof of Proposition 1 follows the same proof idea with the proof of Theorem 4. Replacing the estimation $\hat{\mathcal{X}}$ by $\hat{\mathcal{X}}'$ in the proof of Theorem 4, we have

$$
\min_{\pi\in\Pi} \sum_{i:z^{(0)}(i)\neq\pi(z(i))} \theta(i)^2 \lesssim \left(\sum_{i\in S}\|\boldsymbol{X}_{i:}\|^2 + \sum_{i\in S_0}\|\boldsymbol{X}_{i:}\|^2\right) p^{-(K-1)}r^{K-1}. \tag{51}
$$

By inequalities (43) and (45), we have

$$
\sum_{i\in S}\|\boldsymbol{X}_{i:}\|^2 \leq \left(\frac{16(1+\eta)}{c_0^2\Delta^2_{\min}} + 2\right)\|\hat{\mathcal{X}}' - \mathcal{X}\|_F^2, \tag{52}
$$

$$
\sum_{i\in S_0}\|\boldsymbol{X}_{i:}\|^2 \leq \|\hat{\mathcal{X}}' - \mathcal{X}\|_F^2. \tag{53}
$$

Hence, it suffices to find the upper bound of the estimation error $\|\hat{\mathcal{X}}' - \mathcal{X}\|_F^2$ to complete our proof. Note that the matricization $\text{Mat}_{sq}(\mathcal{X}) \in \mathbb{R}^{p^{\lfloor K/2\rfloor}\times p^{\lceil K/2\rceil}}$ has $\text{rank}(\text{Mat}_{sq}(\mathcal{X})) \leq r^{\lceil K/2\rceil}$, and Bernoulli random variables follow the sub-Gaussian distribution with bounded variance $\sigma^2 = 1/4$. Apply Lemma 9 to $\boldsymbol{Y} = \text{Mat}_{sq}(\mathcal{Y}), \boldsymbol{X} = \text{Mat}_{sq}(\mathcal{X})$, and $\hat{\boldsymbol{X}} = \text{Mat}_{sq}(\hat{\mathcal{X}}')$. Then, with probability tending to 1 as $p\to\infty$, we have

$$
\|\hat{\mathcal{X}}' - \mathcal{X}\|_F^2 = \|\text{Mat}_{sq}(\hat{\mathcal{X}}') - \text{Mat}_{sq}(\mathcal{X})\|_F^2 \lesssim p^{\lceil K/2\rceil}. \tag{54}
$$

Combining the estimation error (54) with inequalities (52), (53), and (51), we obtain

$$
\min_{\pi\in\Pi} \sum_{i:z^{(0)}(i)\neq\pi(z(i))} \theta(i)^2 \lesssim \frac{\sigma^2 r^{K-1}}{\Delta^2_{\min}p^{K-1}}p^{\lceil K/2\rceil}. \tag{55}
$$

Replace the inequality (46) in the proof of Theorem 4 by inequality (55). With the the same procedures to obtain $\ell(\hat{z}^{(0)}, z)$ and $L(\hat{z}^{(0)}, z)$ for Theorem 4 , we finish the proof of Proposition 1. $\square$

Useful Definitions and Lemmas for the Proof of Theorem 4

Lemma 4 (Basic inequality)**.** For any two nonzero vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$ of same dimension, we have

$$\sin(\boldsymbol{v}_1, \boldsymbol{v}_2) \le \|\boldsymbol{v}_1^s - \boldsymbol{v}_2^s\| \le \frac{2\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|}{\max(\|\boldsymbol{v}_1\|, \|\boldsymbol{v}_2\|)}.$$

Proof of Lemma 4. For the first inequality, let $\alpha \in [0, \pi]$ denote the angle between $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. We have

$$\|\boldsymbol{v}_1^s - \boldsymbol{v}_2^s\| = \sqrt{2(1-\cos\alpha)} = 2\sin\frac{\alpha}{2} \ge \sin\alpha,$$

where the equations follow from the properties of trigonometric function and the inequality follows from the fact the $\cos\frac{\alpha}{2} \le 1$ and $\sin\alpha = 2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2} > 0$ for $\alpha \in [0, \pi]$.

For the second inequality, without loss of generality, we assume $\|\boldsymbol{v}_1\| \ge \|\boldsymbol{v}_2\|$. Then

$$\begin{aligned}
\|\boldsymbol{v}_1^s - \boldsymbol{v}_2^s\| &= \left\| \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} - \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_1\|} + \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_1\|} - \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|} \right\| \\
&\le \frac{\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|}{\|\boldsymbol{v}_1\|} + \frac{\|\boldsymbol{v}_2\| \, \|\boldsymbol{v}_1\| - \|\boldsymbol{v}_2\|}{\|\boldsymbol{v}_1\| \, \|\boldsymbol{v}_2\|} \\
&\le \frac{2\|\boldsymbol{v}_1 - \boldsymbol{v}_2\|}{\|\boldsymbol{v}_2\|}.
\end{aligned}$$

Therefore, Lemma 4 is proved. □

Definition 1 (Weighted padding vectors)**.** For a vector $\boldsymbol{a} = [\![a_i]\!] \in \mathbb{R}^d$, we define the padding vector of $\boldsymbol{a}$ with the weight collection $\boldsymbol{w} = \{\boldsymbol{w}_i \colon \boldsymbol{w}_i = [\![w_{ik}]\!] \in \mathbb{R}^{p_i}\}_{i=1}^d$ as

$$\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}) = [a_1 \circ \boldsymbol{w}_1, \ldots, a_d \circ \boldsymbol{w}_d]^T, \tag{56}$$

where $a_i \circ \boldsymbol{w}_i = [a_i w_{i1}, \ldots, a_i w_{ip_i}]^T$, for all $i \in [d]$. Here we also view $\mathrm{Pad}_{\boldsymbol{w}}(\cdot) \colon \mathbb{R}^d \mapsto \mathbb{R}^{\sum_{i\in[d]} p_i}$ as an operator. We have the bounds of the weighted padding vector

$$\min_{i\in[d]} \|\boldsymbol{w}_i\|^2 \|\boldsymbol{a}\|^2 \le \|\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a})\|^2 \le \max_{i\in[d]} \|\boldsymbol{w}_i\|^2 \|\boldsymbol{a}\|^2. \tag{57}$$

Further, we define the inverse weighted padding operator $\mathrm{Pad}^{-1} : \mathbb{R}^{\sum_{i\in[d]} p_i} \mapsto \mathbb{R}^d$ which satisfies

$$\mathrm{Pad}_{\boldsymbol{w}}^{-1}(\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a})) = \boldsymbol{a}.$$

Lemma 5 (Angle for weighted padding vectors)**.** Suppose that we have two non-zero vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$. Given the weight collection $\boldsymbol{w}$, we have

$$\frac{\min_{i\in[d]} \|\boldsymbol{w}_i\|}{\max_{i\in[d]} \|\boldsymbol{w}_i\|} \sin(\boldsymbol{a}, \boldsymbol{b}) \overset{*}{\le} \sin(\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}), \mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})) \overset{**}{\le} \frac{\max_{i\in[d]} \|\boldsymbol{w}_i\|}{\min_{i\in[d]} \|\boldsymbol{w}_i\|} \sin(\boldsymbol{a}, \boldsymbol{b}). \tag{58}$$

Proof of Lemma 5. We prove the two inequalities separately with similar ideas.

First, we prove the inequality ** in (58). Decomposing $\boldsymbol{b}$ yields

$$\boldsymbol{b} = \cos(\boldsymbol{a}, \boldsymbol{b}) \frac{\|\boldsymbol{b}\|}{\|\boldsymbol{a}\|} \boldsymbol{a} + \sin(\boldsymbol{a}, \boldsymbol{b}) \frac{\|\boldsymbol{b}\|}{\|\boldsymbol{a}^\perp\|} \boldsymbol{a}^\perp,$$

where $\boldsymbol{a}^\perp \in \mathbb{R}^d$ is in the orthogonal complement space of $\boldsymbol{a}$. By the Definition 1, we have

$$\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b}) = \cos(\boldsymbol{a}, \boldsymbol{b}) \frac{\|b\|}{\|a\|} \mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}) + \sin(\boldsymbol{a}, \boldsymbol{b}) \frac{\|b\|}{\|a^\perp\|} \mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}^\perp).$$

Note that $\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}^\perp)$ is not necessary equal to the orthogonal vector of $\mathrm{Pad}(\boldsymbol{a})$; i.e., $\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}^\perp) \neq (\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}))^\perp$. By the geometry property of trigonometric functions, we obtain

$$\sin(\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}), \mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})) \leq \frac{\|\boldsymbol{b}\| \|\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}^\perp)\|}{\|\boldsymbol{a}^\perp\| \|\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})\|} \sin(\boldsymbol{a}, \boldsymbol{b}) \leq \frac{\max_{i \in [d]} \|\boldsymbol{w}_i\|}{\min_{i \in [d]} \|\boldsymbol{w}_i\|} \sin(\boldsymbol{a}, \boldsymbol{b}),$$

where the second inequality follows by applying the property (57) to vectors $\boldsymbol{b}$ and $\boldsymbol{a}^\perp$.

Next, we prove inequality * in (58). With the decomposition of $\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})$ and the inverse weighted padding operator, we have

$$\boldsymbol{b} = \cos(\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}), \mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})) \frac{\|\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})\|}{\|\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a})\|} \boldsymbol{a} + \sin(\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}), \mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})) \frac{\|\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})\|}{\|(\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}))^\perp\|} \mathrm{Pad}_{\boldsymbol{w}}^{-1}((\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}))^\perp).$$

Therefore, we obtain

$$\begin{aligned}
\sin(\boldsymbol{a}, \boldsymbol{b}) &\leq \frac{\|\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})\| \|\mathrm{Pad}_{\boldsymbol{w}}^{-1}((\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}))^\perp)\|}{\|(\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}))^\perp\| \|\boldsymbol{b}\|} \sin(\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}), \mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})) \\
&\leq \frac{\max_{i \in [d]} \|\boldsymbol{w}_i\|}{\min_{i \in [d]} \|\boldsymbol{w}_i\|} \sin(\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}), \mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{b})),
\end{aligned}$$

where the second inequality follows by applying the property (57) to vectors $\boldsymbol{b}$ and $\mathrm{Pad}_{\boldsymbol{w}}^{-1}((\mathrm{Pad}_{\boldsymbol{w}}(\boldsymbol{a}))^\perp)$. □

Lemma 6 (Singular value of weighted membership matrix)**.** Under the parameter space (2) and assumption that $\min_{i \in [p]} \theta(i) \geq c$ for some constant $c > 0$, the singular values of $\boldsymbol{\Theta M}$ are bounded as

$$\sqrt{p/r} \lesssim \sqrt{\min_{a \in [r]} \|\boldsymbol{\theta}_{z^{-1}(a)}\|^2} \leq \lambda_r(\boldsymbol{\Theta M}) \leq \|\boldsymbol{\Theta M}\|_\sigma \leq \sqrt{\max_{a \in [r]} \|\boldsymbol{\theta}_{z^{-1}(a)}\|^2} \lesssim p/r.$$

Proof of Lemma 6. Note that

$$(\boldsymbol{\Theta M})^T \boldsymbol{\Theta M} = \boldsymbol{D},$$

with $\boldsymbol{D} = \mathrm{diag}(D_1, \ldots, D_r)$ where $D_a = \|\boldsymbol{\theta}_{z^{-1}(a)}\|^2, a \in [r]$. By the definition of singular values, we have

$$\sqrt{\min_{a \in [r]} \|\boldsymbol{\theta}_{z^{-1}(a)}\|^2} \leq \lambda_r(\boldsymbol{\Theta M}) \leq \|\boldsymbol{\Theta M}\|_\sigma \leq \sqrt{\max_{a \in [r]} \|\boldsymbol{\theta}_{z^{-1}(a)}\|^2}.$$

Since that $\min_{i\in[p]}\theta(i) \geq c$ by the constraints in parameter space, we have

$$\min_{a\in[r]}\|\boldsymbol{\theta}_{z^{-1}(a)}\|^2 \geq c^2 \min_{a\in[r]}|z^{-1}(a)| \gtrsim \frac{p}{r},$$

where the last inequality follows from the constraint in parameter space (2). Finally, notice that

$$\sqrt{\max_{a\in[r]}\|\boldsymbol{\theta}_{z^{-1}(a)}\|^2} \leq \max_{a\in[r]}\sqrt{\|\boldsymbol{\theta}_{z^{-1}(a)}\|_1^2} \lesssim \frac{p}{r}.$$

Therefore, we complete the proof of Lemma 6. $\square$

Lemma 7 (Singular-value gap-free tensor estimation error bound)**.** Consider an order-K tensor $\mathcal{A} = \mathcal{X} + \mathcal{Z} \in \mathbb{R}^{p\times\cdots\times p}$, where $\mathcal{X}$ has Tucker rank $(r, ...r)$ and $\mathcal{Z}$ has independent sub-Gaussian entries with parameter σ^2. Let $\hat{\mathcal{X}}$ denote the double projection estimated tensor in Step 2 of Sub-algorithm 1 in the main paper. Then with probability at least $1 - C\exp(-cp)$, we have

$$\|\hat{\mathcal{X}} - \mathcal{X}\|_F^2 \leq C\sigma^2\left(p^{K/2}r + pr^2 + r^K\right),$$

where C, c are some positive constants.

Proof of Lemma 7. See Han et al. (2022a, Proposition 1). $\square$

Lemma 8 (Upper bound of misclustering error)**.** Let $z : [p] \mapsto [r]$ be a cluster assignment such that $|z^{-1}(a)| \asymp p/r$ for all $a \in [r]$ with $r \geq 2$. Let node i correspond to a vector $\boldsymbol{x}_i = \theta(i)\boldsymbol{v}_{z(i)} \in \mathbb{R}^d$, where $\{\boldsymbol{v}_a\}_{a=1}^r$ are the cluster centers and $\boldsymbol{\theta} = [\![\theta(i)]\!] \in \mathbb{R}_+^p$ is the positive degree heterogeneity. Assume that $\boldsymbol{\theta}$ satisfies the balanced assumption (6) such that $\frac{\max_{a\in[r]}\|\boldsymbol{\theta}_{z^{-1}(a)}\|^2}{\min_{a\in[r]}\|\boldsymbol{\theta}_{z^{-1}(a)}\|^2} = 1 + o(1)$. Consider an arbitrary estimate $\hat{z}$ with $\hat{\boldsymbol{x}}_i = \hat{\boldsymbol{v}}_{\hat{z}(i)}$ for all $i \in S$. Then, if

$$\min_{a\neq b\in[r]}\|\boldsymbol{v}_a - \boldsymbol{v}_b\| \geq 2c, \tag{59}$$

for some constant $c > 0$, we have

$$\min_{\pi\in\Pi}\sum_{i:\hat{z}(i)\neq\pi(z(i))}\theta(i)^2 \leq \sum_{i\in S_0}\theta(i)^2 + 4\sum_{i\in S}\theta(i)^2,$$

where S_0 is defined in Step 4 of Sub-algorithm 1 and

$$S = \{i \in S_0^c : \|\hat{\boldsymbol{x}}_i - \boldsymbol{v}_{z(i)}\| \geq c\}.$$

Proof of Lemma 8. For each cluster $u \in [r]$, we use C_u to collect the subset of points for which the estimated and true positions $\hat{\boldsymbol{x}}_i, \boldsymbol{x}_i$ are within distance c. Specifically, define

$$C_u = \{i \in z^{-1}(u) \cap S_0^c : \|\hat{\boldsymbol{x}}_i - \boldsymbol{v}_{z(i)}\| < c\},$$

and divide $[r]$ into three groups based on C_u as

$$R_1 = \{u \in [r] : C_u = \emptyset\},$$

$$R_2 = \{u \in [r] : C_u \neq \emptyset, \text{ for all } i, j \in C_u, \hat{z}(i) = \hat{z}(j)\},$$
$$R_3 = \{u \in [r] : C_u \neq \emptyset, \text{ there exist } i, j \in C_u, \hat{z}(i) \neq \hat{z}(j)\}.$$

Note that $\cup_{u\in[r]} C_u = S_0^c / S^c$ and $C_u \cap C_v = \emptyset$ for any $u \neq v$. Suppose there exist $i \in C_u$ and $j \in C_v$ with $u \neq v \in [r]$ and $\hat{z}(i) = \hat{z}(j)$. Then we have

$$\|\boldsymbol{v}_{z(i)} - \boldsymbol{v}_{z(j)}\| \leq \|\boldsymbol{v}_{z(i)} - \hat{\boldsymbol{x}}_i\| + \|\boldsymbol{v}_{z(j)} - \hat{\boldsymbol{x}}_j\| < 2c,$$

which contradicts to the assumption (59). Hence, the estimates $\hat{z}(i) \neq \hat{z}(j)$ for the nodes $i \in C_u$ and $j \in C_v$ with $u \neq v$. By the definition of R_2, the nodes in $\cup_{u\in R_2} C_u$ have the same assignment with z and $\hat{z}$. Then, we have

$$\min_{\pi\in\Pi} \sum_{i:\hat{z}(i)\neq\pi(z(i))} \theta(i)^2 \leq \sum_{i\in S_0} \theta(i)^2 + \sum_{i\in S} \theta(i)^2 + \sum_{i\in\cup_{u\in R_3} C_u} \theta(i)^2.$$

