Greed is good: correspondence recovery for unlabeled linear regression

Hang Zhang, Ping Li

Cognitive Computing Lab
Baidu Research
10900 NE 8th St, Bellevue, WA 98004, USA
{zhanghanghitomi, pingli98}@gmail.com

1 NOTATIONS

We start our discussion by defining $\widehat{\mathbf{B}}$ and $\widetilde{\mathbf{B}}$ respectively as

$$\widetilde{\mathbf{B}} = (n-h)^{-1}\,\mathbf{X}^\top \mathbf{\Pi}^* \mathbf{X}\mathbf{B}^*,$$
$$\widehat{\mathbf{B}} = (n-h)^{-1}\,\mathbf{X}^\top \mathbf{Y} = \widetilde{\mathbf{B}} + \ (n-h)^{-1}\,\mathbf{X}^\top \mathbf{W},$$

where h is denoted as the Hamming distance between identity matrix $\mathbf{I}$ and the ground truth selection matrix $\mathbf{\Pi}^*$, i.e., $h = \mathsf{d}_{\mathsf{H}}(\mathbf{I}, \mathbf{\Pi}^*)$.

Here we modify the *leave-one-out* trick, which is previously used in Karoui [2013], Karoui et al. [2013], Karoui [2018], Chen et al. [2020], Sur et al. [2019]. First, we construct an independent copy $\mathbf{X}^{'}_{s,:}$ for each row $\mathbf{X}_{s,:}$ (sth row of the sensing matrix $\mathbf{X}$). Building on these independent copies, we construct the leave-one-out sample $\mathbf{X}_{\setminus(s)}$ by replacing the sth row in the sensing matrix $\mathbf{X}$ with its independent copy $\mathbf{X}^{'}_{s,:}$. The detailed construction of independent copies $\{\widetilde{\mathbf{B}}_{\setminus(s)}\}_{s=1}^n$ proceeds as

$$\widetilde{\mathbf{B}}_{\setminus(s)} = (n-h)^{-1}\Bigg(\sum_{\substack{k\neq s\\ \pi^*(k)\neq s}} \mathbf{X}_{\pi(k),:}\mathbf{X}_{k,:}^\top + \sum_{\substack{k=s \text{ or}\\ \pi^*(k)=s}} \mathbf{X}^{'}_{\pi(k),:}\mathbf{X}^{'\top}_{k,:}\Bigg)\mathbf{B}^*.$$

Easily we can verify that $\widetilde{\mathbf{B}}_{\setminus(i)}$ is independent of the ith row $\mathbf{X}_{i,:}$. Similarly, we construct the matrices $\{\widetilde{\mathbf{B}}_{\setminus(s,t)}\}_{1\le s\neq t\le n}$ as

$$\widetilde{\mathbf{B}}_{\setminus(s,t)} = (n-h)^{-1}\Bigg(\sum_{\substack{k\neq s,t\\ \pi^*(k)\neq s,t}} \mathbf{X}_{\pi(k),:}\mathbf{X}_{k,:}^\top + \sum_{\substack{k=s \text{ or } k=t \text{ or}\\ \pi^*(k)=s \text{ or } \pi^*(k)=t}} \mathbf{X}^{'}_{\pi(k),:}\mathbf{X}^{'\top}_{k,:}\Bigg)\mathbf{B}^*,$$

and verify the independence between $\widetilde{\mathbf{B}}_{\setminus(s,t)}$ and the rows $\mathbf{X}_{s,:}, \mathbf{X}_{t,:}$.

Moreover, we define the events $\mathcal{E}_i$ as

$$\mathcal{E}_1(\mathbf{M}) \triangleq \left\{\left\|\mathbf{M}^\top\mathbf{X}_{i,:}\right\|_2 \lesssim \sqrt{\log n}\,\|\mathbf{M}\|_{\mathrm{F}} \text{ and } \left\|\mathbf{M}^\top\mathbf{X}^{'}_{i,:}\right\|_2 \lesssim \sqrt{\log n}\,\|\mathbf{M}\|_{\mathrm{F}}\ \ \forall\, 1\le i\le n\right\};$$
$$\mathcal{E}_{2,1} \triangleq \left\{\left\langle \mathbf{X}_{i,:}, \mathbf{X}^{'}_{j,:}\right\rangle \lesssim \sqrt{p\log n},\ \ 1\le i,j\le n\right\};$$
$$\mathcal{E}_{2,2} \triangleq \left\{\left\langle \mathbf{X}_{i,:}, \mathbf{X}_{j,:}\right\rangle \lesssim \sqrt{p\log n},\ \ 1\le i\neq j\le n\right\};$$
$$\mathcal{E}_{2,3} \triangleq \left\{\left\langle \mathbf{X}^{'}_{i,:}, \mathbf{X}^{'}_{j,:}\right\rangle \lesssim \sqrt{p\log n},\ \ 1\le i,j\le n\right\};$$
$$\mathcal{E}_2 = \mathcal{E}_{2,1}\bigcap\mathcal{E}_{2,2}\bigcap\mathcal{E}_{2,3};$$

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

$$
\begin{aligned}
\mathcal{E}_3 &= \left\{\|\mathbf{X}_{s,:}\|_2 \le \sqrt{p\log n} \text{ and } \|\mathbf{X}'_{s,:}\|_2 \le \sqrt{p\log n},\ \forall\, 1\le s\le n\right\};\\
\mathcal{E}_4 &= \left\{\|\mathbf{X}\|_{\mathrm{F}} \le \sqrt{2np} \text{ and } \left\|\mathbf{X}_{\backslash(s)}\right\|_{\mathrm{F}} \le \sqrt{2np},\ \forall\, 1\le s\le n\right\};\\
\mathcal{E}_5 &= \left\{\|\mathbf{X}\mathbf{X}_{s,:}\|_2 \lesssim (\log n)\sqrt{np},\ \forall\, 1\le s\le n\right\};\\
\mathcal{E}_{6,1} &= \left\{\left\|\mathbf{B}^* - \widetilde{\mathbf{B}}_{\backslash(s)}\right\|_{\mathrm{F}} \lesssim \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}},\ \forall\, 1\le s\le n\right\};\\
\mathcal{E}_{6,2} &= \left\{\left\|\mathbf{B}^* - \widetilde{\mathbf{B}}_{\backslash(s,t)}\right\|_{\mathrm{F}} \lesssim \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}},\ \forall\, 1\le s\ne t\le n\right\};\\
\mathcal{E}_6 &= \mathcal{E}_{6,1}\bigcap\mathcal{E}_{6,2};\\
\mathcal{E}_7 &= \left\{\left\|(\widetilde{\mathbf{B}} - \widetilde{\mathbf{B}}_{\backslash(s)})^\top\mathbf{X}_{s,:}\right\|_2 \lesssim \frac{p\log n}{n}\|\mathbf{B}^*\|_{\mathrm{F}},\ \forall\, 1\le s\le n\right\};\\
\mathcal{E}_8 &= \left\{\left\|(\widetilde{\mathbf{B}} - \widetilde{\mathbf{B}}_{\backslash(s,t)})^\top\mathbf{X}_{s,:}\right\|_2 \lesssim \frac{p\log n}{n}\|\mathbf{B}^*\|_{\mathrm{F}},\ \forall\, 1\le s\ne t\le n\right\};\\
\mathcal{E}_9 &= \left\{\left\|(\widetilde{\mathbf{B}} - \mathbf{B}^*)^\top\mathbf{X}_{s,:}\right\|_2 \lesssim \frac{(\log n)^{3/2}(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}},\ \forall\, 1\le s\le n\right\}.
\end{aligned}
$$

In addition, we define the quantities Δ_1, Δ_2, and Δ_3 as

$$
\Delta_1 = c_0\sigma(\log n)^{5/2}\sqrt{\frac{p}{n}}\|\mathbf{B}^*\|_{\mathrm{F}}; \tag{1}
$$

$$
\Delta_2 = c_1\sigma(\log^2 n)\|\mathbf{B}^*\|_{\mathrm{F}}; \tag{2}
$$

$$
\Delta_3 = c_2\left[\frac{mp(\log n)^2\sigma^2}{n} + \sigma^2(\log n)^2\sqrt{\frac{mp}{n}}\right], \tag{3}
$$

respectively. Besides, we define the summary Δ as $\Delta_1 + \Delta_2 + \Delta_3$.

2 APPENDIX: PROOF OF THEOREM 2

Proof. We define the error event $\mathcal{E}$ as

$$
\mathcal{E} \triangleq \left\{\left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}\right\|_2^2 + \left\langle \mathbf{W}_i, \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}\right\rangle \le \left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}, \mathbf{B}^{*\top}\mathbf{X}_j\right\rangle + \left\langle \mathbf{W}_i, \mathbf{B}^{*\top}\mathbf{X}_j\right\rangle,\ \forall\, j\ne\pi^*(i)\right\},
$$

and complete the proof by showing $\mathbb{P}(\mathcal{E}) \lesssim n^{-c}$. To start with, we define three events $\mathcal{E}_1, \mathcal{E}_2$ and $\mathcal{E}_3$ as

$$
\begin{aligned}
\mathcal{E}_1 &\triangleq \left\{\left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}\right\|_2 \le \frac{1}{2}\|\mathbf{B}^*\|_{\mathrm{F}}\right\};\\
\mathcal{E}_2 &\triangleq \left\{\langle\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}, \mathbf{B}^{*\top}\mathbf{X}_j\rangle \gtrsim \log n\left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}}, \forall j\ne\pi^*(i)\right\};\\
\mathcal{E}_3 &\triangleq \left\{\langle\mathbf{W}_i, \mathbf{B}^{*\top}\left(\mathbf{X}_j - \mathbf{X}_{\pi^*(i)}\right)\rangle \gtrsim \sigma\log n\|\mathbf{B}^*\|_{\mathrm{F}}, \forall j\ne\pi^*(i)\right\},
\end{aligned}
$$

respectively. The proof begins with the following decomposition, which reads as

$$
\mathbb{E}\mathbb{1}\left(\mathcal{E}\right) = \mathbb{E}\mathbb{1}\left(\mathcal{E}\bigcap\bigcap_{i=1}^3\overline{\mathcal{E}}_i\right) + \mathbb{E}\mathbb{1}\left(\bigcup_{i=1}^3\mathcal{E}_i\right).
$$

The subsequent proof can be divided into two parts.

