A. Notations

We begin the appendix with a restatement of the notations. Denote c, $c^{'}$, c_i as some universal positive constants. Notice that their values may not necessarily the same even for those with same notations. We denote $a \lesssim b$ if there exists some positive constant $c_0 > 0$ such that $a \leq c_0 b$. Similarly we define $a \gtrsim b$ provided $a \geq c_0 b$ for some positive constant c_0. We write $a \asymp b$ when $a \lesssim b$ and $a \gtrsim b$ hold simultaneously.

For an arbitrary matrix $\mathbf{X}$, we denote $\mathbf{X}_{i,:}$ as the i-th row, $\mathbf{X}_{:,i}$ as its i-th column, and X_{ij} as the (i,j)-th element. The Frobenius norm of $\mathbf{X}$ is defined as $\|\mathbf{X}\|_{\mathsf{F}}$ while the operator norm is denoted as $\|\mathbf{X}\|_{\mathsf{OP}}$, whose definition can be found in Section 2.3 of Golub and Loan (2013) (P71). Its stable rank $\rho(\mathbf{X})$ is defined as the ratio $\|\mathbf{X}\|_{\mathsf{F}}^2/\|\mathbf{X}\|_{\mathsf{OP}}^2$ (Section 2.1.15 in Tropp (2015)). The inner product $\langle \mathbf{A}, \mathbf{C}\rangle$ is defined as $\sum_{ij} A_{ij} C_{ij}$.

Associate with each permutation matrix $\mathbf{\Pi}$, we define the operator $\pi(\cdot)$ that transforms index i to $\pi(i)$. The Hamming distance $\mathsf{d_H}(\mathbf{\Pi}_1, \mathbf{\Pi}_2)$ between permutation matrix $\mathbf{\Pi}_1$ and $\mathbf{\Pi}_2$ is defined as $\mathsf{d_H}\left(\mathbf{\Pi}_1, \mathbf{\Pi}_2\right) = \sum_{i=1}^n \mathbb{1}\left(\pi_1(i) \neq \pi_2(i)\right)$. Additionally, we denote $\overline{\mathcal{E}}$ as the complement of the event $\mathcal{E}$ and the *signal-to-noise-ratio* (SNR) as $\mathsf{SNR} = \left\|\mathbf{B}^{\natural}\right\|_{\mathsf{F}}^2/(m\sigma^2)$.

B. Problem Restatement

To begin with, we recall the problem formulation, which reads as

$$\mathbf{Y} = \mathbf{\Pi}^{\natural}\mathbf{X}\mathbf{B}^{\natural} + \mathbf{W},$$

where $\mathbf{Y} \in \mathbb{R}^{n\times m}$ represents the observation, $\mathbf{\Pi} \in \mathbb{R}^{n\times n}$ denotes the unknown permutation matrix, $\mathbf{X} \in \mathbb{R}^{n\times p}$ is the sensing matrix (design matrix) with $X_{ij} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$ being a standard normal random variable (RV), $\mathbf{B}^{\natural} \in \mathbb{R}^{p\times m}$ is the matrix of regression coefficients, and $\mathbf{W} \in \mathbb{R}^{n\times m}$ is the additive Gaussian noise matrix such that $W_{ij} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,\sigma^2)$.

Our goal is to reconstruct the pair $(\widehat{\mathbf{\Pi}}, \widehat{\mathbf{B}})$ from the observation $\mathbf{Y}$ and sensing matrix (design matrix) $\mathbf{X}$. The proposed one-step estimator can be written as

$$\begin{aligned}
\widehat{\mathbf{\Pi}} &= \operatorname{argmax}_{\mathbf{\Pi}\in\mathcal{P}_n} \left\langle \mathbf{\Pi}, \mathbf{Y}\mathbf{Y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle, \\
\widehat{\mathbf{B}} &= (\mathbf{X})^{\dagger}\, \widehat{\mathbf{\Pi}}^{\top}\mathbf{Y},
\end{aligned}$$

where $\mathbf{X}^{\dagger} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}$ denotes the pseudo-inverse of $\mathbf{X}$. In the following, we will separately investigate its properties under the single observation model $(m=1)$ and multiple observations model $(m>1)$. The formal statement is packaged in Theorem 1 and Theorem 2.

C. Appendix for Section 3

This section focuses on the special case where $p=1, m=1$. Consider $\mathbf{X} \in \mathbb{R}^n$ to be a Gaussian distributed RV such that $\mathbf{X} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}_{n\times n}\right)$, and permutation matrix $\mathbf{\Pi}^{\natural}$ which satisfies $\mathsf{d_H}(\mathbf{I}, \mathbf{\Pi}^{\natural}) = h \leq n/4$.

C.1. Notations

First we define the following events $\mathcal{E}_i$, $(1 \leq i \leq 5)$, which reads

$$\begin{aligned}
\mathcal{E}_1 &\triangleq \left\{\left\langle \mathbf{X}, \mathbf{\Pi}^{\natural}\mathbf{X}\right\rangle \geq c_0 n\right\}, \\
\mathcal{E}_2 &\triangleq \left\{\|\mathbf{X}\|_2 \leq 2\sqrt{n}\right\} \\
\mathcal{E}_3(\mathbf{\Pi}) &\triangleq \left\{\mathbf{W}^{\top}\mathbf{X}\mathbf{X}^{\top}\left(\mathbf{\Pi}^{\natural} - \mathbf{\Pi}\right)\mathbf{W} \lesssim \sigma^2 n^2 \log n\right\}, \\
\mathcal{E}_4(\mathbf{\Pi}) &\triangleq \left\{\left|\langle \mathbf{W}, \mathbf{X}\rangle \left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \left(\mathbf{\Pi}^{\natural} - \mathbf{\Pi}\right)^{\top}\mathbf{X}\right\rangle + \left\langle \mathbf{W}, \left(\mathbf{\Pi}^{\natural} - \mathbf{\Pi}\right)^{\top}\mathbf{X}\right\rangle \left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \mathbf{X}\right\rangle\right| \lesssim \sigma n^2\sqrt{\log n}\right\} \\
\mathcal{E}_5(\mathbf{\Pi}; \ell) &\triangleq \left\{\|\mathbf{X} - \mathbf{\Pi}\mathbf{X}\|_2^2 \geq \frac{12\ell}{5en^{20}},\ \mathsf{d_H}\left(\mathbf{I}, \mathbf{\Pi}\right) = \ell\right\},
\end{aligned}$$

where $\mathbf{\Pi}$ is an arbitrary permutation matrix, and $c_0 > 0$ is some positive constant.

C.2. Outline of proof

We will prove that ground truth permutation matrix $\mathbf{\Pi}^{\natural}$ will be returned with high probability under the assumptions in Theorem 1. The formal statement is shown in Theorem 1. Before we delve into the proof details, we give a roadmap of the proof, which is

- **Step I:** Under the events $\mathcal{E}_1 \bigcap_{\mathbf{\Pi}} \left(\mathcal{E}_3(\mathbf{\Pi}) \bigcap \mathcal{E}_4(\mathbf{\Pi}) \bigcap \mathcal{E}_5(\mathbf{\Pi}; \ell)\right)$, we have
$$\left\langle \mathbf{\Pi}^{\natural}, \mathbf{y}\mathbf{y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle - \left\langle \mathbf{\Pi}, \mathbf{y}\mathbf{y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle \gtrsim \frac{c_0\beta^2}{n^{19}} - c_1\beta\sigma n^2\sqrt{\log n} - c_2\sigma^2 n^2\log n.$$
Notice that under assumptions in Theorem 1, we conclude that $\left\langle \mathbf{\Pi}^{\natural}, \mathbf{y}\mathbf{y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle > \left\langle \mathbf{\Pi}, \mathbf{y}\mathbf{y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle$, $\forall\, \mathbf{\Pi}$, which suggests that $\mathbf{\Pi}^{\natural}$ will always be returned by our estimator in Eq. (3).
- **Step II:** We upper-bound the probability $\mathbb{P}(\widehat{\mathbf{\Pi}} \neq \mathbf{\Pi}^{\natural})$ by $\mathbb{P}\left(\overline{\mathcal{E}}_1 \bigcup_{\mathbf{\Pi}} \left(\overline{\mathcal{E}}_3(\mathbf{\Pi}) \bigcup \overline{\mathcal{E}}_4(\mathbf{\Pi}) \bigcup \overline{\mathcal{E}}_5(\mathbf{\Pi}; \ell)\right)\right)$ and complete the proof by showing it is at most cn^{-1}.

Having illustrated the proof strategy, we turn to the proof details. The main proof is attached in Section C.3 while the supporting lemmas bounding $\mathbb{P}(\mathcal{E}_i)$, $(1 \le i \le 5)$, are put in Section C.4.