We only need to bound $\sum_{i\in\cup_{u\in R_3} C_u} \theta(i)^2$ to finish the proof. Note that every C_u with $u \in R_3$ contains at least two nodes assigned to different clusters by $\hat{z}$. Then, we have $|R_2| + 2|R_3| \leq r$. Since $|R_1| + |R_2| + |R_3| = r$, we have $|R_3| \leq |R_1|$. Hence, we obtain

$$\begin{aligned}
\sum_{i\in\cup_{u\in R_3} C_u} \theta(i)^2 &\leq |R_3| \max_{a\in[r]} \|\boldsymbol{\theta}_{z^{-1}(a)}\|^2 \\
&\leq |R_1| \max_{a\in[r]} \|\boldsymbol{\theta}_{z^{-1}(a)}\|^2 \\
&\leq \frac{\max_{a\in[r]} \|\boldsymbol{\theta}_{z^{-1}(a)}\|^2}{\min_{a\in[r]} \|\boldsymbol{\theta}_{z^{-1}(a)}\|^2} \sum_{i\in\cup_{u\in R_1} z^{-1}(u)} \theta(i)^2 \\
&\leq 2 \sum_{i\in S} \theta(i)^2,
\end{aligned}$$

where the last inequality holds by the balanced assumption on $\boldsymbol{\theta}$ when p is large enough, and the fact that $\cup_{u\in R_1} z^{-1}(u) \subset S$. □

Lemma 9 (Low-rank matrix estimation)**.** Let $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{E} \in \mathbb{R}^{m\times n}$, where $n > m$ and $\boldsymbol{E}$ contains independent mean-zero sub-Gaussian entries with bounded variance σ^2. Suppose $\text{rank}(\boldsymbol{X}) = r$. Consider the least square estimator

$$\hat{\boldsymbol{X}} = \underset{\boldsymbol{X}'\in\mathbb{R}^{m\times n}, \text{rank}(\boldsymbol{X}')\leq r}{\arg\min} \|\boldsymbol{X}' - \boldsymbol{Y}\|_F^2.$$

There exist positive constants C_1, C_2 such that

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_F^2 \leq C_1 \sigma^2 n r,$$

with probability at least $1 - \exp(-C_2 n r)$.

Proof of Lemma 9. Note that $\|\hat{\boldsymbol{X}} - \boldsymbol{Y}\|_F^2 \leq \|\boldsymbol{X} - \boldsymbol{Y}\|_F^2$ by the definition of least square estimator. We have

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_F^2 \leq 2 \left\langle \hat{\boldsymbol{X}} - \boldsymbol{X}, \boldsymbol{Y} - \boldsymbol{X} \right\rangle$$

$$\leq 2\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_F \sup_{\boldsymbol{T}\in\mathbb{R}^{m\times n}, \text{rank}(\boldsymbol{T})\leq 2r, \|\boldsymbol{T}\|_F=1} \langle \boldsymbol{T}, \boldsymbol{Y} - \boldsymbol{X}\rangle \tag{60}$$

with probability at least $1 - \exp(-C_2 nr)$, where the second inequality follows by re-arrangement.

Consider the SVD for matrix $\boldsymbol{T} = \boldsymbol{U}\Sigma\boldsymbol{V}^T$ with orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{m\times 2r}, \boldsymbol{V} \in \mathbb{R}^{n\times 2r}$ and diagonal matrix $\Sigma \in \mathbb{R}^{2r\times 2r}$. We have

$$\begin{aligned}\sup_{\boldsymbol{T}\in\mathbb{R}^{m\times n}, \text{rank}(\boldsymbol{T})\leq 2r, \|\boldsymbol{T}\|_F=1} \langle \boldsymbol{T}, \boldsymbol{Y} - \boldsymbol{X}\rangle &= \sup_{\boldsymbol{T}\in\mathbb{R}^{m\times n}, \text{rank}(\boldsymbol{T})\leq 2r, \|\boldsymbol{T}\|_F=1} \langle \boldsymbol{U}\Sigma, \boldsymbol{E}\boldsymbol{V}\rangle \\ &= \sup_{\boldsymbol{v}\in\mathbb{R}^{2nr}} \boldsymbol{v}^T\boldsymbol{e} \leq C\sigma\sqrt{nr}, \end{aligned} \tag{61}$$

with probability $1 - \exp(-C_2 nr)$, where C, C_2 are two positive constants, the vectorization $\boldsymbol{e} = \text{Vec}(\boldsymbol{E}\boldsymbol{V}) \in \mathbb{R}^{2nr}$ has independent mean-zero sub-Gaussian entries with bounded variance σ^2 due to the orthogonality of $\boldsymbol{V}$, and the last inequality follows from Rigollet and Hütter (2015, Theorem 1.19).

Combining inequalities (60) and (61), we obtain the desired conclusion.

□

B.7 Proofs of Theorem 2 (Achievability) and Theorem 5

Proof of Theorem 2 (Achievability) and Theorem 5. The proofs of Theorem 2 (Achievability) and Theorem 5 share the same idea. We prove the contraction step by step. In each step, we show the specific procedures for the algorithm loss and address the MLE loss by stating the difference.

We consider dTBM (1) with symmetric mean tensor, parameters $(z, \mathcal{S}, \boldsymbol{\theta})$, fixed $r \geq 1, K \geq 2$, and i.i.d. noise. Let $(\hat{z}, \hat{\mathcal{S}}, \hat{\boldsymbol{\theta}})$ denote the MLE in (9), and $(z_k^{(0)}, \mathcal{S}^{(0)}, \boldsymbol{\theta}_k^{(0)})$ denote parameters related to the initialization. For the case $r = 1$, $\ell(z_k^{(t)}, z) = 0$ trivially for all $t \geq 0, k \in [k]$. Hence, we focus on the proof of the first mode clustering $z_1^{(t+1)}$ with $r \geq 2$; the extension for other modes can be obtained similarly. We drop the subscript k in the matricizations $\boldsymbol{\Theta}, \boldsymbol{M}_k, \boldsymbol{S}_k, \boldsymbol{X}_k$ and in estimates $z_k^{(0)}, z_k^{(t+1)}, z_k^{(t)}$ for ease of the notation. Without loss of generality, we assume that the variance $\sigma = 1$, and that the identity permutation minimizes the initial misclustering error; i.e., $\pi^{(0)} = \arg\min_{\pi\in\Pi} \sum_{i\in[p]} \mathbb{1}\left\{z^{(0)}(i) \neq \pi \circ z(i)\right\}$ and $\pi^{(0)}(a) = a$ for all $a \in [r]$, and so for $\hat{z}$.

Step 1 (Notation and conditions). We first introduce additional notations and the necessary conditions used in the proof. We will verify that the conditions hold in our context under high probability in the last step of the proof.

Notation.

1. Projection. We use $\boldsymbol{I}_d$ to denote the identity matrix of dimension d. For a vector $\boldsymbol{v} \in \mathbb{R}^d$, let $\text{Proj}(\boldsymbol{v}) \in \mathbb{R}^{d\times d}$ denote the projection matrix to $\boldsymbol{v}$. Then, $\boldsymbol{I}_d - \text{Proj}(\boldsymbol{v})$ is the projection matrix to the orthogonal complement $\boldsymbol{v}^\perp$.

2. We define normalized membership matrices

$$\boldsymbol{W} = \boldsymbol{M}\left(\text{diag}(\boldsymbol{1}_p^T\boldsymbol{M})\right)^{-1}, \boldsymbol{W}^{(t)} = \boldsymbol{M}^{(t)}\left(\text{diag}(\boldsymbol{1}_p^T\boldsymbol{M}^{(t)})\right)^{-1},$$

weighted normalized membership matrices

$$\boldsymbol{P} = \boldsymbol{\Theta}\boldsymbol{M}(\mathrm{diag}(\|\boldsymbol{\theta}_{z^{-1}(1)}\|^2, \ldots, \|\boldsymbol{\theta}_{z^{-1}(r)}\|^2))^{-1}, \ \hat{\boldsymbol{P}} = \hat{\boldsymbol{\Theta}}\hat{\boldsymbol{M}}(\mathrm{diag}(\|\hat{\boldsymbol{\theta}}_{z^{-1}(1)}\|^2, \ldots, \|\hat{\boldsymbol{\theta}}_{z^{-1}(r)}\|^2))^{-1},$$

and the dual normalized and dual weighted normalized membership matrices

$$\boldsymbol{V} = \boldsymbol{W}^{\otimes(K-1)}, \quad \boldsymbol{V}^{(t)} = \left(\boldsymbol{W}^{(t)}\right)^{\otimes(K-1)}, \quad \boldsymbol{Q} = \boldsymbol{P}^{\otimes K-1}, \quad \hat{\boldsymbol{Q}} = \hat{\boldsymbol{P}}^{\otimes K-1}.$$

Also, let $\boldsymbol{B} = (\boldsymbol{\Theta}\boldsymbol{M})^{\otimes(K-1)}, \hat{\boldsymbol{B}} = (\hat{\boldsymbol{\Theta}}\hat{\boldsymbol{M}})^{\otimes(K-1)}$. By the definition, we have $\boldsymbol{B}^T\boldsymbol{Q} = \hat{\boldsymbol{B}}^T\hat{\boldsymbol{Q}} = \boldsymbol{I}_{r^{K-1}}$.

3. We use $\mathcal{S}^{(t)}$ to denote the estimator of $\mathcal{S}$ in the t-th iteration, $\hat{\mathcal{S}}$ for MLE, $\tilde{\mathcal{S}}$ to denote the oracle estimator of $\mathcal{S}$ given true assignment z, and $\bar{\mathcal{S}}$ for weighted oracle estimator; i.e.,

$$\mathcal{S}^{(t)} = \mathcal{Y} \times_1 \left(\boldsymbol{W}^{(t)}\right)^T \times_2 \cdots \times_K \left(\boldsymbol{W}^{(t)}\right)^T, \qquad \tilde{\mathcal{S}} = \mathcal{Y} \times_1 \boldsymbol{W}^T \times_2 \cdots \times_K \boldsymbol{W}^T,$$
$$\hat{\mathcal{S}} = \mathcal{Y} \times_1 \hat{\boldsymbol{P}}^T \times_2 \cdots \times_K \hat{\boldsymbol{P}}^T, \qquad \bar{\mathcal{S}} = \mathcal{Y} \times_1 \boldsymbol{P}^T \times_2 \cdots \times_K \boldsymbol{P}^T.$$

4. We define the matricizations of tensors

$$\boldsymbol{S} = \mathrm{Mat}(\mathcal{S}), \ \boldsymbol{Y} = \mathrm{Mat}(\mathcal{Y}), \ \boldsymbol{X} = \mathrm{Mat}(\mathcal{X}), \ \boldsymbol{E} = \mathrm{Mat}(\mathcal{E}),$$

$$\boldsymbol{S}^{(t)} = \mathrm{Mat}(\mathcal{S}^{(t)}), \ \hat{\boldsymbol{S}} = \mathrm{Mat}(\hat{\mathcal{S}}), \ \tilde{\boldsymbol{S}} = \mathrm{Mat}(\tilde{\mathcal{S}}), \ \bar{\boldsymbol{S}} = \mathrm{Mat}(\bar{\mathcal{S}}).$$

5. We define the extended core tensor on $K-1$ modes

$$\boldsymbol{A} = \boldsymbol{S}\boldsymbol{B}^T, \quad \bar{\boldsymbol{A}} = \bar{\boldsymbol{S}}\boldsymbol{B}^T, \quad \hat{\boldsymbol{A}} = \hat{\boldsymbol{S}}\hat{\boldsymbol{B}}^T.$$

By the assumption in parameter space (2), we have $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{X} = \boldsymbol{W}\boldsymbol{X}, \quad \hat{\boldsymbol{A}} = \hat{\boldsymbol{P}}\hat{\boldsymbol{X}} = \hat{\boldsymbol{W}}\hat{\boldsymbol{X}}$.

6. We define the angle-based misclustering loss in the t-th iteration and loss for MLE

$$L^{(t)} = \frac{1}{p}\sum_{i\in[p]}\theta(i)\sum_{b\in[r]}\mathbb{1}\{z^{(t)}(i) = b\}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2,$$
$$L(\hat{z}) = \frac{1}{p}\sum_{i\in[p]}\theta(i)^2\sum_{b\in[r]}\mathbb{1}\{\hat{z}(i) = b\}\|[\boldsymbol{A}_{z(i):}]^s - [\boldsymbol{A}_{b:}]^s\|^2.$$

We also define the loss for oracle and weighted oracle estimators

$$\xi = \frac{1}{p}\sum_{i\in[p]}\theta(i)\sum_{b\in[r]}\mathbb{1}\left\{\left\langle \boldsymbol{E}_{i:}\boldsymbol{V}, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s\right\rangle \le -\frac{\theta(i)m}{4}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2\right\}\cdot\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2,$$
$$\xi' = \frac{1}{p}\sum_{i\in[p]}\theta(i)^2\sum_{b\in[r]}\mathbb{1}\left\{\langle \boldsymbol{E}_{i:}, [\bar{\boldsymbol{A}}_{z(i):}]^s - [\bar{\boldsymbol{A}}_{b:}]^s\rangle \le -\frac{m'}{4}\sqrt{\frac{p^{K-1}}{r^{K-1}}}\|[\boldsymbol{A}_{z(i):}]^s - [\boldsymbol{A}_{b:}]^s\|_F^2\right\}\cdot\|[\boldsymbol{A}_{z(i):}]^s - [\boldsymbol{A}_{b:}]^s\|^2.$$

where m and m' are some positive universal constants.

Then we introduce the necessary conditions in Condition 1.

Condition 1. (Intermediate results) Let $\mathbb{O}_{p,r}$ denote the collection of all the p-by-r matrices with orthonormal columns. We have

$$\|\boldsymbol{E}\boldsymbol{V}\|_\sigma \lesssim \sqrt{\frac{r^{K-1}}{p^{K-1}}}\left(p^{1/2}+r^{(K-1)/2}\right),\ \|\boldsymbol{E}\boldsymbol{V}\|_F \lesssim \sqrt{\frac{r^{2(K-1)}}{p^{K-2}}},\ \|\boldsymbol{W}_{a:}^T\boldsymbol{E}\boldsymbol{V}\| \lesssim \frac{r^K}{p^{K/2}},\ a\in[r], \tag{62}$$

$$\sup_{\boldsymbol{U}_k\in\mathbb{O}_{p,r},k=2,\ldots,K}\|\boldsymbol{E}(\boldsymbol{U}_2\otimes\cdots\otimes\boldsymbol{U}_K)\|_\sigma \lesssim \left(\sqrt{r^{K-1}}+K\sqrt{pr}\right), \tag{63}$$

$$\sup_{\boldsymbol{U}_k\in\mathbb{O}_{p,r},k=2,\ldots,K}\|\boldsymbol{E}(\boldsymbol{U}_2\otimes\cdots\otimes\boldsymbol{U}_K)\|_F \lesssim \left(\sqrt{pr^{K-1}}+K\sqrt{pr}\right), \tag{64}$$

$$\xi \leq \exp\left(-M\frac{\Delta_{\min}^2 p^{K-1}}{r^{K-1}}\right),\quad \xi' \lesssim \exp\left(-\frac{\Delta_{\min}^2 p^{K-1}}{r^{K-1}}\right), \tag{65}$$

$$L^{(t)} \leq \frac{\bar{C}}{\tilde{C}}\frac{\Delta_{\min}^2}{r\log p},\quad \text{for}\quad t=0,1,\ldots,T,\quad L(\hat{z}) \leq \frac{\bar{C}}{\tilde{C}}\frac{\Delta_{\min}^2}{r\log p}, \tag{66}$$

where M is a positive universal constant in inequality (81), $\bar{C},\tilde{C}$ are positive universal constants in the proof of Theorem 4 and assumption SNR $\geq \tilde{C}p^{-K/2}\log p$, respectively. Further, inequality (62) holds by replacing $\boldsymbol{V}$ to $\boldsymbol{V}^{(t)},\boldsymbol{Q},\hat{\boldsymbol{Q}}$ and $\boldsymbol{W}_{:a}$ to $\boldsymbol{W}_{:a}^{(t),T},\boldsymbol{P}_{:a}^T,\hat{\boldsymbol{P}}_{:a}^T$ when initialization condition (66) holds.