Part I. We prove that $\mathbb{E}\mathbb{1}\left(\mathcal{E}\bigcap\bigcap_{i=1}^3\overline{\mathcal{E}}_i\right)$ is zero provided that $\mathrm{srank}(\mathbf{B}^*) \gtrsim \log^2 n$ and $\mathsf{SNR} \ge c$. The underlying reason is as the following. To begin with, we obtain

$$
\left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}\right\|_2^2 \overset{①}{\gtrsim} \|\mathbf{B}^*\|_{\mathrm{F}}^2 \overset{②}{\gtrsim} \frac{\log n}{\sqrt{\mathrm{srank}(\mathbf{B}^*)}}\|\mathbf{B}^*\|_{\mathrm{F}}^2 + \sigma\log n\|\mathbf{B}^*\|_{\mathrm{F}}
$$

$$\overset{③}{\geq} \log n \left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}} + \sigma \log n \|\mathbf{B}^*\|_{\mathrm{F}}$$

where ① is due to $\overline{\mathcal{E}}_1$, ② is because of the assumption $\mathsf{srank}(\mathbf{B}^*) \gtrsim \log^2 n$ and $\mathsf{SNR} \geq c$, and ③ results from the relation

$$\left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}} \leq \|\mathbf{B}^*\|_{\mathrm{OP}}\|\mathbf{B}^*\|_{\mathrm{F}} = \frac{\|\mathbf{B}^*\|_{\mathrm{F}}^2}{\sqrt{\mathsf{srank}(\mathbf{B}^*)}}.$$

Condition on the event $\overline{\mathcal{E}}_2 \bigcap \overline{\mathcal{E}}_3$, we conclude

$$\left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}\right\|_2^2 \gtrsim \langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}, \mathbf{B}^{*\top}\mathbf{X}_j\rangle + \left\langle \mathbf{W}_i, \mathbf{B}^{*\top}\left(\mathbf{X}_j - \mathbf{X}_{\pi^*(i)}\right)\right\rangle,$$

which is contradictory to the definition of $\mathcal{E}$ and hence leads to $\mathbb{E}\mathbb{1}(\mathcal{E}\bigcap\bigcap_{i=1}^3 \overline{\mathcal{E}}_i) = 0$. Therefore we can invoke the union bound and upper-bound the error probability $\mathbb{E}\mathbb{1}\left(\mathcal{E}\right)$ as $\sum_{i=1}^3 \mathbb{E}\mathbb{1}(\mathcal{E}_i)$.

Part II. The following context separately bound the three terms $\mathbb{E}\mathbb{1}\left(\mathcal{E}_i\right)$, $1 \leq i \leq 3$. For $\mathbb{E}\mathbb{1}(\mathcal{E}_1)$, we can simply invoke Lemma 15 and bound it as

$$\mathbb{E}\mathbb{1}\mathcal{E}_1 \lesssim e^{-\mathsf{srank}(\mathbf{B}^*)} \overset{④}{\lesssim} n^{-c},$$

where ④ is due to the assumption $\mathsf{srank}(\mathbf{B}^*) \gg \log^2 n$.

Then we turn to bounding $\mathbb{E}\mathbb{1}(\mathcal{E}_2)$, which proceeds as

$$\begin{aligned}
&\mathbb{E}\mathbb{1}\left(\mathcal{E}_2\right) \leq \mathbb{P}\left(\left\|\mathbf{B}^*\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}\right\|_2 \gtrsim \sqrt{\log n}\left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}}\right) \\
&+ n\mathbb{E}_{\mathbf{X}_{\pi^*(i)}}\mathbb{1}\left(\left\|\mathbf{B}^*\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}\right\|_2 \lesssim \sqrt{\log n}\left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}}, \langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}, \mathbf{B}^{*\top}\mathbf{X}_j\rangle \gtrsim \log n\left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}}\right).
\end{aligned} \tag{4}$$

For the first term in (4), we have

$$\mathbb{P}\left(\left\|\mathbf{B}^*\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}\right\|_2 \gtrsim \sqrt{\log n}\left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}}\right) \lesssim n^{-c_0}.$$

While for the second term in (4), we exploit the independence between $\mathbf{X}_{\pi^*(i)}$ and $\mathbf{X}_j$, which yields

$$\begin{aligned}
&\mathbb{E}_{\mathbf{X}_{\pi^*(i)}}\mathbb{1}\left(\left\|\mathbf{B}^*\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}\right\|_2 \lesssim \sqrt{\log n}\left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}}, \langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i)}, \mathbf{B}^{*\top}\mathbf{X}_j\rangle \gtrsim \log n\left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}}\right) \\
&\lesssim \exp\left(-\frac{c_1 \log^2 n\left\|\mathbf{B}^*\mathbf{B}^{*\top}\right\|_{\mathrm{F}}^2}{\log n\|\mathbf{B}^*\mathbf{B}^{*\top}\|_{\mathrm{F}}^2}\right) \leq n^{-c_1}.
\end{aligned}$$

Hence we conclude $\mathbb{E}\mathbb{1}\left(\mathcal{E}_2\right) \lesssim n^{-c_0} + n \cdot n^{-c_1} \lesssim n^{-c_2}$. In the end, we consider $\mathbb{E}\mathbb{1}(\mathcal{E}_3)$, which is written as

$$\mathbb{E}\mathbb{1}(\mathcal{E}_3) \leq \mathbb{P}\big(\left\|\mathbf{B}^{*\top}\left(\mathbf{X}_j - \mathbf{X}_{\pi^*(i)}\right)\right\|_2 \leq \frac{\|\mathbf{B}^*\|_{\mathrm{F}}}{2}, \exists\, j\big) + \mathbb{P}\big(\mathcal{E}_3, \left\|\mathbf{B}^{*\top}\left(\mathbf{X}_j - \mathbf{X}_{\pi^*(i)}\right)\right\|_2 \geq \frac{\|\mathbf{B}^*\|_{\mathrm{F}}}{2}, \forall\, j\big). \tag{5}$$

For the first term in (5), we invoke Lemma 15 and have

$$\mathbb{P}\left(\left\|\mathbf{B}^{*\top}\left(\mathbf{X}_j - \mathbf{X}_{\pi^*(i)}\right)\right\|_2 \leq \frac{\|\mathbf{B}^*\|_{\mathrm{F}}}{2}, \exists\, j\right) \overset{⑤}{\leq} n\exp\left(-c \cdot \mathsf{srank}(\mathbf{B}^*)\right) \overset{⑥}{\lesssim} n^{-c},$$

where ⑤ is due to the union bound and ⑥ is due to the assumption such that $\mathsf{srank}(\mathbf{B}^*) \gg \log^2 n$.

For the second term in (5), we exploit the independence across $\mathbf{X}$ and $\mathbf{W}$ and have

$$\mathbb{P}\left(\mathcal{E}_3, \left\|\mathbf{B}^{*\top}\left(\mathbf{X}_j - \mathbf{X}_{\pi^*(i)}\right)\right\|_2 \geq \frac{\|\mathbf{B}^*\|_{\mathrm{F}}}{2}, \forall\, j\right) \leq n\exp\left(-\frac{c\log^2 n\|\mathbf{B}^*\|_{\mathrm{F}}^2}{\|\mathbf{B}^*\|_{\mathrm{F}}^2}\right) \lesssim n^{-c}.$$

Summarizing the above discussion then completes the proof.

□

3 PROOF OF THEOREM 3

Notice the reconstruction error, i.e., $\pi^*(i) \neq \widehat{\pi}^*(i)$, will occur as long as there exists $j \neq \pi^*(i)$ such that

$$\left\langle \mathbf{Y}_{i,:}, \widehat{\mathbf{B}}^\top \mathbf{X}_{\pi^*(i),:} \right\rangle \leq \left\langle \mathbf{Y}_{i,:}, \widehat{\mathbf{B}}^\top \mathbf{X}_{j,:} \right\rangle. \tag{6}$$

With the relation $\mathbf{Y}_{i,:} = \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:} + \mathbf{W}_{i,:}$ and $\widehat{\mathbf{B}} = \widetilde{\mathbf{B}} + (n-h)^{-1}\mathbf{X}^\top\mathbf{W}$, we can rewrite (6) as

$$\begin{aligned} &\left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:} + \mathbf{W}_{i,:}, \left(\widetilde{\mathbf{B}} + (n-h)^{-1}\mathbf{X}^\top\mathbf{W}\right)^\top \mathbf{X}_{\pi^*(i),:} \right\rangle \\ \leq &\left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:} + \mathbf{W}_{i,:}, \left(\widetilde{\mathbf{B}} + (n-h)^{-1}\mathbf{X}^\top\mathbf{W}\right)^\top \mathbf{X}_{j,:} \right\rangle. \end{aligned} \tag{7}$$

For the notation conciseness, we define terms Term_i $(1 \leq i \leq 4)$ as

$$\mathsf{Term}_{\text{tot}} = \left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{\pi^*(i),:} - \mathbf{X}_{j,:}\right)\right\rangle; \tag{8}$$

$$\mathsf{Term}_1 = (n-h)^{-1}\left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \mathbf{W}^\top\mathbf{X}\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^*(i),:}\right)\right\rangle; \tag{9}$$

$$\mathsf{Term}_2 = \left\langle \mathbf{W}_{i,:}, \widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^*(i),:}\right)\right\rangle; \tag{10}$$

$$\mathsf{Term}_3 = (n-h)^{-1}\left\langle \mathbf{W}_{i,:}, \mathbf{W}^\top\mathbf{X}\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^*(i),:}\right)\right\rangle. \tag{11}$$

Then (7) is equivalent to $\mathsf{Term}_{\text{tot}} \leq \mathsf{Term}_1 + \mathsf{Term}_2 + \mathsf{Term}_3$. With the union bound, we conclude

$$\begin{aligned} \mathbb{P}\left(\pi^*(i) \neq \widehat{\pi}(i), \exists\, i\right) &= \mathbb{E}\left[\mathbb{1}\left(\mathsf{Term}_{\text{tot}} \leq \mathsf{Term}_1 + \mathsf{Term}_2 + \mathsf{Term}_3,\ \exists\, i,j\right)\mathbb{1}\left(\bigcap_{a=1}^{9}\mathcal{E}_a\right)\right] + \sum_{a=1}^{9}\mathbb{P}\left(\overline{\mathcal{E}}_a\right) \\ &\overset{①}{\leq} n^2\mathbb{E}\left[\mathbb{1}\left(\mathsf{Term}_{\text{tot}} \leq \mathsf{Term}_1 + \mathsf{Term}_2 + \mathsf{Term}_3\right)\mathbb{1}\left(\bigcap_{a=1}^{9}\mathcal{E}_a\right)\right] + c_0 p^{-c_1} + c_2 n^{-c_3}, \end{aligned} \tag{12}$$

where in ① we invoke Lemma 5, Lemma 6, Lemma 7, Lemma 8, Lemma 9, Lemma 10, Lemma 11, and Lemma 12.

