Supplementary Materials
"Partially Observed Dynamic Tensor Response Regression"

Jie Zhou, Will Wei Sun, Jingfei Zhang, and Lexin Li

In this supplement, we begin with the detailed proofs of the main theorems, followed by some supporting lemmas and their proofs. We conclude with some additional numerical results.

S1 Proof of Theorem 1

For the rank $r = 1$ case, the true model reduces to

$$\mathcal{Y}_i = w_1^*(\boldsymbol{\beta}_{1,4}^{*T}\mathbf{x}_i)\boldsymbol{\beta}_{1,1}^* \circ \boldsymbol{\beta}_{1,2}^* \circ \boldsymbol{\beta}_{1,3}^* + \mathcal{E}_i, \quad i = 1, \ldots, n.$$

We first bound the estimators from Steps 1, 2, and 4 of Algorithm 1, respectively. Putting together these bounds, we bound the estimator from the t-th iteration.

S1.1 Error bound of the estimator from Step 1 of Algorithm 1

In the first step, we derive the error bound for the unconstrained estimator from Step 1 of Algorithm 1. We obtain the bound for $\widetilde{\boldsymbol{\beta}}_{1,3}$ as an example, while the bounds for $\widetilde{\boldsymbol{\beta}}_{1,1}$ and $\widetilde{\boldsymbol{\beta}}_{1,2}$ can be derived similarly. When we derive the error bound for $\widetilde{\boldsymbol{\beta}}_{1,3}$, we fix other parameters. We assume $\widehat{\boldsymbol{\beta}}_{1,1}$, $\widehat{\boldsymbol{\beta}}_{1,2}$ from the previous iteration are μ-mass unit vectors, $|\widehat{w}_1 - w_1^*| < w_1^*\epsilon$, $\|\widehat{\boldsymbol{\beta}}_{1,1} - \boldsymbol{\beta}_{1,1}^*\| < \epsilon$, $\|\widehat{\boldsymbol{\beta}}_{1,2} - \boldsymbol{\beta}_{1,2}^*\| < \epsilon$, and $\|\widehat{\boldsymbol{\beta}}_{1,4} - \boldsymbol{\beta}_{1,4}^*\| < \epsilon$.

Recall $\widetilde{\boldsymbol{\beta}}_{1,3}$ is the solution of the optimization in (5), and is of the form,

$$\widetilde{\boldsymbol{\beta}}_{1,3,l} = \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}^2 \sum_{l_1,l_2} \delta_{i,l_1,l_2,l}\, \widehat{\mathcal{R}}_{i,l_1,l_2,l}\, \widehat{\boldsymbol{\beta}}_{1,1,l_1}\widehat{\boldsymbol{\beta}}_{1,2,l_2}}{\sum_{i=1}^n \widehat{\alpha}_{i,1}^2 \sum_{l_1,l_2} \widehat{w}_1\delta_{i,l_1,l_2,l}\, \widehat{\boldsymbol{\beta}}_{1,1,l_1}^2\widehat{\boldsymbol{\beta}}_{1,2,l_2}^2}$$

where $\widehat{\mathcal{R}}_i = \mathcal{Y}_i/\widehat{\alpha}_{i,1}$ and $\widehat{\alpha}_{i,1} = \widehat{\boldsymbol{\beta}}_{1,4}^\top\mathbf{x}_i$ when the rank $r = 1$. Denote $F_1 = \mathrm{supp}(\boldsymbol{\beta}_{1,1}^*) \cup \mathrm{supp}(\widehat{\boldsymbol{\beta}}_{1,1})$, $F_2 = \mathrm{supp}(\boldsymbol{\beta}_{1,2}^*) \cup \mathrm{supp}(\widehat{\boldsymbol{\beta}}_{1,2})$, and $F_3 = \mathrm{supp}(\boldsymbol{\beta}_{1,3}^*) \cup \mathrm{supp}(\widehat{\boldsymbol{\beta}}_{1,3})$, where $\mathrm{supp}(\mathbf{v})$ refers to the set of indices in $\mathbf{v}$ that are nonzero. Let $F = F_1 \circ F_2 \circ F_3$. Consider the following "equivalent" estimator,

$$\widetilde{\boldsymbol{\beta}}_{1,3,l}^* = \frac{\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,1}^2 \sum_{l_1,l_2} \delta_{i,l_1,l_2,l}(\widehat{\mathcal{R}}_{iF})_{l_1,l_2,l}\widehat{\boldsymbol{\beta}}_{1,1,l_1}\widehat{\boldsymbol{\beta}}_{1,2,l_2}}{\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,1}^2 \sum_{l_1,l_2} \widehat{w}_1\delta_{i,l_1,l_2,l}\widehat{\boldsymbol{\beta}}_{1,1,l_1}^2\widehat{\boldsymbol{\beta}}_{1,2,l_2}^2},$$

where $\mathcal{R}_{iF}$ denotes the restricted version of tensor $\mathcal{R}_i$ on the three modes indexed by F_1, F_2 and F_3. We note that, replacing $\widetilde{\boldsymbol{\beta}}_{1,3}$ by $\widetilde{\boldsymbol{\beta}}^*_{1,3}$ does not affect the iteration of $\widehat{\boldsymbol{\beta}}_{1,3}$ due to the sparsity restriction and the scaling-invariant truncation operation (Yuan and Zhang, 2013; Sun et al., 2017). Therefore, in the sequel, we assume $\widetilde{\boldsymbol{\beta}}_{1,3}$ has been replaced by $\widetilde{\boldsymbol{\beta}}^*_{1,3}$.

By the definition of $\widehat{\mathcal{R}}_{iF}$, $\boldsymbol{\beta}_{1,3}$ can be expanded as

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}_{1,3,l} = & \frac{w_1^* \frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}\sum_{l_1,l_2}\delta_{i,l_1,l_2,l}\boldsymbol{\beta}^*_{1,1,l_1}\boldsymbol{\beta}^*_{1,2,l_2}\widehat{\boldsymbol{\beta}}_{1,1,l_1}\widehat{\boldsymbol{\beta}}_{1,2,l_2}\boldsymbol{\beta}^*_{1,3,l}}{\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}^2_{i,1}\sum_{l_1,l_2}\widehat{w}_1\delta_{i,l_1,l_2,l}\widehat{\boldsymbol{\beta}}^2_{1,1,l_1}\widehat{\boldsymbol{\beta}}^2_{1,2,l_2}} \\ & + \frac{\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,1}\sum_{l_1,l_2}\delta_{i,l_1,l_2,l}(\mathcal{E}_{iF})_{l_1,l_2,l}\widehat{\boldsymbol{\beta}}_{1,1,l_1}\widehat{\boldsymbol{\beta}}_{1,2,l_2}}{\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}^2_{i,1}\sum_{l_1,l_2}\widehat{w}_1\delta_{i,l_1,l_2,l}\widehat{\boldsymbol{\beta}}^2_{1,1,l_1}\widehat{\boldsymbol{\beta}}^2_{1,2,l_2}}.\end{aligned} \tag{S1}$$

In vector form, it can be written as,

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}_{1,3} = & \frac{w_1^*\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\widehat{w}_1\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n}\langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle\langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle \boldsymbol{\beta}^*_{1,3} \\ & + \frac{w_1^*}{\widehat{w}_1}\mathbf{A}^{-1}\left\{\mathbf{B} - \mathbf{A}\frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n}\langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle\right\}\boldsymbol{\beta}^*_{1,3} \\ & + \frac{1}{\widehat{w}_1}\mathbf{A}^{-1}\left\{\sum_{i=1}^n \frac{\widehat{\alpha}_{i,1}}{n}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_1 \times_2 \widehat{\boldsymbol{\beta}}_2\right\},\end{aligned}$$

where $\mathbf{A}$ and $\mathbf{B}$ are diagonal matrices with diagonal entry,

$$\mathbf{A}_{ll} = \sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n\sum_{l_1,l_2}\delta_{i,l_1,l_2,l}\widehat{\boldsymbol{\beta}}^2_{1,1,l_1}\widehat{\boldsymbol{\beta}}^2_{1,2,l_2},\ \mathbf{B}_{ll} = \sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n\sum_{l_1,l_2}\delta_{i,l_1,l_2,l}\boldsymbol{\beta}^*_{1,1,l_1}\boldsymbol{\beta}^*_{1,2,l_2}\widehat{\boldsymbol{\beta}}_{1,1,l_1}\widehat{\boldsymbol{\beta}}_{1,2,l_2}.$$

The distance between $\widetilde{\boldsymbol{\beta}}_{1,3}$ and $(w_1^*\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1})/(\widehat{w}_1\sum_{i=1}^n \widehat{\alpha}^2_{i,1})\boldsymbol{\beta}^*_{1,3}$ is decomposed as,

$$\widetilde{\boldsymbol{\beta}}_{1,3} - \frac{w_1^*\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\widehat{w}_1\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n}\boldsymbol{\beta}^*_{1,3} = I_1 + II_1 + III_1, \tag{S2}$$

where

$$\begin{aligned}I_1 & = \frac{w_1^*\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\widehat{w}_1\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}^2_{i,1}/n}\left\{\langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle\langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle - 1\right\}\boldsymbol{\beta}^*_{1,3} \\ II_1 & = \frac{w_1^*}{\widehat{w}_1}\mathbf{A}^{-1}\left\{\mathbf{B} - \mathbf{A}\frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n}\langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle\langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle\right\}\boldsymbol{\beta}^*_{1,3} \\ III_1 & = \frac{1}{\widehat{w}_1}\mathbf{A}^{-1}\sum_{i=1}^n \frac{\widehat{\alpha}_{i,1}}{n}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_1 \times_2 \widehat{\boldsymbol{\beta}}_2.\end{aligned}$$

Next, we bound I_1, II_1, and III_1 respectively.

Bound for I_1: By definition, we have that,

$$\|I_1\| \le \left|\frac{w_1^*\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\widehat{w}_1\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n}\right|\left|1 - \langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle\langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle\right|\|\boldsymbol{\beta}^*_{1,3}\| \le \frac{2\lambda_{\max}}{\lambda_{\min}}\epsilon^2. \tag{S3}$$

The second inequality is due to the following facts. Since $|\widehat{w}_1 - w_1^*| < \epsilon w_1^* < 1/2 w_1^*$, we have $|\widehat{w}_1| > 1/2 w_1^*$. By Assumption 1 (i), $\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha_{i,1}^*/n = \widehat{\boldsymbol{\beta}}_{1,4}^\top \{\sum_i \mathbf{x}_i\mathbf{x}_i^\top/n\}\boldsymbol{\beta}_{1,4}^* \le \lambda_{\max}$, and similarly we have $\sum_{i=1}^n \widehat{\alpha}_{i,1}^2/n \ge \lambda_{\min}$. Besides, since $\|\widehat{\boldsymbol{\beta}}_{1,1} - \boldsymbol{\beta}_{1,1}^*\|$ and $\|\widehat{\boldsymbol{\beta}}_{1,2} - \boldsymbol{\beta}_{1,2}^*\| < \epsilon$, we have $\Big|1 - \langle \boldsymbol{\beta}_{1,1}^*, \widehat{\boldsymbol{\beta}}_{1,1}\rangle\langle \boldsymbol{\beta}_{1,2}^*, \widehat{\boldsymbol{\beta}}_{1,2}\rangle\Big| \le \epsilon^2$.

Bound for II_1: II_1 can be equivalently written as

$$II_1 = \frac{w_1^*}{\widehat{w}_1}\left\{\frac{1}{p}\mathbf{A}\right\}^{-1}\frac{1}{p}\left\{\mathbf{B} - \mathbf{A}\frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha_{i,1}^*/n}{\sum_{i=1}^n \widehat{\alpha}_{i,1}^2/n}\langle \boldsymbol{\beta}_{1,1}^*, \widehat{\boldsymbol{\beta}}_{1,1}\rangle\langle \boldsymbol{\beta}_{1,2}^*, \widehat{\boldsymbol{\beta}}_{1,2}\rangle\right\}\boldsymbol{\beta}_{1,3}^*.$$

We first bound each diagonal entry of the matrix $p^{-1}\mathbf{A}$. Let $Z_{i,l_1,l_2} = p^{-1}n^{-1}\widehat{\alpha}_{i,1}^2\delta_{i,l_1,l_2,l}\widehat{\boldsymbol{\beta}}_{1,1,l_1}^2\widehat{\boldsymbol{\beta}}_{1,2,l_2}^2$. Then $p^{-1}\mathbf{A}_{ll}$ has the form,

$$\frac{1}{p}\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,1}^2\sum_{l_1,l_2}\delta_{i,l_1,l_2,l}\widehat{\boldsymbol{\beta}}_{1,1,l_1}^2\widehat{\boldsymbol{\beta}}_{1,2,l_2}^2 = \sum_{i,l_1,l_2} Z_{i,l_1,l_2}.$$

Note that,

$$\Big|Z_{i,l_1,l_2} - \mathbb{E}(Z_{i,l_1,l_2})\Big| \le \Big|\Big(\frac{1}{p}\delta_{i,l_1,l_2,l} - 1\Big)\frac{1}{n}\Big| c_1^2\frac{\mu^4}{s^2} \le \frac{c_1^2\mu^4}{nps^2},$$

where the inequality holds due to the facts that $|\widehat{\alpha}_{i,1}| = |\widehat{\boldsymbol{\beta}}_{1,4}^\top\mathbf{x}_i| \le c_1$ by Assumption 1 and $\widehat{\boldsymbol{\beta}}_{1,1}$, $\widehat{\boldsymbol{\beta}}_{1,2}$ are μ-mass vectors. Here, to ensure the initials satisfy similar μ-mass assumption as in Assumption 1 (*iii*), we perform the same thresholding step as in Jain and Oh (2014). Lemma 7 ensures that the thresholded initial estimators satisfy the required Assumption 4 and 2μ-mass condition. Also, we have that,

$$\begin{aligned}\sum_{i,l_1,l_2}\mathbb{E}\Big[Z_{i,l_1,l_2} - \mathbb{E}(Z_{i,l_1,l_2})\Big]^2 &= \sum_{i,l_1,l_2}\mathbb{E}(Z_{i,l_1,l_2}^2) - \mathbb{E}^2(Z_{i,l_1,l_2})\\ &\le \frac{1}{p}\sum_{i,l_1,l_2}\frac{1}{n^2}\widehat{\alpha}_{i,1}^4\widehat{\boldsymbol{\beta}}_{1,1,l_1}^4\widehat{\boldsymbol{\beta}}_{1,2,l_2}^4 \le \frac{c_1^2\lambda_{\max}\mu^4}{nps^2}.\end{aligned}$$

By Bernstein's inequality, we have

$$P\left(\Big|\sum_{i,l_1,l_2} Z_{i,l_1,l_2} - \sum_{i,l_1,l_2}\mathbb{E}(Z_{i,l_1,l_2})\Big| \ge \gamma\right) \le 2\exp\left\{\frac{-\gamma^2/2}{\lambda_{\max} + 1/3\gamma}\times\frac{pns^2}{c_1^2\mu^4}\right\}.$$

Here γ is a fixed positive constant satisfying

$$\gamma = \frac{1}{2}\min\left\{\frac{\lambda_{\min}}{2}, c_2, \frac{\lambda_{\min}^2}{48\sqrt{10}\lambda_{\max}}, \frac{\lambda_{\min}^3}{48\sqrt{10}c_2\lambda_{\max}}\right\}, \tag{S4}$$

where $\lambda_{\min}, \lambda_{\max}, c_2$ are the same constants as defined in Assumption 1.

By (S4), it is obvious that $\gamma < 3\lambda_{\max}$. Then with probability at least $1 - 2/d^{10}$, we have,

$$\left| \frac{1}{p} \sum_{i=1}^{n} \frac{1}{n} \widehat{\alpha}_{i,1}^2 \sum_{l_1,l_2} \delta_{i,l_1,l_2,l} \widehat{\boldsymbol{\beta}}_{1,1,l_1}^2 \widehat{\boldsymbol{\beta}}_{1,2,l_2}^2 - \sum_i \widehat{\alpha}_{i,1}^2/n \right| < \gamma.$$

This implies that,

$$\frac{1}{p} \sum_{i=1}^{n} \frac{1}{n} \widehat{\alpha}_{i,1}^2 \sum_{l_1,l_2} \delta_{i,l_1,l_2,l} \widehat{\boldsymbol{\beta}}_{1,1,l_1}^2 \widehat{\boldsymbol{\beta}}_{1,2,l_2}^2 \geq \sum_i \widehat{\alpha}_{i,1}^2/n - \gamma. \tag{S5}$$

Next, we bound the norm of the vector in II_1, which is

$$\left\| 1/p \left\{ \mathbf{B} - \mathbf{A}^{-1} \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1} \alpha_{i,1}^*/n}{\sum_{i=1}^n \widehat{\alpha}_{i,1}^2/n} \langle \boldsymbol{\beta}_{1,1}^*, \widehat{\boldsymbol{\beta}}_{1,1} \rangle \langle \boldsymbol{\beta}_{1,2}^*, \widehat{\boldsymbol{\beta}}_{1,2} \rangle \right\} \boldsymbol{\beta}_{1,3}^* \right\|.$$

Denote

$$\begin{aligned} Z_{i,l_1,l_2,l} \;\; = \;\; & \left\{ \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1} \alpha_{i,1}^*/n}{\sum_{i=1}^n \widehat{\alpha}_{i,1}^2/n} \widehat{\alpha}_{i,1}^2 \langle \boldsymbol{\beta}_{1,1}^*, \widehat{\boldsymbol{\beta}}_{1,1} \rangle \langle \boldsymbol{\beta}_{1,2}^*, \widehat{\boldsymbol{\beta}}_{1,2} \rangle \widehat{\boldsymbol{\beta}}_{1,1,l_1}^2 \widehat{\boldsymbol{\beta}}_{1,2,l_2}^2 \right. \\ & \left. - \widehat{\alpha}_{i,1} \alpha_{i,1}^* \boldsymbol{\beta}_{1,1,l_1}^* \boldsymbol{\beta}_{1,2,l_2}^* \widehat{\boldsymbol{\beta}}_{1,1,l_1} \widehat{\boldsymbol{\beta}}_{1,2,l_2} \right\} \delta_{i,l_1,l_2,l} p^{-1} n^{-1} \boldsymbol{\beta}_{1,3,l}^* \mathbf{e}_l, \end{aligned}$$

where $\mathbf{e}_l$ is the column vector whose lth entry is 1 and others are 0.

By definition, we have that,

$$\frac{1}{p} \left\{ \mathbf{B} - \mathbf{A} \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1} \alpha_{i,1}^*/n}{\sum_{i=1}^n \widehat{\alpha}_{i,1}^2/n} \langle \boldsymbol{\beta}_{1,1}^*, \widehat{\boldsymbol{\beta}}_{1,1} \rangle \langle \boldsymbol{\beta}_{1,2}^*, \widehat{\boldsymbol{\beta}}_{1,2} \rangle \right\} \boldsymbol{\beta}_{1,3}^* = \sum_{i,l_1,l_2,l} Z_{i,l_1,l_2,l}.$$

Since $\boldsymbol{\beta}_{1,3}^*$, $\widehat{\boldsymbol{\beta}}_{1,1}$, $\widehat{\boldsymbol{\beta}}_{1,2}$ are μ-mass vectors, we have that,

$$\begin{aligned} \| Z_{i,l_1,l_2,l} - \mathbb{E}(Z_{i,l_1,l_2,l}) \| \;\; \leq \;\; & \frac{1}{np} \max \left| \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1} \alpha_{i,1}^*/n}{\sum_{i=1}^n \widehat{\alpha}_{i,1}^2/n} \widehat{\alpha}_{i,1} \langle \boldsymbol{\beta}_{1,1}^*, \widehat{\boldsymbol{\beta}}_{1,1} \rangle \langle \boldsymbol{\beta}_{1,2}^*, \widehat{\boldsymbol{\beta}}_{1,2} \rangle \widehat{\boldsymbol{\beta}}_{1,1,l_1} \widehat{\boldsymbol{\beta}}_{1,2,l_2} \right. \\ & \left. - \alpha_{i,1}^* \boldsymbol{\beta}_{1,1,l_1}^* \boldsymbol{\beta}_{1,2,l_2}^* \right| \max \left| \widehat{\boldsymbol{\beta}}_{1,1,l_1} \widehat{\boldsymbol{\beta}}_{1,2,l_2} \boldsymbol{\beta}_{1,3,l}^* \widehat{\alpha}_{i,1} \right|. \end{aligned} \tag{S6}$$

By the μ-mass assumption and Assumption 1 (i), $\|\mathbf{x}_i\| \leq c_1$, and $\left| \widehat{\boldsymbol{\beta}}_{1,1,l_1} \widehat{\boldsymbol{\beta}}_{1,2,l_2} \boldsymbol{\beta}_{1,3,l}^* \widehat{\alpha}_{i,1} \right| \leq c_1 \mu^3 / s^{1.5}$. We next bound the first absolute-value term on the right-hand-side of (S6).