C.3. Proof of Theorem 1

Proof 1 *For an arbitrary permutation matrix $\mathbf{\Pi}$, we can expand the term $\left\langle \mathbf{\Pi}, \mathbf{y}\mathbf{y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle$ as*
$$\left\langle \mathbf{\Pi}, \mathbf{y}\mathbf{y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle = \mathcal{T}_1(\mathbf{\Pi}) + \beta\mathcal{T}_2(\mathbf{\Pi}) + \beta^2\mathcal{T}_3(\mathbf{\Pi}),$$
where $\mathcal{T}_i(\mathbf{\Pi})$, $(1 \le i \le 3)$, are defined as
$$\begin{aligned}
\mathcal{T}_1(\mathbf{\Pi}) &= \left\langle \mathbf{W}, \mathbf{\Pi}^{\top}\mathbf{X}\right\rangle \left\langle \mathbf{X}, \mathbf{W}\right\rangle; \\
\mathcal{T}_2(\mathbf{\Pi}) &= \left\langle \mathbf{W}, \mathbf{X}\right\rangle \left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \mathbf{\Pi}^{\top}\mathbf{X}\right\rangle + \left\langle \mathbf{W}, \mathbf{\Pi}^{\top}\mathbf{X}\right\rangle \left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \mathbf{X}\right\rangle; \\
\mathcal{T}_3(\mathbf{\Pi}) &= \left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \mathbf{\Pi}\mathbf{X}\right\rangle \left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \mathbf{X}\right\rangle.
\end{aligned}$$

Step I: *We rewrite the difference $\left\langle \mathbf{\Pi}^{\natural}, \mathbf{y}\mathbf{y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle - \left\langle \mathbf{\Pi}, \mathbf{y}\mathbf{y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle$ as*
$$\begin{aligned}
&\left\langle \mathbf{\Pi}^{\natural}, \mathbf{y}\mathbf{y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle - \left\langle \mathbf{\Pi}, \mathbf{y}\mathbf{y}^{\top}\mathbf{X}\mathbf{X}^{\top}\right\rangle \\
&= \mathcal{T}_1(\mathbf{\Pi}^{\natural}) - \mathcal{T}_1(\mathbf{\Pi}) + \beta\left(\mathcal{T}_2\left(\mathbf{\Pi}^{\natural}\right) - \mathcal{T}_2(\mathbf{\Pi})\right) + \beta^2\left(\mathcal{T}_3\left(\mathbf{\Pi}^{\natural}\right) - \mathcal{T}_3(\mathbf{\Pi})\right) \\
&\overset{①}{=} \frac{\beta^2}{2}\left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \mathbf{X}\right\rangle \left\|\mathbf{X} - \mathbf{\Pi}^{\natural\top}\mathbf{\Pi}\mathbf{X}\right\|_2^2 + \beta\left(\mathcal{T}_2\left(\mathbf{\Pi}^{\natural}\right) - \mathcal{T}_2(\mathbf{\Pi})\right) + \mathcal{T}_1(\mathbf{\Pi}^{\natural}) - \mathcal{T}_1(\mathbf{\Pi}) \\
&\overset{②}{\ge} \frac{\beta^2}{2}c_0 n\frac{24}{5en^{20}} - \beta\left|\mathcal{T}_2\left(\mathbf{\Pi}^{\natural}\right) - \mathcal{T}_2(\mathbf{\Pi})\right| - \left|\mathcal{T}_1(\mathbf{\Pi}^{\natural}) - \mathcal{T}_1(\mathbf{\Pi})\right| \\
&\overset{③}{\gtrsim} \frac{c_0\beta^2}{n^{19}} - c_1\beta\sigma n^2\sqrt{\log n} - c_2\sigma^2 n^2\log n \overset{④}{>} 0,
\end{aligned}$$
where in ① we rewrite $\|\mathbf{X}\|_2^2 - \left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \mathbf{\Pi}\mathbf{X}\right\rangle$ as
$$\|\mathbf{X}\|_2^2 - \left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \mathbf{\Pi}\mathbf{X}\right\rangle = \frac{1}{2}\left(\|\mathbf{X}\|_2^2 + \left\|\mathbf{\Pi}^{\natural\top}\mathbf{\Pi}\mathbf{X}\right\|_2^2 - 2\left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \mathbf{\Pi}\mathbf{X}\right\rangle\right) = \frac{1}{2}\left\|\mathbf{X} - \mathbf{\Pi}^{\natural\top}\mathbf{\Pi}\mathbf{X}\right\|_2^2,$$
in ② we condition on event $\mathcal{E}_1, \mathcal{E}_5(\mathbf{\Pi}; \ell)$ and have $\|\mathbf{X} - \mathbf{\Pi}\mathbf{X}\|_2^2 \ge \frac{12\ell}{5en^{20}} \ge \frac{24}{5en^{20}}$, in ③ we condition on $\mathcal{E}_3(\mathbf{\Pi}), \mathcal{E}_4(\mathbf{\Pi})$, and in ④ we use the assumption $\log(\mathsf{SNR}) \gtrsim \log n$ in Theorem 1.

Step II: *The error probability* $\mathbb{P}\left(\widehat{\mathbf{\Pi}} \neq \mathbf{\Pi}^{\natural}\right)$ *is hence be upper-bounded as*

$$
\begin{aligned}
&\mathbb{P}\left(\widehat{\mathbf{\Pi}} \neq \mathbf{\Pi}^{\natural}\right) \leq \mathbb{P}\left(\overline{\mathcal{E}}_1 \bigcup_{\mathbf{\Pi}} \left(\overline{\mathcal{E}}_3(\mathbf{\Pi}) \bigcup \overline{\mathcal{E}}_4(\mathbf{\Pi}) \bigcup \overline{\mathcal{E}}_5(\mathbf{\Pi};\ell)\right)\right) \\
&\overset{⑤}{\leq} \mathbb{P}\left(\bigcup_{\mathbf{\Pi}} \left(\overline{\mathcal{E}}_3(\mathbf{\Pi}) \bigcup \overline{\mathcal{E}}_4(\mathbf{\Pi}) \bigcup \overline{\mathcal{E}}_5(\mathbf{\Pi})\right) \bigcap \mathcal{E}_1 \bigcap \mathcal{E}_2\right) + \mathbb{P}\left(\overline{\mathcal{E}}_1\right) + \mathbb{P}\left(\overline{\mathcal{E}}_2\right) \\
&\overset{⑥}{\leq} \sum_{\mathbf{\Pi}^{\natural} \neq \mathbf{\Pi}} \mathbb{P}\left(\overline{\mathcal{E}}_3(\mathbf{\Pi}) \bigcap \mathcal{E}_1 \bigcap \mathcal{E}_2\right) + \sum_{\mathbf{\Pi}^{\natural} \neq \mathbf{\Pi}} \mathbb{P}\left(\overline{\mathcal{E}}_4(\mathbf{\Pi}) \bigcap \mathcal{E}_1 \bigcap \mathcal{E}_2\right) \\
&+ \sum_{\ell \geq 2} \mathbb{P}\left(\overline{\mathcal{E}}_5(\mathbf{\Pi};\ell) \bigcap \mathcal{E}_1 \bigcap \mathcal{E}_2\right) + 8n^{-1} + 2e^{-c_0 n} \\
&\overset{⑦}{\leq} 2n^{-n} + 3\sum_{\ell \geq 2} \binom{n}{\ell} \ell! n^{-2\ell} + 8n^{-1} + 2e^{-c_0 n} \\
&\overset{⑧}{\lesssim} c_0 n^{-n} + n^{-1} + 3\sum_{\ell \geq 2} n^{\ell} n^{-2\ell} \lesssim c_0 n^{-1} + \frac{3}{n(n-1)} \lesssim n^{-1},
\end{aligned}
$$

where in ⑤ we use the union bound, in ⑥ we complete the proof with Lemma 1 and the fact $\mathbb{P}\left(\overline{\mathcal{E}}_2\right) \leq e^{-0.8n}$*, in ⑦ we invoke Lemma 2, Lemma 3, Lemma 4, and in ⑧ we use* $n!/(n-\ell)! \leq n^{\ell}$ *and complete the proof.*

C.4. Supporting Lemmas for Theorem 1

This subsection collects the supporting lemmas for the proof of Theorem 1.

Lemma 1 *We have* $\mathbb{P}\left(\overline{\mathcal{E}}_1\right) \leq 8n^{-1} + e^{-0.238n}$ *when* n *is sufficiently large.*

Proof 2 *Different from the proof in Lemma 9, we consider the case where* $\mathbf{X} \in \mathbb{R}^n$ *is a vector and would lower-bound* $\left\langle \mathbf{X}, \mathbf{\Pi}^{\natural}\mathbf{X} \right\rangle$*. W.l.o.g, we assume the first* h *entries are permuted and expand the inner product* $\left\langle \mathbf{X}, \mathbf{\Pi}^{\natural}\mathbf{X} \right\rangle$ *as*

$$
\left\langle \mathbf{X}, \mathbf{\Pi}^{\natural}\mathbf{X} \right\rangle = \sum_{i=1}^{h} X_i X_{\pi(i)} + \sum_{i=h+1}^{n} X_i^2.
$$

With union bound, we can upper bound $\mathbb{P}\left(\left\langle \mathbf{X}, \mathbf{\Pi}^{\natural}\mathbf{X} \right\rangle \leq c_0 n\right)$ *as*

$$
\mathbb{P}\left(\left\langle \mathbf{X}, \mathbf{\Pi}^{\natural}\mathbf{X} \right\rangle \leq c_0 n\right) \overset{①}{\leq} \underbrace{\mathbb{P}\left(\sum_{i=h+1}^{n} X_i^2 \leq \frac{1}{4}(n-h)\right)}_{\zeta_1} + \underbrace{\mathbb{P}\left(\sum_{i=1}^{h} X_i X_{\pi(i)} \leq -\frac{4\sqrt{2}+\sqrt{35}}{\sqrt{2}}\sqrt{n\log n}\right)}_{\zeta_2},
$$

where $c_0 > 0$ *is some positive constant, in ① we use the fact*

$$
\frac{n-h}{4} - \frac{4\sqrt{2}+\sqrt{35}}{\sqrt{2}}\sqrt{n\log n} \overset{(h \leq \frac{n}{4})}{\geq} \frac{3n}{16} - \frac{4\sqrt{2}+\sqrt{35}}{\sqrt{2}}\sqrt{n\log n} \geq c_0 n,
$$

when n *is large. We finish the proof by separately upper-bounding* $\zeta_1 \leq e^{-0.2386n}$ *and* $\zeta_2 \leq 8n^{-1}$*. The detailed computation comes as follows.*

Phase I: *For* ζ_1*, we can view* $\sum_{i=h+1}^{n} X_i^2$ *as a* χ^2*-RV with* $(n-h)$ *freedom and have*

$$
\zeta_1 \overset{②}{\leq} \exp\left(\frac{n-h}{2}\left(\log\frac{1}{4} - \frac{1}{4} + 1\right)\right) \overset{③}{\leq} e^{-0.2386n},
$$

where in ② we use Lemma 11, and ③ is because $h \le n/4$.