Regarding the term $\mathbb{E}\left[\mathbb{1}\left(\mathsf{Term}_{\text{tot}} \leq \mathsf{Term}_1 + \mathsf{Term}_2 + \mathsf{Term}_3,\ \exists\, i,j\right)\mathbb{1}\left(\bigcap_{a=1}^{9}\mathcal{E}_a\right)\right]$, we further decompose it as the summary of two terms reading as

$$\begin{aligned} &\mathbb{E}\left[\mathbb{1}\left(\mathsf{Term}_{\text{tot}} \leq \mathsf{Term}_1 + \mathsf{Term}_2 + \mathsf{Term}_3\right)\mathbb{1}\left(\bigcap_{a=1}^{9}\mathcal{E}_a\right)\right] \\ \leq &\,\mathbb{E}\left[\mathbb{1}\left(\mathsf{Term}_{\text{tot}} \leq \Delta\right)\mathbb{1}\left(\bigcap_{a=1}^{9}\mathcal{E}_a\right)\right] \\ &+ \mathbb{E}\left[\mathbb{1}\left(\mathsf{Term}_1 + \mathsf{Term}_2 + \mathsf{Term}_3 \geq \Delta\right)\mathbb{1}\left(\bigcap_{a=1}^{9}\mathcal{E}_a\right)\right], \\ \leq &\,\mathbb{E}\left[\mathbb{1}\left(\mathsf{Term}_{\text{tot}} \leq \Delta\right)\mathbb{1}\left(\bigcap_{a=1}^{9}\mathcal{E}_a\right)\right] + \mathbb{E}\left[\mathbb{1}\left(\mathsf{Term}_1 \geq \Delta_1\right)\mathbb{1}\left(\bigcap_{a=1}^{9}\mathcal{E}_a\right)\right] \\ &+ \mathbb{E}\left[\mathbb{1}\left(\mathsf{Term}_2 \geq \Delta_2\right)\mathbb{1}\left(\bigcap_{a=1}^{9}\mathcal{E}_a\right)\right] + \mathbb{E}\left[\mathbb{1}\left(\mathsf{Term}_3 \geq \Delta_3\right)\mathbb{1}\left(\bigcap_{a=1}^{9}\mathcal{E}_a\right)\right], \end{aligned} \tag{13}$$

where the definitions of Δ_1, Δ_2, Δ_3, and Δ are referred to Section 1. The proof is then completed by combining (12) and (13) and invoking Lemma 1, Lemma 2, Lemma 3, and Lemma 4.

Lemma 1. *Assume that* $\text{srank}(\mathbf{B}^*) \gg \log^4 n$, $n \gtrsim p\log^6 n$, *and* $\mathsf{SNR} \geq c$ *and conditional on the intersection of events* $\mathcal{E}_1(\mathbf{B}^*)\bigcap\mathcal{E}_1\left(\mathbf{B}^*\widetilde{\mathbf{B}}^\top_{\setminus(\pi^*(i),j)}\right)\bigcap\mathcal{E}_6\bigcap\mathcal{E}_7$*, where indices* $\pi^*(i)$ *and* j *are fixed. we have* $\mathsf{Term}_{\text{tot}} \geq \Delta$ *hold with probability exceeding* $1 - n^{-c}$ *when* n *and* p *are sufficiently large, where* $\mathsf{Term}_{\text{tot}}$ *and* Δ *are defined in* (8) *and Section 1, respectively.*

Proof. We start the discussion by decomposing $\mathsf{Term}_{\text{tot}}$ as

$$\mathsf{Term}_{\text{tot}} = \left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}\right\|_2^2 + \underbrace{\left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \left(\widetilde{\mathbf{B}} - \mathbf{B}^*\right)^\top \mathbf{X}_{\pi^*(i),:}\right\rangle}_{\triangleq \mathsf{Term}_{\text{tot},1}} - \underbrace{\left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \widetilde{\mathbf{B}}^\top \mathbf{X}_{j,:}\right\rangle}_{\triangleq \mathsf{Term}_{\text{tot},2}}.$$

Then we obtain

$$\begin{aligned}\mathbb{P}\left(\mathsf{Term}_{\text{tot}} \leq \Delta\right) &= \mathbb{P}\left(\frac{\Delta}{\left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}\right\|_2^2} - \frac{\mathsf{Term}_{\text{tot},1}}{\left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}\right\|_2^2} + \frac{\mathsf{Term}_{\text{tot},2}}{\left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}\right\|_2^2} \geq 1\right) \\ &\leq \underbrace{\mathbb{P}\left(\left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}\right\|_2 \leq \delta\right)}_{\triangleq \zeta_1} + \underbrace{\mathbb{P}\left(\frac{\Delta}{\delta^2} + \frac{|\mathsf{Term}_{\text{tot},1}|}{\delta^2} + \frac{|\mathsf{Term}_{\text{tot},2}|}{\delta^2} \geq 1\right)}_{\triangleq \zeta_2}. \end{aligned} \tag{14}$$

We separately bound the probabilities ζ_1 and ζ_2 by setting δ as $1/2\|\mathbf{B}^*\|_{\text{F}}$. For the term ζ_1, we invoke the small ball probability (Lemma 15) and conclude

$$\mathbb{P}\left(\left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}\right\|_2 \leq \frac{1}{2}\|\mathbf{B}^*\|_{\text{F}}\right) \leq e^{-c\,\text{srank}(\mathbf{B}^*)}. \tag{15}$$

For probability ζ_2, we will prove it to be zero provided $\mathsf{SNR} \geq c$. The proof is completed by showing

$$\frac{\Delta}{\delta^2} + \frac{|\mathsf{Term}_{\text{tot},1}|}{\delta^2} + \frac{|\mathsf{Term}_{\text{tot},2}|}{\delta^2} < 1$$

hold with probability $1 - n^{-c}$. Detailed calculation proceeds as follows.

Phase I. First, we consider term $\mathsf{Term}_{\text{tot},1}$. Conditional on the intersection of events $\mathcal{E}_1(\mathbf{B}^*) \bigcap \mathcal{E}_7 \bigcap \mathcal{E}_9$, we have

$$\begin{aligned}|\mathsf{Term}_{\text{tot},1}| &\leq \left\|\mathbf{B}^{\top *}\mathbf{X}_{i,:}\right\|_2 \left\|\left(\widetilde{\mathbf{B}} - \mathbf{B}^*\right)^\top \mathbf{X}_{\pi^*(i),:}\right\|_2 \lesssim \sqrt{\log n}\|\mathbf{B}^*\|_{\text{F}} \frac{(\log n)^{3/2}(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\text{F}} \\ &= (\log^2 n)(\log n^2 p^3)\sqrt{\frac{p}{n}}\|\mathbf{B}^*\|_{\text{F}}^2. \end{aligned}$$

Phase II. Then we turn to term $\mathsf{Term}_{\text{tot},2}$. Adopting the leave-out-out trick, we can expand it as

$$\mathsf{Term}_{\text{tot},2} = \underbrace{\left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \left(\widetilde{\mathbf{B}} - \widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\right)^\top \mathbf{X}_{j,:}\right\rangle}_{\mathsf{Term}_{\text{tot},2,1}} + \underbrace{\left\langle \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}, \widetilde{\mathbf{B}}^\top_{\setminus(\pi^*(i),j)}\mathbf{X}_{j,:}\right\rangle}_{\mathsf{Term}_{\text{tot},2,2}}.$$

For term $\mathsf{Term}_{\text{tot},2,1}$, we have

$$\begin{aligned}\mathsf{Term}_{\text{tot},2,1} &\leq \left\|\mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:}\right\|_2 \left\|\left(\widetilde{\mathbf{B}} - \widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\right)^\top \mathbf{X}_{j,:}\right\|_2 \overset{①}{\lesssim} \sqrt{\log n}\|\mathbf{B}^*\|_{\text{F}} \frac{p\log n}{n}\|\mathbf{B}^*\|_{\text{F}} \\ &= \frac{p(\log n)^{3/2}}{n}\|\mathbf{B}^*\|_{\text{F}}^2, \end{aligned}$$

where in ① we condition on event $\mathcal{E}_7$. Regarding the term $\mathsf{Term}_{2,2,2}$, we notice that $\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}$ is independent of the rows $\mathbf{X}_{\pi^*(i),:}$ and $\mathbf{X}_{j,:}$ due to its construction method. Then we can bound the term $\mathsf{Term}_{2,2,2}$ by fixing the rows $\{\mathbf{X}_{s,:}\}_{s \neq \pi^*}$ and viewing $\mathbf{X}_{\pi^*(i),:}$ as the RV, which yields

$$\mathsf{Term}_{\text{tot},2,2} \lesssim \sqrt{\log n}\left\|\mathbf{B}^*\widetilde{\mathbf{B}}^\top_{\setminus(\pi^*(i),j)}\mathbf{X}_{j,:}\right\|_2 \tag{16}$$

holds with probability $1 - n^{-c}$. Conditional on event $\mathcal{E}_1(\mathbf{B}^*\widetilde{\mathbf{B}}^\top_{\setminus(\pi^*(i),j)})$, we have

$$\mathsf{Term}_{\text{tot},2,2} \lesssim (\log n)\left\|\mathbf{B}^*\widetilde{\mathbf{B}}^\top_{\setminus(\pi^*(i),j)}\right\|_{\text{F}} \lesssim (\log n)\|\mathbf{B}^*\|_{\text{OP}}\left\|\widetilde{\mathbf{B}}^\top_{\setminus(\pi^*(i),j)}\right\|_{\text{F}}$$

$$\overset{②}{\leq} (\log n)\|\mathbf{B}^*\|_{\mathsf{OP}} \left[\left\|\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)} - \mathbf{B}^*\right\|_{\mathrm{F}} + \|\mathbf{B}^*\|_{\mathrm{F}}\right] \overset{③}{\lesssim} \frac{(\log n)\|\mathbf{B}^*\|_{\mathrm{F}}^2}{\sqrt{\mathsf{srank}(\mathbf{B}^*)}},$$

where in ② we use the definition of stable rank, and in ③ we conditional on event $\mathcal{E}_6$, $n \geq p$, and $n \gtrsim p\log^6 n$.