Instead of bounding the absolute value, we bound the square of the term as,

$$
\begin{aligned}
& \left| \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n} \widehat{\alpha}_{i,1} \langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle \langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle \widehat{\boldsymbol{\beta}}_{1,1,l_1} \widehat{\boldsymbol{\beta}}_{1,2,l_2} - \alpha^*_{i,1} \boldsymbol{\beta}^*_{1,1,l_1} \boldsymbol{\beta}^*_{1,2,l_2} \right|^2 \\
\leq \; & \langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle^2 \langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle^2 \left[\left\{ \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n} \right\}^2 \widehat{\alpha}^2_{i,1} - 2 \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n} \widehat{\alpha}_{i,1}\alpha^*_{i,1} \right] + \alpha^{*2}_{i,1} \\
\leq \; & \langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle^2 \langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle^2 \left[\left\{ \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n} \right\}^2 \{\widehat{\alpha}_{i,1} - \alpha^*_{i,1} + \alpha^*_{i,1}\}^2 + \alpha^{*2}_{i,1} \right. \\
& \left. -2 \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n} \{\widehat{\alpha}_{i,1} - \alpha^*_{i,1} + \alpha^*_{i,1}\}\alpha^*_{i,1} \right] + \alpha^{*2}_{i,1}\{1 - \langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle^2 \langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle^2\} \\
\leq \; & \left[\frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\{\alpha^*_{i,1} - \widehat{\alpha}_{i,1}\}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n} \right]^2 \alpha^{*2}_{i,1} + 2(\widehat{\alpha}_{i,1} - \alpha^*_{i,1})\alpha^*_{i,1} \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}} \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\{\alpha^*_{i,1} - \widehat{\alpha}_{i,1}\}}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}} \\
& + \left\{ \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n} \right\}^2 \{\widehat{\alpha}_{i,1} - \alpha^*_{i,1}\}^2 + \alpha^{*2}_{i,1}\{1 - \langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle^2 \langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle\}^2 \\
\leq \; & \frac{\lambda^2_{\max}}{\lambda^2_{\min}} c_1^2\epsilon^2 + 2\frac{\lambda^2_{\max}}{\lambda^2_{\min}} c_1^2\epsilon^2 + \frac{\lambda^2_{\max}}{\lambda^2_{\min}} c_1^2\epsilon^2 + c_1^2\epsilon^2 = \left\{ 4\frac{\lambda^2_{\max}}{\lambda^2_{\min}} + 1 \right\} c_1^2\epsilon^2, \qquad \text{(S7)}
\end{aligned}
$$

where the last inequality holds by Assumption 1 (i), i.e., $\|x_i\| \leq c_1$, $\|\widehat{\boldsymbol{\beta}}_{1,1} - \boldsymbol{\beta}^*_{1,1}\|$, $\|\widehat{\boldsymbol{\beta}}_{1,2} - \boldsymbol{\beta}^*_{1,2}\|$, and $\|\widehat{\boldsymbol{\beta}}_{1,4} - \boldsymbol{\beta}^*_{1,4}\| < \epsilon$. Therefore, we have that,

$$
\|Z_{i,l_1,l_2,l} - \mathbb{E}(Z_{i,l_1,l_2,l})\| \;\leq\; \frac{1}{np}\frac{\mu^3}{s^{1.5}} c_1\epsilon \sqrt{\left\{ 4\frac{\lambda^2_{\max}}{\lambda^2_{\min}} + 1 \right\} c_1^2}.
$$

By Assumption 2, i.e., $p \geq c_5\{\log(d)\}^4\mu^3/ns^{1.5}$, we know that the above norm is bounded. Also, we have

$$
\begin{aligned}
& \sum_{i,l,l_1,l_2} \mathbb{E}\left\{ \|Z_{i,l_1,l_2,l} - \mathbb{E}(Z_{i,l_1,l_2,l})\|^2 \right\} \\
= \; & \frac{1}{p} \sum_{i,l,l_1,l_2} \frac{1}{n^2} \widehat{\boldsymbol{\beta}}^2_{1,1,l_1} \widehat{\boldsymbol{\beta}}^2_{1,2,l_2} \widehat{\alpha}^2_{i,1} \left\{ \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n} \widehat{\alpha}_{i,1} \langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle \langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle \widehat{\boldsymbol{\beta}}_{1,1,l_1} \widehat{\boldsymbol{\beta}}_{1,2,l_2} \right. \\
& \left. -\alpha^*_{i,1} \boldsymbol{\beta}^*_{1,1,l_1} \boldsymbol{\beta}^*_{1,2,l_2} \right\}^2 \\
\leq \; & \frac{1}{np} \sum_i \frac{\widehat{\alpha}_{i,1}}{n^2} \frac{\mu^4}{s^2} \sum_l \boldsymbol{\beta}^{*2}_{1,3,l} \sum_{l_1,l_2} \left\{ \frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha^*_{i,1}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,1}/n} \widehat{\alpha}_{i,1} \langle \boldsymbol{\beta}^*_{1,1}, \widehat{\boldsymbol{\beta}}_{1,1}\rangle \langle \boldsymbol{\beta}^*_{1,2}, \widehat{\boldsymbol{\beta}}_{1,2}\rangle \widehat{\boldsymbol{\beta}}_{1,1,l_1} \widehat{\boldsymbol{\beta}}_{1,2,l_2} \right. \\
& \left. -\alpha^*_{i,1} \boldsymbol{\beta}^*_{1,1,l_1} \boldsymbol{\beta}^*_{1,2,l_2} \right\}^2 \leq \frac{\lambda_{\max}}{np} \frac{\mu^3}{s^{1.5}} \left\{ 4\frac{\lambda^2_{\max}}{\lambda^2_{\min}} + 1 \right\} c_1^2\epsilon^2.
\end{aligned}
$$

Using vector Bernstein inequality and the fact that $\sum_{i,l_1,l_2,l} \mathbb{E}(Z_{i,l_1,l_2,l}) = 0$, we have that,

$$P\left(\left\| \sum_{i,l_1,l_2,l} Z_{i,l_1,l_2,l} \right\| \leq \gamma\epsilon \right) \geq 1 - \exp\left\{ \frac{1}{4} - \frac{\gamma^2}{8\lambda_{\max}\mu^3 \left\{ 4\frac{\lambda_{\max}^2}{\lambda_{\min}^2} + 1 \right\} c_1^2/(pns^{1.5})} \right\}.$$

By Assumption 2, we have $p \geq c_4\mu^3 \log(d)/\{ns^{1.5}\} \geq c\mu^3 \log(d)/\{ns^{1.5}\gamma^2\}$ for some positive constants c. Here γ is the constant defined in (S4). Then the following holds with probability at least $1 - e^{1/4}/d^{10}$,

$$\left\| 1/p \left\{ B - A\frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha_{i,1}^*/n}{\sum_{i=1}^n \widehat{\alpha}_{i,1}^2/n} \langle \boldsymbol{\beta}_{1,1}^*, \widehat{\boldsymbol{\beta}}_{1,1} \rangle \langle \boldsymbol{\beta}_{1,2}^*, \widehat{\boldsymbol{\beta}}_{1,2} \rangle \right\} \boldsymbol{\beta}_{1,3}^* \right\| \leq \gamma\epsilon. \tag{S8}$$

Henceforthy, by (S5) and (S8) the bound of II_1 can be simplified as,

$$\begin{aligned} \|II_1\| &\leq \left| \frac{w_1^*}{\widehat{w}_1} \right| \left\| \left\{ \frac{1}{p}\mathbf{A} \right\}^{-1} \right\| \left\| \frac{1}{p} \left\{ \mathbf{B} - \mathbf{A}\frac{\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha_{i,1}^*/n}{\sum_{i=1}^n \widehat{\alpha}_{i,1}^2/n} \langle \boldsymbol{\beta}_{1,1}^*, \widehat{\boldsymbol{\beta}}_{1,1} \rangle \langle \boldsymbol{\beta}_{1,2}^*, \widehat{\boldsymbol{\beta}}_{1,2} \rangle \right\} \boldsymbol{\beta}_{1,3}^* \right\| \\ &\leq \frac{2\gamma}{\sum_i \widehat{\alpha}_{i,1}^2/n - \gamma}\epsilon \leq \frac{2\gamma}{\lambda_{\min} - \gamma}\epsilon, \end{aligned} \tag{S9}$$

where the last inequality holds since $\sum_i \widehat{\alpha}_{i,1}/n - \gamma = \widehat{\boldsymbol{\beta}}_{1,4}^\top \frac{\sum_i \mathbf{x}_i\mathbf{x}_i^\top}{n} \widehat{\boldsymbol{\beta}}_{1,4} - \gamma \geq \lambda_{\min} - \gamma$ and $\gamma < \lambda_{\min}$ by (S4).

Bound for III_1: we expand III_1 as,

$$\|III_1\| \leq \left| \frac{1}{\widehat{w}_1} \right| \left\| \left\{ \frac{1}{p}\mathbf{A} \right\}^{-1} \right\| \left\| \sum_i \frac{1}{pn} \widehat{\alpha}_{i,1} \Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_{1,1} \times_2 \widehat{\boldsymbol{\beta}}_{1,2} \right\|.$$

By (S5), each entry of diagonal matrix $p^{-1}A$ can be bounded as

$$\frac{1}{p}\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,1}^2 \sum_{l_1,l_2} \delta_{i,l_1,l_2,l} \widehat{\boldsymbol{\beta}}_{1,1,l_1}^2 \widehat{\boldsymbol{\beta}}_{1,2,l_2}^2 \geq \sum_i \widehat{\alpha}_{i,1}^2/n - \gamma.$$

Therefore, we have that

$$\|III_1\| \leq \frac{2}{w_1^*\{\lambda_{\min} - \gamma\}} \left\| \sum_i \frac{1}{pn} \widehat{\alpha}_{i,1} \Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_{1,1} \times_2 \widehat{\boldsymbol{\beta}}_{1,2} \right\|, \tag{S10}$$

where $\gamma < \lambda_{\min}$ by (S4).

Next, we bound the term $\left\|\sum_i \frac{1}{pn}\widehat{\alpha}_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_{1,1} \times_2 \widehat{\boldsymbol{\beta}}_{1,2}\right\|$. We write

$$\begin{aligned}
&\sum_i \frac{1}{pn}\widehat{\alpha}_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_{1,1} \times_2 \widehat{\boldsymbol{\beta}}_{1,2} \\
= & \underbrace{\sum_i \frac{1}{pn}\{\widehat{\alpha}_{i,1} - \alpha^*_{i,1}\}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_{1,1} \times_2 \widehat{\boldsymbol{\beta}}_{1,2}}_{III_{11}} + \underbrace{\sum_i \frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \{\widehat{\boldsymbol{\beta}}_{1,1} - \boldsymbol{\beta}^*_{1,1}\} \times_2 \widehat{\boldsymbol{\beta}}_{1,2}}_{III_{12}} \\
& + \underbrace{\sum_i \frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \boldsymbol{\beta}^*_{1,1} \times_2 \{\widehat{\boldsymbol{\beta}}_{1,2} - \boldsymbol{\beta}^*_{1,2}\}}_{III_{13}} + \underbrace{\sum_i \frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \boldsymbol{\beta}^*_{1,1} \times_2 \boldsymbol{\beta}^*_{1,2}}_{III_{14}}.
\end{aligned}$$

Bounding III_{12}: Note that

$$\|III_{12}\| \le \left\|\sum_i \frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_2 \widehat{\boldsymbol{\beta}}_{1,2}\right\| \|\widehat{\boldsymbol{\beta}}_{1,1} - \boldsymbol{\beta}^*_{1,1}\| \le \left\|\sum_i \frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_2 \widehat{\boldsymbol{\beta}}_{1,2}\right\| \epsilon.$$

Therefore, it suffices to prove the upper bound of

$$\left\|\sum_i \frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_2 \widehat{\boldsymbol{\beta}}_{1,2}\right\|.$$

We write

$$\begin{aligned}
\left\|\sum_i \frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_2 \widehat{\boldsymbol{\beta}}_{1,2}\right\| &= \left\|\sum_i \frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_2 \{P_{\boldsymbol{\beta}^*_{1,2}} + P^\perp_{\boldsymbol{\beta}^*_{1,2}}\}\widehat{\boldsymbol{\beta}}_{1,2}\right\| \\
\le \quad & \left\|\sum_i \frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_2 P_{\boldsymbol{\beta}^*_{1,2}}\widehat{\boldsymbol{\beta}}_{1,2}\right\| + \left\|\sum_i \frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_2 P^\perp_{\boldsymbol{\beta}^*_{1,2}}\widehat{\boldsymbol{\beta}}_{1,2}\right\|, \qquad \text{(S11)}
\end{aligned}$$

where we denote $P_A = AA^\top$ as the projection onto the column space of matrix A. Next we prove the upper bound of two terms separately. Observe that $P_{\boldsymbol{\beta}^*_{1,2}}\widehat{\boldsymbol{\beta}}_{1,2} = \boldsymbol{\beta}^*_{1,2}(\boldsymbol{\beta}^{*\top}_{1,2}\widehat{\boldsymbol{\beta}}_{1,2})$ and $|\boldsymbol{\beta}^*_{1,2}\widehat{\boldsymbol{\beta}}_{1,2}| \le 1$. Therefore, to bound the first term in (S11), it suffices to bound

$$\left\|\frac{1}{pn}\sum_i \alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_2 \boldsymbol{\beta}^*_{1,2}\right\|.$$

We write

$$\frac{1}{pn}\sum_{i=1}^n \alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_2 \boldsymbol{\beta}^*_{1,2} = \frac{1}{pn}\sum_{i\in[n],j\in F_1,k\in F_2,l\in F_3} \alpha^*_{i,1}\delta_{i,j,k,l}\mathcal{E}_{i,j,k,l}\boldsymbol{\beta}^*_{1,2,k}\mathbf{e}_{j,l},$$

where $\mathbf{e}_{j,l}$ is a $\mathbb{R}^{d_1\times d_3}$ matrix with all zero entries except that the (j,l)-th entry is 1. Note that

$$\left\|\sum_{i\in[n],j\in F_1,k\in F_2,l\in F_3}\mathbb{E}(\delta_{i,j,k,l}\mathcal{E}^2_{i,j,k,l})\alpha^{*2}_{i,1}\boldsymbol{\beta}^{*2}_{1,2,k}\mathbf{e}_{j,l}\mathbf{e}_{j,l}^\top\right\|\le nc_1^2sp\sigma^2,$$

$$\left\|\sum_{i\in[n],j\in F_1,k\in F_2,l\in F_3}\mathbb{E}(\delta_{i,j,k,l}\mathcal{E}^2_{i,j,k,l})\alpha^{*2}_{i,1}\boldsymbol{\beta}^{*2}_{1,2,k}\mathbf{e}_{j,l}^\top\mathbf{e}_{j,l}\right\|\le nc_1^2sp\sigma^2,$$

and

$$\left\|\left\|\alpha^*_{i,1}\delta_{i,j,k,l}\mathcal{E}_{i,j,k,l}\boldsymbol{\beta}^*_{1,2,k}\mathbf{e}_{j,l}\right\|\right\|_{\psi_2}\le\|\mathcal{E}_{i,j,k,l}\|_{\psi_2}\left\|\alpha^*_{i,1}\boldsymbol{\beta}^*_{1,2,k}\mathbf{e}_{j,l}\right\|\le c_1\sigma\frac{\mu}{\sqrt{s}},$$

where we used the $\mu-$mass assumption and Assumption 1 (i) i.e. $\|\mathbf{x}_i\|\le c_1$. By the matrix Bernstein inequality (Lemma 4), the following bound holds with probability at least $1-d^{-10}$,

$$\left\|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\times_2\boldsymbol{\beta}^*_{1,2}\right\|\le\widetilde{C}_1\max\left\{\sigma\sqrt{\frac{s\log(d)}{np}},\frac{\sigma\log(d)}{np\sqrt{s}}\log\left(\sqrt{\frac{s}{p}}\right)\right\},\qquad\text{(S12)}$$

for some large enough constant $\widetilde{C}_1$.

Next, we bound the second term in (S11). Note that

$$\begin{aligned}\left\|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\times_2 P^{\perp}_{\boldsymbol{\beta}^*_{1,2}}\widehat{\boldsymbol{\beta}}_{1,2}\right\| &\le \left\|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\right\|\left\|(\boldsymbol{\beta}^{*\perp}_{1,2})^\top\widehat{\boldsymbol{\beta}}_{1,2}\right\|\\ &\le \epsilon\left\|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\right\|,\end{aligned}$$

where we used the fact $\left\|(\boldsymbol{\beta}^{*\perp}_{1,2})^\top\widehat{\boldsymbol{\beta}}_{1,2}\right\|=\left\|(\boldsymbol{\beta}^{*\perp}_{1,2})^\top(\boldsymbol{\beta}^*_{1,2}-\widehat{\boldsymbol{\beta}}_{1,2})\right\|\le\epsilon$. It suffices to bound

$$\left\|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\right\|=\sup_{\boldsymbol{u}_1\in S_1,\boldsymbol{u}_2\in S_2,\boldsymbol{u}_3\in S_3}\left|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\times_1\boldsymbol{u}_1\times_2\boldsymbol{u}_2\times_3\boldsymbol{u}_3\right|,$$

where $S_i=\{\boldsymbol{u}\in\mathbb{R}^{d_i}:\|\boldsymbol{u}\|=1,\|\boldsymbol{u}\|_0\le s_i\}$, for $i=1,2,3$ is the sparse unit sphere in $\mathbb{R}^{d_i}$. Similar to the proof of Theorem 1 of Ryota and Taiji (2014), we use a covering number argument. For each given subset $U_i\subseteq[d_i]$, we define the set $S_{U_i}:=\{\boldsymbol{v}\in\mathbb{R}^{d_i}:\|\boldsymbol{v}\|=1,\text{supp}(\boldsymbol{v})\subseteq U_i\}$. Let C_1,C_2,C_3 be $\widetilde{\epsilon}$-covers of S_{U_1},S_{U_2},S_{U_3}. Let $\widetilde{\epsilon}=\log(3/2)/3$ and Theorem 1 of Ryota and Taiji (2014) has proved that

$$\mathbb{P}\left(\left\|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\right\|\ge\bar{t}\right)\le\sum_{\bar{\boldsymbol{u}}_j\in C_j}\mathbb{P}\left(\left|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\times_1\bar{\boldsymbol{u}}_1\times_2\bar{\boldsymbol{u}}_2\times_3\bar{\boldsymbol{u}}_3\right|\ge\bar{t}/2\right).$$

Next we bound each term in the summation. For each fixed $\bar{\boldsymbol{u}}_1$, $\bar{\boldsymbol{u}}_2, \bar{\boldsymbol{u}}_3$, we write

$$\frac{1}{pn}\sum_{i=1}^{n}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\times_1\bar{\boldsymbol{u}}_1\times_2\bar{\boldsymbol{u}}_2\times_3\bar{\boldsymbol{u}}_3=\frac{1}{pn}\sum_{i\in[n],j\in\bar{F}_1,k\in\bar{F}_2,l\in\bar{F}_3}\alpha^*_{i,1}\delta_{i,j,k,l}\mathcal{E}_{i,j,k,l}\bar{\boldsymbol{u}}_{1,j}\bar{\boldsymbol{u}}_{2,k}\bar{\boldsymbol{u}}_{3,l},$$

where $\bar{F}_j=\text{supp}(\bar{\boldsymbol{u}}_j)$. It is easy to check the following bounds

$$\sum_{i\in[n],j\in\bar{F}_1,k\in\bar{F}_2,l\in\bar{F}_3}\mathbb{E}(\delta_{i,j,k,l}\mathcal{E}^2_{i,j,k,l})\alpha^{*2}_{i,1}\bar{\boldsymbol{u}}^2_{1,j}\bar{\boldsymbol{u}}^2_{2,k}\bar{\boldsymbol{u}}^2_{3,l}\le c_1^2np\sigma^2,$$

and

$$\left\|\alpha^*_{i,1}\delta_{i,j,k,l}\mathcal{E}_{i,j,k,l}\bar{\boldsymbol{u}}_{1,j}\bar{\boldsymbol{u}}_{2,k}\bar{\boldsymbol{u}}_{3,l}\right\|_{\psi_2}\le c_1\|\mathcal{E}_{i,j,k,l}\|_{\psi_2}\le c_1\sigma.$$

By Lemma 4, we have

$$\mathbb{P}\left(\left|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\times_1\bar{\boldsymbol{u}}_1\times_2\bar{\boldsymbol{u}}_2\times_3\bar{\boldsymbol{u}}_3\right|\ge\widetilde{C}_2\max\left\{\frac{\sigma\sqrt{t}}{\sqrt{np}},\frac{\sigma t}{np}\log\left(\sqrt{\frac{s_1s_2s_3}{p}}\right)\right\}\right)\le e^{-t}.$$

Let $\bar{t}=\widetilde{C}_2\max\left\{\frac{\sigma\sqrt{t}}{\sqrt{np}},\frac{\sigma t}{np}\log\left(\sqrt{\frac{s_1s_2s_3}{p}}\right)\right\}$. Then we have,

$$\mathbb{P}\left(\left\|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\right\|\ge\bar{t}\right)\le\sum_{\bar{\boldsymbol{u}}_1\in C_1,\bar{\boldsymbol{u}}_2\in C_2,\bar{\boldsymbol{u}}_3\in C_3}e^{-t}\le\left\{\frac{6}{\log(3/2)}\right\}^{s_1+s_2+s_3}\cdot 2e^{-t}.$$

Taking a union bound over $\binom{d_1}{s_1}\binom{d_2}{s_2}\binom{d_3}{s_3}\le d^{s_1+s_2+s_3}$ choices of $U_1\circ U_2\circ U_3$ yields,

$$\mathbb{P}\left(\left\|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\right\|\ge\widetilde{C}_2\max\left\{\frac{\sigma\sqrt{t}}{\sqrt{np}},\frac{\sigma t}{np}\log\left(\sqrt{\frac{s_1s_2s_3}{p}}\right)\right\}\right)\le 2e^{-t+\{s_1+s_2+s_3\}\log(d)}.$$

Let $t=13s\log(d)$. Then the following bound holds with probability at least $1-2d^{-10s}$,

$$\left\|\sum_i\frac{1}{pn}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\right\|\le\widetilde{C}_2\max\left\{\sigma\sqrt{\frac{s\log(d)}{np}},\frac{\sigma s\log(d)}{np}\log\left(\sqrt{\frac{s_1s_2s_3}{p}}\right)\right\}.\tag{S13}$$

We now construct the upper bound for III_{12}. Combining (S11), (S12), and (S13), the following bound holds with probability at least $1-3d^{-10}$,

$$\|III_{12}\|\le\widetilde{C}_1\sigma\sqrt{\frac{s\log(d)}{np}}\epsilon+\widetilde{C}_1\frac{\sigma\log(d)}{np\sqrt{s}}\log\left(\sqrt{\frac{s}{p}}\right)\epsilon+\widetilde{C}_2\sigma\sqrt{\frac{s\log(d)}{np}}\epsilon^2+\widetilde{C}_2\frac{\sigma s\log(d)}{np}\log\left(\sqrt{\frac{s^3}{p}}\right)\epsilon^2.\tag{S14}$$

Bounding III_{13}: It is easy to check that

$$\|III_{13}\|\le\left\|\sum_i\frac{1}{np}\alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF})\times_1\boldsymbol{\beta}^*_{1,1}\right\|\epsilon\le\widetilde{C}_1\sigma\sqrt{\frac{s\log(d)}{np}}\epsilon+\widetilde{C}_1\frac{\sigma\log(d)}{np\sqrt{s}}\log\left(\sqrt{\frac{s}{p}}\right)\epsilon,\tag{S15}$$

where the derivation of last inequality is similar to (S12).

Bounding III_{14}: We write

$$\frac{1}{np}\sum_i \alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \boldsymbol{\beta}^*_{1,1} \times_2 \boldsymbol{\beta}^*_{1,2} = \frac{1}{np}\sum_{i\in[n],j\in F_1,k\in F_2,l\in F_3} \alpha^*_{i,1}\delta_{i,j,k,l}\mathcal{E}_{i,j,k,l}\boldsymbol{\beta}^*_{1,1,j}\boldsymbol{\beta}^*_{1,2,k}\mathbf{e}_l,$$

where $\mathbf{e}_l$ is the vector with all zero entries except that the l-th entry is 1. Clearly, we have

$$\left\|\sum_{i\in[n],j\in F_1,k\in F_2,l\in F_3} \mathbb{E}(\delta_{i,j,k,l}\mathcal{E}^2_{i,j,k,l})\alpha^{*2}_{i,1}\boldsymbol{\beta}^{*2}_{1,1,j}\boldsymbol{\beta}^{*2}_{1,2,k}\mathbf{e}_l^\top\mathbf{e}_l\right\| \le c_1^2 np\sigma^2 s$$
$$\left\|\sum_{i\in[n],j\in F_1,k\in F_2,l\in F_3} \mathbb{E}(\delta_{i,j,k,l}\mathcal{E}^2_{i,j,k,l})\alpha^{*2}_{i,1}\boldsymbol{\beta}^{*2}_{1,1,j}\boldsymbol{\beta}^{*2}_{1,2,k}\mathbf{e}_l\mathbf{e}_l^\top\right\| \le c_1^2 np\sigma^2 s,$$

and

$$\left\| \|\alpha^*_{i,1}\mathcal{E}_{i,j,k,l}\boldsymbol{\beta}^*_{1,1,j}\boldsymbol{\beta}^*_{1,2,k}\mathbf{e}_l\| \right\|_{\psi_2} \le \frac{c_1\mu^2\sigma}{s},$$

where the inequality holds due to Assumption 1 (i) and μ-mass assumption. By Lemma 4, with probability at least $1-d^{-10}$ we have,

$$\|III_{14}\| = \left\|\frac{1}{np}\sum_i \alpha^*_{i,1}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \boldsymbol{\beta}^*_{1,1} \times_2 \boldsymbol{\beta}^*_{1,2}\right\| \le \widetilde{C}_3 \max\left\{\sigma\sqrt{\frac{s\log(d)}{np}}, \frac{\sigma\log(d)}{nps}\log\left(\frac{1}{\sqrt{p}}\right)\right\}. \tag{S16}$$

Bounding III_{11}: It easy to check that

$$\|III_{11}\| \le \left\|\sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_{1,1} \times_2 \widehat{\boldsymbol{\beta}}_{1,2}\mathbf{x}_i^\top\right\| \|\widehat{\boldsymbol{\beta}}_{1,4} - \boldsymbol{\beta}^*_{1,4}\|.$$

We write

$$\begin{aligned}\sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_{1,1} \times_2 \widehat{\boldsymbol{\beta}}_{1,2}\mathbf{x}_i^\top &= \sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 (P_{\boldsymbol{\beta}^*_{1,1}} + P^\perp_{\boldsymbol{\beta}^*_{1,1}})\widehat{\boldsymbol{\beta}}_{1,1} \times_2 (P_{\boldsymbol{\beta}^*_{1,2}} + P^\perp_{\boldsymbol{\beta}^*_{1,2}})\widehat{\boldsymbol{\beta}}_{1,2}\mathbf{x}_i^\top \\ &= \sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 P_{\boldsymbol{\beta}^*_{1,1}}\widehat{\boldsymbol{\beta}}_{1,1} \times_2 P_{\boldsymbol{\beta}^*_{1,2}}\widehat{\boldsymbol{\beta}}_{1,2} + \widehat{D}. \end{aligned} \tag{S17}$$

We next bound the two terms separately. Observe that $\|\boldsymbol{\beta}^{*\top}_{1,1}\widehat{\boldsymbol{\beta}}_{1,1}\| \le 1$. Therefore, it suffices to prove the upper bound of

$$\sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \boldsymbol{\beta}^*_{1,1} \times_2 \boldsymbol{\beta}^*_{1,2}\mathbf{x}_i^\top.$$