Step 2 (Misclustering loss decomposition). Next, we derive the upper bound of $L^{(t+1)}$ for $t=0,1,\ldots,T-1$. By Sub-algorithm 2, we update the assignment in t-th iteration via

$$z^{(t+1)}(i) = \underset{a\in[r]}{\arg\min}\|[\boldsymbol{Y}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{a:}^{(t)}]^s\|^2,$$

following the facts that $\|\boldsymbol{a}^s-\boldsymbol{b}^s\|^2 = 1-\cos(\boldsymbol{a},\boldsymbol{b})$ for vectors $\boldsymbol{a},\boldsymbol{b}$ of same dimension and $\mathrm{Mat}(\mathcal{Y}^{\mathrm{d}}) = \boldsymbol{Y}\boldsymbol{V}^{(t)}$ where $\mathcal{Y}^{\mathrm{d}}$ is the reduced tensor defined in Step 8 of Sub-algorithm 2. Then the event $z^{(t+1)}(i)=b$ implies

$$\|[\boldsymbol{Y}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{b:}^{(t)}]^s\|^2 \leq \|[\boldsymbol{Y}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{z(i):}^{(t)}]^s\|^2. \tag{67}$$

Note that the event (67) also holds for the degenerate entity i with $\|\boldsymbol{Y}_{i:}\boldsymbol{V}^{(t)}\|=0$ due to the convention that $\boldsymbol{a}^s=\boldsymbol{0}$ if $\boldsymbol{a}=\boldsymbol{0}$. Arranging the terms in (67) yields the decomposition

$$2\left\langle \boldsymbol{E}_{i:}\boldsymbol{V}, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s\right\rangle \leq \|\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}\|\left(-\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 + G_{ib}^{(t)} + H_{ib}^{(t)}\right) + F_{ib}^{(t)},$$

where

$$\begin{aligned}
F_{ib}^{(t)} &= 2\left\langle \boldsymbol{E}_{i:}\boldsymbol{V}^{(t)}, \left([\tilde{\boldsymbol{S}}_{z(i):}]^s - [\boldsymbol{S}_{z(i):}^{(t)}]^s\right) - \left([\tilde{\boldsymbol{S}}_{b:}]^s - [\boldsymbol{S}_{b:}^{(t)}]^s\right)\right\rangle + 2\left\langle \boldsymbol{E}_{i:}\left(\boldsymbol{V}-\boldsymbol{V}^{(t)}\right), [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s\right\rangle,\\
G_{ib}^{(t)} &= \left(\|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{z(i):}^{(t)}]^s\|^2 - \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:z(i)}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2\right)\\
&\qquad - \left(\|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{b:}^{(t)}]^s\|^2 - \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2\right),\\
H_{ib}^{(t)} &= \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:z(i)}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2 - \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2 + \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2.
\end{aligned}$$

Therefore, the event $\mathbb{1}\left\{z^{(t+1)}(i)=b\right\}$ can be upper bounded as

$$\mathbb{1}\left\{z^{(t+1)}(i)=b\right\} \leq \mathbb{1}\left\{z^{(t+1)}(i)=b, \left\langle \boldsymbol{E}_{j:}\boldsymbol{V}, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s \right\rangle \leq -\frac{1}{4}\|\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}\|\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \right\}$$

$$+ \mathbb{1}\left\{z^{(t+1)}(i)=b, \frac{1}{2}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \leq \|\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}\|^{-1}F_{ib}^{(t)} + G_{ib}^{(t)} + H_{ib}^{(t)} \right\}. \quad (68)$$

Note that

$$\begin{aligned}
\|\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}\| &= \theta(i)\|\boldsymbol{S}_{i:}(\boldsymbol{\Theta}\boldsymbol{M})^{\otimes(K-1),T}\boldsymbol{W}^{(t),\otimes^{K-1}}\| \\
&\geq \theta(i)\|\boldsymbol{S}_{z(i):}\|\lambda_r^{K-1}(\boldsymbol{\Theta}\boldsymbol{M})\lambda_r^{K-1}(\boldsymbol{W}^{(t)}) \\
&\geq \theta(i)m, \qquad (69)
\end{aligned}$$

where the first inequality follows from the property of eigenvalues; the last inequality follows from Lemma 6, Lemma 10, and assumption that $\min_{a\in[r]}\|\boldsymbol{S}_{z(i):}\| \geq c_3 > 0$; and $m > 0$ is a positive constant related to c_3. Plugging the lower bound of $\|\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}\|$ (69) into the inequality (68) gives

$$\mathbb{1}\left\{z^{(t+1)}(i)=b\right\} \leq A_{ib} + B_{ib}, \qquad (70)$$

where

$$\begin{aligned}
A_{ib} &= \mathbb{1}\left\{z^{(t+1)}(i)=b, \left\langle \boldsymbol{E}_{i:}\boldsymbol{V}, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s \right\rangle \leq -\frac{\theta(i)m}{4}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \right\}, \\
B_{ib} &= \mathbb{1}\left\{z^{(t+1)}(i)=b, \frac{1}{2}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \leq (\theta(i)m)^{-1}F_{ib}^{(t)} + G_{ib}^{(t)} + H_{ib}^{(t)} \right\}.
\end{aligned}$$

Taking the weighted summation of (70) over $i \in [p]$ yields

$$L^{(t+1)} \leq \xi + \frac{1}{p}\sum_{i\in[p]}\sum_{b\in[r]/z(i)}\zeta_{ib}^{(t)},$$

where ξ is the oracle loss such that

$$\xi = \frac{1}{p}\sum_{i\in[p]}\theta(i)\sum_{b\in[r]/z(i)}A_{ib}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2. \qquad (71)$$

Similarly to ξ in (71), we define

$$\zeta_{ib}^{(t)} = \theta(i)B_{ib}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2.$$

Now, we show the decomposition for MLE loss.

By the definition of Gaussian MLE, the estimator $\hat{\boldsymbol{\theta}}$ satisfies $\hat{\theta}(i) = \left\langle \boldsymbol{Y}_{i:}, \hat{\boldsymbol{A}}_{\hat{z}(i):} \right\rangle / \|\hat{\boldsymbol{A}}_{\hat{z}(i):}\|_F^2$ for all $i \in [p]$. Hence, we have

$$\hat{z}(i) = \arg\min_{a\in[r_1]}\|[\boldsymbol{Y}_{i:}]^s - [\hat{\boldsymbol{A}}_{a:}]^s\|_F^2,$$

and the decomposition

$$L(\hat{z}) \leq \xi' + \frac{1}{p}\sum_{i\in[p]}\sum_{b\in[r]/z(i)} \zeta'_{ib},$$

where $\zeta'_{ib} = \theta(i)^2 B'_{ib} \|[\boldsymbol{A}_{z(i):}]^s - [\boldsymbol{A}_{b:}]^s\|^2$ and

$$A'_{ib} = \mathbb{1}\left\{\hat{z}(i) = b, \left\langle \boldsymbol{E}_{i:}, [\bar{\boldsymbol{A}}_{z(i):}]^s - [\bar{\boldsymbol{A}}_{b:}]^s \right\rangle \leq -\frac{m'}{4}\sqrt{\frac{p^{K-1}}{r^{K-1}}}\|[\boldsymbol{A}_{z(i):}]^s - [\boldsymbol{A}_{b:}]^s\|_F^2\right\},$$
$$B'_{ib} = \mathbb{1}\left\{\hat{z}(i) = b, -\frac{1}{2}\|[\boldsymbol{A}_{z(i):}]^s - [\boldsymbol{A}_{b:}]^s\|_F^2 \leq \sqrt{\frac{r^{K-1}}{(m')^2 p^{K-1}}}\hat{F}_{ib} + \hat{G}_{ib} + \hat{H}_{ib}\right\}$$

with terms

$$\hat{F}_{ib} = 2\left\langle \boldsymbol{E}_{i:}, ([\bar{\boldsymbol{A}}_{z(i):}]^s - [\hat{\boldsymbol{A}}_{a:}]^s) - ([\bar{\boldsymbol{A}}_{b:}]^s - [\hat{\boldsymbol{A}}_{b:}]^s)\right\rangle,$$
$$\hat{G}_{ib} = \left(\|\boldsymbol{X}_{i:}^s - [\hat{\boldsymbol{A}}_{z(i):}]^s\|_F^2 - \|\boldsymbol{X}_{i:}^s - [\boldsymbol{P}_{:z(i)}^T \boldsymbol{Y}\hat{\boldsymbol{Q}}\hat{\boldsymbol{B}}^T]^s\|_F^2\right) - \left(\|\boldsymbol{X}_{i:}^s - [\hat{\boldsymbol{A}}_{b:}]^s\|_F^2 - \|\boldsymbol{X}_{i:}^s - [\boldsymbol{P}_{:b}^T \boldsymbol{Y}\hat{\boldsymbol{Q}}\hat{\boldsymbol{B}}^T]^s\|_F^2\right),$$

$$\hat{H}_{ib} = \|\boldsymbol{X}_{i:}^s - [\boldsymbol{P}_{:z(i)}^T \boldsymbol{Y}\hat{\boldsymbol{Q}}\hat{\boldsymbol{B}}^T]^s\|_F^2 - \|\boldsymbol{X}_{i:}^s - [\boldsymbol{P}_{:b}^T \boldsymbol{Y}\hat{\boldsymbol{Q}}\hat{\boldsymbol{B}}^T]^s\|_F^2 + \|\boldsymbol{A}_{z(i):}^s - \boldsymbol{A}_{b:}^s\|_F^2.$$

Step 3 (Derivation of contraction inequality). In this step we derive the upper bound of ζ_{ib} and obtain the contraction inequality (24).

Choose the constant $\tilde{C}$ in the condition $\text{SNR} \geq \tilde{C}p^{-K/2}\log p$ that satisfies the condition of Lemma 11, inequalities (95), and (99). Note that

$$\zeta_{ib}^{(t)} = \theta(i)\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \mathbb{1}\left\{z^{(t+1)}(i) = b, \frac{1}{2}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \leq (\theta(i)m)^{-1}F_{ib}^{(t)} + G_{ib}^{(t)} + H_{ib}^{(t)}\right\}$$

$$\leq \theta(i)\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \mathbb{1}\left\{z^{(t+1)}(i) = b, \frac{1}{4}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \leq (\theta(i)m)^{-1}F_{ib}^{(t)} + G_{ib}^{(t)}\right\}$$
$$\leq 64\mathbb{1}\left\{z^{(t+1)}(i) = b\right\}\left(\frac{(F_{ib}^{(t)})^2}{cm^2\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2} + \frac{\theta(i)(G_{ib}^{(t)})^2}{\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2}\right)$$

where the first inequality follows from the inequality (86) in Lemma 11, and the last inequality follows from the assumption that $\min_{i\in[p]}\theta(i) \geq c > 0$. Following Han et al. (2022a, Step 4, Proof of Theorem 2) and Lemma 11, we have

$$\frac{1}{p}\sum_{i\in[p]}\sum_{b\in[r]/z(i)} \mathbb{1}\left\{z^{(t+1)}(i) = b\right\}\frac{(F_{ib}^{(t)})^2}{cm^2\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2} \leq \frac{C_0\bar{C}}{cm^2\tilde{C}^2}L^{(t)},$$

for a positive universal constant C and

$$\frac{1}{p}\sum_{i\in[p]}\sum_{b\in[r]/z(i)} \mathbb{1}\left\{z^{(t+1)}(i) = b\right\}\frac{\theta(i)(G_{ib}^{(t)})^2}{\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2} \leq \frac{1}{512}\frac{1}{p}\sum_{i\in[p]}\theta(i)\sum_{b\in[r]/z(i)} \mathbb{1}\left\{z^{(t+1)}(i) = b\right\}(\Delta_{\min}^2 + L^{(t)})$$

$$\leq \frac{1}{512}(L^{(t+1)} + L^{(t)}),$$

where the last inequality follows from the definition of $L^{(t)}$ and the constraint of $\boldsymbol{\theta}$ in parameter space (2). For $\tilde{C}$ also satisfies

$$\frac{C_0 \bar{C}}{cm^2 \tilde{C}^2} \leq \frac{1}{512}, \tag{72}$$

we have

$$\frac{1}{p} \sum_{i \in [p]} \sum_{b \in [r]/z(i)} \zeta_{ib}^{(t)} \leq \frac{1}{8} L^{(t+1)} + \frac{1}{4} L^{(t)}. \tag{73}$$

Plugging the inequality (73) into the decomposition (71), we obtain the contraction inequality

$$L^{(t+1)} \leq \frac{3}{2}\xi + \frac{1}{2} L^{(t)}, \tag{74}$$

where $\frac{1}{2}$ is the contraction parameter.

Therefore, with $\tilde{C}$ satisfying inequalities (72), (95) and (99), we obtain the conclusion in Theorem 5 via inequality (74) combining the inequality (65) in Condition 1 and Lemma 2.

We also have the contraction inequality for MLE.

Following the same derivation of (74) with the upper bound of $\hat{F}_{ib}, \hat{G}_{ib}, \hat{H}_{ib}$ in Lemma 12, we also have

$$L(\hat{z}) \leq \frac{3}{2}\xi' + \frac{1}{2} L(\hat{z}),$$

which indicates the conclusion $\ell(\hat{z}, z) \lesssim \Delta_{\min}^2 \exp\left(-\frac{p^{K-1}}{r^{K-1}} \Delta_{\min}^2\right)$.

Step 4 (Verification of Condition 1). Last, we verify the Condition 1 under high probability to finish the proof. Note that the inequalities (62), (63), and (64) describe the property of the sub-Gaussian noise tensor $\mathcal{E}$, and the readers can find the proof directly in Han et al. (2022a, Step 5, Proof of Theorem 2). The initial condition (66) for MLE is satisfied by Lemma 13. Here, we include only the verification of inequalities (65) and (66) for algorithm estimators.

Now, we verify the oracle loss condition (65). Recall the definition of ξ,

$$\xi = \frac{1}{p} \sum_{i \in [p]} \theta(i) \sum_{b \in [r]} \mathbb{1}\left\{ \left\langle \boldsymbol{E}_{i:} \boldsymbol{V}, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s \right\rangle \leq -\frac{\theta(i) m}{4} \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \right\} \cdot \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2.$$

Let $e_i = \boldsymbol{E}_{i:} \boldsymbol{V}$ denote the aggregated noise vector for all $i \in [p]$, and e_i's are independent zero-mean sub-Gaussian vector in $\mathbb{R}^{r^{K-1}}$. The entries in e_i are independent zero-mean sub-Gaussian variables with sub-Gaussian norm upper bounded by $m_1 \sqrt{r^{K-1}/p^{K-1}}$ with some positive constant m_1. We have the probability inequality

$$\mathbb{P}\left(\left\langle e_i, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s \right\rangle \leq -\frac{\theta(i) m}{4} \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \right) \leq P_1 + P_2 + P_3,$$

where

$$P_1 = \mathbb{P}\left(\left\langle e_i, [\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s \right\rangle \leq -\frac{\theta(i)m}{8}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2\right),$$
$$P_2 = \mathbb{P}\left(\left\langle e_i, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\boldsymbol{S}_{z(i):}]^s \right\rangle \leq -\frac{\theta(i)m}{16}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2\right),$$
$$P_3 = \mathbb{P}\left(\left\langle e_i, [\boldsymbol{S}_{b:}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s \right\rangle \leq -\frac{\theta(i)m}{16}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2\right).$$

For P_1, notice that the inner product $\left\langle e_j, \boldsymbol{S}^s_{z(j):} - \boldsymbol{S}^s_{b:}\right\rangle$ is a sub-Gaussian variable with sub-Gaussian norm bounded by $m_2\sqrt{r^{K-1}/p^{K-1}}\|\boldsymbol{S}^s_{z(i):} - \boldsymbol{S}^s_{b:}\|$ with some positive constant m_2. Then, by Chernoff bound, we have

$$P_1 \lesssim \exp\left(-\frac{p^{K-1}}{r^{K-1}}\|[\boldsymbol{S}_{z(j):}]^s - [\boldsymbol{S}_{b:}]^s\|^2\right). \tag{75}$$

For P_2 and P_3, we only need to derive the upper bound of P_2 due to the symmetry. By the law of total probability, we have

$$P_2 \leq P_{21} + P_{22}, \tag{76}$$

where with some positive constant $t > 0$,

$$P_{21} = \mathbb{P}\left(t \leq \|[\tilde{\boldsymbol{S}}_{z(i):}]^s - [\boldsymbol{S}_{z(i):}]^s\|\right),$$
$$P_{22} = \mathbb{P}\left(\left\langle e_i, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\boldsymbol{S}_{z(i):}]^s \right\rangle \leq -\frac{\theta(i)m}{16}\cdot\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \middle| \|[\tilde{\boldsymbol{S}}_{z(i):}]^s - [\boldsymbol{S}_{z(i):}]^s\| < t\right).$$

For P_{21}, note that the term $\boldsymbol{W}^T_{:z(i)}\boldsymbol{E}\boldsymbol{V} = \frac{\sum_{j\neq i, j\in[p]}\mathbb{1}\{z(j)=z(i)\}e_j}{\sum_{j\in[p]}\mathbb{1}\{z(j)=z(i)\}}$ is a sub-Gaussian vector with sub-Gaussian norm bounded by $m_3\sqrt{r^K/p^K}$ with some positive constant m_3. This implies

$$P_{21} \leq \mathbb{P}\left(t\|\boldsymbol{S}_{z(i):}\| \leq \|\tilde{\boldsymbol{S}}_{z(i):} - \boldsymbol{S}_{z(i):}\|\right)\mathbb{P}\left(c_3 t \leq \|\boldsymbol{W}^T_{:z(i)}\boldsymbol{E}\boldsymbol{V}\|\right) \lesssim \exp\left(-\frac{p^K t^2}{r^K}\right), \tag{77}$$

where the first inequality follows from the basic inequality in Lemma 4, the second inequality follows from the assumption that $\min_{a\in[r]}\|\boldsymbol{S}_{z(i):}\| \geq c_3 > 0$ in (2), and the last inequality follows from the Bernstein inequality.