Phase III. Conditional on (16), we can expand the sum $\Delta/\delta^2 + \mathsf{Term}_{\mathsf{tot},1}/\delta^2 + \mathsf{Term}_{\mathsf{tot},2}/\delta^2$ as

$$\begin{aligned}
\frac{\Delta}{\delta^2} + \frac{\mathsf{Term}_{\mathsf{tot},1}}{\delta^2} + \frac{\mathsf{Term}_{\mathsf{tot},2}}{\delta^2} &= c_0\sigma(\log n)^{5/2}\sqrt{\frac{p}{n}}\frac{1}{\|\mathbf{B}^*\|_{\mathrm{F}}} + \frac{c_1\sigma(\log^2 n)}{\|\mathbf{B}^*\|_{\mathrm{F}}} + c_2\left(\frac{pm}{n} + \sqrt{\frac{mp}{n}}\right)\frac{(\log n)^2\sigma^2}{\|\mathbf{B}^*\|_{\mathrm{F}}^2} \\
&+ \frac{c_3(\log^2 n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}} + \frac{c_4 p(\log n)^{3/2}}{n} + \frac{c_5\log n}{\sqrt{\mathsf{srank}(\mathbf{B}^*)}} \\
&\asymp c_0\sqrt{\frac{p}{nm}}\frac{(\log n)^{5/2}}{\sqrt{\mathsf{SNR}}} + \frac{c_1\log^2 n}{\sqrt{m\cdot\mathsf{SNR}}} + \frac{c_2 p\,(\log n)^2}{n\cdot\mathsf{SNR}} + c_2\sqrt{\frac{p}{mn}}\frac{(\log n)^2}{\mathsf{SNR}} \\
&+ \frac{c_3(\log^2 n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}} + \frac{c_4 p(\log n)^{3/2}}{n} + \frac{c_5\log n}{\sqrt{\mathsf{srank}(\mathbf{B}^*)}}.
\end{aligned}$$

Provided that $\mathsf{SNR} \geq c$, $\mathsf{srank}(\mathbf{B}^*) \gg \log^4 n$ and $n \gtrsim p\log^6 n$, we can verify the sum $\Delta/\delta^2 + \mathsf{Term}_{\mathsf{tot},1}/\delta^2 + \mathsf{Term}_{\mathsf{tot},2}/\delta^2$ to be significantly smaller than 1 when n and p are sufficiently large, which suggests

$$\zeta_2 \leq \mathbb{P}\left(\mathsf{Term}_{\mathsf{tot},2,2} \gtrsim \sqrt{\log n}\left\|\mathbf{B}^*\widetilde{\mathbf{B}}^{\top}_{\setminus(\pi^*(i),j)}\mathbf{X}_{j,:}\right\|_2\right) \leq n^{-c}.$$

Hence the proof is completed by combining (14) and (15). □

Remark 1. *If we strength the requirement on* SNR *from* $\mathsf{SNR} \geq c$ *to* $\mathsf{SNR} \gtrsim \log^2 n$*, we can relax the requirement on the stable rank* $\mathsf{srank}(\mathbf{B}^*)$ *from* $\mathsf{srank}(\mathbf{B}^*) \gg \log^4 n$ *to* $\mathsf{srank}(\mathbf{B}^*) \gg \log^2 n$*.*

Lemma 2. *Conditional on the intersection of events* $\mathcal{E}_3 \bigcap \mathcal{E}_4 \bigcap \mathcal{E}_5$ *and fixing the indices* $\pi^*(i)$ *and* j*, we have*

$$\mathsf{Term}_1 \lesssim \sigma(\log n)^{5/2}\sqrt{\frac{p}{n}}\|\mathbf{B}^*\|_{\mathrm{F}}.$$

hold with probability at least $1 - n^{-c}$*.*

Proof. Define vectors $\boldsymbol{u}_{\mathbf{X}}$ and $\boldsymbol{v}_{\mathbf{X}}^{\top}$ as

$$\begin{aligned}
\boldsymbol{u}_{\mathbf{X}} &= \mathbf{X}\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^*(i),:}\right), \\
\boldsymbol{v}_{\mathbf{X}} &= \mathbf{B}^{*\top}\mathbf{X}_{\pi^*(i),:},
\end{aligned}$$

respectively. We can rewrite Term_1 as

$$\mathsf{Term}_1 = (n-h)^{-1}\,\mathrm{Tr}\left[\mathbf{X}\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^*(i),:}\right)\mathbf{X}^{\top}_{\pi^*(i),:}\mathbf{B}^*\mathbf{W}^{\top}\right] = (n-h)^{-1}\boldsymbol{u}_{\mathbf{X}}^{\top}\mathbf{W}\boldsymbol{v}_{\mathbf{X}}.$$

Invoking the union bound, we conclude

$$\begin{aligned}
&\mathbb{P}\left(\mathsf{Term}_1 \gtrsim \sigma(\log n)^{5/2}\sqrt{\frac{p}{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right) \\
&\leq \mathbb{P}\left(\mathsf{Term}_1 \gtrsim \sigma(\log n)^{5/2}\sqrt{\frac{p}{n}}\|\mathbf{B}^*\|_{\mathrm{F}},\ \|\boldsymbol{u}_{\mathbf{X}}\|_2\|\boldsymbol{v}_{\mathbf{X}}\|_2 \lesssim (\log n)^{3/2}\sqrt{np}\|\mathbf{B}^*\|_{\mathrm{F}}\right) \\
&+ \mathbb{P}\left(\|\boldsymbol{u}_{\mathbf{X}}\|_2\|\boldsymbol{v}_{\mathbf{X}}\|_2 \gtrsim (\log n)^{3/2}\sqrt{np}\|\mathbf{B}^*\|_{\mathrm{F}}\right) \\
&\leq \underbrace{\mathbb{P}\left(\mathsf{Term}_1 \gtrsim \frac{\sigma(\log n)\|\boldsymbol{u}_{\mathbf{X}}\|_2\|\boldsymbol{v}_{\mathbf{X}}\|_2}{n-h}\right)}_{\triangleq\zeta_1} + \underbrace{\mathbb{P}\left(\|\boldsymbol{u}_{\mathbf{X}}\|_2\|\boldsymbol{v}_{\mathbf{X}}\|_2 \gtrsim (\log n)^{3/2}\sqrt{np}\|\mathbf{B}^*\|_{\mathrm{F}}\right)}_{\triangleq\zeta_2}. \qquad (17)
\end{aligned}$$

Then we separately bound the probabilities ζ_1 and ζ_2.

Phase I. For probability ζ_1, we exploit the independence between $\mathbf{X}$ and $\mathbf{W}$ and can view Term_1 as a Gaussian RV conditional on $\mathbf{X}$, since it is a linear combination of Gaussian RVs $\{\mathbf{W}_{i,j}\}_{1\le i\le n, 1\le j\le m}$. Easily we can calculate its mean to be zero and its variance as

$$\mathbb{E}_{\mathbf{W}}(\mathsf{Term}_1)^2 = \frac{\sigma^2}{(n-h)^2}\|\boldsymbol{u}_{\mathbf{X}}\|_2\|\boldsymbol{v}_{\mathbf{X}}\|_2^2.$$

Thus we can upper-bound ζ_1 as

$$\zeta_1 = \mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{W}}\mathbb{1}\left(\mathsf{Term}_1 \gtrsim \frac{\sigma(\log n)\|\boldsymbol{u}_{\mathbf{X}}\|_2\|\boldsymbol{v}_{\mathbf{X}}\|_2}{n-h}\right) \overset{①}{\le} \mathbb{E}_{\mathbf{X}}\exp\left(-c_0\log n\right) = n^{-c}, \tag{18}$$

where ① is due to the bound on the tail-probability of Gaussian RV.

