

A Novel and Optimal Spectral Method for Permutation Synchronization

Duc Nguyen and Anderson Ye Zhang^{ID}

Abstract—Permutation synchronization is an important problem in computer science that constitutes the key step of many computer vision tasks. The goal is to recover n latent permutations from their noisy and incomplete pairwise measurements. In recent years, spectral methods have gained increasing popularity thanks to their simplicity and computational efficiency. Spectral methods utilize the leading eigenspace U of the data matrix and its block submatrices U_1, U_2, \dots, U_n to recover the permutations. In this paper, we propose a novel and statistically optimal spectral algorithm. Unlike the existing methods which use $\{U_j U_1^T\}_{j \geq 2}$, ours constructs an anchor matrix M by aggregating useful information from all of the block submatrices and estimates the latent permutations through $\{U_j M^T\}_{j \geq 1}$. This modification overcomes a crucial limitation of the existing methods caused by the repetitive use of U_1 and leads to an improved numerical performance. To establish the optimality of the proposed method, we carry out a fine-grained spectral analysis and obtain a sharp exponential error bound that matches the minimax rate.

Index Terms—Spectral method, permutation synchronization, spectral perturbation, minimax rate.

I. INTRODUCTION

IN PERMUTATION synchronization, the objective is to estimate n latent permutations using noisy and potentially incomplete pairwise measurements among them. It is an important problem in computer vision and graphics where finding correspondence between sets of features across multiple images is a fundamental task with wide-ranging applications including image registration [1], shape matching [2], multi-view matching [3], [4], detecting structures from motion [5], and so on. Various methods have been proposed for permutation synchronization including iterative algorithms [6], [7], [8], semi-definite programming (SDP) [9], [10], and spectral methods [4], [11], [12], [13]. Compared to other approaches, spectral methods have gained increasing popularity and have been widely used in permutation synchronization thanks to their simplicity, fast computation speed, and impressive numerical performance. Despite the popularity, it remains

unclear how well spectral methods perform theoretically and whether they achieve statistical optimality or not. In this paper, we address these questions by proposing a new and provably optimal spectral algorithm.

A. Problem Formulation

The permutation synchronization problem is formulated as follows. Let $Z_1^*, \dots, Z_n^* \in \Pi_d$ where Π_d is the permutation group in d dimension defined as:

$$\Pi_d := \{P \in \{0, 1\}^{d \times d} : P^T P = P P^T = I_d\}. \quad (1)$$

We introduce missing and noisy data by assuming that, for each pair $1 \leq j < k \leq n$, the observation $X_{jk} \in \mathbb{R}^{d \times d}$ satisfies

$$X_{jk} := \begin{cases} Z_j^* (Z_k^*)^T + \sigma W_{jk}, & \text{if } A_{jk} = 1, \\ 0_{d \times d}, & \text{otherwise,} \end{cases} \quad (2)$$

where $A_{jk} \in \{0, 1\}$ independently follows $\text{Bernoulli}(p)$ for some $p \in (0, 1]$; $\sigma \in \mathbb{R}^+$ controls the amount of noise; $W_{jk} \in \mathbb{R}^{d \times d}$ is a random matrix with each entry independently distributed according to the standard normal distribution; and $0_{d \times d}$ is the $d \times d$ matrix of all zeros. Roughly speaking, each X_{jk} block, if not missing, is a noisy measurement of $Z_j^* (Z_k^*)^T$ which is the ‘difference’ between two permutation matrices. Denote $Z^* := (Z_1^{*T}, \dots, Z_n^{*T})^T \in \Pi_d^n$. The goal is to estimate Z^* given $\{A_{jk}\}_{1 \leq j < k \leq n}$ and $\{X_{jk}\}_{1 \leq j < k \leq n}$. Note that Z_1^*, \dots, Z_n^* are identifiable only up to a global permutation. For any estimator $\hat{Z} = (\hat{Z}_1^T, \dots, \hat{Z}_n^T)^T \in \Pi_d^n$, its performance can be measured by the following normalized Hamming loss (modulo a global permutation transformation):

$$\ell(\hat{Z}, Z^*) := \min_{P \in \Pi_d} \frac{1}{n} \sum_{j=1}^n \mathbb{I}\{\hat{Z}_j \neq Z_j^* P^T\}. \quad (3)$$

Note that model (2) has a matrix representation. The observation matrix $X \in \mathbb{R}^{nd \times nd}$ can be written as

$$X = (A \otimes J_d) \circ Z^* (Z^*)^T + \sigma (A \otimes J_d) \circ W, \quad (4)$$

where $W \in \mathbb{R}^{nd \times nd}$ is a block-symmetric matrix with $W_{jj} := 0, W_{kj} := W_{jk}^T, \forall 1 \leq j < k \leq n$; $A \in \{0, 1\}^{n \times n}$ is a symmetric matrix with $A_{jj} := 0, A_{kj} := A_{jk}, \forall 1 \leq j < k \leq n$; J_d is the $d \times d$ matrix of all ones; \otimes is the Kronecker product and \circ is the Hadamard product.

B. Spectral Methods

Existing spectral methods [4], [11], [12], [13] use the eigen-decomposition of X followed by a rounding step to estimate

Received 31 March 2023; revised 9 May 2024; accepted 13 February 2025. Date of publication 25 February 2025; date of current version 23 April 2025. The work of Anderson Ye Zhang was supported in part by NSF under Grant DMS-2112988. (Corresponding author: Anderson Ye Zhang.)

Duc Nguyen was with the Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104 USA.

Anderson Ye Zhang is with the Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: ayz@wharton.upenn.edu).

Communicated by Y. M. Lu, Associate Editor for Machine Learning and Statistics.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TIT.2025.3545377>.

Digital Object Identifier 10.1109/TIT.2025.3545377

0018-9448 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

Authorized licensed use limited to: University of Pennsylvania. Downloaded on April 30, 2025 at 17:24:46 UTC from IEEE Xplore. Restrictions apply.

the latent permutations. Let $U = (U_1^\top, \dots, U_n^\top)^\top \in \mathbb{R}^{nd \times d}$ be the matrix composing the top d eigenvectors of X . That is, the columns of U are the eigenvectors corresponding to the d largest eigenvalues of X . Its $d \times d$ blocks are denoted as U_1, \dots, U_n . The first block $U_1 \in \mathbb{R}^{d \times d}$ is then used as an ‘anchor’ to obtain an estimator $\tilde{Z}_1, \dots, \tilde{Z}_n \in \Pi_d$:

$$\tilde{Z}_1 := I_d, \quad \tilde{Z}_j := \operatorname{argmax}_{P \in \Pi_d} \langle P, U_j U_1^\top \rangle, \quad \forall j = 2, \dots, n, \quad (5)$$

where the optimization subproblem serves to round $U_j U_1^\top$ into a permutation matrix and can be efficiently solved using the Kuhn-Munkres algorithm [14] (see Section II-A for an intuitive description of the innerworking of this algorithm). To distinguish the existing algorithm from the one proposed in this paper, we refer to $\tilde{Z} := (\tilde{Z}_1^\top, \dots, \tilde{Z}_n^\top)^\top$ as the ‘vanilla spectral estimator’. It is computationally efficient and has decent numerical performance.

Despite of all the aforementioned advantages, the vanilla spectral method suffers from the repeated use of U_1 in constructing the estimator \tilde{Z}_j for all $j \geq 2$. From a perturbation theoretical point of view, U (along with its blocks U_1, \dots, U_n) are approximations of their population counterparts. By using $U_j U_1^\top$, the estimation accuracy of \tilde{Z}_j is determined by the approximation errors of *both* U_j and U_1 . As a result, the approximation error of U_1 is carried forward in $\{\tilde{Z}_j\}_{j \geq 2}$ and deteriorates the overall numerical performance.

To overcome this crucial limitation of \tilde{Z} , we propose a new spectral method that avoids the use of U_1 as the anchor. Instead, we construct an anchor matrix $M \in \mathbb{R}^{d \times d}$ by carefully aggregating useful information from all of U and estimate the latent permutations by $\hat{Z} := (\hat{Z}_1^\top, \dots, \hat{Z}_n^\top)^\top$ where

$$\hat{Z}_j := \operatorname{argmax}_{P \in \Pi_d} \langle P, U_j M^\top \rangle, \quad \forall j \in [n].$$

The construction of M is built on an intuition that ‘averaging’ information across all n blocks of U leads to a more accurate anchor with smaller variance than U_1 . As a result, the estimation accuracy of \hat{Z}_j is mostly determined by the approximation error of U_j only, which leads to an improved numerical performance. See Algorithm 1 for the detailed implementation of the proposed method and Figure 1 for comparisons of numerical performances between the vanilla spectral estimator and ours.

C. Statistical Optimality

By carrying out fine-grained spectral analysis, we establish a sharp upper bound for the theoretical performance of the proposed method, summarized below in Theorem 1. We note that in this paper, p, σ^2, d are not constants but functions of n . This dependence can be more explicitly represented as p_n, σ_n^2, d_n . However, for simplicity of notation and readability, we choose to denote them as p, σ^2, d throughout the paper. See Theorem 3 for a non-asymptotic and refined version where d is also allowed to grow with n .

Theorem 1: Assume $\frac{np}{\sigma^2} \rightarrow \infty$, $\frac{np}{\log^3 n} \rightarrow \infty$ and $d = O(1)$. Then the proposed spectral method \hat{Z} satisfies

$$\mathbb{E} \ell(\hat{Z}, Z^*) \leq \exp\left(-(1 - o(1)) \frac{np}{2\sigma^2}\right) + n^{-8}.$$

The upper bound in Theorem 1 consists of an exponential error term and a polynomial error term n^{-8} . Note

that by properties of the normalized Hamming loss ℓ , the polynomial error term is negligible. Considering this, our spectral algorithm achieves the minimax lower bound [6] which states that if $\frac{np}{\sigma^2} \rightarrow \infty$, then $\inf_{Z'} \sup_{Z^* \in \Pi_d^n} \mathbb{E} \ell(Z', Z^*) \geq \exp\left(-(1 + o(1)) \frac{np}{2\sigma^2}\right)$. This establishes the statistical optimality of the proposed method for the partial recovery of the latent permutations. Theorem 1 immediately implies the threshold for exact recovery. When $np/(2\sigma^2) > (1 + \gamma) \log n$ for some constant $\gamma > 0$, we have $\ell(\hat{Z}, Z^*) = 0$ holds with high probability. According to the minimax lower bound, no estimator is able to recover Z^* exactly with vanishing error if $np/(2\sigma^2) < (1 - \gamma) \log n$. As a result, simple but powerful, the proposed spectral method \hat{Z} is an optimal procedure.

Theorem 1 allows the observations $\{X_{jk}\}$ to be missing at random as long as the probability p satisfies $np \gg \log^3 n$. Note that in order to have a connected comparison graph A , np needs to be at least of order $\log n$. Compared to this condition, our assumption $np \gg \log^3 n$ requires some additional logarithm factor. The assumption $np \gg \sigma^2$ is the necessary and sufficient condition to achieve estimation consistency according to the minimax lower bound. Theorem 1 assumes that d , the dimension of each permutation matrix, is a constant. To establish Theorem 1, we first provide a block-wise ℓ_∞ perturbation analysis for all block submatrices U_1, \dots, U_n , quantifying the maximum deviation between them and their population counterparts (see Theorem 2). In addition, we give a theoretical justification for the usage of the anchor M by showing that it achieves a negligible error (see Proposition 1). With both results, we investigate the tail behavior of each $U_j M^\top$ and eventually obtain the upper bound in Theorem 1. We leverage the leave-one-out technique [15], [16] in our proofs.

D. Related Literature

Permutation synchronization belongs to a broader class of group synchronization problems where the goal is to identify n group objects based on pairwise measurements among them. In recent years, spectral methods have been widely used and studied in group synchronization problems. In [17], spectral methods are proved to be optimal for phase synchronization and orthogonal group synchronization in terms of squared ℓ_2 losses. To obtain this result, [17] develops perturbation analysis toolkits to show that the leading eigenstructures can be well-approximated by its first-order approximation with a small ℓ_2 error. However, the difference that permutations are discrete-valued while phases and orthogonal matrices are continuous is critical. For permutation synchronization, instead of ℓ_2 perturbation analysis, we need to develop block-wise analysis in order to obtain sharp exponential rates. Reference [16] considers a \mathbb{Z}_2 synchronization problem where each object is ± 1 and assumes that there is no missing data. It proves that a simple spectral procedure using signs of coordinates of the first eigenvector of the data matrix achieves the optimal threshold for the exact recovery of objects by ℓ_∞ analysis of the leading eigenvector. Reference [12] extends [16]’s ℓ_∞ analysis to a permutation synchronization setting where there is no missing data and each observation is corrupted with probability q by a random permutation matrix. It shows that the vanilla spectral method achieves exact recovery when q satisfies certain conditions. Our work’s novelty relative to [12]

is two-fold. Firstly, we propose a novel spectral algorithm that addresses a subtle but important limitation of the vanilla spectral algorithm, leading to a significant improvement in empirical performance. Secondly, our ℓ_∞ analysis is different from those in [12] and [16] as we need to consider the low signal-to-noise ratio regime where exact recovery is impossible but partial recovery is possible. We go beyond ℓ_∞ analysis and study the tail behavior of each block of U in order to obtain exponential error bounds for partial recovery. In addition, the presence of missing data in our model complicates the theoretical analysis as the magnitude and tail behavior of each block U_j is not only related to the additive Gaussian noises but also the randomness of the Bernoulli random variables $\{A_{jk}\}_{k \neq j}$.

E. Notation

For any positive integer n , define $[n] := \{1, 2, \dots, n\}$. Let I_d denote the $d \times d$ identity matrix and J_d denote the $d \times d$ matrix of all ones. Define $\mathcal{O}_d := \{O \in \mathbb{R}^{d \times d} : OO^\top = O^\top O = I_d\}$ to be the set of all $d \times d$ orthogonal matrices. Given $a, b \in \mathbb{R}$, let $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. For a matrix $B \in \mathbb{R}^{d_1 \times d_2}$, the Frobenius norm and the operator norm of B are defined as $\|B\|_F := \left(\sum_{i=1}^{d_1} \sum_{j=1}^{d_2} B_{ij}^2\right)^{1/2}$ and $\|B\|_{\text{op}} := \max_{u \in \mathbb{S}_{d_1}, v \in \mathbb{S}_{d_2}} u^\top B v$ where $\mathbb{S}_d := \{v \in \mathbb{R}^d : \|v\| = 1\}$ is the unit sphere in d dimension and $\|\cdot\|$ is the Euclidean norm. For two matrices $A, B \in \mathbb{R}^{d_1 \times d_2}$, let $\langle A, B \rangle := \sum_{i \in [d_1]} \sum_{j \in [d_2]} A_{ij} B_{ij}$ denote the matrix inner product (this reduces to the usual vector inner product when $d_2 = 1$). For some $d_1, d_2 \in \mathbb{N}$, we use $\mathcal{N}(\mu, \Sigma)$ to denote the normal distribution with mean $\mu \in \mathbb{R}^{d_1}$ and covariance $\Sigma \in \mathbb{R}^{d_1 \times d_1}$ and $\mathcal{MN}(S, \Sigma_1, \Sigma_2)$ to denote the matrix normal distribution with mean parameter $S \in \mathbb{R}^{d_1 \times d_1}$ and covariance parameters $\Sigma_1 \in \mathbb{R}^{d_1 \times d_1}$, $\Sigma_2 \in \mathbb{R}^{d_2 \times d_2}$. Define $(a)_+ := \max\{a, 0\}$. For the rest of the paper, we will use $e_i \in \{0, 1\}^d$ to denote the vector with 1 in the i -th entry and zero everywhere else. For any matrix Y and any positive integer i , we denote $\lambda_i(Y)$ to be the i th largest eigenvalue of Y . For two positive sequences $\{a_n\}$ and $\{b_n\}$, $b_n \gtrsim a_n$ and $a_n = O(b_n)$ both mean $a_n \leq C b_n$ for some constant $C > 0$ independent of n . We also write $a_n = o(b_n)$ or $\frac{b_n}{a_n} \rightarrow \infty$ when $\limsup_n \frac{a_n}{b_n} = 0$. Lastly, we use $\mathbb{I}\{\cdot\}$ to denote the indicator function.

II. A NOVEL SPECTRAL ALGORITHM

In this section, we first give a detailed implementation of the proposed method. In Section II-A, we provide intuitions to explain why it works and how it improves upon the existing algorithm. In Section II-B, we compare the numerical performances of the proposed method with the vanilla spectral method on synthetic datasets.