For P_{22}, the inner product $\left\langle e_i, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\boldsymbol{S}_{z(i):}]^s\right\rangle$ is also a sub-Gaussian variable with sub-Gaussian norm $m_4\sqrt{r^{K-1}/p^{K-1}}t$, conditioned on $\|[\tilde{\boldsymbol{S}}_{z(i):}]^s - [\boldsymbol{S}_{z(i):}]^s\| < t$ with some positive constant m_4. Then, by Chernoff bound, we have

$$P_{22} \lesssim \exp\left(-\frac{p^{K-1}}{r^{K-1}t^2}\|[\boldsymbol{S}_{z(j):}]^s - [\boldsymbol{S}_{b:}]^s\|^4\right). \tag{78}$$

We take $t = \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|$ in P_{21} and P_{22}, and plug the inequalities (77) and (78) into to the upper bound for P_2 in (76). We obtain that

$$P_2 \lesssim \exp\left(-\frac{p^{K-1}}{r^{K-1}}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2\right). \tag{79}$$

Combining the upper bounds (75) and (79) gives

$$\mathbb{P}\left(\left\langle e_i, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s\right\rangle \leq -\frac{\theta(i)m}{4}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2\right) \lesssim \exp\left(-\frac{p^{K-1}}{r^{K-1}}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2\right) \tag{80}$$

Hence, we have

$$\begin{aligned}
\mathbb{E}\xi &= \frac{1}{p}\sum_{i\in[p]}\theta(i)\sum_{b\in[r]}\mathbb{P}\left\{\left\langle \boldsymbol{E}_{i:}\boldsymbol{V}, [\tilde{\boldsymbol{S}}_{z(i):}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s\right\rangle \leq -\frac{\theta(i)m}{4}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2\right\}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \\
&\lesssim \frac{1}{p}\sum_{i\in[p]}\theta(i)\max_{i\in[p],b\in[r]}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \cdot \exp\left(-\frac{p^{K-1}}{r^{K-1}}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2\right) \\
&\leq \exp\left(-M\frac{p^{K-1}}{r^{K-1}}\Delta_{\min}^2\right), \qquad (81)
\end{aligned}$$

where M is a positive constant, the first inequality follows from the constraint that $\sum_{i\in[p]}\theta(i) = p$, and the last inequality follows from (80).

By Markov's inequality, we have

$$\mathbb{P}\left(\xi \lesssim \mathbb{E}\xi + \exp\left(-\frac{Mp^{K-1}}{2r^{K-1}}\Delta_{\min}^2\right)\right) \geq 1 - C\exp\left(-\frac{Mp^{K-1}}{2r^{K-1}}\Delta_{\min}^2\right),$$

and thus the condition (65) holds with probability at least $1 - C\exp\left(-\frac{Mp^{K-1}}{2r^{K-1}}\Delta_{\min}^2\right)$ for some constant $C > 0$.

The initialization condition for MLE also holds.

For ξ', notice that $\langle \boldsymbol{E}_i, \boldsymbol{A}^s_{a:} - \boldsymbol{A}^s_{b:}\rangle$ is a sub-Gaussian vector with variance bounded by $\|\boldsymbol{A}^s_{a:} - \boldsymbol{A}^s_{b:}\|^2$ and

$$\begin{aligned}
\mathbb{P}\left(t \leq \|[\bar{\boldsymbol{A}}_{a:}]^s - \boldsymbol{A}^s_{a:}\|\right) &\leq \left(t \leq \|[\boldsymbol{P}_{:a}^T\boldsymbol{Y}\boldsymbol{Q}]^s - [\boldsymbol{P}_{:a}^T\boldsymbol{X}\boldsymbol{Q}]^s\|\right) \\
&\leq \mathbb{P}(t\min_{a\in[r]}\|\boldsymbol{S}_{a:}\| \leq \|\boldsymbol{P}_{:a}^T\boldsymbol{E}\boldsymbol{Q}\|) \\
&\lesssim \exp\left(-\frac{p^K t^2}{r^K}\right),
\end{aligned}$$

where the first inequality follows from the property in later inequality (102). We also have

$$\xi' \lesssim \left(-\frac{p^{K-1}}{r^{K-1}}\Delta_{\min}^2\right).$$

Finally, we verify the bounded loss condition (66) for algorithm estimator by induction. With output $z^{(0)}$ from Sub-algorithm 2 and the assumption SNR $\geq \tilde{C}p^{-K/2}\log p$, by Theorem 4, we have

$$L^{(0)} \leq \frac{\bar{C}\Delta_{\min}^2}{\tilde{C}r\log p}, \qquad \text{when } p \text{ is large enough.}$$

Therefore, the condition (66) holds for $t = 0$. Assume that the condition (66) also holds for all $t \leq t_0$. Then, by the decomposition (74), we have

$$
\begin{aligned}
L^{(t_0+1)} &\leq \frac{3}{2}\xi + \frac{1}{2}L^{(t_0)} \\
&\leq \exp\left(-M\frac{p^{K-1}}{r^{K-1}}\Delta_{\min}^2\right) + \frac{\Delta_{\min}^2}{r\log p} \\
&\leq \frac{\bar{C}}{\tilde{C}}\frac{\Delta_{\min}^2}{r\log p},
\end{aligned}
$$

where the second inequality follows from the condition (65) and the last inequality follows from the assumption that $\Delta_{\min}^2 \gtrsim p^{-K/2}\log p$. Thus, the condition (66) holds for t_0+1, and the condition (66) is proved by induction. $\square$

Useful Lemmas for the Proof of Theorem 5

Lemma 10 (Singular-value property of membership matrices)**.** Under the setup of Theorem 5, suppose that the condition (66) holds. Then, for all $a \in [r]$, we have $|\left(z^{(t)}\right)^{-1}(a)| \asymp p/r$. Moreover, we have

$$
\begin{aligned}
&\lambda_r(\boldsymbol{M}) \asymp \|\boldsymbol{M}\|_\sigma \asymp \sqrt{p/r},\ \lambda_r(\boldsymbol{W}) \asymp \|\boldsymbol{W}\|_\sigma \asymp \sqrt{r/p}, \\
&\lambda_r(\boldsymbol{P}) \asymp \|\boldsymbol{P}\|_\sigma \asymp \min_{a\in[r]}\|\boldsymbol{\theta}_{z^{-1}(a)}\|^{-1} \lesssim \sqrt{r/p}.
\end{aligned} \tag{82}
$$

The inequalities (82) also hold by replacing $\boldsymbol{M}$ and $\boldsymbol{W}$ to $\boldsymbol{M}^{(t)}$ and $\boldsymbol{W}^{(t)}$ respectively. Further, we have

$$
\lambda_r(\boldsymbol{W}\boldsymbol{W}^T) \asymp \left\|\boldsymbol{W}\boldsymbol{W}^T\right\|_\sigma \asymp r/p, \tag{83}
$$

which is also true for $\boldsymbol{W}^{(t)}\boldsymbol{W}^{(t),T}$.

Proof of Lemma 10. The proof for the inequality (82) for $\boldsymbol{M}, \boldsymbol{W}$ can be found in Han et al. (2022a, Proof of Lemma 4). The inequalities for $\boldsymbol{P}$ follows the same derivation with balance assumption on $\boldsymbol{\theta}$ and $\min_{i\in[p]}\theta(i) \geq c$.

For inequality (83), note that for all $k \in [r]$,

$$
\lambda_k(\boldsymbol{W}\boldsymbol{W}^T) = \sqrt{\text{eigen}_k(\boldsymbol{W}\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{W}^T)} \asymp \sqrt{\frac{r}{p}\text{eigen}_k(\boldsymbol{W}\boldsymbol{W}^T)} = \sqrt{\frac{r}{p}\lambda_k^2(\boldsymbol{W})} \asymp \frac{r}{p},
$$

where $\text{eigen}_k(\boldsymbol{A})$ denotes the k-th largest eigenvalue of the square matrix $\boldsymbol{A}$, the first inequality follows the fact that $\boldsymbol{W}^T\boldsymbol{W}$ is a diagonal matrix with elements of order r/p, and the second equation follows from the definition of singular value. $\square$

Lemma 11 (Upper bound for $F_{ib}^{(t)}, G_{ib}^{(t)}$ and $H_{ib}^{(t)}$)**.** Under the Condition 1 and the setup of Theorem 5 with fixed $r \geq 2$, assume the constant $\tilde{C}$ in the condition $\text{SNR} \geq \tilde{C}p^{-K/2}\log p$ is large enough to satisfy the inequalities (95) and (99). As $p \to \infty$, we have

$$
\max_{i\in[p]}\max_{b\neq z(i)}\frac{\left(F_{ib}^{(t)}\right)^2}{\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2} \lesssim \frac{rL^{(t)}}{\Delta_{\min}^2}\|\boldsymbol{E}_{i:}\boldsymbol{V}\|^2 + \left(1 + \frac{rL^{(t)}}{\Delta_{\min}^2}\right)\|\boldsymbol{E}_{i:}(\boldsymbol{V} - \boldsymbol{V}^{(t)})\|^2, \tag{84}
$$

$$\max_{i\in[p]}\max_{b\neq z(i)}\frac{\left(G_{ib}^{(t)}\right)^2}{\|[\boldsymbol{S}_{z(i):}]^s-[\boldsymbol{S}_{b:}]^s\|^2}\leq\frac{1}{512}\left(\Delta_{\min}^2+L^{(t)}\right),\tag{85}$$

$$\max_{i\in[p]}\max_{b\neq z(i)}\frac{\left|H_{ib}^{(t)}\right|}{\|[\boldsymbol{S}_{z(i):}]^s-[\boldsymbol{S}_{b:}]^s\|^2}\leq\frac{1}{4}.\tag{86}$$

Similarly, when the SNR $\geq\tilde{C}p^{-(K-1)}\log p$ with a large constant $\tilde{C}$, we have

$$\max_{i\in[p]}\max_{b\neq z(i)}\frac{\left(\hat{F}_{ib}\right)^2}{\|[\boldsymbol{A}_{z(i):}]^s-[\boldsymbol{A}_{b:}]^s\|^2}\lesssim p^{K-1}\frac{rL(\hat{z})}{\Delta_{\min}^2}$$

$$\max_{i\in[p]}\max_{b\neq z(i)}\frac{\left(\hat{G}_{ib}\right)^2}{\|[\boldsymbol{A}_{z(i):}]^s-[\boldsymbol{A}_{b:}]^s\|^2}\leq\frac{1}{512}\left(\Delta_{\min}^2+L(\hat{z})\right),$$

$$\max_{i\in[p]}\max_{b\neq z(i)}\frac{\left|\hat{H}_{ib}\right|}{\|[\boldsymbol{A}_{z(i):}]^s-[\boldsymbol{A}_{b:}]^s\|^2}\leq\frac{1}{4}.$$

Proof of Lemma 11. We prove the the first three inequalities in Lemma 11 separately.

1. Upper bound for $F_{ib}^{(t)}$, i.e., inequality (84). Recall the definition of $F_{ib}^{(t)}$,

$$F_{ib}^{(t)}=2\left\langle\boldsymbol{E}_{i:}\boldsymbol{V}^{(t)},\left([\tilde{\boldsymbol{S}}_{z(i):}]^s-[\boldsymbol{S}_{z(i):}^{(t)}]^s\right)-\left([\tilde{\boldsymbol{S}}_{b:}]^s-[\boldsymbol{S}_{b:}^{(t)}]^s\right)\right\rangle+2\left\langle\boldsymbol{E}_{i:}(\boldsymbol{V}-\boldsymbol{V}^{(t)}),[\tilde{\boldsymbol{S}}_{z(i):}]^s-[\tilde{\boldsymbol{S}}_{b:}]^s\right\rangle.$$

By Cauchy-Schwartz inequality, we have

$$\begin{aligned}\left(F_{ib}^{(t)}\right)^2&\leq 8\left(\left\langle\boldsymbol{E}_{i:}\boldsymbol{V}^{(t)},\left([\tilde{\boldsymbol{S}}_{z(i):}]^s-[\boldsymbol{S}_{z(i):}^{(t)}]^s\right)-\left([\tilde{\boldsymbol{S}}_{b:}]^s-[\boldsymbol{S}_{b:}^{(t)}]^s\right)\right\rangle\right)^2\\&\quad+8\left(\left\langle\boldsymbol{E}_{i:}(\boldsymbol{V}-\boldsymbol{V}^{(t)}),[\tilde{\boldsymbol{S}}_{z(i):}]^s-[\tilde{\boldsymbol{S}}_{b:}]^s\right\rangle\right)^2\\&\leq 8\left(\|\boldsymbol{E}_{i:}\boldsymbol{V}\|^2+\|\boldsymbol{E}_{i:}(\boldsymbol{V}-\boldsymbol{V}^{(t)})\|^2\right)\max_{a\in[r]^s}\|[\tilde{\boldsymbol{S}}_{a:}]^s-[\boldsymbol{S}_{a:}^{(t)}]^s\|\\&\quad+\|\boldsymbol{E}_{i:}(\boldsymbol{V}-\boldsymbol{V}^{(t)})\|^2\|[\tilde{\boldsymbol{S}}_{z(i):}]^s-[\tilde{\boldsymbol{S}}_{b:}]^s\|.\end{aligned}\tag{87}$$

Note that for all $a\in[r]$,

$$\begin{aligned}\|[\tilde{\boldsymbol{S}}_{a:}]^s-[\boldsymbol{S}_{a:}^{(t)}]^s\|^2&=\|[\boldsymbol{W}_{:a}^T\boldsymbol{Y}\boldsymbol{V}]^s-[\boldsymbol{W}_{:a}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2\\&\leq 2\|[\boldsymbol{W}_{:a}^T\boldsymbol{Y}\boldsymbol{V}]^s-[\boldsymbol{W}_{:a}^{(t),T}\boldsymbol{Y}\boldsymbol{V}]^s\|^2+2\|[\boldsymbol{W}_{:a}^{(t),T}\boldsymbol{Y}\boldsymbol{V}]^s-[\boldsymbol{W}_{:a}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2\\&\lesssim\frac{r^2(L^{(t)})^2}{\Delta_{\min}^2}+\frac{rr^{2K}+pr^{K+2}}{p^K}\frac{L^{(t)}}{\Delta_{\min}^2}\\&\lesssim rL^{(t)}+\frac{rr^{2K}+pr^{K+2}}{p^K}\frac{L^{(t)}}{\Delta_{\min}^2}\\&\lesssim rL^{(t)},\end{aligned}\tag{88}$$

where the second inequality follows from the inequalities (105) and (106) in Lemma 12, the third inequality follows from the condition (66) in Condition 1, and the last inequality follows from the assumption that $\Delta_{\min}^2 \geq \tilde{C} p^{-K/2} \log p$.

Note that

$$
\begin{aligned}
\|[\tilde{\boldsymbol{S}}_{z(i):}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s\|^2 &= \|[\tilde{\boldsymbol{S}}_{z(i):}]^s - [\boldsymbol{S}_{z(i):}]^s + [\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s + [\boldsymbol{S}_{b:}]^s - [\tilde{\boldsymbol{S}}_{b:}]^s\|^2 \\
&\lesssim \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 + \max_{a \in [r]} \|[\boldsymbol{S}_{a:}]^s - [\tilde{\boldsymbol{S}}_{a:}]^s\|^2 \\
&\lesssim \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 + \max_{a \in [r]} \frac{1}{\|\boldsymbol{S}_{a:}\|^2} \|\boldsymbol{W}_{:a}^T \boldsymbol{E} \boldsymbol{V}\|^2 \\
&\lesssim \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2, \qquad (89)
\end{aligned}
$$

where the second inequality follows from Lemma 4, and the last inequality follows from the assumptions on $\|\boldsymbol{S}_{a:}\|$ in the parameter space (2), the inequality (62) in Condition 1 and the assumption $\Delta_{\min}^2 \gtrsim p^{-K/2} \log p$.

