

THE PHASE TRANSITION PHENOMENON OF SHUFFLED REGRESSION

Anonymous authors

Paper under double-blind review

ABSTRACT

We study the phase transition phenomenon inherent in the shuffled (permuted) regression problem, which has found numerous applications in databases, privacy, data analysis, etc. For the permuted regression task: $\mathbf{Y} = \mathbf{\Pi}^{\natural} \mathbf{X} \mathbf{B}^{\natural}$, the goal is to recover the permutation matrix $\mathbf{\Pi}^{\natural}$ as well as the coefficient matrix \mathbf{B}^{\natural} . It has been empirically observed in prior studies that when recovering $\mathbf{\Pi}^{\natural}$, there exists a phase transition phenomenon: the error rate drops to zero rapidly once the parameters reach certain thresholds. In this study, we aim to precisely identify the locations of the phase transition points by leveraging techniques from *message passing* (MP).

In our analysis, we first transform the permutation recovery problem into a probabilistic graphical model. Then, we leverage the analytical tools rooted in the message passing (MP) algorithm and derive an equation to track the convergence of the MP algorithm. By linking this equation to the branching random walk process, we are able to characterize the impact of the *signal-to-noise-ratio* (snr) on the permutation recovery. Depending on whether the signal is given or not, we separately investigate the oracle case and the non-oracle case. The bottleneck in identifying the phase transition regimes lies in deriving closed-form formulas for the corresponding critical points, but only in rare scenarios can one obtain such precise expressions. To tackle this challenge, we propose the Gaussian approximation method, which allows us to obtain the closed-form formulas in almost all scenarios. In the oracle case, our method can fairly accurately predict the phase transition snr. In the non-oracle case, our proposed algorithm can predict the maximum allowed number of permuted rows and uncover its dependency on the sample number.

Numerical experiments reveal that the observed phase transition points are well aligned with our theoretical predictions. Our study will motivate exploiting MP algorithms (and related techniques) as an effective tool for permuted regression problems, which have found applications in machine learning, privacy, and databases.

1 INTRODUCTION

In this paper, we consider the following permuted (shuffled) linear regression problem:

$$\mathbf{Y} = \mathbf{\Pi}^{\natural} \mathbf{X} \mathbf{B}^{\natural} + \sigma \mathbf{W}, \quad (1)$$

where $\mathbf{Y} \in \mathbb{R}^{n \times m}$ denotes the matrix of observations, $\mathbf{\Pi}^{\natural} \in \{0, 1\}^{n \times n}$ is the permutation matrix, $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the design matrix, $\mathbf{B}^{\natural} \in \mathbb{R}^{p \times m}$ is the matrix of signals (regressors), $\mathbf{W} \in \mathbb{R}^{n \times m}$ denotes the additive noise matrix (with unit variance), and σ^2 is the noise variance. The task is to recover both the signal matrix \mathbf{B}^{\natural} and the permutation matrix $\mathbf{\Pi}^{\natural}$. The research on this challenging permuted regression problem dates back at least to 1970s under the name “broken sample problem” (DeGroot et al., 1971; Goel, 1975; DeGroot & Goel, 1976; 1980; Bai & Hsing, 2005). Recent years have witnessed a revival of this problem due to its broad spectrum of applications in (e.g.,) privacy protection, data integration, etc. (Unnikrishnan et al., 2015; Pananjady et al., 2018; Slawski & Ben-David, 2019; Pananjady et al., 2017; Slawski et al., 2020; Zhang & Li, 2020).

Specifically, this paper will focus on studying the “phase transition” phenomenon in recovering the whole permutation matrix $\mathbf{\Pi}^{\natural}$: the error rate for the permutation recovery sharply drops to zero

once the parameters reach certain thresholds. In particular, we leverage techniques in the *message passing* (MP) algorithm literature to identify the precise positions of the phase transition thresholds. The bottleneck in identifying the phase transition regimes lies in deriving closed-form formulas for the corresponding critical points. This is a highly challenging task because only in rare scenarios can one obtain such precise expressions. To tackle the difficulty, we propose the Gaussian approximation method which allows us to obtain the closed-form formula in almost all scenarios. We should mention that, in previous studies (Slawski et al., 2020; Pananjady et al., 2017; Zhang et al., 2022; Zhang & Li, 2020), this phase transition phenomenon was empirically observed.

Related work. The problem we study simultaneously touches two distinct areas of research: (A) permutation recovery, and (B) *message passing* (MP). In the literature of permuted linear regression, essentially all existing works used the same setting (1). Pananjady et al. (2018); Slawski & Ben-David (2019) consider the single observation model (i.e., $m = 1$) and prove that the *signal-to-noise-ratio* (snr) for the correct permutation recovery is $\mathbb{O}_p(n^c)$, where $c > 0$ is some positive constant. Slawski et al. (2020); Zhang & Li (2020); Zhang et al. (2022) investigate the multiple observations model (i.e., $m > 1$) and suggest that the snr requirement can be significantly decreased, from $\mathbb{O}_p(n^c)$ to $\mathbb{O}_p(n^{c/m})$. In particular, Zhang & Li (2020) develop an estimator which we will leverage and analyze for studying the phase transition phenomenon. Compared with the above work, our analysis can identify the precise locations of the phase transition thresholds. In this February, there comes a paper (Lufkin et al., 2024) considering the same problem as ours but in a much simpler setting (single measurement with $m = 1$). Compared with their work, our framework can easily reproduce their predicted phase transition points, answer the questions they treat as open, and predict the phase transition points in a unified framework. Another line of related research comes from the field of statistical physics. For example, using the replica method, Mézard & Parisi (1985; 1986) study the *linear assignment problem* (LAP), i.e., $\min_{\Pi} \sum_{i,j} \Pi_{ij} \mathbf{E}_{ij}$ where Π denotes a permutation matrix and \mathbf{E}_{ij} is i.i.d random variable uniformly distributed in $[0, 1]$. Martin et al. (2005) then generalize LAP to multi-index matching and presented an investigation based on MP algorithm. Recently, Caracciolo et al. (2017); Malatesta et al. (2019) extend the distribution of \mathbf{E}_{ij} to a broader class. However, all the above works exhibit no phase transition. Chertkov et al. (2010) extend it to the particle tracking problem and observe a phase transition phenomenon. Later, Semerjian et al. (2020) modify it to fit the graph matching problem, which paves way for our work in studying the permuted linear regression problem.

Our contributions. We propose the first framework to identify the precise locations of phase transition thresholds associated with permuted linear regression. In the oracle case where \mathbf{B}^{\natural} is known, our scheme is able to determine the phase transition snr. In the non-oracle case where \mathbf{B}^{\natural} is not given, our method will predict the maximum allowed number of permuted rows and uncover its dependence on the ratio p/n . In our analysis, we identify the precise positions of the phase transition points in the large-system limit, e.g., n, m, p all approach to infinity with $m/n \rightarrow \tau_m, p/n \rightarrow \tau_p$. Interestingly, numerical results well match predictions even when n, m, p are in the hundreds.

Here, we would also like to briefly mention the technical challenges. Compared with the previous works (Mezard & Montanari, 2009; Talagrand, 2010; Linusson & Wästlund, 2004; Mézard & Parisi, 1987; 1986; Parisi & Ratiéville, 2002; Semerjian et al., 2020), where the edge weights are relatively simple, our edge weights usually involve high-order interactions across Gaussian random variables and are densely correlated. To tackle this issue, our proposed approximation method to compute the phase transition thresholds consists of three parts: 1) performing Gaussian approximation; 2) modifying the leave-one-out technique; and 3) performing size correction. A detailed explanation can be found in Section 4. Hopefully, our approximation method will serve independent technical interests for researchers in the machine learning community.

Notations. $a \xrightarrow{\text{a.s.}} b$ denotes a converges almost surely to b . \mathcal{P}_n denotes the set of all possible permutation matrices: $\mathcal{P}_n \triangleq \{\Pi \in \{0, 1\}^{n \times n}, \sum_i \Pi_{ij} = 1, \sum_j \Pi_{ij} = 1\}$. The *signal-to-noise-ratio* is $\text{snr} = \frac{\|\mathbf{B}^{\natural}\|_F^2}{m \cdot \sigma^2}$, where $\|\cdot\|_F$ is the Frobenius norm and σ^2 is the variance of the sensing noise.

2 PERMUTATION RECOVERY USING THE MESSAGE PASSING ALGORITHM

Inspired by Mezard & Montanari (2009); Chertkov et al. (2010); Semerjian et al. (2020), we leverage tools from the statistical physics to identify the locations of the phase transition threshold. We start

this section with a brief review of the *linear assignment problem* (LAP), which reads as

$$\widehat{\Pi} = \operatorname{argmin}_{\Pi \in \mathcal{P}_n} \langle \Pi, \mathbf{E} \rangle, \quad (2)$$

where $\mathbf{E} \in \mathbb{R}^{n \times n}$ is a fixed matrix and \mathcal{P}_n denotes the set of all possible permutation matrices.

In our work, we first establish the link between the LAP and the permuted linear recovery, to be more specific, formulating the permutation recovery of (1) in the form of (2). Next, we predict the phase transition points by studying the matrix \mathbf{E} , which is our major contribution.

We follow the approach in Mezard & Montanari (2009); Semerjian et al. (2020) and introduce a probability measure over the permutation matrix Π , which is written as

$$\mu(\Pi) = (1/Z) \prod_i \mathbb{1}(1 - \sum_j \Pi_{ij}) \prod_j \mathbb{1}(1 - \sum_i \Pi_{ij}) \times \exp\left(-\beta \sum_{i,j} \Pi_{ij} \mathbf{E}_{ij}\right), \quad (3)$$

where $\mathbb{1}(\cdot)$ is the indicator function, Z is the normalization constant of the probability measure $\mu(\Pi)$, and $\beta > 0$ is an auxiliary parameter. It is easy to verify the following two properties, e.g., 1) ML estimator in (2) can be rewritten as $\widehat{\Pi} = \operatorname{argmax}_{\Pi} \mu(\Pi)$ ¹; and 2) the probability measure $\mu(\Pi)$ concentrates on $\widehat{\Pi}$ when letting $\beta \rightarrow \infty$.

Then, we study the impact of $\{\mathbf{E}_{ij}\}$ on the reconstructed permutation $\widehat{\Pi}$ with the *message passing* (MP) algorithm. First, we associate a probabilistic graphical model with the probability measure defined in (3). Then, we rewrite the solution in (2) in the language of the MP algorithm. Finally, we derive an equation (4) to track the convergence of the MP algorithm. By exploiting relation of (4) to the *branching random walk* (BRW) process, we can identify the phase transition points corresponding to the LAP in (2). Since this procedure is fairly standard and space is limited, we defer the detailed derivation to the Appendix.

2.1 IDENTIFICATION OF THE PHASE TRANSITION THRESHOLD

Following the approach of Semerjian et al. (2020), we split the pairs (i, j) into two groups, depending on whether Π_{ij}^{\natural} equals zero. Within each group, we model the edge weights as independent and identically distributed random variables. For the set $\{(i, j) \mid \Pi_{ij}^{\natural} = 1\}$, we denote the random variable corresponding to the edge weight \mathbf{E}_{ij} by Ω , and for all other edge weights we use $\widehat{\Omega}$.²

After linking the solution $\widehat{\Pi} = \operatorname{argmax}_{\Pi} \mu(\Pi)$ with the message passing algorithm, we adopt an approach in the spirit of *density evolution* and *state evolution*, classical techniques used to analyze the convergence of message passing and approximate message passing algorithms, respectively (Chung, 2000; Richardson & Urbanke, 2001; 2008; Donoho et al., 2009; Maleki, 2010; Bayati & Montanari, 2011; Rangan, 2011). This allows us to study the phase transition point, i.e., the condition under which $\widehat{\Pi} = \Pi^{\natural}$. Consider the following iterative updates of the random variables H and \widehat{H}

$$\widehat{H}^{(t+1)} = \min(\Omega - H^{(t)}, H'^{(t)}), \quad H^{(t+1)} = \min_{1 \leq i \leq n-1} \widehat{\Omega}_i - \widehat{H}_i^{(t)}, \quad (4)$$

where $(\cdot)^{(t)}$ denotes the update in the t -th iteration, H' is an independent copy of H , $\{H_i^{(t)}\}_{1 \leq i \leq n-1}$ and $\{\widehat{\Omega}_i\}_{1 \leq i \leq n-1}$ denote the i.i.d. copies of random variables $H_{(\cdot)}^{(t)}$ and $\widehat{\Omega}_{(\cdot)}$.

Proposition 1 (Informal). *The phase transition point, under which $\mathbb{P}(\widehat{\Pi} \neq \Pi^{\natural})$, corresponds to the point in which the distributions of H and \widehat{H} begin to diverge to infinity.*

It's noteworthy that the incorrect recovery, i.e., $\mathbb{P}(\widehat{\Pi} \neq \Pi^{\natural})$, does not mean the reconstructed correspondences are simultaneously incorrect. Numerical experiments also confirm this claim.

¹Notice that the requirement $\Pi \in \mathcal{P}_n$ is incorporated in $\mu(\Pi)$ implicitly and thus we do not need an explicit constraint.

²We conjecture that the distribution difference in the edges' weights is a necessary component in capturing the phase transition. On one hand, according to Mézard & Parisi (1986; 1987); Parisi & Ratiéville (2002); Linusson & Wästlund (2004); Mezard & Montanari (2009); Talagrand (2010), there is no phase transition phenomenon in LAP if the edges' weights, i.e., \mathbf{E}_{ij} , are assumed to be i.i.d uniformly distributed in $[0, 1]$. On the other hand, Semerjian et al. (2020) show a phase transition phenomenon when assuming the weights \mathbf{E}_{ij} follow different distributions among the edges associated with the ground-truth correspondence $\Pi_{ij}^{\natural} = 1$ and the rest edges.

Relation to branching random walk (BRW) process. Conditional on the event that the permutation can be perfectly reconstructed, we have $H + \widehat{H}' > \Omega$ (c.f. (19)) and can simplify (4) as

$$H^{(t+1)} = \min_{1 \leq i \leq n-1} H_i^{(t)} + \Xi_i, \quad (5)$$

where Ξ is defined as the difference between $\widehat{\Omega}$ and Ω , which is written as $\Xi \triangleq \widehat{\Omega} - \Omega$, and $\{H_i^{(t)}\}_{1 \leq i \leq n-1}$ and $\{\Xi_i\}_{1 \leq i \leq n-1}$ denote the i.i.d. copies of random variables $H_{(\cdot)}^{(t)}$ and $\Xi_{(\cdot)}$.

Adopting the same viewpoint of Semerjian et al. (2020), we treat (5) as a *branching random walk* (BRW) process, which enjoys the following property.

Theorem 1 (Hammersley (1974); Kingman (1975); Semerjian et al. (2020)). *Consider the recursive distributional equation $K^{(t+1)} = \min_{1 \leq i \leq n} K_i^{(t)} + \Xi_i$, where $K_i^{(t)}$ and Ξ_i are i.i.d. copies of random variables $K_{(\cdot)}^{(t)}$ and $\Xi_{(\cdot)}$, we have $\frac{K^{(t+1)}}{t} \xrightarrow{\text{a.s.}} -\inf_{\theta > 0} \frac{1}{\theta} \log [\sum_{i=1}^n \mathbb{E} e^{-\theta \Xi_i}]$, conditional on the event that $\lim_{t \rightarrow \infty} K^{(t)} \neq \infty$.*

With Theorem 1, we can compute phase transition point for the correct (full) permutation recovery, i.e., $H + H' > \Omega$, by letting $\inf_{\theta > 0} \frac{1}{\theta} \log [\sum_{i=1}^n \mathbb{E} e^{-\theta \Xi_i}] = 0$, since otherwise the condition in (19) will be violated (see a detailed explanation in Appendix). In practice, directly computing the infimum of $\inf_{\theta > 0} \frac{1}{\theta} \log [\sum_{i=1}^n \mathbb{E} e^{-\theta \Xi_i}]$ is only possible for limited scenarios. The next section proposes an approximate computation method for the phase transition points, which is capable of covering a broader class of scenarios.

3 ANALYSIS OF THE PHASE TRANSITION POINTS

Recall that, in this paper, we consider the following linear regression problem with permuted labels

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

where $\mathbf{Y} \in \mathbb{R}^{n \times m}$ represents the matrix of observations, $\mathbf{\Pi}^{\natural} \in \mathcal{P}_n$ denotes the permutation matrix to be reconstructed, $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the sensing matrix with each entry \mathbf{X}_{ij} following the i.i.d standard normal distribution, $\mathbf{B}^{\natural} \in \mathbb{R}^{p \times m}$ is the matrix of signals, and $\mathbf{W} \in \mathbb{R}^{n \times m}$ represents the additive noise matrix and its entries \mathbf{W}_{ij} are i.i.d standard normal random variables. In addition, we denote h as the number of permuted rows corresponding to the permutation matrix $\mathbf{\Pi}^{\natural}$.

In this work, we focus on studying the ‘‘phase transition’’ phenomenon in recovering $\mathbf{\Pi}^{\natural}$ from the pair (\mathbf{Y}, \mathbf{X}) . That is, the error rate for the permutation recovery sharply drops to zero once certain parameters reach the thresholds. In particular, our analysis will identify the precise positions of the phase transition points in the large-system limit, i.e., n, m, p , and h all approach to infinity with $m/n \rightarrow \tau_m, p/n \rightarrow \tau_p, h/n \rightarrow \tau_h$. We will separately study the phase transition phenomenon in 1) the oracle case where \mathbf{B}^{\natural} is given as a prior, and 2) the non-oracle case where \mathbf{B}^{\natural} is unknown.

In this section, we consider the oracle scenario, as a warm-up example. To reconstruct the permutation matrix $\mathbf{\Pi}^{\natural}$, we adopt the following *maximum-likelihood* (ML) estimator:

$$\widehat{\mathbf{\Pi}}^{\text{oracle}} = \operatorname{argmin}_{\mathbf{\Pi} \in \mathcal{P}_n} \langle \mathbf{\Pi}, -\mathbf{Y} \mathbf{B}^{\natural \top} \mathbf{X}^{\top} \rangle. \quad (6)$$

Denoting the variable $\mathbf{E}_{ij}^{\text{oracle}}$ as $-\mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{B}^{\natural} \mathbf{B}^{\natural \top} \mathbf{X}_j - \sigma \mathbf{W}_i^{\top} \mathbf{B}^{\natural \top} \mathbf{X}_j$, ($1 \leq i, j \leq n$), we can transform the objective function in (6) as the canonical form of LAP, i.e., $\sum_{i,j} \mathbf{\Pi}_{ij} \mathbf{E}_{ij}^{\text{oracle}}$.

3.1 THE PHASE TRANSITION THRESHOLD FOR THE ORACLE CASE

In the oracle case where \mathbf{B}^{\natural} is known, we define the following random variable Ξ :

$$\Xi = \mathbf{x}^{\top} \mathbf{B}^{\natural} \mathbf{B}^{\natural \top} (\mathbf{x} - \mathbf{y}) + \sigma \mathbf{w} \mathbf{B}^{\natural \top} (\mathbf{x} - \mathbf{y}), \quad (7)$$

where \mathbf{x} and \mathbf{y} follow the distribution $\mathbf{N}(\mathbf{0}, \mathbf{I}_{p \times p})$, and \mathbf{w} follows the distribution $\mathbf{N}(\mathbf{0}, \mathbf{I}_{m \times m})$.

Assumption 1. *We ignore the weak correlation across the $\mathbf{E}_{ij}^{\text{oracle}}$ and view the corresponding Ξ_i as i.i.d. copies of (7).*

Numerical experiments show that we can safely adopt Assumption 1 without much sacrifice in the prediction accuracy, see Table 1 and 2. Recalling Theorem 1, we predict the critical points by letting

$$\inf_{\theta > 0} 1/\theta \cdot \log \left(\sum_{i=1}^n \mathbb{E} e^{-\theta \Xi_i} \right) = \inf_{\theta > 0} 1/\theta \cdot (\log n + \log \mathbb{E} e^{-\theta \Xi}) = 0. \quad (8)$$

The computation procedure consists of two stages:

- **Step I.** We compute the optimal θ_* , which is written as $\theta_* = \operatorname{argmin}_{\theta > 0} 1/\theta \cdot (\log n + \log \mathbb{E} e^{-\theta \Xi_i})$.
- **Step II.** We plug the optimal θ^* into (8) and obtain the phase transition snr accordingly.

The following context illustrates the computation details.

Step I: Determine θ_* . The key in determining θ_* lies in the computation of $\mathbb{E} e^{-\theta \Xi}$, which is summarized in the following proposition.

Lemma 1. For the random variable Ξ defined in (7), we can write its expectation as

$$\mathbb{E} e^{-\theta \Xi} = \prod_{i=1}^{\operatorname{rank}(\mathbf{B}^{\natural})} [1 + 2\theta \lambda_i^2 - \theta^2 \lambda_i^2 (\lambda_i^2 + 2\sigma^2)]^{-\frac{1}{2}}, \quad (9)$$

provided that

$$\theta^2 \sigma^2 \lambda_i^2 < 1 \text{ and } \theta^2 \lambda_i^2 (\lambda_i^2 + 2\sigma^2) \leq 1 + 2\theta \lambda_i^2 \quad (10)$$

hold for all singular values λ_i of \mathbf{B}^{\natural} , $1 \leq i \leq \operatorname{rank}(\mathbf{B}^{\natural})$.

Remark 1. When the conditions in (10) is violated, we have the expectation $\mathbb{E} e^{-\theta \Xi}$ diverge to infinity, which suggests the optimal θ_* for $\inf_{\theta > 0} \log(n \cdot \mathbb{E} e^{-\theta \Xi})/\theta$ cannot be achieved.

With (9), we can compute the optimal θ_* by setting the gradient $\frac{\partial [\log(n \cdot \mathbb{E} e^{-\theta \Xi})/\theta]}{\partial \theta} = 0$. However, a closed-form of the exact solution for θ^* is out of reach. As a mitigation, we resort to approximating $\log \mathbb{E} e^{-\theta \Xi}$ by its lower-bound, which reads as

$$\log \mathbb{E} e^{-\theta \Xi} \geq \frac{\theta^2}{2} \left(\|\mathbf{B}^{\natural \top} \mathbf{B}^{\natural}\|_{\mathbb{F}}^2 + 2\sigma^2 \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2 \right) - \theta \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2.$$

The corresponding minimum value $\tilde{\theta}_*$ is thus obtained by minimizing the lower-bound, which is written as $\tilde{\theta}_* = 2 \log n / \left(\|\mathbf{B}^{\natural \top} \mathbf{B}^{\natural}\|_{\mathbb{F}}^2 + 2\sigma^2 \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2 \right)$.

Step II: Compute the phase transition snr. We predict the phase transition point $\operatorname{snr}_{\text{oracle}}$ by letting the lower bound being zero, which can be written as

$$\frac{\log n}{\theta^*} - \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2 + \frac{\theta^*}{2} \left(\|\mathbf{B}^{\natural \top} \mathbf{B}^{\natural}\|_{\mathbb{F}}^2 + 2\sigma^2 \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2 \right) = 0.$$

With standard algebraic manipulations, we have

Proposition 2. The predicted phase transition for the oracle case in (6) can be computed as

$$2(\log n) \operatorname{snr}_{\text{oracle}} \cdot \|\mathbf{B}^{\natural \top} / \|\mathbf{B}^{\natural}\|_{\mathbb{F}} \cdot \mathbf{B}^{\natural} / \|\mathbf{B}^{\natural}\|_{\mathbb{F}}\|_{\mathbb{F}}^2 + 4 \log n / m = \operatorname{snr}_{\text{oracle}}. \quad (11)$$

We then evaluate the accuracy of our predicted phase transition threshold by comparing the predicted values with the numerical values.

3.2 NUMERICAL METHOD FOR PHASE TRANSITION POINTS

Noticing that the correct recovery rate is in monotonic non-decreasing relation with the snr, that is, the error rate $\mathbb{P}(\hat{\mathbf{\Pi}} \neq \mathbf{\Pi}^{\natural})$ for a larger snr is at least equal to, if not less than, that for a smaller snr, we propose a binary-search-based method to compute the phase transition points. The detailed description is in Algorithm 1.

Complexity analysis. For a given precision threshold ε , each iteration (Line 2 to Line 12 in Algorithm 1) takes $O(\log \frac{1}{\varepsilon})$ rounds to converge and it runs the permutation recovery algorithm T_{iter} times in each round. That is, the computational complexity of phase transition point for each configuration is $O(T_{\text{iter}} \cdot \log \frac{1}{\varepsilon})$.

Experiment result. The results are shown in Table 1, from which we can conclude the phase transition threshold snr can be predicted to a good extent. In addition, we observe that the gap between the theoretical values and the numerical values keeps shrinking as m increases.

Algorithm 1 Numerical method to compute the phase transition points.