Algorithm 1 describes our proposed spectral algorithm. The first step computes the top d eigenvectors of the observation matrix X , which can be done using any off-the-shelf numerical eigendecomposition routine. The output of Step 1 is the empirical eigenspace $U \in \mathbb{R}^{nd \times d}$ corresponding to the top d eigenvectors. Step 2 constructs the anchor M used to recover the permutations by performing d -means clustering on the nd rows of U and extracting the estimated cluster centers. Step 3 recovers the underlying permutations via the Kuhn-Munkres algorithm. Note that (8) in Step 3 can be interpreted as a

projection of $U_j M^\top$ onto Π_d . This is because all permutation matrices in Π_d have the same Frobenius norm and (8) can be equivalently written as

$$\hat{Z}_j = \underset{P \in \Pi_d}{\operatorname{argmin}} \|P - U_j M^\top\|_F, \forall j \in [n]. \quad (6)$$

Algorithm 1 A New Spectral Method for Permutation Synchronization

Input: Data matrix $X \in \mathbb{R}^{nd \times nd}$

Output: n permutation matrices $\hat{Z}_1, \hat{Z}_2, \dots, \hat{Z}_n \in \Pi_d$

- 1 Obtain the top d eigenvectors of X as $U \in \mathbb{R}^{nd \times d}$;
- 2 Denote the rows of U as $v_1, \dots, v_{nd} \in \mathbb{R}^{d \times 1}$ such that $U = (v_1^\top, \dots, v_{nd}^\top)^\top$. Run d -means on v_1, \dots, v_{nd} and denote $\hat{\mu}_1, \dots, \hat{\mu}_d \in \mathbb{R}^{d \times 1}$ as the cluster centers:

$$(\hat{\mu}_1, \dots, \hat{\mu}_d) := \underset{\mu_1, \dots, \mu_d \in \mathbb{R}^{d \times 1}}{\operatorname{argmin}} \min_{z \in [d]^{nd}} \sum_{j=1}^{nd} \|v_j - \mu_{z_j}\|^2. \quad (7)$$

Define $M := (\sqrt{n}\hat{\mu}_1^\top, \dots, \sqrt{n}\hat{\mu}_d^\top)^\top \in \mathbb{R}^{d \times d}$ such that the rows of M are the d centers multiplied by a \sqrt{n} scaling;

- 3 Compute for each $j \in [n]$,

$$\hat{Z}_j := \underset{P \in \Pi_d}{\operatorname{argmax}} \langle P, U_j M^\top \rangle. \quad (8)$$

A. Intuition

To understand Algorithm 1, we need to study U through the lens of spectral perturbation theory. Denote

$$U^* := (U_1^{*\top}, \dots, U_n^{*\top})^\top \in \mathbb{R}^{nd \times n}, \quad (9)$$

where $U_j^* := Z_j^* / \sqrt{n}$, $\forall j \in [n]$. Note that the expected value of the data matrix X is equal to $\mathbb{E}X = p((J_n - I_n) \otimes J_d) \circ Z^*(Z^*)^\top$. One could verify that $(\mathbb{E}X)U^* = (n-1)pU^*$. As a result, U^* is the leading eigenspace of $\mathbb{E}X$ that includes its top d eigenvectors. Since X is a perturbed version of $\mathbb{E}X$, U can also be seen as a perturbed version of U^* . However, this correspondence only holds with respect to an orthogonal transformation, as the leading eigenvalue of $\mathbb{E}X$, $(n-1)p$, has a multiplicity of d . That is, $UO^* \approx U^*$ for some $O^* \in \mathcal{O}_d$ (see Section III-A for the definition of O^*). As a result, the blocks $\{U_j\}$ are close to $\{U_j^*\}$ only up to a global orthogonal transformation, i.e., $U_j O^* \approx U_j^*$ for all $j \in [n]$.

To estimate the latent permutation Z_j^* , the vanilla spectral method uses the product $U_j U_1^\top$ as $U_j U_1^\top = (U_j O^*)(U_1 O^*)^\top \approx U_j^* (U_1^*)^\top = Z_j^* (Z_1^*)^\top / n$. Intuitively, if $U_j U_1^\top$ is very close to $Z_j^* Z_1^* / n$ then by (5), we have $\tilde{Z}_j = Z_j^* (Z_1^*)^\top$. As a result, when the perturbation between UO^* and U^* is sufficiently small, the vanilla spectral method is able to recover Z^* up to a global permutation Z_1^* . However, as hinted at before, the use of the product $\{U_j U_1^\top\}$ in the vanilla spectral method inevitably leads to a crucial and fundamental limitation. For each $j \in [n]$, write $U_j O^* = U_j^* + \xi_j$ such that ξ_j can be interpreted as the approximation ‘noise’ of $U_j O^*$ with respect to U_j^* . Then

$$\begin{aligned} U_j U_1^\top &= (U_j O^*)(U_1 O^*)^\top = (U_j^* + \xi_j)(U_1^* + \xi_1)^\top \\ &= U_j^* (U_1^*)^\top + \xi_j (U_1^*)^\top + U_j^* \xi_1^\top + \xi_j \xi_1^\top. \end{aligned}$$

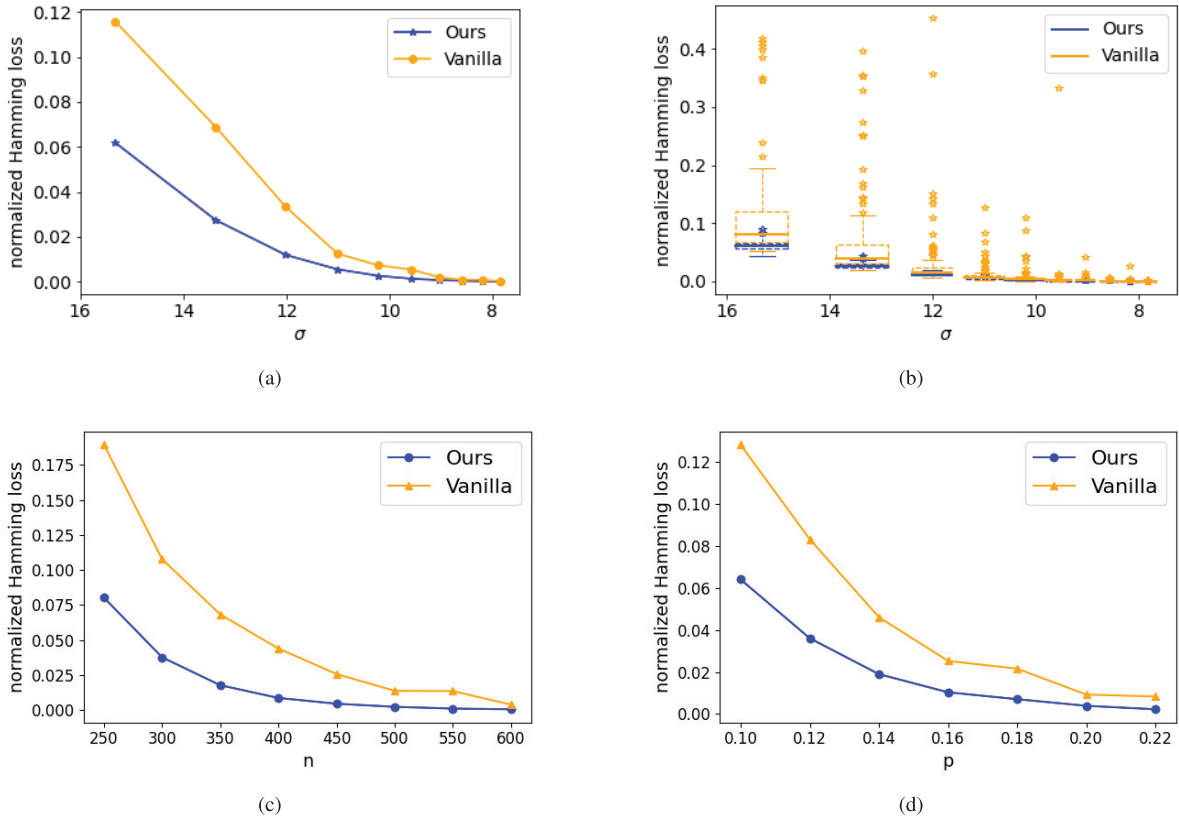


Fig. 1. Comparisons between our method and the vanilla spectra method on synthetic data.

That is, $U_j U_1^\top$ is an approximation of $U_j^* (U_1^*)^\top$ with an error $\xi_j (U_1^*)^\top + U_j^* \xi_1^\top$ (the higher-order term $\xi_j \xi_1^\top$ is ignored). As a function of $U_j U_1^\top$, the estimation accuracy of \hat{Z}_j is determined by both ξ_j and ξ_1 . As a result, the error caused by ξ_1 is carried forward in $\{\hat{Z}_j\}_{j \geq 2}$ and impairs the numerical performance of the overall algorithm (see Figure 1). Additionally, using U_1 as the anchor makes the performance of the vanilla spectral algorithm less stable because it highly depends on the accuracy of U_1 .

Algorithm 1 overcomes this crucial limitation of the vanilla spectral method by constructing an anchor that ‘averages’ information across all rows of U instead of just using its first block U_1 . The key insight into our construction is to recognize the special structure of the permutation synchronization problem where the true eigenspace U^* consists of d unique rows $e_1^\top / \sqrt{n}, \dots, e_d^\top / \sqrt{n}$, each of cardinality exactly n . The empirical eigenspace U , being a noisy estimate of $U^* O^{*\top}$, thus exhibits clustering structures where the cluster centers are the transformed rows $\{e_i^\top O^{*\top} / \sqrt{n}\}_{i=1}^d$. Clustering algorithms such as the d -means algorithm can be used to estimate the cluster centers, and thereby the rotation matrix O^* , accurately. With M being an accurate approximation of $O^{*\top}$ (up to a row permutation), we have

$$U_j M^\top \approx U_j O^* = U_j^* + \xi_j,$$

and consequently the estimation accuracy of \hat{Z}_j is only related to ξ_j whereas the accuracy of $U_j U_1^\top$ depends on both ξ_j and ξ_1 . The reduced noise in $U_j M^\top$ leads to an improved and more stable numerical performance.

B. Numerical Analysis

To verify our intuitions and showcase the improved performance of our spectral method over the vanilla spectral method, we perform experiments on synthetic data following the model in (2).

In Figures 1a-1d, we compare the performance of Algorithm 1 against the vanilla spectral algorithm. In each figure, we vary a single parameter while keeping all of the other parameters fixed: in Figures 1a-1b, we set $d = 2, p = 0.5, n = 2048$ and vary σ ; in Figure 1c, we set $\sigma = 1, p = 0.1, d = 3$ and vary n ; in Figure 1a, we set $n = 200, \sigma = 2, d = 2$ and vary p . In Figures 1a and 1c-1d, each line shows the average over 100 independent trials. One can see that our spectral method outperforms the vanilla spectral method across all settings.

The box plot in Figure 1b shows the distribution of the losses across 100 trials for both methods. The boxes extend from the first quartile to the third quartile of the observed losses with colored bolded lines at the medians and dotted points as the outliers. It is clear that the performance of the vanilla algorithm is highly variable across trials. On the contrary, our method has a smaller variance and is more precise. Figure 1 reflects our intuition that because the vanilla algorithm repeatedly uses U_1 to estimate the latent permutations, its performance is highly dependent on the approximation error of a single block U_1 . On the other hand, because our algorithm constructs an anchor that averages information across all n blocks, it is more stable and consistently outperforms the vanilla algorithm with a much tighter error spread.

III. THEORETICAL GUARANTEES

In this section, we establish theoretical guarantees for Algorithm 1. Similarly to previous analysis [12], without loss of generality¹, we assume that $Z_j^* = I_d$ for all $j \in [n]$. That is, all the latent permutations are the identity matrix. In this way, the population eigenspace U^* has a simpler expression with

$$U_j^* = I_d / \sqrt{n}, \forall j \in [n]. \quad (10)$$

We first give a justification for the choice of the anchor M , the key algorithmic novelty of the proposed algorithm. We then establish the statistical optimality of the proposed method in Section III-B. As a preview of our proofs, we then present an intuitive analysis of the block-wise perturbation analysis for U which will serve as the starting point of our detailed proofs. Last, we propose a practical modification of Algorithm 1 in Section III-D which uses approximate clustering instead of exact clustering, allowing the algorithm to remain scalable when d diverges. The modified spectral algorithm enjoys similar theoretical guarantee as Algorithm 1 under mild additional assumptions.

A. Accuracy of the Anchor M

At a high level, the effectiveness of M as the anchor stems from the property that UM^T is close to U^* up to a global permutation. However, this is not obvious at first glance.

Note that there is a close relationship between the empirical eigenspace U and the population eigenspace U^* , modulo an appropriate orthogonal transformation. Readers familiar with spectral analysis literature might anticipate the natural choice for this transformation to be the sign matrix [18], defined as $O^* := \operatorname{argmin}_{O \in \mathcal{O}_d} \|UO - U^*\|_F$. However, our analysis employs a different approach, using the following matrix [16]:

$$H := U^T U^*. \quad (11)$$

This choice is strategic; although H is not orthonormal, it offers a straightforward and explicit expression that facilitates easier manipulation than O^* . Additionally, O^* , by definition, represents the orthogonal matrix that most closely approximates H , and the distance between them is small. We first show (see (29) of Lemma 1):

$$\|UH - U^*\|_{\text{op}} = O\left(\frac{1 + \sigma\sqrt{d}}{\sqrt{np}}\right) \quad (12)$$

holds with high probability. Then following the intuition given in Section II-A and applying state-of-the-art clustering analysis, we have the following proposition which states that the anchor M is in fact an approximation of H^T up to some permutation matrix \hat{P} where $\hat{P} := \operatorname{argmin}_{P \in \Pi_d} \|M - PH^T\|_F$.

Proposition 1: There exist constants $C_1, C_2, C_3 > 0$ such that if $\frac{np}{\log n} > C_1$ and $\frac{np}{(\sqrt{d} + \sigma d)^2} > C_2$ then the anchor M constructed in Algorithm 1 satisfies

$$\min_{P \in \Pi_d} \|M - PH^T\|_F \leq \frac{C_3(\sqrt{d} + \sigma d)}{\sqrt{np}}, \quad (13)$$

with probability at least $1 - n^{-10}$.

¹This is because for any X as defined in (2) with an arbitrary $Z^* \in \Pi_d^n$, it can be transformed into $X' \in \mathbb{R}^{nd \times nd}$ by letting $X'_{jk} := (Z_j^*)^T X_{jk} Z_k^*$, $\forall 1 \leq j, k \leq n$. One can show that implementing Algorithm 7 on either X or X' yields identical results, up to a global permutation. In addition, it can be shown that X' also adheres to the definition in (2), where all latent permutation matrices are the identity matrix I_d .

The existence of \hat{P} is due to the fact that the clusters are identifiable only up to a permutation in cluster analysis. More importantly, the exact value \hat{P} does not affect the final error bound in (3) because the Hamming loss is defined with respect to an unknown optimal global permutation. (13) shows that the estimation error goes to 0 when $\frac{np}{(\sqrt{d} + \sigma d)^2}$ grows. As a comparison, the operator norm of the matrix H is 'of a constant order' as it is close to an orthogonal matrix (see (31) of Lemma 1). Hence, the error incurred by M as an estimate of H^T is of diminishing proportion. It is also worth mentioning that Proposition 1 holds under weaker assumptions compared to Theorem 1. It only requires $np \gtrsim \log n$ and allows d to grow as long as $np \gtrsim (\sqrt{d} + \sigma d)^2$. With Proposition 1, we can have a decomposition of the quantity $U_j M^T$, the key quantity in (8).

$$\begin{aligned} U_j M^T - U_j^* \hat{P}^T &= U_j (\hat{P} H^T + M - \hat{P} H^T)^T - U_j^* \hat{P}^T \\ &= (U_j H - U_j^*) \hat{P}^T + U_j (M - \hat{P} H^T)^T \\ &\approx (U_j H - U_j^*) \hat{P}^T. \end{aligned} \quad (14)$$

The approximation (14) is due to (13) as the term $U_j (M - \hat{P} H^T)^T$ turns out to be negligible. Hence, the difference $U_j M^T - U_j^* \hat{P}^T$ is primarily about the perturbation $U_j H - U_j^*$. To analyze our spectral method, we need to have a deep understanding on the behaviors of the blockwise perturbation $U_j H - U_j^*$.

B. Statistical Optimality

In Theorem 2, we first derive a block-wise ℓ_∞ upper bound for $UH - U^*$, i.e., an upper bound for $\|U_j H - U_j^*\|_{\text{op}}$ that holds uniformly across all $j \in [n]$.

Theorem 2: There exist constants $C, C', C'' > 0$ such that if $\frac{np}{(\sigma^{4/3} \sqrt{1}) \log n} > C$ and $\frac{np}{(\sqrt{d} + \sigma d)^2} \geq C'$, we have

$$\max_{j \in [n]} \|U_j H - U_j^*\|_{\text{op}} \leq \frac{C''}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma \sqrt{d}}{\sqrt{np}} + \frac{\sigma \sqrt{\log n}}{\sqrt{np}} \right) \quad (15)$$

with probability at least $1 - 6dn^{-9}$.

The upper bound in (15) is equal to the upper bound of $\|UH - U^*\|_{\text{op}}$ in (12) multiplied by a $\sqrt{(\log n)/n}$ factor. This is because $UH - U^*$ consists of n blocks $\{U_j H - U_j^*\}_{j \in [n]}$ and all blocks behave similarly. The magnitude of each block is then on average $1/\sqrt{n}$ of that of the whole matrix, and the $\sqrt{\log n}$ factor is due to the use of union bound to control the supremum. Theorem 2 assumes $\frac{np}{(\sigma^{4/3} \sqrt{1}) \log n} \gtrsim 1$. A sufficient condition is $np/\log^3 n \gtrsim 1$ as $np/\sigma^2 \gtrsim 1$ is implied by the other assumption $np/(\sqrt{d} + \sigma d)^2 \gtrsim 1$.

With the upper bound for each $\|U_j H - U_j^*\|_{\text{op}}$ derived, we further study the tail behavior of each difference $U_j H - U_j^*$. This leads to a sharp theoretical analysis of the performance of the proposed spectral method. The main theoretical result of this paper is stated below in Theorem 3.

Theorem 3: There exist constants $\bar{C}', \bar{C}'', \bar{C}''', \bar{C}'''' > 0$ such that if $\frac{np}{d^2 \log n} > \bar{C}'$, $\frac{np}{\sigma^2 d^3} > \bar{C}''$ and $\frac{np}{\log^3 n} > \bar{C}'''$ then the estimates $\hat{Z}_1, \dots, \hat{Z}_n$ of Algorithm 1 satisfy

$$\begin{aligned} \mathbb{E} \ell(\hat{Z}, Z^*) &\leq \exp \left(- \left(1 - \bar{C}'''' \left(\left(\frac{\sigma^2 d^3}{np} \right)^{\frac{1}{4}} + \left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}} \right) \right) \frac{np}{2\sigma^2} \right) \\ &\quad + n^{-8}. \end{aligned}$$

Theorem 3 is a non-asymptotic and refined version of Theorem 1 stated in the introduction. By letting $np/(\sigma^2 d^3)$ and $np/(d^2 \log n)$ go to infinity, the exponential term in Theorem 3 takes an asymptotic form of $\exp(-(1-o(1))\frac{np}{2\sigma^2})$ and matches with the minimax lower bound. In this way, we establish the statistical optimality of the proposed spectral method. In Theorem 3, d is allowed to grow with n but not too fast. Reference [6] considers the full observation setting where $p = 1$ and studies an EM-like iterative algorithm, requiring that $n/(\sigma^2 d \log d) \rightarrow \infty$. Our conditions are more stringent, reflecting the added complexity introduced by missing data, which necessitates stronger assumptions to apply concentration inequalities effectively.