Phase II. As for ζ_2, easily we can verify it to be zero conditional on the intersection of events $\mathcal{E}_3\bigcap\mathcal{E}_4\bigcap\mathcal{E}_5$ as

$$\|\boldsymbol{u}_{\mathbf{X}}\|_2\|\boldsymbol{v}_{\mathbf{X}}\|_2 \lesssim \sqrt{\log n}\|\mathbf{B}^*\|_{\mathrm{F}}\cdot\left(\|\mathbf{X}\mathbf{X}_{j,:}\|_2 + \left\|\mathbf{X}\mathbf{X}_{\pi^*(i),:}\right\|_2\right) \lesssim (\log n)^{3/2}\sqrt{np}\|\mathbf{B}^*\|_{\mathrm{F}}.$$

The proof is then completed by combining (17) and (18). □

Lemma 3. *Conditional on the intersection of events $\mathcal{E}_2\bigcap\mathcal{E}_3\bigcap\mathcal{E}_4\bigcap\mathcal{E}_6$ and fixing the indices $\pi^*(i)$ and j, we have $\mathsf{Term}_2 \le \sigma(\log n)^2\|\mathbf{B}^*\|_{\mathrm{F}}$ hold with probability at least $1-n^{-c}$.*

Proof. Following a similar proof strategy as in Lemma 3, we first invoke the union bound and obtain

$$\begin{aligned}
&\mathbb{P}\left(\mathsf{Term}_2 \gtrsim \sigma(\log n)^2\|\mathbf{B}^*\|_{\mathrm{F}}\right)\\
&\le \mathbb{P}\left(\mathsf{Term}_2 \gtrsim \sigma(\log n)^2\|\mathbf{B}^*\|_{\mathrm{F}},\ \left\|\widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2 \lesssim (\log n)\|\mathbf{B}^*\|_{\mathrm{F}}\right)\\
&+\mathbb{P}\left(\left\|\widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2 \gtrsim (\log n)\|\mathbf{B}^*\|_{\mathrm{F}}\right)\\
&\le \underbrace{\mathbb{P}\left(\mathsf{Term}_2 \gtrsim \sigma(\log n)\left\|\widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2\right)}_{\zeta_1} + \underbrace{\mathbb{P}\left(\left\|\widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2 \gtrsim (\log n)\|\mathbf{B}^*\|_{\mathrm{F}}\right)}_{\zeta_2}.
\end{aligned}\tag{19}$$

The following analysis separately investigates the two probabilities ζ_1 and ζ_2.

Phase I. Exploiting the independence between $\mathbf{X}$ and $\mathbf{W}$, we can bound ζ_1 as

$$\zeta_1 = \mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{W}}\mathbb{1}\left(\mathsf{Term}_2 \gtrsim \sigma(\log n)\left\|\widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2\right) \overset{①}{\le} \mathbb{E}_{\mathbf{X}}\exp\left(-c_0\log n\right) = n^{-c_0}, \tag{20}$$

where in ① we use the fact that Term_2 is a Gaussian RV with zero mean and $\left\|\widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2$ conditional on $\mathbf{X}$.

Phase II. Then we bound term ζ_2. Notice

$$\begin{aligned}
\left\|\widetilde{\mathbf{B}}^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2 &\le \left\|\left(\widetilde{\mathbf{B}}-\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\right)^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2 + \left\|\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2\\
&\le \left\|\left(\widetilde{\mathbf{B}}-\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\right)^\top\mathbf{X}_{j,:}\right\|_2 + \left\|\left(\widetilde{\mathbf{B}}-\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\right)^\top\mathbf{X}_{\pi^*(i),:}\right\|_2\\
&+ \left\|\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2,
\end{aligned}$$

we conclude

$$\begin{aligned}
\zeta_2 &\overset{②}{\le} \underbrace{\mathbb{P}\left(\left\|\left(\widetilde{\mathbf{B}}-\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\right)^\top\mathbf{X}_{j,:}\right\|_2 + \left\|\left(\widetilde{\mathbf{B}}-\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\right)^\top\mathbf{X}_{\pi^*(i),:}\right\|_2 \gtrsim \frac{p\log n}{n}\|\mathbf{B}^*\|_{\mathrm{F}}\right)}_{\zeta_{2,1}}\\
&+ \underbrace{\mathbb{P}\left(\left\|\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}^\top\left(\mathbf{X}_{j,:}-\mathbf{X}_{\pi^*(i),:}\right)\right\|_2 \gtrsim (\log n)\|\mathbf{B}^*\|_{\mathrm{F}}\right)}_{\zeta_{2,2}},
\end{aligned}\tag{21}$$

where in ② we use the fact $n \gtrsim p$. Invoking Lemma 11 then yields $\zeta_{2,1} = 0$. For term $\zeta_{2,2}$, we exploit the independence between $\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}$ and $\mathbf{X}_{j,:}$, $\mathbf{X}_{\pi^*(i),:}$. Via the Hanson-wright inequality [Vershynin, 2018], we have

$$\zeta_{2,2} \leq \exp\left[-c_0\left(\frac{(\log n)^2 \|\mathbf{B}^*\|_{\mathrm{F}}^2}{\left\|\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}^\top \widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\right\|_{\mathrm{OP}}} \wedge \frac{(\log n)^4 \|\mathbf{B}^*\|_{\mathrm{F}}^4}{\left\|\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}^\top \widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\right\|_{\mathrm{F}}^2}\right)\right] \overset{③}{\leq} n^{-c}, \tag{22}$$

where ③ is due to the fact

$$\left\|\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)}\right\|_{\mathrm{F}} \leq \|\mathbf{B}^*\|_{\mathrm{F}} + \left\|\widetilde{\mathbf{B}}_{\setminus(\pi^*(i),j)} - \mathbf{B}^*\right\|_{\mathrm{F}} \overset{④}{\lesssim} \|\mathbf{B}^*\|_{\mathrm{F}},$$

and in ④ we condition on event $\mathcal{E}_6$. Combining (19), (20), (21), and (22) then completes the proof. □

Lemma 4. *Conditional on event $\mathcal{E}_2$ and fixing the indices $\pi^*(i)$ and j, we have* $\mathsf{Term}_3 \lesssim \frac{mp(\log n)^2\sigma^2}{n} + \sigma^2(\log n)^2\sqrt{\frac{mp}{n}}$ *hold with probability exceeding* $1 - c_0 n^{-c_1}$.

Proof. For the benefits of presentation, we first define $\mathbf{\Xi}^{\pi^*(i),j}$ as $\mathbf{\Xi}^{\pi^*(i),j} = \mathbf{X}\left(\mathbf{X}_{\pi^*(i),:} - \mathbf{X}_{j,:}\right)$. Then we can rewrite Term_3 as $(n-h)^{-1}\mathbf{W}_{i,:}^\top \mathbf{W}^\top \mathbf{\Omega}^{\pi^*(i),j}$ and expand it as

$$\begin{aligned}
|\mathsf{Term}_3| &= (n-h)^{-1}\left|\Xi_i^{\pi^*(i),j}\mathbf{W}_{i,:}^\top \mathbf{W}_{i,:} + \mathbf{W}_{i,:}^\top\left(\sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\right)\right| \\
&\leq \frac{1}{n-h}\left|\Xi_i^{\pi^*(i),j}\right| \cdot \|\mathbf{W}_{i,:}\|_2^2 + \frac{1}{n-h}\left|\left\langle \mathbf{W}_{i,:}, \sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\right\rangle\right| \\
&\overset{①}{\leq} \frac{p\log n}{n-h}\|\mathbf{W}_{i,:}\|_2^2 + \frac{1}{n-h}\left|\left\langle \mathbf{W}_{i,:}, \sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\right\rangle\right|,
\end{aligned}$$

where in ① we condition on event $\mathcal{E}_2$ and have $\left|\Xi_i^{\pi^*(i),j}\right| \leq \left\|\mathbf{X}_{\pi^*(i),:}\right\|_2^2 + \|\mathbf{X}_{j,:}\|_2^2 \lesssim p\log n$. With the union bound, we obtain

$$\begin{aligned}
&\mathbb{P}\left(\mathsf{Term}_3 \gtrsim \frac{mp(\log n)^2\sigma^2}{n} + \sigma(\log n)^2\sqrt{\frac{mp}{n}}\right) \\
&\overset{②}{\leq} \underbrace{\mathbb{P}\left(\frac{p\log n}{n-h}\|\mathbf{W}_{i,:}\|_2^2 \gtrsim \frac{mp(\log n)^2\sigma^2}{n}\right)}_{\triangleq \zeta_1} + \underbrace{\mathbb{P}\left(\frac{1}{n-h}\left|\left\langle \mathbf{W}_{i,:}, \sum_{k\neq i}\Omega_k^{\pi^*(i),j}\mathbf{W}_{k,:}\right\rangle\right| \gtrsim \sigma^2(\log n)^2\sqrt{\frac{mp}{n}}\right)}_{\triangleq \zeta_2}.
\end{aligned} \tag{23}$$

Then we separately bound the two terms ζ_1 and ζ_2.

Phase I. For term ζ_1, we have

$$\zeta_1 \leq \mathbb{P}\left(\|\mathbf{W}_{i,:}\|_2^2 \gtrsim m(\log n)\sigma^2\right) \overset{③}{=} e^{-c_0\log n} = n^{-c_0}, \tag{24}$$

where in ③ we use the fact that $\|\mathbf{W}_{i,:}\|_2^2/\sigma^2$ is a χ^2-RV with freedom m and invoke Lemma 13.

Phase II. Then we upper-bound ζ_2 as

$$\zeta_2 \leq \underbrace{\mathbb{P}\left(\frac{1}{n-h}\left|\left\langle \mathbf{W}_{i,:}, \sum_{k\neq i}\Omega_k^{\pi^*(i),j}\mathbf{W}_{k,:}\right\rangle\right| \gtrsim \frac{\sigma\sqrt{\log n}}{n}\left\|\sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\right\|_2\right)}_{\triangleq \zeta_{2,1}}$$

$$
+ \underbrace{\mathbb{P}\left(\Big\|\sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\Big\|_2^2 \gtrsim mnp(\log n)^3\sigma^2\right)}_{\triangleq \zeta_{2,2}}. \tag{25}
$$

For term $\zeta_{2,1}$, we exploit the independence across the rows of the matrix $\mathbf{W}$. Conditional on $\{\mathbf{W}_{k,:}\}_{k\neq i}$, we conclude the inner-product $\left\langle \mathbf{W}_{i,:}, \sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\right\rangle$ to be a Gaussian RV with zero mean and $\left\|\sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\right\|_2^2$ variance, which yields $\zeta_{2,1}\leq n^{-c}$. For term $\zeta_{2,2}$, we analyze the variance $\left\|\sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\right\|_2^2$, which reads as

$$
\begin{aligned}
\zeta_{2,2} \leq\ & \underbrace{\mathbb{P}\left(\Big\|\sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\Big\|_2^2 \gtrsim m(\log n)\sigma^2\Big[\sum_{k\neq i}(\Xi_k^{\pi^*(i),j})^2\Big],\ \sum_{k\neq i}(\Xi_k^{\pi^*(i),j})^2 \lesssim (\log n)^2 np\right)}_{\triangleq \zeta_{2,2,1}} \\
& + \underbrace{\mathbb{P}\left(\sum_{k\neq i}(\Xi_k^{\pi^*(i),j})^2 \gtrsim (\log n)^2 np\right)}_{\triangleq \zeta_{2,2,2}}.
\end{aligned} \tag{26}
$$