- 1: **Initialization.** Set the initial search range for snr as $[l, r]$. Define the precision threshold ε , the error rate Err (we pick it as 20) and iterations T_{iter} .
 - 2: **while** $|l - r| > \varepsilon$ **do**
 - 3: Set $\text{snr}_{\text{middle}} = \frac{l+r}{2}$.
 - 4: Given $\text{snr}_{\text{middle}}$, we run (2) for T_{iter} times.
 - 5: Compute the error rate of full permutation recovery, namely, $\mathbb{P}(\hat{\Pi} \neq \Pi^{\natural})$.
 - 6:
 - 7: **if** the error rate is below Err **then**
 - 8: $r \rightarrow \text{snr}_{\text{middle}} - \varepsilon$, # we have $\text{snr}_{\text{middle}}$ be greater than the phase transition point
 - 9: **else**
 - 10: $l \rightarrow \text{snr}_{\text{middle}} + \varepsilon$. # we have $\text{snr}_{\text{middle}}$ be no greater than the phase transition point
 - 11: **end if**
 - 12: **end while**
 - 13: **Output.** Return the phase transition point $\text{snr}_{\text{middle}}$.
-

Table 1: Comparison between the predicted value of the phase transition threshold $\text{snr}_{\text{oracle}}$ in Proposition 2 and its simulated value when $n = 500$. \mathbf{P} denotes the predicted value while \mathbf{S} denotes the simulated value (i.e., mean \pm std). \mathbf{S} corresponds to the snr when the error rate drops below 0.05. A detailed description of the numerical method can be found in the appendix (code also included).

m	20	30	40	50	60	70
\mathbf{P}	3.283	1.415	0.902	0.662	0.523	0.432
\mathbf{S}	2.529 ± 0.079	1.290 ± 0.054	0.872 ± 0.034	0.649 ± 0.012	0.515 ± 0.016	0.429 ± 0.015
m	100	110	120	130	140	150
\mathbf{P}	0.284	0.255	0.231	0.211	0.195	0.181
\mathbf{S}	0.282 ± 0.008	0.256 ± 0.006	0.232 ± 0.006	0.212 ± 0.004	0.196 ± 0.006	0.183 ± 0.005

3.3 GAUSSIAN APPROXIMATION OF THE PHASE TRANSITION THRESHOLD

From the above analysis, we can see that deriving a closed-form expression of the infimum value θ of $\log(n\mathbb{E}e^{-\theta\Xi})/\theta$ can be difficult. In fact, in certain scenarios, even obtaining a closed-form expression of $\mathbb{E}e^{-\theta\Xi}$ is difficult. To handle such challenge, we propose to approximate random variable Ξ by a Gaussian $\mathcal{N}(\mathbb{E}\Xi, \text{Var}\Xi)$, namely,

$$\mathbb{E}e^{-\theta\Xi} \approx \exp\left(-\theta\mathbb{E}\Xi + \frac{\theta^2}{2}\text{Var}\Xi\right). \quad (12)$$

With this approximation, we can express $\theta_* \triangleq \inf \log(n\mathbb{E}e^{-\theta\Xi})/\theta$ in a closed form, which is $\sqrt{2 \log n / \text{Var}\Xi}$.

Theorem 2. For the random variable Ξ defined in (7), its mean and variance can be computed as

$$\mathbb{E}\Xi = \|\mathbf{B}^{\natural}\|_{\text{F}}^2, \quad \text{Var}\Xi = 3\|\mathbf{B}^{\natural}\mathbf{B}^{\natural\top}\|_{\text{F}}^2 + 2\sigma^2\|\mathbf{B}^{\natural}\|_{\text{F}}^2. \quad (13)$$

Then, we can predict the phase transition point as follows.

Proposition 3. With Gaussian approximation, we can predict the critical point corresponding to the phase transition in (8) as

$$2(\log n) \cdot \text{Var}\Xi = (\mathbb{E}\Xi)^2, \quad (14)$$

where $\mathbb{E}\Xi$ and $\text{Var}\Xi$ can be found in Theorem 2.

Example 1 (Scaled identity matrix). We consider the scenario where $\mathbf{B}^{\natural} = \lambda\mathbf{I}_{m \times m}$. Then, we have $\mathbf{B}^{\natural}/\|\mathbf{B}^{\natural}\|_{\text{F}} = m^{-1/2}\mathbf{I}$. The phase transition threshold $\text{snr}_{\text{oracle}}$ in (11) is then $4 \log n / (m - 2 \log n)$, and the phase transition threshold $\widetilde{\text{snr}}_{\text{oracle}}$ in (14) as $4 \log n / (m - 6 \log n)$. This solution is almost identical to (11) in the limit as $\text{snr}_{\text{oracle}} \approx \widetilde{\text{snr}}_{\text{oracle}} \approx 4 \log n / m \simeq n^{\frac{4}{m}} - 1$.

Moreover, we should mention that 1) our approximation method applies to a general matrix \mathbf{B}^\natural , not limited to a scaled identity matrix; and 2) our approximation method can also predict the phase transition thresholds to a good extent when the entries \mathbf{X}_{ij} are sub-Gaussian. The numerical experiments are given in Table 2, from which we can conclude that the predicted values are well aligned with the simulation results.

Table 2: Comparison between the predicted value of the phase transition threshold $\widetilde{\text{snr}}_{\text{oracle}}$ in Proposition 3 and its simulated value when $n = 600$. In **(Case 1)**, half of singular values are with λ and the other half are with $\lambda/2$; while in **(Case 2)**, half of the singular values are with λ and the other half are with $(3\cdot\lambda)/4$. **Gauss** refers to $\mathbf{X}_{ij} \stackrel{\text{i.i.d}}{\sim} \mathcal{N}(0, 1)$ and **Unif** refers to $\mathbf{X}_{ij} \stackrel{\text{i.i.d}}{\sim} \text{Unif}[-1, 1]$. **P** denotes the predicted value and **S** denotes the simulated value (i.e., mean \pm std). **S** corresponds to the snr when the error rate drops below 0.05. Averaged over 20 repetitions.

m		100	110	120	130	140	150
(Case 1) P		0.297	0.266	0.241	0.220	0.203	0.188
(Gauss) S		0.307 ± 0.009	0.275 ± 0.005	0.246 ± 0.006	0.227 ± 0.007	0.210 ± 0.005	0.194 ± 0.004
(Unif) S		0.294 ± 0.008	0.266 ± 0.005	0.239 ± 0.008	0.216 ± 0.004	0.201 ± 0.005	0.189 ± 0.006
(Case 2) P		0.310	0.276	0.249	0.227	0.209	0.193
(Gauss) S		0.294 ± 0.008	0.266 ± 0.006	0.241 ± 0.005	0.220 ± 0.004	0.204 ± 0.006	0.190 ± 0.003
(Unif) S		0.287 ± 0.007	$0.255 \pm .0043$	0.234 ± 0.007	0.213 ± 0.005	0.197 ± 0.003	0.185 ± 0.005

3.4 RELATED WORK COMPARISON

In the end, we want to discuss one closely related work (Lufkin et al., 2024), which studies the same topic but in a much simpler setting (i.e., single measurement with $m = 1$). Compared with their work, our framework can easily produce results that the rigorous method in Lufkin et al. (2024) regards as an open problem. For example, the results (on the oracle case) which we treat as a warm-up example, are unsolved and specifically mentioned in the last paragraph. In particular, our framework can derive their proposed conjectured phase transition $\text{snr} = n^{4/m}$. Moreover, our framework can also tell when the conjecture holds, i.e., \mathbf{B}^\natural is an identity matrix, and our predictions' accuracy has been extensively verified by numerical experiments.

Other advantages of our work include 1) our ability to more accurately pinpoint phase transition points (their work can only obtain the lower bound of the phase transition point, while our work can predict the precise location); 2) our applicability to a wider array of cases (we cover the case when $m > 1$ while their method only works for $m = 1$); and 3) our consolidation into a more cohesive framework to predict the phase transition point.

4 EXTENSION TO NON-ORACLE CASE

Having analyzed the oracle case in the previous section, we now extend the analysis to the non-oracle case, where the value of \mathbf{B}^\natural is not given. Different from the oracle case, the ML estimator reduces to a *quadratic assignment problem* (QAP) as opposed to LAP. As a mitigation, we adopt the estimator in Zhang & Li (2020), which reconstructs the permutation matrix within the LAP framework, i.e.,

$$\widehat{\Pi}^{\text{non-oracle}} = \underset{\Pi \in \mathcal{P}_n}{\text{argmin}} \left\langle \Pi, -\mathbf{Y}\mathbf{Y}^\top \mathbf{X}\mathbf{X}^\top \right\rangle. \quad (15)$$

We expect this estimator can yield good insights of the permuted linear regression since 1) this estimator can reach the statistical optimality in a broad range of parameters; and 2) estimator exhibits a phase transition phenomenon, a similar pattern as the oracle case. The technical details of the above claims can be found in Zhang & Li (2020).

Following the same procedure as in Section 3, we identify the phase transition threshold snr with Theorem 1. First, we rewrite the random variable Ξ as

$$\Xi \cong \mathbf{Y}_i \mathbf{Y}^\top \mathbf{X} (\mathbf{X}_{\pi^\natural(i)} - \mathbf{X}_j)^\top, \quad (16)$$

where i and j are uniformly distributed among the set $\{1, 2, \dots, n\}$. Afterwards, we adopt the Gaussian approximation scheme illustrated in Subsection 3.3 and determine the phase transition points by first computing $\mathbb{E}\Xi$ and $\text{Var}\Xi$, respectively.

Theorem 3. For the random variable Ξ defined in (16), its mean $\mathbb{E}\Xi$ and variance $\text{Var}\Xi$ are

$$\begin{aligned} \mathbb{E}\Xi &\simeq n(1 - \tau_h) \left[(1 + \tau_p) \left\| \mathbf{B}^{\natural} \right\|_F^2 + m\tau_p\sigma^2 \right], \\ \text{Var}\Xi &\simeq n^2\tau_h(1 - \tau_h)\tau_p^2 \left[\left\| \mathbf{B}^{\natural} \right\|_F^2 + m\sigma^2 \right]^2 + n^2 \left[2\tau_p + 3(1 - \tau_h)^2 \right] \left\| \mathbf{B}^{\natural\top} \mathbf{B}^{\natural} \right\|_F^2 \\ &\quad + n^2 \left[6\tau_p(1 - \tau_h)^2 + (3 - \tau_h)\tau_p^2 \right] \left\| \mathbf{B}^{\natural\top} \mathbf{B}^{\natural} \right\|_F^2, \end{aligned}$$

respectively, where the definitions of τ_p and τ_h can be found in Section 3.

The proof of Theorem 3 is quite complicated, involving Wick's theorem, Stein's lemma, the conditional technique, and the leave-one-out technique, etc. For clarity, we provide only an outline of the proof strategy here.

Proof outline. We decompose the random variable Ξ as $\Xi = \Xi_1 + \sigma(\Xi_2 + \Xi_3) + \sigma^2\Xi_4$, where Ξ_i ($1 \leq i \leq 4$) are respectively defined as

$$\begin{aligned} \Xi_1 &\triangleq \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{B}^{\natural} \mathbf{B}^{\natural\top} \mathbf{X}^{\top} \mathbf{\Pi}^{\natural\top} \mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j), & \Xi_2 &\triangleq \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{B}^{\natural} \mathbf{W}^{\top} \mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j), \\ \Xi_3 &\triangleq \mathbf{W}_i^{\top} \mathbf{B}^{\natural\top} \mathbf{X}^{\top} \mathbf{\Pi}^{\natural\top} \mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j), & \Xi_4 &\triangleq \mathbf{W}_i^{\top} \mathbf{W}^{\top} \mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j). \end{aligned}$$

Unlike the oracle case, obtaining a closed-form expression of $\mathbb{E}e^{-\theta\Xi}$ would be too difficult. Hence, we adopt the Gaussian approximation method as presented in Section 3.3. The task then transforms to computing the expectation and variance of Ξ .

Computation of the mean $\mathbb{E}\Xi$. For the computation of the mean $\mathbb{E}\Xi$, we can verify that $\mathbb{E}\Xi_2$ and $\mathbb{E}\Xi_3$ are both zero, due to the independence between \mathbf{X} and \mathbf{W} . For $\mathbb{E}\Xi_1$ and $\mathbb{E}\Xi_4$, we adopt Wick's theorem to obtain

$$\mathbb{E}\Xi_1 = n(1 - \tau_h)(1 + \tau_p)[1 + o_p(1)] \left\| \mathbf{B}^{\natural} \right\|_F^2, \quad \mathbb{E}\Xi_4 = nm\tau_p(1 - \tau_h)[1 + o_p(1)].$$

Computation of the variance $\text{Var}\Xi$. Since $\text{Var}\Xi = \mathbb{E}\Xi^2 - (\mathbb{E}\Xi)^2$, we just need to compute $\mathbb{E}\Xi^2$, which can be expanded into the following six terms

$$\mathbb{E}\Xi^2 = \mathbb{E}\Xi_1^2 + \sigma^2\mathbb{E}\Xi_2^2 + \sigma^2\mathbb{E}\Xi_3^2 + \sigma^4\mathbb{E}\Xi_4^2 + 2\sigma^2\mathbb{E}\Xi_1\Xi_4 + 2\sigma^2\mathbb{E}\Xi_2\Xi_3.$$

The computation of above terms turns out to be quite complex due to the high order Gaussian random variables. For example, the term $\mathbb{E}\Xi_1^2$ involves the eighth-order Gaussian moments, the terms $\mathbb{E}\Xi_2^2$, $\mathbb{E}\Xi_3^2$, $\mathbb{E}\Xi_1\Xi_4$ and $\mathbb{E}\Xi_2\Xi_3$ all involve the sixth-order Gaussian variables, etc. To handle the difficulties in computing $\mathbb{E}\Xi^2$, we propose the following computation procedure, which can be roughly divided into 3 phases.

- **Phase I: Leave-one-out decomposition.** The major technical difficulty comes from the correlation between the product $\mathbf{X}^{\top} \mathbf{\Pi}^{\natural} \mathbf{X}$ and the difference $\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j$. We decouple this correlation by first rewriting the matrix $\mathbf{X}^{\top} \mathbf{\Pi}^{\natural} \mathbf{X}$ as the sum $\sum_{\ell} \mathbf{X}_{\ell} \mathbf{X}_{\ell}^{\top}$. Then we collect all terms $\mathbf{X}_{\ell} \mathbf{X}_{\ell}^{\top}$ independent of $\mathbf{X}_{\pi^{\natural}(i)}$ and \mathbf{X}_j in the matrix $\mathbf{\Sigma}$ and leave the remaining terms to the matrix $\mathbf{\Delta}$, i.e., $\mathbf{\Delta} \triangleq \mathbf{X}^{\top} \mathbf{\Pi}^{\natural} \mathbf{X} - \mathbf{\Sigma}$. This decomposition shares the same spirit as the leave-one-out technique (Karoui, 2013; Bai & Silverstein, 2010; Karoui, 2018; Sur et al., 2019). Then, we divide all terms in $\mathbb{E}\Xi^2$ into 3 categories: 1) terms only containing matrix $\mathbf{\Sigma}$; 2) terms containing both $\mathbf{\Sigma}$ and $\mathbf{\Delta}$; and 3) terms only containing $\mathbf{\Delta}$.
- **Phase II: Conditional technique.** Concerning the terms in the first two categories, which covers the majority of terms, we can exploit the independence among the rows in the sensing matrix \mathbf{X} . With the conditional technique, we can reduce the order of Gaussian moments by separately taking the expectation w.r.t $\mathbf{\Sigma}$ and w.r.t vectors $\mathbf{X}_{\pi^{\natural}(i)}$ and \mathbf{X}_j .
- **Phase III: Direct computation.** For the few terms in the third category (i.e., terms only containing $\mathbf{\Delta}$), we compute the high-order Gaussian moments by exhausting all terms and iterative applying of Wick's Theorem and Stein's Lemma, which can reduce the higher-order Gaussian moments to lower-orders.

For the interested readers, the technical details can be found in Section D.

4.1 AN ILLUSTRATING EXAMPLE

Afterwards, we predict the phase transition points. Unlike the oracle case, we notice the edge weights \mathbf{E}_{ij} are strongly correlated, especially when $j = \pi^{\sharp}(i)$, which corresponds to the non-permuted rows. To factor out these dependencies, we only take the permuted rows into account and correct the sample size from n to $\tau_h n$.

Proposition 4. *The predicted $\text{snr}_{\text{non-oracle}}$ for the non-oracle case in (15) can be computed by solving*

$$2 \log(n\tau_h) \text{Var}\Xi = (\mathbb{E}\Xi)^2,$$

where $\mathbb{E}\Xi$ and $\text{Var}\Xi$ are in Theorem 3.

To illustrate the prediction accuracy, we consider the case where \mathbf{B}^{\sharp} 's singular values are of the same order, i.e., $\frac{\lambda_i(\mathbf{B}^{\sharp})}{\lambda_j(\mathbf{B}^{\sharp})} = O(1)$, $1 \leq i, j \leq m$, where $\lambda_i(\cdot)$ denotes the i -th singular value. Then, we obtain the $\text{snr}_{\text{non-oracle}}$, which is written as

$$\text{snr}_{\text{non-oracle}} \approx \eta_1 / \eta_2. \quad (17)$$

Here, η_1 and η_2 are defined as

$$\eta_1 \triangleq 2\tau_h\tau_p^2 \log(n\tau_h) - \tau_p(\tau_p + 1)(1 - \tau_h) + \tau_p \sqrt{2(1 - \tau_h)\tau_h \cdot \log(n\tau_h)},$$

$$\eta_2 \triangleq (1 - \tau_h)(\tau_p + 1)^2 - 2\tau_h\tau_p^2 \log(n\tau_h).$$

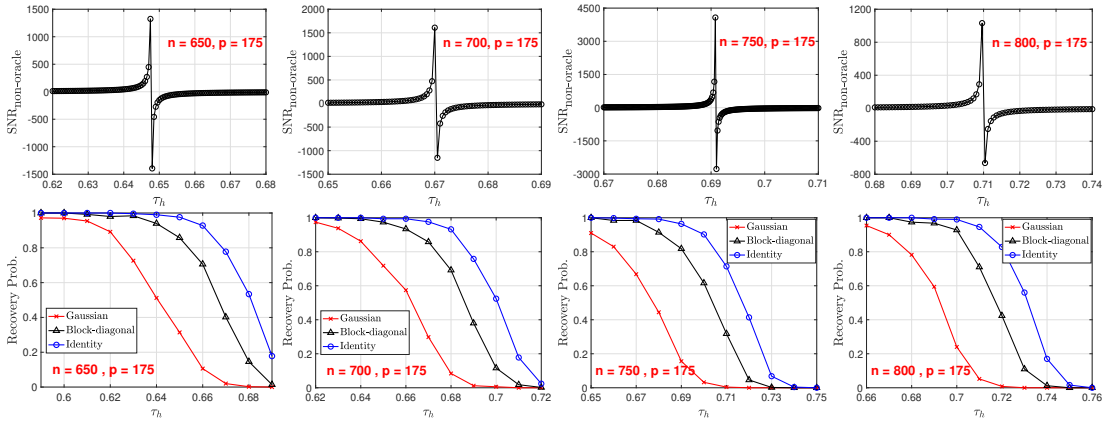


Figure 1: **(Upper panel)** Predicted phase transition points $\text{snr}_{\text{non-oracle}}$. **(Lower panel)** Plot of the recovery rate under the noiseless setting, i.e., $\text{snr} = \infty$. **Gaussian:** $\mathbf{B}_{ij}^{\sharp} \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$; **Identity:** $\mathbf{B}^{\sharp} = \mathbf{I}_{p \times p}$; **Block-diagonal:** $\mathbf{B}^{\sharp} = \text{diag}\{1, \dots, 1, 0.5, \dots, 0.5\}$. We observe that the correct recovery rates drop sharply within the regions of our predicted value.

Note that the predicted $\text{snr}_{\text{non-oracle}}$ varies for different τ_h and τ_p . Viewing $\text{snr}_{\text{non-oracle}}$ as a function of τ_h , we observe a singularity point of τ_h , which corresponds to the case when $\eta_2 = 0$. This suggests a potential phase transition phenomenon w.r.t. τ_h . To validate the predicted phenomenon, we consider the noiseless case, i.e., $\text{snr} = \infty$, and reconstruct the permutation matrix $\mathbf{\Pi}^{\sharp}$ with (2). Numerical experiments in Figure 1 confirm our prediction.

Due to the space limit, this section only presents a glimpse of our results in the non-oracle case. The technical details along with the additional numerical experiments can be found in Section C.

5 CONCLUSIONS

This is the first work that can identify the precise location of phase transition thresholds of permuted linear regressions. For the oracle case where the signal \mathbf{B}^{\sharp} is given as a prior, our analysis can predict the phase transition threshold $\text{snr}_{\text{oracle}}$ to a good extent. For the non-oracle case where \mathbf{B}^{\sharp} is not given, we modified the leave-one-out technique to approximately compute the phase critical $\text{snr}_{\text{non-oracle}}$ value for the phase transition, as the precise computation becomes significantly complicated as the high-order interaction between Gaussian random variables is involved. Moreover, we associated the singularity point in $\text{snr}_{\text{non-oracle}}$ with a phase transition point w.r.t the allowed number of permuted rows. Moreover, we present numerous numerical experiments to confirm the accuracy of our theoretical predictions.

REFERENCES

- 486
487
488 Zhidong Bai and Tailen Hsing. The broken sample problem. *Probability Theory and Related Fields*,
489 131(4):528–552, 2005.
- 490 Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, vol-
491 ume 20. Springer, 2010.
- 492
493 Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with
494 applications to compressed sensing. *IEEE Trans. Inf. Theory*, 57(2):764–785, 2011.
- 495 George T. Cantwell and M. E. J. Newman. Message passing on networks with loops. *Proceedings*
496 *of the National Academy of Sciences*, 116(47):23398–23403, 2019.
- 497
498 Sergio Caracciolo, Matteo P D’Achille, Enrico M Malatesta, and Gabriele Sicuro. Finite-size cor-
499 rections in the random assignment problem. *Physical Review E*, 95(5):052129, 2017.
- 500 Michael Chertkov, Lukas Kroc, F Krzakala, M Vergassola, and L Zdeborová. Inference in particle
501 tracking experiments by passing messages between images. *Proceedings of the National Academy*
502 *of Sciences*, 107(17):7663–7668, 2010.
- 503
504 Sae-Young Chung. *On the construction of some capacity-approaching coding schemes*. PhD thesis,
505 Massachusetts Institute of Technology, 2000.
- 506 Morris H. DeGroot and Prem K. Goel. The matching problem for multivariate normal data. *Sankhyā:*
507 *The Indian Journal of Statistics, Series B (1960-2002)*, 38(1):14–29, 1976.
- 508
509 Morris H. DeGroot and Prem K. Goel. Estimation of the correlation coefficient from a broken
510 random sample. *The Annals of Statistics*, 8(2):264–278, 03 1980.
- 511
512 Morris H. DeGroot, Paul I. Feder, and Prem K. Goel. Matchmaking. *The Annals of Mathematical*
513 *Statistics*, 42(2):578–593, 04 1971.
- 514
515 David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for com-
516 pressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.
- 517
518 Prem K. Goel. On re-pairing observations in a broken random sample. *The Annals of Statistics*, 3
519 (6):1364–1369, 11 1975.
- 520
521 John M Hammersley. Postulates for subadditive processes. *The Annals of Probability*, 2(4):652–680,
522 1974.
- 523
524 Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge university press, 1990.
- 525
526 Svante Janson. *Gaussian Hilbert Spaces*. Cambridge Tracts in Mathematics. Cambridge University
527 Press, 1997. doi: 10.1017/CBO9780511526169.
- 528
529 Nouredine El Karoui. Asymptotic behavior of unregularized and ridge-regularized high-
530 dimensional robust regression estimators: rigorous results. *arXiv preprint arXiv:1311.2445*, 2013.
- 531
532 Nouredine EL Karoui. On the impact of predictor geometry on the performance on high-
533 dimensional ridge-regularized generalized robust regression estimators. *Probability Theory and*
534 *Related Fields*, 170(1-2):95–175, 2018.
- 535
536 John Frank Charles Kingman. The first birth problem for an age-dependent branching process. *The*
537 *Annals of Probability*, pp. 790–801, 1975.
- 538
539 Alec Kirkley, George T. Cantwell, and M. E. J. Newman. Belief propagation for networks with
loops. *Science Advances*, 7(17):eabf1211, 2021. doi: 10.1126/sciadv.abf1211.
- 536
537 Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-product algo-
538 rithm. *IEEE Transactions on information theory*, 47(2):498–519, 2001.
- 539
Svante Linusson and Johan Wästlund. A proof of parisi’s conjecture on the random assignment
problem. *Probability Theory and Related Fields*, 128(3):419–440, 2004.