Phase II: *To bound ζ_2, we divide the index set $\{j : j \neq \pi(j)\}$ into 3 disjoint sets $\mathcal{I}_i$, $1 \le i \le 3$, as in Lemma 8 in Pananjady et al. (2017a) (restated as Lemma 13). This division has two properties: (i) indices j and $\pi(j)$ lies in different sets; (ii) the cardinality h_i of each $\mathcal{I}_i$ satisfies $\lfloor h/5 \rfloor \le h_i \le h/3$. Then we obtain*

$$\zeta_2 \le \mathbb{P}\left(\sum_{i=1}^{h} X_i X_{\pi(i)} \le -\frac{4\sqrt{2}+\sqrt{35}}{\sqrt{2}}\sqrt{n\log n},\ |X_i| \le 2\sqrt{\log n},\ \forall\, i\right) + \mathbb{P}\left(|X_i| \ge 2\sqrt{\log n},\ \exists\, i\right)$$
$$\overset{④}{\le} \sum_{i=1}^{3} \underbrace{\mathbb{P}\left(\sum_{j\in\mathcal{I}_i} X_j X_{\pi(j)} \le -\frac{4\sqrt{2}+\sqrt{35}}{3\sqrt{2}}\sqrt{n\log n},\ |X_i| \le 2\sqrt{\log n},\ \forall\, i\right)}_{\zeta_{2,i}} + n\underbrace{\mathbb{P}\left(|X_i| \ge 2\sqrt{\log n}\right)}_{\le 2n^{-2}},$$

where in ④ we use the union bound for $\sum_{i=1}^{h} X_i X_{\pi(i)}$ and the tail bounds for Gaussian distributed X_i.

Then we define $Z_i = \sum_{j\in\mathcal{I}_i} X_j X_{\pi(j)}$ and bound $\zeta_{2,i}$ via the Bernstein inequality (Theorem 2.8.4 in Vershynin (2018)). First, we verify that $\mathbb{E}\left(X_j X_{\pi(j)}\right) = \left(\mathbb{E}X_j\right)\left(\mathbb{E}X_{\pi(j)}\right) = 0$. Meanwhile we compute $\sigma^2 = \sum_{j\in\mathcal{I}_i}\mathbb{E}\left(X_j X_{\pi(j)}\right)^2 = h_i$. According to the Bernstein inequality, we have

$$\left|\sum_{j\in\mathcal{I}_i} X_j X_{\pi(j)}\right| \ge \frac{4}{3}(\log n)^2 + \sqrt{\frac{16}{9}(\log n)^4 + 2(\log n)h_i},$$

holds with probability $2n^{-1}$. Meanwhile, we can upper bound as

$$\frac{4}{3}(\log n)^2 + \sqrt{\frac{16}{9}(\log n)^4 + 2(\log n)h_i} \le \frac{4}{3}(\log n)^2 + \sqrt{\frac{16}{9}(\log n)^4 + \frac{n\log n}{6}} \overset{⑤}{\le} \frac{4\sqrt{2}+\sqrt{35}}{3\sqrt{2}}\sqrt{n\log n},$$

where ⑤ is because $n \ge \log^3(n)$ for $n \ge 95$. Hence, we conclude that $\zeta_{2,i} \le 2n^{-1}$ and complete the proof by combining the bound for ζ_1 and ζ_2.

Lemma 2 *We have $\mathbb{P}\left(\overline{\mathcal{E}}_3(\mathbf{\Pi})\bigcap\mathcal{E}_2\right) \le n^{-2n}$.*

Proof 3 *For the conciseness of notation, we define $\mathbf{\Xi}$ as $\mathbf{\Xi} \triangleq \mathbf{X}\mathbf{X}^\top\left(\mathbf{\Pi}^\natural - \mathbf{\Pi}\right)$. Due to the independence of the $\mathbf{X}$ and $\mathbf{W}$, we can condition on $\mathbf{X}$ and bound $\mathbb{P}(\overline{\mathcal{E}}_3(\mathbf{\Pi})\bigcap\mathcal{E}_2)$ as*

$$\mathbb{P}\left(\overline{\mathcal{E}}_3(\mathbf{\Pi})\bigcap\mathcal{E}_2\right) \overset{①}{\le} \mathbb{P}\left(\mathbf{W}^\top\mathbf{\Xi}\mathbf{W} \ge \mathbb{E}\mathbf{W}^\top\mathbf{\Xi}\mathbf{W} + c\sigma^2 n^2\log n\right)$$
$$\overset{②}{\le} \exp\left(-\left(\frac{c_0 n^4\log^2 n}{\|\mathbf{\Xi}\|_{\mathrm{F}}^2} \wedge \frac{c_1 n^2\log n}{\|\mathbf{\Xi}\|_2}\right)\right) \overset{③}{\le} n^{-2n},$$

where in ① we condition on $\mathcal{E}_2$ and use the fact

$$\mathbb{E}\mathbf{W}^\top\mathbf{\Xi}\mathbf{W} + c\sigma^2 n^2\log n \lesssim \sigma^2\|\mathbf{X}\|_2^2 + c\sigma^2 n^2\log n \lesssim \sigma^2 n^2\log n,$$

in ② we use Hanson-Wright inequality (Theorem 6.2.1 in Vershynin (2018)), and in ③ we condition on $\mathcal{E}_2$ and use $\|\mathbf{\Xi}\|_2 \lesssim \|\mathbf{X}\|_2^2 \lesssim n$.

Lemma 3 *We have $\mathbb{P}\left(\overline{\mathcal{E}}_4(\mathbf{\Pi})\bigcap\mathcal{E}_2\right) \le n^{-2n}$.*

Proof 4 *Due to the independence between $\mathbf{W}$ and $\mathbf{X}$, we would like to condition on $\mathbf{X}$ and bound $\mathbb{P}(\overline{\mathcal{E}}_4(\mathbf{\Pi})\bigcap\mathcal{E}_2)$ as*

$$\mathbb{P}\left(\overline{\mathcal{E}}_4(\mathbf{\Pi})\bigcap\mathcal{E}_2\right) \le \exp\left(-\frac{4c\sigma^2 n^4\log n}{2\sigma_{\mathbf{\Pi}}^2}\right),$$

where $\sigma_{\mathbf{\Pi}}^2$ is defined as

$$\sigma_{\mathbf{\Pi}}^2 = \sigma^2 \left\| \left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \left(\mathbf{\Pi}^{\natural} - \mathbf{\Pi}\right)^{\top} \mathbf{X} \right\rangle \mathbf{X} + \left\langle \mathbf{\Pi}^{\natural}\mathbf{X}, \mathbf{X} \right\rangle \left(\mathbf{\Pi}^{\natural} - \mathbf{\Pi}\right) \mathbf{X} \right\|_{\mathrm{F}}^2,$$

Notice under $\mathcal{E}_2$, we have $\sigma_{\mathbf{\Pi}}^2 \lesssim \sigma^2 \left(4\|\mathbf{X}\|_2^3\right)^2 = c\sigma^2 n^3$, and complete the proof by showing

$$\exp\left(-\frac{4c\sigma^2 n^4 \log n}{2\sigma_{\mathbf{\Pi}}^2}\right) \leq \exp\left(-\frac{4c\sigma^2 n^4 \log n}{2c\sigma^2 n^3}\right) = n^{-2n}.$$

Lemma 4 *We have $\mathbb{P}\left(\overline{\mathcal{E}}_5(\mathbf{\Pi}); \ell\right) \leq 3n^{-2\ell}$.*