Due to the independence across $\mathbf{X}$ and $\mathbf{W}$, we can verify $\left\|\sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\right\|_2^2/[\sigma^2\sum_{k\neq i}(\Xi_k^{\pi^*(i),j})^2]$ to be a χ^2-RV with freedom m conditional on $\mathbf{X}$. Invoking Lemma 13, we can upper-bound ξ_1 as

$$
\zeta_{2,2,1} \leq \mathbb{P}\left(\Big\|\sum_{k\neq i}\Xi_k^{\pi^*(i),j}\mathbf{W}_{k,:}\Big\|_2^2 \gtrsim m(\log n)\sigma^2\Big[\sum_{k\neq i}(\Xi_k^{\pi^*(i),j})^2\Big]\right) \leq n^{-c}. \tag{27}
$$

As for ξ_2, we condition on event $\mathcal{E}_5$ and have

$$
\zeta_{2,2,2} \leq \mathbb{P}\left(\left\|\mathbf{X}\mathbf{X}_{\pi^*(i),:}\right\|_2 + \|\mathbf{X}\mathbf{X}_{j,:}\|_2 \gtrsim (\log n)\sqrt{np}\right) = 0. \tag{28}
$$

Then the proof is complete by combining (23), (24), (25), (26), (27), and (28).

□

4 SUPPORTING LEMMAS

Lemma 5. *For an arbitrary row $\mathbf{X}_{i,:}$, we have*

$$
\left\|\mathbf{B}^{*\top}\mathbf{X}_{i,:}\right\|_2 \lesssim \sqrt{\log n}\|\mathbf{B}^*\|_{\mathrm{F}},
$$

with probability exceeding $1-n^{-c}$.

Proof. This lemma is a direct consequence of the Hanson-wright inequality [Vershynin, 2018]. Easily we can verify $\mathbb{E}\left\|\mathbf{B}^{*\top}\mathbf{X}_{i,:}\right\|_2^2 = \|\mathbf{M}\|_{\mathrm{F}}^2$ and hence

$$
\begin{aligned}
\mathbb{P}\left(\left\|\mathbf{B}^{*\top}\mathbf{X}_{i,:}\right\|_2^2 \gtrsim \log n\|\mathbf{B}^*\|_{\mathrm{F}}^2\right) &\leq \mathbb{P}\left(\left|\left\|\mathbf{B}^{*\top}\mathbf{X}_{i,:}\right\|_2^2 - \|\mathbf{B}^*\|_{\mathrm{F}}^2\right| \gtrsim (\log n)\|\mathbf{B}^*\|_{\mathrm{F}}^2\right) \\
&\leq \exp\left(-c_0\min\left(\frac{\log n\|\mathbf{B}^*\|_{\mathrm{F}}^2}{\|\mathbf{B}^*\|_{\mathrm{OP}}^2}\wedge\frac{(\log^2 n)\|\mathbf{B}^*\|_{\mathrm{F}}^4}{\|\mathbf{B}^*\|_{\mathrm{F}}^4}\right)\right) \leq n^{-1-c}.
\end{aligned}
$$

Adopting the union bound, we have

$$
\mathbb{P}\left(\left\|\mathbf{B}^{*\top}\mathbf{X}_{i,:}\right\|_2^2 \gtrsim \log n\|\mathbf{B}^*\|_{\mathrm{F}}^2,\ \forall\, i\right) \leq n\cdot n^{-1-c} = n^{-c}.
$$

□

Lemma 6. *For an arbitrary row $\mathbf{X}_{i,:}$ (or $\mathbf{X}^{'}_{i,:}$), we have*

$$\begin{aligned}\left\langle \mathbf{X}_{i_1,:}, \mathbf{X}^{'}_{j_1,:}\right\rangle &\lesssim \sqrt{p\log n};\\ \langle \mathbf{X}_{i_2,:}, \mathbf{X}_{j_2,:}\rangle &\lesssim \sqrt{p\log n},\ \ i_2\neq j_2;\\ \left\langle \mathbf{X}^{'}_{i_3,:}, \mathbf{X}^{'}_{j_3,:}\right\rangle &\lesssim \sqrt{p\log n},\ \ i_3\neq j_3,\end{aligned}$$

hold with probability $1-n^{-c}$.

Lemma 7. *We conclude $\mathbb{P}\left(\mathcal{E}_4\right)\geq 1-1-ne^{-cnp}$.*

This lemma is a direct consequence of Lemma 13 and hence its proof is omitted.

Lemma 8. *Conditional on the intersection of events $\mathcal{E}_2\bigcap\mathcal{E}_3\bigcap\mathcal{E}_4$, we have $\mathbb{P}\left(\mathcal{E}_5\right)\geq 1-c_0 n^{-c_1}$.*

Proof. For a fixed row index s ($1\leq s\leq n$), we have

$$\begin{aligned}&\mathbb{P}\left(\|\mathbf{X}\mathbf{X}_{s,:}\|_2 \gtrsim (\log n)\sqrt{np}\right)\\ \overset{①}{\leq}\ &\mathbb{P}\left(\left\|\left(\mathbf{X}-\mathbf{X}_{\setminus(s)}\right)\mathbf{X}_{s,:}\right\|_2\gtrsim p\log n\right)+\mathbb{P}\left(\left\|\mathbf{X}_{\setminus(s)}\mathbf{X}_{s,:}\right\|_2\gtrsim(\log n)\sqrt{np}\right)\\ \overset{②}{\leq}\ &\underbrace{\mathbb{P}\left(\left(\|\mathbf{X}_{s,:}\|_2+\|\mathbf{X}^{'}_{s,:}\|_2\right)\|\mathbf{X}_{s,:}\|_2\gtrsim p\log n\right)}_{\triangleq\zeta_1}+\underbrace{\mathbb{P}\left(\left\|\mathbf{X}_{\setminus(s)}\mathbf{X}_{s,:}\right\|_2\gtrsim(\log n)\sqrt{np}\right)}_{\triangleq\zeta_2},\end{aligned}$$

where in ① we use the union bound and the fact $n\geq p$; and in ② we use the definition of $\mathbf{X}_{\setminus(s)}$ such that the difference $\mathbf{X}-\mathbf{X}_{\setminus(s)}$ only have non-zero elements in the sth column. Conditional on the intersection of events $\mathcal{E}_2\bigcap\mathcal{E}_3\bigcap\mathcal{E}_4$, we conclude that probability ζ_1 is zero and probability ζ_2 is upper-bounded as

$$\begin{aligned}\mathbb{P}\left(\left\|\mathbf{X}_{\setminus(s)}\mathbf{X}_{s,:}\right\|_2\gtrsim(\log n)\sqrt{np}\right)&\leq\mathbb{P}\left(\left|\left\|\mathbf{X}_{\setminus(s)}\mathbf{X}_{s,:}\right\|_2^2-\left\|\mathbf{X}_{\setminus(s)}\right\|_{\mathrm{F}}^2\right|\gtrsim(\log^2 n)np\right)\\ &\leq\exp\left(-c_0\left(\frac{(\log^2 n)np}{\left\|\mathbf{X}^{\top}_{\setminus(s)}\mathbf{X}_{\setminus(s)}\right\|_{\mathrm{OP}}}\wedge\frac{(\log n)^4n^2p^2}{\left\|\mathbf{X}^{\top}_{\setminus(s)}\mathbf{X}_{\setminus(s)}\right\|_{\mathrm{F}}^2}\right)\right)\leq n^{-c}.\end{aligned}$$

Thus the proof is completed by invoking the union bound since

$$\mathbb{P}\left(\|\mathbf{X}\mathbf{X}_{s,:}\|_2\gtrsim(\log n)\sqrt{np},\ \forall\ s\right)\leq n\cdot\mathbb{P}\left(\|\mathbf{X}\mathbf{X}_{s,:}\|_2\gtrsim(\log n)\sqrt{np}\right)\ \leq n\left(\zeta_1+\zeta_2\right)\leq n^{1-c}=n^{-c^{'}}.$$

□

Lemma 9. *Conditional on $\mathcal{E}_4$, we have $\mathbb{P}(\mathcal{E}_6)\geq 1-c_0p^{-2}$.*

Proof. We assume that the first h rows of $\mathbf{X}$ are permuted w.l.o.g. Due to the iid distribution of $\{\mathbf{X}_{i,:}\}_{i=1}^n$ and $\{\mathbf{X}^{'}_{i,:}\}_{i=1}^n$, we conclude

$$\mathbb{P}(\mathcal{E}_6)\leq n^2\mathbb{P}\left(\left\|\mathbf{B}^*-\widetilde{\mathbf{B}}\right\|_2\gtrsim\ \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right).\tag{29}$$

First, we expand $\mathbf{X}^{\top}\mathbf{\Pi}^*\mathbf{X}$ as

$$\mathbf{X}^{\top}\mathbf{\Pi}^*\mathbf{X}=\sum_{i=1}^h\mathbf{X}_{\pi(i),:}\mathbf{X}_{i,:}^{\top}+\sum_{i=h+1}^n\mathbf{X}_{i,:}\mathbf{X}_{i,:}^{\top},$$

and obtain

$$\mathbb{P}\left(\left\|\mathbf{B}^*-\widetilde{\mathbf{B}}\right\|_2\gtrsim\ \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right)$$

$$
\begin{aligned}
&\leq \mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{i=1}^{h}\mathbf{X}_{\pi(i),:}\mathbf{X}_{i,:}^{\top}\mathbf{B}^*\right\|_{\mathrm{F}} + \frac{1}{n-h}\left\|\sum_{i=h+1}^{n}\left(\mathbf{X}_{i,:}\mathbf{X}_{i,:}^{\top}-\mathbf{I}\right)\mathbf{B}^*\right\|_{\mathrm{F}} \gtrsim \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right)\\
&\overset{①}{\leq} \underbrace{\mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{i=1}^{h}\mathbf{X}_{\pi(i),:}\mathbf{X}_{i,:}^{\top}\mathbf{B}^*\right\|_{\mathrm{F}} \gtrsim \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right)}_{\zeta_1}\\
&+ \underbrace{\mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{i=h+1}^{n}\left(\mathbf{X}_{i,:}\mathbf{X}_{i,:}^{\top}-\mathbf{I}\right)\mathbf{B}^*\right\|_{\mathrm{F}} \gtrsim \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right)}_{\zeta_2},
\end{aligned}
$$

where ① is because of the union bound. The proof is complete by proving $\zeta_1 \leq 6n^{-2}p^{-2}$ and $\zeta_2 \leq 4n^{-2}p^{-2}$. The computation details come as follows.