- 540 Leon Lufkin, Yihong Wu, and Jiaming Xu. Sharp information-theoretic thresholds for shuffled linear
541 regression. *arXiv preprint arXiv:2402.09693*, 2024.
542
- 543 David JC MacKay, David JC Mac Kay, et al. *Information theory, inference and learning algorithms*.
544 Cambridge university press, 2003.
- 545 Enrico M Malatesta, Giorgio Parisi, and Gabriele Sicuro. Fluctuations in the random-link matching
546 problem. *Physical Review E*, 100(3):032102, 2019.
547
- 548 Arian Maleki. *Approximate message passing algorithms for compressed sensing*. PhD thesis, Stan-
549 ford University, 2010.
- 550 OC Martin, M Mézard, and Olivier Rivoire. Random multi-index matching problems. *Journal of*
551 *Statistical Mechanics: Theory and Experiment*, 2005(09):P09006, 2005.
552
- 553 Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University
554 Press, 2009.
- 555 Marc Mézard and Giorgio Parisi. Replicas and optimization. *Journal de Physique Lettres*, 46(17):
556 771–778, 1985.
557
- 558 Marc Mézard and Giorgio Parisi. Mean-field equations for the matching and the travelling salesman
559 problems. *EPL (Europhysics Letters)*, 2(12):913, 1986.
560
- 561 Marc Mézard and Giorgio Parisi. On the solution of the random link matching problems. *Journal*
562 *de Physique*, 48(9):1451–1459, 1987.
- 563 Heinz Neudecker and Tom Wansbeek. Fourth-order properties of normally distributed random ma-
564 trices. *Linear Algebra and Its Applications*, 97:13–21, 1987.
565
- 566 Ashwin Pananjady, Martin J Wainwright, and Thomas A Courtade. Denoising linear models with
567 permuted data. In *Proceedings of the 2017 IEEE International Symposium on Information Theory*
568 *(ISIT)*, pp. 446–450, Aachen, Germany, 2017.
- 569 Ashwin Pananjady, Martin J. Wainwright, and Thomas A. Courtade. Linear regression with shuffled
570 data: Statistical and computational limits of permutation recovery. *IEEE Trans. Inf. Theory*, 64
571 (5):3286–3300, 2018.
572
- 573 Giorgio Parisi and Matthieu Ratiéville. On the finite size corrections to some random matching
574 problems. *The European Physical Journal B-Condensed Matter and Complex Systems*, 29(3):
575 457–468, 2002.
- 576 Sundeep Rangan. Generalized approximate message passing for estimation with random linear
577 mixing. In *Proceedings of the 2011 IEEE International Symposium on Information Theory Pro-*
578 *ceedings (ISIT)*, pp. 2168–2172, St. Petersburg, Russia, 2011.
579
- 580 Thomas J. Richardson and Rüdiger L. Urbanke. The capacity of low-density parity-check codes
581 under message-passing decoding. *IEEE Trans. Inf. Theory*, 47(2):599–618, 2001.
- 582 Tom Richardson and Ruediger Urbanke. *Modern coding theory*. Cambridge university press, 2008.
583
- 584 Guilhem Semerjian, Gabriele Sicuro, and Lenka Zdeborová. Recovery thresholds in the sparse
585 planted matching problem. *Physical Review E*, 102(2):022304, 2020.
586
- 587 Martin Slawski and Emanuel Ben-David. Linear regression with sparsely permuted data. *Electronic*
588 *Journal of Statistics*, 1:1–36, 2019.
- 589 Martin Slawski, Emanuel Ben-David, and Ping Li. Two-stage approach to multivariate linear regres-
590 sion with sparsely mismatched data. *J. Mach. Learn. Res.*, 21(204):1–42, 2020.
591
- 592 Pragya Sur, Yuxin Chen, and Emmanuel J Candès. The likelihood ratio test in high-dimensional
593 logistic regression is asymptotically a rescaled chi-square. *Probability Theory and Related Fields*,
175(1):487–558, 2019.

594 Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54.
595 Springer Science & Business Media, 2010.
596

597 Jayakrishnan Unnikrishnan, Saeid Haghighatshoar, and Martin Vetterli. Unlabeled sensing: Solv-
598 ing a linear system with unordered measurements. In *Proceedings of the 53rd Annual Allerton*
599 *Conference on Communication, Control, and Computing (Allerton)*, pp. 786–793, Monticello, IL,
600 2015.

601 Hang Zhang and Ping Li. Optimal estimator for unlabeled linear regression. In *Proceedings of the*
602 *37th International Conference on Machine Learning (ICML)*, pp. 11153–11162, Virtual Event,
603 2020.

604 Hang Zhang, Martin Slawski, and Ping Li. The benefits of diversity: Permutation recovery in
605 unlabeled sensing from multiple measurement vectors. *IEEE Trans. Inf. Theory*, 68(4):2509–
606 2529, 2022.
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647

A NOTATIONS

We repeat the notations for the self-containing of the appendix. We denote $a \xrightarrow{\text{a.s.}} b$ when a converges almost surely to b . $f(n) \simeq g(n)$ denotes $\lim_{n \rightarrow \infty} f(n)/g(n) = 1$, and $f(n) = \mathcal{O}_P(g(n))$ if the sequence $f(n)/g(n)$ is bounded in probability, and $f(n) = o_P(g(n))$ if $f(n)/g(n)$ converges to zero in probability. The inner product between two vectors (resp. matrices) are denoted as $\langle \cdot, \cdot \rangle$. For two distributions d_1 and d_2 , we write $d_1 \cong d_2$ if they are equal up to normalization. Moreover, \mathcal{P}_n corresponds to the set of all possible permutation matrices: $\mathcal{P}_n \triangleq \{\mathbf{\Pi} \in \{0, 1\}^{n \times n}, \sum_i \mathbf{\Pi}_{ij} = 1, \sum_j \mathbf{\Pi}_{ij} = 1\}$. The *signal-to-noise-ratio* is $\text{snr} = \frac{\|\mathbf{B}^s\|_F^2}{m \cdot \sigma^2}$, where $\|\cdot\|_F$ is the Frobenius norm and σ^2 is the variance of the sensing noise.

B LINK BETWEEN PERMUTATION RECOVERY WITH MESSAGE PASSING

B.1 CONSTRUCTION OF THE GRAPHICAL MODEL

First, we construct the factor graph associated with the probability measure in (3). Adopting the same strategy as in Chapter 16 of Mezard & Montanari (2009), we conduct the following operations, e.g., 1) associating each variable $\mathbf{\Pi}_{ij}$ a variable node v_{ij} ; 2) associating the variable node v_{ij} a function node representing the term $e^{-\beta \mathbf{\Pi}_{ij} \mathbf{E}_{ij}}$; and 3) linking each constraint $\sum_i \mathbf{\Pi}_{ij} = 1$ to a function node and similarly for the constraint $\sum_j \mathbf{\Pi}_{ij} = 1$. A graphical representation is available in Figure 2.

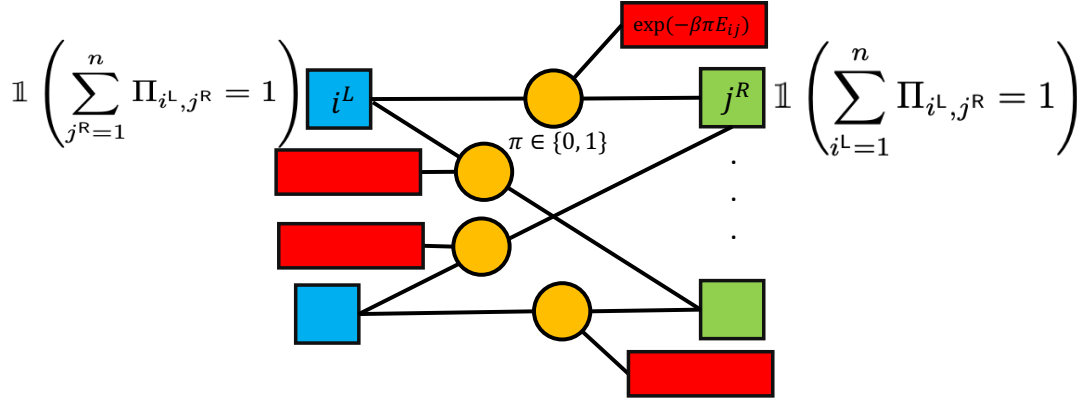


Figure 2: The constructed graphical model. Circle icons denote the variable nodes and square icons denote the function nodes: blue squares (green squares resp.) for the constraints on the rows (columns resp.) of $\mathbf{\Pi}$, and red squares for the function $e^{-\beta \pi \mathbf{E}_{ij}}$.

Now we briefly review the MP algorithm. Informally speaking, MP is a local algorithm to compute the marginal probabilities over the graphical model. In each iteration, the variable node v transmits the message to its incident function node f by multiplying all incoming messages except the message along the edge (v, f) . The function node f transmits the message to its incident variable node v by computing the weighted summary of all incoming messages except the message along the edge (f, v) . For a detailed introduction to MP, we refer readers to Kschischang et al. (2001), Chapter 16 in MacKay et al. (2003), and Chapter 14 in Mezard & Montanari (2009).

It is known that MP can obtain the exact marginals (Mezard & Montanari, 2009) for singly connected graphical models. For other types of graphs, however, whether MP can obtain the exact solution still remains an open problem (Cantwell & Newman, 2019; Kirkley et al., 2021). At the same time, numerical evidences have been witnessed to show that MP can yield meaningful results for graphs with loops; particular examples include applications in the coding theory (Chung, 2000; Richardson & Urbanke, 2001; 2008) and the LAP (which happens to be our case) (Mezard & Montanari, 2009; Chertkov et al., 2010; Caracciolo et al., 2017; Malatesta et al., 2019; Semerjian et al., 2020).

B.2 THE MESSAGE PASSING (MP) ALGORITHM

Next, we perform permutation recovery via MP. The following derivation follows the standard procedure, which can be found in the previous works (Mezard & Montanari, 2009; Semerjian et al., 2020). We denote the message flow from the node i^L to the variable node (i^L, j^R) as $\widehat{m}_{i^L \rightarrow (i^L, j^R)}(\cdot)$ and that from the edge (i^L, j^R) to node i^L as $m_{(i^L, j^R) \rightarrow i^L}(\cdot)$. Similarly, we define $\widehat{m}_{j^R \rightarrow (i^L, j^R)}(\cdot)$ and $m_{(i^L, j^R) \rightarrow j^R}(\cdot)$ as the message flow transmitted between the functional node j^R and the variable node (i^L, j^R) . Here the superscripts L and R are used to indicate the positions of the node (left and right, respectively). Roughly speaking, these transmitted messages can be viewed as (unnormalized) conditional probability $\mathbb{P}(\Pi_{i,j} = \{0, 1\} | \cdot)$ with the joint PDF being defined in (3). The message transmission process is to iteratively compute these conditional probabilities.

First, we consider the message flows transmitted between the functional node i^L and the variable node (i^L, j^R) , which are written as

$$\begin{aligned} m_{(i^L, j^R) \rightarrow i^L}(\pi) &\cong \widehat{m}_{j^R \rightarrow (i^L, j^R)}(\pi) e^{-\beta \pi \mathbf{E}_{i^L, j^R}}, \\ \widehat{m}_{i^L \rightarrow (i^L, j^R)}(\pi) &\cong \sum_{\pi_{i^L, k^R}} \prod_{k^R \neq j^R} \widehat{m}_{k^R \rightarrow (i^L, k^R)}(\pi_{i^L, k^R}) \times e^{-\beta \pi_{i^L, k^R} \mathbf{E}_{i^L, k^R}} \mathbb{1}(\pi + \sum_k \pi_{i^L, k^R} = 1), \end{aligned} \quad (18)$$

where $\pi \in \{0, 1\}$ is a binary value. Similarly, we can write the message flows between the functional node j^R and the variable node (i^L, j^R) , which are denoted as $m_{(i^L, j^R) \rightarrow j^R}(\pi)$ and $\widehat{m}_{j^R \rightarrow (i^L, j^R)}(\pi)$, respectively. With the parametrization approach, we define

$$h_{i^L \rightarrow (i^L, j^R)} \triangleq \frac{1}{\beta} \log \frac{\widehat{m}_{i^L \rightarrow (i^L, j^R)}(1)}{\widehat{m}_{i^L \rightarrow (i^L, j^R)}(0)}, \quad h_{j^R \rightarrow (i^L, j^R)} \triangleq \frac{1}{\beta} \log \frac{\widehat{m}_{j^R \rightarrow (i^L, j^R)}(1)}{\widehat{m}_{j^R \rightarrow (i^L, j^R)}(0)}.$$

Following the routine derivations in MP, we get the edge selection criteria, i.e., we pick $\widehat{\pi}(i^L) = j^R$ if

$$h_{i^L \rightarrow (i^L, j^R)} + h_{j^R \rightarrow (i^L, j^R)} > \mathbf{E}_{i^L, j^R}; \quad (19)$$

otherwise, we have $\widehat{\pi}(i^L) \neq j^R$. Due to the fact that $\mu(\mathbf{\Pi})$ concentrates on $\widehat{\mathbf{\Pi}}$ when β is sufficiently large, we can thus rewrite the MP update equation as

$$h_{i^L \rightarrow (i^L, j^R)} = \min_{k^R \neq j^R} \mathbf{E}_{i^L, k^R} - h_{k^R \rightarrow (i^L, k^R)}, \quad h_{j^R \rightarrow (i^L, j^R)} = \min_{k^L \neq i^L} \mathbf{E}_{k^L, j^R} - h_{k^L \rightarrow (k^L, j^R)}, \quad (20)$$

which is attained by letting $\beta \rightarrow \infty$.

C ANALYSIS OF ORACLE CASE: PROOF OF LEMMA 1

Proof. Denote the singular values of \mathbf{B}^\natural as $\{\lambda_i\}_{i=1}^{\text{rank}(\mathbf{B}^\natural)}$. We exploit the rotation invariance property of Gaussian random variables; and have Ξ be identically distributed as

$$\Xi = \sum_{i=1}^{\text{rank}(\mathbf{B}^\natural)} \lambda_i^2 x_i (x_i - y_i) + \sigma \sum_{i=1}^{\text{rank}(\mathbf{B}^\natural)} \lambda_i w_i (x_i - y_i).$$

Due to the independence across w , x , and y , we have

$$\begin{aligned} \mathbb{E} e^{-\theta \Xi} &= \prod_{i=1}^{\text{rank}(\mathbf{B}^\natural)} \mathbb{E}_{x,y,w} \exp[-\theta \lambda_i^2 x (x - y) - \theta \sigma \lambda_i w (x - y)] \\ &= \prod_{i=1}^{\text{rank}(\mathbf{B}^\natural)} \mathbb{E}_{x,y} \exp\left(\frac{\theta \lambda_i^2 (x - y) (\theta \sigma^2 (x - y) - 2x)}{2}\right) \\ &\stackrel{\textcircled{1}}{=} \prod_{i=1}^{\text{rank}(\mathbf{B}^\natural)} \mathbb{E}_x \frac{\exp\left(\frac{\theta \lambda_i^2 x^2 (\theta (\lambda_i^2 + \sigma^2) - 2)}{2 - 2\theta^2 \lambda_i^2 \sigma^2}\right)}{\sqrt{1 - \theta^2 \lambda_i^2 \sigma^2}} \end{aligned}$$

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

$$\stackrel{\textcircled{2}}{=} \prod_{i=1}^{\text{rank}(\mathbf{B}^\natural)} (1 + 2\theta\lambda_i^2 - \theta^2\lambda_i^2 (\lambda_i^2 + 2\sigma^2))^{-\frac{1}{2}},$$

where in ① we use the fact $\theta^2\sigma^2\lambda_i^2 < 1$ and in ② we use the fact $\theta^2\lambda_i^2 (\lambda_i^2 + 2\sigma^2) \leq 1 + 2\theta\lambda_i^2$. \square

D ANALYSIS OF THE NON-ORACLE CASE

This section presents the technical details in analyzing the non-oracle case.

D.1 ADDITIONAL NUMERICAL RESULTS

We consider the same settings as in Subsection 4.1. Here, we present additional numerical results to evaluate the prediction accuracy of our method.

D.1.1 VERIFICATION OF PHASE TRANSITION POINTS

For the predicted phase transition $\text{snr}_{\text{non-oracle}}$, we notice an increasing gap between the predicted value and the simulated value, unlike in the oracle case. This might be caused by the strong correlation across the edge weights $\{\mathbf{E}_{ij}\}_{1 \leq i, j \leq n}$, or due to the error with the approximation relation $\mathbb{E}e^{-\theta\Xi} \approx \mathbb{E} \exp(\theta\mathbb{E}\Xi - \theta^2\text{Var}\Xi/2)$.

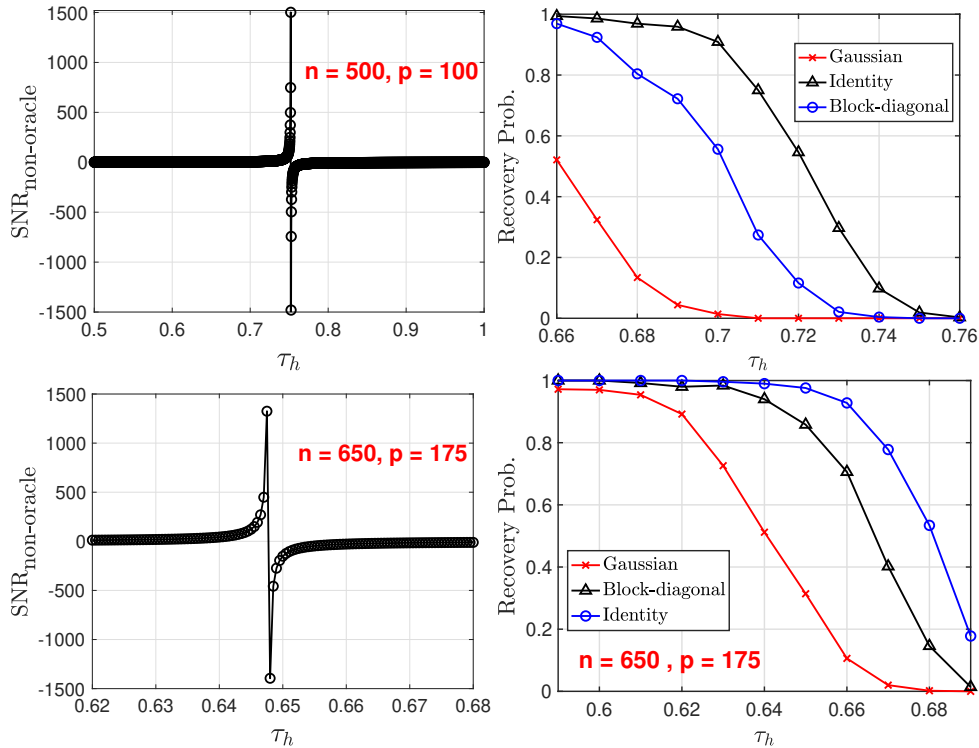


Figure 3: **(Left panel)** Predicted $\text{snr}_{\text{non-oracle}}$. **(Right panel)** Plot of recovery rate under the noiseless setting, i.e., $\text{snr} = \infty$. **Gaussian:** $\mathbf{B}_{ij}^\natural \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$; **Identity:** $\mathbf{B}^\natural = \mathbf{I}_{p \times p}$; **Block-diagonal:** $\mathbf{B}^\natural = \text{diag}\{1, \dots, 1, 0.5, \dots, 0.5\}$.

Additional experiments are available in Table 3, from which we conclude the solution (17) can predict the phase transition point w.r.t. τ_h to a good extent.

D.1.2 IMPACT OF n ON THE PHASE TRANSITION POINT

We study the impact of n on τ_h . The numerical experiment is shown in the top row of Figure 3, from which we can see the predicted phase transition τ_h matches to a good extent to the numerical

Table 3: Comparison between the predicted value of the phase transition threshold τ_h and its simulated value when $n = 500$. **P** denotes the **predicted value** while **S** denotes the **simulated value**. **S** corresponds to the τ_h when the error rate drops below 0.05. We adopt a similar algorithm as in Algorithm 1. The only difference is we replace (2) in Line 5 with (15).

p	75	100	125	150	175	200
P	0.82	0.73	0.68	0.62	0.56	0.52
S	0.77	0.75	0.7	0.66	0.61	0.57

experiments. Then, we fix the p and study the impact of n on τ_h . We observe that the phase transition τ_h increases together with the sample number n , which is also captured by our formula in (17).

D.1.3 LIMITS OF τ_h

We consider the limiting behavior of τ_h when τ_p approaches 0, or equivalently, $p = o_p(n)$. We can simplify $\mathbb{E}\Xi$ and $\text{Var}\Xi$ in Theorem 3 as

$$\begin{aligned}\mathbb{E}\Xi &\simeq n(1 - \tau_h) \|\mathbf{B}^\natural\|_{\text{F}}^2, \\ \text{Var}\Xi &\simeq 3n^2(1 - \tau_h)^2 \|\mathbf{B}^\natural\|_{\text{F}}^2.\end{aligned}$$

We notice that the singularity point in (17) disappears. In other words, we can have the correct permutation matrix $\mathbf{\Pi}^\natural$ even when $h \approx n$. This is (partly) verified by Figure 4, from which we observe that the phase transition point w.r.t. τ_h approaches to one, or equivalently, h approaches n , as τ_p decreases to zero.

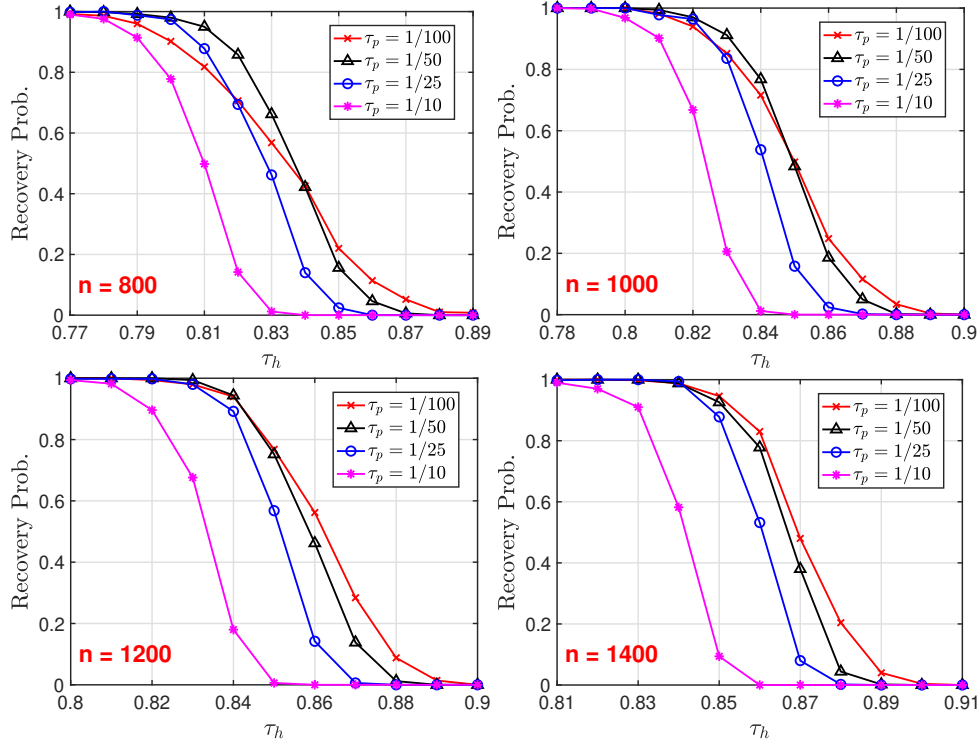


Figure 4: Plot of correct recovery rate w.r.t. τ_h . We consider the noiseless scenario (i.e., $\text{snr} = \infty$) and pick $n = \{800, 1000, 1200, 1400\}$.

D.2 ANALYSIS OF NON-ORACLE CASE: PROOF OF THEOREM 3

This subsection presents the computational details of Theorem 3. To begin with, we decompose the random variable Ξ as

$$\Xi = \Xi_1 + \sigma(\Xi_2 + \Xi_3) + \sigma^2\Xi_4, \quad (21)$$

where Ξ_i ($1 \leq i \leq 4$) are respectively defined as

$$\begin{aligned} \Xi_1 &\triangleq \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{B}^{\natural} \mathbf{B}^{\natural\top} \mathbf{X}^{\top} \mathbf{\Pi}^{\natural\top} \mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j), \\ \Xi_2 &\triangleq \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{B}^{\natural} \mathbf{W}^{\top} \mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j), \\ \Xi_3 &\triangleq \mathbf{W}_i^{\top} \mathbf{B}^{\natural\top} \mathbf{X}^{\top} \mathbf{\Pi}^{\natural\top} \mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j), \\ \Xi_4 &\triangleq \mathbf{W}_i^{\top} \mathbf{W}^{\top} \mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j). \end{aligned}$$

Unlike the oracle case, obtaining a closed-form expression of $\mathbb{E}e^{-\theta\Xi}$ would be too difficult. Hence, we adopt the Gaussian approximation method as presented in Section 3.3. The task then transforms to computing the expectation and variance of Ξ with the computation details as follows.