We write

$$\sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF} \times_1 \boldsymbol{\beta}^*_{1,1} \times_2 \boldsymbol{\beta}^*_{1,2}\mathbf{x}_i^\top) = \sum_{i\in[n],j\in F_1,k\in F_2,l\in F_3} \frac{1}{pn}\delta_{i,j,k,l}\mathcal{E}_{i,j,k,l}\boldsymbol{\beta}^*_{1,1,j}\boldsymbol{\beta}^*_{1,2,k}\mathbf{e}_l\mathbf{x}_i^\top.$$

Observe that

$$\left\|\sum_{i\in[n],j\in F_1,k\in F_2,l\in F_3} \mathbb{E}(\delta_{i,j,k,l}\mathcal{E}^2_{i,j,k,l})\boldsymbol{\beta}^{*2}_{1,1,j}\boldsymbol{\beta}^{*2}_{1,2,k}\mathbf{e}_l\mathbf{x}_i^\top\mathbf{x}_i\mathbf{e}_l^\top\right\| \le p\sigma^2\left\|\sum_{i\in[n],l\in F_3}\mathbf{e}_l\mathbf{x}_i^\top\mathbf{x}_i\mathbf{e}_l^\top\right\| \le c_1^2 p\sigma^2 n,$$
$$\left\|\sum_{i\in[n],j\in F_1,k\in F_2,l\in F_3} \mathbb{E}(\delta_{i,j,k,l}\mathcal{E}^2_{i,j,k,l})\boldsymbol{\beta}^{*2}_{1,1,j}\boldsymbol{\beta}^{*2}_{1,2,k}\mathbf{x}_i\mathbf{e}_l^\top\mathbf{e}_l\mathbf{x}_i^\top\right\| \le p\sigma^2 s\left\|\sum_{i\in[n]}\mathbf{x}_i\mathbf{x}_i^\top\right\| \le c_1^2 ps\sigma^2 n,$$

and

$$\left\|\|\delta_{i,j,k,l}\mathcal{E}_{i,j,k,l}\boldsymbol{\beta}^*_{1,1,j}\boldsymbol{\beta}^*_{1,2,k}\mathbf{e}_l\mathbf{x}_i^\top\|\right\|_{\psi_2} \le \frac{c_1\mu^2\sigma}{s}.$$

By Lemma 4, with probability at least $1-d^{-10}$ we have,

$$\left\|\frac{1}{pn}\sum_i \Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \boldsymbol{\beta}^*_{1,1} \times_2 \boldsymbol{\beta}^*_{1,2}\mathbf{x}_i^\top\right\| \le \widetilde{C}_3 \max\left\{\sigma\sqrt{\frac{s\log(d)}{np}}, \frac{\sigma\log(d)}{nps}\log\left(\sqrt{\frac{1}{p}}\right)\right\}.$$

Next we prove the upper bound of $\widehat{D}$ in (S17), where it can be explicitly expressed as

$$\begin{aligned}\widehat{D} &= \sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 P_{\boldsymbol{\beta}^*_{1,1}}\widehat{\boldsymbol{\beta}}_{1,1} \times_2 P^\perp_{\boldsymbol{\beta}^*_{1,2}}\widehat{\boldsymbol{\beta}}_{1,2}\mathbf{x}_i^\top + \sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 P^\perp_{\boldsymbol{\beta}^*_{1,1}}\widehat{\boldsymbol{\beta}}_{1,1} \times_2 P_{\boldsymbol{\beta}^*_{1,2}}\widehat{\boldsymbol{\beta}}_{1,2}\mathbf{x}_i^\top \\ &\quad + \sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 P^\perp_{\boldsymbol{\beta}^*_{1,1}}\widehat{\boldsymbol{\beta}}_{1,1} \times_2 P^\perp_{\boldsymbol{\beta}^*_{1,2}}\widehat{\boldsymbol{\beta}}_{1,2}\mathbf{x}_i^\top.\end{aligned}$$

We know that

$$\begin{aligned}&\left\|\sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 P_{\boldsymbol{\beta}^*_{1,1}}\widehat{\boldsymbol{\beta}}_{1,1} \times_2 P^\perp_{\boldsymbol{\beta}^*_{1,2}}\widehat{\boldsymbol{\beta}}_{1,2}\mathbf{x}_i^\top\right\| \le c_1\left\|\sum_i \frac{1}{pn}\Pi_{\Omega_i}(\mathcal{E}_{iF})\right\| \|(\boldsymbol{\beta}^{*\perp}_{1,2})^\top\widehat{\boldsymbol{\beta}}_{1,2}\| \\ &\le \widetilde{C}_2\sigma\sqrt{\frac{s\log(d)}{np}}\epsilon + \widetilde{C}_2\frac{\sigma s\log(d)}{np}\log\left(\sqrt{\frac{s_1s_2s_3}{p}}\right)\epsilon,\end{aligned}$$

where the last inequality used the bound in (S13) and the aforementioned fact that $\|(\boldsymbol{\beta}^{*\perp}_{1,2})^\top\widehat{\boldsymbol{\beta}}_{1,2}\| \le \epsilon$. The other terms in the expression of $\widehat{D}$ can be bounded similarly. Therefore, with probability at least $1-d^{-10}$ we have

$$\|\widehat{D}\| \le \widetilde{C}_2\sigma\sqrt{\frac{s\log(d)}{np}}\epsilon + \widetilde{C}_2\frac{\sigma s\log(d)}{np}\log\left(\sqrt{\frac{s_1s_2s_3}{p}}\right)\epsilon.$$

Now we conclude that with probability at least $1 - 2/d^{-10}$, we have that

$$\begin{aligned}\|III_{11}\| \ \le\ & \widetilde{C}_3\sigma\sqrt{\frac{s\log(d)}{np}}\epsilon + \widetilde{C}_3\frac{\sigma\log(d)}{nps}\log\left(\sqrt{\frac{1}{p}}\right)\epsilon + \widetilde{C}_2\sigma\sqrt{\frac{s\log(d)}{np}}\epsilon^2 \\ & + \widetilde{C}_2\frac{\sigma s\log(d)}{np}\log\left(\sqrt{\frac{s_1s_2s_3}{p}}\right)\epsilon^2. \end{aligned} \tag{S18}$$

Finally, plug the bounds in (S18), (S14), (S15), and (S16) into (S10), we bound $\|III_1\|$ as

$$\begin{aligned}\|III_1\| \ \le\ & \frac{2\{2\widetilde{C}_1 + \widetilde{C}_3\}\sigma\epsilon}{w_1^*\{\lambda_{\min} - \gamma\}}\left(\sqrt{\frac{s\log(d)}{np}} + \frac{\log(d)}{np\sqrt{s}}\log\left(\sqrt{\frac{s}{p}}\right) + \sqrt{\frac{s\log(d)}{np}} + \frac{\log(d)}{nps}\log\left(\frac{1}{\sqrt{p}}\right)\right) \\ & + \frac{4\widetilde{C}_2\sigma\epsilon^2}{w_1^*\{\lambda_{\min} - \gamma\}}\left(\sqrt{\frac{s\log(d)}{np}} + \frac{s\log(d)}{np}\log\left(\sqrt{\frac{s^3}{p}}\right)\right) + \frac{2\widetilde{C}_3\sigma}{w_1^*\{\lambda_{\min} - \gamma\}}\sqrt{\frac{s\log(d)}{np}} \end{aligned} \tag{S19}$$

By (S4) it is obvious that $\gamma < \lambda_{\min}/2$ and $1/\{\lambda_{\min} - \gamma\} < 2\lambda_{\min}$. By Assumption 5, it is easy to check that

$$\frac{4\{2\widetilde{C}_1 + \widetilde{C}_3\}\sigma\epsilon}{\lambda_{\min}w_1^*}\sqrt{\frac{s\log(d)}{np}} \le \gamma'\epsilon,$$

for some positive constant γ'. The value of constant γ' will be determined later. Similarly, we have the following bounds,

$$\begin{aligned}&\frac{4\{2\widetilde{C}_1 + \widetilde{C}_3\}\sigma\epsilon}{\lambda_{\min}w_1^*}\max\left\{\frac{\log(d)}{np\sqrt{s}}\log\left(\sqrt{\frac{s}{p}}\right), \frac{\log(d)}{nps}\log\left(\sqrt{\frac{1}{p}}\right)\right\} \le \gamma'\epsilon, \\ &\frac{8\widetilde{C}_2\sigma\epsilon^2}{\lambda_{\min}w_1^*}\max\left\{\sqrt{\frac{s\log(d)}{np}}, \frac{s\log(d)}{np}\log\left(\sqrt{\frac{s^3}{p}}\right)\right\} \le \gamma'\epsilon.\end{aligned}$$

Therefore the bound in (S19) can be simplified as

$$\|III_1\| \le \frac{\widetilde{C}\sigma}{w_1^*\{\lambda_{\min} - \gamma\}}\sqrt{\frac{s\log(d)}{np}} + 6\gamma'\epsilon, \tag{S20}$$

for some large enough $\widetilde{C} > 0$.

Combining the bounds of I_1, II_1, III_1, with probability at least $1 - 1/d^9$, we have,

$$\left\|\widetilde{\boldsymbol{\beta}}_{1,3} - \frac{w_1^*\frac{1}{n}\sum_{i=1}^n\widehat{\alpha}_{i,1}\alpha_{i,1}^*}{\widehat{w}_1\frac{1}{n}\sum_{i=1}^n\widehat{\alpha}_{i,1}^2}\boldsymbol{\beta}_{1,3}^*\right\| \le \left\{\frac{2\lambda_{\max}}{\lambda_{\min}}\epsilon + \frac{2\gamma}{\lambda_{\min} - \gamma} + 6\gamma'\right\}\epsilon + \frac{\widetilde{C}\sigma}{w_1^*\{\lambda_{\min} - \gamma\}}\sqrt{\frac{s\log(d)}{np}}.$$

Finally, we bound the distance between the normalized $\widetilde{\boldsymbol{\beta}}_{1,3}$ and the true parameter $\beta_{1,3}^*$. For the normalized vectors $\widetilde{\boldsymbol{\beta}}_{1,3}/\|\widetilde{\boldsymbol{\beta}}_{1,3}\|$ and $\boldsymbol{\beta}_{1,3}^*$ we have

$$\left|\|\widetilde{\boldsymbol{\beta}}_{1,3}\| - \frac{w_1^*\frac{1}{n}\sum_{i=1}^n\widehat{\alpha}_{i,1}\alpha_{i,1}^*}{\widehat{w}_1\frac{1}{n}\sum_{i=1}^n\widehat{\alpha}_{i,1}^2}\right| \le \left\|\|\widetilde{\boldsymbol{\beta}}_{1,3}\|\frac{\widetilde{\boldsymbol{\beta}}_{1,3}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \frac{w_1^*\frac{1}{n}\sum_{i=1}^n\widehat{\alpha}_{i,1}\alpha_{i,1}^*}{\widehat{w}_1\frac{1}{n}\sum_{i=1}^n\widehat{\alpha}_{i,1}^2}\boldsymbol{\beta}_{1,3}^*\right\|.$$

Next, it is clear that

$$\left|\frac{w_1^* \frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha_{i,1}^*}{\widehat{w}_1 \frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,1}^2}\right| \left\|\frac{\widetilde{\boldsymbol{\beta}}_{1,3}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}_{1,3}^*\right\|$$
$$\leq \left|\|\widetilde{\boldsymbol{\beta}}_{1,3}\| - \frac{w_1^* \frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha_{i,1}^*}{\widehat{w}_1 \frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,1}^2}\right| + \left\|\widetilde{\boldsymbol{\beta}}_{1,3} - \frac{w_1^* \frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha_{i,1}^*}{\widehat{w}_1 \frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,1}^2}\boldsymbol{\beta}_{1,3}^*\right\|.$$

Therefore, the following bound holds with probability at least $1 - 1/d^9$

$$\left\|\frac{\widetilde{\boldsymbol{\beta}}_{1,3}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}_{1,3}^*\right\| \leq \frac{3\lambda_{\max}}{\lambda_{\min}}\left\|\widetilde{\boldsymbol{\beta}}_{1,3} - \frac{w_1^* \frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,1}\alpha_{i,1}^*}{\widehat{w}_1 \frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,1}^2}\boldsymbol{\beta}_{1,3}^*\right\|$$
$$\leq \frac{6\lambda_{\max}}{\lambda_{\min}}\left\{\frac{\lambda_{\max}}{\lambda_{\min}}\epsilon + \frac{\gamma}{\lambda_{\min} - \gamma} + 3\gamma'\right\}\epsilon + \frac{3\lambda_{\max}\widetilde{C}\sigma}{w_1^*\lambda_{\min}\{\lambda_{\min} - \gamma\}}\sqrt{\frac{s\log(d)}{np}}. \qquad \text{(S21)}$$

S1.2 Error bound of the estimator from Step 2 of Algorithm 1

In the second step, we derive the error bound for the constrained estimator from Step 2 of Algorithm 1. After obtaining normalized $\widetilde{\boldsymbol{\beta}}_{1,3}/\|\widetilde{\boldsymbol{\beta}}_{1,3}\|$ from Step 1, we apply the `Truncatefuse` operator to $\widetilde{\boldsymbol{\beta}}_{1,3}/\|\widetilde{\boldsymbol{\beta}}_{1,3}\|$ to obtain the sparse and fused estimator, then normalize it to get a unit estimator $\widehat{\boldsymbol{\beta}}_{1,3}$.

We first obtain a result for the estimator after applying the fusion operator to $\widetilde{\boldsymbol{\beta}}_{1,3}/\|\widetilde{\boldsymbol{\beta}}_{1,3}\|$,

$$\widehat{\boldsymbol{\beta}}_{1,3}^f = \arg\min_{\boldsymbol{\beta}} \left\|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}_{1,3}/\|\widetilde{\boldsymbol{\beta}}_{1,3}\|\right\| \quad \text{such that } \|D\boldsymbol{\beta}\|_0 \leq \tau_{f_3},$$

where τ_{f_3} is the fusion parameter in our Algorithm 1, and f_3 is the true fusion parameter. Then by Lemma 8, we have that, $\|\widehat{\boldsymbol{\beta}}_{1,3}^f - \boldsymbol{\beta}_{1,3}^*\| \leq \left\|\widetilde{\boldsymbol{\beta}}_{1,3}/\|\widetilde{\boldsymbol{\beta}}_{1,3}\| - \boldsymbol{\beta}_{1,3}^*\right\|$. In other words, if the true parameter has a fusion structure, then adding the fusion step in our algorithm is guaranteed to reduce the estimation error.

Next, we apply the `Truncate` operator to $\widehat{\boldsymbol{\beta}}_{1,3}^f$. By Lemma 1, we have

$$\left|\texttt{Truncate}\left(\widehat{\boldsymbol{\beta}}_{1,3}^f, \tau_{s_3}\right)^\top \boldsymbol{\beta}_{1,3}^*\right| \geq \left|\widehat{\boldsymbol{\beta}}_{1,3}^{f\top}\boldsymbol{\beta}_{1,3}^*\right| - \sqrt{\frac{s_3}{\tau_{s_3}}}\left(1 + \sqrt{\frac{s_3}{\tau_{s_3}}}\right)\left\{1 - \left(\widehat{\boldsymbol{\beta}}_{1,3}^{f\top}\boldsymbol{\beta}_{1,3}^*\right)^2\right\},$$

where the right-hand-side is an increasing function in terms of $|\widehat{\boldsymbol{\beta}}_{1,3}^{f\top}\boldsymbol{\beta}_{1,3}^*|$, when it is in $[0, 1]$ and $\tau_{s_3} \geq s_3$. The estimator from Step 2 of our algorithm has the form $\widehat{\boldsymbol{\beta}}_{1,3} = \texttt{Truncate}(\widehat{\boldsymbol{\beta}}_{1,3}^f, \tau_{s_3})/\left\|\texttt{Truncate}(\widehat{\boldsymbol{\beta}}_{1,3}^f, \tau s_3)\right\|$. Note that $\left\|\texttt{Truncate}(\widehat{\boldsymbol{\beta}}_{1,3}^f, \tau_{s_3})\right\| \leq 1$ due to the facts that $\|\widehat{\boldsymbol{\beta}}^f_{1,3}\| = 1$ and `Truncate` operator sets some entries in $\widehat{\boldsymbol{\beta}}^f_{1,3}$ to 0. Therefore,

$$
\begin{aligned}
\|\widehat{\boldsymbol{\beta}}_{1,3} - \boldsymbol{\beta}^*_{1,3}\| &\leq \sqrt{2}\sqrt{1 - (\widehat{\boldsymbol{\beta}}^\top_{1,3}\boldsymbol{\beta}^*_{1,3})^2} \leq \sqrt{2}\sqrt{1 - \left\{\texttt{Truncate}(\widehat{\boldsymbol{\beta}}^f_{1,3}, \tau_{s_3})^\top \boldsymbol{\beta}^*_{1,3}\right\}^2} \\
&\leq \sqrt{2}\left\{1 + 2\sqrt{\frac{s_3}{\tau_{s_3}}}\left(1 + \sqrt{\frac{s_3}{\tau_{s_3}}}\right)\right\}^{1/2}\sqrt{1 - (\widehat{\boldsymbol{\beta}}^{f\top}_{1,3}\boldsymbol{\beta}^*_{1,3})^2} \\
&\leq \sqrt{10}\|\widehat{\boldsymbol{\beta}}^f_{1,3} - \boldsymbol{\beta}^*_{1,3}\| \leq \sqrt{10}\left\|\widetilde{\boldsymbol{\beta}}_{1,3}/\|\widetilde{\boldsymbol{\beta}}_{1,3}\| - \boldsymbol{\beta}^*_{1,3}\right\|. \qquad \text{(S22)}
\end{aligned}
$$

Next, we establish the error bound for the constrained estimator $\widehat{\boldsymbol{\beta}}_{1,3}$ from Step 2 of Algorithm 1. Combining (S21) and (S22), with probability at least $1 - 1/d^9$, we have,

$$
\begin{aligned}
\left\|\widehat{\boldsymbol{\beta}}_{1,3} - \boldsymbol{\beta}^*_{1,3}\right\| &\leq \frac{6\sqrt{10}\lambda_{\max}}{\lambda_{\min}}\left\{\frac{\lambda_{\max}}{\lambda_{\min}}\epsilon + \frac{\gamma}{\lambda_{\min} - \gamma} + 3\gamma'\right\}\epsilon + \frac{3\sqrt{10}\lambda_{\max}\widetilde{C}\sigma}{w^*_1\lambda_{\min}\{\lambda_{\min} - \gamma\}}\sqrt{\frac{s\log(d)}{np}} \\
&\leq \frac{6\sqrt{10}\lambda_{\max}}{\lambda_{\min}}\left\{\frac{\lambda_{\max}}{\lambda_{\min}}\epsilon + \frac{2\gamma}{\lambda_{\min}} + 3\gamma'\right\}\epsilon + \frac{6\sqrt{10}\lambda_{\max}\widetilde{C}\sigma}{w^*_1\lambda^2_{\min}}\sqrt{\frac{s\log(d)}{np}}, \qquad \text{(S23)}
\end{aligned}
$$

where the second inequality is due to the fact that constant $\gamma < \lambda_{\min}/2$ based on (S4).

Finally, we prove that, if the true parameter $\boldsymbol{\beta}^*_{1,3}$ is a μ-mass vector, then the estimator $\widehat{\boldsymbol{\beta}}_{1,3}$ after each iteration is also a $\widetilde{c}\mu$-mass vector. Note that the μ-mass of the update does not increase beyond a global constant. By (S1), each entry of $\widehat{\boldsymbol{\beta}}_{1,3}$ can be bounded as,

$$
\begin{aligned}
|\widetilde{\boldsymbol{\beta}}_{1,3,l}| &\leq \frac{|w^*_1\boldsymbol{B}_{ll}|}{|\widehat{w}_1\boldsymbol{A}_{ll}|}\frac{\mu}{\sqrt{s}} + \frac{\left|\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,1}\sum_{l_1,l_2}\delta_{i,l_1,l_2,l}(\mathcal{E}_{iF})_{l_1,l_2,l}\widehat{\boldsymbol{\beta}}_{1,1,l_1}\widehat{\boldsymbol{\beta}}_{1,2,l_2}\right|}{|\widehat{w}_1\boldsymbol{A}_{ll}|} \\
&\leq 2\frac{\frac{1}{n}\sum_i \widehat{\alpha}_{i,1}\alpha^*_{i,1} + \gamma}{\frac{1}{n}\sum_i \widehat{\alpha}^2_{i,1} - \gamma}\frac{\mu}{\sqrt{s}} + \left\|\frac{1}{\widehat{w}_1}\boldsymbol{A}^{-1}\sum_{i=1}^n \frac{\widehat{\alpha}_{i,1}}{n}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_1 \times_2 \widehat{\boldsymbol{\beta}}_2\right\| \\
&\leq 2\frac{\lambda_{\max} + \gamma}{\lambda_{\min} - \gamma}\frac{\mu}{\sqrt{s}} + \frac{\widetilde{C}\sigma}{w^*_1\{\lambda_{\min} - \gamma\}}\sqrt{\frac{s\log(d)}{np}},
\end{aligned}
$$

where the last inequality is from the bound for III_1. By (S4), $\gamma < \lambda_{\min}/2$, and then we have

$$
2\frac{\lambda_{\max} + \gamma}{\lambda_{\min} - \gamma}\frac{\mu}{\sqrt{s}} \leq \frac{6\lambda_{\max}}{\lambda_{\min}}\frac{\mu}{\sqrt{s}}.
$$

And by Assumption 5, $np \geq \widetilde{C}^2\sigma^2 s^2\log(d)/\{w^{*2}\lambda^2_{\min}\}$, then we have

$$
\frac{\widetilde{C}\sigma}{w^*_1\{\lambda_{\min} - \gamma\}}\sqrt{\frac{s\log(d)}{np}} \leq \frac{\mu}{\sqrt{s}}.
$$

Therefore there is some global constant $\widetilde{c} > 0$, and we have

$$
\max_l\left\{|\widehat{\boldsymbol{\beta}}_{k,3,l}|\right\} \leq \widetilde{c}\frac{\mu}{\sqrt{s}}.
$$

Given that the true parameter $\boldsymbol{\beta}^*_{k,3}$ is a μ-mass vector, the update from each iteration is a $\widetilde{c}\mu$-mass vector, where $\widetilde{c}$ is a global constant.

S1.3 Error bound of the estimator from Step 4 of Algorithm 1

In the third step, we derive the error bound for the estimator $\widehat{\boldsymbol{\beta}}_{1,4}$ from Step 4 of Algorithm 1. That is, we aim to bound $\|\widehat{\boldsymbol{\beta}}_{1,4} - \boldsymbol{\beta}^*_{1,4}\|$, given the other estimators $\widehat{w}_1, \widehat{\boldsymbol{\beta}}_{1,1}, \widehat{\boldsymbol{\beta}}_{1,2}, \widehat{\boldsymbol{\beta}}_{1,3}$.