While the proof of Theorem 3 is complicated, one can still develop an intuition as to why we can achieve the error bound $\exp(-(1-o(1))\frac{np}{2\sigma^2})$. At a high level, we will prove and use a variant of the linear approximation [16] $UH - U^* \approx XU^* \Lambda^{-1} - U^*$ where Λ is the diagonal matrix of the leading eigenvalues of X (see (18)). A further decomposition using the structure of X reveals that

$$U_j H - U_j^* \approx \frac{\sigma}{\sqrt{n}} \sum_{k \neq j} A_{jk} W_{jk} \Lambda^{-1} \approx \frac{1}{\sqrt{n}} \frac{\sigma}{np} \sum_{k \neq j} A_{jk} W_{jk}. \quad (16)$$

Recall that $U_j^* = I_d / \sqrt{n}$ according to (10). Then $\sqrt{n} U_j H - I_d \approx \frac{\sigma}{np} \sum_{k \neq j} A_{jk} W_{jk}$ which follows a Gaussian distribution conditioned on $\{A_{jk}\}_{k \neq j}$. Since $\sum_{k \neq j} A_{jk}$ concentrates around $(n-1)p$, roughly speaking, $\sqrt{n} U_j H - I_d$ is a random matrix with each entry i.i.d. following $\mathcal{N}(0, \frac{\sigma^2}{np})$. The Gaussian tail is used to characterize the probability of the event when $\sqrt{n} U_j H$ considerably deviates away from I_d and eventually leads to a probability bound of $\exp(-(1-o(1))\frac{np}{2\sigma^2})$.

C. Block-Wise Decomposition of $UH - U^*$

In this section, we provide a block-wise decomposition of $UH - U^*$. The decomposition is the key towards the block-wise analysis in Section III-B and provides insights on how Theorem 2 and Theorem 3 are established. Recall that we let $Z_j^* = I_d$ for all $j \in [n]$. Then, (4) becomes

$$X = (A \otimes I_d) + \sigma(A \otimes J_d) \circ W.$$

For each $j \in [n]$, define $X_j := (X_{j1}, \dots, X_{jn}) \in \mathbb{R}^{d \times nd}$ to be the j -th block row of X and define $W_j \in \mathbb{R}^{d \times nd}$ analogously for W . Then we have

$$X_j = (A_j \otimes I_d) + \sigma(A_j \otimes J_d) \circ W_j. \quad (17)$$

As a result, $X = (X_1^\top, \dots, X_n^\top)^\top$. Define $\Lambda \in \mathbb{R}^{d \times d}$ to be the diagonal matrix of the leading eigenvalues of X . That is, $\Lambda_{ii} := \lambda_i(X)$ and $\Lambda_{ik} := 0$ for all $1 \leq i \neq k \leq d$. Then we have

$$\begin{aligned} UH\Lambda - XU^* &= U(H\Lambda - \Lambda H) + U\Lambda H - XU^* \\ &= U(H\Lambda - \Lambda H) + XUH - XU^* \\ &= U(H\Lambda - \Lambda H) + X(UH - U^*). \end{aligned}$$

Multiplying both sides by Λ^{-1} and rearranging the terms, we have

$$\begin{aligned} UH - U^* &= U(H\Lambda - \Lambda H)\Lambda^{-1} + X(UH - U^*)\Lambda^{-1} \\ &\quad + XU^*\Lambda^{-1} - U^*. \end{aligned} \quad (18)$$

The above display involves $X(UH - U^*)$ where X and $UH - U^*$ are dependent on each other. To decouple the dependence, we approximate $UH - U^*$ by its leave-one-out counterparts. Consider any $j \in [n]$. Define $X^{(j)} \in \mathbb{R}^{nd \times nd}$ such that

$$X_{ik}^{(j)} := \begin{cases} X_{ik}, & \forall i, k \neq j, \\ 0_{d \times d}, & \text{otherwise.} \end{cases}$$

In addition, let $U^{(j)} \in \mathbb{R}^{nd \times d}$ be the matrix including the leading d eigenvectors of $X^{(j)}$. As a consequence, $X^{(j)}, U^{(j)}$ are independent of $\{A_{jk}\}_{k \neq j}$ and $\{W_{jk}\}_{k \neq j}$. Define $H^{(j)} := U^{(j)\top} U^* \in \mathbb{R}^{d \times d}$. Then

$$\begin{aligned} UH - U^* &= UH - U^{(j)} H^{(j)} + U^{(j)} H^{(j)} - U^* \\ &= (UU^\top - U^{(j)} U^{(j)\top}) U^* + U^{(j)} H^{(j)} - U^*. \end{aligned}$$

After plugging it into the right-hand side of (18), we have

$$\begin{aligned} UH - U^* &= U(H\Lambda - \Lambda H)\Lambda^{-1} + X(UU^\top - U^{(j)} U^{(j)\top}) U^* \Lambda^{-1} \\ &\quad + X(U^{(j)} H^{(j)} - U^*) \Lambda^{-1} + XU^* \Lambda^{-1} - U^*. \end{aligned}$$

Then, the j th block matrix of $UH - U^*$ satisfies

$$\begin{aligned} U_j H - U_j^* &= U_j (H\Lambda - \Lambda H)\Lambda^{-1} + X_j (UU^\top - U^{(j)} U^{(j)\top}) U^* \Lambda^{-1} \\ &\quad \underbrace{\hspace{10em}}_{=: B_j} \\ &\quad + X_j (U^{(j)} H^{(j)} - U^*) \Lambda^{-1} + X_j U^* \Lambda^{-1} - U_j^*. \end{aligned} \quad (19)$$

The last two terms in (19) can be further decomposed. Using (17), we have

$$\begin{aligned} &X_j (U^{(j)} H^{(j)} - U^*) \Lambda^{-1} \\ &= ((A_j \otimes I_d) + \sigma(A_j \otimes J_d) \circ W_j) (U^{(j)} H^{(j)} - U^*) \Lambda^{-1} \\ &= \underbrace{\sum_{k \neq j} A_{jk} (U_k^{(j)} H^{(j)} - U_k^*) \Lambda^{-1}}_{=: F_{j1}} \\ &\quad + \underbrace{\sigma \sum_{k \neq j} A_{jk} W_{jk} (U_k^{(j)} H^{(j)} - U_k^*) \Lambda^{-1}}_{=: F_{j2}}, \end{aligned}$$

and

$$\begin{aligned} X_j U^* \Lambda^{-1} - U_j^* &= ((A_j \otimes I_d) + \sigma(A_j \otimes J_d) \circ W_j) U^* \Lambda^{-1} - U_j^* \\ &= \left(\sum_{k \neq j} A_{jk} U_k^* \right) \Lambda^{-1} - U_j^* + \sigma \sum_{k \neq j} A_{jk} W_{jk} U_k^* \Lambda^{-1} \\ &= \underbrace{\frac{1}{\sqrt{n}} \left(\left(\sum_{k \neq j} A_{jk} \right) \Lambda^{-1} - I_d \right)}_{=: G_{j1}} + \underbrace{\frac{\sigma}{\sqrt{n}} \sum_{k \neq j} A_{jk} W_{jk} \Lambda^{-1}}_{=: G_{j2}}, \end{aligned}$$

where the last equation is due to Lemma 10. As a result, we have a decomposition of $U_j H - U_j^*$

$$U_j H - U_j^* = B_j + F_{j1} + F_{j2} + G_{j1} + G_{j2}, \quad (20)$$

holds for all $j \in [n]$.

The block-wise decomposition (20) of $UH - U^*$ is the starting point to establishing both Theorem 2 and Theorem 3. Note that we have a mutual independence among $\{A_{jk}\}_{k \neq i}$, $\{W_{jk}\}_{k \neq i}$, and $U^{(j)}H^{(j)} - U^*$ in the definitions of F_{j1} and F_{j2} , which is crucial to obtaining sharp bounds and tail probabilities for them. Theorem 2 is proved by establishing upper bounds for the operator norm of $B_j, F_{j1}, F_{j2}, G_{j1}$, and G_{j2} . By further analyzing their tail bounds, we establish Theorem 3. Among these terms, G_{j2} is the one contributing to the exponential error bound in Theorem 3, as we illustrate in (16).

D. Approximate Clustering

Algorithm 1 involves a d -means clustering (7). There exists an $O(n^{O(d^2)})$ algorithm using a weighted Voronoi diagram [19] that finds the globally optimal solution. In the case where d is a constant, this algorithm enjoys polynomial time complexity. However, this approach quickly becomes impractical as d grows. In fact, solving the general d -means optimization problem exactly is NP-hard [20], [21]. A practical solution is to use an approximate algorithm which guarantees a $(1 + \epsilon)$ -optimal solution to (7) under a polynomial time constraint. That is, let $\check{\mu}_1, \dots, \check{\mu}_d \in \mathbb{R}^{d \times 1}$ and $\check{z} \in [d]^{nd}$ denote a $(1 + \epsilon)$ -approximate solution. That is, they satisfy

$$\begin{aligned} & \sum_{j=1}^{nd} \|v_j - \check{\mu}_{z_j}\|^2 \\ & \leq (1 + \epsilon) \min_{\mu_1, \dots, \mu_d \in \mathbb{R}^{d \times 1}} \min_{z \in [d]^{nd}} \sum_{j=1}^{nd} \|v_j - \mu_{z_j}\|^2, \end{aligned} \quad (21)$$

Define an approximate anchor as

$$\check{M} := (\sqrt{n}\check{\mu}_1^\top, \dots, \sqrt{n}\check{\mu}_d^\top)^\top \in \mathbb{R}^{d \times d}. \quad (22)$$

We will refer to the anchor M used in Algorithm 1 as the exact anchor. Using the randomized d -means algorithm of [22], we obtain a $(1 + \epsilon)$ -approximate solution to (7) in $O(nd2^{(d/\epsilon)^{O(1)}})$ time, which can be considerably faster than the exact clustering approach. We also emphasize that the third step of Algorithm 1 is agnostic of the anchor. Therefore, we simply replace M by \check{M} with no change to the remaining of the algorithm, as shown in the following.

Algorithm 2 Spectral Method With Approximate Clustering.

Input: Data matrix $X \in \mathbb{R}^{nd \times nd}$, approximation factor $\epsilon > 0$

Output: n permutation matrices $\check{Z}_1, \check{Z}_2, \dots, \check{Z}_n \in \Pi_d$

- 1 Obtain the top d eigenvectors of X as $U \in \mathbb{R}^{nd \times d}$;
- 2 Denote the rows of U as $v_1, \dots, v_{nd} \in \mathbb{R}^{d \times 1}$. Run $(1 + \epsilon)$ -approximate d -means of clustering on v_1, \dots, v_{nd} and denote $\check{\mu}_1, \dots, \check{\mu}_d \in \mathbb{R}^{d \times 1}$ as the approximate cluster centers which satisfy (21);
- 3 Compute for each $j \in [n]$,

$$\check{Z}_j := \operatorname{argmax}_{P \in \Pi_d} \langle P, U_j \check{M}^\top \rangle, \quad (23)$$

where \check{M} is defined per (22).

With a small additional assumption regarding $(1 + \epsilon)$, in Theorem 4, we show that the theoretical guarantee of the modified spectral algorithm is the same as that of Algorithm 1.

The additional assumption $\frac{np}{\sigma^2 d(1+\epsilon)} \gtrsim 1$ is due to the use of the approximate anchor \check{M} instead of the exact one M . The proof of Theorem 4 is nearly identical to that of Theorem 3 with minor modifications.

Theorem 4: There exist constants $\check{C}_1, \check{C}_2, \check{C}_3, \check{C}_4, \check{C}_5 > 0$ such that if $\frac{np}{d^2 \log n} > \check{C}_1$, $\frac{np}{\sigma^2 d^3} > \check{C}_2$, $\frac{np}{\log^3 n} > \check{C}_3$, and $\frac{np}{\sigma^2 d} \geq \check{C}_4(1 + \epsilon)$, then the estimates $\check{Z}_1, \dots, \check{Z}_n$ of Algorithm 2 satisfy

$$\begin{aligned} & \mathbb{E} \ell(\check{Z}, Z^*) \\ & \leq \exp \left(- \left(1 - \check{C}_5 \left(\left(\frac{\sigma^2 d^3}{np} \right)^{\frac{1}{4}} + \left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}} \right) \right) \frac{np}{2\sigma^2} \right) \\ & \quad + n^{-8}. \end{aligned}$$

IV. PROOFS

A. Preliminaries and Useful Inequalities

We first present a lemma which enumerates a number of useful intermediate inequalities. Conditioned on certain inequalities, the statements of many subsequent lemmas and theorems hold deterministically. This allows us to present the proofs of the lemmas and theorems in more intuitive ways. The proof of Lemma 1 is deferred to a later section.

Lemma 1: There exist constants $C'_0 > 0, C_0 > 7$ such that if $\frac{np}{\log n} > C'_0$, then the following event holds

$$\begin{aligned} \mathcal{E} = & \left\{ \|A - \mathbb{E}A\|_{\text{op}} \leq C_0 \sqrt{np}, \right. \\ & \|(A \otimes I_d) \circ W\|_{\text{op}} \leq C_0 \sqrt{npd}, \\ & \left. \max_{j \in [n]} \left| \sum_{k \neq j} A_{jk} - np \right| \leq C_0 \sqrt{np \log n} \right\} \end{aligned}$$

with probability at least $1 - n^{-10}$. Under the event \mathcal{E} , we have

$$\begin{aligned} & \max \left\{ \|X - (\mathbb{E}A \otimes I_d)\|_{\text{op}}, \max_{i \in [d]} |\lambda_i(X) - (n-1)p|, \right. \\ & \quad \left. \max_{d+1 \leq i \leq nd} |\lambda_i(X) + p| \right\} \leq C_0 (1 + \sigma \sqrt{d}) \sqrt{np}, \end{aligned} \quad (24)$$

$$\max_{j \in [n]} \|X_j\|_{\text{op}} \leq p \sqrt{n} + C_0 (1 + \sigma \sqrt{d}) \sqrt{np}. \quad (25)$$

If $\frac{np}{(1 + \sigma \sqrt{d})^2} \geq 64C_0^2$ is further assumed, under the event \mathcal{E} , the following hold for Λ, U , and H :

$$\|\Lambda\|_{\text{op}} = \lambda_1(X) \leq \frac{9np}{8}, \quad (26)$$

$$\|\Lambda^{-1}\|_{\text{op}}^{-1} = \lambda_d(X) \geq \frac{7np}{8}, \quad (27)$$

$$\lambda_d(X) - \lambda_{d+1}(X) \geq \frac{3np}{4}, \quad (28)$$

$$\max \left\{ \min_{O \in \mathcal{O}_d} \|H - O\|_{\text{op}}, \|UH - U^*\|_{\text{op}}, \right.$$

$$\left. \|UU^\top - U^*U^{*\top}\|_{\text{op}} \right\} \leq \frac{8C_0(1 + \sigma \sqrt{d})}{7\sqrt{np}}, \quad (29)$$

$$\|H\Lambda - \Lambda H\|_{\text{op}} \leq 2C_0(1 + \sigma \sqrt{d}) \sqrt{np}, \quad (30)$$

$$\|H^{-1}\|_{\text{op}} \leq \frac{4}{3}, \quad (31)$$

the following hold any $j \in [n]$:

$$\|G_{j1}\|_{\text{op}} \leq \frac{2C_0}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma\sqrt{d}}{\sqrt{np}} \right), \quad (32)$$

$$\|U_j\|_{\text{op}} \leq \frac{4}{3} \left(\|U_j H - U_j^*\|_{\text{op}} + \frac{1}{\sqrt{n}} \right), \quad (33)$$

$$\|U^{(j)}(U^{(j)})^\top - UU^\top\|_{\text{op}} \leq 6 \left(\|U_j H - U_j^*\|_{\text{op}} + \frac{1}{\sqrt{n}} \right), \quad (34)$$

$$\|U^{(j)}H^{(j)} - U^*\|_{\text{op}} \leq \frac{9C_0(1 + \sigma\sqrt{d})}{\sqrt{np}}, \quad (35)$$

$$\|U^{(j)}H^{(j)} - U^{*(j)}\|_{\text{op},\infty} \leq 7 \left(\|UH - U^*\|_{\text{op},\infty} + \frac{1}{\sqrt{n}} \right). \quad (36)$$

B. Proof of Proposition 1

We first state and prove a deterministic version of Proposition 1. The proof of Proposition 1 follows from a simple probabilistic argument.

Lemma 2: Assume that (29) holds. If $\frac{np}{(\sqrt{d} + \sigma d)^2} \geq 32C_0^2$, we have

$$\min_{P \in \Pi_d} \|M - PH^\top\|_{\text{F}} \leq \frac{4C_0(\sqrt{d} + \sigma d)}{\sqrt{np}}. \quad (37)$$

Proof: Note that

$$\begin{aligned} \|U - U^*H^\top\|_{\text{op}} &= \|UU^\top U - U^*U^{*\top}U\|_{\text{op}} \\ &\leq \|UU^\top - U^*U^{*\top}\|_{\text{op}} \|U\|_{\text{op}} \\ &\leq \|UU^\top - U^*U^{*\top}\|_{\text{op}}. \end{aligned}$$

Since (29) holds, we have

$$\begin{aligned} \|U - U^*H^\top\|_{\text{F}} &\leq \sqrt{d} \|U - U^*H^\top\|_{\text{op}} \\ &\leq \sqrt{d} \frac{8C_0(1 + \sigma\sqrt{d})}{7\sqrt{np}}. \end{aligned}$$

Note that $\sqrt{n}U^*$ has only k unique rows $e_1^\top, \dots, e_d^\top$ and each is of size n . Then $\sqrt{n}U^*H^\top$ also has k unique rows $e_1^\top H^\top, \dots, e_d^\top H^\top$ and each is of size n . From (29), there exists a matrix $O \in \mathcal{O}_d$ such that $\|H - O\|_{\text{op}} \leq \frac{8C_0(1 + \sigma\sqrt{d})}{7\sqrt{np}} \leq \frac{1}{7}$. Then

$$\begin{aligned} &\min_{a,b \in [d]} \|(e_a - e_b)^\top H^\top\| \\ &\geq \min_{a,b \in [d]} \|(e_a - e_b)^\top (O^\top + (H - O)^\top)\| \\ &\geq \min_{a,b \in [d]} \|(e_a - e_b)^\top O^\top\| - \|H - O\|_{\text{op}} \\ &\geq \sqrt{2} - \frac{1}{7} \\ &> \frac{6}{5}. \end{aligned}$$

That is, the minimum distance among the unique rows of U^*H^\top , is at least $\frac{6}{5\sqrt{n}}$.