D.2.1 NOTATIONS

Note that our analysis can involve the terms containing $(\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j)$ and $\mathbf{X}^{\top} \mathbf{\Pi}^{\natural} \mathbf{X}$ simultaneously. To decouple the dependence between $(\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j)$ and $\mathbf{X}^{\top} \mathbf{\Pi}^{\natural} \mathbf{X}$, we first rewrite the matrix $\mathbf{X}^{\top} \mathbf{\Pi}^{\natural} \mathbf{X}$ as the sum $\sum_{\ell} \mathbf{X}_{\ell} \mathbf{X}_{\pi^{\natural}(\ell)}^{\top}$ and then collect all terms $\mathbf{X}_{\ell} \mathbf{X}_{\pi^{\natural}(\ell)}^{\top}$ independent of $\mathbf{X}_{\pi^{\natural}(i)}$ and \mathbf{X}_j in the matrix Σ , which is written as

$$\Sigma \triangleq \sum_{\ell, \pi^{\natural}(\ell) \neq \pi^{\natural}(i), j} \mathbf{X}_{\ell} \mathbf{X}_{\pi^{\natural}(\ell)}^{\top}. \quad (22)$$

The rest terms are then put in the matrix Δ such that $\mathbf{X}^{\top} \mathbf{\Pi}^{\natural} \mathbf{X} = \Sigma + \Delta$. Note that the expression of Δ varies under different cases such that

- **Case (s, s) :** $i = \pi^{\natural}(i)$ and $j = \pi^{\natural}(j)$. We have

$$\Delta = \Delta^{(s,s)} = \mathbf{X}_i \mathbf{X}_i^{\top} + \mathbf{X}_j \mathbf{X}_j^{\top}. \quad (23)$$

- **Case (s, d) :** $i = \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. We have

$$\Delta = \Delta^{(s,d)} = \mathbf{X}_i \mathbf{X}_i^{\top} + \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^{\top} + \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^{\top}. \quad (24)$$

- **Case (d, s) :** $i \neq \pi^{\natural}(i)$ and $j = \pi^{\natural}(j)$. We have

$$\Delta = \Delta^{(d,s)} = \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} + \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^{\top} + \mathbf{X}_j \mathbf{X}_j^{\top}. \quad (25)$$

- **Case (d, d) :** $i \neq \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. We have

$$\Delta = \Delta^{(d,d)} = \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} + \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^{\top} + \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^{\top} + \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^{\top}. \quad (26)$$

In addition, we define the matrix \mathbf{M} as $\mathbf{B}^{\natural} \mathbf{B}^{\natural\top}$, and define the index sets \mathcal{S} , \mathcal{D} , and $\mathcal{D}_{\text{pair}}$ as

$$\mathcal{S} \triangleq \{\ell \mid \ell \neq i \text{ or } j, \ell = \pi^{\natural}(\ell)\}, \quad (27)$$

$$\mathcal{D} \triangleq \{\ell \mid \ell, \pi^{\natural}(\ell) \neq i \text{ or } j, \ell \neq \pi^{\natural}(\ell)\}, \quad (28)$$

$$\mathcal{D}_{\text{pair}} \triangleq \{(\ell_1, \ell_2) : \ell_1 = \pi^{\natural}(\ell_2), \ell_2 = \pi^{\natural}(\ell_1), \ell_1, \ell_2 \in \mathcal{D}\}, \quad (29)$$

respectively.

918 D.2.2 MAIN COMPUTATION

919 In this case, we can write Ξ as

$$\begin{aligned}
920 \Xi &= \underbrace{\mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{B}^{\natural} \mathbf{B}^{\natural\top} \mathbf{X}^{\top} \mathbf{\Pi}^{\natural\top} \mathbf{X} [\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j]}_{\triangleq \Xi_1} + \underbrace{\sigma \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{B}^{\natural} \mathbf{W}^{\top} \mathbf{X} [\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j]}_{\triangleq \Xi_2} \\
921 &+ \underbrace{\sigma \mathbf{W}_i^{\top} \mathbf{B}^{\natural\top} \mathbf{X}^{\top} \mathbf{\Pi}^{\natural\top} \mathbf{X} [\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j]}_{\triangleq \Xi_3} + \underbrace{\sigma^2 \mathbf{W}_i^{\top} \mathbf{W}^{\top} \mathbf{X} [\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j]}_{\triangleq \Xi_4}.
\end{aligned}$$

922 The following context separately computes its expectation $\mathbb{E}\Xi$ and its variance $\text{Var}\Xi$.

923 **Expectation.** We can easily verify that both $\mathbb{E}\Xi_2$ and $\mathbb{E}\Xi_3$ are zero. Then our goal turns to calculating the expectation of $\mathbb{E}\Xi_1$ and $\mathbb{E}\Xi_4$. First, we have

$$\mathbb{E}\Xi_1 = \mathbb{E} \sum_{\ell=\pi^{\natural}(\ell)} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\ell} \mathbf{X}_{\ell}^{\top} \mathbf{X}_{\pi^{\natural}(i)} - \mathbb{E} \sum_{\ell} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(\ell)} \mathbf{X}_{\ell}^{\top} \mathbf{X}_j.$$

924 With Lemma 16 and Lemma 17, we conclude

$$\begin{aligned}
925 \mathbb{E}\Xi_1 &= (n-h) \text{Tr}(\mathbf{M}) + (p+1) \mathbb{E} \mathbb{1}_{i=\pi^{\natural}(i)} \text{Tr}(\mathbf{M}) - (p \mathbb{E} \mathbb{1}_{i=j} + \mathbb{E} \mathbb{1}_{j=\pi^{\natural}(i)}) \text{Tr}(\mathbf{M}) \\
926 &= (n+p-h-hp/n) [1+o(1)] \text{Tr}(\mathbf{M}).
\end{aligned} \tag{30}$$

927 Meanwhile, we have

$$\begin{aligned}
928 \mathbb{E}\Xi_4 &= \mathbb{E} [\mathbf{W}_i^{\top} \mathbf{W}_1 \cdots \mathbf{W}_i^{\top} \mathbf{W}_i \cdots \mathbf{W}_i^{\top} \mathbf{W}_n] \mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j) \\
929 &= m \mathbb{E} \mathbf{X}_i^{\top} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j) = mp (\mathbb{E} \mathbb{1}_{i=\pi^{\natural}(i)} - \mathbb{E} \mathbb{1}_{i=j}) = \frac{mp(n-h)\sigma^2}{n} [1+o(1)].
\end{aligned} \tag{31}$$

930 Combining (30) and (31) and neglecting the $o(1)$ terms yields

$$\mathbb{E}\Xi \approx (n+p)(1-h/n) \|\mathbf{B}^{\natural}\|_{\text{F}}^2 + \frac{mp(n-h)\sigma^2}{n}.$$

931 **Variance.** Then we study the variance of Ξ . With the relation $\text{Var}(\Xi) = \mathbb{E}\Xi^2 - (\mathbb{E}\Xi)^2$, our goal reduces to computing $\mathbb{E}\Xi^2$, which can be written as

$$\mathbb{E}\Xi^2 = \mathbb{E}\Xi_1^2 + \sigma^2 \mathbb{E}\Xi_2^2 + \sigma^2 \mathbb{E}\Xi_3^2 + \sigma^4 \mathbb{E}\Xi_4^2 + 2\sigma^2 \mathbb{E}\Xi_1 \Xi_4 + 2\sigma^2 \mathbb{E}\Xi_2 \Xi_3.$$

932 The following context separately computes each terms

$$\begin{aligned}
933 \mathbb{E}\Xi_1^2 &\approx (n-h)^2 \left(1 + \frac{2p}{n} + \frac{p^2}{n(n-h)} \right) [\text{Tr}(\mathbf{M})]^2 \\
934 &+ n^2 \left[\frac{2p}{n} + 3 \left(1 - \frac{h}{n} \right)^2 + \frac{6(n-h)^2 p}{n^3} + \frac{(3n-h)p^2}{n^3} \right] \text{Tr}(\mathbf{M}\mathbf{M}), \\
935 \mathbb{E}\Xi_2^2 &\approx 2np(1+p/n) \text{Tr}(\mathbf{M}), \\
936 \mathbb{E}\Xi_3^2 &\approx 2n^2 \left(\frac{p}{n} + \left(1 - \frac{h}{n} \right)^2 + \frac{p^2}{n^2} + \frac{4p(n-h)^2}{n^3} \right) \text{Tr}(\mathbf{M}), \\
937 \mathbb{E}\Xi_4^2 &\approx \frac{(n-h)m^2 p^2}{n}, \\
938 \mathbb{E}\Xi_1 \Xi_4 &\approx \frac{mp(n-h)(n+p-h)}{n} \text{Tr}(\mathbf{M}), \\
939 \mathbb{E}\Xi_2 \Xi_3 &\approx \frac{p(n-h)(n+p-h)}{n} \text{Tr}(\mathbf{M}).
\end{aligned}$$

940 The detailed computation is attached as follows.

972 **Lemma 2.** *We have*

$$973 \mathbb{E}\Xi_1^2 = (n-h)^2 \left(1 + \frac{2p}{n} + \frac{p^2}{n(n-h)} + o(1) \right) [\text{Tr}(\mathbf{M})]^2$$

$$974 + n^2 \left[\frac{2p}{n} + 3 \left(1 - \frac{h}{n} \right)^2 + \frac{6(n-h)^2 p}{n^3} + \frac{(3n-h)p^2}{n^3} + o(1) \right] \text{Tr}(\mathbf{M}\mathbf{M}),$$

975 where Ξ_1 is defined in (21).

976 *Proof.* We begin the proof by decomposing Ξ_1^2 as

$$977 \mathbb{E}\Xi_1^2 = \underbrace{\mathbb{E} (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \boldsymbol{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \boldsymbol{\Sigma}^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)}_{\Lambda_1}$$

$$978 + 2 \underbrace{\mathbb{E} (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \boldsymbol{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \boldsymbol{\Delta}^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)}_{\Lambda_2}$$

$$979 + \underbrace{\mathbb{E} (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \boldsymbol{\Delta} \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \boldsymbol{\Delta}^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)}_{\Lambda_3},$$

980 and separately bound each term as in Lemma 3, Lemma 4, and Lemma 5.

□

981 **Lemma 3.** *We have*

$$982 \mathbb{E} (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \boldsymbol{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \boldsymbol{\Sigma}^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)$$

$$983 = (n-h)^2 (1 + o(1)) [\text{Tr}(\mathbf{M})]^2 + n^2 \left[\frac{2p}{n} + 3 \left(1 - \frac{h}{n} \right)^2 + o(1) \right] \text{Tr}(\mathbf{M}\mathbf{M}).$$

984 *Proof.* Due to the independence among different rows of the sensing matrix \mathbf{X} , we condition on $\boldsymbol{\Sigma}$ and take expectation w.r.t. $\mathbf{X}_{\pi^h(i)}$ and \mathbf{X}_j , which leads to

$$985 \mathbb{E}\Lambda_1 = \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \boldsymbol{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \boldsymbol{\Sigma}^\top \mathbf{X}_{\pi^h(i)}}_{\Lambda_{1,1}} + \underbrace{\mathbb{E} \mathbf{X}_j^\top \boldsymbol{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \boldsymbol{\Sigma}^\top \mathbf{X}_j}_{\Lambda_{1,2}}.$$

986 For $\mathbb{E}\Lambda_{1,1}$, we obtain

$$987 \mathbb{E}\Lambda_{1,1} \stackrel{\textcircled{1}}{=} \mathbb{E} [\text{Tr}(\boldsymbol{\Sigma} \mathbf{M}) \text{Tr}(\boldsymbol{\Sigma} \mathbf{M})] + \mathbb{E} \text{Tr}(\boldsymbol{\Sigma} \mathbf{M} \mathbf{M}^\top \boldsymbol{\Sigma}^\top) + \mathbb{E} \text{Tr}(\boldsymbol{\Sigma} \mathbf{M} \boldsymbol{\Sigma} \mathbf{M})$$

$$988 \stackrel{\textcircled{2}}{=} (n-h)^2 [1 + o(1)] [\text{Tr}(\mathbf{M})]^2 + n^2 \left[\frac{p}{n} + 2 \left(1 - \frac{h}{n} \right)^2 + o(1) \right] \text{Tr}(\mathbf{M}\mathbf{M}),$$

989 where $\textcircled{1}$ is due to (65), and $\textcircled{2}$ is due to Lemma 13, Lemma 14, and Lemma 15. As for $\mathbb{E}\Lambda_{1,2}$, we have

$$990 \mathbb{E}\Lambda_{1,2} = \mathbb{E} \text{Tr}(\boldsymbol{\Sigma} \mathbf{M} \mathbf{M}^\top \boldsymbol{\Sigma}^\top) = n^2 \left[\frac{p}{n} + \left(1 - \frac{h}{n} \right)^2 + o(1) \right] \text{Tr}(\mathbf{M}^\top \mathbf{M}),$$

991 and hence complete the proof. □

992 **Lemma 4.** *We have*

$$993 \mathbb{E} (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \boldsymbol{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \boldsymbol{\Delta}^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j) \approx \frac{(n-h)^2 p}{n} \left[(\text{Tr}(\mathbf{M}))^2 + 3 \text{Tr}(\mathbf{M}\mathbf{M}) \right].$$

1026 *Proof.* Similar as above, we first expand Λ_2 as

$$1027 \mathbb{E}\Lambda_2 = (n-h)\mathbb{E}\underbrace{\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \Delta^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\Lambda_{2,1}} + (n-h)\mathbb{E}\underbrace{\mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \Delta^\top \mathbf{X}_j}_{\Lambda_{2,2}}$$

$$1028 - (n-h)\mathbb{E}\underbrace{\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \Delta^\top \mathbf{X}_j}_{\Lambda_{2,3}} - (n-h)\mathbb{E}\underbrace{\mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \Delta^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\Lambda_{2,4}}.$$

1034 **Case (s, s) :** $i = \pi^{\natural}(i)$ and $j = \pi^{\natural}(j)$. We first compute $\Lambda_{2,1}$ as

$$1035 \mathbb{E}\Lambda_{2,1} = \underbrace{\mathbb{E}\mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i}_{\mathbb{E}\|\mathbf{X}_i\|_2^2 (\mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i)^2} + \underbrace{\mathbb{E}\mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_i}_{\mathbb{E}(\mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i)^2}$$

$$1036 = (p+5) \left[(\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M}\mathbf{M}) + \text{Tr}(\mathbf{M}^\top \mathbf{M}) \right].$$

1041 We consider $\Lambda_{2,2}$ as

$$1042 \mathbb{E}\Lambda_{2,2} = \underbrace{\mathbb{E}(\mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i)}_{\mathbb{E}(\mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i)^2} + \underbrace{\mathbb{E}(\mathbf{X}_j^\top \mathbf{M} \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j)}_{\mathbb{E}\|\mathbf{X}_j\|_2^2 \mathbf{X}_j^\top \mathbf{M} \mathbf{M} \mathbf{X}_j}$$

$$1043 = (\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M}\mathbf{M}) + \text{Tr}(\mathbf{M}^\top \mathbf{M}) + (p+2) \text{Tr}(\mathbf{M}\mathbf{M}).$$

1048 As for $\Lambda_{2,3}$ and $\Lambda_{2,4}$, we can verify that they are both zero, which gives

$$1049 \mathbb{E}\Lambda_2 = (n-h)p[1+o(1)] [\text{Tr}(\mathbf{M})]^2 + 3(n-h)p[1+o(1)] \text{Tr}(\mathbf{M}\mathbf{M}). \quad (32)$$

1052 **Case (s, d) :** $i = \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. We can compute $\Lambda_{2,1}$ as

$$1053 \mathbb{E}\Lambda_{2,1} = \mathbb{E}\mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i + \underbrace{\mathbb{E}\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)}}_0$$

$$1054 + \underbrace{\mathbb{E}\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^\top \mathbf{X}_{\pi^{\natural}(i)}}_0$$

$$1055 = (p+4) \left[(\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M}\mathbf{M}) + \text{Tr}(\mathbf{M}^\top \mathbf{M}) \right].$$

1061 We consider $\Lambda_{2,2}$ as

$$1062 \mathbb{E}\Lambda_{2,2} = \mathbb{E}\mathbf{X}_j^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_j + \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top \mathbf{X}_j}_0$$

$$1063 + \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^\top \mathbf{X}_j}_0$$

$$1064 = \mathbb{E}(\mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i)^2 = (\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M}\mathbf{M}) + \text{Tr}(\mathbf{M}^\top \mathbf{M}).$$

1070 Similarly, we can verify that both $\mathbb{E}\Lambda_{2,3}$ and $\mathbb{E}\Lambda_{2,4}$ are zero and hence have

$$1071 \mathbb{E}\Lambda_2 = (n-h)p[1+o(1)] [\text{Tr}(\mathbf{M})]^2 + 2(n-h)p[1+o(1)] \text{Tr}(\mathbf{M}\mathbf{M}). \quad (33)$$

1074 **Case (d, s) :** $i \neq \pi^{\natural}(i)$ and $j = \pi^{\natural}(j)$. We compute $\Lambda_{2,1}$ as

$$1075 \mathbb{E}\Lambda_{2,1} = \underbrace{\mathbb{E}\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)}}_0 + \underbrace{\mathbb{E}\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)}}_0$$

$$1076 + \underbrace{\mathbb{E}\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{E}\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}} = (\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M}\mathbf{M}) + \text{Tr}(\mathbf{M}^\top \mathbf{M}).$$

1080
1081
1082
1083
1084
1085
1086
1087
1088
1089

We consider $\Lambda_{2,2}$ as

$$\begin{aligned} \mathbb{E}\Lambda_{2,2} &= \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_j}_{0} + \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^{h2}(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j}_{0} \\ &+ \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j}_{\mathbb{E}\|\mathbf{X}_j\|_2^2 \mathbf{X}_j^\top \mathbf{M}\mathbf{M}\mathbf{X}_j} = (p+2) \text{Tr}(\mathbf{M}\mathbf{M}). \end{aligned}$$

As for $\mathbb{E}\Lambda_{2,3}$ and $\Lambda_{2,4}$, we can follow the same strategy and prove they are both zero, which yields

1091
1092

$$\mathbb{E}\Lambda_2 = (n-h) [\text{Tr}(\mathbf{M})]^2 + (n-h)p [1 + o(1)] \text{Tr}(\mathbf{M}\mathbf{M}). \quad (34)$$

1093
1094
1095

Case (d, d) : $i \neq \pi^h(i)$ and $j \neq \pi^h(j)$. Contrary to the previous cases, we have $\mathbb{E}\Lambda_{2,1}$ and $\mathbb{E}\Lambda_{2,2}$ to be zero in this case rather than $\mathbb{E}\Lambda_{2,3}$ and $\mathbb{E}\Lambda_{2,4}$.

1096
1097

Hence our focus turns to the calculation of $\mathbb{E}\Lambda_{2,3}$ and that of $\mathbb{E}\Lambda_{2,4}$. For $\Lambda_{2,3}$, we have

1098
1099
1100
1101
1102
1103
1104
1105

$$\begin{aligned} \mathbb{E}\Lambda_{2,3} &= \underbrace{\mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_j}_{p \mathbb{1}_{i=j} \mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)}} + \underbrace{\mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^{h2}(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j}_{\mathbb{1}_{j=\pi^{h2}(i)} \mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)}} \\ &+ \underbrace{\mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(j)} \mathbf{X}_j^\top \mathbf{X}_j}_{p \mathbb{1}_{i=j} \mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)}} + \underbrace{\mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_j \mathbf{X}_{\pi^{h-1}(j)}^\top \mathbf{X}_j}_{\mathbb{1}_{j=\pi^{h2}(i)} \mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)}} \\ &= 2(p \mathbb{1}_{i=j} + \mathbb{1}_{j=\pi^{h2}(i)}) \left[(\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M}\mathbf{M}) + \text{Tr}(\mathbf{M}^\top \mathbf{M}) \right]. \end{aligned}$$

1106
1107

Then we turn to the calculation of $\mathbb{E}\Lambda_{2,4}$, which proceeds as

1108
1109
1110
1111
1112
1113
1114
1115
1116

$$\begin{aligned} \mathbb{E}\Lambda_{2,4} &= \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)}}_{\mathbb{1}(i=j) \mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)}} + \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^{h2}(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^h(i)}}_{\mathbb{1}_{j=\pi^{h2}(i)} \mathbb{E}\|\mathbf{X}_{\pi^h(i)}\|_2^2 \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{M}\mathbf{X}_{\pi^h(i)}} \\ &+ \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)}}_{\mathbb{1}(i=j) \mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_{\pi^h(i)}} + \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{X}_j \mathbf{X}_{\pi^{h-1}(j)}^\top \mathbf{X}_{\pi^h(i)}}_{\mathbb{1}_{j=\pi^{h2}(i)} \mathbb{E}\|\mathbf{X}_{\pi^h(i)}\|_2^2 \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\mathbf{M}\mathbf{X}_{\pi^h(i)}} \\ &= 2 \mathbb{1}_{j=\pi^{h2}(i)} (p+2) \text{Tr}(\mathbf{M}\mathbf{M}) + 2 \mathbb{1}(i=j) \left[(\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M}\mathbf{M}) + \text{Tr}(\mathbf{M}^\top \mathbf{M}) \right]. \end{aligned}$$

1117
1118

Then we conclude

1119
1120
1121

$$\begin{aligned} \mathbb{E}\Lambda_2 &= -2(n-h) \left[(p+1) \mathbb{1}_{i=j} + \mathbb{1}_{j=\pi^{h2}(i)} \right] \left[(\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M}\mathbf{M}) + \text{Tr}(\mathbf{M}^\top \mathbf{M}) \right] \\ &- 2(n-h)(p+2) \mathbb{1}_{j=\pi^{h2}(i)} \text{Tr}(\mathbf{M}\mathbf{M}). \end{aligned} \quad (35)$$

1122
1123

The proof is thus completed by combining (32), (33), (34), and (35). \square

1124
1125

Lemma 5. *We have*

1126
1127
1128
1129

$$\begin{aligned} &\mathbb{E}(\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \Delta \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\Delta^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j) \\ &= \left(1 - \frac{h}{n} + o(1)\right) p^2 [\text{Tr}(\mathbf{M})]^2 + \left(3 - \frac{h}{n} + o(1)\right) p^2 \text{Tr}(\mathbf{M}\mathbf{M}). \end{aligned}$$

1130
1131

Proof. We begin the proof by expanding Λ_3 as

1132
1133

$$\mathbb{E}\Lambda_3 = \underbrace{\mathbb{E}\mathbf{X}_{\pi^h(i)}^\top \Delta \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\Delta^\top \mathbf{X}_{\pi^h(i)}}_{\Lambda_{3,1}} + \underbrace{\mathbb{E}\mathbf{X}_j^\top \Delta \mathbf{M}\mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M}\Delta^\top \mathbf{X}_j}_{\Lambda_{3,2}}$$

$$- \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \Delta \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \Delta^{\top} \mathbf{X}_j}_{\Lambda_{3,3}} - \underbrace{\mathbb{E} \mathbf{X}_j^{\top} \Delta \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \Delta^{\top} \mathbf{X}_{\pi^{\natural}(i)}}_{\Lambda_{3,4}}.$$

Case (s, s): $i = \pi^{\natural}(i)$ and $j = \pi^{\natural}(j)$. First we compute $\Lambda_{3,1}$ as

$$\begin{aligned} \mathbb{E} \Lambda_{3,1} &= \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{X}_i}_{\mathbb{E} \|\mathbf{X}_i\|_2^4 (\mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i)^2} + \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_j^{\top} \mathbf{X}_i}_{\mathbb{E} \|\mathbf{X}_i\|_2^2 (\mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i)^2} \\ &+ \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_j \mathbf{X}_j^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{X}_i}_{\mathbb{E} \|\mathbf{X}_i\|_2^2 (\mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i)^2} + \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_j \mathbf{X}_j^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_j^{\top} \mathbf{X}_i}_{\mathbb{E} (\mathbf{X}_j^{\top} \mathbf{X}_i)^2 \mathbf{X}_j^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_j} \\ &= (p+4)(p+8) \left[(\text{Tr}(\mathbf{M}))^2 + 2 \text{Tr}(\mathbf{M}\mathbf{M}) \right] + 2 (\text{Tr}(\mathbf{M}))^2 + (p+6) \text{Tr}(\mathbf{M}\mathbf{M}). \end{aligned}$$