Proof 5 *Adopting a similar approach as in proving Lemma 1, we can decompose the index sets $\{j : j \neq \pi(j)\}$ into 3 disjoint sets $\mathcal{I}_i$ $(1 \leq i \leq 3)$ such that: (1) j and $\pi(j)$ do not lie within the same index set $\mathcal{I}_i$; and (2) the cardinality ℓ_i of $\mathcal{I}_i$ satisfies $\lfloor \ell/5 \rfloor \leq \ell_i \leq \ell/3$. Then we can bound $\mathbb{P}\left(\mathcal{E}_5(\mathbf{\Pi}; \ell)\right)$ as*

$$\mathbb{P}\left(\left\|\mathbf{X} - \mathbf{\Pi}^{\natural}\mathbf{X}\right\|_2^2 \leq \frac{12\ell}{5en^{20}}\right) \overset{①}{=} \sum_{i=1}^{3} \mathbb{P}\left(\sum_{j\in\mathcal{I}_i} \left(X_j - X_{\pi(j)}\right)^2 \leq \frac{4\ell}{5en^{20}}\right)$$
$$\overset{②}{\leq} \sum_{i=1}^{3} \exp\left(\frac{\ell_i}{2}\left(\log\frac{2l}{5en^{20}\ell_i} - \frac{2l}{5en^{20}\ell_i} + 1\right)\right) \overset{③}{\leq} 3n^{-2\ell}.$$

where ① is due to the decomposition $\mathcal{I}_i$, $1 \leq i \leq 3$, ② is because $\sum \left(X_j - X_{\pi(j)}\right)^2/2$ is a χ^2 RV with freedom ℓ_i and Lemma 11, and ③ is due to $\lfloor \ell/5 \rfloor \leq \ell_i \leq \ell/3$ and hence

$$\frac{\ell_i}{2}\left(\log\frac{2l}{5en^{20}\ell_i} - \frac{2l}{5en^{20}\ell_i} + 1\right) \leq \frac{\ell_i}{2}\left(\log\frac{2l}{5\ell_i} - 20\log n\right) \leq -10\ell_i \log n \leq -2\ell \log n.$$

D. Appendix for Section 4

This section provides theoretical analysis for the multiple observations model, i.e., $m > 1$. We will show that our estimator in Eq. (3) gives correct permutation matrix $\mathbf{\Pi}^{\natural}$ once

$$\log(\mathsf{SNR}) \gtrsim \frac{\log n}{\rho(\mathbf{B}^{\natural})} + \log\log n.$$

The formal statement is packaged in Theorem 2.

D.1. Notations

Before our discussion, first we define $\widetilde{\mathbf{B}}$ and $\mathbf{B}^*$ respectively as

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

where h is denoted as the Hamming distance between identity matrix $\mathbf{I}$ and the ground truth permutation matrix $\mathbf{\Pi}^{\natural}$, i.e., $h = \mathsf{d}_{\mathsf{H}}(\mathbf{I}, \mathbf{\Pi}^{\natural})$. Similar as in Section C, we define events $\mathcal{E}_i$, $(6 \leq i \leq 9)$ as

$$\mathcal{E}_6 \triangleq \left\{\|\mathbf{X}_{i,:}\|_2 \leq 2\sqrt{p\log n},\ \forall i\right\};$$
$$\mathcal{E}_7 \triangleq \left\{\left\|\mathbf{X}_{i,:}\left(\mathbf{B}^* - \mathbf{B}^{\natural}\right)\right\|_2 \lesssim c_0\frac{p(\log n)^{3/2}(\log p)}{\sqrt{n}}\left\|\mathbf{B}^{\natural}\right\|_{\mathrm{F}} + c_1\sqrt{m}(\log n)\sigma\left(1 + \frac{p}{n}\right),\ \forall i\right\};$$
$$\mathcal{E}_8 \triangleq \left\{\left\langle \mathbf{W}_{i,:}, \left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^{\natural}(i),:}\right)\mathbf{B}^*\right\rangle \geq \Delta,\ \exists\, i, j\right\};$$
$$\mathcal{E}_9 \triangleq \left\{\left\|\left(\mathbf{X}_{\pi^{\natural}(i),:} - \mathbf{X}_{j,:}\right)\mathbf{B}^{\natural}\right\|_2^2 + 2\left\langle\left(\mathbf{X}_{\pi^{\natural}(i),:} - \mathbf{X}_{j,:}\right)\mathbf{B}^{\natural}, \mathbf{X}_{j,:}\left(\mathbf{B}^{\natural} - \mathbf{B}^*\right)\right\rangle - \left\|\mathbf{X}_{\pi^{\natural}(i),:}\left(\mathbf{B}^{\natural} - \mathbf{B}^*\right)\right\|_2^2 \leq \Delta,\ \exists\, i, j\right\},$$

where Δ is defined as

$$\Delta = 16\sqrt{2}c_0\sigma\frac{p(\log n)^{3/2}(\log p)}{\sqrt{n}}\left\|\mathbf{B}^{\natural}\right\|_{\mathrm{F}} + 16c_1\sqrt{2m}(\log n)\sigma^2\left(1 + \frac{p}{n}\right) + 4\sqrt{2}c_2(\log n)\sigma\left\|\mathbf{B}^{\natural}\right\|_{\mathrm{F}}.$$

D.2. Outline of proof

In front of the rigorous proof in Section D.3, we first illustrate our proof strategy as

- **Step I:** We relax the wrong recovery $\left\{\widehat{\mathbf{\Pi}} \neq \mathbf{\Pi}^{\natural}\right\}$ to event $\mathcal{E}$, i.e. $\left\{\widehat{\mathbf{\Pi}} \neq \mathbf{\Pi}^{\natural}\right\} \subseteq \mathcal{E}$, which reads as

$$\mathcal{E} \triangleq \left\{\left\|\mathbf{Y}_{i,:} - \mathbf{X}_{\pi^{\natural}(i),:}\mathbf{B}^*\right\|_2^2 \geq \|\mathbf{Y}_{i,:} - \mathbf{X}_{j,:}\mathbf{B}^*\|_2^2, \ \exists\, i, j\right\}. \tag{14}$$

  The physical meaning of $\mathcal{E}$ is that we may reduce the residual $\|\mathbf{Y} - \mathbf{\Pi}^{\natural}\mathbf{X}\mathbf{B}^*\|_{\mathrm{F}}$ by changing $\pi^{\natural}(i)$ to j. Same relaxation has been previously used in Collier and Dalalyan (2016); Slawski et al. (2019a); Zhang et al. (2019a;b).

- **Step II:** The core in this step lies in how to lower bound $\mathbb{P}(\mathcal{E}_7)$. First we decompose $\mathcal{E}$ into $\mathcal{E}_8 \bigcup \mathcal{E}_9$ with some simple algebraic manipulations. Under the SNR assumption in Eq. (7), we show both $\mathbb{P}(\mathcal{E}_8)$ and $\mathbb{P}(\mathcal{E}_9)$ are approximately $\mathbb{P}(\overline{\mathcal{E}}_7)$, as in Lemma 5 and Lemma 6, respectively.

  To show $\mathbb{P}(\overline{\mathcal{E}}_7)$ is with low probability, in another words, $\mathbb{P}(\mathcal{E}_7)$ is highly likely, we prove the following relations hold with high probability under $\mathcal{E}_6$,

$$\left\|\mathbf{X}_{i,:}\left(\widetilde{\mathbf{B}} - \mathbf{B}^{\natural}\right)\right\|_2 \lesssim \frac{p(\log n)^{3/2}(\log p)}{\sqrt{n}}\|\mathbf{B}^{\natural}\|_{\mathrm{F}};$$
$$\left\|\mathbf{X}_{i,:}\mathbf{X}^{\top}\mathbf{W}\right\|_2 \lesssim \sqrt{m}(\log n)\sigma(n+p),$$

  whose proof are in Lemma 9 and Lemma 10, respectively, and hence finish the proof by

$$\left\|\mathbf{X}_{i,:}\left(\mathbf{B}^* - \mathbf{B}^{\natural}\right)\right\|_2 \leq \left\|\mathbf{X}_{i,:}\left(\widetilde{\mathbf{B}} - \mathbf{B}^{\natural}\right)\right\|_2 + \frac{1}{n-h}\left\|\mathbf{X}_{i,:}\mathbf{X}^{\top}\mathbf{W}\right\|_2.$$

  In particular, we would like to mention the technique used in bounding $\left\|\mathbf{X}_{i,:}\mathbf{X}^{\top}\mathbf{W}\right\|_2$. First we review the widely-used bounding procedure, which proceeds as

$$\left\|\mathbf{X}_{i,:}\mathbf{X}^{\top}\mathbf{W}\right\|_2 \leq \|\mathbf{X}_{i,:}\|_2\|\mathbf{X}\|_2\|\mathbf{W}\|_2 \overset{①}{\lesssim} \sqrt{p\log n}\left(\sqrt{n}+\sqrt{p}\right)\sigma\left(\sqrt{n}+\sqrt{m}\right) \overset{②}{\asymp} \sqrt{\log n}(n^{3/2})\sigma + \sqrt{mn\log n}\sigma,$$

  where in ① we use the fact $\|\mathbf{X}_{i,:}\|_2 \lesssim \sqrt{p\log n}$, $\|\mathbf{X}\|_2 \lesssim \sqrt{n}+\sqrt{p}$, $\|\mathbf{W}\|_2 \lesssim \sigma(\sqrt{n}+\sqrt{m})$ hold with high probability, and in ② we use $p \asymp n$. Comparing with our results in Lemma 10, this bound experience inflations when $m \ll n$ and will lift the SNR requirement to $\log(\mathsf{SNR}) \gtrsim \log n$, which hides the role of $\rho(\mathbf{B}^{\natural})$ compared with our current result in Theorem 2. To handle such problem, we adopt the leave-one-out trick as in El Karoui (2013; 2018); Chen et al. (2019); Sur et al. (2019) and refer to Lemma 10 for the technical details.

Having illustrated our proof strategies, we leave the detailed calculation to Section D.3.