Let $\hat{z} \in [d]^{nd}$ be the minimizer of (7) with $\{\hat{\mu}_1, \dots, \hat{\mu}_d\}$. Denote $\hat{\Theta} := (\hat{\mu}_{\hat{z}_1}^\top, \dots, \hat{\mu}_{\hat{z}_{nd}}^\top)^\top \in \mathbb{R}^{nd \times d}$. According to (7), we have $\|\hat{\Theta} - U\|_{\text{F}} \leq \|U^*H^\top - U\|_{\text{F}}$. Hence,

$$\|\hat{\Theta} - U^*H^\top\|_{\text{F}} \leq \|\hat{\Theta} - U\|_{\text{F}} + \|U^*H^\top - U\|_{\text{F}}$$

$$\begin{aligned} &\leq 2 \|U - U^*H^\top\|_{\text{F}} \\ &\leq \sqrt{d} \frac{16C_0(1 + \sigma\sqrt{d})}{7\sqrt{np}}. \end{aligned} \quad (38)$$

Define $z^* \in [d]^{nd}$ such that the i th row of U^*H^\top is equal to $e_{z_i^*}^\top H^\top / \sqrt{n}$ for each $i \in [n]$. Define the set S as

$$S := \left\{ i \in [nd] : \|\hat{\mu}_{z_i} - e_{z_i^*}^\top H^\top / \sqrt{n}\| > \frac{3}{5\sqrt{n}} \right\}.$$

Then we have

$$|S| \leq \frac{\|\hat{\Theta} - U^*H^\top\|_{\text{F}}^2}{\left(\frac{3}{5\sqrt{n}}\right)^2} \leq \frac{(4C_0(\sqrt{d} + \sigma d))^2}{p}.$$

Under the assumption $\frac{np}{(\sqrt{d} + \sigma d)^2} \geq 32C_0^2$, we have $|S| \leq n/2$. Then by the same argument as in the proof of Proposition 3.1 of [23], there exists a bijection $\phi : [d] \rightarrow [d]$ such that $\hat{z}_i = \phi(z_i^*)$ for all $i \notin S$. Hence, for each $a \in [d]$, we have

$$\begin{aligned} &\|\hat{\mu}_{\phi(a)} - e_a^\top H^\top / \sqrt{n}\|^2 \\ &= \frac{\sum_{i \in [nd]: \hat{z}_i = \phi(a), z_i^* = a} \|\hat{\mu}_{z_i} - e_{z_i^*}^\top H^\top\|^2}{|i \in [nd] : \hat{z}_i = \phi(a), z_i^* = a|} \\ &\leq \frac{\sum_{i \in [nd]: \hat{z}_i = \phi(a)} \|\hat{\mu}_{z_i} - e_{z_i^*}^\top H^\top\|^2}{n - |S|} \\ &\leq \frac{\sum_{i \in [nd]: \hat{z}_i = \phi(a)} \|\hat{\mu}_{z_i} - e_{z_i^*}^\top H^\top\|^2}{n/2}. \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{a \in [d]} \|\hat{\mu}_{\phi(a)} - e_a^\top H^\top / \sqrt{n}\|^2 &\leq \frac{\sum_{i \in [nd]} \|\hat{\mu}_{z_i} - e_{z_i^*}^\top H^\top\|^2}{n/2} \\ &= \frac{\|\hat{\Theta} - U^*H^\top\|_{\text{F}}^2}{n/2} \\ &\leq \frac{(4C_0(\sqrt{d} + \sigma d))^2}{n^2 p}. \end{aligned}$$

That is, there exists a permutation matrix P such that

$$\|M - PH^\top\|_{\text{F}}^2 \leq \frac{(4C_0(\sqrt{d} + \sigma d))^2}{np},$$

since rows of M are $\sqrt{n}\hat{\mu}_1, \dots, \sqrt{n}\hat{\mu}_d$. \square

Proof of Proposition 1: According to Lemma 1, there exist constants $C, C_0 > 0$ such that if $\frac{np}{\log n} > C$ and $\frac{np}{(1 + \sigma\sqrt{d})^2} \geq 64C_0^2$, we have (29) hold with probability at least $1 - n^{-10}$. The proof is complete by Lemma 2. \square

The following lemma establishes the accuracy of the anchor constructed using approximate clustering in Algorithm 2. It serves as a counterpart of Lemma 2 and will be used in the proof of Theorem 4. Its proof is nearly identical to that of Lemma 2 with minor modifications.

Lemma 3: Assume that (29) holds. If $\frac{np}{(\sqrt{d}+\sigma d)^2} \geq 32(1+\epsilon)C_0^2$, we have

$$\min_{P \in \Pi_d} \|\check{M} - PH^\top\|_F \leq \frac{4C_0 \sqrt{1+\epsilon} (\sqrt{d} + \sigma d)}{\sqrt{np}}. \quad (39)$$

Proof: We follow a similar argument to that in the proof of Lemma 2. The difference starts from the derivation of (38). Define $\check{\Theta} = (\check{\mu}_{z_1}^\top, \dots, \check{\mu}_{z_{nd}}^\top)^\top$ to be the assigned centers given by the approximate clustering algorithm. Then, by (21), we have $\|\check{\Theta} - U\|_F^2 \leq (1+\epsilon) \|U^* H^\top - U\|_F^2$. Hence,

$$\begin{aligned} \|\check{\Theta} - U^* H^\top\|_F &\leq \sqrt{1+\epsilon} \|\check{\Theta} - U\|_F + \|U^* H^\top - U\|_F \\ &\leq 2\sqrt{1+\epsilon} \|U - U^* H^\top\|_F \\ &\leq \sqrt{1+\epsilon} \frac{16C_0 \sqrt{d} (1+\sigma \sqrt{d})}{7\sqrt{np}}. \end{aligned}$$

The rest of the proof proceeds similarly, with a minor modification of the condition $\frac{np}{(\sqrt{d}+\sigma d)^2} \geq 32C_0^2$ to $\frac{np}{(\sqrt{d}+\sigma d)^2} \geq 32C_0^2(1+\epsilon)$. \square

C. Proof of Theorem 2

We first give a deterministic upper bound for $\|U_j H - U_j^*\|_{\text{op}}$, using the decomposition (20).

Lemma 4: Assume (24)–(35) hold. Under the assumption $\frac{np}{(1+\sigma \sqrt{d})^2} \geq 22^2 C_0^2$, for each $j \in [n]$, we have

$$\begin{aligned} \|U_j H - U_j^*\|_{\text{op}} &\leq \frac{22C_0}{\sqrt{n}} \left(\frac{1+\sigma \sqrt{d}}{\sqrt{np}} \right) + 2 \|F_{j1}\|_{\text{op}} \\ &\quad + 2 \|F_{j2}\|_{\text{op}} + 2 \|G_{j1}\|_{\text{op}} + 2 \|G_{j2}\|_{\text{op}}, \end{aligned} \quad (40)$$

and

$$\begin{aligned} \|B_j\|_{\text{op}} &\leq 22C_0 \left(\frac{1+\sigma \sqrt{d}}{\sqrt{np}} \right) \\ &\quad \times \left(\frac{1}{\sqrt{n}} + \|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} + \|G_{j1}\|_{\text{op}} + \|G_{j2}\|_{\text{op}} \right). \end{aligned} \quad (41)$$

Proof: Consider any $j \in [n]$. We have

$$\begin{aligned} \|B_j\|_{\text{op}} &\leq \|U_j\|_{\text{op}} \|H\Lambda - \Lambda H\|_{\text{op}} \|\Lambda^{-1}\|_{\text{op}} \\ &\quad + \|X_j\|_{\text{op}} \|UU^\top - U^{(j)}U^{(j)\top}\|_{\text{op}} \|\Lambda^{-1}\|_{\text{op}} \\ &\leq \frac{4}{3} \left(\|U_j H - U_j^*\|_{\text{op}} + \frac{1}{\sqrt{n}} \right) (2C_0 (1+\sigma \sqrt{d}) \sqrt{np}) \frac{8}{7np} \\ &\quad + (p\sqrt{n} + C_0 (1+\sigma \sqrt{d}) \sqrt{np}) \\ &\quad \times \left(6 \left(\|U_j H - U_j^*\|_{\text{op}} + \frac{1}{\sqrt{n}} \right) \right) \frac{8}{7np} \\ &\leq \frac{8}{7np} \left(\|U_j H - U_j^*\|_{\text{op}} + \frac{1}{\sqrt{n}} \right) \\ &\quad \times \left(6p\sqrt{n} + \frac{26}{3} C_0 (1+\sigma \sqrt{d}) \sqrt{np} \right) \\ &\leq 11C_0 \left(\|U_j H - U_j^*\|_{\text{op}} + \frac{1}{\sqrt{n}} \right) \left(\frac{1+\sigma \sqrt{d}}{\sqrt{np}} \right), \end{aligned}$$

where the second inequality is by (24)–(35) and the last inequality is by $C_0 > 7$. From (20), we also have

$$\begin{aligned} \|U_j H - U_j^*\|_{\text{op}} &\leq \|B_j\|_{\text{op}} + \|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} \\ &\quad + \|G_{j1}\|_{\text{op}} + \|G_{j2}\|_{\text{op}}. \end{aligned}$$

Plug the upper bound (41) of $\|B_j\|_{\text{op}}$ into the above display. After rearrangement, we have

$$\begin{aligned} &\left(1 - 11C_0 \left(\frac{1+\sigma \sqrt{d}}{\sqrt{np}} \right) \right) \|U_j H - U_j^*\|_{\text{op}} \\ &\leq \frac{11C_0}{\sqrt{n}} \left(\frac{1+\sigma \sqrt{d}}{\sqrt{np}} \right) + \|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} \\ &\quad + \|G_{j1}\|_{\text{op}} + \|G_{j2}\|_{\text{op}}. \end{aligned}$$

Under the assumption $\frac{np}{(1+\sigma \sqrt{d})^2} \geq 22^2 C_0^2$, we have

$$\begin{aligned} \|U_j H - U_j^*\|_{\text{op}} &\leq \frac{22C_0}{\sqrt{n}} \left(\frac{1+\sigma \sqrt{d}}{\sqrt{np}} \right) + 2 \|F_{j1}\|_{\text{op}} \\ &\quad + 2 \|F_{j2}\|_{\text{op}} + 2 \|G_{j1}\|_{\text{op}} + 2 \|G_{j2}\|_{\text{op}}. \end{aligned}$$

Plugging it into the upper bound of $\|B_j\|_{\text{op}}$, we have

$$\begin{aligned} \|B_j\|_{\text{op}} &\leq 11C_0 \left(\left(\frac{22C_0}{\sqrt{n}} \left(\frac{1+\sigma \sqrt{d}}{\sqrt{np}} \right) + 2 \|F_{j1}\|_{\text{op}} + 2 \|F_{j2}\|_{\text{op}} \right. \right. \\ &\quad \left. \left. + 2 \|G_{j1}\|_{\text{op}} + 2 \|G_{j2}\|_{\text{op}} \right) + \frac{1}{\sqrt{n}} \right) \left(\frac{1+\sigma \sqrt{d}}{\sqrt{np}} \right) \\ &\leq 22C_0 \left(\frac{1}{\sqrt{n}} + \|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} + \|G_{j1}\|_{\text{op}} \right. \\ &\quad \left. + \|G_{j2}\|_{\text{op}} \right) \left(\frac{1+\sigma \sqrt{d}}{\sqrt{np}} \right). \end{aligned}$$

\square

1) Useful Short-Hand Notations: In our proofs, some key quantities are repeated and some are lexicographically cumbersome. In order to declutter the presentation, wherever convenient without sacrificing clarity, we will use the following shorthand notations. Define

$$\begin{aligned} \Delta &:= UH - U^* \in \mathbb{R}^{nd \times d}, \\ \Delta^{(j)} &:= U^{(j)}H^{(j)} - U^* \in \mathbb{R}^{nd \times d}, \forall j \in [n], \end{aligned} \quad (42)$$

where $H^{(j)} = U^{(j)\top}U^*$, such that block submatrices $\Delta_k = U_k H - U_k^*$ and $\Delta_k^{(j)} = U_k^{(j)}H^{(j)} - U_k^*$ for each $j, k \in [n]$. We further introduce $\|\cdot\|_{\text{op},\infty}$ norm such that

$$\|\Delta\|_{\text{op},\infty} := \max_{k \in [n]} \|\Delta_k\|_{\text{op}}, \quad \|\Delta^{(j)}\|_{\text{op},\infty} := \max_{k \in [n]} \|\Delta_k^{(j)}\|_{\text{op}}. \quad (43)$$

2) Helper Tail Bound Inequalities and Proof Sketch: As a preview, the terms $\|F_{j1}\|_{\text{op}}$, $\|F_{j2}\|_{\text{op}}$, $\|G_{j1}\|_{\text{op}}$ and $\|G_{j2}\|_{\text{op}}$ in (40) and (41) can be bounded using the helper Lemma 5, Lemma 6, Lemma 7 and Lemma 8 below. We defer the proof of these helper lemmas to a later section. Note that all of these inequalities require $\|\Delta^{(j)}\|_{\text{op},\infty}$ which in turn can be bounded in terms of $\|\Delta\|_{\text{op},\infty}$ using (36). The key strategy is to construct an inequality in which $\|\Delta\|_{\text{op},\infty}$ appears on both the left and

the right side of the inequality. After some manipulation and under appropriate assumptions, we reduce this to an inequality where the left side of the inequality consists of only

$$\|\Delta\|_{\text{op},\infty}.$$

Lemma 5: For any $j \in [n]$, both of the following tail bounds hold for any $t > 0$.

$$\mathbb{P}\left(\left\|\sum_{k \neq j} A_{jk} \Delta_k^{(j)}\right\|_{\text{op}} \geq p \sqrt{n} \|\Delta^{(j)}\|_{\text{op}} + t \left|\Delta^{(j)}\right|\right) \leq 2d \times \exp\left(-\frac{t^2/2}{p \sqrt{nd} \|\Delta^{(j)}\|_{\text{op}} \|\Delta^{(j)}\|_{\text{op},\infty} + \|\Delta^{(j)}\|_{\text{op},\infty} t/3}\right), \quad (44)$$

and

$$\mathbb{P}\left(\left\|\sum_{k \neq j} A_{jk} \Delta_k^{(j)}\right\|_{\text{op}} \geq p \sqrt{n} \|\Delta^{(j)}\|_{\text{op}} + t \left|\Delta^{(j)}\right|\right) \leq 2d \exp\left(-\frac{t^2/2}{np \|\Delta^{(j)}\|_{\text{op},\infty}^2 + \|\Delta^{(j)}\|_{\text{op},\infty} t/3}\right). \quad (45)$$

Lemma 6: For any $j \in [n]$, we have

$$\mathbb{P}\left(\left\|\sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)}\right\|_{\text{op}} \geq p \|\Delta^{(j)}\|_{\text{op}}^2 + t \left|\Delta^{(j)}\right|\right) \leq 2d \exp\left(-\frac{t^2/2}{p \|\Delta^{(j)}\|_{\text{op},\infty}^2 \|\Delta^{(j)}\|_{\text{op}}^2 + \|\Delta^{(j)}\|_{\text{op},\infty}^2 t/3}\right),$$

for any $t \geq 0$.

Lemma 7: There exists some constant $c > 0$, such that for any $t \geq 4\sqrt{d} \left\|\sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)}\right\|_{\text{op}}^{\frac{1}{2}}$ and for any $j \in [n]$, we have

$$\mathbb{P}\left(\left\|\sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)}\right\|_{\text{op}} \geq t \left|\{A_{jk}\}_{k \neq j}, \Delta^{(j)}\right|\right) \leq 2 \exp\left(-\frac{ct^2}{\left\|\sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)}\right\|_{\text{op}}}\right).$$

Lemma 8 (Corollary 7.3.3 of [24]) Let W be a $d \times d$ matrix with independent $\mathcal{N}(0, 1)$ entries. Then there exists some constant $c > 0$ such that for every $t \geq 0$, we have

$$\mathbb{P}\left(\|W\|_{\text{op}} \geq 2\sqrt{d} + t\right) \leq 2 \exp(-ct^2).$$

We are now ready to prove Theorem 2.

Proof of Theorem 2: From Lemma 1, there exist constants $C'_0, C_0 > 0$ such that if $\frac{np}{\log n} > C'_0$, then the event \mathcal{E} holds with probability at least $1 - n^{-10}$. Assume \mathcal{E} holds and $\frac{np}{(\sqrt{d} + \sigma d)^2} \geq 22^2 C_0^2$. Then according to Lemma 1, we have (24)–(36) hold. As a consequence, by Lemma 4, we have (40) hold as well.

Consider any $j \in [n]$. Note that (40) can be written as

$$\|\Delta_j\|_{\text{op}} \leq \frac{22C_0}{\sqrt{n}} \left(\frac{1 + \sigma \sqrt{d}}{\sqrt{np}}\right) + 2 \|F_{j1}\|_{\text{op}} + 2 \|F_{j2}\|_{\text{op}} + 2 \|G_{j1}\|_{\text{op}} + 2 \|G_{j2}\|_{\text{op}}. \quad (46)$$

Hence, to upper bound $\|\Delta_j\|_{\text{op}}$, we need to study $\|F_{j1}\|_{\text{op}}$, $\|F_{j2}\|_{\text{op}}$, $\|G_{j1}\|_{\text{op}}$, and $\|G_{j2}\|_{\text{op}}$.