Then, we consider $\Lambda_{3,2}$ as

$$\begin{aligned} \mathbb{E} \Lambda_{3,2} &= \underbrace{\mathbb{E} \mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{X}_j}_{\mathbb{E} \|\mathbf{X}_i\|_2^2 (\mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i)^2} + \underbrace{\mathbb{E} \mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_j^{\top} \mathbf{X}_j}_{(p+2) \mathbb{E} (\mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i)^2} \\ &+ \underbrace{\mathbb{E} \mathbf{X}_j^{\top} \mathbf{X}_j \mathbf{X}_j^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{X}_j}_{(p+2) \mathbb{E} (\mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i)^2} + \underbrace{\mathbb{E} \mathbf{X}_j^{\top} \mathbf{X}_j \mathbf{X}_j^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_j^{\top} \mathbf{X}_j}_{\mathbb{E} \|\mathbf{X}_j\|_2^4 \mathbf{X}_j^{\top} \mathbf{M} \mathbf{M} \mathbf{X}_j} \\ &= (3p+8) \left[(\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M}\mathbf{M}) + \text{Tr}(\mathbf{M}^{\top} \mathbf{M}) \right] + (p+2)(p+4) \text{Tr}(\mathbf{M}\mathbf{M}). \end{aligned}$$

In addition, we can verify that $\mathbb{E} \Lambda_{3,3}$ and $\mathbb{E} \Lambda_{3,4}$ are both zero. Hence we conclude

$$\mathbb{E} \Lambda_3 = (p^2 + 15p + 42) [\text{Tr}(\mathbf{M})]^2 + (3p^2 + 37p + 94) \text{Tr}(\mathbf{M}\mathbf{M}). \quad (36)$$

Case (s, d): $i = \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. We can compute $\Lambda_{3,1}$ as

$$\begin{aligned} \mathbb{E} \Lambda_{3,1} &= \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{X}_i}_{\mathbb{E} \|\mathbf{X}_i\|_2^4 (\mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i)^2} + \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^{\top} \mathbf{X}_i}_0 \\ &+ \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^{\top} \mathbf{X}_i}_0 + \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{X}_i}_0 \\ &+ \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^{\top} \mathbf{X}_i}_{\mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^{\top} \mathbf{M} \mathbf{M} \mathbf{X}_i} + \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^{\top} \mathbf{X}_i}_{\mathbb{1}_{j=\pi^{\natural}2(j)} \mathbb{E} (\mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i)^2} \\ &+ \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_{\pi^{\natural}-1(j)} \mathbf{X}_j^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{X}_i}_0 + \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_{\pi^{\natural}-1(j)} \mathbf{X}_j^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^{\top} \mathbf{X}_i}_{\mathbb{1}_{j=\pi^{\natural}2(j)} \mathbb{E} (\mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i)^2} \\ &+ \underbrace{\mathbb{E} \mathbf{X}_i^{\top} \mathbf{X}_{\pi^{\natural}-1(j)} \mathbf{X}_j^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^{\top} \mathbf{X}_i}_{\mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^{\top} \mathbf{M} \mathbf{M} \mathbf{X}_i} \\ &= (p^2 + 10p + 24 + 2 \mathbb{1}_{j=\pi^{\natural}2(j)}) \left[(\text{Tr}(\mathbf{M}))^2 + 2 \text{Tr}(\mathbf{M}\mathbf{M}) \right] + 2(p+2) \text{Tr}(\mathbf{M}\mathbf{M}). \end{aligned}$$

We consider $\Lambda_{3,2}$ as

$$\begin{aligned} \mathbb{E} \Lambda_{3,2} &= \underbrace{\mathbb{E} \mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{X}_j}_{\mathbb{E} \|\mathbf{X}_i\|_2^2 (\mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i)^2} + \underbrace{\mathbb{E} \mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^{\top} \mathbf{X}_j}_0 \\ &+ \underbrace{\mathbb{E} \mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^{\top} \mathbf{X}_j}_0 \\ &+ \underbrace{\mathbb{E} \mathbf{X}_j^{\top} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{X}_j}_0 + \underbrace{\mathbb{E} \mathbf{X}_j^{\top} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^{\top} \mathbf{X}_j}_{\mathbb{E} \|\mathbf{X}_i\|_2^4 \text{Tr}(\mathbf{M}\mathbf{M})} \end{aligned}$$

$$\begin{aligned}
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^\top \mathbf{X}_j}_{\mathbb{1}_{j=\pi^{\natural}2(j)} \mathbb{E} \|\mathbf{X}_j\|_2^2 \mathbf{X}_j^\top \mathbf{M} \mathbf{M} \mathbf{X}_j} + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}-1(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}-1(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top \mathbf{X}_j}_{\mathbb{1}_{j=\pi^{\natural}2(j)} \mathbb{E} \|\mathbf{X}_j\|_2^2 \mathbf{X}_j^\top \mathbf{M} \mathbf{M} \mathbf{X}_j} + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}-1(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^\top \mathbf{X}_j}_{\mathbb{E} \|\mathbf{X}_j\|_2^2 \mathbf{X}_j^\top \mathbf{M} \mathbf{M} \mathbf{X}_j} \\
& = (p+4) \left[(\text{Tr}(\mathbf{M}))^2 + 2 \text{Tr}(\mathbf{M}\mathbf{M}) \right] + (p+2) (p+1 + 2\mathbb{1}_{j=\pi^{\natural}2(j)}) \text{Tr}(\mathbf{M}\mathbf{M}).
\end{aligned}$$

As for $\Lambda_{3,3}$ and $\Lambda_{3,4}$, we can prove that they are both zero in this case. Hence we conclude

$$\begin{aligned}
\mathbb{E} \Lambda_3 & = (p^2 + 11p + 28 + 2\mathbb{1}_{j=\pi^{\natural}2(j)}) (\text{Tr}(\mathbf{M}))^2 \\
& + [3p^2 + 27p + 62 + 2(p+4) \mathbb{1}_{j=\pi^{\natural}2(j)}] \text{Tr}(\mathbf{M}\mathbf{M}). \tag{37}
\end{aligned}$$

Case (d, s) : $i \neq \pi^{\natural}(i)$ and $j = \pi^{\natural}(j)$. We consider the term $\Lambda_{3,1}$ as

$$\begin{aligned}
\mathbb{E} \Lambda_{3,1} & = \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^2 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}2(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=\pi^{\natural}2(i)} \mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^2 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}2(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=\pi^{\natural}2(i)} \mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^2 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}2(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}2(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{M} \mathbf{X}_i} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}2(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}2(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{E} (\mathbf{X}_i^\top \mathbf{X}_j)^2 \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_j} \\
& = (p+6) (\text{Tr}(\mathbf{M}))^2 + (p^2 + 9p + 22) \text{Tr}(\mathbf{M}\mathbf{M}) + 2\mathbb{1}_{i=\pi^{\natural}2(i)} (p+4) \left[(\text{Tr}(\mathbf{M}))^2 + 2 \text{Tr}(\mathbf{M}\mathbf{M}) \right].
\end{aligned}$$

We consider the term $\Lambda_{3,2}$ as

$$\begin{aligned}
\mathbb{E} \Lambda_{3,2} & = \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_j}_{p \times \mathbb{E} (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}2(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j}_{\mathbb{1}_{i=\pi^{\natural}2(i)} \mathbb{E} (\mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i)^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j}_0
\end{aligned}$$

$$\begin{aligned}
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_j}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} \mathbb{E} (\mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i)^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j}_{\mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{M} \mathbf{X}_i} \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j}_{\mathbb{E} \|\mathbf{X}_j\|_F^4 \times \mathbf{X}_j^\top \mathbf{M} \mathbf{M} \mathbf{X}_j} \\
& = p (\text{Tr}(\mathbf{M}))^2 + (p^2 + 9p + 10) \text{Tr}(\mathbf{M}\mathbf{M}) + 2 \mathbb{1}_{i=\pi^{\natural 2}(i)} \left[(\text{Tr}(\mathbf{M}))^2 + 2 \text{Tr}(\mathbf{M}\mathbf{M}) \right].
\end{aligned}$$

As for $\Lambda_{3,3}$ and $\Lambda_{3,4}$, easily we can verify they are both zero and hence

$$\begin{aligned}
\mathbb{E} \Lambda_3 & = 2(p+3) (\text{Tr}(\mathbf{M}))^2 + 2(p^2 + 9p + 16) \text{Tr}(\mathbf{M}\mathbf{M}) \\
& + 2 \mathbb{1}_{i=\pi^{\natural 2}(i)} (p+5) \left[(\text{Tr}(\mathbf{M}))^2 + 2 \text{Tr}(\mathbf{M}\mathbf{M}) \right]. \tag{38}
\end{aligned}$$

Case (d, d): $i \neq \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. First, We compute $\mathbb{E} \Lambda_{3,1}$ as

$$\begin{aligned}
\mathbb{E} \Lambda_{3,1} & = \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^2 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} \mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^2 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=j} \mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^2 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{j=i} \mathbb{1}_{j=\pi^{\natural 2}(i)} \mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^2 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} \mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^2 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^4 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{j=i} \mathbb{1}_{j=\pi^{\natural 2}(i)} \mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^2 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{j=\pi^{\natural 2}(i)} \mathbb{E} \|\mathbf{X}_{\pi^{\natural}(i)}\|_F^4 (\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)})}
\end{aligned}$$

$$\begin{aligned}
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{h-1}(j)}^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^{h+2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^{h+2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{h+2}(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^{h+2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(j)} \mathbf{X}_j^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^{h+2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{h-1}(j)}^\top \mathbf{X}_j}_{\mathbb{1}_{j=\pi^{h+3}(i)} \mathbb{E} \left\| \mathbf{X}_{\pi^h(i)} \right\|_{\text{F}}^2 \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{M} \mathbf{X}_{\pi^h(i)}} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_j}_{\mathbb{1}_{i=\pi^h(j)} \mathbb{E} \left[\mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \right]^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{h+2}(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(j)} \mathbf{X}_j^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{h-1}(j)}^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^{h-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^{h-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{h+2}(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j}_{\mathbb{1}_{j=\pi^{h+3}(i)} \mathbb{E} \left[\mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \right]^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^{h-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(j)} \mathbf{X}_j^\top \mathbf{X}_j}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^{h-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{h-1}(j)}^\top \mathbf{X}_j}_0 \\
& = \left[(p+1) \mathbb{1}_{i=\pi^h(j)} + \mathbb{1}_{j=\pi^{h+3}(i)} \right] [\text{Tr}(\mathbf{M})]^2 \\
& + \left[2(p+1) \mathbb{1}_{i=\pi^h(j)} + (p+4) \mathbb{1}_{j=\pi^{h+3}(i)} \right] \text{Tr}(\mathbf{M}\mathbf{M}). \tag{41}
\end{aligned}$$

Then, we consider the term $\mathbb{E}\Lambda_{3,4}$, which can be written as

$$\begin{aligned}
\mathbb{E}\Lambda_{3,4} & = \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{h+2}(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^h(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)}}_{\mathbb{1}_{i=\pi^h(j)} \mathbb{E} \left[\mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \right]^2} \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{h-1}(j)}^\top \mathbf{X}_{\pi^h(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^{h+2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)}}_0
\end{aligned}$$

$$\begin{aligned}
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}2}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\mathfrak{h}-1}(j)}^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& \quad \mathbb{1}_{j=\pi^{\mathfrak{h}3}(i)} \mathbb{E} \left[\mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \right]^2 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^{\mathfrak{h}}(j)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& \quad \mathbb{1}_{i=\pi^{\mathfrak{h}}(j)} p \mathbb{E} \left[\mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \right]^2 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^{\mathfrak{h}}(j)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}2}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^{\mathfrak{h}}(j)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^{\mathfrak{h}}(j)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\mathfrak{h}-1}(j)}^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\mathfrak{h}-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\mathfrak{h}-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}2}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& \quad \mathbb{1}_{j=\pi^{\mathfrak{h}3}(i)} \mathbb{E} \left\| \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \right\|_{\text{F}}^2 \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\mathfrak{h}-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\mathfrak{h}-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\mathfrak{h}-1}(j)}^\top \mathbf{X}_{\pi^{\mathfrak{h}}(i)}}_0 \\
& = \left[(p+1) \mathbb{1}_{i=\pi^{\mathfrak{h}}(j)} + \mathbb{1}_{j=\pi^{\mathfrak{h}3}(i)} \right] [\text{Tr}(\mathbf{M})]^2 \\
& \quad + \left[2(p+1) \mathbb{1}_{i=\pi^{\mathfrak{h}}(j)} + (p+4) \mathbb{1}_{j=\pi^{\mathfrak{h}3}(i)} \right] \text{Tr}(\mathbf{M}\mathbf{M}). \tag{42}
\end{aligned}$$

Combing (39), (40), (41), and (42) together then yields

$$\mathbb{E} \Lambda_3 = \frac{4hp(p+2)}{n^2} [1 + o(1)] [\text{Tr}(\mathbf{M})]^2 + 2p^2 [1 + o(1)] \text{Tr}(\mathbf{M}\mathbf{M}). \tag{43}$$

The proof is thus completed by summarizing the computations thereof. \square

Lemma 6. *We have*

$$\mathbb{E} \Xi_2^2 = 2[(p+2)(p+3) + (n-2)(p+1)] \left\| \mathbf{B}^{\mathfrak{h}} \right\|_{\text{F}}^2 = 2np(1 + p/n + o(1)) \left\| \mathbf{B}^{\mathfrak{h}} \right\|_{\text{F}}^2,$$

where Ξ_2 is defined in (21).

Proof. We have

$$\begin{aligned}
\mathbb{E} \Xi_2^2 &= \mathbb{E} \left[\mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{B}^{\mathfrak{h}} \mathbf{W}^\top \mathbf{X} (\mathbf{X}_{\pi^{\mathfrak{h}}(i)} - \mathbf{X}_j) (\mathbf{X}_{\pi^{\mathfrak{h}}(i)} - \mathbf{X}_j)^\top \mathbf{X}^\top \mathbf{W} \mathbf{B}^{\mathfrak{h}\top} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \right] \\
&= \mathbb{E} \left[\mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{B}^{\mathfrak{h}} \text{Tr} \left[\mathbf{X} (\mathbf{X}_{\pi^{\mathfrak{h}}(i)} - \mathbf{X}_j) (\mathbf{X}_{\pi^{\mathfrak{h}}(i)} - \mathbf{X}_j)^\top \mathbf{X}^\top \right] \mathbf{B}^{\mathfrak{h}\top} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \right] \\
&= \mathbb{E} \left[\left\| \mathbf{X} (\mathbf{X}_{\pi^{\mathfrak{h}}(i)} - \mathbf{X}_j) \right\|_2^2 \mathbf{X}_{\pi^{\mathfrak{h}}(i)}^\top \mathbf{B}^{\mathfrak{h}} \mathbf{B}^{\mathfrak{h}\top} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \right] \\
&= \mathbb{E} \left[\left\| \mathbf{X} (\mathbf{X}_{\pi^{\mathfrak{h}}(i)} - \mathbf{X}_j) \right\|_2^2 \times \left\| \mathbf{B}^{\mathfrak{h}\top} \mathbf{X}_{\pi^{\mathfrak{h}}(i)} \right\|_2^2 \right].
\end{aligned}$$

For the conciseness of notation, we assume $\pi^{\natural}(i) = 1$ and $j = 2$ w.l.o.g. Decomposing the term $\|\mathbf{X}(\mathbf{X}_1 - \mathbf{X}_2)\|_2^2$ as

$$\|\mathbf{X}(\mathbf{X}_1 - \mathbf{X}_2)\|_F^2 = \underbrace{[\mathbf{X}_1^\top (\mathbf{X}_1 - \mathbf{X}_2)]^2}_{\mathcal{T}_1} + \underbrace{[\mathbf{X}_2^\top (\mathbf{X}_1 - \mathbf{X}_2)]^2}_{\mathcal{T}_2} + \underbrace{\sum_{i=3}^n [\mathbf{X}_i^\top (\mathbf{X}_1 - \mathbf{X}_2)]^2}_{\mathcal{T}_3},$$

we then separately bound the above three terms. For the first term $\mathbb{E}\mathcal{T}_1 \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2$, we have

$$\begin{aligned} \mathbb{E}\mathcal{T}_1 \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2 &= \mathbb{E} \left[\left(\|\mathbf{X}_1\|_2^4 + (\mathbf{X}_1^\top \mathbf{X}_2)^2 \right) \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2 \right] \\ &= \underbrace{\mathbb{E} \|\mathbf{X}_1\|_2^4 \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2}_{(p+2)(p+4)\|\mathbf{B}^{\natural}\|_F^2} + \underbrace{\mathbb{E} (\mathbf{X}_1^\top \mathbf{X}_2)^2 \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2}_{(p+2)\|\mathbf{B}^{\natural}\|_F^2} \stackrel{\textcircled{1}}{=} (p+2)(p+5)\|\mathbf{B}^{\natural}\|_F^2, \end{aligned} \quad (44)$$

where $\textcircled{1}$ is due to (66) and (67).

Similarly, for term $\mathbb{E}\mathcal{T}_2 \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2$, we invoke (66) and (67), which gives

$$\begin{aligned} \mathbb{E}\mathcal{T}_2 \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2 &= \underbrace{\mathbb{E} \left[(\mathbf{X}_1^\top \mathbf{X}_2)^2 \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2 \right]}_{(p+2)\|\mathbf{B}^{\natural}\|_F^2} + \underbrace{\mathbb{E} \|\mathbf{X}_2\|_2^4 \times \mathbb{E} \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2}_{p(p+2)\|\mathbf{B}^{\natural}\|_F^2} \\ &\stackrel{\textcircled{2}}{=} (p+1)(p+2)\|\mathbf{B}^{\natural}\|_F^2, \end{aligned} \quad (45)$$

where $\textcircled{2}$ is due to (66).

For the last term $\mathbb{E}\mathcal{T}_3 \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2$, we exploit the independence among the rows of matrix \mathbf{X} and have

$$\begin{aligned} \mathbb{E}\mathcal{T}_3 \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2 &= \sum_{i \geq 3} \mathbb{E} \left[(\mathbf{X}_i^\top (\mathbf{X}_1 - \mathbf{X}_2))^2 \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2 \right] \\ &= \sum_{i \geq 3} \mathbb{E} \left[\|\mathbf{X}_1 - \mathbf{X}_2\|_2^2 \cdot \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2 \right] \\ &= \sum_{i \geq 3} \mathbb{E} \left[\left(\|\mathbf{X}_1\|_2^2 + \|\mathbf{X}_2\|_2^2 \right) \cdot \|\mathbf{B}^{\natural\top} \mathbf{X}_1\|_F^2 \right] \\ &= 2 \sum_{i \geq 3} (p+1) \|\mathbf{B}^{\natural}\|_F^2 = 2(n-2)(p+1) \|\mathbf{B}^{\natural}\|_F^2. \end{aligned} \quad (47)$$

The proof is then completed by combining (44), (46), and (47). \square

Lemma 7. *We have*

$$\mathbb{E}\Xi_3^2 = 2n^2 \left[\frac{p}{n} + \left(1 - \frac{h}{n}\right)^2 + \frac{p^2}{n^2} + \frac{4p(n-h)^2}{n^3} + o(1) \right] \text{Tr}(\mathbf{M}),$$

where Ξ_3 is defined in (21).

Proof. To begin with, we decompose the term $\mathbb{E}\Xi_3^2$ as

$$\begin{aligned} \mathbb{E}\Xi_3^2 &= \mathbb{E} \left[\underbrace{(\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j)^\top \Sigma \mathbf{M} \Sigma^\top (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j)}_{\triangleq \Lambda_1} \right] + 2\mathbb{E} \left[\underbrace{(\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j)^\top \Sigma \mathbf{M} \Delta^\top (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j)}_{\triangleq \Lambda_2} \right] \\ &\quad + \mathbb{E} \left[\underbrace{(\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j)^\top \Delta \mathbf{M} \Delta^\top (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j)}_{\triangleq \Lambda_3} \right]. \end{aligned} \quad (48)$$

1566

1567

1568

Step I. First we consider $\mathbb{E}\Lambda_1$, which can be written as

1569

1570

1571

1572

$$\mathbb{E}\Lambda_1 = 2\mathbb{E} \text{Tr} \left(\mathbf{\Sigma} \mathbf{M} \mathbf{\Sigma}^\top \right) \stackrel{\textcircled{1}}{=} 2n^2 \left[\frac{p}{n} + \left(1 - \frac{h}{n} \right)^2 + o(1) \right] \text{Tr}(\mathbf{M}), \quad (49)$$

1573

where $\textcircled{1}$ is due to Lemma 13.

1574

1575

Step II. Then we turn to $\mathbb{E}\Lambda_2$, which can be written as

1576

1577

1578

1579

1580

1581

1582

$$\begin{aligned} \mathbb{E}\Lambda_2 = & (n-h)\mathbb{E} \underbrace{\left[\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{\Delta}^\top \mathbf{X}_{\pi^{\natural}(i)} \right]}_{\Lambda_{2,1}} + (n-h)\mathbb{E} \underbrace{\left[\mathbf{X}_j^\top \mathbf{M} \mathbf{\Delta}^\top \mathbf{X}_j \right]}_{\Lambda_{2,2}} \\ & - (n-h)\mathbb{E} \underbrace{\left[\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{\Delta}^\top \mathbf{X}_j \right]}_{\Lambda_{2,3}} - (n-h)\mathbb{E} \underbrace{\left[\mathbf{X}_j^\top \mathbf{M} \mathbf{\Delta}^\top \mathbf{X}_{\pi^{\natural}(i)} \right]}_{\Lambda_{2,4}}. \end{aligned}$$

1583

1584

Case (s, s) : $i = \pi^{\natural}(i)$ and $j = \pi^{\natural}(j)$. We have

1585

1586

1587

$$\begin{aligned} \mathbb{E}\Lambda_{2,1} &= \mathbb{E} \mathbf{X}_i^\top \mathbf{M} (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top) \mathbf{X}_i = (p+3) \text{Tr}(\mathbf{M}), \\ \mathbb{E}\Lambda_{2,2} &= \mathbb{E} \mathbf{X}_j^\top \mathbf{M} (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top) \mathbf{X}_j = (p+3) \text{Tr}(\mathbf{M}). \end{aligned}$$

1588

In addition, we can verify that $\mathbb{E}\Lambda_{2,2}$ and $\Lambda_{2,3}$ are both zero, which suggests that

1589

1590

$$\mathbb{E}\Lambda_2 = 2(n-h)(p+3) \text{Tr}(\mathbf{M}). \quad (50)$$

1591

1592

Case (s, d) : $i = \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. We have

1593

1594

1595

1596

$$\begin{aligned} \mathbb{E}\Lambda_{2,1} &= \mathbb{E} \mathbf{X}_i^\top \mathbf{M} (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top + \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top) \mathbf{X}_i = (p+2) \text{Tr}(\mathbf{M}); \\ \mathbb{E}\Lambda_{2,2} &= \mathbb{E} \mathbf{X}_j^\top \mathbf{M} (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top + \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top) \mathbf{X}_j = \text{Tr}(\mathbf{M}). \end{aligned}$$

1597

Moreover, we have both $\mathbb{E}\Lambda_{2,3}$ and $\mathbb{E}\Lambda_{2,4}$ be zero, which suggests that

1598

1599

$$\mathbb{E}\Lambda_2 = (n-h)(p+3) \text{Tr}(\mathbf{M}). \quad (51)$$

1600

1601

1602

Case (d, s) : $i \neq \pi^{\natural}(i)$ and $j = \pi^{\natural}(j)$. We have

1603

1604

1605

$$\begin{aligned} \mathbb{E}\Lambda_{2,1} &= \mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \right) = \text{Tr}(\mathbf{M}), \\ \mathbb{E}\Lambda_{2,2} &= \mathbb{E} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j = (p+2) \text{Tr}(\mathbf{M}). \end{aligned}$$

1606

Similar as above, we can verify both $\mathbb{E}\Lambda_{2,3}$ and $\mathbb{E}\Lambda_{2,4}$ are zero, which suggests that

1607

1608

$$\mathbb{E}\Lambda_2 = (n-h)(p+3) \text{Tr}(\mathbf{M}). \quad (52)$$

1609

1610

1611

Case (d, d) : $i \neq \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. Different from the above three cases, we have $\mathbb{E}\Lambda_{2,1}$ and $\mathbb{E}\Lambda_{2,2}$ be zero and focus on the calculation of $\mathbb{E}\Lambda_{2,3}$ and $\mathbb{E}\Lambda_{2,4}$, which proceeds as