Denote $\widehat{\mathcal{A}}_1 = \widehat{w}_1 \widehat{\boldsymbol{\beta}}_{1,1} \circ \widehat{\boldsymbol{\beta}}_{1,2} \circ \widehat{\boldsymbol{\beta}}_{1,3}$, and $\mathcal{A}^*_1 = w^*_1 \boldsymbol{\beta}^*_{1,1} \circ \boldsymbol{\beta}^*_{1,2} \circ \boldsymbol{\beta}^*_{1,3}$. When the rank $r = 1$, the true model is $\mathcal{Y}_i = \boldsymbol{\beta}^{*\top}_{1,4} \mathbf{x}_i \mathcal{A}^*_1 + \mathcal{E}_i$. Then the closed-form solution of $\widehat{\boldsymbol{\beta}}_{1,4}$ in (9) becomes

$$\widehat{\boldsymbol{\beta}}_{1,4} = \left\{ \frac{1}{n} \sum_{i=1}^n \|\Pi_{\Omega_i}(\widehat{\mathcal{A}}_1)\|_F^2 \mathbf{x}_i \mathbf{x}_i^\top \right\}^{-1} \frac{1}{n} \sum_{i=1}^n \left\langle \Pi_{\Omega_i}(\boldsymbol{\beta}^{*\top}_{1,4} \mathbf{x}_i \mathcal{A}^*_1 + \mathcal{E}_i), \Pi_{\Omega_i}(\widehat{\mathcal{A}}_1) \right\rangle \mathbf{x}_i.$$

By the fact that $\|\Pi_{\Omega_i}(\widehat{\mathcal{A}}_1)\|_F^2 = \left\langle \Pi_{\Omega_i}(\widehat{\mathcal{A}}_1), \Pi_{\Omega_i}(\widehat{\mathcal{A}}_1) \right\rangle$, we have,

$$\begin{aligned}
\|\widehat{\boldsymbol{\beta}}_{1,4} - \boldsymbol{\beta}^*_{1,4}\| &= \Bigg\| \left\{ \frac{1}{n} \sum_i \|\Pi_{\Omega_i}(\widehat{\mathcal{A}}_1)\|_F^2 \mathbf{x}_i \mathbf{x}_i^\top \right\}^{-1} \left\{ \frac{1}{n} \sum_i \left\langle \Pi_{\Omega_i}(\mathcal{A}^*_1 - \widehat{\mathcal{A}}_1), \Pi_{\Omega_i}(\widehat{\mathcal{A}}_1) \right\rangle \mathbf{x}_i \mathbf{x}_i^\top \boldsymbol{\beta}^*_{1,4} \right. \\
&\quad \left. + \frac{1}{n} \sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \Pi_{\Omega_i}(\widehat{\mathcal{A}}_1) \right\rangle \mathbf{x}_i \right\} \Bigg\| \\
&\leq \underbrace{\left\| \left\{ \frac{1}{np} \sum_i \|\Pi_{\Omega_i}(\widehat{\mathcal{A}}_1)\|_F^2 \mathbf{x}_i \mathbf{x}_i^\top \right\}^{-1} \right\|}_{I_2} \Bigg(\underbrace{\left\| \frac{1}{np} \sum_i \left\langle \Pi_{\Omega_i}(\mathcal{A}^*_1 - \widehat{\mathcal{A}}_1), \Pi_{\Omega_i}(\widehat{\mathcal{A}}_1) \right\rangle \mathbf{x}_i \mathbf{x}_i^\top \boldsymbol{\beta}^*_{1,4} \right\|}_{II_2} \\
&\quad + \underbrace{\left\| \frac{1}{np} \sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \Pi_{\Omega_i}(\widehat{\mathcal{A}}_1) \right\rangle \mathbf{x}_i \right\|}_{III_2} \Bigg).
\end{aligned}$$

We next bound the three terms, I_2, II_2 and III_2, respectively.

Bound for I_2: We first show that,

$$\left\| \frac{1}{n} \left\{ \frac{1}{p} \sum_i \|\Pi_{\Omega_i}(\widehat{\boldsymbol{\beta}}_{1,1} \circ \widehat{\boldsymbol{\beta}}_{1,2} \circ \widehat{\boldsymbol{\beta}}_{1,3})\|_F^2 - \|\widehat{\boldsymbol{\beta}}_{1,1} \circ \widehat{\boldsymbol{\beta}}_{1,2} \circ \widehat{\boldsymbol{\beta}}_{1,3}\|_F^2 \right\} \mathbf{x}_i \mathbf{x}_i^\top \right\| \leq \gamma,$$

where γ is the same constant as defined in (S4). Denote

$$\begin{aligned}
Z_i &= \frac{1}{n} \left\{ \frac{1}{p} \|\Pi_{\Omega_i}(\widehat{\boldsymbol{\beta}}_{1,1} \circ \widehat{\boldsymbol{\beta}}_{1,2} \circ \widehat{\boldsymbol{\beta}}_{1,3})\|_F^2 - \|\widehat{\boldsymbol{\beta}}_{1,1} \circ \widehat{\boldsymbol{\beta}}_{1,2} \circ \widehat{\boldsymbol{\beta}}_{1,3}\|_F^2 \right\} \mathbf{x}_i \mathbf{x}_i^\top \\
&= \frac{1}{n} \left\{ \frac{1}{p} \|\Pi_{\Omega_i}(\widehat{\boldsymbol{\beta}}_{1,1} \circ \widehat{\boldsymbol{\beta}}_{1,2} \circ \widehat{\boldsymbol{\beta}}_{1,3} 1)\|_F^2 - 1 \right\} \mathbf{x}_i \mathbf{x}_i^\top.
\end{aligned}$$

Then, it holds that $\mathbb{E}(Z_i) = 0$ where the expectation is taken with respect to $\delta_{i,l_1,l_2,l}$. Besides,

$$\|Z_i\| = \frac{1}{n} \left\{ \frac{1}{p} \|\Pi_{\Omega_i}(\widehat{\boldsymbol{\beta}}_{1,1} \circ \widehat{\boldsymbol{\beta}}_{1,2} \circ \widehat{\boldsymbol{\beta}}_{1,3})\|_F^2 - 1 \right\} \|\mathbf{x}_i \mathbf{x}_i^\top\| \leq c_1^2 \frac{1}{n} \left\{ \frac{1}{p} - 1 \right\} \leq \frac{c_1^2}{np}.$$

In addition, we have that,

$$\Big\|\sum_i \mathbb{E}(Z_i^2)\Big\| = \Big\|\sum_i \frac{1}{n^2}\mathbb{E}\left(\left\{\frac{1}{p}\|\Pi_{\Omega_i}(\widehat{\boldsymbol{\beta}}_{1,1}\circ\widehat{\boldsymbol{\beta}}_{1,2}\circ\widehat{\boldsymbol{\beta}}_{1,3})\|_F^2 - 1\right\}^2\right)\mathbf{x}_i\mathbf{x}_i^\top\mathbf{x}_i\mathbf{x}_i^\top\Big\|$$

$$= \Big\|\sum_i\sum_{j,k,l}\frac{1}{n^2 p}\widehat{\boldsymbol{\beta}}_{1,1,j}^4\widehat{\boldsymbol{\beta}}_{1,2,k}^4\widehat{\boldsymbol{\beta}}_{1,3,l}^4\mathbf{x}_i\mathbf{x}_i^\top\mathbf{x}_i\mathbf{x}_i^\top\Big\| \le \frac{1}{np}\sum_{j,k,l}\widehat{\boldsymbol{\beta}}_{1,1,j}^4\widehat{\boldsymbol{\beta}}_{1,2,k}^4\widehat{\boldsymbol{\beta}}_{1,3,l}^4\|1/n\sum_i\mathbf{x}_i\mathbf{x}_i^\top\mathbf{x}_i\mathbf{x}_i^\top\|$$

$$\le \frac{c_1^2c_2\mu^6}{nps^3} \le \frac{c_1^2c_2\mu^3}{nps^{1.5}}.$$

By matrix Bernstein inequality, we have that,

$$P\left(\left\|\frac{1}{n}\frac{1}{p}\sum_i\|\Pi_{\Omega_i}(\widehat{\boldsymbol{\beta}}_{1,1}\circ\widehat{\boldsymbol{\beta}}_{1,2}\circ\widehat{\boldsymbol{\beta}}_{1,3})\|_F^2\mathbf{x}_i\mathbf{x}_i^\top - \sum_i\frac{1}{n}\mathbf{x}_i\mathbf{x}_i^\top\right\| \le \gamma\right)$$
$$\ge 1 - 2q\exp\left\{\frac{-\gamma^2/2}{c_1^2c_2\mu^3/\{nps^{1.5}\} + c_1^2\gamma/\{3np\}}\right\}.$$

By Assumption 2, i.e., $p \ge c_4\mu^3\{\log(d)\}/(n\,s^{1.5}) \ge c\mu^3\{\log(d)\}/(n\,s^{1.5}\gamma^2)$ for some constant c. Here γ is the constant as defined in (S4). With probability at least $1-2q/d^{10}$, we have

$$\left\|\frac{1}{n}\frac{1}{p}\sum_i\|\Pi_{\Omega_i}(\widehat{\boldsymbol{\beta}}_{1,1}\circ\widehat{\boldsymbol{\beta}}_{1,2}\circ\widehat{\boldsymbol{\beta}}_{1,3})\|_F^2\mathbf{x}_i\mathbf{x}_i^\top - \sum_i\frac{1}{n}\mathbf{x}_i\mathbf{x}_i^\top\right\| \le \gamma.$$

For two matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, since $\|\mathbf{A}_1^{-1}-\mathbf{A}_2^{-1}\| \le \|\mathbf{A}_1-\mathbf{A}_2\|\|\mathbf{A}_1^{-1}\|\|\mathbf{A}_2^{-1}\|$, we have that,

$$\|\mathbf{A}_1^{-1}\| \le \frac{\|\mathbf{A}_2^{-1}\|}{1-\|\mathbf{A}_1-\mathbf{A}_2\|\|\mathbf{A}_2^{-1}\|}. \tag{S24}$$

Letting $\mathbf{A}_1$ denote the matrix $\sum_i\|\Pi_{\Omega_i}(\widehat{\boldsymbol{\beta}}_{1,1}\circ\widehat{\boldsymbol{\beta}}_{1,2}\circ\widehat{\boldsymbol{\beta}}_{1,3})\|_F^2\mathbf{x}_i\mathbf{x}_i^\top$, and $\mathbf{A}_2$ denote the matrix $\sum_i\frac{1}{n}\mathbf{x}_i\mathbf{x}_i^\top$, we then have that,

$$\begin{aligned} I_2 &= \left\{\frac{1}{np}\sum_i\|\Pi_{\Omega_i}(\widehat{\mathcal{A}}_1)\|_F^2\mathbf{x}_i\mathbf{x}_i^\top\right\}^{-1} \\ &\le \frac{1}{\widehat{w}_1^2}\frac{\left\|\left\{1/n\sum_i\mathbf{x}_i\mathbf{x}_i^\top\right\}^{-1}\right\|}{1-\gamma\left\|\left\{1/n\sum_i\mathbf{x}_i\mathbf{x}_i^\top\right\}^{-1}\right\|} \le \frac{4}{\lambda_{\min}w_1^{*2}\{1-\gamma/\lambda_{\min}\}} \le \frac{8}{\lambda_{\min}w_1^{*2}}, \end{aligned} \tag{S25}$$

where the first inequality is from (S24), the second inequality is due to the fact that $|\widehat{w}_1 - w_1^*| < w_1^*/2$, and the last inequality is due to the fact that constant $\gamma < \lambda_{\min}/2$ based on (S4).

Bound for II_2: Let $Z_{i,l_1,2,l} = p^{-1}n^{-1}\delta_{i,l_1,l_2,l}(\mathcal{A}_1^* - \widehat{\mathcal{A}}_1)_{i,l_1,l_2,l}(\widehat{\mathcal{A}}_1)_{i,l_1,l_2,l}\mathbf{x}_i\mathbf{x}_i^\top\boldsymbol{\beta}_{1,4}^*$. Then,

$$\frac{1}{np}\sum_i\left\langle\Pi_{\Omega_i}(\mathcal{A}_1^*-\widehat{\mathcal{A}}_1), \Pi_{\Omega_i}(\widehat{\mathcal{A}}_1)\right\rangle\mathbf{x}_i\mathbf{x}_i^\top\boldsymbol{\beta}_{1,4}^* = \sum_{i,l,l_1,l_2} Z_{i,l_1,2,l}.$$

Note that,

$$\|Z_{i,l_1,2,l} - \mathbb{E}(Z_{i,l_1,2,l})\| \le \frac{1}{pn}\left|(\mathcal{A}_1^* - \widehat{\mathcal{A}}_1)_{i,l_1,l_2,l}\right|\left|(\widehat{\mathcal{A}}_1)_{i,l_1,l_2,l}\right|\|\mathbf{x}_i\mathbf{x}_i^\top \boldsymbol{\beta}_{1,4}^*\|$$

By Assumption 1, we have $\|\mathbf{x}_i\mathbf{x}_i^\top \boldsymbol{\beta}_{1,4}^*\| \le c_1^2$. Besides, we have $\left|(\widehat{\mathcal{A}}_1)_{i,l_1,l_2,l}\right| \le \frac{3}{2}w_1^*\mu^3/s^{1.5}$, since $|\widehat{w}| \le 3/2w_1^*$, and the $\mu-$mass condition. Furthermore, we have that,

$$\begin{aligned}
&\left|(\mathcal{A}_1^* - \widehat{\mathcal{A}}_1)_{i,l_1,l_2,l}\right| \le \left|w_1^*\boldsymbol{\beta}_{1,1,l_1}^*\boldsymbol{\beta}_{1,2,l_2}^*\boldsymbol{\beta}_{1,3,l}^* - \widehat{w}_1\boldsymbol{\beta}_{1,1,l_1}^*\boldsymbol{\beta}_{1,2,l_2}^*\boldsymbol{\beta}_{1,3,l}^*\right| \\
&+ \left|\widehat{w}_1\boldsymbol{\beta}_{1,1,l_1}^*\boldsymbol{\beta}_{1,2,l_2}^*\boldsymbol{\beta}_{1,3,l}^* - \widehat{w}_1\widehat{\boldsymbol{\beta}}_{1,1,l_1}\boldsymbol{\beta}_{1,2,l_2}^*\boldsymbol{\beta}_{1,3,l}^*\right| + \left|\widehat{w}_1\widehat{\boldsymbol{\beta}}_{1,1,l_1}\boldsymbol{\beta}_{1,2,l_2}^*\boldsymbol{\beta}_{1,3,l}^* - \widehat{w}_1\widehat{\boldsymbol{\beta}}_{1,1,l_1}\widehat{\boldsymbol{\beta}}_{1,2,l_2}\boldsymbol{\beta}_{1,3,l}^*\right| \\
&+ \left|\widehat{w}_1\widehat{\boldsymbol{\beta}}_{1,1,l_1}\widehat{\boldsymbol{\beta}}_{1,2,l_2}\boldsymbol{\beta}_{1,3,l}^* - \widehat{w}_1\widehat{\boldsymbol{\beta}}_{1,1,l_1}\widehat{\boldsymbol{\beta}}_{1,2,l_2}\widehat{\boldsymbol{\beta}}_{1,3,l}\right| \le \frac{6\mu^2 w_1^*\epsilon}{s}.
\end{aligned}$$

Therefore, we have,

$$\|Z_{i,l_1,2,l} - \mathbb{E}(Z_{i,l_1,2,l})\| \le \frac{1}{pn}6w_1^*\frac{\mu^2}{s}\frac{3}{2}w_1^*\frac{\mu^3}{s^{1.5}}c_1^2.$$

Also we have

$$\begin{aligned}
\sum_{i,l,l_1,l_2}\mathbb{E}(\|Z_{i,l_1,2,l} - \mathbb{E}(Z_{i,l_1,2,l})\|^2) &\le \frac{1}{pn}\sum_i \frac{1}{n}\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i\mathbf{x}_i^\top\mathbf{x}_i\mathbf{x}_i^\top\boldsymbol{\beta}_{1,4}^*\widehat{w}_1^2\left\{6w_1^*\epsilon\frac{\mu}{s}\right\}^2 \\
&\le \frac{81c_1^2c_2w_1^{*4}\mu^3\epsilon^2}{pns^{1.5}}.
\end{aligned}$$

By vector Bernstein's inequality, we have that,

$$P\left(\left\|\sum_{i,l,l_1,l_2}Z_{i,l_1,l_2,l} - \frac{1}{n}\sum_i\langle\mathcal{A}_1^* - \widehat{\mathcal{A}}_1, \widehat{\mathcal{A}}_1\rangle\mathbf{x}_i\mathbf{x}_i^\top\boldsymbol{\beta}_{1,4}^*\right\| \le \gamma w_1^{*2}\epsilon\right) \ge 1 - \exp\left(\frac{1}{4} - \frac{\gamma^2}{\frac{8(c_4+3)^2c_1^2c_2\mu^3}{pns^{1.5}}}\right).$$

Therefore, by Assumption 2, with probability at least $1 - e^{\frac{1}{4}}/d^{10}$,

$$\begin{aligned}
II_2 &\le \left\|\frac{1}{np}\sum_i\left\langle\Pi_{\Omega_i}(\mathcal{A}_1^* - \widehat{\mathcal{A}}_1), \Pi_{\Omega_i}(\widehat{\mathcal{A}}_1)\right\rangle\mathbf{x}_i\mathbf{x}_i^\top\boldsymbol{\beta}_{1,4}^*\right\| \le \frac{1}{n}\sum_i\langle\mathcal{A}_1^* - \widehat{\mathcal{A}}_1, \widehat{\mathcal{A}}_1\rangle\mathbf{x}_i\mathbf{x}_i^\top\boldsymbol{\beta}_{1,4}^* + \gamma w_1^{*2}\epsilon \\
&\le \frac{1}{n}\sum_i\left\|\mathcal{A}_1^* - \widehat{\mathcal{A}}_1\right\|_F\left\|\widehat{\mathcal{A}}_1\right\|_F\|\mathbf{x}_i\mathbf{x}_i^\top\| + \gamma w_1^{*2}\epsilon.
\end{aligned}$$

Note that $\|\widehat{\mathcal{A}}_1\|_F = |\widehat{w}_1| \le 3/2w_1^*$. We next bound $\|\mathcal{A}_1^* - \widehat{\mathcal{A}}_1\|_F$ as,

$$\begin{aligned}
&\|\widehat{\mathcal{A}}_1 - \mathcal{A}_1^*\|_F = \left\|\widehat{w}_1\widehat{\boldsymbol{\beta}}_{1,1}\circ\widehat{\boldsymbol{\beta}}_{1,2}\circ\widehat{\boldsymbol{\beta}}_{1,3} - w_1^*\boldsymbol{\beta}_{1,1}^*\circ\boldsymbol{\beta}_{1,2}^*\circ\boldsymbol{\beta}_{1,3}^*\right\|_F \\
\le\ &\underbrace{\left\|\widehat{w}_1\widehat{\boldsymbol{\beta}}_{1,1}\circ\widehat{\boldsymbol{\beta}}_{1,2}\circ\widehat{\boldsymbol{\beta}}_{1,3} - w_1^*\widehat{\boldsymbol{\beta}}_{1,1}\circ\widehat{\boldsymbol{\beta}}_{1,2}\circ\widehat{\boldsymbol{\beta}}_{1,3}\right\|_F}_{II_{21}} + \underbrace{\left\|w_1^*\widehat{\boldsymbol{\beta}}_{1,1}\circ\widehat{\boldsymbol{\beta}}_{1,2}\circ\widehat{\boldsymbol{\beta}}_{1,3} - w_1^*\boldsymbol{\beta}_{1,1}^*\circ\widehat{\boldsymbol{\beta}}_{1,2}\circ\widehat{\boldsymbol{\beta}}_{1,3}\right\|_F}_{II_{22}} \\
&+ \underbrace{\left\|w_1^*\boldsymbol{\beta}_{1,1}^*\circ\widehat{\boldsymbol{\beta}}_{1,2}\circ\widehat{\boldsymbol{\beta}}_{1,3} - w_1^*\boldsymbol{\beta}_{1,1}^*\circ\boldsymbol{\beta}_{1,2}^*\circ\widehat{\boldsymbol{\beta}}_{1,3}\right\|_F}_{II_{23}} + \underbrace{\left\|w_1^*\boldsymbol{\beta}_{1,1}^*\circ\boldsymbol{\beta}_{1,2}^*\circ\widehat{\boldsymbol{\beta}}_{1,3} - w_1^*\boldsymbol{\beta}_{1,1}^*\circ\boldsymbol{\beta}_{1,2}^*\circ\boldsymbol{\beta}_{1,3}^*\right\|_F}_{II_{24}}.
\end{aligned}$$

Note that $II_{21} = |\widehat{w}_1 - w_1^*| < w_1^*\epsilon$, $II_{22} \le w_1^* \|\widehat{\boldsymbol\beta}_{1,1} - \boldsymbol\beta_{1,1}^*\| \le w_1^*\epsilon$, $II_{23} \le w_1^*\epsilon$, and $II_{24} \le w_1^*\epsilon$. Therefore, we have $\|\widehat{\mathcal{A}}_1 - \mathcal{A}_1^*\|_F \le 4w_1^*\epsilon$. By Assumption 1(i) that $n^{-1}\sum_i \|\mathbf{x}_i\mathbf{x}_i^\top\| \le c_2$,

$$II_2 \le \{6c_2 + \gamma\} w_1^{*2}\epsilon. \tag{S26}$$

Bound for III_2: By definition,

$$III_2 = \left\| \frac{1}{np}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \widehat{\mathcal{A}}_1 \right\rangle \mathbf{x}_i \right\| = \left\| \widehat{w}_1 \frac{1}{np}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \widehat{\boldsymbol\beta}_{1,1}\circ\widehat{\boldsymbol\beta}_{1,2}\circ\widehat{\boldsymbol\beta}_{1,3} \right\rangle \mathbf{x}_i \right\|.$$

The jth entry of the vector $\frac{1}{np}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \widehat{\mathcal{A}}_1 \right\rangle \mathbf{x}_i$ for each $j \in [q]$ can be written as

$$\frac{1}{np}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \widehat{\mathcal{A}}_1 \right\rangle \mathbf{x}_{i,j} = \frac{c_1}{np}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \widehat{\mathcal{A}}_1 \right\rangle \mathbf{x}_{i,l}/c_1.$$

Our goal is to prove the upper bound of

$$\left\| \frac{1}{np}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \widehat{\boldsymbol\beta}_{1,1}\circ\widehat{\boldsymbol\beta}_{1,2}\circ\widehat{\boldsymbol\beta}_{1,3} \right\rangle \mathbf{x}_{i,j} \right\|.$$

Note that $|\mathbf{x}_{i,j}/c_1| \le 1$. Therefore, it suffices to bound

$$\left\| \frac{c_1}{np}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \widehat{\boldsymbol\beta}_{1,1}\circ\widehat{\boldsymbol\beta}_{1,2}\circ\widehat{\boldsymbol\beta}_{1,3} \right\rangle \right\| \le \left\| \frac{c_1}{np}\sum_i \Pi_{\Omega_i}(\mathcal{E}_i) \right\|.$$

Similary as the upper bound of $\left\|\sum_i \frac{1}{pn}\alpha_{i,1}^*\Pi_{\Omega_i}(\mathcal{E}_{iF})\right\|$ in (S13), we obtain for each $j \in [q]$, with probability at least $1 - 2d^{-10s}$, that

$$\left\| \frac{c_1}{np}\sum_i \Pi_{\Omega_i}(\mathcal{E}_i) \right\| \le \widetilde{C}_2\sigma\sqrt{\frac{s\log(d)}{np}},$$

for $\widetilde{C}_2 > 0$. Therefore, we conclude that with probability at least $1 - 2q/d^{10s}$

$$III_2 \le |\widehat{w}_1| \left\| \frac{1}{np}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \widehat{\boldsymbol\beta}_{1,1}\circ\widehat{\boldsymbol\beta}_{1,2}\circ\widehat{\boldsymbol\beta}_{1,3} \right\rangle \mathbf{x}_i \right\| \le \frac{3\widetilde{C}_2\sigma w_1^*}{2}\sqrt{\frac{qs\log(d)}{np}}. \tag{S27}$$

Combining the bounds of I_2, II_2 and III_2, with probability at least $1 - 1/d^9$, we obtain that,

$$\begin{aligned} \|\widehat{\boldsymbol\beta}_{1,4} - \boldsymbol\beta_{1,4}^*\| &\le \frac{8}{\lambda_{\min} w_1^{*2}}\left[\{6c_2 + \gamma\} w_1^{*2}\epsilon + \frac{3\widetilde{C}_2\sigma w_1^*}{2}\sqrt{\frac{qs\log(d)}{np}} \right] \\ &\le \kappa_2\epsilon + \frac{\widetilde{C}_2\sigma}{\lambda_{\min} w_1^*}\sqrt{\frac{qs\log(d)}{np}}. \end{aligned} \tag{S28}$$

Here, the constant $\gamma < c_2$ by (S4). Let $\kappa_2 = c_2/\lambda_{\min}$ where some constant multiplier is merged into c_2 and $\widetilde{C}_2$.