Bounding $\|F_{j1}\|_{\text{op}}$. By (45) of Lemma 5, the following holds with probability at least $1 - 2dn^{-10}$

$$\begin{aligned} & \left\|\sum_{k \neq j} A_{jk} \Delta_k^{(j)}\right\|_{\text{op}} \\ & \leq p \sqrt{n} \|\Delta^{(j)}\|_{\text{op}} + \sqrt{40np \|\Delta^{(j)}\|_{\text{op},\infty}^2 \log n} \\ & \quad + \frac{40}{3} \log n \|\Delta^{(j)}\|_{\text{op},\infty} \\ & \leq p \sqrt{n} \|\Delta^{(j)}\|_{\text{op}} \\ & \quad + \left(\sqrt{40np \log n} + \frac{40}{3} \log n\right) \|\Delta^{(j)}\|_{\text{op},\infty}. \end{aligned}$$

Then with (27), we have

$$\begin{aligned} \|F_{j1}\|_{\text{op}} & \leq \left\|\sum_{k \neq j} A_{jk} \Delta_k^{(j)}\right\|_{\text{op}} \|\Lambda^{-1}\|_{\text{op}} \\ & \leq \frac{8}{7np} \left(p \sqrt{n} \|\Delta^{(j)}\|_{\text{op}}\right. \\ & \quad \left.+ \left(\sqrt{40np \log n} + \frac{40}{3} \log n\right) \|\Delta^{(j)}\|_{\text{op},\infty}\right) \\ & \leq \frac{C_1}{\sqrt{n}} \|\Delta^{(j)}\|_{\text{op}} + C_1 \sqrt{\frac{\log n}{np}} \|\Delta^{(j)}\|_{\text{op},\infty}, \quad (47) \end{aligned}$$

for some constant $C_1 > 0$.

Bounding $\|F_{j2}\|_{\text{op}}$. By Lemma 7, we have

$$\left\|\sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)}\right\|_{\text{op}} \leq C_2 \sqrt{\log n} \left\|\sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)}\right\|_{\text{op}}^{\frac{1}{2}}$$

holds with probability at least $1 - n^{-10}$ for some constant $C_2 > 0$. In addition, by Lemma 6, we have

$$\begin{aligned} & \left\|\sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)}\right\|_{\text{op}} \\ & \leq p \|\Delta^{(j)}\|_{\text{op}}^2 + \sqrt{40p \|\Delta^{(j)}\|_{\text{op}}^2 \|\Delta^{(j)}\|_{\text{op},\infty}^2 \log n} \\ & \quad + \frac{40}{3} \log n \|\Delta^{(j)}\|_{\text{op},\infty}^2 \\ & \leq p \|\Delta^{(j)}\|_{\text{op}}^2 + \sqrt{40np \|\Delta^{(j)}\|_{\text{op},\infty}^4 \log n} \\ & \quad + \frac{40}{3} \log n \|\Delta^{(j)}\|_{\text{op},\infty}^2 \\ & \leq p \|\Delta^{(j)}\|_{\text{op}}^2 + \left(\sqrt{40np \log n} + \frac{40}{3} \log n\right) \|\Delta^{(j)}\|_{\text{op},\infty}^2 \end{aligned}$$

holds with probability at least $1 - 2dn^{-10}$. Then with (27), we have

$$\begin{aligned} & \|F_{j2}\|_{\text{op}} \\ & \leq \sigma \left\|\sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)}\right\|_{\text{op}} \|\Lambda^{-1}\|_{\text{op}} \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{8\sigma}{7np} C_2 \sqrt{\log n} \left(p \|\Delta^{(j)}\|_{\text{op}}^2 \right. \\
 &\quad \left. + \left(\sqrt{40np \log n} + \frac{40}{3} \log n \right) \|\Delta^{(j)}\|_{\text{op},\infty}^2 \right)^{\frac{1}{2}} \\
 &\leq \frac{8\sigma}{7np} C_2 \sqrt{\log n} \left(\sqrt{p} \|\Delta^{(j)}\|_{\text{op}} \right. \\
 &\quad \left. + \sqrt{\sqrt{40np \log n} + \frac{40}{3} \log n} \|\Delta^{(j)}\|_{\text{op},\infty} \right) \\
 &\leq \frac{C_3\sigma}{\sqrt{n}} \sqrt{\frac{\log n}{np}} \|\Delta^{(j)}\|_{\text{op}} + C_3\sigma \left(\frac{\log n}{np} \right)^{\frac{3}{4}} \|\Delta^{(j)}\|_{\text{op},\infty}, \quad (48)
 \end{aligned}$$

for some constant $C_3 > 0$.

For $\|G_{j1}\|$, its upper bound is given in (32).

Bounding $\|G_{j2}\|_{\text{op}}$. By Lemma 8, using the fact that $\{A_{jk}\}_{k \neq j}$ and $\{W_{jk}\}_{k \neq j}$ are independent, we have

$$\left\| \sum_{k \neq j} A_{jk} W_{jk} \right\|_{\text{op}} \leq \sqrt{\sum_{k \neq j} A_{jk}} \left(2\sqrt{d} + C_4 \sqrt{\log n} \right),$$

with probability at least $1 - n^{-10}$, for some constant $C_4 > 0$. Then

$$\begin{aligned}
 \|G_{j2}\|_{\text{op}} &\leq \frac{\sigma}{\sqrt{n}} \left\| \sum_{k \neq j} A_{jk} W_{jk} \right\|_{\text{op}} \|\Lambda^{-1}\|_{\text{op}} \\
 &\leq \frac{\sigma}{\sqrt{n}} \sqrt{\sum_{k \neq j} A_{jk}} \left(2\sqrt{d} + C_4 \sqrt{\log n} \right) \|\Lambda^{-1}\|_{\text{op}} \\
 &\leq C_5 \frac{\sigma}{\sqrt{n}} \left(\sqrt{\frac{d}{np}} + \sqrt{\frac{\log n}{np}} \right), \quad (49)
 \end{aligned}$$

for some constant $C_5 > 0$, where the last inequality is by \mathcal{E} and (27).

Putting things together. Plugging (47), (48), (32), and (49) into (46), we have

$$\begin{aligned}
 \|\Delta_j\|_{\text{op}} &\leq \frac{22C_0}{\sqrt{n}} \left(\frac{1 + \sigma\sqrt{d}}{\sqrt{np}} \right) \\
 &\quad + 2 \left(\frac{C_1}{\sqrt{n}} \|\Delta^{(j)}\|_{\text{op}} + C_1 \sqrt{\frac{\log n}{np}} \|\Delta^{(j)}\|_{\text{op},\infty} \right) \\
 &\quad + 2 \left(\frac{C_3\sigma}{\sqrt{n}} \sqrt{\frac{\log n}{np}} \|\Delta^{(j)}\|_{\text{op}} \right. \\
 &\quad \left. + C_3\sigma \left(\frac{\log n}{np} \right)^{\frac{3}{4}} \|\Delta^{(j)}\|_{\text{op},\infty} \right) \\
 &\quad + \frac{4C_0}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma\sqrt{d}}{\sqrt{np}} \right) \\
 &\quad + 2C_5 \frac{\sigma}{\sqrt{n}} \left(\sqrt{\frac{d}{np}} + \sqrt{\frac{\log n}{np}} \right) \\
 &\leq \frac{2}{\sqrt{n}} \left(C_1 + C_3\sigma \sqrt{\frac{\log n}{np}} \right) \|\Delta^{(j)}\|_{\text{op}} \\
 &\quad + 2 \left(C_1 \sqrt{\frac{\log n}{np}} + C_3\sigma \left(\frac{\log n}{np} \right)^{\frac{3}{4}} \right) \|\Delta^{(j)}\|_{\text{op},\infty}
 \end{aligned}$$

$$+ \frac{26C_0 + 2C_5}{\sqrt{n}} \frac{\sqrt{\log n} + \sigma\sqrt{d} + \sigma\sqrt{\log n}}{\sqrt{np}}.$$

By (36) and (35), we can replace the terms $\|\Delta^{(j)}\|_{\text{op}}$, $\|\Delta^{(j)}\|_{\text{op},\infty}$ with their respective upper bounds and have

$$\begin{aligned}
 \|\Delta_j\|_{\text{op}} &\leq \frac{2}{\sqrt{n}} \left(C_1 + C_3\sigma \sqrt{\frac{\log n}{np}} \right) \frac{9C_0(1 + \sigma\sqrt{d})}{\sqrt{np}} \\
 &\quad + 14 \left(C_1 \sqrt{\frac{\log n}{np}} + C_3\sigma \left(\frac{\log n}{np} \right)^{\frac{3}{4}} \right) \\
 &\quad \times \left(\|\Delta\|_{\text{op},\infty} + \frac{1}{\sqrt{n}} \right) \\
 &\quad + \frac{26C_0 + 2C_5}{\sqrt{n}} \frac{\sqrt{\log n} + \sigma\sqrt{d} + \sigma\sqrt{\log n}}{\sqrt{np}}.
 \end{aligned}$$

By a union bound, with probability at least $1 - 6dn^{-9}$, the above inequality holds for all $j \in [n]$. Then

$$\begin{aligned}
 \|\Delta\|_{\text{op},\infty} &\leq \frac{2}{\sqrt{n}} \left(C_1 + C_3\sigma \sqrt{\frac{\log n}{np}} \right) \frac{9C_0(1 + \sigma\sqrt{d})}{\sqrt{np}} \\
 &\quad + 14 \left(C_1 \sqrt{\frac{\log n}{np}} + C_3\sigma \left(\frac{\log n}{np} \right)^{\frac{3}{4}} \right) \\
 &\quad \times \left(\|\Delta\|_{\text{op},\infty} + \frac{1}{\sqrt{n}} \right) \\
 &\quad + \frac{26C_0 + 2C_5}{\sqrt{n}} \frac{\sqrt{\log n} + \sigma\sqrt{d} + \sigma\sqrt{\log n}}{\sqrt{np}}.
 \end{aligned}$$

After a rearrangement so that $\|\Delta\|_{\text{op},\infty}$ only appears on the left side, we have

$$\begin{aligned}
 &\left(1 - 14 \left(C_1 \sqrt{\frac{\log n}{np}} + C_3\sigma \left(\frac{\log n}{np} \right)^{\frac{3}{4}} \right) \right) \|\Delta\|_{\text{op},\infty} \\
 &\leq \frac{2}{\sqrt{n}} \left(C_1 + C_3\sigma \sqrt{\frac{\log n}{np}} \right) \frac{9C_0(1 + \sigma\sqrt{d})}{\sqrt{np}} \\
 &\quad + \frac{14}{\sqrt{n}} \left(C_1 \sqrt{\frac{\log n}{np}} + C_3\sigma \left(\frac{\log n}{np} \right)^{\frac{3}{4}} \right) \\
 &\quad + \frac{26C_0 + 2C_5}{\sqrt{n}} \frac{\sqrt{\log n} + \sigma\sqrt{d} + \sigma\sqrt{\log n}}{\sqrt{np}}.
 \end{aligned}$$

Then, there exists a constant $C > 0$ such that if $\frac{np}{(\sigma^{4/3}\sqrt{1})\log n} > C$, we have

$$14 \left(C_1 \sqrt{\frac{\log n}{np}} + C_3\sigma \left(\frac{\log n}{np} \right)^{\frac{3}{4}} \right) \leq \frac{1}{2}.$$

Consequently, we have

$$\begin{aligned}
 \|\Delta\|_{\text{op},\infty} &\leq \frac{4}{\sqrt{n}} \left(C_1 + C_3\sigma \sqrt{\frac{\log n}{np}} \right) \frac{9C_0(1 + \sigma\sqrt{d})}{\sqrt{np}} \\
 &\quad + \frac{28}{\sqrt{n}} \left(C_1 \sqrt{\frac{\log n}{np}} + C_3\sigma \left(\frac{\log n}{np} \right)^{\frac{3}{4}} \right) \\
 &\quad + \frac{52C_0 + 4C_5}{\sqrt{n}} \frac{\sqrt{\log n} + \sigma\sqrt{d} + \sigma\sqrt{\log n}}{\sqrt{np}}
 \end{aligned}$$

$$\leq \frac{C_6}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma \sqrt{d}}{\sqrt{np}} + \frac{\sigma \sqrt{\log n}}{\sqrt{np}} \right),$$

for some constant $C_6 > 0$. \square

D. Proofs of Theorem 3 and Theorem 4

For the proof of Theorem 3, we start with a proof sketch in Section IV-D1. This sketch illustrates that the proof can be divided into four distinct steps. Subsequent detailed discussions of each step are elaborated upon in four separate sections, from Section IV-D2 to Section IV-D5, with each section dedicated to one specific step of the proof. Following these, Section IV-D6 presents the proof of Theorem 4.

1) *Proof Sketch:* In the previous section, the proof of Theorem 2 boils down to obtaining the operator norm bound on the terms $B_j, F_{j1}, F_{j2}, G_{j1}, G_{j2}$. In the proof of Theorem 3, we go further by showing an *exponential tail bound* on the operator norm of these terms.

The proof of Theorem 3 is quite involved but can be broken down into the following key steps:

Step 1. Let \mathcal{F} (to be defined in (53)) be a high-probability event such that the expected value of the normalized Hamming loss $\mathbb{E}\ell(\hat{Z}, Z^*) \approx \mathbb{E}\ell(\hat{Z}, Z^*) \mathbb{I}\{\mathcal{F}\}$. We first show an upper bound on $\mathbb{E}\ell(\hat{Z}, Z^*) \mathbb{I}\{\mathcal{F}\}$ as a sum of n terms, one for each block of U :

$$\mathbb{E}\ell(\hat{Z}, Z^*) \mathbb{I}\{\mathcal{F}\} \leq \frac{1}{n} \sum_{j \in [n]} \sum_{R \in \Pi_d: R \neq I_d} \mathbb{E} \left\{ \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I}\{\mathcal{F}\}, \quad (50)$$

where we recall the definition $\hat{P} := \operatorname{argmin}_{P \in \Pi_d} \|M - PH^\top\|_F$. To understand it, note that each \hat{Z}_j is obtained by finding a permutation matrix closest to $U_j M^\top$ as in (6). If $\sqrt{n} U_j^* \hat{P}^\top$ is the closest permutation matrix to $U_j M^\top$, which can equivalently be stated as

$$\left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F < \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F, \quad (51)$$

$\forall R \in \Pi_d$ s.t. $R \neq I_d$, then according to (10), $\hat{Z}_j = \sqrt{n} U_j^* \hat{P}^\top = \hat{P}^\top$. If the event (51) holds for all $j \in [n]$, then all \hat{Z}_j are \hat{P}^\top , and consequently \hat{Z} is equal to Z^* up to a global permutation \hat{P}^\top . Hence, the distance between \hat{Z} and Z^* essentially depends on the number of $j \in [n]$ for which (51) is not satisfied, which can be further upper bounded and leads to (50).

Step 2. From (51), upper bounding $\mathbb{E}\ell(\hat{Z}, Z^*) \mathbb{I}\{\mathcal{F}\}$ is about upper bounding each individual term on its right-hand side. It turns out each error term can be further decomposed into four tail probabilities. Consider any $j \in [n]$ and any $R \in \Pi_d$ such that $R \neq I_d$, we are able to show

$$\begin{aligned} & \mathbb{E} \left\{ \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I}\{\mathcal{F}\} \\ & \leq \text{a tail probability of } \|F_{j1}\|_{\text{op}} \\ & \quad + \text{a tail probability of } \|F_{j2}\|_{\text{op}} \\ & \quad + \text{a tail probability of } \|G_{j2}\|_{\text{op}} \\ & \quad + \text{a tail probability of } \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle. \end{aligned} \quad (52)$$

This step is summarized in Lemma 9 and explicit expressions of tail probabilities are shown in (57). Among the four tail probabilities, the last one that involves $\sum_{k \neq j} A_{jk} W_{jk}$ will lead to the exponential error rate. There is a variable ρ that trade-offs the contributions between the four terms which will be determined in the last step.

Step 3. In this step, we leverage Lemma 5–8 to obtain exponential bound on each of the four tail probabilities of (52). Despite $\|F_{j1}\|_{\text{op}}, \|F_{j2}\|_{\text{op}},$ and $\|G_{j2}\|_{\text{op}}$ have been analyzed in the proof of Theorem 2, results established there cannot directly applied here. In fact, to derive sharp tail bounds here, we need leverage the obtained blockwise maximum deviation in Theorem 2. The final tail bounds are given in (64), (66), (69), and (71), respectively.

Step 4. In the last step, we combine the four upper bounds in Step 3 to obtain a tight upper bound on (52), and consequently on $\mathbb{E}\ell(\hat{Z}, Z^*) \mathbb{I}\{\mathcal{F}\}$. At this point, we will select the variable ρ introduced in Step 2 to ensure the tail probability involving $\sum_{k \neq j} A_{jk} W_{jk}$ dominates the other three, and consequently the final error bound matches the desired form.