1612

1613

1614

1615

1616

1617

1618

1619

$$\begin{aligned} \mathbb{E}\Lambda_{2,3} &= \underbrace{\mathbb{E} \left[\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_j \right]}_{p \mathbb{1}_{i=j} \|\mathbf{B}^{\natural}\|_{\text{F}}^2} + \underbrace{\mathbb{E} \left[\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j \right]}_{\mathbb{1}_{j=\pi^{\natural 2}(i)} \|\mathbf{B}^{\natural}\|_{\text{F}}^2} \\ &+ \underbrace{\mathbb{E} \left[\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top \mathbf{X}_j \right]}_{p \mathbb{1}_{i=j} \|\mathbf{B}^{\natural}\|_{\text{F}}^2} + \underbrace{\mathbb{E} \left[\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural-1}(j)}^\top \mathbf{X}_j \right]}_{\mathbb{1}_{\pi^{\natural}(i)=\pi^{\natural-1}(j)} \|\mathbf{B}^{\natural}\|_{\text{F}}^2} \\ &= 2 \left[p \mathbb{1}_{i=j} + \mathbb{1}_{j=\pi^{\natural 2}(i)} \right] \text{Tr}(\mathbf{M}); \end{aligned}$$

$$\begin{aligned}
1620 \quad \mathbb{E}\Lambda_{2,4} &= \underbrace{\mathbb{E} \left[\mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \right]}_{\mathbb{1}_{i=j} \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2} + \underbrace{\mathbb{E} \left[\mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \right]}_{p \mathbb{1}_{j=\pi^{\natural 2}(i)} \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2} \\
1621 & \\
1622 & \\
1623 & \\
1624 &+ \underbrace{\mathbb{E} \left[\mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \right]}_{\mathbb{1}_{i=j} \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2} + \underbrace{\mathbb{E} \left[\mathbf{X}_j^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural-1}(j)}^\top \mathbf{X}_{\pi^{\natural}(i)} \right]}_{p \mathbb{1}_{j=\pi^{\natural 2}(i)} \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2} \\
1625 & \\
1626 & \\
1627 &= 2 \left[p \mathbb{1}_{j=\pi^{\natural 2}(i)} + \mathbb{1}_{i=j} \right] \text{Tr}(\mathbf{M}), \\
1628 & \\
1629 &
\end{aligned}$$

which suggests that

$$\mathbb{E}\Lambda_2 = -2(n-h)(p+1) (\mathbb{1}_{j=\pi^{\natural 2}(i)} + \mathbb{1}_{i=j}) \text{Tr}(\mathbf{M}). \quad (53)$$

Combing (50), (51), (52), and (53), we conclude

$$\mathbb{E}\Lambda_2 = \frac{2p(n-h)^2}{n} \text{Tr}(\mathbf{M}) [1 + o(1)]. \quad (54)$$

Step III. Then we turn to the calculation of $\mathbb{E}\Lambda_3$. First we perform the following decomposition

$$\begin{aligned}
1637 \quad \Lambda_3 &= \underbrace{\mathbf{X}_{\pi^{\natural}(i)}^\top \Delta \mathbf{M} \Delta^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\Lambda_{3,1}} + \underbrace{\mathbf{X}_j^\top \Delta \mathbf{M} \Delta^\top \mathbf{X}_j}_{\Lambda_{3,2}} - \underbrace{\mathbf{X}_{\pi^{\natural}(i)}^\top \Delta \mathbf{M} \Delta^\top \mathbf{X}_j}_{\Lambda_{3,3}} - \underbrace{\mathbf{X}_j^\top \Delta \mathbf{M} \Delta^\top \mathbf{X}_{\pi^{\natural}(i)}}_{\Lambda_{3,4}}. \\
1638 & \\
1639 & \\
1640 &
\end{aligned}$$

Case (s, s): $i = \pi^{\natural}(i)$ and $j = \pi^{\natural}(j)$. We have

$$\begin{aligned}
1642 \quad \mathbb{E}\Lambda_{3,1} &= \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i \right)}_{\mathbb{E} \|\mathbf{X}_i\|_2^4 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} + \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_i \right)}_{(p+2) \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2} \\
1643 & \\
1644 & \\
1645 &+ \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i \right)}_{(p+2) \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2} + \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_i \right)}_{(p+2) \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2} = (p+2)(p+7) \text{Tr}(\mathbf{M}); \\
1646 & \\
1647 &
\end{aligned}$$

$$\mathbb{E}\Lambda_{3,2} = \mathbb{E} \mathbf{X}_j^\top \left(\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top \right) \mathbf{M} \left(\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top \right) \mathbf{X}_j = (p+2)(p+7) \text{Tr}(\mathbf{M}).$$

As for $\mathbb{E}\Lambda_{3,3}$ and $\mathbb{E}\Lambda_{3,4}$, easily we can verify that they are both zero and hence have

$$\mathbb{E}\Lambda_3 = 2(p+2)(p+7) \text{Tr}(\mathbf{M}) = 2p^2 \text{Tr}(\mathbf{M}) [1 + o(1)]. \quad (55)$$

Case (s, d): $i = \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. We can write $\Lambda_{3,1}$ as

$$\begin{aligned}
1653 \quad \mathbb{E}\Lambda_{3,1} &= \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i \right)}_{\mathbb{E} \|\mathbf{X}_i\|_2^4 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} + \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top \mathbf{X}_i \right)}_0 \\
1654 & \\
1655 & \\
1656 & \\
1657 & \\
1658 &+ \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural-1}(j)}^\top \mathbf{X}_i \right)}_0 \\
1659 & \\
1660 & \\
1661 &+ \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i \right)}_0 + \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top \mathbf{X}_i \right)}_{p \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2} \\
1662 & \\
1663 & \\
1664 &+ \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural-1}(j)}^\top \mathbf{X}_i \right)}_{\mathbb{1}_{j=\pi^{\natural 2}(j)} \text{Tr}(\mathbf{M})} \\
1665 & \\
1666 & \\
1667 &+ \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i \right)}_0 + \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top \mathbf{X}_i \right)}_{\mathbb{1}_{j=\pi^{\natural 2}(j)} \text{Tr}(\mathbf{M})} \\
1668 & \\
1669 & \\
1670 & \\
1671 &+ \underbrace{\mathbb{E} \left(\mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural-1}(j)}^\top \mathbf{X}_i \right)}_{p \|\mathbf{B}^{\natural}\|_{\mathbb{F}}^2} \\
1672 & \\
1673 & \\
&= (p^2 + 8p + 8 + 2 \mathbb{1}_{j=\pi^{\natural 2}(j)}) \text{Tr}(\mathbf{M}).
\end{aligned}$$

Mean $\Lambda_{3,2}$ can be written as

$$\begin{aligned}
\mathbb{E}\Lambda_{3,2} &= \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_j \right)}_{(p+2)\|\mathbf{B}^\natural\|_F^2} + \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_{\pi^\natural(j)} \mathbf{X}_j^\top \mathbf{X}_j \right)}_0 \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural-1}(j)}^\top \mathbf{X}_j \right)}_0 \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^\natural(j)}^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_j \right)}_0 + \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^\natural(j)}^\top \mathbf{M} \mathbf{X}_{\pi^\natural(j)} \mathbf{X}_j^\top \mathbf{X}_j \right)}_{\mathbb{E}\|\mathbf{X}_j\|_F^4 \text{Tr}(\mathbf{M})} \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^\natural(j)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural-1}(j)}^\top \mathbf{X}_j \right)}_{\mathbb{1}_{\pi^{\natural-1}(j)=\pi^\natural(j)}(p+2)\|\mathbf{B}^\natural\|_F^2} \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_j \right)}_0 + \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^\natural(j)} \mathbf{X}_j^\top \mathbf{X}_j \right)}_{\mathbb{1}_{\pi^{\natural-1}(j)=\pi^\natural(j)}(p+2)\|\mathbf{B}^\natural\|_F^2} \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural-1}(j)}^\top \mathbf{X}_j \right)}_{(p+2)\|\mathbf{B}^\natural\|_F^2} \\
&= (p+2) \left(p+2 + 2\mathbb{1}_{j=\pi^{\natural 2}(j)} \right) \text{Tr}(\mathbf{M}).
\end{aligned}$$

And for $\mathbb{E}\Lambda_{3,3}$ and $\mathbb{E}\Lambda_{3,4}$, easily we can verify that they are both zero. Then we conclude

$$\mathbb{E}\Lambda_3 = 2 \left(p^2 + 6p + 6 + (p+3)\mathbb{1}_{j=\pi^{\natural 2}(j)} \right) \text{Tr}(\mathbf{M}) = 2p^2 \text{Tr}(\mathbf{M}) [1 + o(1)]. \quad (56)$$

Case (d, s) : $i \neq \pi^\natural(i)$ and $j = \pi^\natural(j)$. In this case, we can write $\Lambda_{3,1}$ as

$$\begin{aligned}
\mathbb{E}\Lambda_{3,1} &= \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^\natural(i)}^\top \mathbf{M} \mathbf{X}_{\pi^\natural(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^\natural(i)} \right)}_{\mathbb{E}\|\mathbf{X}_i\|_F^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^\natural(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_{\pi^\natural(i)} \right)}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} \mathbb{E}\|\mathbf{X}_i\|_F^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^\natural(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^\natural(i)} \right)}_0 \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_{\pi^\natural(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^\natural(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^\natural(i)} \right)}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} \mathbb{E}\|\mathbf{X}_i\|_F^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_{\pi^\natural(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_{\pi^\natural(i)} \right)}_{\mathbb{E}\|\mathbf{X}_i\|_F^4 \text{Tr}(\mathbf{M})} \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_{\pi^\natural(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^\natural(i)} \right)}_0 \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^\natural(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^\natural(i)} \right)}_0 \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_{\pi^\natural(i)} \right)}_0 \\
&+ \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^\natural(i)}^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^\natural(i)} \right)}_{\mathbb{E}\|\mathbf{X}_i\|_F^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i}
\end{aligned}$$

1728

$$= (p+2)(p+2+2\mathbb{1}_{i=\pi^{\natural 2}(i)}) \text{Tr}(\mathbf{M}).$$

1729

We consider $\Lambda_{3,2}$ as

1730

1731

1732

1733

1734

1735

1736

1737

1738

1739

1740

1741

1742

1743

1744

1745

1746

1747

Similarly, as above, we can verify that $\mathbb{E}\Lambda_{3,3} = 0$ and $\mathbb{E}\Lambda_{3,4} = 0$. Hence, we can conclude

1748

$$\mathbb{E}\Lambda_3 = 2(p^2 + 6p + 6 + (p+3)\mathbb{1}_{i=\pi^{\natural 2}(i)}) \text{Tr}(\mathbf{M}) = 2p^2 \text{Tr}(\mathbf{M}) [1 + o(1)]. \quad (57)$$

1749

1750

Case (d, d): $i \neq \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. We write $\Lambda_{3,1}$ as

1751

1752

1753

1754

1755

1756

1757

1758

1759

1760

1761

1762

1763

1764

1765

1766

1767

1768

1769

1770

1771

1772

1773

1774

1775

1776

1777

1778

1779

We consider $\Lambda_{3,2}$ as

1780

1781

$$\mathbb{E}\Lambda_{3,2} = \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_j}_{p \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} \text{Tr}(\mathbf{M})}$$

$$+ \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j}_0$$

$$+ \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_j}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j}_{p \text{Tr}(\mathbf{M})}$$

$$+ \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j}_0$$

$$+ \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_j}_0 + \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j}_0 + \underbrace{\mathbb{E}\mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j}_{\mathbb{E}\|\mathbf{X}_i\|_{\mathbb{F}}^4 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i}$$

$$= (p^2 + 8p + 8 + 2\mathbb{1}_{i=\pi^{\natural 2}(i)}) \text{Tr}(\mathbf{M}).$$

$$\begin{aligned}
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)}}_{\mathbb{1}_{i=\pi^h(j)} \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^h-1(j)}^\top \mathbf{X}_{\pi^h(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h-1(j)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)}}_0 + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h-1(j)}^\top \mathbf{M} \mathbf{X}_{\pi^h-1(i)} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h-1(j)}^\top \mathbf{M} \mathbf{X}_{\pi^h(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)}}_0 + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h-1(j)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^h-1(j)}^\top \mathbf{X}_{\pi^h(i)}}_{\mathbb{1}_{j=\pi^h-1(i)} \text{Tr}(\mathbf{M})} \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)}}_{p \mathbb{1}_{i=\pi^h(j)} \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top \mathbf{M} \mathbf{X}_{\pi^h-1(i)} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top \mathbf{M} \mathbf{X}_{\pi^h(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)}}_0 + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^h-1(j)}^\top \mathbf{X}_{\pi^h(i)}}_0 \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^h-1(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^h(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)}}_0 + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^h-1(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^h-1(i)} \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top}_{p \mathbb{1}_{j=\pi^h-1(i)} \text{Tr}(\mathbf{M})} \\
& + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^h-1(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^h(j)} \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)}}_0 + \underbrace{\mathbb{E} \mathbf{X}_j^\top \mathbf{X}_{\pi^h-1(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^h-1(j)}^\top \mathbf{X}_{\pi^h(i)}}_0 \\
& = (p+1) (\mathbb{1}_{i=\pi^h(j)} + \mathbb{1}_{j=\pi^h-1(i)}) \text{Tr}(\mathbf{M}).
\end{aligned}$$

In summary, we have

$$\begin{aligned}
\mathbb{E} \Lambda_3 & = 2(p^2 + 5p + 2) \text{Tr}(\mathbf{M}) + 2(p+3) [\mathbb{1}_{i=\pi^h-1(i)} + \mathbb{1}_{j=\pi^h-1(j)}] \text{Tr}(\mathbf{M}) \\
& + (3p^2 + 7p + 6) (\mathbb{1}_{i=j} + \mathbb{1}_{j=\pi^h-1(i)}) \text{Tr}(\mathbf{M}) + \mathbb{1}_{i=j} \mathbb{1}_{i=\pi^h-1(i)} 4(3p+5) \text{Tr}(\mathbf{M}) \\
& - 2(p+1) [\mathbb{1}_{i=\pi^h(j)} + \mathbb{1}_{j=\pi^h-1(i)}] \text{Tr}(\mathbf{M}) = 2p^2 \text{Tr}(\mathbf{M}) [1 + o(1)]. \tag{58}
\end{aligned}$$

Combining (55), (56), (57), and (58) then yields

$$\mathbb{E} \Lambda_3 = 2p^2 \text{Tr}(\mathbf{M}) [1 + o(1)]. \tag{59}$$

The proof is then completed by (48), (49), (54), and (59). \square

Lemma 8. *We have*

$$\begin{aligned}
\mathbb{E} \Xi_4^2 & = m(m+1) [p(p+2) (\mathbb{1}_{i=\pi^h(i)} + \mathbb{1}_{i=j}) + p (\mathbb{1}_{i \neq \pi^h(i)} + \mathbb{1}_{i \neq j})] + 2mp(n+p+1) \\
& = \frac{(n-h)m^2 p^2}{n} [1 + o(1)],
\end{aligned}$$

where Ξ_4 is defined in (21).

Proof. For the conciseness of notation, we define Γ as $\mathbf{X} (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j) (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \mathbf{X}^\top$ and hence have

$$\mathbb{E} \Xi_4^2 = \mathbb{E} \mathbf{W}_i^\top \mathbf{W}^\top \Gamma \mathbf{W} \mathbf{W}_i.$$

We begin the discussion by expanding $\mathbf{W} \mathbf{W}_i$ as

$$\begin{bmatrix} \mathbf{W}_1^\top \\ \mathbf{W}_2^\top \\ \dots \\ \mathbf{W}_n^\top \end{bmatrix} \mathbf{W}_i = \begin{bmatrix} \mathbf{W}_1^\top \mathbf{W}_i \\ \mathbf{W}_2^\top \mathbf{W}_i \\ \dots \\ \mathbf{W}_n^\top \mathbf{W}_i \end{bmatrix}.$$

Then we obtain

$$\mathbb{E} \Xi_4^2 = \sum_{s=1}^n \sum_{t=1}^n \Gamma_{ij} \mathbb{E} [(\mathbf{W}_s^\top \mathbf{W}_i) (\mathbf{W}_t^\top \mathbf{W}_i)] = \Gamma_{ii} \mathbb{E} (\mathbf{W}_i^\top \mathbf{W}_i)^2 + \underbrace{\sum_{s \neq i} \sum_{t \neq i} \Gamma_{st} \mathbb{E} (\mathbf{W}_s^\top \mathbf{W}_i \mathbf{W}_t^\top \mathbf{W}_i)}_{\sum_{s \neq i} \Gamma_{ss} \mathbb{E} (\mathbf{W}_s^\top \mathbf{W}_i)^2}$$

$$\begin{aligned}
&= \Gamma_{ii} \mathbb{E} \left(\sum_{j=1}^m W_{ij}^2 \right)^2 + \sum_{s \neq i} \Gamma_{ss} \cdot m = m(m+1) \mathbb{E} \Gamma_{ii} + m \mathbb{E} \text{Tr}(\mathbf{\Gamma}). \tag{60}
\end{aligned}$$

We can thus complete the proof by separately computing $\mathbb{E} \text{Tr}(\mathbf{M})$ and $\mathbb{E} \Gamma_{ii}$. First we compute $\mathbb{E} \Gamma_{ii}$, which proceeds as

$$\begin{aligned}
\mathbb{E} \Gamma_{ii} &= \mathbb{E} (\mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)})^2 + \mathbb{E} (\mathbf{X}_i^\top \mathbf{X}_j)^2 \\
&= \mathbb{1}_{i=\pi^h(i)} p(p+2) + \mathbb{1}_{i \neq \pi^h(i)} p + \mathbb{1}_{i=j} p(p+2) + \mathbb{1}_{i \neq j} p \\
&= p(p+2) [\mathbb{1}_{i=\pi^h(i)} + \mathbb{1}_{i=j}] + p [\mathbb{1}_{i \neq \pi^h(i)} + \mathbb{1}_{i \neq j}]. \tag{61}
\end{aligned}$$

Then we turn to the computation of $\mathbb{E} \text{Tr}(\mathbf{M})$, which proceeds as

$$\begin{aligned}
\mathbb{E} \text{Tr}(\mathbf{\Gamma}) &= \|\mathbf{X} (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)\|_F^2 \\
&= \mathbb{E} \|\mathbf{X}_{\pi^h(i)}^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)\|_2^2 + \mathbb{E} \|\mathbf{X}_j^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)\|_2^2 + \sum_{s \neq \pi^h(i), j} \mathbb{E} \|\mathbf{X}_s^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)\|_2^2 \\
&= 2\mathbb{E} \|\mathbf{X}_{\pi^h(i)}\|_2^4 + 2\mathbb{E} (\mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j)^2 + 2 \sum_{s \neq \pi^h(i), j} \mathbb{E} \|\mathbf{X}_s\|_2^2 \\
&= 2p(p+3) + 2(n-2)p = 2p(n+p+1). \tag{62}
\end{aligned}$$

The proof is thus completed by combing (60), (62), and (61). \square

Lemma 9. *We have*

$$\mathbb{E} \Xi_1 \Xi_4 = \frac{mp(n-h)(n+p-h)}{n} [1 + o(1)] \text{Tr}(\mathbf{M}),$$

where Ξ_1 and Ξ_4 are defined in (21).

Proof. We have

$$\mathbb{E} \Xi_1 \Xi_4 = \mathbb{E} \underbrace{\mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}^\top \mathbf{\Pi}^h \mathbf{X} (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j) (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \mathbf{X}^\top \mathbf{W} \mathbf{W}_i}_{\triangleq \mathbf{v}^\top}.$$

First we conditional on \mathbf{X} . Expanding the product $\mathbf{W} \mathbf{W}_i$ as

$$\mathbb{E} \begin{bmatrix} \mathbf{W}_1^\top \mathbf{W}_i \\ \mathbf{W}_2^\top \mathbf{W}_i \\ \dots \\ \mathbf{W}_i^\top \mathbf{W}_i \\ \dots \\ \mathbf{W}_n^\top \mathbf{W}_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ m \\ \dots \\ 0 \end{bmatrix},$$

we can compute $\mathbb{E} \Xi_1 \Xi_2$ w.r.t. \mathbf{W} as

$$\mathbb{E} (\Xi_1 \Xi_2) = \mathbb{E} \mathbf{v}^\top \mathbf{W} \mathbf{W}_i = m \mathbb{E} \mathbf{v}_i,$$

where \mathbf{v}_i denotes the i -th entry of \mathbf{v} and can be written as

$$\mathbf{v}_i = \underbrace{\mathbf{X}_i^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j) (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \mathbf{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)}}_{\Lambda_1} + \underbrace{\mathbf{X}_i^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j) (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \mathbf{\Delta} \mathbf{M} \mathbf{X}_{\pi^h(i)}}_{\Lambda_2}.$$

For Λ_1 , we conclude

$$\begin{aligned}
\mathbb{E} \Lambda_1 &= \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)} + \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)} \\
&\quad - \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_j^\top \mathbf{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)} - \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_{\pi^h(i)}^\top \mathbf{\Sigma} \mathbf{M} \mathbf{X}_{\pi^h(i)} \\
&= \mathbb{1}_{i=\pi^h(i)} (p+3) \mathbb{E} \text{Tr}(\mathbf{\Sigma} \mathbf{M}) - \mathbb{1}_{i=j} (p+1) \mathbb{E} \text{Tr}(\mathbf{\Sigma} \mathbf{M})
\end{aligned}$$

$$= (n - h) (\mathbb{1}_{i=\pi^h(i)}(p + 3) - \mathbb{1}_{i=j}(p + 1)) \text{Tr}(\mathbf{M}) = \frac{p(n - h)^2}{n} [1 + o(1)] \text{Tr}(\mathbf{M}).$$

Then we turn to $\mathbb{E}\Lambda_2$ and obtain

$$\begin{aligned} \mathbb{E}\Lambda_2 = & \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \Delta \mathbf{M} \mathbf{X}_{\pi^h(i)}}_{\Lambda_{2,1}} + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \Delta \mathbf{M} \mathbf{X}_{\pi^h(i)}}_{\Lambda_{2,2}} \\ & - \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_j^\top \Delta \mathbf{M} \mathbf{X}_{\pi^h(i)}}_{\Lambda_{2,3}} - \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_{\pi^h(i)}^\top \Delta \mathbf{M} \mathbf{X}_{\pi^h(i)}}_{\Lambda_{2,4}}. \end{aligned}$$

We compute the value of $\mathbb{E}\Lambda_2$ under the four different cases.

Case (s, s) : $i = \pi^h(i)$ and $j = \pi^h(j)$. In this case, we have Δ be

$$\Delta = \Delta^{(s,s)} = \mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top.$$

We have

$$\begin{aligned} \mathbb{E}\Lambda_{2,1} &= \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top) \mathbf{M} \mathbf{X}_i = (p + 2)(p + 5) \text{Tr}(\mathbf{M}), \\ \mathbb{E}\Lambda_{2,2} &= \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top) \mathbf{M} \mathbf{X}_i = 2(p + 2) \text{Tr}(\mathbf{M}), \\ \mathbb{E}\Lambda_{2,3} &= \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_j^\top (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top) \mathbf{M} \mathbf{X}_i = 0, \\ \mathbb{E}\Lambda_{2,4} &= \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_i^\top (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top) \mathbf{M} \mathbf{X}_i = 0, \end{aligned}$$

which implies

$$\mathbb{E}\Lambda_2 = (p + 2)(p + 7) \text{Tr}(\mathbf{M}).$$

Case (s, d) : $i = \pi^h(i)$ and $j \neq \pi^h(j)$. First we write Δ as

$$\Delta^{(s,d)} = \mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top + \mathbf{X}_{\pi^{h-1}(j)} \mathbf{X}_j^\top.$$

Then we conclude

$$\begin{aligned} \mathbb{E}\Lambda_{2,1} &= \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top + \mathbf{X}_{\pi^{h-1}(j)} \mathbf{X}_j^\top) \mathbf{M} \mathbf{X}_i = (p + 2)(p + 4) \text{Tr}(\mathbf{M}), \\ \mathbb{E}\Lambda_{2,2} &= \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top + \mathbf{X}_{\pi^{h-1}(j)} \mathbf{X}_j^\top) \mathbf{M} \mathbf{X}_i = (p + 2) \text{Tr}(\mathbf{M}), \\ \mathbb{E}\Lambda_{2,3} &= \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_j^\top (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top + \mathbf{X}_{\pi^{h-1}(j)} \mathbf{X}_j^\top) \mathbf{M} \mathbf{X}_i = 0, \\ \mathbb{E}\Lambda_{2,4} &= \mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_i^\top (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_{\pi^h(j)}^\top + \mathbf{X}_{\pi^{h-1}(j)} \mathbf{X}_j^\top) \mathbf{M} \mathbf{X}_i = 0, \end{aligned}$$

which suggests that

$$\mathbb{E}\Lambda_2 = (p + 2)(p + 5) \text{Tr}(\mathbf{M}).$$

Case (d, s) : $i \neq \pi^h(i)$ and $j = \pi^h(j)$. In this case, Δ reduces to

$$\Delta^{(d,s)} = \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top + \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^{h2}(i)}^\top + \mathbf{X}_j \mathbf{X}_j^\top.$$

We have

$$\begin{aligned} \mathbb{E}\Lambda_{2,1} &= \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)}}_{\mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^{h2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)}}_{\mathbb{1}_{i=\pi^{h2}(i)} \mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} \\ &+ \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^h(i)}}_0 = (p + 2) [1 + \mathbb{1}_{i=\pi^{h2}(i)}] \text{Tr}(\mathbf{M}), \\ \mathbb{E}\Lambda_{2,2} &= \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)}}_{p \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^h(i)} \mathbf{X}_{\pi^{h2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^h(i)}}_{\mathbb{1}_{i=\pi^{h2}(i)} \text{Tr}(\mathbf{M})} \end{aligned}$$

$$+ \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_0 = (p + \mathbb{1}_{i=\pi^{\natural 2}(i)}) \text{Tr}(\mathbf{M}),$$

$$\mathbb{E} \Lambda_{2,3} = 0,$$

$$\mathbb{E} \Lambda_{2,4} = 0,$$

which suggests

$$\mathbb{E} \Lambda_2 = 2(p+1) \text{Tr}(\mathbf{M}) + \mathbb{1}_{i=\pi^{\natural 2}(i)} (p+3) \text{Tr}(\mathbf{M}).$$

Case (d, d) : $i \neq \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. In this case, Δ is written as

$$\Delta^{(d,d)} = \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top + \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top + \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top + \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top.$$

We have

$$\mathbb{E} \Lambda_{2,1} = \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} \mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i}$$

$$+ \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=j} \mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=j} \mathbb{1}_{i=\pi^{\natural 2}(i)} \mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i};$$

$$\mathbb{E} \Lambda_{2,2} = \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=j} p(p+2) \text{Tr}(\mathbf{M}) + \mathbb{1}_{i \neq j} p \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} [\mathbb{1}_{i=j} \mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i + \mathbb{1}_{i \neq j} \text{Tr}(\mathbf{M})]}$$

$$+ \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=j} p(p+2) \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=j} \mathbb{1}_{j=\pi^{\natural 2}(i)} \mathbb{E} \|\mathbf{X}_i\|_2^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i};$$

$$\mathbb{E} \Lambda_{2,3} = \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_0 + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_0$$

$$+ \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_j^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=\pi^{\natural}(j)} p \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_0;$$

$$\mathbb{E} \Lambda_{2,4} = \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_0 + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_{\pi^{\natural 2}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_0$$

$$+ \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(j)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_{\mathbb{1}_{i=\pi^{\natural}(j)} \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \mathbf{X}_i^\top \mathbf{X}_j \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural-1}(j)} \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)}}_0.$$

Hence we conclude

$$\mathbb{E} \Lambda_2 = 2(p+1) \text{Tr}(\mathbf{M}) + \mathbb{1}_{i=j} 2(p+1)^2 \text{Tr}(\mathbf{M}) + \mathbb{1}_{i=\pi^{\natural}(j)} (p+1) \text{Tr}(\mathbf{M}) \\ + \mathbb{1}_{i=\pi^{\natural 2}(i)} (p+3) \text{Tr}(\mathbf{M}) + \mathbb{1}_{i=j} \mathbb{1}_{i=\pi^{\natural 2}(i)} (3p+5) \text{Tr}(\mathbf{M}).$$

□

Lemma 10. We have

$$\mathbb{E} \Xi_2 \Xi_3 = \frac{p(n-h)(n+p-h)}{n} \text{Tr}(\mathbf{M}) [1 + o(1)],$$

where Ξ_2 and Ξ_3 are defined in (21).