D.3. Proof of Theorem 2

Proof 6 *We restate the definition of event $\mathcal{E}$ as*

$$\mathcal{E} \triangleq \left\{\left\|\mathbf{Y}_{i,:} - \mathbf{X}_{\pi^{\natural}(i),:}\mathbf{B}^*\right\|_2^2 \geq \|\mathbf{Y}_{i,:} - \mathbf{X}_{j,:}\mathbf{B}^*\|_2^2, \ \exists\, i, j\right\}.$$

Step I: *First we verify that*

$$\widehat{\mathbf{\Pi}} = \mathrm{argmin}_{\mathbf{\Pi}}\ \|\mathbf{Y} - \mathbf{\Pi}\mathbf{X}\mathbf{B}^*\|_{\mathrm{F}}$$

returns the same permutation matrix $\widehat{\mathbf{\Pi}}$ as that by Eq. (3). Hence, correct recovery of the ground truth permutation matrix $\mathbf{\Pi}^{\natural}$ suggests that

$$\left\|\mathbf{Y} - \mathbf{\Pi}^{\natural}\mathbf{X}\mathbf{B}^*\right\|_{\mathrm{F}} < \|\mathbf{Y} - \mathbf{\Pi}\mathbf{X}\mathbf{B}^*\|_{\mathrm{F}}, \ \forall\, \mathbf{\Pi} \neq \mathbf{\Pi}^{\natural}.$$

Then we finish the proof by showing that $\overline{\mathcal{E}} \subseteq \left\{\widehat{\mathbf{\Pi}} = \mathbf{\Pi}^\natural\right\}$. Assuming the claim is not true, which means we have matrix $\mathbf{\Pi}$ such that

$$\left\|\mathbf{Y} - \mathbf{\Pi}^\natural \mathbf{X}\mathbf{B}^*\right\|_{\mathrm{F}}^2 \geq \left\|\mathbf{Y} - \mathbf{\Pi}\mathbf{X}\mathbf{B}^*\right\|_{\mathrm{F}}^2,$$

conditional on event $\overline{\mathcal{E}}$. Meanwhile we have

$$\left\|\mathbf{Y} - \mathbf{\Pi}^\natural \mathbf{X}\mathbf{B}^*\right\|_{\mathrm{F}}^2 = \sum_{i=1}^n \left\|\mathbf{Y}_{i,:} - \mathbf{X}_{\pi^\natural(i),:}\mathbf{B}^*\right\|_2^2 \overset{①}{<} \sum_{i=1}^n \left\|\mathbf{Y}_{i,:} - \mathbf{X}_{\pi(i),:}\mathbf{B}^*\right\|_2^2 = \left\|\mathbf{Y} - \mathbf{\Pi}\mathbf{X}\mathbf{B}^*\right\|_{\mathrm{F}}^2,$$

which leads to contradiction, where in ① we use the definition of $\overline{\mathcal{E}}$.

Step II: *We verify that $\left\|\mathbf{Y}_{i,:} - \mathbf{X}_{\pi^\natural(i),:}\mathbf{B}^*\right\|_2^2 \geq \left\|\mathbf{Y}_{i,:} - \mathbf{X}_{j,:}\mathbf{B}^*\right\|_2^2$ is equivalent to*

$$\begin{aligned} 2\left\langle \mathbf{W}_{i,:},\ \left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^\natural(i),:}\right)\mathbf{B}^*\right\rangle \geq{}& \left\|\left(\mathbf{X}_{\pi^\natural(i),:} - \mathbf{X}_{j,:}\right)\mathbf{B}^\natural\right\|_2^2 + \left\|\mathbf{X}_{j,:}\left(\mathbf{B}^\natural - \mathbf{B}^*\right)\right\|_2^2 \\ &+ 2\left\langle \left(\mathbf{X}_{\pi^\natural(i),:} - \mathbf{X}_{j,:}\right)\mathbf{B}^\natural, \mathbf{X}_{j,:}\left(\mathbf{B}^\natural - \mathbf{B}^*\right)\right\rangle - \left\|\mathbf{X}_{\pi^\natural(i),:}\left(\mathbf{B}^\natural - \mathbf{B}^*\right)\right\|_2^2, \end{aligned}$$

which suggests that $\mathbb{P}(\mathcal{E}) \leq \mathbb{P}(\mathcal{E}_8) + \mathbb{P}(\mathcal{E}_9)$ and completes the proof with Lemma 5 and Lemma 6.

Lemma 5 *We have $\mathbb{P}(\mathcal{E}_8) \leq c_0 e^{-\left((\log n)^4 \wedge (\log n)^2 \rho(\mathbf{B}^\natural)\right)} + c_1 n^{-1} + c_2 n e^{-c_3 n} + c_4 n e^{-c_0 m} + 2e^{-p} + 6p^{-2}$.*

Proof 7 *For the conciseness of notation, we define Δ_1 and Δ_2 as*

$$\begin{aligned} \Delta_1 ={}& 4c_0 \frac{p(\log n)^{3/2}(\log p)}{\sqrt{n}}\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}} + 4c_1\sqrt{m}(\log n)\sigma\left(1 + \frac{p}{n}\right); \\ \Delta_2 ={}& c_2(\log n)\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}. \end{aligned}$$

Then we can bound $\mathbb{P}(\mathcal{E}_8)$ as

$$\begin{aligned} \mathbb{P}(\mathcal{E}_8) &\overset{①}{\leq} \mathbb{P}\left(\left\|\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^\natural(i),:}\right)\mathbf{B}^*\right\|_2 \geq \Delta_1 + \Delta_2,\ \exists\, i,j\right) + \exp\left(-\frac{\Delta^2}{2\sigma^2\left(\Delta_1 + \Delta_2\right)^2}\right) \\ &\overset{②}{\leq} \underbrace{\mathbb{P}\left(\left\|\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^\natural(i),:}\right)\left(\mathbf{B}^* - \mathbf{B}^\natural\right)\right\|_2 \geq \Delta_1,\ \exists\, i,j\right)}_{\zeta_1} + \underbrace{\mathbb{P}\left(\left\|\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^\natural(i),:}\right)\mathbf{B}^\natural\right\|_2 \geq \Delta_2,\ \exists\, i,j\right)}_{\zeta_2} + n^{-8}, \end{aligned} \tag{15}$$

where in ① we use the independence between $\mathbf{W}$ and $\mathbf{X}$ and condition on $\mathbf{X}$, in ② we use the relation $\Delta = 4\sqrt{2}\sigma\left(\Delta_1 + \Delta_2\right)$. Then we will prove that $\zeta_1 \leq \mathbb{P}(\overline{\mathcal{E}}_7)$ and $\zeta_2 \asymp e^{-\left((\log n)^4 \wedge (\log n)^2 \rho(\mathbf{B}^\natural)\right)}$.

Phase I: bounding ζ_1 *Conditional on $\mathcal{E}_7$, we have*

$$\begin{aligned} &\left\|\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^\natural(i),:}\right)\left(\mathbf{B}^* - \mathbf{B}^\natural\right)\right\|_2 \leq \left\|\mathbf{X}_{j,:}\left(\mathbf{B}^* - \mathbf{B}^\natural\right)\right\|_2 + \left\|\mathbf{X}_{\pi^\natural(i),:}\left(\mathbf{B}^* - \mathbf{B}^\natural\right)\right\|_2 \\ &\overset{③}{\leq} 2c_0 \frac{p(\log n)^{3/2}(\log p)}{\sqrt{n}}\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}} + 2c_1\sqrt{m}(\log n)\sigma\left(1 + \frac{p}{n}\right) < \frac{\Delta_1}{2}, \end{aligned}$$

and obtain $\zeta_1 = 0$, where ③ is due to the definition of $\mathcal{E}_7$. Then we conclude that $\zeta_1 \leq \mathbb{P}(\overline{\mathcal{E}}_7)$.

Phase II: bounding ζ_2 *For ζ_2, we upper-bound it as*

$$\begin{aligned} \zeta_2 &\overset{④}{\leq} \sum_{\pi^\natural(i),j} \mathbb{P}\left(Z \geq c_2(\log n)^2\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}^2\right) \overset{⑤}{\leq} n^2\,\mathbb{P}\left(\left|Z - \mathbb{E}Z\right| \geq c_3(\log n)^2\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}^2\right) \\ &\overset{⑥}{\leq} n^2 \exp\left(-\left(\frac{(\log n)^4\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}^4}{\left\|\mathbf{B}^\natural\mathbf{B}^{\natural\top}\right\|_{\mathrm{F}}^2} \wedge \frac{(\log n)^2\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}^2}{\left\|\mathbf{B}^\natural\mathbf{B}^{\natural\top}\right\|_{\mathrm{OP}}}\right)\right) = n^2 e^{-\left((\log n)^4 \wedge (\log n)^2 \rho(\mathbf{B}^\natural)\right)} \\ &\asymp e^{-\left((\log n)^4 \wedge (\log n)^2 \rho(\mathbf{B}^\natural)\right)}, \end{aligned} \tag{16}$$

where in ④ we define $Z \triangleq \left\|\left(\mathbf{X}_{j,:} - \mathbf{X}_{\pi^\natural(i),:}\right)\mathbf{B}^\natural\right\|_2^2$, *in ⑤ we have* $\mathbb{E}Z = 4\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}^2$ *and use* $c_2(\log n)^2\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}^2 \geq \left(4 + c_3(\log n)^2\right)\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}^2$ *when* n *is sufficiently large, and in ⑥ we use the Hanson-Wright inequality (Theorem* 6.2.1 *in Vershynin (2018)). Combining Eq. (15), Eq. (16) and Lemma 8 together, we complete the proof.*