2) *Step 1. Decomposition of The Hamming Loss:* From Lemma 1, there exist constants $C'_0, C_0 > 0$ such that if $\frac{np}{\log n} > C'_0$, then the event \mathcal{E} holds with probability at least $1 - n^{-10}$. By Theorem 2, there exist some constants $C, C', C'' > 0$ such that if $\frac{np}{(\sigma^{4/3}\sqrt{1})\log n} > C$ and $\frac{np}{(\sqrt{d} + \sigma d)^2} \geq C'$, we have (15) holds with probability at least $1 - 6dn^{-9}$. Denote \mathcal{F} to be the event that both \mathcal{E} and (15) hold. That is,

$$\mathcal{F} := \{\mathcal{E}, (15) \text{ holds}\}. \quad (53)$$

By union bound, we have

$$\mathbb{P}(\mathcal{F}) \geq 1 - 7dn^{-9}. \quad (54)$$

Assume \mathcal{F} holds and $\frac{np}{(\sqrt{d} + \sigma d)^2} \geq 22^2 C_0^2$. Then according to Lemma 1, we have (24)–(36) hold. As a consequence, by Lemma 4 and Lemma 2, we have (40)–(41) and (37) all hold as well. Note that

$$\begin{aligned} \mathbb{E}\ell(\hat{Z}, Z^*) & \leq \mathbb{E}\ell(\hat{Z}, Z^*) \mathbb{I}\{\mathcal{F}\} + \mathbb{P}(\mathcal{F}^c) \\ & \leq \mathbb{E}\ell(\hat{Z}, Z^*) \mathbb{I}\{\mathcal{F}\} + 7dn^{-9}. \end{aligned} \quad (55)$$

We will later show that the $7dn^{-9}$ is dominated by $\mathbb{E}\ell(\hat{Z}, Z^*) \mathbb{I}\{\mathcal{F}\}$. Define

$$\hat{P} := \operatorname{argmin}_{P \in \Pi_d} \|M - PH^\top\|_F.$$

Since we let $Z_j^* = I_d$ for all $j \in [n]$, we have

$$\begin{aligned} & \mathbb{E}\ell(\hat{Z}, Z^*) \mathbb{I}\{\mathcal{F}\} \\ & \leq \mathbb{E} \left(\frac{1}{n} \sum_{j \in [n]} \mathbb{I}\{\hat{Z}_j \neq Z_j^* \hat{P}^\top\} \mathbb{I}\{\mathcal{F}\} \right) \\ & = \frac{1}{n} \sum_{j \in [n]} \mathbb{E} \mathbb{I}\{\hat{Z}_j \neq Z_j^* \hat{P}^\top\} \mathbb{I}\{\mathcal{F}\} \\ & \leq \frac{1}{n} \sum_{j \in [n]} \mathbb{E} \left\{ \exists R \in \Pi_d \text{ s.t. } R \neq I_d \text{ and } \left\| \sqrt{n} U_j M^\top - R \hat{P}^\top \right\|_F^2 \leq \left\| \sqrt{n} U_j M^\top - Z_j^* \hat{P}^\top \right\|_F^2 \right\} \mathbb{I}\{\mathcal{F}\} \end{aligned}$$

$$= \frac{1}{n} \sum_{j \in [n]} \mathbb{E} \left\{ \exists R \in \Pi_d \text{ s.t. } R \neq I_d \text{ and } \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F^2 \leq \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F^2 \right\} \mathbb{I}\{\mathcal{F}\}.$$

It can be further upper bounded by

$$\begin{aligned} & \mathbb{E} \ell(\hat{Z}, Z^*) \mathbb{I}\{\mathcal{F}\} \\ & \leq \frac{1}{n} \sum_{j \in [n]} \sum_{R \in \Pi_d: R \neq I_d} \mathbb{E} \left\{ \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F^2 \leq \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F^2 \right\} \mathbb{I}\{\mathcal{F}\} \\ & = \frac{1}{n} \sum_{j \in [n]} \sum_{R \in \Pi_d: R \neq I_d} \mathbb{E} \left\{ \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I}\{\mathcal{F}\}, \quad (56) \end{aligned}$$

where the second inequity is by the definition of \hat{Z}_j in (6) and the second equation is by that we let $Z_j^* = I_d$.

3) Step 2. Decomposition of Each Individual Error:

The reader can see that the key to obtaining our final error bound is to obtain a tight upper bound on (56), which is all about analyzing each individual error $\mathbb{E} \left\{ \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I}\{\mathcal{F}\}$. The next step is the following inequality which shows an upper bound on it consists in four terms. Note that we introduce a ρ parameter which trades off the contribution among the three terms. ρ will be carefully chosen in the third step. Recall the decomposition (20).

Lemma 9: Consider any $j \in [n]$ and any $R \in \Pi_d$ such that $R \neq I_d$. Define

$$h_R := \|R - I_d\|_F^2 / 2.$$

Suppose that (32), (33), (37), (41), and $\frac{np}{(1+\sigma\sqrt{d})^2} \geq 22^2 C_0^2$ hold. Then there exists some $C_2 > 0$ such that for any $\rho > 0$, the following inequality holds deterministically:

$$\begin{aligned} & \mathbb{E} \left\{ \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I}\{\mathcal{F}\} \\ & \leq \mathbb{E} \left\{ 4 \sqrt{2dh_R} \|F_{j1}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I}\{\mathcal{F}\} \\ & + \mathbb{E} \left\{ 4 \sqrt{2dh_R} \|F_{j2}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I}\{\mathcal{F}\} \\ & + \mathbb{E} \left\{ 46C_0 \left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}} \right) \sqrt{2dh_R} \|G_{j2}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I}\{\mathcal{F}\} \\ & + \mathbb{E} \left\{ \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \geq \left(1 - 3\rho - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma\sqrt{d}}{\sqrt{np}} \right) \right) \frac{h_R}{\sqrt{n}} \right\} \mathbb{I}\{\mathcal{F}\}. \quad (57) \end{aligned}$$

Proof: By the definition of h_R , we have $2 \leq h_R \leq d$. We have

$$\mathbb{I} \left\{ \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I}\{\mathcal{F}\}$$

$$\begin{aligned} & = \mathbb{I} \left\{ \left\langle U_j M^\top - U_j^* \hat{P}^\top, \frac{R \hat{P}^\top}{\sqrt{n}} - U_j^* \hat{P}^\top \right\rangle \geq \frac{1}{2} \left\| \frac{R \hat{P}^\top}{\sqrt{n}} - U_j^* \hat{P}^\top \right\|_F^2 \right\} \mathbb{I}\{\mathcal{F}\} \\ & = \mathbb{I} \left\{ \left\langle U_j M^\top \hat{P} - U_j^*, \frac{R}{\sqrt{n}} - U_j^* \right\rangle \geq \frac{1}{2} \left\| \frac{R}{\sqrt{n}} - U_j^* \right\|_F^2 \right\} \mathbb{I}\{\mathcal{F}\} \\ & = \mathbb{I} \left\{ \left\langle U_j M^\top \hat{P} - U_j^*, R - I_d \right\rangle \geq \frac{h_R}{\sqrt{n}} \right\} \mathbb{I}\{\mathcal{F}\} \\ & = \mathbb{I} \left\{ \left\langle U_j H - U_j^*, R - I_d \right\rangle + \left\langle U_j (M - \hat{P} H^\top)^\top \hat{P}, R - I_d \right\rangle \geq \frac{h_R}{\sqrt{n}} \right\} \mathbb{I}\{\mathcal{F}\}. \quad (58) \end{aligned}$$

We have to obtain a more refined decomposition of both inner products in (58). For the second inner product of (58) we have

$$\begin{aligned} & \left\langle U_j (M - \hat{P} H^\top)^\top \hat{P}, R - I_d \right\rangle \\ & \leq \left\| U_j (M - \hat{P} H^\top)^\top \right\|_F \|R - I_d\|_F \\ & \leq \|U_j\|_{\text{op}} \|M - \hat{P} H^\top\|_F \|R - I_d\|_F \\ & \leq \frac{4}{3} \left(\|U_j H - U_j^*\|_{\text{op}} + \frac{1}{\sqrt{n}} \right) \left(\frac{4C_0 (\sqrt{d} + \sigma d)}{\sqrt{np}} \right) \sqrt{2h_R} \\ & \leq \frac{4}{3} \left(\|B_j\|_{\text{op}} + \|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} + \|G_{j1}\|_{\text{op}} + \|G_{j2}\|_{\text{op}} \right. \\ & \quad \left. + \frac{1}{\sqrt{n}} \right) \left(\frac{4C_0 (\sqrt{d} + \sigma d)}{\sqrt{np}} \right) \sqrt{2h_R}, \quad (59) \end{aligned}$$

where the third inequality is by (33) and (37) and the fourth inequality is by (20).

For the first inner product of (58), by (20), we have

$$\begin{aligned} & \left\langle U_j H - U_j^*, R - I_d \right\rangle \\ & = \langle B_j + F_{j1} + F_{j2} + G_{j1}, R - I_d \rangle + \langle G_{j2}, R - I_d \rangle \\ & \leq \|B_j + F_{j1} + F_{j2} + G_{j1}\|_F \|R - I_d\|_F + \langle G_{j2}, R - I_d \rangle \\ & \leq \sqrt{d} \|B_j + F_{j1} + F_{j2} + G_{j1}\|_{\text{op}} \sqrt{2h_R} + \langle G_{j2}, R - I_d \rangle \\ & \leq \left(\|B_j\|_{\text{op}} + \|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} + \|G_{j1}\|_{\text{op}} \right) \sqrt{2dh_R} \\ & \quad + \langle G_{j2}, R - I_d \rangle. \end{aligned}$$

From (41), since $\frac{np}{(1+\sigma\sqrt{d})^2} \geq 22^2 C_0^2$, we have

$$\begin{aligned} \|B_j\|_{\text{op}} & \leq \frac{22C_0}{\sqrt{n}} \left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}} \right) + \|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} \\ & \quad + \|G_{j1}\|_{\text{op}} + 22C_0 \left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}} \right) \|G_{j2}\|_{\text{op}}. \quad (60) \end{aligned}$$

The reader can see that thanks to this upper bound, we only have to contend with establishing upper bounds on the operator norm of $F_{j1}, F_{j2}, G_{j1}, G_{j2}$. From the above three displays, we proceed with a series of inequalities where we use the established upper bound results obtained previously

to obtain an upper bound that can be written in terms of $\|F_{j1}\|_{\text{op}}$, $\|F_{j2}\|_{\text{op}}$, $\|G_{j1}\|_{\text{op}}$ and $\langle G_{j2}, R - I_d \rangle$.

$$\begin{aligned}
& \langle U_j H - U_j^*, R - I_d \rangle + \langle U_j (M - \hat{P} H^\top)^\top \hat{P}, R - I_d \rangle \\
& \leq \left(1 + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) \\
& \quad \times \left(\|B_j\|_{\text{op}} + \|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} + \|G_{j1}\|_{\text{op}} \right) \sqrt{2dh_R} \\
& \quad + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \sqrt{2dh_R} \|G_{j2}\|_{\text{op}} + \langle G_{j2}, R - I_d \rangle \\
& \quad + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \frac{\sqrt{2dh_R}}{\sqrt{n}} \\
& \leq 2 \left(1 + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) \\
& \quad \times \left(\|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} + \|G_{j1}\|_{\text{op}} \right) \sqrt{2dh_R} \\
& \quad + \left(\frac{41}{33} + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) 22C_0 \left(\frac{1 + \sigma\sqrt{d}}{\sqrt{np}} \right) \\
& \quad \times \|G_{j2}\|_{\text{op}} \sqrt{2dh_R} + \langle G_{j2}, R - I_d \rangle \\
& \quad + \left(\frac{41}{33} + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) 22C_0 \left(\frac{1 + \sigma\sqrt{d}}{\sqrt{np}} \right) \frac{\sqrt{2dh_R}}{\sqrt{n}}.
\end{aligned}$$

where the second inequality is by the upper bound on $\|B_j\|_{\text{op}}$ from (60). We then have

$$\begin{aligned}
& \langle U_j H - U_j^*, R - I_d \rangle + \langle U_j (M - \hat{P} H^\top)^\top \hat{P}, R - I_d \rangle \\
& \leq 2 \left(1 + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) \left(\|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} \right) \sqrt{2dh_R} \\
& \quad + \left(\frac{41}{33} + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) 22C_0 \left(\frac{1 + \sigma\sqrt{d}}{\sqrt{np}} \right) \\
& \quad \times \|G_{j2}\|_{\text{op}} \sqrt{2dh_R} + \langle G_{j2}, R - I_d \rangle \\
& \quad + \left(\frac{41}{33} + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) 22C_0 \left(\frac{1 + \sigma\sqrt{d}}{\sqrt{np}} \right) \frac{\sqrt{2dh_R}}{\sqrt{n}} \\
& \quad + 4 \left(1 + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) C_0 \left(\frac{\sqrt{\log n} + \sigma\sqrt{d}}{\sqrt{np}} \right) \\
& \quad \times \frac{\sqrt{2dh_R}}{\sqrt{n}} \\
& \leq 2 \left(1 + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) \left(\|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} \right) \sqrt{2dh_R} \\
& \quad + \left(\frac{41}{33} + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) 22C_0 \left(\frac{1 + \sigma\sqrt{d}}{\sqrt{np}} \right) \\
& \quad \times \|G_{j2}\|_{\text{op}} \sqrt{2dh_R} + \langle G_{j2}, R - I_d \rangle \\
& \quad + \left(\frac{41}{33} + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) 26C_0 \left(\frac{\sqrt{\log n} + \sigma\sqrt{d}}{\sqrt{np}} \right) \\
& \quad \times \frac{\sqrt{2dh_R}}{\sqrt{n}},
\end{aligned}$$

where the first inequality is by the upper bound on $\|G_{j1}\|_{\text{op}}$ in (32) and the last inequality is obtained by combining the two terms containing $\frac{\sqrt{2dh_R}}{\sqrt{n}}$.

In addition to $\|F_{j1}\|_{\text{op}}$, $\|F_{j2}\|_{\text{op}}$, and $\|G_{j1}\|_{\text{op}}$, the above display involves $\langle G_{j2}, R - I_d \rangle$ which must be analyzed closely in order to achieve the optimal error bound. Note the following decomposition.

$$\begin{aligned}
& \langle G_{j2}, R - I_d \rangle \\
& = \frac{\sigma}{\sqrt{n}} \left\langle \sum_{k \neq j} A_{jk} W_{jk} \Lambda^{-1}, R - I_d \right\rangle \\
& = \frac{\sigma}{\sqrt{n}} \left\langle \sum_{k \neq j} A_{jk} W_{jk} \frac{1}{np}, R - I_d \right\rangle \\
& \quad + \frac{\sigma}{\sqrt{n}} \left\langle \sum_{k \neq j} A_{jk} W_{jk} \left(\Lambda^{-1} - \frac{1}{np} I_d \right), R - I_d \right\rangle \\
& = \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \\
& \quad + \frac{\sigma}{\sqrt{n}} \left\langle \sum_{k \neq j} A_{jk} W_{jk} \Lambda^{-1} (np I_d - \Lambda) \frac{1}{np}, R - I_d \right\rangle \\
& = \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \\
& \quad + \frac{1}{np} \langle G_{j2}(np I_d - \Lambda), R - I_d \rangle.
\end{aligned}$$

It can be further upper bounded by

$$\begin{aligned}
& \langle G_{j2}, R - I_d \rangle \\
& \leq \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \\
& \quad + \frac{\sqrt{d}}{np} \|G_{j2}(np I_d - \Lambda)\|_{\text{op}} \|R - I_d\|_F \\
& \leq \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \\
& \quad + \frac{\sqrt{d}}{np} \|G_{j2}\|_{\text{op}} \max_{i \in [n]} |\lambda_i(X) - np| \sqrt{2h_R} \\
& \leq \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \\
& \quad + \frac{C_0 \sqrt{2dh_R}(1 + \sigma\sqrt{d})}{\sqrt{np}} \|G_{j2}\|_{\text{op}}.
\end{aligned}$$

Hence, plugging the above inequality into the display for $\langle U_j H - U_j^*, R - I_d \rangle + \langle U_j (M - \hat{P} H^\top)^\top \hat{P}, R - I_d \rangle$ gives

$$\begin{aligned}
& \langle U_j H - U_j^*, R - I_d \rangle + \langle U_j (M - \hat{P} H^\top)^\top \hat{P}, R - I_d \rangle \\
& \leq 2 \left(1 + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) \left(\|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} \right) \sqrt{2dh_R} \\
& \quad + \left(\frac{41}{33} + \frac{16C_0(1 + \sigma\sqrt{d})}{3\sqrt{np}} \right) 23C_0 \left(\frac{1 + \sigma\sqrt{d}}{\sqrt{np}} \right)
\end{aligned}$$

$$\begin{aligned}
 & \times \|G_{j2}\|_{\text{op}} \sqrt{2dh_R} \\
 & + \left(\frac{41}{33} + \frac{16C_0(1+\sigma\sqrt{d})}{3\sqrt{np}} \right) 26C_0 \left(\frac{\sqrt{\log n} + \sigma\sqrt{d}}{\sqrt{np}} \right) \\
 & \times \frac{\sqrt{2dh_R}}{\sqrt{n}} \\
 & + \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \\
 & \leq 4 \left(\|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} \right) \sqrt{2dh_R} \\
 & + 46C_0 \left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}} \right) \|G_{j2}\|_{\text{op}} \sqrt{2dh_R} \\
 & + 52C_0 \left(\frac{\sqrt{\log n} + \sigma\sqrt{d}}{\sqrt{np}} \right) \frac{\sqrt{2dh_R}}{\sqrt{n}} \\
 & + \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle, \tag{61}
 \end{aligned}$$

where the last inequality holds when $\frac{\sqrt{np}}{1+\sigma\sqrt{d}} > \frac{176}{25}C_0$.

As a result, by (61), (58) becomes

$$\begin{aligned}
 & \mathbb{I} \left\{ \|U_j M^\top - U_j^* \hat{P}^\top\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I} \{\mathcal{F}\} \\
 & \leq \mathbb{I} \left\{ 4 \sqrt{2dh_R} \left(\|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} \right) \right. \\
 & \quad + 46C_0 \left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}} \right) \sqrt{2dh_R} \|G_{j2}\|_{\text{op}} \\
 & \quad + \left. \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \geq \right. \\
 & \quad \left. \left(1 - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma\sqrt{d}}{\sqrt{np}} \right) \right) \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\},
 \end{aligned}$$

for some constant $C_2 > 0$. Let $\rho > 0$ be some quantity whose value will be determined later. We have

$$\begin{aligned}
 & \mathbb{I} \left\{ \|U_j M^\top - U_j^* \hat{P}^\top\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I} \{\mathcal{F}\} \\
 & \leq \mathbb{I} \left\{ 4 \sqrt{2dh_R} \|F_{j1}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\} \\
 & \quad + \mathbb{I} \left\{ 4 \sqrt{2dh_R} \|F_{j2}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\} \\
 & \quad + \mathbb{I} \left\{ 46C_0 \left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}} \right) \sqrt{2dh_R} \|G_{j2}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\} \\
 & \quad + \mathbb{I} \left\{ \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \geq \right. \\
 & \quad \left. \left(1 - 3\rho - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma\sqrt{d}}{\sqrt{np}} \right) \right) \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\}.
 \end{aligned}$$

Taking expected values on both sides, we complete the proof. \square

4) *Step 3. Bounding The Four Tail Probabilities:* In the following, we are going to establish upper bounds for each of the four tail probabilities in (57). Recall the definitions of $\Delta, \Delta^{(j)}$ in (42) and those of $\|\Delta\|_{\text{op},\infty}, \|\Delta^{(j)}\|_{\text{op},\infty}$ in (43).