Proof. To start with, we write the expectation as $\mathbb{E} \Xi_2 \Xi_3$

$$\mathbb{E} \Xi_2 \Xi_3 = \mathbb{E} \underbrace{\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{B}^\natural \mathbf{W}^\top}_{\mathbf{u}^\top} \underbrace{\mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j)}_{\mathbf{p} \in \mathbb{R}^{n \times 1}} \underbrace{\mathbf{W}_i^\top \mathbf{B}^{\natural \top} \mathbf{X}^\top \Pi^{\natural \top} \mathbf{X} (\mathbf{X}_{\pi^{\natural}(i)} - \mathbf{X}_j)}_{\mathbf{v}} \\ = \mathbb{E} \mathbf{u}^\top \mathbf{W}^\top \mathbf{p} \mathbf{W}_i^\top \mathbf{v} = \mathbb{E} \langle \mathbf{W}_i, \mathbf{u}^\top \mathbf{W}^\top \mathbf{p} \mathbf{v} \rangle.$$

Exploiting the independence among \mathbf{X} and \mathbf{W} , we condition on \mathbf{X} and have

$$\mathbb{E}_{\mathbf{W}} \langle \mathbf{W}_i, \mathbf{u}^\top \mathbf{W}^\top \mathbf{p} \mathbf{v} \rangle = \mathbb{E}_{\mathbf{W}} \text{Tr} (\nabla_{\mathbf{W}_i} \mathbf{u}^\top \mathbf{W}^\top \mathbf{p} \mathbf{v}).$$

Note that only the diagonal entries of the Hessian matrix $\nabla_{\mathbf{W}_i} \mathbf{u}^\top \mathbf{W}^\top \mathbf{p} \mathbf{v}$ matters. For an arbitrary index s , we can compute the gradient of the s -th entry of $\mathbf{u}^\top \mathbf{W}^\top \mathbf{p} \mathbf{v}$ w.r.t. $\mathbf{W}_{i,s}$ as

$$\frac{d}{dW_{i,s}} (\mathbf{u}^\top \mathbf{W}^\top \mathbf{p} \mathbf{v}_s) = \mathbf{v}_s \frac{d}{dW_{i,s}} (\mathbf{u}^\top \mathbf{W}^\top \mathbf{p}) = \mathbf{v}_s \sum_{t=1}^n \frac{d}{dW_{i,s}} (\mathbf{p}_t \mathbf{W}_t^\top \mathbf{u}) = \mathbf{v}_s \frac{d}{dW_{i,s}} \mathbf{p}_i \mathbf{W}_i^\top \mathbf{u} = \mathbf{p}_i \mathbf{v}_s \mathbf{u}_s.$$

Invoking the definitions of \mathbf{p} , \mathbf{v} and \mathbf{u} , we have

$$\begin{aligned} \mathbb{E}_{\mathbf{W}, \mathbf{X}} \langle \mathbf{W}_i, \mathbf{u}^\top \mathbf{W}^\top \mathbf{p} \mathbf{v} \rangle &= \mathbb{E}_{\mathbf{X}} (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \mathbf{X}_i \sum_{s=1}^m \left[\mathbf{X}_{\pi^h(i)}^\top (\mathbf{B}^{h\top})_s (\mathbf{B}^{h\top})_s^\top \mathbf{X}^\top \mathbf{\Pi}^{h\top} \mathbf{X} (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j) \right] \\ &\stackrel{\textcircled{1}}{=} \mathbb{E} \left[\underbrace{(\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Sigma}^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)}_{\Lambda_1} \right] + \mathbb{E} \left[\underbrace{(\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Delta}^\top (\mathbf{X}_{\pi^h(i)} - \mathbf{X}_j)}_{\Lambda_2} \right], \end{aligned}$$

where in $\textcircled{1}$ we use the relation $\sum_{s=1}^m (\mathbf{B}^{h\top})_s (\mathbf{B}^{h\top})_s^\top = \mathbf{B}^h \mathbf{B}^{h\top} = \mathbf{M}$.

For the first term Λ_1 , we obtain

$$\begin{aligned} \mathbb{E} \Lambda_1 &= \mathbb{E} \left(\underbrace{\mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Sigma}^\top \mathbf{X}_{\pi^h(i)}}_{\mathbb{1}_{i=\pi^h(i)} \mathbb{E} \|\mathbf{X}_i\|_F^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{\Sigma}^\top \mathbf{X}_i} \right) + \mathbb{E} \left(\underbrace{\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Sigma}^\top \mathbf{X}_j}_{\mathbb{1}_{i=\pi^h(i)} \mathbb{E} \mathbf{X}_i^\top \mathbf{M} \mathbf{\Sigma}^\top \mathbf{X}_i} \right) \\ &\quad - \mathbb{E} \left(\underbrace{\mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Sigma}^\top \mathbf{X}_j}_{\mathbb{1}_{i=j} \mathbb{1}_{i \neq \pi^h(i)} \mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Sigma}^\top \mathbf{X}_{\pi^h(i)}} \right) - \mathbb{E} \left(\underbrace{\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Sigma}^\top \mathbf{X}_{\pi^h(i)}}_{p \mathbb{1}_{i=j} \mathbb{1}_{i \neq \pi^h(i)} \mathbb{E} \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Sigma}^\top \mathbf{X}_{\pi^h(i)}} \right) \\ &= (n-h) \left[\mathbb{1}_{i=\pi^h(i)} (p+3) - (p+1) \mathbb{1}_{i=j} \mathbb{1}_{i \neq \pi^h(i)} \right] \text{Tr}(\mathbf{M}). \end{aligned}$$

Then we consider the second term Λ_2 , which can be decomposed further into four sub-terms reading as

$$\begin{aligned} \mathbb{E} \Lambda_2 &= \mathbb{E} \left(\underbrace{\mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Delta}^\top \mathbf{X}_{\pi^h(i)}}_{\Lambda_{2,1}} \right) + \mathbb{E} \left(\underbrace{\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Delta}^\top \mathbf{X}_j}_{\Lambda_{2,2}} \right) \\ &\quad - \mathbb{E} \left(\underbrace{\mathbf{X}_{\pi^h(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Delta}^\top \mathbf{X}_j}_{\Lambda_{2,3}} \right) - \mathbb{E} \left(\underbrace{\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^h(i)}^\top \mathbf{M} \mathbf{\Delta}^\top \mathbf{X}_{\pi^h(i)}}_{\Lambda_{2,4}} \right). \end{aligned}$$

Case (s, s) : $i = \pi^h(i)$ and $j = \pi^h(j)$. In this case, we have $\mathbf{\Delta}$ be

$$\mathbf{\Delta}^{(s,s)} = \mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top.$$

Hence we conclude

$$\begin{aligned} \mathbb{E} \Lambda_{2,1} &= \mathbb{E} \left[\mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top) \mathbf{X}_i \right] = \mathbb{E} \|\mathbf{X}_i\|_F^4 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i + \mathbb{E} \|\mathbf{X}_i\|_F^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i \\ &= (p+2)(p+4) \text{Tr}(\mathbf{M}) + (p+2) \text{Tr}(\mathbf{M}), \\ \mathbb{E} \Lambda_{2,2} &= \mathbb{E} \left[\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top) \mathbf{X}_j \right] = \mathbb{E} \|\mathbf{X}_i\|_F^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i + \mathbb{E} \|\mathbf{X}_j\|_F^2 \mathbf{X}_j^\top \mathbf{M} \mathbf{X}_j \\ &= 2 \mathbb{E} \|\mathbf{X}_i\|_F^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i = 2(p+2) \text{Tr}(\mathbf{M}), \\ \mathbb{E} \Lambda_{2,3} &= \mathbb{E} \left[\mathbf{X}_i^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top) \mathbf{X}_j \right] = 0, \\ \mathbb{E} \Lambda_{2,4} &= \mathbb{E} \left[\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{M} (\mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_j \mathbf{X}_j^\top) \mathbf{X}_i \right] = 0, \end{aligned}$$

which suggests $\mathbb{E} \Lambda_2 = (p+2)(p+7) \text{Tr}(\mathbf{M})$.

Case (s, d) : $i = \pi^h(i)$ and $j \neq \pi^h(j)$. First we write $\mathbf{\Delta}$ as

2106
2107

$$\Delta^{(s,d)\top} = \mathbf{X}_i \mathbf{X}_i^\top + \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top + \mathbf{X}_j \mathbf{X}_{\pi^{\natural-1}(j)}^\top.$$

2108

Then we conclude

2109

2110

2111

2112

2113

2114

2115

2116

2117

2118

2119

2120

2121

2122

2123

2124

2125

2126

2127

2128

2129

2130

2131

2132

2133

2134

2135

2136

2137

2138

2139

2140

2141

2142

2143

2144

2145

2146

2147

2148

2149

2150

2151

2152

2153

2154

2155

2156

2157

2158

2159

which suggests $\mathbb{E}\Lambda_2 = (p+2)(p+5) \text{Tr}(\mathbf{M})$.**Case (d, s):** $i \neq \pi^{\natural}(i)$ and $j = \pi^{\natural}(j)$. In this case, Δ reduces to

$$\Delta^{(d,s)\top} = \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top + \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top + \mathbf{X}_j \mathbf{X}_j^\top.$$

Then we obtain

$$\mathbb{E}\Lambda_{2,1} = \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \right)}_{\mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} + \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \right)}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} \mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i}$$

$$+ \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_{\pi^{\natural}(i)} \right)}_0 = (1 + \mathbb{1}_{i=\pi^{\natural 2}(i)}) (p+2) \text{Tr}(\mathbf{M}),$$

$$\mathbb{E}\Lambda_{2,2} = \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_j \right)}_{\mathbb{1}_{i=j} \mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^4 \text{Tr}(\mathbf{M}) + \mathbb{1}_{i \neq j} p \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_j \right)}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} [\mathbb{1}_{i=j} \mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i + \mathbb{1}_{i \neq j} \text{Tr}(\mathbf{M})]}$$

$$+ \underbrace{\mathbb{E} \left(\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_j \mathbf{X}_j^\top \mathbf{X}_j \right)}_0 = (p + \mathbb{1}_{i=\pi^{\natural 2}(i)}) \text{Tr}(\mathbf{M}),$$

$$\mathbb{E}\Lambda_{2,3} = \mathbb{E} \left[\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \left(\mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top + \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top + \mathbf{X}_j \mathbf{X}_j^\top \right) \mathbf{X}_j \right] = 0,$$

$$\mathbb{E}\Lambda_{2,4} = \mathbb{E} \left[\mathbf{X}_j^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \left(\mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top + \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top + \mathbf{X}_j \mathbf{X}_j^\top \right) \mathbf{X}_{\pi^{\natural}(i)} \right] = 0,$$

which suggests

$$\mathbb{E}\Lambda_2 = 2(p+1) \text{Tr}(\mathbf{M}) + (p+3) \mathbb{1}_{i=\pi^{\natural 2}(i)} \text{Tr}(\mathbf{M}).$$

Case (d, d): $i \neq \pi^{\natural}(i)$ and $j \neq \pi^{\natural}(j)$. In this case, Δ is written as

$$\Delta^{(d,d)\top} = \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top + \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top + \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^\top + \mathbf{X}_j \mathbf{X}_{\pi^{\natural-1}(j)}^\top.$$

Then we have

$$\mathbb{E}\Lambda_{2,1} = \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^\top \mathbf{X}_{\pi^{\natural}(i)} \right)}_{\mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i} + \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{M} \mathbf{X}_{\pi^{\natural 2}(i)} \mathbf{X}_{\pi^{\natural}(i)}^\top \mathbf{X}_{\pi^{\natural}(i)} \right)}_{\mathbb{1}_{i=\pi^{\natural 2}(i)} \mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^2 \mathbf{X}_i^\top \mathbf{M} \mathbf{X}_i}$$

$$\begin{aligned}
& + \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^{\top} \mathbf{X}_{\pi^{\natural}(i)} \right)}_{\mathbb{1}_{i=j} \mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^2 \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i} + \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^{\top} \mathbf{X}_{\pi^{\natural}(i)} \right)}_{\mathbb{1}_{i=j} \mathbb{1}_{i=\pi^{\natural}2(i)} \mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^2 \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i} \\
& = (p+2) [1 + \mathbb{1}_{i=\pi^{\natural}2(i)} + \mathbb{1}_{i=j} + \mathbb{1}_{i=\pi^{\natural}2(i)} \mathbb{1}_{i=j}] \text{Tr}(\mathbf{M}), \\
\mathbb{E} \Lambda_{2,2} & = \underbrace{\mathbb{E} \left(\mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^{\top} \mathbf{X}_j \right)}_{\mathbb{1}_{i=j} \mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^4 \text{Tr}(\mathbf{M}) + p \mathbb{1}_{i \neq j} \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \left(\mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}2(i)} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{X}_j \right)}_{\mathbb{1}_{i=\pi^{\natural}2(i)} [\mathbb{1}_{i=j} (p+2) \text{Tr}(\mathbf{M}) + \mathbb{1}_{i \neq j} \text{Tr}(\mathbf{M})]} \\
& + \underbrace{\mathbb{E} \left(\mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^{\top} \mathbf{X}_j \right)}_{\mathbb{1}_{i=j} \mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^4 \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \left(\mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^{\top} \mathbf{X}_j \right)}_{\mathbb{1}_{i=j} \mathbb{1}_{i=\pi^{\natural}2(i)} \mathbb{E} \|\mathbf{X}_i\|_{\mathbb{F}}^2 \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i} \\
& = 2 \mathbb{1}_{i=j} p (p+2) \text{Tr}(\mathbf{M}) + p \mathbb{1}_{i \neq j} \text{Tr}(\mathbf{M}) + \mathbb{1}_{i=\pi^{\natural}2(i)} [\mathbb{1}_{i=j} 2(p+2) \text{Tr}(\mathbf{M}) + \mathbb{1}_{i \neq j} \text{Tr}(\mathbf{M})], \\
\mathbb{E} \Lambda_{2,3} & = \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^{\top} \mathbf{X}_j \right)}_0 + \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}2(i)} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{X}_j \right)}_0 \\
& + \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^{\top} \mathbf{X}_j \right)}_{p \mathbb{1}_{i=\pi^{\natural}(j)} \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \left(\mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^{\top} \mathbf{X}_j \right)}_0, \\
\mathbb{E} \Lambda_{2,4} & = \underbrace{\mathbb{E} \left(\mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(i)} \mathbf{X}_i^{\top} \mathbf{X}_{\pi^{\natural}(i)} \right)}_0 + \underbrace{\mathbb{E} \left(\mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}2(i)} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{X}_{\pi^{\natural}(i)} \right)}_0 \\
& + \underbrace{\mathbb{E} \left(\mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(j)} \mathbf{X}_j^{\top} \mathbf{X}_{\pi^{\natural}(i)} \right)}_{\mathbb{1}_{i=\pi^{\natural}(j)} \text{Tr}(\mathbf{M})} + \underbrace{\mathbb{E} \left(\mathbf{X}_j^{\top} \mathbf{X}_i \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_j \mathbf{X}_{\pi^{\natural}-1(j)}^{\top} \mathbf{X}_{\pi^{\natural}(i)} \right)}_0,
\end{aligned}$$

which gives

$$\mathbb{E} \Lambda_2 = (p+1) [2 + 2(p+1) \mathbb{1}_{i=j} + \mathbb{1}_{i=\pi^{\natural}(j)}] \text{Tr}(\mathbf{M}) + \mathbb{1}_{i=\pi^{\natural}2(i)} [p+3 + (3p+5) \mathbb{1}_{i=j}] \text{Tr}(\mathbf{M}).$$

□

D.2.3 SUPPORTING LEMMAS

First, we study the higher order expectations of Gaussian random vectors' inner product, which hopefully will serve independent interests.

Lemma 11. Assume $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^p$ are Gaussian distributed random vectors whose entries follow the i.i.d. standard normal distribution, then we have

$$\mathbb{E} \text{Tr}(\mathbf{y} \mathbf{y}^{\top} \mathbf{x} \mathbf{x}^{\top} \mathbf{M}) = \text{Tr}(\mathbf{M}), \quad (63)$$

$$\mathbb{E} \text{Tr}(\mathbf{y} \mathbf{y}^{\top} \mathbf{x}^{\top} \mathbf{M} \mathbf{x}) = \mathbb{E} \|\mathbf{y}\|_2^2 \text{Tr}(\mathbf{x}^{\top} \mathbf{M} \mathbf{x}) = p \text{Tr}(\mathbf{M}), \quad (64)$$

$$\mathbb{E} (\mathbf{x}^{\top} \mathbf{M} \mathbf{x})^2 = [\text{Tr}(\mathbf{M})]^2 + \text{Tr}(\mathbf{M} \mathbf{M}) + \text{Tr}(\mathbf{M}^{\top} \mathbf{M}), \quad (65)$$

$$\mathbb{E} \|\mathbf{x}\|_2^2 (\mathbf{x}^{\top} \mathbf{M} \mathbf{x}) = (p+2) \text{Tr}(\mathbf{M}), \quad (66)$$

$$\mathbb{E} \|\mathbf{x}\|_2^4 (\mathbf{x}^{\top} \mathbf{M} \mathbf{x}) = (p+2)(p+4) \text{Tr}(\mathbf{M}), \quad (67)$$

$$\mathbb{E} \|\mathbf{x}\|_2^2 (\mathbf{x}^{\top} \mathbf{M} \mathbf{x})^2 = (p+4) \left[(\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M} \mathbf{M}) + \text{Tr}(\mathbf{M}^{\top} \mathbf{M}) \right], \quad (68)$$

$$\mathbb{E} \|\mathbf{x}\|_2^4 (\mathbf{x}^{\top} \mathbf{M} \mathbf{x})^2 = (p+4)(p+6) \left[(\text{Tr}(\mathbf{M}))^2 + \text{Tr}(\mathbf{M} \mathbf{M}) + \text{Tr}(\mathbf{M}^{\top} \mathbf{M}) \right], \quad (69)$$

$$\mathbb{E} (\mathbf{x}^{\top} \mathbf{y})^2 \mathbf{y}^{\top} \mathbf{M}_1 \mathbf{x} \mathbf{x}^{\top} \mathbf{M}_2 \mathbf{y} = 2 \text{Tr}(\mathbf{M}_1) \text{Tr}(\mathbf{M}_2) + (p+4) \text{Tr}(\mathbf{M}_1 \mathbf{M}_2) + 2 \text{Tr}(\mathbf{M}_1 \mathbf{M}_2^{\top}), \quad (70)$$

where $\mathbf{M} \in \mathbb{R}^{p \times p}$ is a fixed matrix.

Remark 2. If we assume $\mathbf{M} = \mathbf{I}_{p \times p}$, we can get $\mathbb{E} \|\mathbf{x}\|_2^4 = p(p+2)$, $\mathbb{E} \|\mathbf{x}\|_2^6 = p(p+2)(p+4)$, and $\mathbb{E} \|\mathbf{x}\|_2^8 = p(p+2)(p+4)(p+6)$.

Proof. This lemma is proved by iteratively applying the Wick's theorem in Theorem 4, Stein's lemma in Lemma 20, and Lemma 19.

- 2214 • **Proof of (63) and (64).** The proof can be conducted easily with the property such that
 2215 $\text{Tr}(\mathbf{u}\mathbf{v}^\top) = \mathbf{u}^\top \mathbf{v} = \text{Tr}(\mathbf{v}\mathbf{u}^\top)$ holds for arbitrary vectors \mathbf{u} and \mathbf{v} .
 2216
- 2217 • **Proof of (65).** This property is a direct consequence of Neudecker & Wansbeek (1987) (Equa-
 2218 tion (3.2)), which is attached in Lemma 19 for the sake of self-containing.