Therefore, we finish the proof of inequality (84) by plugging the inequalities (88) and (89) into the upper bound (87).

2. Upper bound for $G_{ib}^{(t)}$, i.e., inequality (85). By definition of $G_{ib}^{(t)}$, we rearrange terms and obtain

$$
\begin{aligned}
G_{ib}^{(t)} = &\left(\|[\boldsymbol{X}_{i:} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{z(i):}^{(t)}]^s\|^2 - \|[\boldsymbol{X}_{i:} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:z(i)}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s\|^2 \right) \\
&\quad - \left(\|[\boldsymbol{X}_{i:} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{b:}^{(t)}]^s\|^2 - \|[\boldsymbol{X}_{i:} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s\|^2 \right) \\
&= 2 \left\langle [\boldsymbol{X}_{i:} \boldsymbol{V}^{(t)}]^s, \left([\boldsymbol{W}_{:z(i)}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{z(i):}^{(t)}]^s \right) - \left([\boldsymbol{W}_{:b}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{b:}^{(t)}]^s \right) \right\rangle \\
&= G_1 + G_2 - G_3, \qquad (90)
\end{aligned}
$$

where

$$
\begin{aligned}
G_1 &= \|[\boldsymbol{W}_{:z(i)}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{z(i):}^{(t)}]^s\|^2 - \|[\boldsymbol{W}_{:b}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{b:}^{(t)}]^s\|^2, \\
G_2 &= 2 \left\langle [\boldsymbol{X}_{i:} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:z(i)}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s, [\boldsymbol{W}_{:z(i)}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{z(i):}^{(t)}]^s \right\rangle, \\
G_3 &= 2 \left\langle [\boldsymbol{X}_{i:} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s, [\boldsymbol{W}_{:b}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{b:}^{(t)}]^s \right\rangle.
\end{aligned}
$$

For G_1, we have

$$
\begin{aligned}
|G_1|^2 &\leq \left| \|[\boldsymbol{W}_{:z(i)}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{z(i):}^{(t)}]^s\|^2 - \|[\boldsymbol{W}_{:b}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{b:}^{(t)}]^s\|^2 \right|^2 \\
&\leq \max_{a \in [r]} \|[\boldsymbol{W}_{:a}^T \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:a}^{(t),T} \boldsymbol{Y} \boldsymbol{V}^{(t)}]^s\|^4 \\
&\leq C^4 \frac{r^4}{\Delta_{\min}^4} (L^{(t)})^4 + \frac{r^2 r^{4K} + p^2 r^{2K+4}}{p^{2K}} \frac{(L^{(t)})^2}{\Delta_{\min}^4} \\
&\leq C^4 \frac{\bar{C}}{\tilde{C}^3} \left(\Delta_{\min}^4 + \Delta_{\min}^2 L^{(t)} \right), \qquad (91)
\end{aligned}
$$

where the third inequality follows from the inequality (107) in Lemma 12 and the last inequality follows from the assumption that $\Delta_{\min}^2 \geq \tilde{C}p^{-K/2}\log p$ and inequality (66) in Condition 1.

For G_2, noticing that $[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s = [\boldsymbol{W}_{z(i):}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s$, we have

$$
\begin{aligned}
|G_2|^2 &\leq 2\|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:z(i)}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2\|[\boldsymbol{W}_{:z(i)}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{S}_{z(i):}^{(t)}]^s\|^2 \\
&\leq \frac{2}{\|\boldsymbol{W}_{z(i):}^T\boldsymbol{X}\boldsymbol{V}^{(t)}\|^2}\max_{a\in[r]}\|\boldsymbol{W}_{:a}^T\boldsymbol{E}\boldsymbol{V}^{(t)}\|^2\max_{a\in[r]}\|[\boldsymbol{W}_{:a}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:a}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2 \\
&\leq C'\frac{r^{2K-1}+Kpr^{K+1}}{p^K}\left(\frac{r^2}{\Delta_{\min}^2}(L^{(t)})^2 + \frac{rr^{2K}+pr^{K+2}}{p^K}\frac{L^{(t)}}{\Delta_{\min}^2}\right) \\
&\leq \frac{C'}{\tilde{C}^2}\Delta_{\min}^2 L^{(t)},
\end{aligned}
\tag{92}
$$

where C' is a positive universal constant, the second inequality follows from Lemma 4, the third inequality follows from the inequality (63) in Condition 1, the inequalities (107) and (126) in the proof of Lemma 12, and the last inequality follows from the assumption $\Delta_{\min}^2 \geq \tilde{C}p^{-K/2}\log p$ and inequality (66) in Condition 1.

For G_3, note that by triangle inequality

$$
\begin{aligned}
\|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s\|^2 &\leq \|\boldsymbol{S}_{z(i):}^s - \boldsymbol{S}_{b:}^s\|^2 + 2\max_{a\in[r]}\|[\boldsymbol{W}_{:a}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:a}^T\boldsymbol{X}\boldsymbol{V}]^s\|^2 \\
&\leq \|\boldsymbol{S}_{z(i):}^s - \boldsymbol{S}_{b:}^s\|^2 + C\frac{r^2(L^{(t)})^2}{\Delta_{\min}^2},
\end{aligned}
\tag{93}
$$

where the last inequality follows from the inequality (125) in the proof of Lemma 12 and C is a positive constant. Then we have

$$
\begin{aligned}
|G_3|^2 &\leq 2\|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2\max_{a\in[r]}\|[\boldsymbol{W}_{:a}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:a}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2 \\
&\leq 2\left(\|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s\|^2 + \|[\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s\|^2\right) \\
&\quad\times\max_{a\in[r]}\left\|[\boldsymbol{W}_{:a}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:a}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\right\|^2 \\
&\leq C^2\left(\|\boldsymbol{S}_{z(i):}^s - \boldsymbol{S}_{b:}^s\|^2 + C\frac{r^2(L^{(t)})^2}{\Delta_{\min}^2}\right)\left(\frac{r^2(L^{(t)})^2}{\Delta_{\min}^2} + \frac{rr^{2K}+pr^{K+2}}{p^K}\frac{L^{(t)}}{\Delta_{\min}^2}\right) + \frac{C'}{\tilde{C}^2}\Delta_{\min}^2 L^{(t)} \\
&\leq \frac{C^2\bar{C}^2}{\tilde{C}}\|\boldsymbol{S}_{z(i):}^s - \boldsymbol{S}_{b:}^s\|^2(\Delta_{\min}^2 + L^{(t)}) + \frac{C^3C'\bar{C}^2}{\tilde{C}^2}\left(\Delta_{\min}^4 + \Delta_{\min}^2 L^{(t)}\right),
\end{aligned}
\tag{94}
$$

where the third inequality follows from the same procedure to derive (91) and (92), and the last inequality follows from the assumption $\Delta_{\min}^2 \geq \tilde{C}p^{-K/2}\log p$ and inequality (66) in Condition 1.

Choose the $\tilde{C}$ such that

$$
3\left(C^4\frac{\bar{C}}{\tilde{C}^3} + \frac{C'}{\tilde{C}^2} + \frac{C^2\bar{C}^2}{\tilde{C}} + \frac{C^3C'\bar{C}^2}{\tilde{C}^2}\right) \leq \frac{1}{512}.
\tag{95}
$$

Then, we finish the proof of inequality (85) by plugging the inequalities (91), (92), and (94) into the upper bound (90).

3. Upper bound for $H_{ib}^{(t)}$, i.e., the inequality (86). By definition of H_{ib}, we rearrange terms and obtain

$$
\begin{aligned}
H_{ib} &= \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:z(i)}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2 - \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2 + \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 \\
&= \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:z(i)}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2 \\
&\qquad + \left(\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 - \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s\|\right) \\
&\qquad - \left(\|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\| - \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s\|\right) \\
&= H_1 + H_2 + H_3,
\end{aligned}
$$

where

$$
\begin{aligned}
H_1 &= \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:z(i)}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2 - \|[\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\|^2, \\
H_2 &= \|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2 - \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s\|^2, \\
H_3 &= 2\left\langle [\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s, [\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s \right\rangle.
\end{aligned}
$$

For H_1, we have

$$
|H_1| \leq \frac{4\max_{a\in[r]}\|\boldsymbol{W}_{:a}^T\boldsymbol{E}\boldsymbol{V}^{(t)}\|^2}{\|\boldsymbol{W}_{z(i):}^T\boldsymbol{X}\boldsymbol{V}^{(t)}\|^2} \leq \frac{r^{2K-1} + Kpr^{K+1}}{p^K} \leq \tilde{C}^{-2}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{b:}]^s\|^2, \tag{96}
$$

following the derivation of G_2 in inequality (92) and the assumption that $\Delta_{\min}^2 \geq \tilde{C}p^{-K/2}\log p$.

For H_2, by the inequality (93), we have

$$
|H_2| \lesssim 2\max_{a\in[r]}\|[\boldsymbol{W}_{:a}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:a}^T\boldsymbol{X}\boldsymbol{V}]^s\|^2 \lesssim \frac{r^2(L^{(t)})^2}{\Delta_{\min}^2} \leq C\frac{\bar{C}^2}{\tilde{C}^2}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{a:}]^s\|^2, \tag{97}
$$

where the last inequality follows from the condition (66) in Condition 1.

For H_3, by Cauchy-Schwartz inequality, we have

$$
|H_3| \lesssim \|[\boldsymbol{X}_{i:}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}^{(t)}]^s\||H_1|^{1/2} \leq 2\tilde{C}^{-1}\|[\boldsymbol{S}_{z(i):}]^s - [\boldsymbol{S}_{a:}]^s\|^2, \tag{98}
$$

following the inequalities (93) and (96).

Choose $\tilde{C}$ such that

$$
\tilde{C}^{-2} + C\frac{\bar{C}^2}{\tilde{C}^2} + \tilde{C}^{-1} \leq \frac{1}{4}. \tag{99}
$$

Therefore, we finish the proof of inequality (86) combining inequalities (96), (97), and (98).

Next, we show the upper bounds for $\hat{F}_{ib}, \hat{G}_{ib}$ and $\hat{H}_{ib}$.

By Lemma 1, we have

$$
\|\boldsymbol{S}_{a:}^s - \boldsymbol{S}_{b:}^s\| = (1 + o(1))\|\boldsymbol{A}_{a:}^s - \boldsymbol{A}_{b:}^s\|.
$$

Also, notice that the matrix product of $\boldsymbol{B}^T$ corresponds to the padding operation in Lemma 5, and the padding weights are balanced such that $\|\boldsymbol{v}\boldsymbol{B}\| = (1+o(1))\max_a \|\boldsymbol{\theta}_{z^{-1}(a)}\|^{(K-1)/2}\|\boldsymbol{v}\|$ for all $\boldsymbol{v} \in \mathbb{R}^{r(K-1)}$. For two vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^{r^{K-1}}$, we have

$$\|\boldsymbol{v}_1^s - \boldsymbol{v}_2^s\| = (1+o(1))\|[\boldsymbol{v}_1\boldsymbol{B}^T]^s - [\boldsymbol{v}_2\boldsymbol{B}^T]^s\|. \tag{100}$$

The equation (100) also holds for $\hat{\boldsymbol{B}}^T$.

Note that for all $i \in [p]$ we have

$$\begin{aligned}
\|\boldsymbol{A}_{i:}\hat{\boldsymbol{Q}}\| &= \|\boldsymbol{S}_{z(i:)}\boldsymbol{B}^T\hat{\boldsymbol{Q}}\| \\
&= \|\boldsymbol{S}_{z(i:)}\hat{\boldsymbol{D}}^{\otimes(K-1)}\| \\
&= (1+o(1))\|\boldsymbol{S}_{z(i:)}\| \\
&= (1+o(1))\max_a \|\boldsymbol{\theta}_{z^{-1}(a)}\|^{-(K-1)/2}\|\boldsymbol{A}_{i:}\|,
\end{aligned} \tag{101}$$

where the third inequality follows from the singular property of MLE confusion matrix (132) and the last inequality follows from the fact that $\boldsymbol{A}_i = \boldsymbol{S}_{z(i:)}\boldsymbol{B}^T$ and Lemma 10. Above equation indicates that $\boldsymbol{A}_{i:}$ is the span space of the singular values as $p \to \infty$. Also, notice that the row space of $\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}\hat{\boldsymbol{B}}^T$ is equal to the column space of $\hat{\boldsymbol{Q}}$, and $\boldsymbol{A}_{i:} \neq \boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}\hat{\boldsymbol{B}}^T$ in noisy case.

Hence, for all $a \in [r]$, we have

$$\begin{aligned}
\|[\boldsymbol{X}_i\hat{\boldsymbol{Q}}]^s - [\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}]^s\| &= \left\| \frac{\boldsymbol{A}_{z(i:)}\hat{\boldsymbol{Q}}}{\|\boldsymbol{A}_{z(i:)}\hat{\boldsymbol{Q}}\|} - \frac{\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}}{\|\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}\|} \right\| \\
&= (1+o(1)) \left\| \frac{\boldsymbol{A}_{z(i:)}}{\|\boldsymbol{A}_{z(i:)}\|} - \frac{\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}\hat{\boldsymbol{B}}^T}{\|\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}\hat{\boldsymbol{B}}^T\|} \right\| \\
&= (1+o(1))\|[\boldsymbol{X}_i]^s - [\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}\hat{\boldsymbol{B}}^T]^s\|
\end{aligned} \tag{102}$$

where the second equation follows from (101), $\|\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}\hat{\boldsymbol{B}}^T\| = (1+o(1))\max_a \|\boldsymbol{\theta}_{z^{-1}(a)}\|^{(K-1)/2}\|\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}\|$, and singular property of $\hat{\boldsymbol{B}}^T$. Similar result holds after replacing $\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}$ by $\boldsymbol{P}_{:a}^T\boldsymbol{Y}\boldsymbol{Q}$ or $\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}$.

We are now ready to show the upper bounds for $\hat{F}_{ib}, \hat{G}_{ib}$ and $\hat{H}_{ib}$.

For $\hat{F}_{ib}$, we have

$$\begin{aligned}
(\hat{F}_{ib})^2 &\leq \|\boldsymbol{E}_{i:}\|^2\|[\bar{\boldsymbol{A}}_{a:}]^s - [\hat{\boldsymbol{A}}_{a:}]^s\|^2 \\
&\leq \|\boldsymbol{E}_{i:}\|^2\left[\|[\bar{\boldsymbol{S}}_{a:}\boldsymbol{B}^T]^s - [\bar{\boldsymbol{S}}_{a:}\hat{\boldsymbol{B}}^T]^s\| + \|[\bar{\boldsymbol{S}}_{a:}\hat{\boldsymbol{B}}^T]^s - [\hat{\boldsymbol{S}}_{a:}\hat{\boldsymbol{B}}^T]^s\|\right]^2 \\
&\lesssim \|\boldsymbol{E}_{i:}\|^2\left[\|[\bar{\boldsymbol{S}}_{a:}\boldsymbol{B}^T\hat{\boldsymbol{Q}}]^s - [\bar{\boldsymbol{S}}_{a:}]^s\| + \|[\bar{\boldsymbol{S}}_{a:}]^s - [\hat{\boldsymbol{S}}_{a:}]^s\|\right]^2.
\end{aligned}$$

Following similar derivations in inequalities (88), (89), and the upper bound for J_1 in the proof of Lemma 12, respectively, we have

$$\|[\bar{\boldsymbol{S}}_{a:}]^s - [\hat{\boldsymbol{S}}_{a:}]^s\| \lesssim rL(\hat{z}), \quad \|[\bar{\boldsymbol{S}}_{a:}]^s - [\bar{\boldsymbol{S}}_{b:}]^s\| \lesssim \|\boldsymbol{S}_{a:}^s - \boldsymbol{S}_{b:}^s\|^2,$$

and

$$\|[\bar{\boldsymbol{S}}_{a:}\boldsymbol{B}^T\hat{\boldsymbol{Q}}]^s - [\bar{\boldsymbol{S}}_{a:}]^s\| \lesssim L(\hat{z}).$$

We then obtain the upper bound for $\hat{F}_{ib}$ by noticing that $\|\boldsymbol{E}_i\|^2 \lesssim p^{K-1}$.