Lemma 6 *Consider the same setting of Theorem* 2*. Provided the* SNR *satisfies*

$$\log(\mathsf{SNR}) \gtrsim \frac{6\log n}{\rho\left(\mathbf{B}^\natural\right)} + \log\log n,$$

we have $\mathbb{P}\left(\mathcal{E}_9\right) \leq 2e^{-p} + ne^{-c_1 m} + c_2 p^{-2} + c_3 n e^{-c_4 n}$, *when* n *is sufficiently large, where* $c_i > 0$, $0 \leq i \leq 4$ *are some positive constants.*

Proof 8 *We upper bound* $\mathbb{P}\left(\mathcal{E}_9\right)$ *as*

$$\mathbb{P}\left(\mathcal{E}_9\right) \leq \mathbb{P}\left(\left\|\left(\mathbf{X}_{\pi^\natural(i),:} - \mathbf{X}_{j,:}\right)\mathbf{B}^\natural\right\|_2^2 - 2\left\|\left(\mathbf{X}_{\pi^\natural(i),:} - \mathbf{X}_{j,:}\right)\mathbf{B}^\natural\right\|_2\left\|\mathbf{X}_{j,:}\left(\mathbf{B}^\natural - \mathbf{B}^*\right)\right\|_2 - \left\|\mathbf{X}_{\pi^\natural(i),:}\left(\mathbf{B}^\natural - \mathbf{B}^*\right)\right\|_2^2 \leq \Delta,\ \exists\, i,j\right)$$
$$\leq \underbrace{\mathbb{P}\left(\left\|\left(\mathbf{X}_{\pi^\natural(i),:} - \mathbf{X}_{j,:}\right)\mathbf{B}^\natural\right\|_2 \leq \delta,\ \exists\, i,j\right)}_{\triangleq\,\zeta_1} + \underbrace{\mathbb{P}\left(\frac{\left\|\mathbf{X}_{\pi^\natural(i),:}\left(\mathbf{B}^\natural - \mathbf{B}^*\right)\right\|_2^2}{\delta^2} + \frac{2\left\|\mathbf{X}_{\pi^\natural(i),:}\left(\mathbf{B}^\natural - \mathbf{B}^*\right)\right\|_2}{\delta} + \frac{\Delta}{\delta^2} \geq 1,\ \ \exists\, i,j\right)}_{\triangleq\,\zeta_2}.$$

Setting δ *as* $\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}} n^{-\frac{3}{c\rho(\mathbf{B}^\natural)}}$, *we would like to show* $\zeta_1 \lesssim n^{-1}$ *and* $\zeta_2 \leq \mathbb{P}(\overline{\mathcal{E}}_7)$ *under the assumptions in Lemma 6.*

Phase I: bounding ζ_1 *We set* δ *as* $\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}} n^{-\frac{3}{c\rho(\mathbf{B}^\natural)}}$, *and can upper bound* ζ_1 *as*

$$\zeta_1 \leq \sum_{i=1}^{n}\sum_{j\neq\pi^\natural(i)} \mathbb{P}\left(\left\|\left(\mathbf{X}_{\pi^\natural(i),:} - \mathbf{X}_{j,:}\right)\mathbf{B}^\natural\right\|_2 \leq \delta\right) \overset{①}{\leq} \sum_{i=1}^{n}\sum_{j\neq\pi^\natural(i)} n^{-3} \lesssim n^{-1}, \tag{17}$$

where ① comes from the small ball *probability as in Lemma 2.6 in Latala et al. (2007), which is also stated as Lemma 12.*

Phase II: bounding ζ_2 *Then we prove that* ζ_2 *can be arbitrarily small under the* SNR *requirement in Eq. (7). Conditional on event* $\mathcal{E}_7$, *we have*

$$\frac{\left\|\mathbf{X}_{\pi^\natural(i),:}\left(\mathbf{B}^\natural - \mathbf{B}^*\right)\right\|_2^2}{\delta^2} \leq \frac{2c_0^2\frac{p^2(\log n)^3(\log p)^2}{n}\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}^2 + 2c_1^2 m(\log n)^2\sigma^2\left(1 + p/n\right)^2}{\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}^2 n^{-\frac{6}{c\rho(\mathbf{B}^\natural)}}}$$
$$\overset{②}{\leq} \underbrace{\frac{2c_0^2 p^2(\log n)^3(\log p)^2}{n^{1-6/(c\rho(\mathbf{B}^\natural))}}}_{\eta_1} + \underbrace{8c_1^2\frac{(\log n)^2 n^{\frac{6}{c\rho(\mathbf{B}^\natural)}}}{\mathsf{SNR}}}_{\eta_2}, \tag{18}$$

in ② we use the fact $p \leq n$*. Since we have* $n \geq p^4(\log n)^6(\log p)^4$ *and* $\rho(\mathbf{B}^\natural) \geq 18/c$, *we conclude* $\eta_1 \to 0$ *as* n *goes to infinity. Meanwhile, because of the assumptions in Eq. (7), we have* η_2 *to be a small positive constants.*

Additionally, we can expand Δ/δ^2 *as*

$$\frac{\Delta}{\delta^2} \lesssim \frac{n^{\frac{6}{c\rho(\mathbf{B}^\natural)}}\sigma}{\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}^2}\left(c_0\frac{p(\log n)^{3/2}(\log p)}{\sqrt{n}}\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}} + c_1\sqrt{m}(\log n)\sigma\left(1 + \frac{p}{n}\right) + c_2(\log n)\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}\right)$$
$$\lesssim c_0\frac{p(\log n)^{3/2}(\log p)}{\sqrt{mn}} \times \frac{n^{\frac{6}{c\rho(\mathbf{B}^\natural)}}}{\sqrt{\mathsf{SNR}}} + c_1\frac{\log n}{\sqrt{m}} \times \frac{n^{\frac{6}{c\rho(\mathbf{B}^\natural)}}}{\sqrt{\mathsf{SNR}}} + c_2\frac{\log n}{\sqrt{m}} \times \frac{n^{\frac{6}{c\rho(\mathbf{B}^\natural)}}}{\mathsf{SNR}}. \tag{19}$$

Following similar procedures as above, we can prove Δ/δ^2 *to be a small positive constant given Eq. (7). Combing Eq. (18) and Eq. (19) together, we conclude*

$$\eta_1 + \eta_2 + 2\sqrt{\eta_1 + \eta_2} + \frac{\Delta}{\delta^2} < 1,$$

which suggests that ζ_2 *equals zero conditional on events* $\mathcal{E}_7$*. Therefore, we obtain*

$$\zeta_2 \leq \mathbb{P}\left(\overline{\mathcal{E}}_7\right) \overset{③}{\leq} 2e^{-p} + 6p^{-2} + ne^{-c_0 m} + c_0 n^{-1} + c_1 n e^{-c_2 n} \overset{④}{\lesssim} 2e^{-p} + ne^{-c_0 m} + c_0 p^{-2} + c_1 n e^{-c_2 n}$$

and completes the proof together with Eq. (17), where ③ is due to Lemma 8, and ④ is because of $n \gtrsim p^2$.

D.4. Supporting Lemmas for Theorem 2

Lemma 7 *For arbitrary row $\mathbf{X}_{i,:}$, we have*

$$\|\mathbf{X}_{i,:}\|_2 \leq 2\sqrt{p\log n},$$

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

Proof 9 *Notice that $\|\mathbf{X}_{i,:}\|_2^2$ is a χ^2-RV with freedom p, we have*

$$\mathbb{P}\left(\|\mathbf{X}_{i,:}\|_2^2 \geq 4p\log n\right) \leq \exp\left(\frac{p}{2}\left(\log(4p\log n) - 4\log n + 1\right)\right) \overset{①}{\leq} \exp\left(-p\log n\right) = n^{-p},$$

where in ① we use $2\log n \geq \log\left(4\log n\right) + 1$, when $n \geq 4$.