Phase I: Bounding ζ_1. According to Lemma 8 in Pananjady et al. [2018] (restated as Lemma 14), we can decompose the set $\{j : \pi(j) \neq j\}$ into three disjoint sets $\mathcal{I}_i$, $1 \leq i \leq 3$, such that j and $\pi(j)$ does not lie in the same set. And the cardinality of set $\mathcal{I}_i$ is h_i satisfies $\lfloor h/5 \rfloor \leq h_i \leq h/3$. Adopting the union bound, we can upper-bound ζ_1 as

$$
\begin{aligned}
\zeta_1 &\leq \sum_{i=1}^{3}\mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{j\in\mathcal{I}_i}\mathbf{X}_{\pi(j),:}\mathbf{X}_{j,:}^{\top}\mathbf{B}^*\right\|_{\mathrm{F}} \gtrsim \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right)\\
&\leq \sum_{i=1}^{3}\mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{j\in\mathcal{I}_i}\mathbf{X}_{\pi(j),:}\mathbf{X}_{j,:}^{\top}\right\|_{\mathrm{OP}} \gtrsim \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\right).
\end{aligned} \tag{30}
$$

Defining $\mathbf{Z}_i$ as $\mathbf{Z}_i = \sum_{j\in\mathcal{I}_i}\mathbf{X}_{\pi(j),:}\mathbf{X}_{j,:}^{\top}$, we would bound the above probability by invoking the matrix Bernstein inequality (Theorem 7.3.1 in Tropp [2015]). First, we have

$$
\mathbb{E}\left(\mathbf{X}_{\pi(j),:}\mathbf{X}_{j,:}^{\top}\right) = \left(\mathbb{E}\mathbf{X}_{\pi(j),:}\right)\left(\mathbb{E}\mathbf{X}_{j,:}\right)^{\top} = \mathbf{0},
$$

due to the independence between $\mathbf{X}_{\pi(j),:}$ and $\mathbf{X}_{j,:}$. Then we upper bound $\left\|\mathbf{X}_{\pi(j),:}\mathbf{X}_{j,:}^{\top}\right\|_2$ as

$$
\left\|\mathbf{X}_{\pi(j),:}\mathbf{X}_{j,:}^{\top}\right\|_2 \overset{②}{=} \left\|\mathbf{X}_{\pi(j),:}\mathbf{X}_{j,:}^{\top}\right\|_{\mathrm{F}} \overset{③}{=} \left\|\mathbf{X}_{\pi(j),:}\right\|_2\left\|\mathbf{X}_{j,:}\right\|_2 \overset{④}{\lesssim} p\log n,
$$

where ② is because $\mathbf{X}_{\pi(j),:}\mathbf{X}_{j,:}^{\top}$ is rank-1, ③ is due to the fact $\left\|\boldsymbol{u}\boldsymbol{v}^{\top}\right\|_{\mathrm{F}}^2 = \mathrm{Tr}\left(\boldsymbol{u}\boldsymbol{v}^{\top}\boldsymbol{v}\boldsymbol{u}^{\top}\right) = \|\boldsymbol{u}\|_2^2\|\boldsymbol{v}\|_2^2$ for arbitrary vector $\boldsymbol{u},\boldsymbol{v} \in \mathbb{R}^p$, and ④ is because of event $\mathcal{E}_3$.

In the end, we compute $\mathbb{E}\left(\mathbf{Z}_i\mathbf{Z}_i^{\top}\right)$ and $\mathbb{E}\left(\mathbf{Z}_i^{\top}\mathbf{Z}_i\right)$ as

$$
\begin{aligned}
\mathbb{E}\left(\mathbf{Z}_i\mathbf{Z}_i^{\top}\right) &= \mathbb{E}\left(\sum_{j_1,j_2\in\mathcal{I}_i}\mathbf{X}_{\pi(j_1),:}\mathbf{X}_{j_1,:}^{\top}\mathbf{X}_{j_2,:}\mathbf{X}_{\pi(j_2),:}^{\top}\right) \overset{⑤}{=} \mathbb{E}\left(\sum_{j\in\mathcal{I}_i}\mathbf{X}_{\pi(j),:}\mathbf{X}_{j,:}^{\top}\mathbf{X}_{j,:}\mathbf{X}_{\pi(j),:}^{\top}\right)\\
&\overset{⑥}{=} \mathbb{E}\left(\sum_{j\in\mathcal{I}_i}\mathbf{X}_{\pi(j),:}\mathbb{E}\left(\mathbf{X}_{j,:}^{\top}\mathbf{X}_{j,:}\right)\mathbf{X}_{\pi(j),:}^{\top}\right) = p\left(\sum_{j\in\mathcal{I}_i}\mathbb{E}\mathbf{X}_{\pi(j),:}\mathbf{X}_{\pi(j),:}^{\top}\right) = ph_i\mathbf{I}_{p\times p} = \mathbb{E}\left(\mathbf{Z}\mathbf{Z}^{\top}\right),
\end{aligned}
$$

where ⑤ and ⑥ is because of the fact such that j and $\pi(j)$ are not within the set $\mathcal{I}_i$ simultaneously. To sum up, we invoke the matrix Bernstein inequality (Theorem 7.3.1 in Tropp [2015]) and have

$$
\begin{aligned}
\frac{1}{n-h}\left\|\sum_{j\in\mathcal{I}}\mathbf{X}_{\pi(j),:}\mathbf{X}_{j,:}^{\top}\right\|_{\mathrm{OP}} &\leq \frac{p(\log n)\log(n^2p^3)}{3(n-h)} + \frac{\sqrt{p^2(\log^2 n)\log^2\left(n^2p^3\right) + 18ph_i\log\left(n^2p^3\right)}}{(n-h)}\\
&\overset{⑦}{\lesssim} \frac{p(\log n)\log(n^2p^3)}{n} + \frac{p}{n}\sqrt{(\log^2 n)\log^2\left(n^2p^3\right) + \frac{n}{p}(\log n^2p^3)}
\end{aligned}
$$

$$\overset{⑧}{\lesssim} \frac{p(\log n)\log(n^2p^3)}{n} + \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}} \overset{⑨}{\lesssim} \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}$$

holds with probability $1 - 2(np)^{-2}$, where in ⑦, ⑧, and ⑨ we use the fact such that $h \leq n/4$, $h_i \leq h/3$. Hence we can show ζ_1 in (30) to be less than $6n^{-2}p^{-2}$.

Phase II: Bounding ζ_2. We upper bound ζ_2 as

$$\begin{aligned}\zeta_2 &\leq \mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{i=h+1}^{n}\left(\mathbf{X}_{i,:}\mathbf{X}_{i,:}^\top - \mathbf{I}\right)\mathbf{B}^*\right\|_{\mathrm{F}} \gtrsim \frac{(\log n)(\log n^2p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right) \\ &\leq \mathbb{P}\left(\left\|\sum_{i=h+1}^{n}\left(\mathbf{X}_{i,:}\mathbf{X}_{i,:}^\top - \mathbf{I}\right)\right\|_{\mathrm{OP}} \gtrsim (\log n)(\log n^2p^3)\sqrt{np}\right).\end{aligned}$$

Similar to above, we define $\widetilde{\mathbf{Z}}_i = \mathbf{X}_{i,:}\mathbf{X}_{i,:}^\top - \mathbf{I}$. First, we verify that $\mathbb{E}\widetilde{\mathbf{Z}}_i = \mathbf{0}$ and $\mathbf{Z}_i$ are independent. Then we bound $\|\mathbf{Z}\|_{\mathrm{OP}}$ as

$$\|\mathbf{Z}\|_{\mathrm{OP}} \leq \left\|\mathbf{X}_{i,:}\mathbf{X}_{i,:}^\top\right\|_{\mathrm{OP}} + \|\mathbf{I}\|_{\mathrm{OP}} \overset{Ⓐ}{=} \|\mathbf{X}_{i,:}\|_2^2 + 1 \overset{Ⓑ}{\lesssim} p\log n + 1 \lesssim p\log n,$$

where in Ⓐ we use $\left\|\boldsymbol{u}\boldsymbol{u}^\top\right\|_{\mathrm{OP}} = \|\boldsymbol{u}\|_2^2$ for arbitrary vector $\boldsymbol{u}$, in Ⓑ we condition on event $\mathcal{E}_4$. In the end, we compute $\mathbb{E}(\mathbf{Z}_i\mathbf{Z}_i^\top)$ as

$$\mathbb{E}\left(\mathbf{Z}_i\mathbf{Z}_i^\top\right) = \mathbb{E}\left(\|\mathbf{X}_{i,:}\|_2^2\mathbf{X}_{i,:}\mathbf{X}_{i,:}^\top\right) - \mathbf{I} \preceq p\log n\left(\mathbb{E}\left(\mathbf{X}_{i,:}\mathbf{X}_{i,:}^\top\right)\right) - \mathbf{I} \preceq (p\log n)\mathbf{I}.$$