For $\hat{G}_{ib}$ and $\hat{H}_{ib}$, by the property (102), we have

$$\begin{aligned}(1+o(1))\hat{G}_{ib} &= \left(\|[\boldsymbol{X}_{i:}\hat{\boldsymbol{Q}}]^s - [\hat{\boldsymbol{S}}_{a:}]^s\|_F^2 - \|[\boldsymbol{X}_{i:}\hat{\boldsymbol{Q}}]^s - [\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}]^s\|_F^2\right) \\ &\quad - \left(\|[\boldsymbol{X}_{i:}\hat{\boldsymbol{Q}}]^s - [\hat{\boldsymbol{S}}_{b:}]^s\|_F^2 - \|[\boldsymbol{X}_{i:}\hat{\boldsymbol{Q}}]^s - [\boldsymbol{P}_{:b}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}]^s\|_F^2\right). \\ (1+o(1))\hat{H}_{ib} &= \|[\boldsymbol{X}_{i:}\hat{\boldsymbol{Q}}]^s - [\boldsymbol{P}_{:a}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}]^s\|_F^2 - \|[\boldsymbol{X}_{i:}\hat{\boldsymbol{Q}}]^s - [\boldsymbol{P}_{:b}^T\boldsymbol{Y}\hat{\boldsymbol{Q}}]^s\|_F^2 + \|\boldsymbol{A}_{a:}^s - \boldsymbol{A}_{b:}^s\|_F^2.\end{aligned}$$

We obtain the upper bounds following the proof for inequalities (85) and (86).

□

Lemma 12 (Relationship between misclustering loss and intermediate parameters)**.** Under the Condition 1 and the setup of Theorem 5 with fixed $r \geq 2$, as $p \to \infty$, we have

$$\|\boldsymbol{V} - \boldsymbol{V}^{(t)}\|_\sigma \lesssim \sqrt{\frac{r^{K-1}}{p^{K-1}}}\frac{r}{\Delta_{\min}^2}L^{(t)}, \tag{103}$$

$$\|\boldsymbol{E}(\boldsymbol{V} - \boldsymbol{V}^{(t)})\|_\sigma \lesssim \sqrt{\frac{r^{K-1}(pr^{K-1}+pr)}{p^{K-1}}}\frac{r}{\Delta_{\min}^2}L^{(t)}, \tag{104}$$

$$\max_{b\in[r]}\|[\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}]^s - [\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}]^s\| \leq C\left(\frac{rL^{(t)}}{\Delta_{\min}} + \sqrt{\frac{r^{2K}+pr^{K+1}}{p^K}}\frac{\sqrt{L^{(t)}}}{\Delta_{\min}}\right), \tag{105}$$

$$\max_{b\in[r]}\|[\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}]^s - [\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\| \leq C\left(\sqrt{\frac{rr^{2K}+pr^{K+2}}{p^K}}\frac{\sqrt{L^{(t)}}}{\Delta_{\min}} + \frac{rL^{(t)}}{\Delta_{\min}}\right), \tag{106}$$

$$\max_{b\in[r]}\|[\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\| \leq C\left(\frac{rL^{(t)}}{\Delta_{\min}} + \sqrt{\frac{rr^{2K}+pr^{K+2}}{p^K}}\frac{\sqrt{L^{(t)}}}{\Delta_{\min}}\right), \tag{107}$$

for some positive universal constant C. In addition, the inequality (106) also holds by replacing $\boldsymbol{W}_{:b}^{(t)}$ to $\boldsymbol{W}_{:b}$. Further, the above inequalities holds after replacing $\boldsymbol{W}$ to $\boldsymbol{P}$, $\boldsymbol{V}$ to $\boldsymbol{Q}$, and $L^{(t)}$ to $L(\hat{z})$.

Proof of Lemma 12. We follow and use several intermediate conclusions in Han et al. (2022a, Proof of Lemma 5). We prove each inequality separately.

1. Inequality (103). By Han et al. (2022a, Proof of Lemma 5), we have

$$\|\boldsymbol{V} - \boldsymbol{V}^{(t)}\|_\sigma \lesssim \sqrt{\frac{r^{K-1}}{p^{K-1}}}r\ell^{(t)}.$$

Then, we complete the proof of inequality (103) by applying Lemma 2 to the above inequality.

2. Inequality (104). By Han et al. (2022a, Proof of Lemma 5), we have

$$\|\boldsymbol{E}(\boldsymbol{V}-\boldsymbol{V}^{(t)})\|_\sigma \lesssim \sqrt{\frac{r^{K-1}(pr^{K-1}+pr)}{p^{K-1}}}r\ell^{(t)}.$$

Also, we complete the proof of inequality (103) by applying Lemma 2 to the above inequality.

3. Inequality (105). We upper bound the desired quantity by triangle inequality,

$$\|[\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}]^s - [\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}]^s\| \le I_1 + I_2 + I_3,$$

where

$$I_1 = \left\|\frac{\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}\|} - \frac{\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|}\right\|,$$

$$I_2 = \left\|\left(\frac{1}{\|\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}\|} - \frac{1}{\|\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}\|}\right)\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}\right\|,$$

$$I_3 = \left\|\left(\frac{1}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}\|} - \frac{1}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|}\right)\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}\right\|.$$

Next, we upper bound the quantities I_1, I_2, I_3 separately.

For I_1, we further bound I_1 by triangle inequality,

$$I_1 \le I_{11} + I_{12},$$

where

$$I_{11} = \left\|\frac{\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}\|} - \frac{\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|}\right\|, \quad I_{12} = \left\|\frac{\boldsymbol{W}_{:b}^T\boldsymbol{E}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^T\boldsymbol{X}\boldsymbol{V}\|} - \frac{\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|}\right\|.$$

We first consider I_{11}. Define the confusion matrix $\boldsymbol{D} = \boldsymbol{M}^T\boldsymbol{\Theta}^T\boldsymbol{W}^{(t)} = [\![D_{ab}]\!] \in \mathbb{R}^{r\times r}$ where

$$D_{ab} = \frac{\sum_{i\in[p]}\theta(i)\mathbb{1}\left\{z(i)=a, z^{(t)}(i)=b\right\}}{\sum_{i\in[p]}\mathbb{1}\left\{z^{(t)}(i)=b\right\}}, \text{ for all } a,b\in[r].$$

By Lemma 10, we have $\sum_{i\in[p]}\mathbb{1}\left\{z^{(t)}(i)=b\right\} \gtrsim p/r$. Then, we have

$$\sum_{a\neq b, a,b\in[r]} D_{ab} \lesssim \frac{r}{p}\sum_{i:\, z^{(t)}(i)\neq z(i)}\theta(i) \lesssim \frac{L^{(t)}}{\Delta_{\min}^2} \lesssim \frac{1}{\log p}, \tag{108}$$

and for all $b\in[r]$,

$$D_{bb} = \frac{\sum_{i\in[p]}\theta(i)\mathbb{1}\left\{z(i)=z^{(t)}(i)=b\right\}}{\sum_{i\in[p]}\mathbb{1}\left\{z^{(t)}(i)=b\right\}} \ge \frac{c(\sum_{i\in[p]}\mathbb{1}\left\{z^{(t)}(i)=b\right\} - p\ell^{(t)})}{\sum_{i\in[p]}\mathbb{1}\left\{z^{(t)}(i)=b\right\}} \gtrsim 1 - \frac{1}{\log p}, \tag{109}$$

under the inequality (66) in Condition 1. By the definition of $\boldsymbol{W}, \boldsymbol{W}^{(t)}, \boldsymbol{V}$, we have

$$\frac{\boldsymbol{W}_{:b}^T \boldsymbol{X}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^T \boldsymbol{X}\boldsymbol{V}\|} = [\boldsymbol{S}_{b:}]^s, \quad \frac{\boldsymbol{W}_{:b}^{(t),T} \boldsymbol{X}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^{(t),T} \boldsymbol{X}\boldsymbol{V}\|} = [D_{bb}\boldsymbol{S}_{b:} + \sum_{a\neq b, a\in[r]} D_{ab}\boldsymbol{S}_{a:}]^s.$$

Let α denote the angle between $\boldsymbol{S}_{b:}$ and $D_{bb}\boldsymbol{S}_{b:} + \sum_{a\neq b, a\in[r]} D_{ab}\boldsymbol{S}_{a:}$. To roughly estimate the range of α, we consider the inner product

$$\begin{aligned}
\left\langle \boldsymbol{S}_{b:}, D_{bb}\boldsymbol{S}_{b:} + \sum_{a\neq b, a\in[r]} D_{ab}\boldsymbol{S}_{a:} \right\rangle &= D_{bb}\|\boldsymbol{S}_{b:}\|^2 + \sum_{a\neq b} D_{ab}\langle \boldsymbol{S}_{b:}, \boldsymbol{S}_{a:}\rangle \\
&\geq D_{bb}\|\boldsymbol{S}_{b:}\|^2 - \sum_{a\neq b, a\in[r]} D_{ab}\|\boldsymbol{S}_{b:}\| \max_{a\in[r]}\|\boldsymbol{S}_{a:}\| \\
&\geq C,
\end{aligned}$$

where C is a positive constant, and the last inequality holds when p is large enough following the constraint of $\|\boldsymbol{S}_{b:}\|$ in parameter space (2) and the bounds of $\boldsymbol{D}$ in (108) and (109).

The positive inner product between $\boldsymbol{S}_{b:}$ and $D_{bb}\boldsymbol{S}_{b:} + \sum_{a\neq b, a\in[r]} D_{ab}\boldsymbol{S}_{a:}$ indicates $\alpha \in [0, \pi/2)$, and thus $2\sin\frac{\alpha}{2} \leq \sqrt{2}\sin\alpha$. Then, by the geometry property of trigonometric function, we have

$$\begin{aligned}
\|[D_{bb}\boldsymbol{S}_{b:} + \sum_{a\neq b, a\in[r]} D_{ab}\boldsymbol{S}_{a:}]\sin\alpha\| &= \|(\boldsymbol{I}_d - \mathrm{Proj}(\boldsymbol{S}_{b:}))\sum_{a\neq b, a\in[r]} D_{ab}\boldsymbol{S}_{a:}\| \\
&\leq \sum_{a\neq b, a\in[r]} D_{ab}\|(\boldsymbol{I}_d - \mathrm{Proj}(\boldsymbol{S}_{b:}))\boldsymbol{S}_{a:}\| \\
&= \sum_{a\neq b, a\in[r]} D_{ab}\|\boldsymbol{S}_{a:}\sin(\boldsymbol{S}_{b:}, \boldsymbol{S}_{a:})\| \\
&\leq \sum_{a\neq b, a\in[r]} D_{ab}\|\boldsymbol{S}_{a:}\|\|\boldsymbol{S}_{b:}^s - \boldsymbol{S}_{a:}^s\|, \qquad (110)
\end{aligned}$$

where the first inequality follows from the triangle inequality, and the last inequality follows from Lemma 4. Note that with bounds (108) and (109), when p is large enough, we have

$$\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\| = \|D_{bb}\boldsymbol{S}_{b:} + \sum_{a\neq b, a\in[r]} D_{ab}\boldsymbol{S}_{a:}\| \geq D_{bb}\|\boldsymbol{S}_{b:}\| - \sum_{a\neq b, a\in[r]} D_{ab}\|\boldsymbol{S}_{a:}\| \geq C_1, \qquad (111)$$

for some positive constant C_1. Notice that $I_{11} = \sqrt{1-\cos\alpha} = 2\sin\frac{\alpha}{2}$. Therefore, we obtain

$$\begin{aligned}
I_{11} &\leq \sqrt{2}\sin\alpha \\
&= \frac{\|[D_{bb}\boldsymbol{S}_{b:} + \sum_{a\neq b, a\in[r]} D_{ab}\boldsymbol{S}_{a:}]\sin\alpha\|}{\|D_{bb}\boldsymbol{S}_{b:} + \sum_{a\neq b, a\in[r]} D_{ab}\boldsymbol{S}_{a:}\|} \\
&\leq \frac{1}{C_1}\sum_{a\neq b, a\in[r]} D_{ab}\|\boldsymbol{S}_{a:}\|\|\boldsymbol{S}_{b:}^s - \boldsymbol{S}_{a:}^s\| \\
&\lesssim \frac{r}{p}\sum_{i\in[p]}\theta(i)\sum_{b\in[r]}\mathbb{1}\left\{z^{(t)}(i) = b\right\}\|\boldsymbol{S}_{b:}^s - \boldsymbol{S}_{a:}^s\|
\end{aligned}$$

$$\leq \frac{rL^{(t)}}{\Delta_{\min}}, \tag{112}$$

where the second inequality follows from (110) and (111), and the last two inequalities follow by the definition of D_a and $L^{(t)}$, and the constraint of $\|\boldsymbol{S}_{b:}\|$ in parameter space (2).

We now consider I_{12}. By triangle inequality, we have

$$I_{12} \leq \frac{1}{\|\boldsymbol{W}_{:b}^T \boldsymbol{X}\boldsymbol{V}\|}\|(\boldsymbol{W}_{:b}^T - \boldsymbol{W}_{:b}^{(t),T})\boldsymbol{E}\boldsymbol{V}\| + \frac{\|(\boldsymbol{W}_{:b}^T - \boldsymbol{W}_{:b}^{(t),T})\boldsymbol{X}\boldsymbol{V}\|}{\|\boldsymbol{W}_{:b}^T \boldsymbol{X}\boldsymbol{V}\|\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|}\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}\|.$$

By Han et al. (2022a, Proof of Lemma 5), we have

$$\|(\boldsymbol{W}_{:b}^T - \boldsymbol{W}_{:b}^{(t),T})\boldsymbol{E}\boldsymbol{V}\| \lesssim \sqrt{\frac{r^{2K} + pr^{K+1}}{p^K}}\frac{\sqrt{L^{(t)}}}{\Delta_{\min}}. \tag{113}$$

Notice that

$$\|(\boldsymbol{W}_{:b}^T - \boldsymbol{W}_{:b}^{(t),T})\boldsymbol{X}\boldsymbol{V}\| \leq \|\boldsymbol{W}_{:b}^T - \boldsymbol{W}_{:b}^{(t),T}\|\,\|\boldsymbol{X}\boldsymbol{V}\|_F \lesssim \frac{r^{3/2}L^{(t)}}{\sqrt{p}\Delta_{\min}^2}\|\boldsymbol{S}\|\|\boldsymbol{\Theta}\boldsymbol{M}\|_\sigma \lesssim \frac{\sqrt{rL^{(t)}}}{\Delta_{\min}}, \tag{114}$$

where the second inequality follows from Han et al. (2022a, Inequality (121), Proof of Lemma 5) and the last inequality follows from Lemma 6 and (66) in Condition 1. Note that $\left\|\boldsymbol{W}_{:b}^T \boldsymbol{X}\boldsymbol{V}\right\| = \|\boldsymbol{S}_{b:}\| \geq c_3$ and $\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\| \geq C_1$ by inequality (111). Therefore, we have

$$\begin{aligned}
I_{12} &\lesssim \|(\boldsymbol{W}_{:b}^T - \boldsymbol{W}_{:b}^{(t),T})\boldsymbol{E}\boldsymbol{V}\| + \|(\boldsymbol{W}_{:b}^T - \boldsymbol{W}_{:b}^{(t),T})\boldsymbol{X}\boldsymbol{V}\|\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}\| \\
&\lesssim \sqrt{\frac{r^{2K} + pr^{K+1}}{p^K}}\frac{\sqrt{L^{(t)}}}{\Delta_{\min}} + \frac{\sqrt{rL^{(t)}}}{\Delta_{\min}}\sqrt{\frac{r^{2K}}{p^K}} \\
&\lesssim \sqrt{\frac{r^{2K} + pr^{K+1}}{p^K}}\frac{\sqrt{L^{(t)}}}{\Delta_{\min}},
\end{aligned} \tag{115}$$

where second inequality follows from the inequalities (113), (114), and (62) in Condition 1.

Hence, combining inequalities (112) and (115) yields

$$I_1 \lesssim \frac{rL^{(t)}}{\Delta_{\min}} + \sqrt{\frac{r^{2K} + pr^{K+1}}{p^K}}\frac{\sqrt{L^{(t)}}}{\Delta_{\min}}. \tag{116}$$

For I_2 and I_3, recall that $\left\|\boldsymbol{W}_{:b}^T \boldsymbol{X}\boldsymbol{V}\right\| = \|\boldsymbol{S}_{b:}\| \geq c_3$ and $\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\| \geq C_1$ by inequality (111). By triangle inequality and (62) in Condition 1, we have

$$I_2 \leq \frac{\|\boldsymbol{W}_{:b}^T \boldsymbol{E}\boldsymbol{V}\|}{\|\boldsymbol{W}_{:b}^T \boldsymbol{X}\boldsymbol{V}\|} \lesssim \|\boldsymbol{W}_{:b}^T \boldsymbol{E}\boldsymbol{V}\| \lesssim \frac{r^K}{p^{K/2}}, \tag{117}$$

and

$$I_3 \leq \frac{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}\|}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|} \lesssim \|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}\| \lesssim \frac{r^K}{p^{K/2}}. \tag{118}$$

Therefore, combining the inequalities (116), (117), and (118), we finish the proof of inequality (105).

4. Inequality (106). Here we only show the proof of inequality (106) with $\boldsymbol{W}_{:b}^{(t)}$. The proof also holds by replacing $\boldsymbol{W}_{:b}^{(t)}$ to $\boldsymbol{W}_{:b}$, and we omit the repeated procedures.