For the first term in (57), by (27), we have

$$\begin{aligned}
 & \mathbb{E} \mathbb{I} \left\{ 4 \sqrt{2dh_R} \|F_{j1}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\} \\
 & \leq \mathbb{E} \left\{ 4 \sqrt{2dh_R} \left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \left\| \Lambda^{-1} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\} \\
 & \leq \mathbb{E} \left\{ \frac{32 \sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\}.
 \end{aligned}$$

Define an event

$$\begin{aligned}
 \mathcal{F}_j := & \mathbb{I} \left\{ \|\Delta^{(j)}\|_{\text{op}} \leq \frac{9C_0(1+\sigma\sqrt{d})}{\sqrt{np}}, \|\Delta^{(j)}\|_{\text{op},\infty} \leq \right. \\
 & \left. 7 \left(\frac{C''}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma\sqrt{d}}{\sqrt{np}} + \frac{\sigma\sqrt{\log n}}{\sqrt{np}} \right) + \frac{1}{\sqrt{n}} \right) \right\}.
 \end{aligned}$$

Note that the upper bound for $\|\Delta^{(j)}\|_{\text{op},\infty}$ in \mathcal{F}_j is a direct consequence of (36) and (15). Together with (35), we have $\mathcal{F} \subset \mathcal{F}_j$. The event \mathcal{F}_j has an equivalent expression. Define a set

$$\begin{aligned}
 \mathcal{Y} := & \left\{ Y = (Y_1^\top, Y_2^\top, \dots, Y_n^\top)^\top \in \mathbb{R}^{nd \times d} : \right. \\
 & \|Y\|_{\text{op}} \leq \frac{9C_0(1+\sigma\sqrt{d})}{\sqrt{np}}, \max_{i \in [n]} \|Y_i\|_{\text{op}} \leq \\
 & \left. 7 \left(\frac{C''}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma\sqrt{d}}{\sqrt{np}} + \frac{\sigma\sqrt{\log n}}{\sqrt{np}} \right) + \frac{1}{\sqrt{n}} \right) \right\}.
 \end{aligned}$$

Then $\mathcal{F}_j = \mathbb{I} \{\Delta^{(j)} \in \mathcal{Y}\}$. Using that $\Delta^{(j)}$ is independent of $\{A_{jk}\}_{k \neq j}$, we have

$$\begin{aligned}
 & \mathbb{E} \mathbb{I} \left\{ 4 \sqrt{2dh_R} \|F_{j1}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\} \\
 & \leq \mathbb{E} \left\{ \frac{32 \sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}_j\} \\
 & = \mathbb{P} \left(\frac{32 \sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}}, \mathcal{F}_j \right) \\
 & = \mathbb{E} \left(\mathbb{P} \left(\frac{32 \sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}}, \mathcal{F}_j \middle| \Delta^{(j)} \right) \right) \\
 & = \mathbb{E} \left(\mathbb{P} \left(\frac{32 \sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \middle| \Delta^{(j)} \right) \right. \\
 & \quad \left. \times \mathbb{I} \{\Delta^{(j)} \in \mathcal{F}_j\} \right)
 \end{aligned}$$

$$\begin{aligned}
&\leq \sup_{\Delta^{(j)} \in \mathcal{Y}} \mathbb{P} \left(\frac{32\sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \left| \Delta^{(j)} \right| \right) \\
&\leq \sup_{\Delta^{(j)} \in \mathcal{Y}} \mathbb{P} \left(\left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{7\sqrt{np}}{32\sqrt{d}} \left| \Delta^{(j)} \right| \right), \quad (62)
\end{aligned}$$

where in the last inequality, we use the fact that $h_R \geq 2$. Assume ρ satisfies

$$\rho \frac{7\sqrt{np}}{64\sqrt{d}} \geq 9C_0(1 + \sigma\sqrt{d})\sqrt{dp}. \quad (63)$$

Then by Lemma 5, we have

$$\begin{aligned}
&\mathbb{E} \mathbb{I} \left\{ 4\sqrt{2dh_R} \|F_{j1}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\} \\
&\leq \sup_{\Delta^{(j)} \in \mathcal{Y}} \mathbb{P} \left(\left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \sqrt{n} \|\Delta^{(j)}\|_{\text{op}} + \rho \frac{7\sqrt{np}}{64\sqrt{d}} \left| \Delta^{(j)} \right| \right) \\
&\leq 2d \sup_{\Delta^{(j)} \in \mathcal{Y}} \exp \left(-\frac{1}{2} \left(\rho \frac{7\sqrt{np}}{64\sqrt{d}} \right)^2 \left(\rho \sqrt{nd} \right. \right. \\
&\quad \times \left. \left. \|\Delta^{(j)}\|_{\text{op}} \|\Delta^{(j)}\|_{\text{op},\infty} + \rho \frac{7\sqrt{np}}{32\sqrt{d}} \|\Delta^{(j)}\|_{\text{op},\infty} \frac{1}{3} \right)^{-1} \right) \\
&\leq 2d \sup_{\Delta^{(j)} \in \mathcal{Y}} \exp \left(-\frac{3}{8} \frac{\rho \frac{7\sqrt{np}}{64\sqrt{d}}}{\|\Delta^{(j)}\|_{\text{op},\infty}} \right) \\
&\leq 2d \exp \left(-\frac{3}{512} \rho \sqrt{np} \left(\sqrt{d} \times \right. \right. \\
&\quad \left. \left. \left(\frac{C''}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma\sqrt{d}}{\sqrt{np}} + \frac{\sigma\sqrt{\log n}}{\sqrt{np}} \right) + \frac{1}{\sqrt{n}} \right) \right)^{-1} \right), \quad (64)
\end{aligned}$$

where in the first and third inequality, we have used (63) and the upper bound on $\|\Delta^{(j)}\|_{\text{op}}$ per \mathcal{Y} , and in the last inequality, we have used the upper bound on $\|\Delta^{(j)}\|_{\text{op},\infty}$ per \mathcal{Y} .

For the second term in (57), by (27), we have

$$\begin{aligned}
&\mathbb{E} \mathbb{I} \left\{ 4\sqrt{2dh_R} \|F_{j2}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\} \\
&\leq \mathbb{E} \mathbb{I} \left\{ 4\sigma\sqrt{2dh_R} \left\| \sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \|\Lambda^{-1}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \\
&\quad \times \mathbb{I} \{\mathcal{F}\} \\
&\leq \mathbb{E} \mathbb{I} \left\{ \frac{32\sigma\sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\} \\
&\leq \mathbb{E} \mathbb{I} \left\{ \frac{32\sigma\sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}_j\}.
\end{aligned}$$

By the independence among $\{A_{jk}\}_{k \neq j}$, $\{W_{jk}\}_{k \neq j}$, and $\Delta^{(j)}$, following the same argument as used in (62), we have

$$\mathbb{E} \mathbb{I} \left\{ 4\sqrt{2dh_R} \|F_{j2}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\}$$

$$\begin{aligned}
&= \mathbb{E} \left(\mathbb{P} \left(\frac{32\sigma\sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}}, \mathcal{F}_j \mid \{A_{jk}\}_{k \neq j}, \Delta^{(j)} \right) \right) \\
&= \mathbb{E} \left(\mathbb{P} \left(\frac{32\sigma\sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \mid \{A_{jk}\}_{k \neq j}, \Delta^{(j)} \right) \mathbb{I} \{\Delta^{(j)} \in \mathcal{F}_j\} \right) \\
&\leq \mathbb{E} \left(\mathbb{P} \left(\frac{32\sigma\sqrt{2dh_R}}{7np} \left\| \sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \mid \{A_{jk}\}_{k \neq j}, \Delta^{(j)} \right) \mathbb{I} \{\Delta^{(j)} \in \mathcal{F}_j\} \right) \\
&\leq \mathbb{E} \left(\mathbb{P} \left(\left\| \sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq \rho \frac{7\sqrt{np}}{32\sigma\sqrt{d}} \left| \{A_{jk}\}_{k \neq j}, \Delta^{(j)} \right| \right) \mathbb{I} \{\Delta^{(j)} \in \mathcal{F}_j\} \right) \\
&\leq \mathbb{E} \left(\left(2 \exp \left(-\frac{c \left(\rho \frac{7\sqrt{np}}{32\sigma\sqrt{d}} \right)^2}{\left\| \sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}}} \right) \right. \right. \\
&\quad \left. \left. + \mathbb{I} \left\{ \frac{\rho \frac{7\sqrt{np}}{32\sigma\sqrt{d}}}{\left\| \sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}}^{\frac{1}{2}}} < 4\sqrt{d} \right\} \right) \right. \right. \\
&\quad \left. \left. \times \mathbb{I} \{\Delta^{(j)} \in \mathcal{F}_j\} \right) \right),
\end{aligned}$$

for some constant $c > 0$, where the second to last inequality is by the fact that $h_R \geq 2$ and the last inequality is by Lemma 7. Since

$$\begin{aligned}
&\exp \left(-\frac{c \left(\rho \frac{7\sqrt{np}}{32\sigma\sqrt{d}} \right)^2}{\left\| \sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}}} \right) \leq \exp \left(-\frac{np}{\sigma^2} \right) \\
&+ \mathbb{I} \left\{ \frac{c \left(\rho \frac{7\sqrt{np}}{32\sigma\sqrt{d}} \right)^2}{\left\| \sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}}} \leq \frac{np}{\sigma^2} \right\},
\end{aligned}$$

we have

$$\begin{aligned}
&\mathbb{E} \mathbb{I} \left\{ 2\sqrt{2dh_R} \|F_{j2}\|_{\text{op}} \geq \rho \frac{h_R}{\sqrt{n}} \right\} \mathbb{I} \{\mathcal{F}\} \\
&\leq \mathbb{E} \left(\left(2 \exp \left(-\frac{np}{\sigma^2} \right) \right. \right. \\
&\quad \left. \left. + 3 \mathbb{I} \left\{ \frac{\left(\rho \frac{7\sqrt{np}}{32\sigma\sqrt{d}} \right)^2}{\left\| \sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}}} \leq \frac{np}{c\sigma^2} \wedge 16d \right\} \right) \right. \right. \\
&\quad \left. \left. \times \mathbb{I} \{\Delta^{(j)} \in \mathcal{F}_j\} \right) \right) \\
&\leq 2 \exp \left(-\frac{np}{\sigma^2} \right)
\end{aligned}$$

$$+ 3\mathbb{E}\mathbb{I}\left\{\frac{\left(\rho\frac{7\sqrt{np}}{32\sigma\sqrt{d}}\right)^2}{\left\|\sum_{k\neq j}A_{jk}(\Delta_k^{(j)})^\top\Delta_k^{(j)}\right\|_{\text{op}}}\leq\frac{np}{c\sigma^2}\right\}$$

$$\times \mathbb{I}\{\Delta^{(j)}\in\mathcal{F}_j\},$$

where the last inequality holds as long as $\frac{np}{\sigma^2 d} \geq 16c$. Then characterizing the above display is about controlling the tail probabilities of $\left\|\sum_{k\neq j}A_{jk}(\Delta_k^{(j)})^\top\Delta_k^{(j)}\right\|_{\text{op}}$. Similar to the establishment of (62), we have

$$\mathbb{E}\mathbb{I}\left\{4\sqrt{2dh_R}\|F_{j2}\|_{\text{op}}\geq\rho\frac{h_R}{\sqrt{n}}\right\}\mathbb{I}\{\mathcal{F}\}\leq 2\exp\left(-\frac{np}{\sigma^2}\right)$$

$$+ 3\sup_{\Delta^{(j)}\in\mathcal{Y}}\mathbb{P}\left(\left\|\sum_{k\neq j}A_{jk}(\Delta_k^{(j)})^\top\Delta_k^{(j)}\right\|_{\text{op}}\geq\frac{7^2\rho^2cp}{32^2d}\middle|\Delta^{(j)}\right).$$

Assume ρ satisfies

$$\frac{1}{2}\frac{7^2\rho^2cp}{32^2d}\geq p\left(\frac{9C_0(1+\sigma\sqrt{d})}{\sqrt{np}}\right)^2. \quad (65)$$

By Lemma 6, we have

$$\mathbb{E}\mathbb{I}\left\{4\sqrt{2dh_R}\|F_{j2}\|_{\text{op}}\geq\rho\frac{h_R}{\sqrt{n}}\right\}\mathbb{I}\{\mathcal{F}\}$$

$$\leq 2\exp\left(-\frac{np}{\sigma^2}\right)$$

$$+ 6d\sup_{\Delta^{(j)}\in\mathcal{Y}}\exp\left(\left\|\sum_{k\neq j}A_{jk}(\Delta_k^{(j)})^\top\Delta_k^{(j)}\right\|_{\text{op}}\right)$$

$$\geq p\|\Delta^{(j)}\|_{\text{op}}^2 + \frac{1}{2}\frac{7^2\rho^2cp}{32^2d}$$

$$\leq 2\exp\left(-\frac{np}{\sigma^2}\right)$$

$$+ 6d\sup_{\Delta^{(j)}\in\mathcal{Y}}\exp\left(-\frac{1}{2}\left(\frac{1}{2}\frac{7^2\rho^2cp}{32^2d}\right)^2\left(p\|\Delta^{(j)}\|_{\text{op},\infty}^2\right.\right.$$

$$\left.\left.\times\|\Delta^{(j)}\|_{\text{op}}^2+\|\Delta^{(j)}\|_{\text{op},\infty}^2\frac{1}{2}\frac{7^2\rho^2cp}{16^2d}\frac{1}{3}\right)^{-1}\right)$$

$$\leq 2\exp\left(-\frac{np}{\sigma^2}\right)+6d\sup_{\Delta^{(j)}\in\mathcal{Y}}\exp\left(-\frac{3}{8}\frac{\frac{1}{2}\frac{7^2\rho^2cp}{\|\Delta^{(j)}\|_{\text{op},\infty}^2}}{\|\Delta^{(j)}\|_{\text{op},\infty}^2}\right)$$

$$\leq 2\exp\left(-\frac{np}{\sigma^2}\right)$$

$$+ 6d\exp\left(-\frac{3}{8}\times\frac{1}{2}\frac{7^2\rho^2cp}{32^2d}\left(7^2\left(\frac{C''}{\sqrt{n}}\left(\sqrt{\frac{\log n}{np}}\right.\right.\right.\right.$$

$$\left.\left.\left.\left.+\frac{\sigma\sqrt{d}}{\sqrt{np}}+\frac{\sigma\sqrt{\log n}}{\sqrt{np}}\right)+\frac{1}{\sqrt{n}}\right)^2\right)^{-1}\right)$$

$$\leq 2\exp\left(-\frac{np}{\sigma^2}\right)$$

$$+ 6d\exp\left(-\frac{3c}{16\times 32^2}\left(\rho\sqrt{p}\left(\sqrt{d}\left(\frac{C''}{\sqrt{n}}\left(\sqrt{\frac{\log n}{np}}\right.\right.\right.\right.\right.$$

$$\left.\left.\left.\left.+\frac{\sigma\sqrt{d}}{\sqrt{np}}+\frac{\sigma\sqrt{\log n}}{\sqrt{np}}\right)+\frac{1}{\sqrt{n}}\right)\right)^{-1}\right)^2\right). \quad (66)$$

For the third term in (57), using (27) and the fact that $h_R \geq 2$, we have

$$\mathbb{E}\mathbb{I}\left\{46C_0\left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}}\right)\sqrt{2dh_R}\|G_{j2}\|_{\text{op}}\geq\rho\frac{h_R}{\sqrt{n}}\right\}\mathbb{I}\{\mathcal{F}\}$$

$$\leq \mathbb{E}\mathbb{I}\left\{46C_0\left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}}\right)\sqrt{2dh_R}\frac{\sigma}{\sqrt{n}}\|\Lambda^{-1}\|_{\text{op}}\right.$$

$$\left.\times\left\|\sum_{k\neq j}A_{jk}W_{jk}\right\|_{\text{op}}\geq\rho\frac{h_R}{\sqrt{n}}\right\}$$

$$\leq \mathbb{E}\mathbb{I}\left\{\left\|\sum_{k\neq j}A_{jk}W_{jk}\right\|_{\text{op}}\geq\frac{7}{8\times 46C_0\sqrt{d}\left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}}\right)}\frac{\rho np}{\sigma}\right\}$$

$$= \mathbb{E}\left(\mathbb{P}\left(\left\|\sum_{k\neq j}A_{jk}W_{jk}\right\|_{\text{op}}\geq\frac{7}{8\times 46C_0\sqrt{d}\left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}}\right)}\frac{\rho np}{\sigma}\middle|\{A_{jk}\}_{k\neq j}\right)\right).$$