S1.4 Proof of the theorem

Finally, we iteratively apply the error bound from each step, and obtain the final error bound in Theorem 1. Given the initial estimators $\widehat{\boldsymbol{\beta}}_{1,j}^{(0)}$ and $\widehat{w}_1^{(0)}$ with an initialization error ϵ, the error bound in (S23) implies that, with probability at least $1-1/d^9$,

$$\|\widehat{\boldsymbol{\beta}}_{1,3}^{(1)} - \boldsymbol{\beta}_{1,3}^*\| \leq \kappa_1\epsilon + \frac{6\sqrt{10}\lambda_{\max}\widetilde{C}\sigma}{w_1^*\lambda_{\min}^2}\sqrt{\frac{s\log(d)}{np}},$$

where $\kappa_1 = 6\sqrt{10}\lambda_{\max}^2\epsilon/\lambda_{\min}^2 + 12\sqrt{10}\lambda_{\max}\gamma/(\lambda_{\min}^2) + 18\sqrt{10}\lambda_{\max}\gamma'/\lambda_{\min}$. By (S4) the constant $\gamma < \lambda_{\min}^2/\{48\sqrt{10}\lambda_{\max}\}$, then we have $\frac{12\sqrt{10}\lambda_{\max}\gamma}{\lambda_{\min}^2} < \frac{1}{4}$. Let the positive constantγ' satisfy

$$\gamma' = \frac{1}{2}\min\left\{\frac{\lambda_{\min}}{72\sqrt{10}\lambda_{\max}}, \frac{\lambda_{\min}^2}{144\sqrt{10}\lambda_{\max}c_2}\right\}. \tag{S29}$$

Then it is easy to check $18\sqrt{10}\lambda_{\max}\gamma'/\lambda_{\min} < 1/4$. By Assumption 4 we have $\frac{6\sqrt{10}\lambda_{\max}^2}{\lambda_{\min}^2}\epsilon < \frac{1}{4}$. Therefore, $\kappa_1 < 1$. By a similar derivation, the error bound holds for $\|\widehat{\boldsymbol{\beta}}_{1,1}^{(1)} - \boldsymbol{\beta}_{1,1}^*\|$, $\|\widehat{\boldsymbol{\beta}}_{1,2}^{(1)} - \boldsymbol{\beta}_{1,2}^*\|$, and $|\widehat{w}_1^{(1)} - w_1^*|/w_1^*$. Then by the error bound in (S28), with probability at least $1-2/d^9$, we have that,

$$\|\widehat{\boldsymbol{\beta}}_{1,4}^{(1)} - \boldsymbol{\beta}_{1,4}^*\| \;\leq\; \kappa_2\kappa_1\epsilon + \kappa_2\frac{6\sqrt{10}\widetilde{C}\lambda_{\max}\sigma}{\lambda_{\min}^2 w_1^*}\sqrt{\frac{s\log(d)}{np}} + \frac{\widetilde{C}_2\sigma}{\lambda_{\min}w_1^*}\sqrt{\frac{sq\log(d)}{np}}.$$

The contraction coefficient is

$$\kappa = \kappa_1\kappa_2 = \{6\sqrt{10}\lambda_{\max}^2\epsilon/\lambda_{\min}^2 + 12\sqrt{10}\lambda_{\max}\gamma/\lambda_{\min}^2 + 18\sqrt{10}\lambda_{\max}\gamma'/\lambda_{\min}\}c_2/\lambda_{\min}.$$

By (S4), the constant $\gamma < \lambda_{\min}^3/\{48\sqrt{10}c_2\lambda_{\max}\}$ then we have $\frac{12\sqrt{10}c_2\lambda_{\max}\gamma}{\lambda_{\min}^3} < \frac{1}{4}$. By (S29), we have $18\sqrt{10}\lambda_{\max}c_2\gamma'/\lambda_{\min} < 1/4$. By Assumption 4, we have $\epsilon < \lambda_{\min}^3/\{24\sqrt{10}c_2\lambda_{\max}^2\}$. Therefore, it is easy to check that $\frac{6\sqrt{10}c_2\lambda_{\max}^2\epsilon}{\lambda_{\min}^3} < \frac{1}{4}$, then $\kappa < 1$.

We have now obtained the error bound from the first iteration. After repeatedly plugging the estimation error bound from iteration $(t-1)$ into the error bound from iteration t, with probability at least $1-(t+1)/d^9$, we have that,

$$\begin{aligned}
&\max\left\{|\widehat{w}_1^{(t)} - w_1^*|/w_1^*, \max_j\{\|\widehat{\boldsymbol{\beta}}_{1,j}^{(t)} - \boldsymbol{\beta}_{1,j}^*\|_2\}\right\} \\
\leq\;& \kappa^t\epsilon + \frac{1-\kappa^t}{1-\kappa}\frac{6\sqrt{10}\widetilde{C}\lambda_{\max}\sigma}{\lambda_{\min}^2 w_1^*}\sqrt{\frac{s\log(d)}{np}} + \frac{1-\kappa^{t-1}}{1-\kappa}\frac{\widetilde{C}_2\sigma}{\lambda_{\min}w_1^*}\sqrt{\frac{sq\log(d)}{np}} \\
\leq\;& \kappa^t\epsilon + \frac{1}{1-\kappa}\frac{C_1\sigma}{w_1^*}\sqrt{\frac{s\log(d)}{np}}
\end{aligned}$$

where $C_1 = (6\sqrt{10}\widetilde{C}\lambda_{\max} + \widetilde{C}_2\lambda_{\min}\sqrt{q})/\lambda_{\min}^2$, and q is the dimension of the predictor vector and is fixed under Assumption 1(i). This completes the proof of Theorem 1. □

S2 Proof of Theorem 2

We next extend the theory to the general rank r case. Although this proof follows similar steps as those of Theorem 1, the interplay among parameters from different ranks makes the derivations more challenging and considerably different from Theorem 1.

S2.1 Error bound of the estimator from Step 1 of Algorithm 1

Similar to the rank-1 case, the first step is to bound the error for the unconstrained estimator from Step 1 of Algorithm 1. Now the model under a general rank r is of the form,

$$\mathcal{Y}_i = \sum_{k\in[r]} \boldsymbol{\beta}_{k,4}^{*\top}\mathbf{x}_i w_k^* \boldsymbol{\beta}_{k,1}^* \circ \boldsymbol{\beta}_{k,2}^* \circ \boldsymbol{\beta}_{k,3}^* + \mathcal{E}_i, \quad i = 1, \ldots, n.$$

The unconstrained estimator is of the form,

$$\widetilde{\boldsymbol{\beta}}_{k,3,l} = \frac{\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}^2 \sum_{l_1,l_2} \delta_{i,l_1,l_2,l}\, \widehat{\mathcal{R}}_{i,l_1,l_2,l}\, \widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}}{\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}^2 \sum_{l_1,l_2} \widehat{w}_k\delta_{i,l_1,l_2,l}\, \widehat{\boldsymbol{\beta}}_{k,1,l_1}^2\widehat{\boldsymbol{\beta}}_{k,2,l_2}^2},$$

where $\widehat{\alpha}_{i,k} = \widehat{\boldsymbol{\beta}}_{k,4}^{\top}\mathbf{x}_i$ and $\widehat{\mathcal{R}}_i = \left(\mathcal{Y}_i - \sum_{k'\neq k} \widehat{w}_{k'}\widehat{\alpha}_{i,k'}\widehat{\boldsymbol{\beta}}_{k',1} \circ \widehat{\boldsymbol{\beta}}_{k',2} \circ \widehat{\boldsymbol{\beta}}_{k',1}\right)/\widehat{\alpha}_{i,k}$. Denote $F_1 = \text{supp}(\boldsymbol{\beta}_{k,1}^*) \cup \text{supp}(\widehat{\boldsymbol{\beta}}_{k,1})$, $F_2 = \text{supp}(\boldsymbol{\beta}_{k,2}^*) \cup \text{supp}(\widehat{\boldsymbol{\beta}}_{k,2})$, and $F_3 = \text{supp}(\boldsymbol{\beta}_{k,3}^*) \cup \text{supp}(\widehat{\boldsymbol{\beta}}_{k,3})$. Let $F = F_1 \circ F_2 \circ F_3$. Similar to the rank-1 case, we restrict $\widehat{\mathcal{R}}_i$ on the three modes indexed by F_1, F_2 and F_3. Then $\widetilde{\boldsymbol{\beta}}_{k,3,l}$ can be rewritten as,

$$\widetilde{\boldsymbol{\beta}}_{k,3,l} = \frac{\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}^2 \sum_{l_1,l_2} \delta_{i,l_1,l_2,l}\, (\widehat{\mathcal{R}}_{iF})_{l_1,l_2,l}\, \widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}}{\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}^2 \sum_{l_1,l_2} \widehat{w}_k\delta_{i,l_1,l_2,l}\, \widehat{\boldsymbol{\beta}}_{k,1,l_1}^2\widehat{\boldsymbol{\beta}}_{k,2,l_2}^2}.$$

For the vector $\boldsymbol{\beta}_{k,j}$, denote $\bar{\boldsymbol{\beta}}_{k,j} = \texttt{Truncate}(\boldsymbol{\beta}_{k,j}^*, F_j)$, for $k = 1, \cdots r$ and $j = 1, 2, 3$. By the definition of F_j, we have $\bar{\boldsymbol{\beta}}_{k,j}^* = \boldsymbol{\beta}_{k,j}^*$ and $\bar{\widehat{\boldsymbol{\beta}}}_{k,j} = \widehat{\boldsymbol{\beta}}_{k,j}$, for $j = 1, 2, 3$. By Lemma 2 and substituting the expression of $\widehat{\mathcal{R}}_i$ into $\widetilde{\boldsymbol{\beta}}_{k,3,l}$, the vector $\widetilde{\boldsymbol{\beta}}_{k,3}$ can be expanded as

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}_{k,3} \;=\;& \frac{w_k^* \sum_i \widehat{\alpha}_{i,k}\alpha_{i,k}^*/n}{\widehat{w}_k \sum_i \widehat{\alpha}_{i,k}^2/n} \langle \boldsymbol{\beta}_{k,1}^*, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \boldsymbol{\beta}_{k,2}^*, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \boldsymbol{\beta}_{k,3}^* \\
&+ \frac{\sum_{k'\neq k}\{\sum_i \frac{1}{n}\widehat{\alpha}_{i,k}\alpha_{i,k'}^*\} w_{k'}^* \langle \bar{\boldsymbol{\beta}}_{k',1}^*, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\boldsymbol{\beta}}_{k',2}^*, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \bar{\boldsymbol{\beta}}_{k',3}^*}{\widehat{w}_k \sum_i \widehat{\alpha}_{i,k}^2/n} \\
&- \frac{\sum_{k'\neq k}\{\sum_i \frac{1}{n}\widehat{\alpha}_{i,k}\widehat{\alpha}_{i,k'}\} \widehat{w}_{k'} \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \bar{\widehat{\boldsymbol{\beta}}}_{k',3}}{\widehat{w}_k \sum_i \widehat{\alpha}_{i,k}^2/n} \\
&+ \frac{w_k^*}{\widehat{w}_k}\mathbf{A}^{-1}\left\{\mathbf{B} - \mathbf{A}\frac{\sum_i \widehat{\alpha}_{i,k}\alpha_{i,k}^*/n}{\sum_i \widehat{\alpha}_{i,k}^2/n} \langle \boldsymbol{\beta}_{k,1}^*, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \boldsymbol{\beta}_{k,2}^*, \widehat{\boldsymbol{\beta}}_{k,2}\rangle\right\} \boldsymbol{\beta}_{k,3}^* \\
&+ \sum_{k'\neq k} \mathbf{A}^{-1}\frac{w_{k'}^*}{\widehat{w}_k}\left\{\mathbf{F}_{k'} - \frac{\sum_i \widehat{\alpha}_{i,k}\alpha_{i,k'}^*/n}{\sum_i \widehat{\alpha}_{i,k}^2/n}\mathbf{A}\langle \bar{\boldsymbol{\beta}}_{k',1}^*, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\boldsymbol{\beta}}_{k',2}^*, \widehat{\boldsymbol{\beta}}_{k,2}\rangle\right\} \bar{\boldsymbol{\beta}}_{k',3}^* \\
&- \sum_{k'\neq k} \mathbf{A}^{-1}\frac{\widehat{w}_{k'}}{\widehat{w}_k}\left\{\mathbf{G}_{k'} - \frac{\sum_i \widehat{\alpha}_{i,k}\widehat{\alpha}_{i,k'}/n}{\sum_i \widehat{\alpha}_{i,k}^2/n}\boldsymbol{A}\langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle\right\} \bar{\widehat{\boldsymbol{\beta}}}_{k',3} \\
&+ \frac{1}{\widehat{w}_1}\mathbf{A}^{-1}\sum_{i=1}^n \frac{\widehat{\alpha}_{i,1}}{n}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_{k,1} \times_2 \widehat{\boldsymbol{\beta}}_{k,2},
\end{aligned}$$

where $\mathbf{A}$ and $\mathbf{B}$ are diagonal matrices with the diagonal entry,

$$\begin{aligned}
\boldsymbol{A}_{ll} &= \sum_{i=1}^n \widehat{\alpha}_{i,k}^2/n \sum_{l_1,l_2} \delta_{i,l_1,l_2,l}\widehat{\boldsymbol{\beta}}_{k,1,l_1}^2 \widehat{\boldsymbol{\beta}}_{k,2,l_2}^2, \\
\boldsymbol{B}_{ll} &= \sum_{i=1}^n \widehat{\alpha}_{i,k}\alpha_{i,k}^*/n \sum_{l_1,l_2} \delta_{i,l_1,l_2,l}\boldsymbol{\beta}_{k,1,l_1}^* \boldsymbol{\beta}_{k,2,l_2}^* \widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}, \\
\boldsymbol{F}_{k'll} &= \sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}\alpha_{i,k'}^* \sum_{l_1,l_2} \delta_{i,l_1,l_2,l}\bar{\boldsymbol{\beta}}_{k',1,l_1}^* \bar{\boldsymbol{\beta}}_{k',2,l_2}^* \widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}, \\
\mathbf{G}_{k'll} &= \sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}\widehat{\alpha}_{i,k'} \sum_{l_1,l_2} \delta_{i,l_1,l_2,l}\bar{\widehat{\boldsymbol{\beta}}}_{k',1,l_1} \bar{\widehat{\boldsymbol{\beta}}}_{k',2,l_2} \widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}.
\end{aligned}$$

Then the difference between and $\widetilde{\boldsymbol{\beta}}_{k,3}$ and $\frac{w_k^* \sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}\alpha_{i,k}^*}{\widehat{w}_k \sum_{i=1}^n \widehat{\alpha}_{i,k}^2/n}\boldsymbol{\beta}_{k,3}^*$ is

$$\widetilde{\boldsymbol{\beta}}_{k,3} - \frac{w_k^* \sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}\alpha_{i,k}^*}{\widehat{w}_k \sum_{i=1}^n \widehat{\alpha}_{i,k}^2/n}\boldsymbol{\beta}_{k,3}^* = I_1 + II_1 + III_1 + IV_1 + V_1, \tag{S30}$$

where

$$
\begin{aligned}
I_1 &= \frac{w_k^* \sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}\alpha_{i,k}^*}{\widehat{w}_k \sum_{i=1}^n \widehat{\alpha}_{i,k}^2/n} \left\{\langle \boldsymbol{\beta}_{k,1}^*, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \boldsymbol{\beta}_{k,2}^*, \widehat{\boldsymbol{\beta}}_{k,2}\rangle - 1\right\} \boldsymbol{\beta}_{k,3}^*, \\
II_1 &= \frac{1}{\widehat{w}_k \sum_i \widehat{\alpha}_{i,k}^2/n} \sum_i \frac{1}{n}\widehat{\alpha}_{i,k} \sum_{k'\neq k} \Big\{ \alpha_{i,k'}^* w_{k'}^* \langle \bar{\boldsymbol{\beta}}_{k',1}^*, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\boldsymbol{\beta}}_{k',2}^*, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \bar{\boldsymbol{\beta}}_{k',3}^* - \widehat{\alpha}_{i,k'}\widehat{w}_{k'} \\
&\qquad \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \bar{\widehat{\boldsymbol{\beta}}}_{k',3} \Big\}, \\
III_1 &= \frac{w_k^*}{\widehat{w}_k}\mathbf{A}^{-1}\left\{\mathbf{B} - \mathbf{A}\frac{\sum_i \widehat{\alpha}_{i,k}\alpha_{i,k}^*/n}{\sum_i \widehat{\alpha}_{i,k}^2/n}\langle \boldsymbol{\beta}_{k,1}^*, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \boldsymbol{\beta}_{k,2}^*, \widehat{\boldsymbol{\beta}}_{k,2}\rangle\right\} \boldsymbol{\beta}_{k,3}^* \\
IV_1 &= \sum_{k'\neq k} \mathbf{A}^{-1}\frac{w_{k'}^*}{\widehat{w}_k}\left\{\mathbf{F}_{k'} - \frac{\sum_i \widehat{\alpha}_{i,k}\alpha_{i,k'}^*/n}{\sum_i \widehat{\alpha}_{i,k}^2/n}\mathbf{A}\langle \bar{\boldsymbol{\beta}}_{k',1}^*, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\boldsymbol{\beta}}_{k',2}^*, \widehat{\boldsymbol{\beta}}_{k,2}\rangle\right\} \bar{\boldsymbol{\beta}}_{k',3}^* \\
&\quad - \sum_{k'\neq k} \mathbf{A}^{-1}\frac{\widehat{w}_{k'}}{\widehat{w}_k}\left\{\mathbf{G}_{k'} - \frac{\sum_i \widehat{\alpha}_{i,k}\widehat{\alpha}_{i,k'}/n}{\sum_i \widehat{\alpha}_{i,k}^2/n}\mathbf{A}\langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle\right\} \bar{\widehat{\boldsymbol{\beta}}}_{k',3} \\
V_1 &= \frac{1}{\widehat{w}_k}\mathbf{A}^{-1}\sum_{i=1}^n \frac{\widehat{\alpha}_{i,k}}{n}\Pi_{\Omega_i}(\mathcal{E}_{iF}) \times_1 \widehat{\boldsymbol{\beta}}_{k,1} \times_2 \widehat{\boldsymbol{\beta}}_{k,2}.
\end{aligned}
$$

Comparing (S30) with (S2) in the rank-1 case, we see that, when $r = 1$, the sum includes the terms I_1, III_1 and V_1 . When the rank is $r > 1$, the sum includes two additional terms, II_1 and IV_1, which appear due to the interplay among different ranks.

We have shown in (S3) that

$$
\|I_1\| \le \frac{2\lambda_{\max}}{\lambda_{\min}}\epsilon^2.
$$

By (S9), with probability at least $1 - 4/d^{10}$, we have,

$$
\|III_1\| \le \frac{2\gamma}{\lambda_{\min} - \gamma}\epsilon \le \frac{4\gamma}{\lambda_{\min}}\epsilon,
$$

Here γ is the positive constant satisfying

$$
\gamma = \frac{1}{2}\min\left\{\frac{\lambda_{\min}}{2}, c_2, \frac{\lambda_{\min}^2}{192\sqrt{10}\lambda_{\max}}, \frac{\lambda_{\min}^3 w_{\min}^{*2}}{96\sqrt{10}c_2\lambda_{\max}w_{\max}^{*2}r}\right\}. \tag{S31}
$$

Furthermore, by (S20), with probability at least $1 - 10/d^{10}$, we have,

$$
\|V_1\| \le \frac{2\widetilde{C}\sigma}{\lambda_{\min}w_{\min}^*}\sqrt{\frac{s\log(d)}{np}} + 6\gamma'\epsilon.
$$

The value of γ' will be determined later. Next, we bound II_1 and IV_1, respectively.

Bound for II_1: Note that $\langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle = \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1} - \bar{\boldsymbol{\beta}}^*_{k',1} + \bar{\boldsymbol{\beta}}^*_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2} - \bar{\boldsymbol{\beta}}^*_{k',2} + \bar{\boldsymbol{\beta}}^*_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle$. Henceforth,

$$
\begin{aligned}
& \left\| \sum_{k' \neq k} \left\{ \alpha^*_{i,k'} w^*_{k'} \langle \bar{\boldsymbol{\beta}}^*_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\boldsymbol{\beta}}^*_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \bar{\boldsymbol{\beta}}^*_{k',3} - \alpha^*_{i,k'} w^*_{k'} \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \bar{\widehat{\boldsymbol{\beta}}}_{k',3} \right\} \right\| \\
\leq & \left\| \sum_{k' \neq k} \alpha^*_{i,k'} w^*_{k'} \langle \bar{\boldsymbol{\beta}}^*_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\boldsymbol{\beta}}^*_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \left\{ \bar{\boldsymbol{\beta}}^*_{k',3} - \bar{\widehat{\boldsymbol{\beta}}}_{k',3} \right\} \right\| \\
& + \left\| \sum_{k' \neq k} \alpha^*_{i,k'} w^*_{k'} \langle \bar{\boldsymbol{\beta}}^*_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2} - \bar{\boldsymbol{\beta}}^*_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \bar{\widehat{\boldsymbol{\beta}}}_{k',3} \right\| \\
& + \left\| \sum_{k' \neq k} \alpha^*_{i,k'} w^*_{k'} \langle \bar{\boldsymbol{\beta}}^*_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1} - \bar{\boldsymbol{\beta}}^*_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \bar{\widehat{\boldsymbol{\beta}}}_{k',3} \right\| \\
& + \left\| \sum_{k' \neq k} \alpha^*_{i,k'} w^*_{k'} \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2} - \bar{\boldsymbol{\beta}}^*_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1} - \bar{\boldsymbol{\beta}}^*_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \bar{\widehat{\boldsymbol{\beta}}}_{k',3} \right\| \\
\leq & \; c_1 r w^*_{\max} \{\xi + \epsilon\}^2 \epsilon + 2 c_1 r w^*_{\max} \{\xi + \epsilon\} \epsilon + c_1 r w^*_{\max} \epsilon^2 \leq 4 c_1 r w^*_{\max} \epsilon^2 + 4 c_1 r w^*_{\max} \xi \epsilon,
\end{aligned}
$$

where the second inequality is due to the following facts. First, $\left| \langle \bar{\boldsymbol{\beta}}^*_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \right| = \left| \langle \bar{\boldsymbol{\beta}}^*_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1} - \boldsymbol{\beta}^*_{k,1} + \boldsymbol{\beta}^*_{k,1}\rangle \right| \leq \xi + \epsilon$, and $|\alpha^*_{i,k'}| \leq c_1$ due to Assumption 1 (i). Besides,

$$
\begin{aligned}
& \left\| \sum_{k' \neq k} \alpha^*_{i,k'} w^*_{k'} \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \bar{\widehat{\boldsymbol{\beta}}}_{k',3} - \widehat{\alpha}_{i,k'} \widehat{w}_{k'} \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\widehat{\boldsymbol{\beta}}}_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \bar{\widehat{\boldsymbol{\beta}}}_{k',3} \right\| \\
\leq & \; r |\alpha^*_{i,k'} w^*_{k'} - \widehat{\alpha}_{i,k'} \widehat{w}_{k'}| \xi^2 \leq |\widehat{\alpha}_{i,k'} \widehat{w}_{k'} - \widehat{w}_{k'} \alpha^*_{i,k'} + \widehat{w}_{k'} \alpha^*_{i,k'} - \alpha^*_{i,k'} w^*_{k'}| r \xi^2 \\
\leq & \; \frac{5 c_1 r w^*_{\max} \epsilon \xi^2}{2}.
\end{aligned}
$$

By triangular inequality, we have that,

$$
\begin{aligned}
\|II_1\| & \leq \frac{2c_1^2}{\lambda_{\min} w^*_{\min}} \left[4 r w^*_{\max} \epsilon^2 + 4 r w^*_{\max} \xi \epsilon + 5 w^*_{\max} r \epsilon \xi^2 / 2 \right] \\
& \leq \frac{c_1^2 r w^*_{\max} \{8\epsilon^2 + 8\xi\epsilon + 5\xi^2\epsilon\}}{\lambda_{\min} w^*_{\min}}.
\end{aligned}
$$