Invoking the matrix Bernstein inequality (Theorem 7.3.1 in Tropp [2015]), we conclude

$$\zeta_2 \leq 4p\exp\left(-\frac{3n(\log n)\log^2\left(n^2p^3\right)}{\sqrt{np}(\log n)\log\left(n^2p^3\right) + 6}\right) \overset{Ⓒ}{\leq} 4n^{-2}p^{-2},$$

where in Ⓒ we use the fact $n \gtrsim p$. □

Lemma 10. *Conditional on the intersection of events $\mathcal{E}_2 \bigcap \mathcal{E}_3$, we conclude*

$$\left\|\left(\widetilde{\mathbf{B}} - \widetilde{\mathbf{B}}_{\setminus(s)}\right)^\top \mathbf{X}_{s,:}\right\|_2 \lesssim \frac{p\log n}{n}\|\mathbf{B}^*\|_{\mathrm{F}}.$$

Proof. Here we focus on the case when $\pi(s) = s$. The proof of the case when $\pi(s) \neq s$ can be completed effortless by following a similar strategy. First, we notice

$$\begin{aligned}\left\|\left(\widetilde{\mathbf{B}} - \widetilde{\mathbf{B}}_{\setminus(s)}\right)^\top \mathbf{X}_{s,:}\right\|_2 &= (n-h)^{-1}\left\|\mathbf{B}^{*\top}\left(\widetilde{\mathbf{X}}_{s,:}\widetilde{\mathbf{X}}_{s,:}^\top - \mathbf{X}_{s,:}\mathbf{X}_{s,:}^\top\right)\mathbf{X}_{s,:}\right\|_2 \\ &\leq (n-h)^{-1}\left(\left|\left\langle \mathbf{X}_{s,:}, \widetilde{\mathbf{X}}_{s,:}\right\rangle\right| \left\|\mathbf{B}^{*\top}\widetilde{\mathbf{X}}_{s,:}\right\|_2 + \|\mathbf{X}_{s,:}\|_2^2 \cdot \left\|\mathbf{B}^{*\top}\mathbf{X}_{s,:}\right\|_2\right).\end{aligned}$$

Conditional on the intersection of events $\mathcal{E}_2 \bigcap \mathcal{E}_3$, we conclude

$$\left\|\left(\widetilde{\mathbf{B}}_{\setminus(s)} - \widetilde{\mathbf{B}}\right)^\top \mathbf{X}_{s,:}\right\|_2 \lesssim \frac{p\log n}{n-h}\|\mathbf{B}^*\|_{\mathrm{F}} \asymp \frac{p\log n}{n}\|\mathbf{B}^*\|_{\mathrm{F}}.$$

□

Following the same strategy, we can prove that

Lemma 11. *Conditional on the intersection of events $\mathcal{E}_2 \bigcap \mathcal{E}_3$, we conclude*

$$\left\|\left(\widetilde{\mathbf{B}} - \widetilde{\mathbf{B}}_{\setminus(s,t)}\right)^\top \mathbf{X}_{s,:}\right\|_2 \lesssim \frac{p\log n}{n}\|\mathbf{B}^*\|_{\mathrm{F}}.$$

Lemma 12. *Conditional on the intersection of events $\mathcal{E}_6 \bigcap \mathcal{E}_7 \bigcap \mathcal{E}_8$, we conclude $\mathbb{P}(\mathcal{E}_9) \geq 1 - c_0 n^{-c_1}$.*

Proof. We adopt the leave-one-out trick and construct the matrix $\widetilde{\mathbf{B}}_{\backslash(i)}$ as

$$\widetilde{\mathbf{B}}_{\backslash(i)} = (n-h)^{-1}\Bigg(\sum_{\substack{k\neq i \\ \pi^*(k)\neq i}} \mathbf{X}_{\pi(k),:}\mathbf{X}_{k,:}^{\top} + \sum_{\substack{k=i \\ \pi^*(k)\neq i}} \widetilde{\mathbf{X}}_{\pi(k),:}\widetilde{\mathbf{X}}_{k,:}^{\top}\Bigg)\mathbf{B}^*,$$

where $\widetilde{\mathbf{X}}_{i,:}$ are the independent copy of $\mathbf{X}_{i,:}$. Adopting the union bound, we conclude

$$\begin{aligned}
&\mathbb{P}\left(\left\|(\widetilde{\mathbf{B}} - \mathbf{B}^*)^{\top}\mathbf{X}_{i,:}\right\|_2 \gtrsim \frac{(\log n)^{3/2}(\log n^2 p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right) \\
&\leq \mathbb{P}\left(\left\|(\mathbf{B}^* - \widetilde{\mathbf{B}}_{\backslash(i)})^{\top}\mathbf{X}_{i,:}\right\|_2 + \left\|(\widetilde{\mathbf{B}}_{\backslash(i)} - \widetilde{\mathbf{B}})^{\top}\mathbf{X}_{i,:}\right\|_2 \gtrsim \frac{(\log n)^{3/2}(\log n^2 p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right) \\
&\leq \underbrace{\mathbb{P}\left(\left\|(\mathbf{B}^* - \widetilde{\mathbf{B}}_{\backslash(i)})^{\top}\mathbf{X}_{i,:}\right\|_2 \gtrsim \frac{(\log n)^{3/2}(\log n^2 p^3)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^*\|_{\mathrm{F}}\right)}_{\triangleq \zeta_1} \\
&+ \underbrace{\mathbb{P}\left(\left\|(\widetilde{\mathbf{B}}_{\backslash(i)} - \widetilde{\mathbf{B}})^{\top}\mathbf{X}_{i,:}\right\|_2 \gtrsim \frac{p\log n}{n}\|\mathbf{B}^*\|_{\mathrm{F}}\right)}_{\triangleq \zeta_2}.
\end{aligned}$$

First, we study the probability ζ_1. Due to the construction of $\widetilde{\mathbf{B}}_{\backslash(i)}$, we have $\mathbf{X}_{i,:}$ to be independent of $\mathbf{B}^* - \widetilde{\mathbf{B}}_{\backslash(i)}$. Conditional on $\mathbf{B}^* - \widetilde{\mathbf{B}}_{\backslash(i)}$, we conclude

$$\zeta_1 \overset{①}{\leq} \mathbb{P}\left(\left\|(\mathbf{B}^* - \widetilde{\mathbf{B}}_{\backslash(i)})^{\top}\mathbf{X}_{i,:}\right\|_2 \geq \sqrt{\log n}\left\|\mathbf{B}^* - \widetilde{\mathbf{B}}_{\backslash(i)}\right\|_{\mathrm{F}}\right) \leq n^{-c},$$

where in ① we condition on event $\mathcal{E}_6$ such that $\left\|\mathbf{B}^* - \widetilde{\mathbf{B}}_{\backslash(i)}\right\|_{\mathrm{F}} \lesssim (\log n)(\log n^2 p^3)\sqrt{p/n}\|\mathbf{B}^*\|_{\mathrm{F}}$. As for probability ζ_2, we conclude it to be zero conditional on $\mathcal{E}_7$. Thus the proof is completed. □

5 SUPPLEMENTARY MATERIAL: USEFUL FACTS

This section lists some useful facts for the sake of self-containing.

Lemma 13. *For a χ^2-RV Z with ℓ freedom, we have*

$$\begin{aligned}
\mathbb{P}(Z \leq t) &\leq \exp\left(\frac{\ell}{2}\left(\log\frac{t}{\ell} - \frac{t}{\ell} + 1\right)\right), \ \ t < \ell; \\
\mathbb{P}(Z \geq t) &\leq \exp\left(\frac{\ell}{2}\left(\log\frac{t}{\ell} - \frac{t}{\ell} + 1\right)\right), \ \ t > \ell.
\end{aligned}$$

Lemma 14 (Lemma 8 in Pananjady et al. [2018])**.** *Consider an arbitrary permutation map π with Hamming distance k from the identity map, i.e., $\mathsf{d}_{\mathsf{H}}(\boldsymbol{\pi}, \mathbf{I}) = h$. We define the index set $\{i: \ i \neq \pi(i)\}$ and can decompose it into 3 independent sets $\mathcal{I}_i$ $(1 \leq i \leq 3)$ such that the cardinality of each set satisfies $|\mathcal{I}_i| \geq \lfloor h/3 \rfloor \geq h/5$.*

Lemma 15 (Theorem 1.3 in Paouris [2012])**.** *Let $\boldsymbol{g} \in \mathbb{R}^n$ be an isotropic log-concave random vector with sub-gaussian constant K, and $\mathbf{A}$ is a non-zero $n \times n$ matrix. For any $\boldsymbol{y} \in \mathbb{R}^n$ and $\varepsilon \in (0, c_1)$, one has*

$$\mathbb{P}\left(\|\boldsymbol{y} - \mathbf{A}\boldsymbol{g}\|_2 \leq \varepsilon\|\mathbf{A}\|_{\mathrm{F}}\right) \leq \exp\left(\kappa(K)\mathrm{srank}(\mathbf{A})\log\varepsilon\right),$$

where $\kappa = c_1/K^2$.

References

Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM J. Optim.*, 30(4):3098–3121, 2020.

Noureddine El Karoui. Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results. *arXiv preprint arXiv:1311.2445*, 2013.

Noureddine EL Karoui. On the impact of predictor geometry on the performance on high-dimensional ridge-regularized generalized robust regression estimators. *Probability Theory and Related Fields*, 170(1-2):95–175, 2018.

Noureddine El Karoui, Derek Bean, Peter J Bickel, Chinghway Lim, and Bin Yu. On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences*, 110(36):14557–14562, 2013.

Ashwin Pananjady, Martin J. Wainwright, and Thomas A. Courtade. Linear regression with shuffled data: Statistical and computational limits of permutation recovery. *IEEE Trans. Inf. Theory*, 64(5):3286–3300, 2018.

Grigoris Paouris. Small ball probability estimates for log-concave measures. *Transactions of the American Mathematical Society*, 364(1):287–308, 2012.

Pragya Sur, Yuxin Chen, and Emmanuel J Candès. The likelihood ratio test in high-dimensional logistic regression is asymptotically a rescaled chi-square. *Probability Theory and Related Fields*, 175(1):487–558, 2019.

Joel A. Tropp. An introduction to matrix concentration inequalities. *Found. Trends Mach. Learn.*, 8(1-2):1–230, 2015.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.