- 2219 • **Proof of (66).** Invoking the Stein’s lemma, we have

$$\mathbb{E}\|\mathbf{x}\|_2^2(\mathbf{x}^\top \mathbf{M}\mathbf{x}) = \mathbb{E}[\nabla_{\mathbf{x}}(\mathbf{x}^\top \mathbf{M}\mathbf{x})\mathbf{x}].$$

2221 Then our goal transforms to computing the trace of the Hessian matrix $\text{Tr}[\nabla_{\mathbf{x}} \text{Tr}(\mathbf{x}^\top \mathbf{M}\mathbf{x})\mathbf{x}]$.
 2222 For the i -th entry of the gradient, we have

$$\frac{d}{dx_i} \mathbf{x}^\top \mathbf{M}\mathbf{x} = \langle \mathbf{M}_i, \mathbf{x} \rangle + \langle (\mathbf{M}^\top)_i, \mathbf{x} \rangle,$$

2223 where \mathbf{M}_i is the i -th row (or column) of \mathbf{M} . Then we obtain

$$\frac{d}{dx_i} [x_i \text{Tr}(\mathbf{x}^\top \mathbf{M}\mathbf{x})] = \mathbf{x}^\top \mathbf{M}\mathbf{x} + x_i [\langle \mathbf{M}_i, \mathbf{x} \rangle + \langle (\mathbf{M}^\top)_i, \mathbf{x} \rangle],$$

2224 and hence

$$\begin{aligned} \mathbb{E}\|\mathbf{x}\|_2^2(\mathbf{x}^\top \mathbf{M}\mathbf{x}) &= \sum_{i=1}^p \mathbb{E}(\mathbf{x}^\top \mathbf{M}\mathbf{x}) + \sum_{i=1}^p \mathbb{E}[x_i (\langle \mathbf{M}_i, \mathbf{x} \rangle + \langle (\mathbf{M}^\top)_i, \mathbf{x} \rangle)] \\ &= p \text{Tr}(\mathbf{M}) + 2 \sum_i M_{ii} = (p+2) \text{Tr}(\mathbf{M}). \end{aligned}$$

- 2225
- 2226 • **Proof of (67).** Following the same strategy as in proving (66), we have

$$\mathbb{E}\|\mathbf{x}\|_F^4(\mathbf{x}^\top \mathbf{M}\mathbf{x}) = \mathbb{E}[\nabla_{\mathbf{x}}\|\mathbf{x}\|_2^2(\mathbf{x}^\top \mathbf{M}\mathbf{x})\mathbf{x}].$$

2227 Then our goal transforms to computing the trace of the Hessian matrix $\text{Tr}[\nabla_{\mathbf{x}}\|\mathbf{x}\|_2^2(\mathbf{x}^\top \mathbf{M}\mathbf{x})\mathbf{x}]$.
 2228 For the i -th entry of the gradient, we obtain

$$\begin{aligned} \frac{d}{dx_i} [x_i \|\mathbf{x}\|_2^2(\mathbf{x}^\top \mathbf{M}\mathbf{x})] &= \|\mathbf{x}\|_2^2 \cdot (\mathbf{x}^\top \mathbf{M}\mathbf{x}) + x_i \frac{d}{dx_i} [\|\mathbf{x}\|_2^2(\mathbf{x}^\top \mathbf{M}\mathbf{x})] \\ &= \|\mathbf{x}\|_2^2 \cdot (\mathbf{x}^\top \mathbf{M}\mathbf{x}) + 2x_i^2(\mathbf{x}^\top \mathbf{M}\mathbf{x}) + x_i \|\mathbf{x}\|_2^2 [\langle \mathbf{M}_i, \mathbf{x} \rangle + \langle (\mathbf{M}^\top)_i, \mathbf{x} \rangle], \end{aligned}$$

2229 whose expectation reads as

$$\begin{aligned} &\mathbb{E}\|\mathbf{x}\|_2^2 \cdot (\mathbf{x}^\top \mathbf{M}\mathbf{x}) + 2\mathbb{E}[x_i^2(\mathbf{x}^\top \mathbf{M}\mathbf{x})] + \mathbb{E}x_i \|\mathbf{x}\|_2^2 [\langle \mathbf{M}_i, \mathbf{x} \rangle + \langle (\mathbf{M}^\top)_i, \mathbf{x} \rangle] \\ &= (p+2) \text{Tr}(\mathbf{M}) + 2\mathbb{E}\left[x_i^2 M_{ii} x_i^2 + x_i^2 \left(\sum_{j \neq i} M_{jj} x_j^2\right)\right] + 2M_{ii} \mathbb{E}[x_i^2 \|\mathbf{x}\|_2^2] \\ &= (p+2) \text{Tr}(\mathbf{M}) + 2M_{ii} (\mathbb{E}x_i^4) + 2 \sum_{j \neq i} M_{jj} (\mathbb{E}x_i^2)(\mathbb{E}x_j^2) + 2M_{ii} \left[\mathbb{E}(x_i^4) + \sum_{j \neq i} (\mathbb{E}x_i^2)(\mathbb{E}x_j^2)\right] \\ &= (p+2) \text{Tr}(\mathbf{M}) + 6M_{ii} + 2 \sum_{j \neq i} M_{jj} + 2M_{ii} (3+p-1) \\ &= (p+4) \text{Tr}(\mathbf{M}) + 2(p+4)M_{ii}. \end{aligned}$$

2230 Then we conclude

$$\begin{aligned} \mathbb{E} \text{Tr}[\nabla_{\mathbf{x}}\|\mathbf{x}\|_2^2(\mathbf{x}^\top \mathbf{M}\mathbf{x})\mathbf{x}] &= p(p+2) \text{Tr}(\mathbf{M}) + 2(p+4) \sum_i M_{ii} + 2p \text{Tr}(\mathbf{M}) \\ &= (p+2)(p+4) \text{Tr}(\mathbf{M}). \end{aligned}$$

2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321

• **Proof of (68).** Invoking the Stein’s lemma, we have

$$\mathbb{E}\|\mathbf{x}\|_2^2 (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 = \sum_i \frac{d}{dx_i} \left[x_i (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 \right] = p (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 + 4 \sum_i x_i (\mathbf{x}^\top \mathbf{M}\mathbf{x}) \langle \mathbf{M}_i^{(\text{sym})}, \mathbf{x} \rangle.$$

The proof is then completed by invoking Lemma 12.

• **Proof of (69).** Following the same strategy as in proving (68), we consider the i -th gradient w.r.t x_i , which can be written as

$$\begin{aligned} \mathbb{E}\|\mathbf{x}\|_2^4 (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 &= \sum_i \frac{d}{dx_i} \left[x_i \|\mathbf{x}\|_2^2 (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 \right] \\ &= \sum_i \|\mathbf{x}\|_2^2 (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 + 2 \sum_i x_i^2 (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 + 4 \sum_i x_i \|\mathbf{x}\|_2^2 (\mathbf{x}^\top \mathbf{M}\mathbf{x}) \langle \mathbf{M}^{(\text{sym})}, \mathbf{x} \rangle \\ &= \sum_i \|\mathbf{x}\|_2^2 (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 + 2 \sum_i \mathbb{E} \frac{d}{dx_i} \left[x_i (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 \right] + 4 \sum_i \mathbb{E} \frac{d}{dx_i} \left[\|\mathbf{x}\|_2^2 (\mathbf{x}^\top \mathbf{M}\mathbf{x}) \langle \mathbf{M}_i^{(\text{sym})}, \mathbf{x} \rangle \right]. \end{aligned} \quad (71)$$

Noticing the following relations

$$\frac{d}{dx_i} \left[x_i (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 \right] = (\mathbf{x}^\top \mathbf{M}\mathbf{x})^2 + 4x_i (\mathbf{x}^\top \mathbf{M}\mathbf{x}) \langle \mathbf{M}_i^{(\text{sym})}, \mathbf{x} \rangle, \quad (72)$$

$$\begin{aligned} \frac{d}{dx_i} \left[\|\mathbf{x}\|_2^2 \text{Tr}(\mathbf{x}^\top \mathbf{M}\mathbf{x}) \langle \mathbf{M}_i^{(\text{sym})}, \mathbf{x} \rangle \right] &= 2x_i (\mathbf{x}^\top \mathbf{M}\mathbf{x}) \langle \mathbf{M}_i^{(\text{sym})}, \mathbf{x} \rangle \\ &\quad + 2\|\mathbf{x}\|_2^2 \langle \mathbf{M}_i^{(\text{sym})}, \mathbf{x} \rangle^2 + M_{ii} \|\mathbf{x}\|_2^2 (\mathbf{x}^\top \mathbf{M}\mathbf{x}), \end{aligned} \quad (73)$$

we can conclude the proof by combining (68), (71), (72), (73), and Lemma 12.

• **Proof of (70).** Due to the independence between \mathbf{x} and \mathbf{y} , we first condition on \mathbf{x} and have

$$\begin{aligned} \mathbb{E}(\mathbf{x}^\top \mathbf{y})^2 \mathbf{y}^\top \mathbf{M}_1 \mathbf{x} \mathbf{x}^\top \mathbf{M}_2 \mathbf{y} &= \mathbb{E}_{\mathbf{x}} \mathbb{E}_{\mathbf{y}} \mathbf{y}^\top \mathbf{x} \mathbf{x}^\top \mathbf{y} \mathbf{y}^\top \mathbf{M}_1 \mathbf{x} \mathbf{x}^\top \mathbf{M}_2 \mathbf{y} \\ &\stackrel{\textcircled{1}}{=} \mathbb{E}_{\mathbf{x}} \text{Tr}(\mathbf{x} \mathbf{x}^\top) \text{Tr}(\mathbf{M}_1 \mathbf{x} \mathbf{x}^\top \mathbf{M}_2) + \mathbb{E}_{\mathbf{x}} \text{Tr}(\mathbf{x} \mathbf{x}^\top \mathbf{M}_1 \mathbf{x} \mathbf{x}^\top \mathbf{M}_2) + \mathbb{E}_{\mathbf{x}} \text{Tr}(\mathbf{x} \mathbf{x}^\top \mathbf{M}_2^\top \mathbf{x} \mathbf{x}^\top \mathbf{M}_1^\top) \\ &= \mathbb{E}_{\mathbf{x}} \|\mathbf{x}\|_2^2 \text{Tr}(\mathbf{M}_2 \mathbf{M}_1) + \mathbb{E}_{\mathbf{x}} \mathbf{x}^\top \mathbf{M}_1 \mathbf{x} \mathbf{x}^\top \mathbf{M}_2 \mathbf{x} + \mathbb{E}_{\mathbf{x}} \mathbf{x}^\top \mathbf{M}_1^\top \mathbf{x} \mathbf{x}^\top \mathbf{M}_2^\top \mathbf{x} \\ &\stackrel{\textcircled{2}}{=} 2 \text{Tr}(\mathbf{M}_1) \text{Tr}(\mathbf{M}_2) + (p+4) \text{Tr}(\mathbf{M}_1 \mathbf{M}_2) + 2 \text{Tr}(\mathbf{M}_1 \mathbf{M}_2^\top), \end{aligned}$$

where in ① and ② we both use Lemma 19.

□

Lemma 12. For a fixed matrix $\mathbf{M} \in \mathbb{R}^{p \times p}$, we associate it with a symmetric matrix $\mathbf{M}^{(\text{sym})}$ defined as $(\mathbf{M} + \mathbf{M}^\top)/2$. Consider the Gaussian distributed random vector $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, we have

$$\mathbb{E} \sum_i x_i (\mathbf{x}^\top \mathbf{M}\mathbf{x}) \langle \mathbf{M}_i^{(\text{sym})}, \mathbf{x} \rangle = (\text{Tr}(\mathbf{M}))^2 + \|\mathbf{M}\|_F^2 + \text{Tr}(\mathbf{M}\mathbf{M}).$$

Proof. This lemma is a direct application of Wick’s theorem, which is completed by showing

$$\begin{aligned} \mathbb{E} x_i (\mathbf{x}^\top \mathbf{M}\mathbf{x}) \langle \mathbf{M}_i^{(\text{sym})}, \mathbf{x} \rangle &= \mathbb{E} \sum_j \sum_{\ell_1, \ell_2} M_{i,j}^{(\text{sym})} M_{\ell_1, \ell_2} x_i x_j x_{\ell_1} x_{\ell_2} \\ &= \underbrace{\mathbb{E} \sum_j \sum_{\ell_1, \ell_2} \mathbb{1}_{\ell_1=i} \mathbb{1}_{\ell_2=j} M_{i,j}^{(\text{sym})} M_{\ell_1, \ell_2} x_i x_j x_{\ell_1} x_{\ell_2}}_{\sum_j M_{i,j}^{(\text{sym})} M_{i,j}} + \underbrace{\mathbb{E} \sum_j \sum_{\ell_1, \ell_2} \mathbb{1}_{\ell_2=i} \mathbb{1}_{\ell_1=j} M_{i,j}^{(\text{sym})} M_{\ell_1, \ell_2} x_i x_j x_{\ell_1} x_{\ell_2}}_{\sum_j M_{i,j}^{(\text{sym})} M_{j,i}} \\ &\quad + \underbrace{\mathbb{E} \sum_j \sum_{\ell_1, \ell_2} \mathbb{1}_{i=j} \mathbb{1}_{\ell_1=\ell_2} M_{i,j}^{(\text{sym})} M_{\ell_1, \ell_2} x_i x_j x_{\ell_1} x_{\ell_2}}_{\sum_\ell M_{i,i}^{(\text{sym})} M_{\ell, \ell}} = \sum_j 2 \left[M_{i,j}^{(\text{sym})} \right]^2 + M_{i,i} \text{Tr}(\mathbf{M}), \end{aligned}$$

where $\mathbf{M}^{(\text{sym})}$ is defined as $(\mathbf{M} + \mathbf{M}^\top)/2$.

□

Then we study the properties of Σ , which is defined as $\mathbf{X}^\top \Pi^h \mathbf{X} - \Delta$.

Lemma 13. For a fixed matrix \mathbf{M} , we have

$$\mathbb{E} \text{Tr} \left(\Sigma \mathbf{M} \Sigma^\top \right) = n^2 \left[\frac{p}{n} + \left(1 - \frac{h}{n} \right)^2 + o(1) \right] \text{Tr}(\mathbf{M}),$$

where matrix Σ is defined in (22).

Proof. We conclude the proof by showing

$$\begin{aligned} \mathbb{E} \text{Tr} \left(\Sigma \mathbf{M} \Sigma^\top \right) &\stackrel{\textcircled{1}}{=} \sum_{\ell_1, \ell_2 \in \mathcal{S}} \mathbb{E} \text{Tr} \left[\mathbf{X}_{\ell_1} \mathbf{X}_{\ell_1}^\top \mathbf{M} \mathbf{X}_{\ell_2} \mathbf{X}_{\ell_2}^\top \right] + \sum_{\ell_1, \ell_2 \in \mathcal{D}} \mathbb{E} \text{Tr} \left[\mathbf{X}_{\ell_1} \mathbf{X}_{\pi^h(\ell_1)}^\top \mathbf{M} \mathbf{X}_{\pi^h(\ell_2)} \mathbf{X}_{\ell_2}^\top \right] \\ &= \sum_{\ell \in \mathcal{S}} \mathbb{E} \text{Tr} \left(\mathbf{X}_\ell \mathbf{X}_\ell^\top \mathbf{M} \mathbf{X}_\ell \mathbf{X}_\ell^\top \right) + \sum_{\ell_1, \ell_2 \in \mathcal{S}, \ell_1 \neq \ell_2} \mathbb{E} \text{Tr} \left[\mathbf{X}_{\ell_1} \mathbf{X}_{\ell_1}^\top \mathbf{M} \mathbf{X}_{\ell_2} \mathbf{X}_{\ell_2}^\top \right] \\ &\quad + \sum_{\ell \in \mathcal{D}} \underbrace{\mathbb{E} \text{Tr} \left[\mathbf{X}_\ell \mathbf{X}_{\pi^h(\ell)}^\top \mathbf{M} \mathbf{X}_{\pi^h(\ell)} \mathbf{X}_\ell^\top \right]}_{p \text{Tr}(\mathbf{M})} + \sum_{(\ell_1, \ell_2) \in \mathcal{D}_{\text{pair}}} \underbrace{\mathbb{E} \text{Tr} \left[\mathbf{X}_{\ell_1} \mathbf{X}_{\ell_2}^\top \mathbf{M} \mathbf{X}_{\ell_1} \mathbf{X}_{\ell_2}^\top \right]}_{\text{Tr}(\mathbf{M})} \\ &= (n-h)(p+2) \text{Tr}(\mathbf{M}) + (n-h)(n-h-1) \text{Tr}(\mathbf{M}) + hp \text{Tr}(\mathbf{M}) + |\mathcal{D}_{\text{pair}}| \text{Tr}(\mathbf{M}) \\ &\stackrel{\textcircled{2}}{=} n^2 \left[\frac{p}{n} + \left(1 - \frac{h}{n} \right)^2 + o(1) \right] \text{Tr}(\mathbf{M}), \end{aligned}$$

where $\textcircled{1}$ is due to the definitions of index sets \mathcal{S} and \mathcal{D} (Equation (27) and in Equation (28)), and $\textcircled{2}$ is because $|\mathcal{D}_{\text{pair}}| \leq h$. \square

Lemma 14. For a fixed matrix \mathbf{M} , we have

$$\mathbb{E} \text{Tr}(\Sigma \mathbf{M} \Sigma \mathbf{M}) = (n-h + |\mathcal{D}_{\text{pair}}|) [\text{Tr}(\mathbf{M})]^2 + (n-h)^2 \text{Tr}(\mathbf{M} \mathbf{M}) + n \text{Tr}(\mathbf{M} \mathbf{M}^\top)$$

Proof. Following the same strategy as in proving Lemma 13, we complete the proof by showing

$$\begin{aligned} \mathbb{E} \text{Tr}(\Sigma \mathbf{M} \Sigma \mathbf{M}) &= \sum_{\ell \in \mathcal{S}} \mathbb{E} \text{Tr} \left(\mathbf{X}_\ell^\top \mathbf{M} \mathbf{X}_\ell \mathbf{X}_\ell^\top \mathbf{M} \mathbf{X}_\ell \right) + \sum_{\ell_1, \ell_2 \in \mathcal{S}, \ell_1 \neq \ell_2} \mathbb{E} \text{Tr} \left(\mathbf{X}_{\ell_1} \mathbf{X}_{\ell_1}^\top \mathbf{M} \mathbf{X}_{\ell_2} \mathbf{X}_{\ell_2}^\top \mathbf{M} \right) \\ &\quad + \sum_{\ell \in \mathcal{D}} \mathbb{E} \text{Tr} \left(\mathbf{X}_\ell \mathbf{X}_{\pi^h(\ell)}^\top \mathbf{M} \mathbf{X}_\ell \mathbf{X}_{\pi^h(\ell)}^\top \mathbf{M} \right) + \sum_{\ell \in \mathcal{D}_{\text{pair}}} \mathbb{E} \text{Tr} \left(\mathbf{X}_\ell \mathbf{X}_{\pi^h(\ell)}^\top \mathbf{M} \mathbf{X}_{\pi^h(\ell)} \mathbf{X}_\ell^\top \mathbf{M} \right) \\ &= (n-h + |\mathcal{D}_{\text{pair}}|) [\text{Tr}(\mathbf{M})]^2 + (n-h)^2 \text{Tr}(\mathbf{M} \mathbf{M}) + n \text{Tr}(\mathbf{M} \mathbf{M}^\top). \end{aligned}$$

\square

Lemma 15. For a fixed matrix \mathbf{M} , we have

$$\mathbb{E} [\text{Tr}(\Sigma \mathbf{M})]^2 = (n-h)^2 [\text{Tr}(\mathbf{M})]^2 + n \text{Tr}(\mathbf{M}^\top \mathbf{M}) + (n-h + |\mathcal{D}_{\text{pair}}|) \text{Tr}(\mathbf{M} \mathbf{M}).$$

Proof. We complete the proof by showing

$$\begin{aligned} \mathbb{E} (\text{Tr}(\Sigma \mathbf{M}))^2 &= \sum_{\ell \in \mathcal{S}} \mathbb{E} \left(\mathbf{X}_\ell^\top \mathbf{M} \mathbf{X}_\ell \right)^2 + \sum_{\ell_1, \ell_2 \in \mathcal{S}, \ell_1 \neq \ell_2} (\text{Tr}(\mathbf{M}))^2 + \sum_{\ell \in \mathcal{D}} \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(\ell)}^\top \mathbf{M} \mathbf{X}_\ell \mathbf{X}_\ell^\top \mathbf{M}^\top \mathbf{X}_{\pi^h(\ell)}}_{\text{Tr}(\mathbf{M}^\top \mathbf{M})} \\ &\quad + \sum_{\ell \in \mathcal{D}_{\text{pair}}} \underbrace{\mathbb{E} \mathbf{X}_{\pi^h(\ell)}^\top \mathbf{M} \mathbf{X}_\ell \mathbf{X}_\ell^\top \mathbf{M} \mathbf{X}_{\pi^h(\ell)}}_{\text{Tr}(\mathbf{M} \mathbf{M})} \\ &= (n-h)^2 [\text{Tr}(\mathbf{M})]^2 + n \text{Tr}(\mathbf{M}^\top \mathbf{M}) + (n-h + |\mathcal{D}_{\text{pair}}|) \text{Tr}(\mathbf{M} \mathbf{M}). \end{aligned}$$

\square

2376 **Lemma 16.** For a fixed matrix \mathbf{M} , we have

$$2377 \mathbb{E} \sum_{\ell=\pi^{\natural}(\ell)} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\ell} \mathbf{X}_{\ell}^{\top} \mathbf{X}_{\pi^{\natural}(i)} = (n-h) \text{Tr}(\mathbf{M}) + (p+1) \mathbb{1}_{i=\pi^{\natural}(i)} \text{Tr}(\mathbf{M}).$$

2381 *Proof.* Provided that $i = \pi^{\natural}(i)$, we have

$$2382 \mathbb{E} \Xi_{1,1} = \mathbb{E} \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_i \mathbf{X}_i^{\top} \mathbf{X}_i + \sum_{\ell \neq i, \ell=\pi^{\natural}(\ell)} \mathbb{E} \mathbf{X}_i^{\top} \mathbf{M} \mathbf{X}_{\ell} \mathbf{X}_{\ell}^{\top} \mathbf{X}_i$$

$$2384 = (p+2) \text{Tr}(\mathbf{M}) + \sum_{\ell \neq i, \ell=\pi^{\natural}(\ell)} \text{Tr}(\mathbf{M}) = (n-h+p+1) \text{Tr}(\mathbf{M}) \mathbb{1}_{i=\pi^{\natural}(i)}. \quad (74)$$

2387 Provided that $i \neq \pi^{\natural}(i)$, we have

$$2388 \mathbb{E} \Xi_{1,1} = \sum_{\ell=\pi^{\natural}(\ell)} \mathbb{E} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\ell} \mathbf{X}_{\ell}^{\top} \mathbf{X}_{\pi^{\natural}(i)} = (n-h) \text{Tr}(\mathbf{M}) \mathbb{1}_{i \neq \pi^{\natural}(i)}. \quad (75)$$

2391 Combining (74) and (75) then completes the proof. \square

2392 **Lemma 17.** For a fixed \mathbf{M} , we have

$$2394 \mathbb{E} \sum_{\ell} \mathbf{X}_{\pi^{\natural}(i)}^{\top} \mathbf{M} \mathbf{X}_{\pi^{\natural}(\ell)} \mathbf{X}_{\ell}^{\top} \mathbf{X}_j = (p \mathbb{1}_{i=j} \mathbb{1}_{i \neq \pi^{\natural}(i)} + \mathbb{1}_{j=\pi^{\natural}(i)}) \text{Tr}(\mathbf{M}).$$

2397 We omit its proof as it is a direct application of Wick's theorem (Theorem 4).

2398 **Lemma 18.** We have

$$2399 \mathbb{E} \mathbb{1}(i = \pi^{\natural}(i)) = \frac{n-h}{n} (1 + o_{\mathbb{P}}(1)), \quad \mathbb{E} \mathbb{1}_{i=j} = \frac{h}{n^2} (1 + o_{\mathbb{P}}(1)),$$

$$2402 \mathbb{E} \mathbb{1}_{j=\pi^{\natural}(i)} = \frac{h}{n^2} (1 + o_{\mathbb{P}}(1)), \quad \mathbb{E} \mathbb{1}_{i=j} \mathbb{1}_{i=\pi^{\natural}(i)} = \frac{|\mathcal{D}_{\text{pair}}|}{n^2} (1 + o_{\mathbb{P}}(1)).$$

2404 This lemma can be easily proved by assuming the indices $i, j, \pi^{\natural}(i)$, and $\pi^{\natural}(j)$ are uniformly
2405 sampled from the set $\{1, 2, \dots, n\}$

2407 E USEFUL FACTS

2409 This section collects some useful facts for the sake of self-containing.

2410 **Theorem 4** (Wick's theorem (Theorem 1.28 in Janson (1997))). *Considering the centered jointly*
2411 *normal variables g_1, g_2, \dots, g_n , we conclude*

$$2413 \mathbb{E}(g_1 g_2 \dots g_n) = \sum_{\substack{\text{all possible disjoint} \\ \text{pairs } (i_k, j_k) \in \{1, 2, \dots, n\}}} \prod_k \mathbb{E}(g_{i_k} g_{j_k}).$$

2416 With Wick's theorem, we can reduce the computation of high-order Gaussian moments to calculating
2417 the expectations of a series of low-order Gaussian moments.

2418 **Lemma 19** (Equation (3.2) in Neudecker & Wansbeek (1987)). *For a normally distributed random*
2419 *matrix $\mathbf{G} \in \mathbb{R}^{n \times p}$ which satisfies $\mathbb{E} \mathbf{G} = \mathbf{0}$ and $\text{Evec}(\mathbf{G}) \text{vec}(\mathbf{G})^{\top} = \mathbf{U} \otimes \mathbf{V}$, we have*

$$2421 \mathbb{E}(\mathbf{G}^{\top} \mathbf{A} \mathbf{G} \mathbf{C} \mathbf{G}^{\top} \mathbf{B} \mathbf{G}) = \text{Tr}(\mathbf{A} \mathbf{U}) \text{Tr}(\mathbf{B} \mathbf{U}) \mathbf{V} \mathbf{C} \mathbf{V} + \text{Tr}(\mathbf{A} \mathbf{B}^{\top} \mathbf{U}) \mathbf{V} \mathbf{C}^{\top} \mathbf{V}$$

$$2422 + \text{Tr}(\mathbf{A} \mathbf{U} \mathbf{B} \mathbf{U}) \text{Tr}(\mathbf{C} \mathbf{V}) \mathbf{V},$$

2423 where $\text{vec}(\cdot)$ is the vector operation; \otimes is the Kronecker product (Horn & Johnson, 1990); and \mathbf{A}, \mathbf{B}
2424 and \mathbf{C} are arbitrary fixed matrices.

2425 **Lemma 20** (Stein's Lemma (cf. Section 1.3 in Talagrand (2010))). *Let $g \sim \mathcal{N}(0, 1)$. Then for any*
2426 *differentiable function $f: \mathbb{R} \mapsto \mathbb{R}$ we have*

$$2428 \mathbb{E}[gf(g)] = \mathbb{E}f'(g),$$

2429 where $\lim_{\|g\| \rightarrow \infty} f(g) e^{-a\|g\|_2^2} = 0$ for any $a > 0$.