We upper bound the desired quantity by triangle inequality

$$\|[\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}]^s - [\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\| \leq J_1 + J_2 + J_3,$$

where

$$\begin{aligned}
J_1 &= \left\| \frac{\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|} - \frac{\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{(t)}\|} \right\|, \\
J_2 &= \left\| \left(\frac{1}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}\|} - \frac{1}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|} \right) \boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V} \right\|, \\
J_3 &= \left\| \left(\frac{1}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}\|} - \frac{1}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{(t)}\|} \right) \boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)} \right\|.
\end{aligned}$$

Next, we upper bound the quantities J_1, J_2, J_3 separately.

For J_1, by triangle inequality, we have

$$J_1 \leq J_{11} + J_{12},$$

where

$$J_{11} = \left\| \frac{\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|} - \frac{\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{(t)}}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{(t)}\|} \right\|, \quad J_{12} = \left\| \frac{\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|} - \frac{\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}^{(t)}}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{(t)}\|} \right\|.$$

We first consider J_{11}. Define the matrix $\boldsymbol{V}^k := \boldsymbol{W}^{\otimes(k-1)} \otimes \boldsymbol{W}^{(t),\otimes(K-k)}$ for $k = 2, \ldots, K-1$, and denote $\boldsymbol{V}^1 = \boldsymbol{V}^{(t)}, \boldsymbol{V}^K = \boldsymbol{V}$. Also, define the quantity

$$J_{11}^k = \|[\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^k]^s - [\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{k+1}]^s\|,$$

for $k = 1, \ldots, K-1$. Let β_k denote the angle between $\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^k$ and $\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{k+1}$. With the same idea to prove I_{11} in inequality (112), we bound J_{11}^k by the trigonometric function of β_k.

To roughly estimate the range of β_k, we consider the inner product between $\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^k$ and $\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{k+1}$. Before the specific derivation of the inner product, note that

$$\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^k = \mathrm{Mat}_1(\mathcal{T}_k), \quad \boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{k+1} = \mathrm{Mat}_1(\mathcal{T}_{k+1}),$$

where

$$\begin{aligned}
\mathcal{T}_k &= \mathcal{X} \times_1 \boldsymbol{W}_{:b}^{(t),T} \times_2 \boldsymbol{W}^T \times_3 \cdots \times_k \boldsymbol{W}^T \times_{k+1} \boldsymbol{W}^{(t),T} \times_{k+2} \cdots \times_K \boldsymbol{W}^{(t),T} \\
\mathcal{T}_{k+1} &= \mathcal{X} \times_1 \boldsymbol{W}_{:b}^{(t),T} \times_2 \boldsymbol{W}^T \times_3 \cdots \times_k \boldsymbol{W}^T \times_{k+1} \boldsymbol{W}^T \times_{k+2} \cdots \times_K \boldsymbol{W}^{(t),T}.
\end{aligned}$$

Recall the definition of confusion matrix $\boldsymbol{D} = \boldsymbol{M}^T\boldsymbol{\Theta}^T\boldsymbol{W}^{(t)} = [\![D_{ab}]\!] \in \mathbb{R}^{r\times r}$. We have

$$\left\langle \boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^k, \boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{k+1} \right\rangle = \langle \mathrm{Mat}_{k+1}(\mathcal{T}_k), \mathrm{Mat}_{k+1}(\mathcal{T}_{k+1}) \rangle$$

$$
\begin{aligned}
&= \left\langle \boldsymbol{D}^T \boldsymbol{S}\boldsymbol{Z}^k, \boldsymbol{S}\boldsymbol{Z}^k \right\rangle \\
&= \sum_{b\in[r]} \left(D_{bb} \|\boldsymbol{S}_{b:}\boldsymbol{Z}^k\|^2 + \sum_{a\neq b, a\in[r]} D_{ab} \left\langle \boldsymbol{S}_{a:}\boldsymbol{Z}^k, \boldsymbol{S}_{b:}\boldsymbol{Z}^k \right\rangle \right) \\
&\gtrsim (1-\log p^{-1}) \min_{a\in[r]} \|\boldsymbol{S}_{a:}\boldsymbol{Z}^k\|^2 - \log p^{-1} \max_{a\in[r]} \|\boldsymbol{S}_{a:}\boldsymbol{Z}^k\|^2, \quad (119)
\end{aligned}
$$

where $\boldsymbol{Z}^k = \boldsymbol{D}_{:b} \otimes \boldsymbol{I}_r^{\otimes(k-1)} \otimes \boldsymbol{D}^{\otimes(K-k-1)}$, the equations follow by the tensor algebra and definitions, and the last inequality follows from the bounds of $\boldsymbol{D}$ in (108) and (109).

Note that

$$
\|\boldsymbol{D}\|_\sigma \leq \|\boldsymbol{D}\|_F \leq \sqrt{\sum_{b\in[r]} D_{bb}^2 + (\sum_{a\neq b, a, b\in[r]} D_{ab})^2} \lesssim \sqrt{r + \log^2 p^{-1}} \lesssim 1, \tag{120}
$$

where the second inequality follows from inequality (108), and the fact that for all $b \in [r]$,

$$
D_{bb} \lesssim \frac{r}{p} \sum_{i:\, z(i)=b} \theta(i) \lesssim 1.
$$

Also, we have

$$
\lambda_r(\boldsymbol{D}) \geq \lambda_r(\boldsymbol{W}^{(t)}) \lambda_r(\boldsymbol{\Theta}\boldsymbol{M}) \gtrsim 1, \tag{121}
$$

following the Lemma 6 and Lemma 10. Then, for all $k \in [K]$, we have

$$
1 \lesssim \|\boldsymbol{D}_{:b}\| \, \lambda_r(\boldsymbol{D})^{K-k-1} \leq \lambda_{r^{K-2}}(\boldsymbol{Z}^k) \leq \|\boldsymbol{Z}^k\|_\sigma \leq \|\boldsymbol{D}_{:b}\| \, \|\boldsymbol{D}\|_\sigma^{K-k-1} \lesssim 1. \tag{122}
$$

Thus, we have bounds

$$
\max_{a\in[r]} \|\boldsymbol{S}_{a:}\boldsymbol{Z}^k\| \leq \max_{a\in[r]} \|\boldsymbol{S}_{a:}\| \, \|\boldsymbol{Z}^k\|_\sigma \lesssim 1, \quad \min_{a\in[r]} \|\boldsymbol{S}_{a:}\boldsymbol{Z}^k\| \geq \min_{a\in[r]} \|\boldsymbol{S}_{a:}\| \, \lambda_{r^{K-2}}(\boldsymbol{Z}^k) \gtrsim 1.
$$

Hence, when p is large enough, the inner product (119) is positive, which implies $\beta_k \in [0, \pi/2)$ and thus $2\sin\frac{\beta_k}{2} \leq \sqrt{2}\sin\beta_k$.

Next, we upper bound the trigonometric function $\sin\beta_k$. Note that

$$
\begin{aligned}
\sin\beta_k &= \sin(\boldsymbol{D}_{:b}^T \boldsymbol{S}\boldsymbol{I}_r^{\otimes k-1} \otimes \boldsymbol{D}^{\otimes K-k}, \boldsymbol{D}_{:b}^T \boldsymbol{S}\boldsymbol{I}_r^{\otimes k} \otimes \boldsymbol{D}^{\otimes K-k-1}) \\
&\leq \sin\beta_{k1} + \sin\beta_{k2},
\end{aligned}
$$

where

$$
\begin{aligned}
\sin\beta_{k1} &= \sin(\boldsymbol{D}_{:b}^T \boldsymbol{S}\boldsymbol{I}_r^{\otimes k-1} \otimes \boldsymbol{D}^{\otimes K-k}, \boldsymbol{D}_{:b}^T \boldsymbol{S}\boldsymbol{I}_r^{\otimes k-1} \otimes \tilde{\boldsymbol{D}} \otimes \boldsymbol{D}^{\otimes K-k-1}), \\
\sin\beta_{k2} &= \sin(\boldsymbol{D}_{:b}^T \boldsymbol{S}\boldsymbol{I}_r^{\otimes k-1} \otimes \tilde{\boldsymbol{D}} \otimes \boldsymbol{D}^{\otimes K-k-1}, \boldsymbol{D}_{:b}^T \boldsymbol{S}\boldsymbol{I}_r^{\otimes k} \otimes \boldsymbol{D}^{\otimes K-k-1}),
\end{aligned}
$$

and $\tilde{\boldsymbol{D}}$ is the normalized confusion matrix with entries $\tilde{\boldsymbol{D}}_{ab} = \frac{\sum_{i\in[p]} \theta(i) \mathbb{1}\{z^{(t)}=b, z(i)=a\}}{\sum_{i\in[p]} \theta(i) \mathbb{1}\{z^{(t)}=b\}}$.

To bound $\sin\beta_{k1}$, recall Definition 2 that for any cluster assignment $\bar{z}$ in the ε-neighborhood of true z,

$$
\boldsymbol{p}(\bar{z}) = (|\bar{z}^{-1}(1)|, \ldots, |\bar{z}^{-1}(r)|)^T, \quad \boldsymbol{p}_{\boldsymbol{\theta}}(\bar{z}) = (\|\boldsymbol{\theta}_{\bar{z}^{-1}(1)}\|_1, \ldots, \|\boldsymbol{\theta}_{\bar{z}^{-1}(r)}\|_1)^T.
$$

Note that we have $\ell^{(t)} \leq \frac{L^{(t)}}{\Delta_{\min}^2} \leq \frac{\bar{C}}{C} r \log^{-1}(p)$ by Condition 1 and Lemma 2. Then, with the locally linear stability assumption, the $\boldsymbol{\theta}$ is $\ell^{(t)}$-locally linearly stable; i.e.,

$$\sin(\boldsymbol{p}(z^{(t)}), \boldsymbol{p}_{\boldsymbol{\theta}}(z^{(t)})) \lesssim \frac{L^{(t)}}{\Delta_{\min}}.$$

Note that $\operatorname{diag}(\boldsymbol{p}(z^{(t)}))\boldsymbol{D} = \operatorname{diag}(\boldsymbol{p}_{\boldsymbol{\theta}}(z^{(t)}))\tilde{\boldsymbol{D}}$, and $\sin(\boldsymbol{a}, \boldsymbol{b}) = \min_{c \in \mathbb{R}} \frac{\|\boldsymbol{a} - c\boldsymbol{b}\|}{\|\boldsymbol{a}\|}$ for vectors $\boldsymbol{a}, \boldsymbol{b}$ of same dimension. Let $c_0 = \arg\min_{c \in \mathbb{R}} \frac{\|\boldsymbol{p}(z^{(t)}) - c\boldsymbol{p}_{\boldsymbol{\theta}}(z^{(t)})\|}{\|\boldsymbol{p}(z^{(t)})\|}$. Then, we have

$$\begin{aligned}
\min_{c \in \mathbb{R}} \|\boldsymbol{D} - c\tilde{\boldsymbol{D}}\|_F &\leq \|\boldsymbol{I}_r - c_0 \operatorname{diag}(\boldsymbol{p}(z^{(t)}))\operatorname{diag}^{-1}(\boldsymbol{p}_{\boldsymbol{\theta}}(z^{(t)}))\|_F \|\boldsymbol{D}\|_F \\
&\lesssim \frac{\|\boldsymbol{p}(z^{(t)}) - c_0 \boldsymbol{p}_{\boldsymbol{\theta}}(z^{(t)})\|}{\min_{a \in [r]} \|\boldsymbol{\theta}_{z^{(t),-1}(a)}\|_1} \\
&= \frac{\|\boldsymbol{p}(z^{(t)})\|}{\min_{a \in [r]} \|\boldsymbol{\theta}_{z^{(t),-1}(a)}\|_1} \sin(\boldsymbol{p}(z^{(t)}), \boldsymbol{p}_{\boldsymbol{\theta}}(z^{(t)})) \\
&\lesssim \frac{L^{(t)}}{\Delta_{\min}},
\end{aligned}$$

where the last inequality follows from Lemma 10, the constraint $\min_{i \in [p]} \theta(i) \geq c > 0$, $\|\boldsymbol{p}(z^{(t)})\| \lesssim p$ and $\min_{a \in [r]} \|\boldsymbol{\theta}_{z^{(t),-1}(a)}\|_1 \gtrsim p$.

By the geometry property of trigonometric function, we have

$$\begin{aligned}
\sin \beta_{k1} &= \min_{c \in \mathbb{R}} \frac{\|\boldsymbol{D}_{:b}^T \boldsymbol{S} \boldsymbol{I}_r^{\otimes k-1} \otimes (\boldsymbol{D} - c\tilde{\boldsymbol{D}}) \otimes \boldsymbol{D}^{\otimes K-k-1}\|}{\|\boldsymbol{D}_{:b}^T \boldsymbol{S} \boldsymbol{I}_r^{\otimes k-1} \otimes \boldsymbol{D}^{\otimes K-k}\|} \\
&\leq \frac{\|\boldsymbol{D}_{:b}^T \boldsymbol{S}\| \|\boldsymbol{D} - c_0 \tilde{\boldsymbol{D}}\|_\sigma \|\boldsymbol{D}\|_\sigma^{K-k-1}}{\|\boldsymbol{D}_{:b}^T \boldsymbol{S}\| \lambda_r^{K-k}(\boldsymbol{D})} \\
&\lesssim \|\boldsymbol{D} - c_0 \tilde{\boldsymbol{D}}\|_F \\
&\lesssim \frac{L^{(t)}}{\Delta_{\min}},
\end{aligned} \tag{123}$$

where the second inequality follows from the singular property of $\boldsymbol{D}$ in (120), (121) and the constraint of $\boldsymbol{S}$ in (2).

To bound $\sin \beta_{k2}$, let $\boldsymbol{C} = \operatorname{diag}(\{\|\boldsymbol{S}_{a:}\|\}_{a \in [r]})$. We have

$$\begin{aligned}
\sin \beta_{k2} &\lesssim \frac{\left\| \boldsymbol{D}_{:b}^T \boldsymbol{S} \boldsymbol{I}_r^{\otimes k-1} \otimes (\boldsymbol{I}_r - \tilde{\boldsymbol{D}}) \otimes \boldsymbol{D}^{\otimes K-k-1} \right\|}{\|\boldsymbol{D}_{:b}^T \boldsymbol{S} \boldsymbol{I}_r^{\otimes k} \otimes \boldsymbol{D}^{\otimes K-k-1}\|} \\
&\lesssim \frac{\|(\boldsymbol{I}_r - \tilde{\boldsymbol{D}}^T)\boldsymbol{S}\boldsymbol{Z}^k\|_F}{\|\boldsymbol{D}_{:b}^T \boldsymbol{S}\| \lambda_r^{K-k-1}(\boldsymbol{D})} \\
&\lesssim \|(\boldsymbol{I}_r - \tilde{\boldsymbol{D}}^T)\boldsymbol{S}\boldsymbol{C}^{-1}\|_F \|\boldsymbol{C}\boldsymbol{Z}^k\|_\sigma \\
&\lesssim \frac{r}{p} \sum_{i \in [p]} \theta(i) \sum_{b \in [r]} \mathbb{1}\{z^{(t)}(i) = b\} \|\boldsymbol{S}_{b:}^s - \boldsymbol{S}_{z(i):}^s\|
\end{aligned}$$

$$\lesssim \frac{L^{(t)}}{\Delta_{\min}}, \tag{124}$$

where the third inequality follows from the singular property of $\boldsymbol{D}$ and the boundedness of $\boldsymbol{S}$, and the fourth inequality follows from the definition of $\tilde{\boldsymbol{D}}$, boundedness of $\boldsymbol{S}$, the lower bound of $\boldsymbol{\theta}$, and the singular property of $\boldsymbol{Z}^k$ in inequality (122), and the last line follows from the definition of $L^{(t)}$.

Combining (123) and (124) yields

$$\sin\beta_k \leq \sin\beta_{k1} + \sin\beta_{k2} \lesssim \frac{L^{(t)}}{\Delta_{\min}}.$$

Finally, by triangle inequality, we obtain

$$J_{11} \leq \sum_{k=1}^{K-1} J_{11}^k \lesssim \sum_{k=1}^{K-1} \sin\beta_k \lesssim (K-1)\frac{rL^{(t)}}{\Delta_{\min}}. \tag{125}$$

We now consider J_{12}. By triangle inequality, we have

$$J_{12} \leq \frac{1}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|}\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}(\boldsymbol{V}-\boldsymbol{V}^{(t)})\| + \frac{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}(\boldsymbol{V}-\boldsymbol{V}^{(t)})\|}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{(t)}\|}\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}^{(t)}\|.$$

Note that

$$\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{(t)}\| = \|\boldsymbol{D}^T\boldsymbol{S}\boldsymbol{Z}^1\| \geq \lambda_r(\boldsymbol{D})\,\|\boldsymbol{S}\|\,\lambda_{r^{K-2}}(\boldsymbol{Z}^1) \gtrsim 1, \tag{126}$$

where the inequality follows from the bounds (121) and (122).