Lemma 8 *We have $\mathbb{P}\left(\mathcal{E}_7\right) \geq 1 - 2e^{-p} - 6p^{-2} - ne^{-c_0 m} - c_0 n^{-1} - c_1 n e^{-c_2 n}$.*

Proof 10 *Invoking Lemma 10, we have*

$$\begin{aligned}
&\mathbb{P}\left(\left\|\mathbf{X}_{i,:}\mathbf{X}^\top\mathbf{W}\right\|_2 \leq\ c_0\sqrt{m}(\log n)\sigma\left(n+p\right), \forall\, i\right) \\
&= 1 - \mathbb{P}\left(\left\|\mathbf{X}_{i,:}\mathbf{X}^\top\mathbf{W}\right\|_2 >\ c_0\sqrt{m}(\log n)\sigma\left(n+p\right), \exists\, i\right) \\
&\geq 1 - \sum_i \mathbb{P}\left(\left\|\mathbf{X}_{i,:}\mathbf{X}^\top\mathbf{W}\right\|_2 >\ c_0\sqrt{m}(\log n)\sigma\left(n+p\right)\right) \\
&\geq 1 - n^{1-p} - ne^{-c_0 m} - n^{-1} - c_1 n e^{-c_2 n}.
\end{aligned} \tag{20}$$

Then we conclude

$$\begin{aligned}
&\left\|\mathbf{X}_{i,:}\left(\mathbf{B}^* - \mathbf{B}^\natural\right)\right\|_2 \leq \left\|\mathbf{X}_{i,:}\left(\widetilde{\mathbf{B}} - \mathbf{B}^\natural\right)\right\|_2 + \frac{1}{n-h}\left\|\mathbf{X}_{i,:}\mathbf{X}^\top\mathbf{W}\right\|_2 \\
&\leq \|\mathbf{X}_{i,:}\|_2 \left\|\widetilde{\mathbf{B}} - \mathbf{B}^\natural\right\|_{\mathrm{F}} + \frac{1}{n-h}\left\|\mathbf{X}_{i,:}\mathbf{X}^\top\mathbf{W}\right\|_2 \\
&\overset{①}{\leq} c_0\frac{p(\log n)^{3/2}(\log p)}{\sqrt{n}}\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}} + \frac{c_1\sqrt{m}(\log n)\sigma\left(n+p\right)}{n-h} \\
&\overset{②}{\leq} c_0\frac{p(\log n)^{3/2}(\log p)}{\sqrt{n}}\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}} + \frac{4}{3}c_1\sqrt{m}(\log n)\sigma\left(1+\frac{p}{n}\right),
\end{aligned}$$

where in ① we condition on Lemma 9 and Eq. (20), and in ② we use the fact $h \leq n/4$.

Lemma 9 *Provided that $n \gtrsim p^2$, $h \leq n/4$, we have*

$$\left\|\widetilde{\mathbf{B}} - \mathbf{B}^\natural\right\|_{\mathrm{F}} \leq \sqrt{\frac{p}{n}}\left\|\mathbf{B}^\natural\right\|_{\mathrm{F}}\left(4\sqrt{6} + (\log n)(\log p)\right),$$

with probability at least $1 - 2e^{-p} - 6p^{-2}$ when n, p are sufficiently large.

Proof 11 *We assume that the first h rows of $\mathbf{X}$ are permuted w.l.o.g. First, we expand $\mathbf{X}^\top\mathbf{\Pi}^\natural\mathbf{X}$ as*

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

and obtain

$$\begin{aligned}
&\mathbb{P}\left(\left\|\mathbf{B}^\natural - \widetilde{\mathbf{B}}\right\|_2 \geq \sqrt{\frac{p}{n}}\|\mathbf{B}^\natural\|_{\mathrm{F}}\left(4\sqrt{6} + (\log n)(\log p)\right)\right) \\
\leq\ &\mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{i=1}^{h}\mathbf{X}_{\pi(i),:}^\top\mathbf{X}_{i,:}\mathbf{B}^\natural\right\|_{\mathrm{F}} + \frac{1}{n-h}\left\|\sum_{i=h+1}^{n}\left(\mathbf{X}_{i,:}^\top\mathbf{X}_{i,:} - \mathbf{I}\right)\mathbf{B}^\natural\right\|_{\mathrm{F}} \geq \sqrt{\frac{p}{n}}\|\mathbf{B}^\natural\|_{\mathrm{F}}\left(4\sqrt{6} + (\log n)(\log p)\right)\right) \\
\overset{①}{\leq}\ &\underbrace{\mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{i=1}^{h}\mathbf{X}_{\pi(i),:}^\top\mathbf{X}_{i,:}\mathbf{B}^\natural\right\|_{\mathrm{F}} \geq \frac{(\log n)(\log p)\sqrt{p}}{\sqrt{n}}\|\mathbf{B}^\natural\|_{\mathrm{F}}\right)}_{\zeta_1} \\
+\ &\underbrace{\mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{i=h+1}^{n}\left(\mathbf{X}_{i,:}^\top\mathbf{X}_{i,:} - \mathbf{I}\right)\mathbf{B}^\natural\right\|_{\mathrm{F}} \geq 4\sqrt{\frac{6p}{n}}\|\mathbf{B}^\natural\|_{\mathrm{F}}\right)}_{\zeta_2},
\end{aligned}$$

where ① is because of the union bound. Then we separately bound ζ_1 and ζ_2.

Phase I: Bounding ζ_1 *According to Lemma 8 in Pananjady et al. (2017a) (restated as Lemma 13), 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*

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

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

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

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

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

where ② is because $\mathbf{X}_{\pi(j),:}^\top\mathbf{X}_{j,:}$ is rank-1, ③ is due to the fact $\left\|\mathbf{u}\mathbf{v}^\top\right\|_{\mathrm{F}}^2 = \mathrm{Tr}\left(\mathbf{u}\mathbf{v}^\top\mathbf{v}\mathbf{u}^\top\right) = \|\mathbf{u}\|_2^2\|\mathbf{v}\|_2^2$ for arbitrary vector $\mathbf{u},\mathbf{v}\in\mathbb{R}^p$, and ④ is because of Lemma 7.

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

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

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 (cf. Thm 7.3.1 in Tropp (2015)) and have

$$\frac{1}{n-h}\left\|\sum_{j\in\mathcal{I}}\mathbf{X}_{\pi(j),:}^\top\mathbf{X}_{j,:}\right\|_2 \leq \frac{1}{3}\left(\frac{4p(\log n)(\log p)}{n-h} + \frac{p\sqrt{16(\log n)^2(\log p)^2 + 6h_i\log p/p}}{n-h}\right)$$

holds with probability $1 - 2p^{-2}$.

Exploiting the fact such that $h \leq n/4$, $h_i \leq h/3$, *and* $p \lesssim \sqrt{n}$, *we obtain*

$$\frac{p\sqrt{16(\log n)^2(\log p)^2 + 6h_i \log p/p}}{n-h} \leq \frac{4p}{3n}\sqrt{16(\log n)^2(\log p)^2 + \frac{n}{2p}(\log n)(\log p)} \overset{⑦}{\leq} \frac{4\sqrt{p}}{3\sqrt{n}} \times (\log n)(\log p),$$

in ⑦ we $n \gtrsim p^2 \geq 32p$ *and hence*

$$\frac{1}{n-h}\left\|\sum_{j\in\mathcal{I}} \mathbf{X}_{\pi(j),:}^\top \mathbf{X}_{j,:}\right\|_2 \leq (\log n)(\log p)\left(\frac{16p}{9n} + \frac{4\sqrt{p}}{9\sqrt{n}}\right) \overset{⑧}{\leq} \sqrt{\frac{p}{n}}(\log n)(\log p),$$

holds with probability exceeding $1 - 6p^{-2}$, *where in ⑧ we use* $n \geq 256p/25$.

Phase II: Bounding ζ_2 *We upper bound* ζ_2 *as*

$$\begin{aligned}\zeta_2 \leq\ & \mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{i=h+1}^{n}\left(\mathbf{X}_{i,:}^\top \mathbf{X}_{i,:} - \mathbf{I}\right)\mathbf{B}^\natural\right\|_{\mathsf{F}} \geq 4\sqrt{\frac{6p}{n}}\|\mathbf{B}^\natural\|_{\mathsf{F}}\right) \\ \leq\ & \mathbb{P}\left(\frac{1}{n-h}\left\|\sum_{i=h+1}^{n}\left(\mathbf{X}_{i,:}^\top \mathbf{X}_{i,:} - \mathbf{I}\right)\right\|_{\mathsf{OP}}\|\mathbf{B}^\natural\|_{\mathsf{F}} \geq 4\sqrt{\frac{6p}{n}}\|\mathbf{B}^\natural\|_{\mathsf{F}}\right) \overset{⑨}{\leq} 2e^{-p}.\end{aligned}$$

where ⑨ is because of $(n-h)^{-1}\left\|\sum_{i=h+1}^{n}\left(\mathbf{X}_{i,:}\mathbf{X}_{i,:}^\top - \mathbf{I}\right)\right\|_2 \leq 6\sqrt{2p/(n-h)}$ *with probability* $2e^{-p}$ *in Example 6.1 in Wainwright (2019) (also listed as Lemma 14) and* $h \leq n/4$.

The proof is completed via combing the results in Phase I and Phase II.