Bound for IV_1: Denote

$$
\begin{aligned}
Z_{i,k',l_1,l_2,l} = & \frac{w^*_{k'}}{\widehat{w}_k} \left\{ \frac{\sum_{i=1}^n \widehat{\alpha}_{i,k} \alpha^*_{i,k'}/n}{\sum_{i=1}^n \widehat{\alpha}^2_{i,k}/n} \widehat{\alpha}^2_{i,k} \langle \bar{\boldsymbol{\beta}}^*_{k',1}, \widehat{\boldsymbol{\beta}}_{k,1}\rangle \langle \bar{\boldsymbol{\beta}}^*_{k',2}, \widehat{\boldsymbol{\beta}}_{k,2}\rangle \widehat{\boldsymbol{\beta}}^2_{k,1,l_1} \widehat{\boldsymbol{\beta}}^2_{k,2,l_2} \right. \\
& \left. - \widehat{\alpha}_{i,k} \alpha^*_{i,k'} \bar{\boldsymbol{\beta}}^*_{k',1,l_1} \bar{\boldsymbol{\beta}}^*_{k',2,l_2} \widehat{\boldsymbol{\beta}}_{k,1,l_1} \widehat{\boldsymbol{\beta}}_{k,2,l_2} \right\} \delta_{i,l_1,l_2,l} p^{-1} n^{-1} \bar{\boldsymbol{\beta}}^*_{k',3,l} \mathbf{e}_l,
\end{aligned}
$$

where $\mathbf{e}_l$ is the column vector whose lth entry is 1 and others are 0. Then we have,

$$\frac{1}{p}\sum_{k'\neq k}\frac{w^*_{k'}}{\widehat{w}_k}\left\{\mathbf{A}\frac{\sum_{i=1}^n\widehat{\alpha}_{i,k}\alpha^*_{i,k'}/n}{\sum_{i=1}^n\widehat{\alpha}^2_{i,k}/n}\langle\bar{\boldsymbol{\beta}}^*_{k',1},\widehat{\boldsymbol{\beta}}_{k,1}\rangle\langle\bar{\boldsymbol{\beta}}^*_{k',2},\widehat{\boldsymbol{\beta}}_{k,2}\rangle-\mathbf{F}_{k'}\right\}\bar{\boldsymbol{\beta}}^*_{k',3}=\sum_{i,k',l_1,l_2,l}Z_{i,k',l_1,l_2,l}.$$

Since $\boldsymbol{\beta}^*_{k',3},\ \widehat{\boldsymbol{\beta}}_{k,1},\ \widehat{\boldsymbol{\beta}}_{k,2}$ are μ-mass vectors, we have that,

$$\begin{aligned}\|Z_{i,l_1,l_2,l}-\mathbb{E}(Z_{i,l_1,l_2,l})\| \;\leq\; & \frac{2w^*_{\max}}{w^*_{\min}}\frac{1}{pn}\left|\widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}\bar{\boldsymbol{\beta}}^*_{k',3,l}\widehat{\alpha}_{i,k}\right|\left|\frac{\sum_{i=1}^n\widehat{\alpha}_{i,k}\alpha^*_{i,k'}/n}{\sum_{i=1}^n\widehat{\alpha}^2_{i,k}/n}\widehat{\alpha}_{i,k}\langle\bar{\boldsymbol{\beta}}^*_{k',1},\widehat{\boldsymbol{\beta}}_{k,1}\rangle\right.\\ &\left.\langle\bar{\boldsymbol{\beta}}^*_{k',2},\widehat{\boldsymbol{\beta}}_{k,2}\rangle\widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}-\alpha^*_{i,k'}\bar{\boldsymbol{\beta}}^*_{k',1,l_1}\bar{\boldsymbol{\beta}}^*_{k',2,l_2}\right|.\end{aligned}$$

We have proved in (S7) that

$$\left|\frac{\sum_{i=1}^n\widehat{\alpha}_{i,k}\alpha^*_{i,k'}/n}{\sum_{i=1}^n\widehat{\alpha}^2_{i,k}/n}\widehat{\alpha}_{i,k}\langle\bar{\boldsymbol{\beta}}^*_{k',1},\widehat{\boldsymbol{\beta}}_{1,1}\rangle\langle\bar{\boldsymbol{\beta}}^*_{k',2},\widehat{\boldsymbol{\beta}}_{k,2}\rangle\widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}-\alpha^*_{i,k}\bar{\boldsymbol{\beta}}^*_{k',1,l_1}\bar{\boldsymbol{\beta}}^*_{k',2,l_2}\right|^2\leq\left\{4\frac{\lambda^2_{\max}}{\lambda^2_{\min}}+1\right\}c_1^2\epsilon^2.$$

Therefore, we have

$$\|Z_{i,l_1,l_2,l}-\mathbb{E}(Z_{i,l_1,l_2,l})\| \;\leq\; \frac{2w^*_{\max}}{w^*_{\min}pn}\frac{\mu^3}{s^{1.5}}c_1\epsilon\sqrt{\left\{4\frac{\lambda^2_{\max}}{\lambda^2_{\min}}+1\right\}c_1^2}\;\leq\;\gamma\epsilon,$$

where the last inequality holds by Assumption 6, i.e., $p\geq(c_5\mu^3w^*_{\max})/(w^*_{\min}ns^{1.5})$ and the constant γ is defined in (S31).

In addition, we have that,

$$\begin{aligned}&\sum_{i,k',l,l_1,l_2}\mathbb{E}(\|Z_{i,k',l_1,l_2,l}-\mathbb{E}(Z_{i,k',l_1,l_2,l})\|^2)\\ =\;&\left\{\frac{1}{p}-1\right\}\sum_{i,k',l,l_1,l_2}\frac{4w^{*2}_{\max}}{w^{*2}_{\min}n^2}\widehat{\boldsymbol{\beta}}^2_{k,1,l_1}\widehat{\boldsymbol{\beta}}^2_{k,2,l_2}\widehat{\alpha}^2_{i,k}\left\{\frac{\sum_{i=1}^n\widehat{\alpha}_{i,k}\alpha^*_{i,k'}/n}{\sum_{i=1}^n\widehat{\alpha}^2_{i,k}/n}\widehat{\alpha}_{i,k}\langle\bar{\boldsymbol{\beta}}^*_{k',1},\widehat{\boldsymbol{\beta}}_{k,1}\rangle\langle\bar{\boldsymbol{\beta}}^*_{k',2},\widehat{\boldsymbol{\beta}}_{k,2}\rangle\widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}\right.\\ &\left.-\alpha^*_{i,k'}\bar{\boldsymbol{\beta}}^*_{k',1,l_1}\bar{\boldsymbol{\beta}}^*_{k',2,l_2}\right\}^2\\ \leq\;&\frac{4w^{*2}_{\max}}{w^{*2}_{\min}pn}\sum_i\frac{\widehat{\alpha}^2_{i,k}}{n}\frac{\mu^4}{s^2}\sum_{k',l_1,l_2}\sum_l\bar{\boldsymbol{\beta}}^{*2}_{k',3,l}\left\{\frac{\sum_{i=1}^n\widehat{\alpha}_{i,k}\alpha^*_{i,k'}/n}{\sum_{i=1}^n\widehat{\alpha}^2_{i,k}/n}\widehat{\alpha}_{i,k}\langle\bar{\boldsymbol{\beta}}^*_{k',1},\widehat{\boldsymbol{\beta}}_{k,1}\rangle\langle\bar{\boldsymbol{\beta}}^*_{k',2},\widehat{\boldsymbol{\beta}}_{k,2}\rangle\widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}\right.\\ &\left.-\alpha^*_{i,k'}\bar{\boldsymbol{\beta}}^*_{k',1,l_1}\bar{\boldsymbol{\beta}}^*_{k',2,l_2}\right\}^2\leq\frac{4rw^{*2}_{\max}\lambda_{\max}}{w^{*2}_{\min}pn}\frac{\mu^3}{s^{1.5}}\left\{4\frac{\lambda^2_{\max}}{\lambda^2_{\min}}+1\right\}c_1^2\epsilon^2\end{aligned}$$

By vector Bernstein inequality and the fact that $\sum_{i,l_1,l_2,l}\mathbb{E}(Z_{i,l_1,l_2,l})=0$, we have,

$$P\left(\left\|\sum_{i,l_1,l_2,l}Z_{i,l_1,l_2,l}\right\|\leq\gamma\epsilon\right)\geq1-\exp\left\{\frac{1}{4}-\frac{\gamma^2}{8\frac{4rw^{*2}_{\max}\lambda_{\max}\mu^3\left\{4\frac{\lambda^2_{\max}}{\lambda^2_{\min}}+1\right\}c_1^2}{w^{*2}_{\min}pns^{1.5}}}\right\}.$$

By Assumption 6, we have $p \geq c_7 r\mu^3 w_{\max}^{*2}\log(d)/(ns^{1.5}w_{\min}^{*2}) \geq cr\mu^3 w_{\max}^{*2}\log(d)/(ns^{1.5}w_{\min}^{*2}\gamma^2)$ for some positive constant c. Here γ is the same constant as defined in (S31). Then the following inequality holds with probability at least $1-e^{1/4}/d^{10}$,

$$\left\|\frac{1}{p}\sum_{k'\neq k}\frac{w_{k'}^*}{\widehat{w}_k}\left\{\mathbf{A}\frac{\sum_{i=1}^n \widehat{\alpha}_{i,k}\alpha_{i,k'}^*/n}{\sum_{i=1}^n \widehat{\alpha}_{i,k}^2/n}\langle\bar{\boldsymbol{\beta}}_{k',1}^*,\widehat{\boldsymbol{\beta}}_{k,1}\rangle\langle\bar{\boldsymbol{\beta}}_{k',2}^*,\widehat{\boldsymbol{\beta}}_{k,2}\rangle - \mathbf{F}_{k'}\right\}\bar{\boldsymbol{\beta}}_{k',3}^*\right\| \leq \gamma\epsilon.$$

Similarly, we have that,

$$\left\|\sum_{k'\neq k}\mathbf{A}^{-1}\frac{\widehat{w}_{k'}}{\widehat{w}_k}\left\{\mathbf{G}_{k'} - \frac{\sum_i \widehat{\alpha}_{i,k}\widehat{\alpha}_{i,k'}/n}{\sum_i \widehat{\alpha}_{i,k}^2/n}\mathbf{A}\langle\bar{\widehat{\boldsymbol{\beta}}}_{k',1},\widehat{\boldsymbol{\beta}}_{k,1}\rangle\langle\bar{\widehat{\boldsymbol{\beta}}}_{k',2},\widehat{\boldsymbol{\beta}}_{k,2}\rangle\right\}\bar{\widehat{\boldsymbol{\beta}}}_{k',3}\right\| \leq \gamma\epsilon.$$

As we have shown in (S5), if $p \geq c_5\mu^4\log(d)/\{ns^2\} \geq c\mu^4\log(d)/\{ns^2\gamma^2\}$ for some positive constant c. Here γ is defined in (S31). Then with probability at least $1-2/d^{10}$, each entry of the diagonal matrix A has the lower bound $|1/p\mathbf{A}_{ll}| \geq \sum_i \widehat{\alpha}_{i,1}^2/n - \gamma$. Therefore, by the definition of IV_1 in (S30), IV_1 can be bounded as,

$$\|IV_1\| \leq \frac{2\gamma\epsilon}{\lambda_{\min}-\gamma} \leq \frac{4\gamma\epsilon}{\lambda_{\min}}.$$

Now we are ready to bound the distance in (S30),

$$\begin{aligned}\left\|\widetilde{\boldsymbol{\beta}}_{k,3} - \frac{w_k^*\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}\alpha_{i,k}^*}{\widehat{w}_k\sum_{i=1}^n \widehat{\alpha}_{i,k}^2/n}\boldsymbol{\beta}_{k,3}^*\right\| &\leq \|I_1\| + \|II_1\| + \|III_1\| + \|IV_1\| + \|V_1\| \\ \leq \frac{2\lambda_{\max}}{\lambda_{\min}}\epsilon^2 + \frac{c_1^2 r w_{\max}^*\{8\epsilon^2 + 8\xi\epsilon + 5\xi^2\epsilon\}}{\lambda_{\min}w_{\min}^*} &+ \frac{8\gamma}{\lambda_{\min}}\epsilon + 6\gamma'\epsilon + \frac{2\widetilde{C}\sigma}{w_{\min}^*\lambda_{\min}}\sqrt{\frac{s\log(d)}{np}}.\end{aligned}$$

As shown in the rank-1 case, the error of normalized $\widetilde{\boldsymbol{\beta}}_{k,3}$ can be bounded as,

$$\left|\frac{w_k^*\frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,k}\alpha_{i,k}^*}{\widehat{w}_k\frac{1}{n}\sum_{i=1}^n \widehat{\alpha}_{i,k}^2}\right|\left\|\frac{\widetilde{\boldsymbol{\beta}}_{k,3}}{\|\widetilde{\boldsymbol{\beta}}_{k,3}\|} - \boldsymbol{\beta}_{k,3}^*\right\| \leq 2\left\|\widetilde{\boldsymbol{\beta}}_{k,3} - \frac{w_k^*\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}\alpha_{i,k}^*}{\widehat{w}_k\sum_{i=1}^n \widehat{\alpha}_{i,k}^2/n}\boldsymbol{\beta}_{k,3}^*\right\|.$$

Therefore, we have that,

$$\begin{aligned}\left\|\frac{\widetilde{\boldsymbol{\beta}}_{k,3}}{\|\widetilde{\boldsymbol{\beta}}_{k,3}\|} - \boldsymbol{\beta}_{k,3}^*\right\| \;\leq\; & \frac{6\lambda_{\max}^2}{\lambda_{\min}^2}\epsilon^2 + \frac{3c_1^2\lambda_{\max}r w_{\max}^*\{8\epsilon^2 + 8\xi\epsilon + 5\xi^2\epsilon\}}{\lambda_{\min}^2 w_{\min}^*} + \frac{24\lambda_{\max}\gamma}{\lambda_{\min}^2}\epsilon + \frac{18\lambda_{\max}\gamma'}{\lambda_{\min}}\epsilon \\ & + \frac{6\widetilde{C}\lambda_{\max}\sigma}{w_{\min}^*\lambda_{\min}^2}\sqrt{\frac{s\log(d)}{np}}.\end{aligned}$$

S2.2 Error bound of the estimator from Step 2 of Algorithm 1

In the second step, we derive the error bound for the constrained estimator from Step 2 of Algorithm 1 for the general rank case. Similar to the proof in Section S1.2, we have that,

$$\left\|\widehat{\boldsymbol{\beta}}_{k,3} - \boldsymbol{\beta}^*_{k,3}\right\| \le \sqrt{10}\left\|\frac{\widetilde{\boldsymbol{\beta}}_{k,3}}{\|\widetilde{\boldsymbol{\beta}}_{k,3}\|} - \boldsymbol{\beta}^*_{k,3}\right\| \le \kappa_1\epsilon + \frac{6\sqrt{10}\widetilde{C}\lambda_{\max}\sigma}{w^*_{\min}\lambda^2_{\min}}\sqrt{\frac{s\log(d)}{np}},$$

where

$$\kappa_1 = \frac{6\sqrt{10}\lambda^2_{\max}}{\lambda^2_{\min}}\epsilon + \frac{24\sqrt{10}\lambda_{\max}}{\lambda^2_{\min}}\gamma + \frac{18\sqrt{10}\lambda_{\max}}{\lambda_{\min}}\gamma' + \frac{c_1^2\lambda_{\max}rw^*_{\max}}{\lambda^2_{\min}w^*_{\min}}\epsilon + \frac{c_1^2\lambda_{\max}rw^*_{\max}}{\lambda^2_{\min}w^*_{\min}}\xi.$$

By (S31) $\gamma < \lambda^2_{\min}/\{192\sqrt{10}\lambda_{\max}\}$, then we have $24\sqrt{10}\lambda_{\max}\gamma/\lambda^2_{\min} < 1/8$. Setting constant $\gamma' < \lambda_{\min}/\{144\sqrt{10}\lambda_{\max}\}$ we have $18\sqrt{10}\lambda_{\max}\gamma'/\lambda_{\min} < 1/8$. By Assumption 8, we have $\epsilon < \lambda^2_{\min}/\{\lambda^2_{\max}\}$ and then $6\sqrt{10}\lambda^2_{\max}\epsilon/\lambda^2_{\min} < 1/4$. By Assumption 8, we have $c_1^2\lambda_{\max}rw^*_{\max}\epsilon/\{\lambda^2_{\min}w^*_{\min}\} < 1/4$. By Assumption 9, we have $c_1^2\lambda_{\max}rw^*_{\max}\xi/\lambda^2_{\min}w^*_{\min} < 1/4$. Then it is easy to check $\kappa_1 < 1$.

Finally, we prove that the estimator $\widehat{\boldsymbol{\beta}}_{k,3}$ has the μ-mass property under the assumption that the true parameter $\boldsymbol{\beta}^*_{1,3}$ is a μ-mass vector. This would allow us to bound the error in the subsequent iteration effectively. Toward that end, we show that $\max_l\{|\widehat{\boldsymbol{\beta}}_{k,3,l}|\} \le \widetilde{c}\mu/\sqrt{s}$, where $\widetilde{c}$ is a global constant. By the expression of $\widetilde{\boldsymbol{\beta}}_{k,3,l}$, each entry can be bounded as,

$$\begin{aligned}
|\widetilde{\boldsymbol{\beta}}_{k,3,l}| &\le \frac{|w^*_k\mathbf{B}_{ll}|}{|\widehat{w}_k\mathbf{A}_{ll}|}\frac{\mu}{\sqrt{s}} + \frac{\left|\sum_{i=1}^n \frac{1}{n}\widehat{\alpha}_{i,k}\sum_{l_1,l_2}\delta_{i,l_1,l_2,l}(\mathcal{E}_{iF})_{l_1,l_2,l}\widehat{\boldsymbol{\beta}}_{k,1,l_1}\widehat{\boldsymbol{\beta}}_{k,2,l_2}\right|}{|\widehat{w}_k\mathbf{A}_{ll}|} \\
&\quad + \sum_{k'\neq k}\frac{w^*_{k'}|\mathbf{F}_{k'll}|}{|\widehat{w}_k\mathbf{A}_{ll}|}\frac{\mu}{\sqrt{s}} + \frac{|\widehat{w}_{k'}\mathbf{G}_{k'll}|}{|\widehat{w}_k\mathbf{A}_{ll}|}\frac{\mu}{\sqrt{s}} \\
&\le 2\frac{\frac{1}{n}\sum_i\widehat{\alpha}_{i,k}\alpha^*_{i,k}+\gamma}{\frac{1}{n}\sum_i\widehat{\alpha}^2_{i,k}-\gamma}\frac{\mu}{\sqrt{s}} + \sum_{k'\neq k}\left\{\frac{2w^*_{\max}}{w^*_{\min}}\frac{\sum_i\widehat{\alpha}_{i,k}\alpha^*_{i,k'}/n\epsilon+\gamma}{\frac{1}{n}\sum_i\widehat{\alpha}^2_{i,k}-\gamma}\frac{\mu}{\sqrt{s}}\right. \\
&\quad \left.+\frac{3w^*_{\max}}{w^*_{\min}}\frac{\sum_i\widehat{\alpha}_{i,k}\widehat{\alpha}_{i,k'}/n\epsilon+\gamma}{\frac{1}{n}\sum_i\widehat{\alpha}^2_{i,k}-\gamma}\frac{\mu}{\sqrt{s}}\right\} + \left\|\frac{1}{\widehat{w}_1}\boldsymbol{A}^{-1}\sum_{i=1}^n\frac{\widehat{\alpha}_{i,1}}{n}\Pi_{\Omega_i}(\mathcal{E}_{iF})\times_1\widehat{\boldsymbol{\beta}}_1\times_2\widehat{\boldsymbol{\beta}}_2\right\| \\
&\le 2\frac{\lambda_{\max}+\gamma}{\lambda_{\min}-\gamma}\frac{\mu}{\sqrt{s}} + \frac{5(r-1)w^*_{\max}}{w^*_{\min}}\frac{\lambda_{\max}+\gamma}{\lambda_{\min}-\gamma}\frac{\mu}{\sqrt{s}} + \frac{\widetilde{C}\sigma}{w^*_{\min}\lambda_{\min}}\sqrt{\frac{s\log(d)}{np}}.
\end{aligned}$$

Similar to the proof in Section S1.2, under Assumption 10, we have

$$\frac{\widetilde{C}\sigma}{w^*_{\min}\{\lambda_{\min}-\gamma\}}\sqrt{\frac{s\log(d)}{np}} \le \frac{\mu}{\sqrt{s}}.$$

By (S31) $\gamma < \lambda_{\min}/2$ then we have

$$2\frac{\lambda_{\max}+\gamma}{\lambda_{\min}-\gamma}\frac{\mu}{\sqrt{s}} \leq \frac{4\lambda_{\max}}{\lambda_{\min}}\frac{\mu}{\sqrt{s}}, \quad \frac{5(r-1)w^*_{\max}}{w^*_{\min}}\frac{\lambda_{\max}+\gamma}{\lambda_{\min}-\gamma}\frac{\mu}{\sqrt{s}} \leq \frac{10(r-1)w^*_{\max}\lambda_{\max}}{w^*_{\min}\lambda_{\min}}\frac{\mu}{\sqrt{s}}.$$

Therefore, for some constant $\widetilde{c}$, we have that,

$$\max_l \left\{ |\widehat{\boldsymbol{\beta}}_{k,3,l}| \right\} \leq \widetilde{c}\frac{\mu}{\sqrt{s}}.$$

S2.3 Error bound of the estimator from Step 4 of Algorithm 1

In the third step, we bound $\|\widehat{\boldsymbol{\beta}}_{k,4}-\boldsymbol{\beta}^*_{k,4}\|$ for each k given all other parameters $\widehat{w}_k$, $\widehat{\boldsymbol{\beta}}_{k,1}, \widehat{\boldsymbol{\beta}}_{k,2}, \widehat{\boldsymbol{\beta}}_{k,3}$ and $\widehat{w}_{k'}$, $\widehat{\boldsymbol{\beta}}'_{k',1}, \widehat{\boldsymbol{\beta}}'_{k',2}, \widehat{\boldsymbol{\beta}}'_{k',3}$ for $k' \neq k$. The closed-form estimator in step 4 of our algorithm is

$$\widehat{\boldsymbol{\beta}}_{k,4} = \left(\frac{1}{n}\sum_{i=1}^n \|\Pi_{\Omega_i}(\mathcal{A}_k)\|_F^2 \mathbf{x}_i\mathbf{x}_i^\top \right)^{-1} n^{-1}\sum_{i=1}^n \left\langle \Pi_{\Omega_i}(\widehat{\mathcal{S}}_{ik}), \widehat{\mathcal{A}}_k \right\rangle \mathbf{x}_i \tag{S32}$$

where $\widehat{\mathcal{A}}_k = \widehat{w}_k\widehat{\boldsymbol{\beta}}_{k,1} \circ \widehat{\boldsymbol{\beta}}_{k,2} \circ \widehat{\boldsymbol{\beta}}_{k,3}$, and $\widehat{\mathcal{S}}_{ik} = \mathcal{Y}_i - \sum_{k'\neq k}\widehat{w}_{k'}(\widehat{\boldsymbol{\beta}}_{k',4}^\top \mathbf{x}_i)\widehat{\boldsymbol{\beta}}_{k',1} \circ \widehat{\boldsymbol{\beta}}_{k',1} \circ \widehat{\boldsymbol{\beta}}_{k',3}$. The rank of the coefficient tensor $\mathcal{B}^*$ is r. Plugging $\mathcal{Y}_i$ into (S32), $\widehat{\boldsymbol{\beta}}_{k,4}$ can be rewritten as

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{k,4} &= \left(\frac{1}{n}\sum_{i=1}^n \|\Pi_{\Omega_i}(\mathcal{A}_k)\|_F^2 \mathbf{x}_i\mathbf{x}_i^\top \right)^{-1} \left(\frac{1}{n}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{A}^*_k), \widehat{\mathcal{A}}_k \right\rangle \mathbf{x}_i\mathbf{x}_i^\top \boldsymbol{\beta}^*_{k,4} \right. \\
&\quad \left. + \frac{1}{n}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \widehat{\mathcal{A}}_k \right\rangle \mathbf{x}_i + \frac{1}{n}\sum_i \sum_{k'\neq k, k'\in[r]} \left\langle \Pi_{\Omega_i}(\mathcal{A}^*_{k'} - \widehat{\mathcal{A}}_{k'}), \widehat{\mathcal{A}}_k \right\rangle \mathbf{x}_i\mathbf{x}_i^\top \boldsymbol{\beta}^*_{k,4} \right).
\end{aligned}$$

Therefore, we have,

$$\begin{aligned}
\|\widehat{\boldsymbol{\beta}}_{k,4} - \boldsymbol{\beta}^*_{k,4}\| &\leq \left\| \left(\frac{1}{n}\sum_{i=1}^n \|\Pi_{\Omega_i}(\mathcal{A}_k)\|_F^2 \mathbf{x}_i\mathbf{x}_i^\top \right)^{-1} \right\| \left(\left\| \frac{1}{n}\sum_i \left\langle \Pi_{\Omega_i}(\mathcal{E}_i), \widehat{\mathcal{A}}_k \right\rangle \mathbf{x}_i \right\| \right. \\
&\quad \left. + \left\| \frac{1}{n}\sum_i \sum_{k\in[r]} \left\langle \Pi_{\Omega_i}(\mathcal{A}^*_{k'} - \widehat{\mathcal{A}}_{k'}), \widehat{\mathcal{A}}_k \right\rangle \mathbf{x}_i\mathbf{x}_i^\top \boldsymbol{\beta}^*_{k,4} \right\| \right).
\end{aligned}$$

By (S25), (S26) and (S27), we obtain that,

$$\begin{aligned}
\|\widehat{\boldsymbol{\beta}}_{k,4} - \boldsymbol{\beta}^*_{k,4}\| &\leq \frac{8}{\lambda_{\min} w^{*2}_{\min}} \left[r\left\{6c_2 + \gamma\right\} w^{*2}_{\max}\epsilon + \frac{3\widetilde{C}_2\sigma w^*_{\max}}{2}\sqrt{\frac{qs\log(d)}{np}} \right] \\
&\leq \kappa_2\epsilon + \frac{12\widetilde{C}_2\sigma w^*_{\max}}{\lambda_{\min} w^{*2}_{\min}}\sqrt{\frac{qs\log(d)}{np}}
\end{aligned} \tag{S33}$$

where $\kappa_2 = c_2 w^{*2}_{\max} r/(\lambda_{\min} w^{*2}_{\min})$. Here we use the fact that $\gamma < c_2$ by (S31).