By Han et al. (2022a, Proof of Lemma 5), we have

$$\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}(\boldsymbol{V}-\boldsymbol{V}^{(t)})\| \lesssim \sqrt{\frac{r^{2K+1}+pr^{2+K}}{p^K}}\frac{(K-1)\sqrt{L^{(t)}}}{\Delta_{\min}}. \tag{127}$$

Notice that

$$\begin{aligned}
\|\boldsymbol{X}(\boldsymbol{V}^k-\boldsymbol{V}^{k+1})\|_F &\leq \|(\boldsymbol{I}-\boldsymbol{D}^T)\boldsymbol{S}(\boldsymbol{I}_r^{\otimes(k-1)}\otimes\boldsymbol{D}^{\otimes(K-k-1)})\|_F \\
&\leq \|(\boldsymbol{W}^T-\boldsymbol{W}^{(t),T})\boldsymbol{\Theta}\boldsymbol{M}\|_F\,\|\boldsymbol{S}\|_F\,\|\boldsymbol{D}\|_\sigma^{K-k-1} \\
&\lesssim \|\boldsymbol{W}^T-\boldsymbol{W}^{(t),T}\|\,\|\boldsymbol{\Theta}\boldsymbol{M}\|_\sigma \\
&\lesssim \frac{\sqrt{rL^{(t)}}}{\Delta_{\min}},
\end{aligned} \tag{128}$$

where the first inequality follows from the tensor algebra in inequality (119), the second inequality follows from the fact that $\boldsymbol{I} = \boldsymbol{W}^T\boldsymbol{\Theta}\boldsymbol{M}$, and the last inequality follows from Han et al. (2022a, Proof of Lemma 5). It follows from (128) and Lemma 10 that

$$\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}(\boldsymbol{V}-\boldsymbol{V}^{(t)})\| \leq \|\boldsymbol{W}_{:b}^{(t),T}\|\sum_{k=1}^{K-1}\|\boldsymbol{X}(\boldsymbol{V}^k-\boldsymbol{V}^{k+1})\|_F \lesssim \frac{\sqrt{rL^{(t)}}}{\sqrt{p}\Delta_{\min}}. \tag{129}$$

Note that $\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|$ and $\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{(t)}\|$ are lower bounded by inequalities (111) and (126), respectively. We have

$$
\begin{aligned}
J_{12} &\lesssim \|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}(\boldsymbol{V}-\boldsymbol{V}^{(t)})\| + \|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}(\boldsymbol{V}-\boldsymbol{V}^{(t)})\|\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}^{(t)}\| \\
&\lesssim \sqrt{\frac{r^{2K+1}+pr^{2+K}}{p^K}}\frac{\sqrt{L^{(t)}}}{\Delta_{\min}} + \frac{\sqrt{rL^{(t)}}}{\sqrt{p}\Delta_{\min}}\sqrt{\frac{r^{2K}}{p^K}} \\
&\lesssim \sqrt{\frac{r^{2K+1}+pr^{2+K}}{p^K}}\frac{\sqrt{L^{(t)}}}{\Delta_{\min}},
\end{aligned}
$$

where the second inequality follows from inequalities (127), (129), and the inequality (62) in Condition 1.

For J_2 and J_3, recall that $\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|$ and $\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{(t)}\|$ are lower bounded by inequalities (111) and (126), respectively. By triangle inequality and inequality (62) in Condition 1, we have

$$
J_2 \leq \frac{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}\|}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}\|} \lesssim \|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}\| \lesssim \frac{r^K}{p^{K/2}}, \tag{130}
$$

and

$$
J_3 \leq \frac{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}^{(t)}\|}{\|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{X}\boldsymbol{V}^{(t)}\|} \lesssim \|\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{E}\boldsymbol{V}\| \lesssim \frac{r^K}{p^{K/2}}. \tag{131}
$$

Therefore, combining the inequalities (125), (130), and (131), we finish the proof of inequality (106).

5. Inequality (107). By triangle inequality, we upper bound the desired quantity

$$
\begin{aligned}
&\|[\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\| \\
&\leq \|[\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s - [\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}]^s\| + \|[\boldsymbol{W}_{:b}^T\boldsymbol{Y}\boldsymbol{V}]^s - [\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}]^s\| + \|[\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}]^s - [\boldsymbol{W}_{:b}^{(t),T}\boldsymbol{Y}\boldsymbol{V}^{(t)}]^s\| \\
&\lesssim \frac{rL^{(t)}}{\Delta_{\min}} + \sqrt{\frac{rr^{2K}+pr^{K+2}}{p^K}}\frac{\sqrt{L^{(t)}}}{\Delta_{\min}},
\end{aligned}
$$

following the inequalities (105) and (106). Therefore, we finish the proof of inequality (107).

Next, we show the intermediate inequalities holds with $\boldsymbol{P}, \boldsymbol{Q}$ and $L(\hat{z})$.

Consider the MLE confusion matrix $\hat{\boldsymbol{D}} = \boldsymbol{M}^T\boldsymbol{\Theta}^T\hat{\boldsymbol{P}} = [\![\hat{D}_{ab}]\!] \in \mathbb{R}^{r\times r}$ with entries

$$
\begin{aligned}
\hat{D}_{ab} &= \frac{\sum_{i\in[p]}\theta(i)\hat{\theta}(i)\mathbb{1}\{z(i)=a,\hat{z}(i)=b\}}{\|\hat{\boldsymbol{\theta}}_{\hat{z}^{-1}(b)}\|^2} \\
&= \frac{\sum_{i\in[p]}(1+o(p^{K-2}))(\hat{\theta}(i))^2\mathbb{1}\{z(i)=a,\hat{z}(i)=b\}}{\|\hat{\boldsymbol{\theta}}_{\hat{z}^{-1}(b)}\|^2},
\end{aligned} \tag{132}
$$

where the second equation follows from Lemma 13, and thus $\sum_{a\in[r]} \hat{\boldsymbol{D}}_{ab} = 1+o(1)$. By the derivation of (108), (109), (121), and (120), we have

$$\sum_{a\neq b\in[r]} \hat{D}_{ab} \lesssim \frac{1}{p}\sum_{i\in[p]} \mathbb{1}\{\hat{z}(i)\neq z(i)\}(\hat{\theta}(i))^2 \lesssim \frac{1}{\log p}, \quad \hat{D}_{bb} \gtrsim 1-\frac{1}{\log p}, \quad \lambda_{\min}(\hat{\boldsymbol{D}}) \asymp \|\hat{\boldsymbol{D}}\|_\sigma = (1+o(1)).$$

for all $a\neq b\in[r]$.

Now, we are ready to show the intermediate inequalities. First, by Lemma 1 and $\min_{i\in[p]}\theta(i)\geq c$, we have

$$\|\boldsymbol{S}^s_{a:} - \boldsymbol{S}^s_{b:}\| \asymp \|\boldsymbol{A}^s_{a:} - \boldsymbol{A}^s_{b:}\|.$$

Then we can replace the $L^{(t)}$ by $L(\hat{z})$ in the proof of Lemma 12. The analogies of inequalities (103), (104), (105), (106), and (107) hold by using the MLE confusion matrix and the definition of $L(\hat{z})$.

Particularly, for the analogy of (106), the usage of MLE confusion matrix avoids the stability condition on $\boldsymbol{\theta}$. Let $\bar{\boldsymbol{D}}$ be the normalized version of $\hat{\boldsymbol{D}}$. The angle in inequality (123) decays to 0 at speed $p^{-(K-2)} \lesssim \Delta_{\min}$ when $K\geq 3$, and the inequality (124) holds by the fact that

$$\|(\boldsymbol{I}_r - \bar{\boldsymbol{D}})\boldsymbol{S}\boldsymbol{C}^{-1}\|_F \lesssim \frac{r}{p}\sum_{i\in[p]}(\theta(i))^2\sum_{b\in[r]}\|\boldsymbol{S}^s_{b:} - \boldsymbol{S}^s_{z(i):}\| \lesssim \frac{r}{p}\sum_{i\in[p]}(\theta(i))^2\sum_{b\in[r]}\|\boldsymbol{A}^s_{b:} - \boldsymbol{A}^s_{z(i):}\|.$$

□

Lemma 13 (Polynomial estimation error of MLE)**.** Let $(\hat{z}, \hat{\mathcal{S}}, \hat{\boldsymbol{\theta}})$ denote the MLE in (9) with fixed $K\geq 2$ and symmetric mean tensor, and $\hat{\mathcal{X}}$ denote the mean tensor consisting of parameter $(\hat{z}, \hat{\mathcal{S}}, \hat{\boldsymbol{\theta}})$. With high probability going to 1 as $p\to\infty$, we have

$$\|\mathcal{X} - \hat{\mathcal{X}}\|_F^2 \lesssim \sigma^2\left(r^K + Kpr\right),$$

with probability going to 1. When SNR $\gtrsim p^{-(K-1)}\log p$, $\boldsymbol{\theta}$ is balanced, and $\min_{i\in[p]}\theta(i)\geq c$ for some positive constant c, the MLE satisfies

$$\frac{1}{p}\sum_{i\in[p]}\mathbb{1}\{\hat{z}(i)\neq z(i)\}(\theta(i))^2 \lesssim \frac{1}{r\log p}, \quad \frac{1}{p}\sum_{i\in[p]}\mathbb{1}\{\hat{z}(i)\neq z(i)\}(\hat{\theta}(i))^2 \lesssim \frac{1}{r\log p}, \text{ and } L(\hat{z}) \lesssim \frac{\Delta^2_{\min}}{r\log p},$$

Further, we have

$$\theta(i)^2 = (1+o(p^{-(K-2)}))\hat{\theta}(i)^2.$$

Proof of Lemma 13. Without loss of generality, we assume $\sigma^2 = 1$ and identity mapping minimizes the misclustering error for MLE. For arbitrary two sets of parameters $(z, \mathcal{S}, \boldsymbol{\theta}), (z', \mathcal{S}', \boldsymbol{\theta}') \in \mathcal{P}(\gamma)$ and corresponding mean tensors $\mathcal{X}, \mathcal{X}'$, we have

$$\operatorname{rank}(\operatorname{Mat}_k(\mathcal{X}) - \operatorname{Mat}_k(\mathcal{X}')) \leq \operatorname{rank}(\operatorname{Mat}_k(\mathcal{X})) + \operatorname{rank}(\operatorname{Mat}_k(\mathcal{X})) \leq 2r, \quad k\in[K].$$

Hence, we have

$$\mathcal{X} - \mathcal{X}' \in \mathcal{Q}(2r,\ldots,2r), \tag{133}$$

where $\mathcal{Q}(r,\ldots,r) \coloneqq \{\text{Tucker tensor with rank } (r,\ldots,r)\}$.

Then, we obtain that

$$\begin{aligned}\mathbb{P}(\|\mathcal{X}-\hat{\mathcal{X}}_{ML}\|_F \geq t) &\leq 2\mathbb{P}\left(\sup_{\mathcal{X},\mathcal{X}'\in\mathcal{P}(r,\ldots,r)} \left\langle \frac{\mathcal{X}-\mathcal{X}'}{\|\mathcal{X}-\mathcal{X}'\|_F}, \mathcal{E}\right\rangle \geq t\right)\\ &\leq 2\mathbb{P}\left(\sup_{\mathcal{T}\in\mathcal{Q}(2r,\ldots,2r)\cap\{\|\mathcal{T}\|_F=1\}} \langle \mathcal{T},\mathcal{E}\rangle \geq t\right)\\ &\lesssim \exp(-Kpr),\end{aligned}$$

with the choice $t \asymp \sigma\sqrt{(Kpr+r^K)}$. Here the first inequality follows from Wang and Zeng (2019, Lemma 1), the second inequality follows from (133), and the last inequality follows from Han et al. (2022b, Lemma E5).

When $\Delta^2_{\min} \gtrsim p^{-(K-1)}\log p$, we replace the vector $\hat{x}_{\hat{z}(i)}$ and $\hat{\boldsymbol{X}}$ by our MLE estimator in the proof of Theorem 4. With estimation error $\|\mathcal{X}-\hat{\mathcal{X}}\|_F^2 \lesssim \left(r^K + Kpr\right)$ and $\Delta^2_{\min} \gtrsim p^{-(K-1)}\log p$, we have

$$\frac{1}{p}\sum_{i\in[p]} \mathbb{1}\{\hat{z}(i)\neq z(i)\}(\theta(i))^2 \lesssim \frac{r^{K-1}}{\Delta^2_{\min}p^K}\|\mathcal{X}-\hat{\mathcal{X}}\|_F^2 \lesssim \frac{r^{K-2}}{p^{K-1}\Delta^2_{\min}} \lesssim \frac{1}{r\log p},$$

and

$$L(\hat{z}) \lesssim \frac{\Delta^2_{\min}}{r\log p}.$$

Above result holds for $\hat{\theta}(i)$ after switching the parameters $\boldsymbol{X}$ with $\hat{\boldsymbol{X}}$ and switch $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}$ in the proof.

Last, notice that for all $a\in[r]$

$$(1-O(1))\frac{p^2}{r^2}\|\boldsymbol{W}_{:a}^T\boldsymbol{X}-\hat{\boldsymbol{W}}_{:a}^T\hat{\boldsymbol{X}}\|_F^2 \leq \|\sum_{\hat{z}(i)=z(i)=a}(\theta(i)\boldsymbol{W}_{:a}^T\boldsymbol{X}-\hat{\theta}(i)\hat{\boldsymbol{W}}_{:a}^T\hat{\boldsymbol{X}})\|_F^2 \leq \|\mathcal{X}-\hat{\mathcal{X}}\|_F^2 \leq pr,$$

where the first inequality follows from the facts that $\ell(\hat{z},z) \lesssim \frac{1}{\log p}, |z^{-1}(a)| \asymp p/r$,

$$|z^{-1}(a)| - C\frac{p}{r}\ell(\hat{z},z) \leq |\hat{z}^{-1}(a)| \leq |z^{-1}(a)| + C\frac{p}{r}\ell(\hat{z},z),$$

$$|z^{-1}(a)| - C\frac{p}{r}\ell(\hat{z},z) \leq \sum_{z(i)=z(i)=a}\theta(i) \leq |z^{-1}(a)|, \quad |\hat{z}^{-1}(a)| - C\frac{p}{r}\ell(\hat{z},z) \leq \sum_{\hat{z}(i)=z(i)=a}\hat{\theta}(i) \leq |\hat{z}^{-1}(a)|.$$

Hence, for all $i\in[p]$

$$\begin{aligned}(\theta(i)-\hat{\theta}(i))^2\|\boldsymbol{W}_{:a}^T\boldsymbol{X}\|_F^2 - O(p) &\leq \|(\theta(i)-\hat{\theta}(i))\boldsymbol{W}_{:a}^T\boldsymbol{X}\|_F^2 - \|\hat{\theta}(i)(\boldsymbol{W}_{:a}^T\boldsymbol{X}-\hat{\boldsymbol{W}}_{:a}^T\hat{\boldsymbol{X}})\|_F^2\\ &\leq \|\mathcal{X}-\hat{\mathcal{X}}\|_F^2 \leq pr,\end{aligned}$$

where the first inequality follows from $\|\boldsymbol{W}_{:a}^T\boldsymbol{X}-\hat{\boldsymbol{W}}_{:a}^T\hat{\boldsymbol{X}}\|_F^2 \lesssim 1/p$ and $\hat{\theta}(i) \lesssim \frac{p}{r}$. Notice that for all $a\in[r]$

$$\|\boldsymbol{W}_{:a}^T\boldsymbol{X}\|_F^2 \geq \|\boldsymbol{S}_{a:}\|_F^2\lambda_{\min}^{2(K-1)}(\boldsymbol{\Theta}\boldsymbol{M}) \gtrsim p^{K-1}.$$

The inequality indicates that $\theta(i)^2 = (1+o(p^{-(K-2)}))\hat{\theta}(i)^2$.

□

References

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2018). Community detection in degree-corrected block models. *The Annals of Statistics*, 46(5):2153–2185.

Han, R., Luo, Y., Wang, M., and Zhang, A. R. (2022a). Exact clustering in tensor clock model: Statistical optimality and computational limit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 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.

Ke, Z. T., Shi, F., and Xia, D. (2019). Community detection for hypergraph networks via regularized tensor power iteration. *arXiv preprint arXiv:1909.06503*.

Rigollet, P. and Hütter, J.-C. (2015). High dimensional statistics. *Lecture notes for course 18S997*.

Wang, M. and Zeng, Y. (2019). Multiway clustering via tensor block models. In *Advances in Neural Information Processing Systems*, volume 32.