Lemma 10 *For an arbitrary index* i, *we have*

$$\mathbb{P}\left(\left\|\mathbf{X}_{i,:}\mathbf{X}^\top \mathbf{W}\right\|_2 \geq\ c_0\sqrt{m}(\log n)\sigma\,(n+p)\right) \leq n^{-p} + e^{-c_0 m} + n^{-2} + c_1 e^{-c_2 n}.$$

Proof 12 *For the conciseness of notation, we define* δ *as* $c_0\sqrt{m}(\log n)\sigma\,(n+p)$. *In addition, we assume that* $i = 1$ *w.l.o.g and prove this lemma with the* leave-one-out *trick, which is previously used in El Karoui (2013); El Karoui et al. (2013); El Karoui (2018); Chen et al. (2019); Sur et al. (2019). First we define a perturbed matrix* $\widetilde{\mathbf{X}}$ *such that* $\widetilde{\mathbf{X}}_{j,:} = \mathbf{X}_{j,:}$, $2 \leq j \leq n$, *while* $\widetilde{\mathbf{X}}_{1,:} \in \mathbb{R}^{1\times p}$ *is a independent identically distributed Gaussian vector as* $\mathbf{X}_{1,:}$, *namely,* $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

Then we can upper-bound the probability as

$$\begin{aligned}&\mathbb{P}\left(\left\|\mathbf{X}_{1,:}\mathbf{X}^\top\mathbf{W}\right\|_2 \geq \delta\right) \leq \mathbb{P}\left(\left\|\mathbf{X}_{1,:}\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_2 + \left\|\mathbf{X}_{1,:}\left(\mathbf{X}-\widetilde{\mathbf{X}}\right)^\top\mathbf{W}\right\|_2 \geq \delta\right) \\ \leq\ & \underbrace{\mathbb{P}\left(\left\|\mathbf{X}_{1,:}\left(\mathbf{X}-\widetilde{\mathbf{X}}\right)^\top\mathbf{W}\right\|_2 \geq 4p\,(\log n)\sqrt{m}\sigma\right)}_{\zeta_1} + \underbrace{\mathbb{P}\left(\left\|\mathbf{X}_{i,:}\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_2 \geq \delta - 4p\,(\log n)\sqrt{m}\sigma\right)}_{\zeta_2}.\end{aligned}$$

Phase I: bounding ζ_1 *To bound* ζ_1, *easily we can verify the following relation*

$$\left\|\mathbf{X}_{1,:}\left(\mathbf{X}-\widetilde{\mathbf{X}}\right)^\top\mathbf{W}\right\|_2 \leq \|\mathbf{X}_{1,:}\|_2\left\|\left(\mathbf{X}-\widetilde{\mathbf{X}}\right)^\top\mathbf{W}\right\|_{\mathsf{F}} \overset{①}{=} \|\mathbf{X}_{1,:}\|_2\left\|\mathbf{X}_{1,:}-\widetilde{\mathbf{X}}_{1,:}\right\|_2\|\mathbf{W}_{1,:}\|_2 \overset{②}{\leq} 4p\,(\log n)\sqrt{m}\sigma.$$

with probability exceeding $1 - n^{-p} - e^{-c_0 m}$, *where ① is because only the first row of* $\mathbf{X}-\widetilde{\mathbf{X}}$ *is nonzero, and ② conditions on* $\mathcal{E}_6$ *and* $\|\mathbf{W}_{1,:}\|_2 \leq 2\sqrt{m}\sigma$ *holds with probability at least* $1 - e^{-c_0 m}$.

Phase II: bounding ζ_2 *Since* $\delta - 4p(\log n)\sqrt{m}\sigma \gtrsim n(\log n)\sqrt{m}\sigma$, *we can upper-bound* ζ_2 *as*

$$\zeta_2 \leq \mathbb{P}\left(\left\|\mathbf{X}_{i,:}\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_2 \geq c_1 n(\log n)\sqrt{m}\sigma\right).$$

Due to the construction of $\widetilde{\mathbf{X}}$, we have $\mathbf{X}_{1,:}$ to be independent of $\widetilde{\mathbf{X}}$. Hence, we condition on $\widetilde{\mathbf{X}}^\top\mathbf{W}$ and obtain

$$\begin{aligned}\zeta_2 \leq\ & \mathbb{P}\left(\left\|\mathbf{X}_{i,:}\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_2 \geq c_1 n(\log n)\sqrt{m}\sigma,\ \left\|\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_{\mathrm{F}} < 8n\sqrt{m}\sigma\right) + \mathbb{P}\left(\left\|\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_{\mathrm{F}} \geq 8n\sqrt{m}\sigma\right) \\ \leq\ & \underbrace{\mathbb{E}_{\widetilde{\mathbf{X}}^\top\mathbf{W}}\mathbb{1}\left(\left\|\mathbf{X}_{i,:}\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_2 \geq c_2(\log n)\left\|\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_{\mathrm{F}}\right)}_{\zeta_{2,1}} + \underbrace{\mathbb{P}\left(\left\|\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_{\mathrm{F}} \geq 8n\sqrt{m}\sigma\right)}_{\zeta_{2,2}}.\end{aligned}$$

For $\zeta_{2,1}$, we define $Z = \left\|\mathbf{X}_{i,:}\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_2^2$ and have

$$\begin{aligned}\zeta_{2,1} &\leq \mathbb{E}_{\widetilde{\mathbf{X}}^\top\mathbf{W}}\mathbb{1}\left(|Z-\mathbb{E}Z| \geq c_3(\log n)^2\left\|\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_{\mathrm{F}}^2\right) \\ &\overset{③}{\leq} \mathbb{E}_{\widetilde{\mathbf{X}}^\top\mathbf{W}}\exp\left(-\left(\frac{(\log n)^4\left\|\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_{\mathrm{F}}^4}{\left\|\widetilde{\mathbf{X}}^\top\mathbf{W}\mathbf{W}^\top\widetilde{\mathbf{X}}\right\|_{\mathrm{F}}^2} \wedge \frac{(\log n)^2\left\|\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_{\mathrm{F}}^2}{\left\|\widetilde{\mathbf{X}}^\top\mathbf{W}\mathbf{W}^\top\widetilde{\mathbf{X}}\right\|_{\mathrm{OP}}}\right)\right) \overset{④}{\leq} n^{-2},\end{aligned}$$

where ③ is because of the Hanson-Wright inequality (Theorem 6.2.1 in Vershynin (2018)), and ④ is due to the stable rank $\rho(\widetilde{\mathbf{X}}^\top\mathbf{W}) \geq 1$. Meanwhile we upper-bound $\zeta_{2,2}$ as

$$\begin{aligned}&\mathbb{P}\left(\left\|\widetilde{\mathbf{X}}^\top\mathbf{W}\right\|_2 \geq 8n\sqrt{m}\sigma\right) \leq\ \mathbb{P}\left(\left\|\widetilde{\mathbf{X}}\right\|_{\mathrm{OP}}\|\mathbf{W}\|_{\mathrm{F}} \geq 8n\sqrt{m}\sigma\right) \\ \overset{⑤}{\leq}\ & \mathbb{P}\left(\left\|\widetilde{\mathbf{X}}\right\|_{\mathrm{OP}} \geq 2\left(\sqrt{n}+\sqrt{p}\right)\right) + \mathbb{P}\left(\|\mathbf{W}\|_{\mathrm{F}} \geq \frac{8n\sqrt{m}\sigma}{2\left(\sqrt{n}+\sqrt{p}\right)},\ \left\|\widetilde{\mathbf{X}}\right\|_{\mathrm{OP}} \leq 2\left(\sqrt{n}+\sqrt{p}\right)\right) \\ \overset{⑥}{\leq}\ & \mathbb{P}\left(\left\|\widetilde{\mathbf{X}}\right\|_{\mathrm{OP}} \geq 2\left(\sqrt{n}+\sqrt{p}\right)\right) + \mathbb{P}\left(\|\mathbf{W}\|_{\mathrm{F}} \geq \sqrt{2nm}\sigma\right) \overset{⑦}{\leq} e^{-c_0 n} + e^{-0.8nm},\end{aligned}$$

where ⑤ is because of the union bound, in ⑥ we use $p \leq n$, and in ⑦ we use $\|\mathbf{X}\|_{\mathrm{OP}} \geq 2\left(\sqrt{n}+\sqrt{p}\right)$ with probability less than $e^{-c_0 n}$ (Chandrasekaran et al., 2012) and the fact $\|\mathbf{W}\|_{\mathrm{F}}^2/\sigma^2$ is a χ^2-RV with nm freedom, and Lemma 11.

E. Useful Facts

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

Lemma 11 *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 12 (Small ball probability, Lemma 2.6 in Latala et al. (2007)) *Given an arbitrary fixed vector $\mathbf{y} \in \mathbb{R}^n$, we have*

$$\mathbb{P}\left(\|\mathbf{y}-\mathbf{A}\mathbf{g}\|_2 \leq \alpha\|\mathbf{A}\|_{\mathrm{F}}\right) \leq \exp\left(\kappa\log(\alpha)\varrho(\mathbf{A})\right),\quad \forall\ \alpha \in (0,\alpha_0),$$

where $\mathbf{g}$ is a Gaussian RV following $\mathcal{N}\left(\mathbf{0},\mathbf{I}_{n\times n}\right)$, $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a non-zero matrix, and $\alpha_0 \in (0,1)$ and $\kappa > 0$ are some universal constants.

Lemma 13 (Lemma 8 in Pananjady et al. (2017a)) *Consider an arbitrary permutation map $\boldsymbol{\pi}$ with Hamming distance k from the identity map, i.e., $\mathsf{d}_{\mathsf{H}}\left(\boldsymbol{\pi},\mathbf{I}\right) = k$. We define the index set $\{i:\ i \neq \pi(i)\}$ and can decompose it into 3 independent sets $\mathcal{I}_j$ $(1 \leq j \leq 3)$, i.e., i and $\pi(i)$ are in different sets $\mathcal{I}_j$ for arbitrary $i \in \{i:\ i \neq \pi(i)\}$, such that the cardinality of each set satisfies $|\mathcal{I}_j| \geq \lfloor k/3 \rfloor \geq k/5$.*

Lemma 14 (Example 6.1 in Wainwright (2019)) *Let $\mathbf{G} \in \mathbb{R}^{n_1\times n_2}$ be generated with iid standard normal random variables, we have $\|\mathbf{G}\|_{\mathrm{OP}} \leq 4\sqrt{n_2/n_1}$, hold with probability exceeding $1 - 2e^{-n_2/2}$.*