S2.4 Proof of the theorem

Finally, we provide the error rate after t iterations, by iteratively applying the error bound from each step. We have shown that, with probability at least $1 - 1/d^9$,

$$\|\widehat{\boldsymbol{\beta}}_{k,3}^{(1)} - \boldsymbol{\beta}_{k,3}^*\| \;\leq\; \kappa_1\epsilon + \frac{6\sqrt{10}\widetilde{C}\lambda_{\max}\sigma}{w_{\min}^*\lambda_{\min}^2}\sqrt{\frac{s\log(d)}{np}}.$$

The error bound in (S33) implies that with probability at least $1 - 2/d^9$ we have

$$\|\widehat{\boldsymbol{\beta}}_{k,4}^{(1)} - \boldsymbol{\beta}_{k,4}^*\| \;\leq\; \kappa_2\kappa_1\epsilon + \kappa_2\frac{6\sqrt{10}\widetilde{C}\lambda_{\max}\sigma}{w_{\min}^*\lambda_{\min}^2}\sqrt{\frac{s\log(d)}{np}} + \frac{12\widetilde{C}_2\sigma w_{\max}^*}{\lambda_{\min}w_{\min}^{*2}}\sqrt{\frac{qs\log(d)}{np}},$$

The contraction coefficient is

$$\kappa = \kappa_1\kappa_2 = \left[\frac{6\sqrt{10}\lambda_{\max}^2}{\lambda_{\min}^2}\epsilon + \frac{24\sqrt{10}\lambda_{\max}}{\lambda_{\min}^2}\gamma + \frac{c_1^2\lambda_{\max}rw_{\max}^*}{\lambda_{\min}^2w_{\min}^*}\epsilon + \frac{c_1^2\lambda_{\max}rw_{\max}^*}{\lambda_{\min}^2w_{\min}^*}\xi\right]\frac{c_2w_{\max}^{*2}r}{\lambda_{\min}w_{\min}^{*2}}.$$

Here, by Assumption 8 and Assumption 9, we have

$$\frac{6\sqrt{10}\lambda_{\max}^2}{\lambda_{\min}^2}\epsilon\cdot\frac{c_2w_{\max}^{*2}r}{\lambda_{\min}w_{\min}^{*2}} < \frac{1}{4},\; \frac{c_1^2\lambda_{\max}rw_{\max}^*}{\lambda_{\min}^2w_{\min}^*}\epsilon\cdot\frac{c_2w_{\max}^{*2}r}{\lambda_{\min}w_{\min}^{*2}} < \frac{1}{4},\; \frac{c_1^2\lambda_{\max}rw_{\max}^*}{\lambda_{\min}^2w_{\min}^*}\xi\cdot\frac{c_2w_{\max}^{*2}r}{\lambda_{\min}w_{\min}^{*2}} < \frac{1}{4},$$

and by (S31) we have $\gamma < \lambda_{\min}^3 w_{\min}^{*2}/\{96\sqrt{10}\lambda_{\max}c_2w_{\max}^{*2}r\}$ and

$$\frac{24\sqrt{10}\lambda_{\max}}{\lambda_{\min}^2}\gamma\cdot\frac{c_2w_{\max}^{*2}r}{\lambda_{\min}w_{\min}^{*2}} < \frac{1}{4}.$$

Therefore, the contraction coefficient $\kappa < 1$.

The error bound after t iterations is, with probability at least $1 - (t+1)/d^9$

$$\begin{aligned}
&\max\left\{\max_k|\widehat{w}_k^{(t)} - w_k^*|,\; \max_{k,j}\left(\|\widehat{\boldsymbol{\beta}}_{k,j}^{(t)} - \boldsymbol{\beta}_{k,j}^*\|_2\right)\right\}\\
\leq\;& \kappa^t\epsilon + \frac{1-\kappa^t}{1-\kappa}\frac{6\sqrt{10}\widetilde{C}\lambda_{\max}\sigma}{w_{\min}^*\lambda_{\min}^2}\sqrt{\frac{s\log(d)}{np}} + \frac{1-\kappa^{t-1}}{1-\kappa}\frac{12\widetilde{C}_2\sigma w_{\max}^*}{\lambda_{\min}w_{\min}^{*2}}\sqrt{\frac{qs\log(d)}{np}}\\
\leq\;& \kappa^t\epsilon + \frac{1}{1-\kappa}\max\left\{C_2'\frac{\sigma}{w_{\min}^*}\sqrt{\frac{s\log(d)}{np}},\, \frac{C_2''\sigma w_{\max}^*}{w_{\min}^{*2}}\sqrt{\frac{s\log(d)}{np}}\right\}.\\
\leq\;& \kappa^t\epsilon + \frac{C_2}{1-\kappa}\frac{\sigma w_{\max}^*}{w_{\min}^{*2}}\sqrt{\frac{s\log(d)}{np}}
\end{aligned}$$

where $C_2' = 6\sqrt{10}\widetilde{C}\lambda_{\max}/\lambda_{\min}^2$, $C_2'' = 12\widetilde{C}_2\sqrt{q}/\lambda_{\min}$ and $C_2 = C_2' + C_2''$. This completes the proof of Theorem 2. □

S3 Proof of Proposition 1

We divide the proof into three steps. In Step 1, we show that there exists at least one trial $1 \le \tau \le L$ such that $\boldsymbol{\beta}^*_{1,1}$ is the top left singular vector of the population version of $\mathcal{T} \times_3 \widetilde{\boldsymbol{g}}^\tau_1$. In Step 2, we prove that the top singular vector $\mathbf{v}^\tau_1$ is close to $\boldsymbol{\beta}^*_{1,1}$. We obtain the bound for $\mathbf{v}_1$ as an example, while the bounds for $\mathbf{v}_2$ and $\mathbf{v}_3$ can be derived similarly. In Step 3, we prove that $\widehat{\boldsymbol{\beta}}^{(0)}_{1,4}$ is close to true factor.

Step 1: Let $\mathcal{T}^* = w^*_1 \sum_i n^{-1} \boldsymbol{\beta}^{*\top}_{1,4} \mathbf{x}_i \boldsymbol{\beta}^*_{1,1} \circ \boldsymbol{\beta}^*_{1,2} \circ \boldsymbol{\beta}^*_{1,3}$, and $\gamma^{*\tau} = w^*_1 \sum_i n^{-1} \boldsymbol{\beta}^{*\top}_{1,4} \mathbf{x}_i \langle \boldsymbol{\beta}^*_{1,3}, \widetilde{\mathbf{g}}^\tau_1 \rangle$. Then,

$$\mathcal{T}^* \times_3 \widetilde{g}^\tau = w^*_1 \sum_i \frac{1}{n} \boldsymbol{\beta}^{*\top}_{1,4} \mathbf{x}_i \langle \boldsymbol{\beta}^*_{1,3}, \widetilde{g}^\tau_1 \rangle \boldsymbol{\beta}^*_{1,1} \circ \boldsymbol{\beta}^*_{1,2} = \gamma^{*\tau} \boldsymbol{\beta}^*_{1,1} \boldsymbol{\beta}^{*\top}_{1,2}. \tag{S34}$$

Therefore, $\gamma^{*\tau}$ is the singular value of $T^* \times_3 \widetilde{g}^\tau$ when rank $r = 1$.

Next, we need to prove that there exists some τ such that $\gamma^{*\tau}$ is sufficiently separated from 0. Since $\widetilde{\mathbf{g}}^\tau_1 = \mathbf{U}_1 \mathbf{U}_1^\top \mathbf{g}^\tau_1$, we have,

$$\gamma^{*\tau} = w^*_1 \sum_i \frac{1}{n} \boldsymbol{\beta}^{*\top}_{1,4} \mathbf{x}_i \langle \mathbf{U}_1 \mathbf{U}_1^\top \boldsymbol{\beta}^*_{1,3}, \mathbf{g}^\tau_1 \rangle,$$

where, recall that, $\mathbf{U}_1$ is the rank-1 eigen-decomposition of $\mathbf{B}_1$, and $\mathbf{B}_1 = \Pi_{\text{off-diag}}(\mathbf{A}_1 \mathbf{A} \mathbf{A}_1^\top)$ as defined in Section 4.3. Given that $\mathbf{g}^\tau_1$ is a standard Gaussian vector, $\gamma^{*\tau}$ is zero-mean Gaussian conditional on Ω_i and $\mathcal{E}_i$, with the standard deviation, $w^*_1 \left| \sum_i n^{-1} \boldsymbol{\beta}^{*\top}_{1,4} \mathbf{x}_i \right| \|\mathbf{U}_1 \mathbf{U}_1^\top \boldsymbol{\beta}^*_{1,3}\|$.

Without loss of generity, let $|\gamma^{*1}| \ge |\gamma^{*2}| \ge \cdots \ge |\gamma^{*L}|$. By Lemma 5 and $r = 1$, for any fixed small constant $\delta > 0$, with probability at least $1 - \delta$, we have,

$$\gamma^{*1} \gtrsim w^*_1 \left| \sum_i n^{-1} \boldsymbol{\beta}^{*\top}_{1,4} \mathbf{x}_i \right| \|\mathbf{U}_1 \mathbf{U}_1^\top \boldsymbol{\beta}^*_{1,3}\|, \tag{S35}$$

which holds due to the condition that $L \ge C'_1$ for some constant C'_1.

Let $\mathbf{A}^*_1$ be the mode-3 matricization of $\mathcal{T}^*$. Then,

$$\mathbf{A}^*_1 = w^*_1 \sum_i \frac{1}{n} \boldsymbol{\beta}^{*\top}_{1,4} \mathbf{x}_i \boldsymbol{\beta}^*_{1,3} (\boldsymbol{\beta}^*_{1,1} \otimes \boldsymbol{\beta}^*_{1,2})^\top \in \mathbb{R}^{d_3 \times d_1 d_2}.$$

Let $\mathbf{U}^*_1$ be the basis of the column space of $\mathbf{A}^*_1 \mathbf{A}^{*\top}_1$. Intuitively, the space spanned by $\mathbf{U}_1$ is close to the space spanned by the true tensor factor. Then, by Lemma 6, the bound in (S35) can be simplified as,

$$\gamma^{*1} \gtrsim w^*_1 \sqrt{1 - \|\mathbf{U}_1 - \mathbf{U}^*_1\|^2}, \tag{S36}$$

since $\left|\sum_i n^{-1}\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i\right| \geq C_2'$. Next, we show that $\|\mathbf{U}_1 - \mathbf{U}_1^*\|$ is small. Following the proof of Theorem 1 in Cai et al. (2019a), we have,

$$\|\mathbf{U}_1 - \mathbf{U}_1^*\| \leq \sqrt{2}\|\mathbf{U}_1\mathbf{U}_1^\top - \mathbf{U}_1^*\mathbf{U}_1^{*\top}\| \leq \frac{\left\|\mathbf{A}_1^*\mathbf{A}_1^{*\top} - \mathbf{A}_1\mathbf{A}_1^\top\right\|}{\sigma(\mathbf{A}_1^*)^2}, \tag{S37}$$

where $\sigma(\mathbf{A}_1^*)$ is the singular value of $\mathbf{A}_1^*$. Since $\mathbf{A}_1^*$ is a rank-1 matrix, we have $\sigma(\mathbf{A}_1^*) = w_1^* \sum_i n^{-1}\boldsymbol{\beta}_{1,4}^{*\top}$. By Lemma 1 in Cai et al. (2019a), we have,

$$\begin{aligned}\left\|\mathbf{A}_1^*\mathbf{A}_1^{*\top} - \mathbf{A}_1\mathbf{A}_1^\top\right\| \lesssim &\left\{\frac{\|\mathbf{A}_1^*\|_{2,\infty} + \sigma\sqrt{\widetilde{d}}}{\sqrt{p}} + \frac{\|\mathbf{A}_1^{*\top}\|_{2,\infty} + \sigma\sqrt{\widetilde{d}}}{\sqrt{p}}\right\} \times \\ &\left\{\frac{\|\mathbf{A}_1^{*\top}\|_{2,\infty} + \sigma\sqrt{\widetilde{d}}}{\sqrt{p}} + \|\mathbf{A}_1^{*\top}\|_{2,\infty}\right\}\log(\widetilde{d}) \\ &+ \frac{\|\mathbf{A}_1^{*\top}\|_{2,\infty} + \sigma\sqrt{\widetilde{d}}}{\sqrt{p}}\sqrt{\log(\widetilde{d})}\|\mathbf{A}_1^*\| + \|\mathbf{A}_1^*\|_{2,\infty}^2,\end{aligned} \tag{S38}$$

where $\|\mathbf{A}\|_{2,\infty} = \max_{i\in[m]} \|\mathbf{A}_{i,:}\|_2$ for any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, and $\widetilde{d} = \max\{d_3, d_1 d_2\}$. By the definition of $\mathbf{A}_1^*$, we have,

$$\begin{aligned}\|\mathbf{A}_1^*\|_{2,\infty} =& \max_{l\in[d_3]}\left\|\boldsymbol{\beta}_{1,3,l}^* w_1^* \sum_i \frac{1}{n}\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i(\boldsymbol{\beta}_{1,1}^* \otimes \boldsymbol{\beta}_{1,2}^*)^\top\right\|_2 \\ \leq& \max_{l\in[d_3]}|\boldsymbol{\beta}_{1,3,l}^*|\left|w_1^* \sum_i \frac{1}{n}\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i\right| \left\|(\boldsymbol{\beta}_{1,1}^* \otimes \boldsymbol{\beta}_{1,2}^*)^\top\right\|_2 \leq c_1 w_1^* \frac{\mu}{\sqrt{s}},\end{aligned} \tag{S39}$$

where the last inequality is due to Assumption 1 (i) and (iii). In addition,

$$\|\mathbf{A}_1^{*\top}\|_{2,\infty} \leq c_1 w_1^* \max_{l_1\in[d_1], l_2\in[d_2]}\left|\boldsymbol{\beta}_{1,1,l_1}^*\boldsymbol{\beta}_{1,2,l_2}^*\right| \leq c_1 w_1^* \frac{\mu^2}{s}. \tag{S40}$$

Since $\mathbf{A}_1^*$ is a rank-1 matrix, we have,

$$\|\mathbf{A}_1^*\|_2 = \|\mathbf{A}_1^*\|_F = \left|w_1^* \sum_i \frac{1}{n}\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i\right| \|\boldsymbol{\beta}_{1,3}^*\| \left\|(\boldsymbol{\beta}_{1,1}^* \otimes \boldsymbol{\beta}_{1,2}^*)^\top\right\| \leq c_1 w_1^*. \tag{S41}$$

Combining (S39), (S40) and (S41), (S38) can be simplified as

$$\left\|\mathbf{A}_1^*\mathbf{A}_1^{*\top} - \mathbf{A}_1\mathbf{A}_1^\top\right\| \lesssim \frac{w_1^{*2}\mu^3}{s^{1.5}p}\log(\widetilde{d}) + \frac{\sigma^2\widetilde{d}\log(\widetilde{d})}{p} + \frac{w_1^{*2}\mu^2}{s}\sqrt{\frac{\log(\widetilde{d})}{p}} + w_1^*\sigma\sqrt{\frac{\widetilde{d}\log(\widetilde{d})}{p}} + \frac{w_1^{*2}\mu^2}{s}.$$

Therefore, (S37) can be simplified as

$$\|\mathbf{U}_1 - \mathbf{U}_1^*\| \lesssim \frac{\mu^3}{s^{1.5}p}\log(\widetilde{d}) + \frac{\sigma^2\widetilde{d}\log(\widetilde{d})}{w_1^{*2}p} + \frac{\mu^2}{s}\sqrt{\frac{\log(\widetilde{d})}{p}} + \frac{\sigma}{w_1^*}\sqrt{\frac{\widetilde{d}\log(\widetilde{d})}{p}} + \frac{\mu^2}{s}.$$

Therefore, for any arbitrary small constant $\delta > 0$, with probability greater than $1 - \delta$,

$$\gamma^{*1} \gtrsim w_1^*. \tag{S42}$$

We have shown that there exists some $\tau \in [L]$, such that $\gamma^{*\tau} \gtrsim w_1^*$. This means that $\boldsymbol{\beta}_{1,3}^*$ exhibits the largest correlation with the projected $\boldsymbol{g}_1$, which further implies that $\boldsymbol{\beta}_{1,1}^*$ is the largest left singular vector of $\mathcal{T}^* \times_3 \widetilde{\boldsymbol{g}}_1^\tau$.

Step 2: Recall that $\mathbf{v}_1^\tau$ is the top left singular vector of $\mathbf{M}^\tau$, where

$$\mathbf{M}^\tau = p^{-1}\mathcal{T} \times_3 \widetilde{\mathbf{g}}_1^\tau = \mathcal{T}^* \times_3 \widetilde{\mathbf{g}}_1^\tau + \{\mathcal{T} - \mathcal{T}^*\} \times_3 \widetilde{\mathbf{g}}_1^\tau = \underbrace{\gamma^{*\tau}\boldsymbol{\beta}_{1,1}^*\boldsymbol{\beta}_{1,2}^{*\top}}_{M^{*\tau}} + \{T - T^*\} \times_3 \widetilde{\mathbf{g}}_1^\tau,$$

and the second equality is due to the definition of $\gamma^{*\tau}$ in (S34). By Wedin's Theorem,

$$\|\mathbf{v}_1^\tau - \boldsymbol{\beta}_{1,1}^*\| \le \frac{\|(\mathbf{M}^\tau - \mathbf{M}^{*\tau})\boldsymbol{\beta}_{1,1}^*\|_2}{\gamma^{*\tau} - \|\mathbf{M}^\tau - \mathbf{M}^{*\tau}\|}. \tag{S43}$$

Next, we bound $\|(\mathbf{M}^\tau - \mathbf{M}^{*\tau})\boldsymbol{\beta}_{1,1}^*\|_2$ and $\|\mathbf{M}^\tau - \mathbf{M}^{*\tau}\|$, respectively.

First, to bound $\|\mathbf{M}^\tau - \mathbf{M}^{*\tau}\|$, recall that $\widetilde{\mathbf{g}}_1^\tau = \mathbf{U}_1\mathbf{U}_1^\top\mathbf{g}_1^\tau$. Define $\mathbf{g}_1^{*\tau} = \mathbf{U}_1^*\mathbf{U}_1^{*\top}\mathbf{g}_1^\tau$, and decompose,

$$\mathbf{M}^\tau - \mathbf{M}^{*\tau} = \{p^{-1}\mathcal{T} - \mathcal{T}^*\} \times_3 \widetilde{\mathbf{g}}_1^\tau = \underbrace{\{p^{-1}\mathcal{T} - \mathcal{T}^*\} \times_3 \mathbf{g}_1^{*\tau}}_{\mathbf{V}_1} + \underbrace{\{p^{-1}\mathcal{T} - \mathcal{T}^*\} \times_3 \{\widetilde{\mathbf{g}}_1^\tau - \mathbf{g}_1^{*\tau}\}}_{\mathbf{V}_2}.$$

We next bound the two terms $\mathbf{V}_1$ and $\mathbf{V}_2$, respectively.

Note that $\mathbf{V}_1$ is a zero mean random matrix in $\mathbb{R}^{d_1 \times d_2}$ with independent entries

$$\mathbf{V}_{1,l_1,l_2} = \sum_{l_3 \in [d_3], i \in [n]} n^{-1}\boldsymbol{g}_{1,l_3}^{*\tau}\left[\{p^{-1}\delta_{i,l_1,l_2,l_3} - 1\}\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i w_1^*\boldsymbol{\beta}_{1,1,l_1}^*\boldsymbol{\beta}_{1,2,l_2}^*\boldsymbol{\beta}_{1,3,l_3}^* + p^{-1}\mathcal{E}_{i,l_1,l_2,l_3}\right]$$

By Lemma D.4 in Cai et al. (2019b), with probability $1 - O(d^{-10})$, we have,

$$\|\mathbf{V}_1\| \lesssim \left\{\mathbf{V}_3\log(d) + \mathbf{V}_4\sqrt{\log(d)}\right\}\log(d) + \sqrt{\mathbf{V}_5\log(d)},$$

where

$$\begin{aligned}
\mathbf{V}_3 &= \max_{l_3 \in [d_3], i \in [n]} \left|n^{-1}\boldsymbol{g}_{1,l_3}^{*\tau}\left[\{p^{-1}\delta_{i,l_1,l_2,l_3} - 1\}\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i w_1^*\boldsymbol{\beta}_{1,1,l_1}^*\boldsymbol{\beta}_{1,2,l_2}^*\boldsymbol{\beta}_{1,3,l_3}^* + p^{-1}\mathcal{E}_{i,l_1,l_2,l_3}\right]\right| \\
&\lesssim \frac{\|\boldsymbol{g}_1^{*\tau}\|_\infty}{pn}\left\{w_1^*\frac{\mu^3}{s^{1.5}} + \sigma\sqrt{\log(d)}\right\}. \\
\mathbf{V}_4 &= \left\{\mathbb{E}(E_{l_1,l_2}^2)\right\}^{1/2} \lesssim \frac{1}{\sqrt{pn}}\left\{\|\boldsymbol{g}_1^{*\tau}\|_\infty^2 w_1^{*2}\frac{\mu^4}{s^2} + \|\boldsymbol{g}_1^{*\tau}\|_2^2\sigma^2\right\}^{1/2}. \\
\mathbf{V}_5 &= \max\left\{\max_{l_1}\sum_{l_2}\mathbb{E}(E_{l_1,l_2}^2), \max_{l_2}\sum_{l_1}\mathbb{E}(E_{l_1,l_2}^2)\right\} \le \frac{1}{pn}\left\{w_1^{*2}\|\boldsymbol{g}_1^{*\tau}\|_\infty^2\frac{\mu^2}{s} + \|\boldsymbol{g}_1^{*\tau}\|_2^2\sigma^2 s\right\}.
\end{aligned}$$

Therefore,

$$
\begin{aligned}
\|\mathbf{V}_1\| &\lesssim \frac{\|\boldsymbol{g}_1^{*\tau}\|_\infty \sigma \log^{2.5}(d)}{pn} + \frac{\|\boldsymbol{g}_1^{*\tau}\|_\infty w_1^* \mu \sqrt{\log(d)}}{\sqrt{pns}} + \frac{\|\boldsymbol{g}_1^{*\tau}\|_2 \sigma \sqrt{s\log(d)}}{\sqrt{pn}} \\
&\lesssim \frac{\mu\sigma\log^3(d)}{pn\sqrt{s}} + \frac{w_1^*\mu^2\log(d)}{\sqrt{pns^2}} + \frac{\sigma\sqrt{s}\log(d)}{\sqrt{pn}} \\
&\lesssim \frac{w_1^*\mu^2\log(d)}{\sqrt{pns^2}} + \frac{\sigma\sqrt{s}\log(d)}{\sqrt{pn}},
\end{aligned} \tag{S44}
$$

where the first inequality and the third inequality are satisfied under the assumptions of Proposition 1. The second inequality holds due to the fact that, with probability at least $1-O(d^{-20})$, $\|\boldsymbol{g}_1^{*\tau}\|_\infty = \|\mathbf{U}_1^*\mathbf{U}_1^{*\top}\mathbf{g}_1| \lesssim \|\mathbf{U}_1^*\|_{2,\infty}\sqrt{\log(d)} \lesssim \mu\sqrt{\log(d)/s}$, and $\|\boldsymbol{g}_1^{*\tau}\|_2 \lesssim \|\mathbf{U}_1^*\|_F\sqrt{\log(d)} \lesssim \sqrt{\log(d)}$.

Next, we turn to $\mathbf{V}_2$, and have that,

$$
\|\mathbf{V}_2\| \le \|\{p^{-1}\mathcal{T} - \mathcal{T}^*\} \times_3 \{\widetilde{\mathbf{g}}_1^\tau - \mathbf{g}_1^{*\tau}\}\| \le \|p^{-1}\mathcal{T} - \mathcal{T}^*\| \|\widetilde{\mathbf{g}}_1^\tau - \mathbf{g}_1^{*\tau}\|_2.
$$

Since $\mathbf{U}_1\mathbf{U}_1^\top$ and $\mathbf{U}_1^*\mathbf{U}_1^{*\top}$ are both rank-1 matrices, by the standard result of Gaussian random vectors, we have, with probability at least $1-O(d^{-12})$,

$$
\|\widetilde{\mathbf{g}}_1^\tau - \mathbf{g}_1^{*\tau}\|_2 \lesssim \|\mathbf{U}_1\mathbf{U}_1^\top - \mathbf{U}_1^*\mathbf{U}_1^{*\top}\|\sqrt{d} \ll 1.
$$

Moreover, we have that,

$$
\|p^{-1}\mathcal{T} - \mathcal{T}^*\| \le \left\| \frac{w_1^*}{pn} \sum_i \Pi_{\Omega_i}\left(\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i \boldsymbol{\beta}_{1,1}^* \circ \boldsymbol{\beta}_{1,2}^* \circ \boldsymbol{\beta}_{1,3}^*\right) - \mathcal{T}^* \right\| + \left\| \frac{1}{pn}\sum_i \Pi_{\Omega_i}(\mathcal{E}_i) \right\|.
$$

By Corollary D.3 in Cai et al. (2019b), with probability at least $1-O(d^{-10})$, we have,

$$
\left\| \frac{1}{pn}\sum_i \Pi_{\Omega_i}(\mathcal{E}_i) \right\| \le \frac{1}{pn}\sum_i \|\Pi_{\Omega_i}(\mathcal{E}_i)\| \lesssim \frac{\sigma\log^{7/2}(d)}{p} + \sigma\sqrt{\frac{d\log^5(d)}{p}} \tag{S45}
$$

By Corollary D.4 in Cai et al. (2019b), with probability at least $1-O(d^{-10})$, we have,

$$
\left\| \frac{w_1^*}{pn} \sum_i \Pi_{\Omega_i}\left(\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i \boldsymbol{\beta}_{1,1}^* \circ \boldsymbol{\beta}_{1,2}^* \circ \boldsymbol{\beta}_{1,3}^*\right) - \mathcal{T}^* \right\| \le V_6 \log^3(d) + \sigma_{\text{mode}} \log^{5/2}(d),
$$

where

$$
\begin{aligned}
V_6 &= \max_{l_1\in[d_1],l_2\in[d_2],l_3\in[d_3]} \left| \left(\frac{w_1^*}{pn} \sum_i \Pi_{\Omega_i}\left(\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i \boldsymbol{\beta}_{1,1}^* \circ \boldsymbol{\beta}_{1,2}^* \circ \boldsymbol{\beta}_{1,3}^*\right) - T^* \right)_{l_1 l_2 l_3} \right| \lesssim \frac{w_1^*\mu^3}{s^{1.5}p}, \\
\sigma^2_{\text{mode}} &= \max_{l_1\in[d_1],l_2\in[d_2]} \sum_{l_3\in[d_3],i\in[n]} \mathbb{E}\left[\left(\frac{w_1^*}{pn} \sum_i \Pi_{\Omega_i}\left(\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i \boldsymbol{\beta}_{1,1}^* \circ \boldsymbol{\beta}_{1,2}^* \circ \boldsymbol{\beta}_{1,3}^*\right) - T^* \right)^2_{l_1 l_2 l_3} \right] \lesssim \frac{w_1^{*2}\mu^4}{pns^2}.
\end{aligned}
$$

Therefore,

$$\left\|\frac{w_1^*}{pn}\sum_i \Pi_{\Omega_i}\left(\boldsymbol{\beta}_{1,4}^{*\top}\mathbf{x}_i\boldsymbol{\beta}_{1,1}^*\circ\boldsymbol{\beta}_{1,2}^*\circ\boldsymbol{\beta}_{1,3}^*\right)-T^*\right\| \lesssim \frac{w_1^*\mu^3\log^3(d)}{s^{1.5}p}+\frac{w_1^*\mu^2\log^{5/2}(d)}{\sqrt{pns}}. \qquad \text{(S46)}$$

Combining (S45) and (S46), we have,

$$\|\mathbf{V}_2\| \lesssim \|p^{-1}T-T^*\| \lesssim \frac{w_1^*\mu^3\log^3(d)}{s^{1.5}p}+\frac{w_1^*\mu^2\log^{5/2}(d)}{\sqrt{pns}}+\frac{\sigma\log^{7/2}(d)}{p}+\sigma\sqrt{\frac{d\log^5(d)}{p}}.$$

Combining the bounds of $\mathbf{V}_1$ and $\mathbf{V}_2$, we obtain that,

$$\|\mathbf{M}^\tau-\mathbf{M}^{*\tau}\| \leq \|\mathbf{V}_1\|+\|\mathbf{V}_2\| \ll w_1^*.$$

Second, to bound $\|(\mathbf{M}^\tau-\mathbf{M}^{*\tau})\boldsymbol{\beta}_{1,1}^*\|_2$, we have, by the definition of the operator norm,

$$\|(\mathbf{M}^\tau-\mathbf{M}^{*\tau})\boldsymbol{\beta}_{1,1}^*\|_2 \leq \|\{p^{-1}\mathcal{T}-\mathcal{T}^*\}\times_1\boldsymbol{\beta}_{1,1}^*\times_3\widetilde{\mathbf{g}}_1^\tau\| \leq \|\{p^{-1}\mathcal{T}-\mathcal{T}^*\}\times_1\boldsymbol{\beta}_{1,1}^*\|\|\widetilde{\mathbf{g}}_1^\tau\|.$$

Similar to the proof of (S44), we have that,

$$\begin{aligned}\|\{p^{-1}\mathcal{T}-\mathcal{T}^*\}\times_1\boldsymbol{\beta}_{1,1}^*\| &\lesssim \frac{\|\boldsymbol{\beta}_{1,1}^*\|_\infty\sigma\log^{2.5}(d)}{pn}+\frac{\|\boldsymbol{\beta}_{1,1}^*\|_\infty w_1^*\mu\sqrt{\log(d)}}{\sqrt{pns}}+\frac{\|\boldsymbol{\beta}_{1,1}^*\|_2\sigma\sqrt{s\log(d)}}{\sqrt{pn}}\\ &\lesssim \frac{\mu\sigma\log^{2.5}(d)}{pn\sqrt{s}}+\frac{\mu^2w_1^*\sqrt{\log(d)}}{\sqrt{pns^2}}+\frac{\sigma\sqrt{s\log(d)}}{\sqrt{pn}}\\ &\lesssim \frac{\mu^2w_1^*\sqrt{\log(d)}}{\sqrt{pns^2}}+\frac{\sigma\sqrt{s\log(d)}}{\sqrt{pn}}.\end{aligned}$$

We have proved in (S42) that $\gamma^{*\tau}\gtrsim w_1^*$. Besides, we have $\|\mathbf{M}^\tau-\mathbf{M}^{*\tau}\|\ll w_1^*$. Therefore, the difference in (S43) becomes,

$$\|\mathbf{v}_1^\tau-\boldsymbol{\beta}_{1,1}^*\| \lesssim \mu^2\sqrt{\frac{\log(d)}{pns^2}}+\frac{\sigma}{w_1^*}\sqrt{\frac{s\log(d)}{pn}}. \qquad \text{(S47)}$$

Next, we show that $\mathbf{v}_1^\tau$ is $c\mu-$mass, where c is a general constant. We have that,

$$\max_{l_1\in[d_1]}|\mathbf{v}_{1,l_1}^\tau| \leq \max_{l_1\in[d_1]}|\boldsymbol{\beta}_{1,1,l_1}^*|+c'\mu^2\sqrt{\frac{\log(d)}{pns^2}}+c''\frac{\sigma}{w_1^*}\sqrt{\frac{s\log(d)}{pn}} \leq \frac{\mu}{\sqrt{s}}+\frac{c_1'\mu}{\sqrt{s}}+\frac{c_1''\mu}{\sqrt{s}},$$

where the first inequality holds by (S47). The second inequality holds due to the assumption on the sample size.

Step 3: Now we develop the bound for $\widehat{\boldsymbol{\beta}}_{1,4}$. We have shown in (S28) that, if $p \gtrsim \mu^3 \log(d)/\{ns^{1.5}\}$, with a high probability,

$$\|\widehat{\boldsymbol{\beta}}_{1,4} - \boldsymbol{\beta}^*_{1,4}\| \le \kappa\epsilon + \frac{\widetilde{C}_2\sigma}{\lambda_{\min} w_1^*}\sqrt{\frac{qs\log(d)}{np}},$$

where κ is some constant, and ϵ is the estimation error of $\mathbf{v}_j$ in (S47). Therefore, the final error rate of $\widehat{\boldsymbol{\beta}}_{1,4}$ is,

$$\max_j \left\{\|\widehat{\boldsymbol{\beta}}^{(0)}_{1,j} - \boldsymbol{\beta}^*_{1,j}\|_2\right\} \lesssim \mu^2\sqrt{\frac{\log(d)}{pns^2}} + \frac{\sigma}{w_1^*}\sqrt{\frac{s\log(d)}{np}} \tag{S48}$$

This completes the proof of Proposition 1. □

S4 Auxiliary lemmas

Lemma 1. *(Yuan and Zhang, 2013, Lemma 12) Consider a sparse vector* $\mathbf{x}$ *with supp*$(\mathbf{x}) = F_{\mathbf{x}}$ *and* $F_{\mathbf{x}} = d_0$*. Let* $F_y = supp(\mathbf{y}, s)$*. If* $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$*, then*

$$\left|\texttt{Truncate}(\mathbf{y}, F_y)^\top \mathbf{x}\right| \ge |\mathbf{y}^\top\mathbf{x}| - \sqrt{\frac{d_0}{s}}\min\left[\sqrt{1-(\mathbf{y}^\top\mathbf{x})^2}, \left(1+\sqrt{\frac{d_0}{s}}\right)\left\{1-(\mathbf{y}^\top\mathbf{x})^2\right\}\right].$$

Lemma 2. *(Sun et al., 2017, Lemma S.6.2) For any tensor* $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times d_3}$ *and an index set* $F = F_1 \circ F_2 \circ F_3$ *with* $F_i \subseteq \{1, \cdots d_i\}$*, if* $\mathcal{T} = \sum_{i\in[R]} w_i \mathbf{a}_i \circ \mathbf{b}_i \circ \mathbf{c}_i$*, then*

$$\mathcal{T}_F = \sum_{i\in[R]} w_i \texttt{Truncate}(\mathbf{a}_i, F_1) \circ \texttt{Truncate}(\mathbf{b}_i, F_2) \circ \texttt{Truncate}(\mathbf{c}_i, F_3).$$

Lemma 3. *(Zhang et al., 2019a, Lemma 4) For a scaler* a^* *and any sequences* $(a_1, \cdots, a_S)$*,*

$$\sum_{i=1}^S (a_i - a^*)^2 \ge S\left(S^{-1}\sum_{i=1}^S a_i - s^*\right)^2$$

where the equality holds if and only $a_1 = a_2 = \cdots a_S$*.*

Lemma 4. *(Xia et al., 2020, Lemma 5) Let* $X_1, \cdots, X_n \in \mathbb{R}^{m_1\times m_2}$ *be random matrices with zero mean. Suppose that* $\max_{1<\le i\le n} \|\|X_i\|\|_{\psi_\alpha} \le U^{(\alpha)} < \infty$ *for some* $\alpha \ge 1$*. Let*

$$\sigma^2 := \max\left\{\left\|\sum_{i=1}^n \mathbb{E}X_iX_i^\top\right\|, \left\|\sum_{i=1}^n \mathbb{E}X_i^\top X_i\right\|\right\}.$$

Then there exits a universal constant $C > 1$ *such that for all* $t > 0$*, the following bound holds with probability at least* $1 - e^{-t}$*,*

$$\left\|\frac{X_1 + \cdots X_n}{n}\right\| \le C\max\left\{\sigma\frac{\sqrt{t+\log(m_1+m_2)}}{n}, U^{(\alpha)}\left(\log\frac{\sqrt{n}U^{(\alpha)}}{\sigma}\right)\frac{t+\log(m_1+m_2)}{n}\right\}.$$

Lemma 5. *([Cai et al.], [2019b], Lemma D.5) Let $\{X_{i,j}\}_{1\le i\le r,1\le j\le L}$ be a sequence of i.i.d. standard Gaussian random variables. Consider some quantities $\kappa \ge 1$, $\Delta > 0$ and $0 < \delta < 1/2$. There exists some universal constant $C > 0$ such that if*

$$L \ge C r^{2\kappa^2}(\kappa\sqrt{r} + \delta)\exp(\Delta^2)\log(\frac{1}{\delta}),$$

then with probability at least $1-\delta$, there exists some $1 \le j_0 \le L$ such that $X_{1,j_0} > \kappa \max_{1<i\le r}|X_{i,j_0}| + \Delta$.

Lemma 6. *([Cai et al.], [2019b], Lemma D.6) Let U and V be two $d\times r$ matrices, each with orthonormal columns. Suppose that $\|UU^\top - VV^\top\| \le \delta$. Then for any unit vector $\mathbf{u}_0 \in \mathbb{R}^d$ lying in span(U), we have*

$$\|P_V(\mathbf{u}_0)\| \ge \sqrt{1-\delta^2} \qquad \text{and} \qquad \|P_{V^\perp}(\mathbf{u}_0)\| \le \delta,$$

where $P_V(\mathbf{u}_0)$ denotes $P_V(\mathbf{u}_0) = VV^\top\mathbf{u}_0$.

Lemma 7. *([Jain and Oh], [2014], Lemma A.4) Let $\mathbf{u}_l \in \mathbb{R}^n, 1 \le l \le r$ be such that $\|\mathbf{u}_l - \mathbf{u}_l^*\|_2 \le \alpha$ where $\alpha < 1/4$. Also, let $\mathbf{u}_l^*, 1 \le l \le r$ be μ-incoherent unit vectors. Now define $\widetilde{\mathbf{u}}_l$ as:*

$$\widetilde{\mathbf{u}}_l(i) = \begin{cases} \mathbf{u}_l(i), & \text{if } |\mathbf{u}_l(i)| \le \frac{\mu}{\sqrt{n}} \\ sign(\mathbf{u}_l(i))\frac{\mu}{\sqrt{n}}, & \text{if } |\mathbf{u}_l(i)| > \frac{\mu}{\sqrt{n}}. \end{cases}$$

Also, let $\widehat{\mathbf{u}}_l = \widetilde{\mathbf{u}}_l/\|\widetilde{\mathbf{u}}_l\|_2$. Then, $\|\widehat{\mathbf{u}}_l - \mathbf{u}_l^*\|_2 \le 3\alpha$, $\forall 1 \le l \le r$ and each $\widehat{\mathbf{u}}_l$ is 2μ-incoherent.

Lemma 8. *Suppose Assumption 3 holds. Then,*

$$\|\widehat{\boldsymbol{\beta}}_{1,3}^f - \boldsymbol{\beta}_{1,3}^*\| \le \left\|\widetilde{\boldsymbol{\beta}}_{1,3}/\|\widetilde{\boldsymbol{\beta}}_{1,3}\| - \boldsymbol{\beta}_{1,3}^*\right\|,$$

The equality holds if and only if $\widehat{\boldsymbol{\beta}}_{1,3}^f = \widetilde{\boldsymbol{\beta}}_{1,3}/\|\widetilde{\boldsymbol{\beta}}_{1,3}\|$.

Proof: Denote the estimated fusion groups as $G_1, \cdots, G_{\tau_{f_3}}$, and the true fusion groups in $\boldsymbol{\beta}_{1,3}^*$ as $G_1^*, \cdots, G_{f_3}^*$. If $f_3 = \tau_{f_3}$. Then, with a high probability,

$$\left\|\frac{\widetilde{\boldsymbol{\beta}}_{1,3}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}_{1,3}^*\|_2^2 - \|\widehat{\boldsymbol{\beta}}_{1,3}^f - \boldsymbol{\beta}_{1,3}^*\right\|_2^2 = \sum_{j=1}^{f_3}\sum_{s\in G_j^*}\left\{\left(\frac{\widetilde{\boldsymbol{\beta}}_{1,3,s}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}_{1,3,s}^*\right)^2 - \left(\widehat{\boldsymbol{\beta}}_{1,3,s}^f - \boldsymbol{\beta}_{1,3,s}^*\right)^2\right\}$$

$$= \sum_{j=1}^{f_3}\sum_{s\in G_j^*}\left\{\left(\frac{\widetilde{\boldsymbol{\beta}}_{1,3,s}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}_{1,3,s}^*\right)^2 - \left(\frac{1}{|G_j^*|}\sum_{s\in G_j^*}\frac{\widetilde{\boldsymbol{\beta}}_{1,3,s}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}_{1,3,s}^*\right)^2\right\} \ge 0,$$

where the second equality is due to the definition of the fusion operator, and the inequality is due to Lemma 3.

Without loss of generality, we assume $\tau_{f_3} = f_3 + 1$. By the assumption on Δ^*, one of the fusion groups in $\boldsymbol{\beta}^*_{1,3}$, say, G^*_m, is to be separated into two of the estimated fusion groups G_m and G_{m+1}. Then the difference can be decomposed into three terms,

$$\begin{aligned}
&\left\| \frac{\widetilde{\boldsymbol{\beta}}_{1,3}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}^*_{1,3} \right\|_2^2 - \left\| \widehat{\boldsymbol{\beta}}^f_{1,3} - \boldsymbol{\beta}^*_{1,3} \right\|_2^2 \\
= \ & \sum_{j \neq m} \sum_{s \in G^*_j} \left\{ \left(\frac{\widetilde{\boldsymbol{\beta}}_{1,3,s}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}^*_{1,3,s} \right)^2 - \left(\frac{1}{|G^*_j|} \sum_{s \in G^*_j} \frac{\widetilde{\boldsymbol{\beta}}_{1,3,s}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}^*_{1,3,s} \right)^2 \right\} \\
& + \sum_{s \in G_m} \left\{ \left(\frac{\widetilde{\boldsymbol{\beta}}_{1,3,s}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}^*_{1,3,s} \right)^2 - \left(\frac{1}{|G_m|} \sum_{s \in G_m} \frac{\widetilde{\boldsymbol{\beta}}_{1,3,s}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}^*_{1,3,s} \right)^2 \right\} \\
& + \left\{ \sum_{s \in G_{m+1}} \left(\frac{\widetilde{\boldsymbol{\beta}}_{1,3,s}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}^*_{1,3,s} \right)^2 - \left(\frac{1}{|G_{m+1}|} \sum_{s \in G_{m+1}} \frac{\widetilde{\boldsymbol{\beta}}_{1,3,s}}{\|\widetilde{\boldsymbol{\beta}}_{1,3}\|} - \boldsymbol{\beta}^*_{1,3,s} \right)^2 \right\} \geq 0,
\end{aligned}$$

where the last inequality is due to Lemma 3. Here the equality holds if and only if $\widehat{\boldsymbol{\beta}}^f_{1,3} = \widetilde{\boldsymbol{\beta}}_{1,3}/\|\widetilde{\boldsymbol{\beta}}_{1,3}\|$. The completes the proof of Lemma 8. □

S5 Additional simulations

We report in Table 3 the additional simulation results with block missing and $m_n = 0.9$.

Table 3: Additional simulation example with block missing and $m_n = 0.9$. Five methods are compared: STORE of Sun and Li (2017), MAGEE of Li et al. (2013), our method applied to the complete data only (Complete), our method without the fusion constraint (No-fusion), and our proposed method (POSTER).

(m_n, m_t)	w_k^*	f_0	method	Error of $\mathcal{B}^*$	Error of $\boldsymbol{\beta}_{k,j}^*$	TPR	FPR
(0.9, 0.4)	30	0.3	STORE	0.888 (0.044)	1.264 (0.067)	0.763 (0.017)	0.626 (0.037)
			MAGEE	1.420 (0.018)	NA	NA	NA
			Complete	0.541 (0.065)	0.864 (0.115)	0.859 (0.023)	0.262 (0.040)
			No-fusion	0.145 (0.005)	0.136 (0.007)	1.000 (0.000)	0.120 (0.000)
			POSTER	0.078 (0.004)	0.079 (0.007)	1.000 (0.000)	0.025 (0.004)
		0.7	STORE	0.860 (0.040)	1.290 (0.058)	0.782 (0.016)	0.574 (0.038)
			MAGEE	1.434 (0.004)	NA	NA	NA
			Complete	0.577 (0.057)	0.956 (0.109)	0.885 (0.016)	0.311 (0.036)
			No-fusion	0.138 (0.005)	0.130 (0.010)	1.000 (0.000)	0.073 (0.001)
			POSTER	0.114 (0.005)	0.141 (0.030)	1.000 (0.000)	0.055 (0.002)
	40	0.3	STORE	0.702 (0.056)	1.115 (0.095)	0.825 (0.02)	0.499 (0.042)
			MAGEE	1.248 (0.002)	NA	NA	NA
			Complete	0.308 (0.053)	0.614 (0.116)	0.901 (0.023)	0.170 (0.035)
			No-fusion	0.134 (0.005)	0.155 (0.029)	0.999 (0.001)	0.126 (0.004)
			POSTER	0.070 (0.004)	0.082 (0.008)	1.000 (0.000)	0.026 (0.004)
		0.7	STORE	0.674 (0.068)	1.032 (0.105)	0.837 (0.022)	0.490 (0.051)
			MAGEE	1.265 (0.003)	NA	NA	NA
			Complete	0.379 (0.055)	0.732 (0.119)	0.927 (0.014)	0.250 (0.038)
			No-fusion	0.126 (0.005)	0.150 (0.011)	1.000 (0.000)	0.096 (0.024)
			POSTER	0.105 (0.004)	0.139 (0.032)	1.000 (0.000)	0.054 (0.002)
(0.9, 0.6)	30	0.3	STORE	0.889 (0.030)	1.346 (0.037)	0.775 (0.013)	0.603 (0.028)
			MAGEE	1.570 (0.004)	NA	NA	NA
			Complete	0.511 (0.064)	0.822 (0.116)	0.868 (0.023)	0.246 (0.040)
			No-fusion	0.194 (0.007)	0.189 (0.009)	1.000 (0.000)	0.120 (0.000)
			POSTER	0.112 (0.007)	0.126 (0.010)	1.000 (0.000)	0.022 (0.004)
		0.7	STORE	0.864 (0.040)	1.292 (0.059)	0.785 (0.017)	0.589 (0.039)
			MAGEE	1.584 (0.005)	NA	NA	NA
			Complete	0.539 (0.057)	0.872 (0.113)	0.900 (0.015)	0.294 (0.035)
			No-fusion	0.190 (0.007)	0.212 (0.011)	0.998 (0.001)	0.097 (0.021)
			POSTER	0.167 (0.008)	0.186 (0.015)	0.999 (0.001)	0.082 (0.020)
	40	0.3	STORE	0.685 (0.059)	1.112 (0.096)	0.842 (0.019)	0.515 (0.047)
			MAGEE	1.348 (0.003)	NA	NA	NA
			Complete	0.310 (0.049)	0.700 (0.119)	0.903 (0.020)	0.213 (0.039)
			No-fusion	0.182 (0.008)	0.207 (0.019)	0.997 (0.001)	0.187 (0.034)
			POSTER	0.102 (0.007)	0.249 (0.071)	0.987 (0.007)	0.076 (0.031)
		0.7	STORE	0.678 (0.064)	1.116 (0.095)	0.843 (0.019)	0.520 (0.051)
			MAGEE	1.364 (0.004)	NA	NA	NA
			Complete	0.305 (0.039)	0.820 (0.118)	0.917 (0.015)	0.326 (0.047)
			No-fusion	0.163 (0.006)	0.399 (0.069)	0.993 (0.002)	0.174 (0.042)
			POSTER	0.147 (0.007)	0.242 (0.042)	0.995 (0.002)	0.141 (0.036)