By the independence between $\{A_{jk}\}_{k\neq j}$ and $\{W_{jk}\}_{k\neq j}$, we have $\sum_{k\neq j}A_{jk}W_{jk}|\{A_{jk}\}_{k\neq j} \stackrel{d}{=} \sqrt{\{A_{jk}\}_{k\neq j}}\xi$ for some $\xi \sim \mathcal{MN}(0, I_d, I_d)$ that is independent of $\{A_{jk}\}_{k\neq j}$. Using Lemma 8, we have

$$\mathbb{E}\mathbb{I}\left\{46C_0\left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}}\right)\sqrt{2dh_R}\|G_{j2}\|_{\text{op}}\geq\rho\frac{h_R}{\sqrt{n}}\right\}\mathbb{I}\{\mathcal{F}\}$$

$$\leq \mathbb{E}\left(2\exp\left(-c\left(\frac{1}{2}\frac{7}{8\times 46C_0\sqrt{d}\left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}}\right)}\frac{\rho np}{\sigma}\right)^2\right.\right.$$

$$\left.\times\frac{1}{\sum_{k\neq j}A_{jk}}\right)+\mathbb{I}\left\{\frac{1}{2}\frac{7}{8\times 46C_0\sqrt{d}\left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}}\right)}\frac{\rho np}{\sigma}\right.$$

$$\left.\times\frac{1}{\sqrt{\sum_{k\neq j}A_{jk}}}<2\sqrt{d}\right\}),$$

for some constant $c' > 0$. Note that by Bernstein's inequality, there exists some constant $C_3 > 0$ such that

$$\mathbb{P}\left(\sum_{k\neq j}A_{jk}\geq np+C_3\sqrt{np\log n}\right)\leq n^{-10}. \quad (67)$$

We assume $\frac{np}{\log n} \geq C_3^2$ such that $2np \geq np + C_3\sqrt{np\log n}$ and consequently $\mathbb{P}\left(\sum_{k\neq j}A_{jk} \geq 2np\right) \leq n^{-10}$. Assume ρ satisfies

$$\frac{1}{\sqrt{2np}}\left(\frac{1}{2}\frac{7}{8\times 46C_0\sqrt{d}\left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}}\right)}\frac{\rho np}{\sigma}\right)\geq(c')^{-\frac{1}{2}}\vee(2\sqrt{d}). \quad (68)$$

We have

$$\mathbb{E}\mathbb{I}\left\{46C_0\left(\frac{1+\sigma\sqrt{d}}{\sqrt{np}}\right)\sqrt{2dh_R}\|G_{j2}\|_{\text{op}}\geq\rho\frac{h_R}{\sqrt{n}}\right\}\mathbb{I}\{\mathcal{F}\}$$

$$\leq 2\exp\left(-\frac{np}{\sigma^2}\right)+2\mathbb{P}\left(\sum_{k\neq j}A_{jk}\geq 2np\right)$$

$$\leq 2\exp\left(-\frac{np}{\sigma^2}\right)+2n^{-10}. \quad (69)$$

For the fourth and last term in (57), by properties of the Gaussian distribution, one can verify that $\left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \left| \{A_{jk}\}_{k \neq j} \sim \mathcal{N}\left(0, 4 \sum_{k \neq j} A_{jk}\right) \right.$. Assume ρ satisfies

$$3\rho + C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \leq \frac{1}{2}. \quad (70)$$

By the fact $h_R \geq 2$, we then have

$$\begin{aligned} & \mathbb{E} \left\{ \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \right. \\ & \geq \left(1 - 3\rho - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \right) \frac{h_R}{\sqrt{n}} \Big\} \mathbb{I}\{\mathcal{F}\} \\ & \leq \mathbb{E} \left\{ \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \right. \\ & \geq \left(1 - 3\rho - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \right) \frac{2np}{\sigma} \Big\} \\ & = \mathbb{E} \left(\mathbb{P} \left(\left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \geq \left(1 - 3\rho \right. \right. \right. \\ & \quad \left. \left. \left. - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \right) \frac{2np}{\sigma} \middle| \{A_{jk}\}_{k \neq j} \right) \right) \\ & \leq \mathbb{E} \exp \left(- \left(1 - 3\rho - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \right)^2 \right. \\ & \quad \left. \times \frac{(np)^2}{2\sigma^2 \sum_{k \neq j} A_{jk}} \right) \\ & \leq \exp \left(- \left(1 - 3\rho - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \right)^2 \right. \\ & \quad \left. \times \frac{(np)^2}{2\sigma^2 (np + C_3 \sqrt{np \log n})} \right) \\ & \quad + \mathbb{P} \left(\sum_{k \neq j} A_{jk} \geq np + C_0 \sqrt{np \log n} \right) \\ & \leq \exp \left(- \left(1 - 3\rho - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \right)^2 \right. \\ & \quad \left. \times \left(1 + C_3 \sqrt{\frac{\log n}{np}} \right)^{-1} \frac{np}{2\sigma^2} \right) + n^{-10}, \end{aligned} \quad (71)$$

where the last inequality is by (67).

5) *Step 4. Selecting ρ and Putting Things Together:* Plugging (64), (66), (69), and (71) into (57), we have

$$\begin{aligned} & \mathbb{E} \left\{ \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I}\{\mathcal{F}\} \\ & \leq 2d \exp \left(- \frac{3}{512} \rho \sqrt{np} \left(\sqrt{d} \times \right. \right. \\ & \quad \left. \left. \left(\frac{C''}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma \sqrt{d}}{\sqrt{np}} + \frac{\sigma \sqrt{\log n}}{\sqrt{np}} \right) + \frac{1}{\sqrt{n}} \right) \right)^{-1} \right) \end{aligned}$$

$$\begin{aligned} & + 2 \exp \left(- \frac{np}{\sigma^2} \right) \\ & + 6d \exp \left(- \frac{3c}{16 \times 32^2} \left(\rho \sqrt{p} \left(\sqrt{d} \left(\frac{C''}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} \right. \right. \right. \right. \right. \right. \right. \\ & \quad \left. \left. \left. + \frac{\sigma \sqrt{d}}{\sqrt{np}} + \frac{\sigma \sqrt{\log n}}{\sqrt{np}} \right) + \frac{1}{\sqrt{n}} \right) \right)^{-1} \right)^2 \right) \\ & + 2 \exp \left(- \frac{np}{\sigma^2} \right) + 2n^{-10} \\ & + \exp \left\{ - \left(1 - 3\rho - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \right)^2 \right. \\ & \quad \left. \times \left(1 + C_3 \sqrt{\frac{\log n}{np}} \right)^{-1} \frac{np}{2\sigma^2} \right\} + n^{-10} \\ & \leq \exp \left(- \left(1 - 3\rho - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \right)^2 \right. \\ & \quad \left. \times \left(1 + C_3 \sqrt{\frac{\log n}{np}} \right)^{-1} \frac{np}{2\sigma^2} \right) + 3n^{-10} \\ & + 4 \exp \left(- \frac{np}{\sigma^2} \right) \\ & + 2d \exp \left(- \frac{3}{512} \rho \sqrt{np} \left(\sqrt{d} \times \right. \right. \\ & \quad \left. \left. \left(\frac{C''}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma \sqrt{d}}{\sqrt{np}} + \frac{\sigma \sqrt{\log n}}{\sqrt{np}} \right) + \frac{1}{\sqrt{n}} \right) \right)^{-1} \right) \\ & + 6d \exp \left(- \frac{3c}{16 \times 32^2} \left(\rho \sqrt{p} \left(\sqrt{d} \left(\frac{C''}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} \right. \right. \right. \right. \right. \right. \right. \right. \\ & \quad \left. \left. \left. + \frac{\sigma \sqrt{d}}{\sqrt{np}} + \frac{\sigma \sqrt{\log n}}{\sqrt{np}} \right) + \frac{1}{\sqrt{n}} \right) \right)^{-1} \right)^2 \right). \end{aligned} \quad (72)$$

We will now pick ρ such that we can substantially simplify the exponential terms. Specifically, we let $\rho = \left(\frac{\sigma^2 d^3}{np} \right)^{\frac{1}{4}} + \left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}}$. Suppose that $\frac{np}{d^2 \log n}, \frac{np}{\sigma^2 d^3} \geq C_3$ for some large enough constant $C_3 > 0$, we can make sure that conditions (63), (65), (68), and (70) are all satisfied. Next, we are going to show that with this choice of ρ , the last two terms in (72) are small. We can show the exponent in the second to last term is bounded below by $\log n$, up to a constant factor, as follows.

$$\begin{aligned} & \frac{\log n}{\log n} \cdot \frac{3}{512} \frac{\rho \sqrt{np}}{\sqrt{d} \left(\frac{C''}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma \sqrt{d}}{\sqrt{np}} + \frac{\sigma \sqrt{\log n}}{\sqrt{np}} \right) + \frac{1}{\sqrt{n}} \right)} \\ & = \log n \cdot \frac{3}{512} \frac{\rho}{\frac{\sqrt{d} \log n}{np} \left(C'' \sqrt{\frac{\log n}{np}} + \frac{C'' \sigma \sqrt{d}}{\sqrt{np}} + \frac{C'' \sigma \sqrt{\log n}}{\sqrt{np}} + 1 \right)} \\ & \geq \log n \cdot \frac{3}{512} \frac{\rho}{\frac{\sqrt{d} (\log n)^{3/2}}{np} \left(C'' \sqrt{\frac{\log n}{np}} + \frac{C'' \sigma \sqrt{d}}{\sqrt{np}} + \frac{C'' \sigma}{\sqrt{np}} + 1 \right)} \\ & \geq \log n \cdot \frac{3}{512} \frac{\left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}}}{\frac{\sqrt{d} (\log n)^{3/2}}{np} \left(C'' \sqrt{\frac{\log n}{np}} + \frac{C'' \sigma \sqrt{d}}{\sqrt{np}} + \frac{C'' \sigma}{\sqrt{np}} + 1 \right)} \end{aligned}$$

$$\begin{aligned}
 &= \log n \cdot \frac{3}{512} \frac{1}{\frac{(\log n)^{5/4}}{(np)^{3/4}} \left(C'' \sqrt{\frac{\log n}{np}} + \frac{C'' \sigma \sqrt{d}}{\sqrt{np}} + \frac{C'' \sigma}{\sqrt{np}} + 1 \right)} \\
 &\geq 10 \log n.
 \end{aligned}$$

In the first line, we simply multiply and divide the same quantity by $\log n$. The second and third lines are by expanding and re-arranging the terms. In the second inequality, we use $\rho \geq \left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}}$. The last inequality holds so long as $\frac{np}{d \log n}, \frac{np}{\log^{5/3} n}, \frac{np}{\sigma^2 d} \geq C'_3$ for some sufficiently large constant C'_3 . Similarly, we can also simplify the term inside the square root of the last term in (72) as

$$\begin{aligned}
 &\frac{\sqrt{\log n}}{\sqrt{\log n}} \cdot \frac{\sqrt{3c}}{4 \times 32} \\
 &\cdot \frac{\rho \sqrt{p}}{\sqrt{d} \left(\frac{C''}{\sqrt{n}} \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma \sqrt{d}}{\sqrt{np}} + \frac{\sigma \sqrt{\log n}}{\sqrt{np}} \right) + \frac{1}{\sqrt{n}} \right)} \\
 &= \sqrt{\log n} \cdot \frac{\sqrt{3c}}{4 \times 32} \\
 &\cdot \frac{\rho}{\frac{\sqrt{d \log n}}{\sqrt{np}} \left(C'' \sqrt{\frac{\log n}{np}} + C'' \frac{\sigma \sqrt{d}}{\sqrt{np}} + C'' \frac{\sigma \sqrt{\log n}}{\sqrt{np}} + 1 \right)} \\
 &\geq \sqrt{\log n} \cdot \frac{\sqrt{3c}}{4 \times 32} \\
 &\cdot \frac{\rho}{\frac{\sqrt{d \log n}}{\sqrt{np}} \left(C'' \sqrt{\frac{\log n}{np}} + C'' \frac{\sigma \sqrt{d}}{\sqrt{np}} + C'' \frac{\sigma}{\sqrt{np}} + 1 \right)} \\
 &\geq \sqrt{\log n} \cdot \frac{\sqrt{3c}}{4 \times 32} \\
 &\cdot \frac{\left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}}}{\frac{\sqrt{d \log n}}{\sqrt{np}} \left(C'' \sqrt{\frac{\log n}{np}} + C'' \frac{\sigma \sqrt{d}}{\sqrt{np}} + C'' \frac{\sigma}{\sqrt{np}} + 1 \right)} \\
 &\geq \sqrt{\log n} \cdot \frac{\sqrt{3c}}{4 \times 32} \\
 &\cdot \frac{1}{\frac{(\log n)^{3/4}}{(np)^{1/4}} \left(C'' \sqrt{\frac{\log n}{np}} + C'' \frac{\sigma \sqrt{d}}{\sqrt{np}} + C'' \frac{\sigma}{\sqrt{np}} + 1 \right)} \\
 &\geq \sqrt{10} \sqrt{\log n}.
 \end{aligned}$$

In the first line, we multiply and divide the same quantity by $\sqrt{\log n}$. The last inequality holds so long as $\frac{np}{d \log n}, \frac{np}{\log^{5/3} n}, \frac{np}{\sigma^2 d} \geq C'_3$ for some sufficiently large constant C'_3 .

Putting things together into (72), we have

$$\begin{aligned}
 &\mathbb{E} \left\{ \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I} \{ \mathcal{F} \} \\
 &\leq \exp \left(- \left(1 - 3 \left(\left(\frac{\sigma^2 d^3}{np} \right)^{\frac{1}{4}} + \left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}} \right) \right. \right. \\
 &\quad \left. \left. - C_2 \sqrt{d} \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \right)^2 \right. \\
 &\quad \left. \times \left(1 + C_3 \sqrt{\frac{\log n}{np}} \right)^{-1} \frac{np}{2\sigma^2} \right) \\
 &\quad + 4 \exp \left(- \frac{np}{\sigma^2} \right) + 3n^{-10}
 \end{aligned}$$

$$+ \underbrace{2d \exp(-10 \log n) + 6d \exp(-10 \log n)}_{\leq 9dn^{-10}},$$

By combining the two exponential terms together and the three polynomial terms together, we see that there exists a constant $C_4 > 0$ such that

$$\begin{aligned}
 &\mathbb{E} \left\{ \left\| U_j M^\top - U_j^* \hat{P}^\top \right\|_F \geq \left\| U_j M^\top - \frac{R \hat{P}^\top}{\sqrt{n}} \right\|_F \right\} \mathbb{I} \{ \mathcal{F} \} \\
 &\leq \exp \left(- \left(1 - C_4 \left(\left(\frac{\sigma^2 d^3}{np} \right)^{\frac{1}{4}} + \left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}} \right) \right) \frac{np}{2\sigma^2} \right) \\
 &\quad + 12dn^{-10}.
 \end{aligned} \tag{73}$$

Combining (54), (55), (56) and (73) we have

$$\begin{aligned}
 &\mathbb{E} \ell(\hat{Z}, Z^*) \\
 &\leq \frac{1}{n} \sum_{j \in [n]} \sum_{R \in \Pi_d: R \neq I_d} \left(\exp \left(- \left(1 - C_4 \left(\left(\frac{\sigma^2 d^3}{np} \right)^{\frac{1}{4}} \right. \right. \right. \right. \\
 &\quad \left. \left. \left. + \left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}} \right) \right) \frac{np}{2\sigma^2} \right) + 12dn^{-10} \right) + \mathbb{P}(\mathcal{F}^c) \\
 &\leq d^d \left(\exp \left(- \left(1 - C_4 \left(\left(\frac{\sigma^2 d^3}{np} \right)^{\frac{1}{4}} \right. \right. \right. \right. \right. \\
 &\quad \left. \left. \left. + \left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}} \right) \right) \frac{np}{2\sigma^2} \right) + 12dn^{-10} \right) + 7dn^{-9} \\
 &\leq \exp \left(- \left(1 - C_5 \left(\left(\frac{\sigma^2 d^3}{np} \right)^{\frac{1}{4}} \right. \right. \right. \right. \\
 &\quad \left. \left. \left. + \left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}} \right) \right) \frac{np}{2\sigma^2} \right) + 8dn^{-9} \\
 &\leq \exp \left(- \left(1 - C_5 \left(\left(\frac{\sigma^2 d^3}{np} \right)^{\frac{1}{4}} \right. \right. \right. \right. \\
 &\quad \left. \left. \left. + \left(\frac{d^2 \log n}{np} \right)^{\frac{1}{4}} \right) \right) \frac{np}{2\sigma^2} \right) + 8dn^{-9},
 \end{aligned}$$

for some constant $C_5 > 0$, where the last inequality holds when $\frac{np}{\sigma^2 d^3}, \frac{np}{d^2 \log n}$ exceed a sufficiently large constant. The proof of Theorem 3 is complete.

6) *Proof of Theorem 4:* The proof closely parallels that of Theorem 3, with substitutions where \hat{Z} and M are replaced by \check{Z} and \check{M} , respectively, and incorporates two minor modifications. Firstly, it is important to note that the error in the anchor affects only the second inner product in (58). Specifically, this influence introduces an additional factor of $(1 + \epsilon)$, leading to a modified formulation in (59) for the scenario involving \check{M} :

$$\begin{aligned}
 &\left\langle U_j (\check{M} - \hat{P} H^\top)^\top \hat{P}, R - I_d \right\rangle \\
 &\leq \left\| U_j (\check{M} - \hat{P} H^\top)^\top \right\|_F \|R - I_d\|_F \\
 &\leq \|U_j\|_{\text{op}} \|\check{M} - \hat{P} H^\top\|_F \|R - I_d\|_F \\
 &\leq \frac{4}{3} \left(\|U_j H - U_j^*\|_{\text{op}} + \frac{1}{\sqrt{n}} \right)
 \end{aligned}$$

$$\begin{aligned}
& \times \left(\frac{4C_0 \sqrt{1+\epsilon} (\sqrt{d} + \sigma d)}{\sqrt{np}} \right) \sqrt{2h_R} \\
& \leq \frac{4}{3} \left(\|B_j\|_{\text{op}} + \|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} + \|G_{j1}\|_{\text{op}} \right. \\
& \quad \left. + \|G_{j2}\|_{\text{op}} + \frac{1}{\sqrt{n}} \right) \left(\frac{4C_0 \sqrt{1+\epsilon} (\sqrt{d} + \sigma d)}{\sqrt{np}} \right) \sqrt{2h_R},
\end{aligned}$$

where in the second inequality, we use (39) from Lemma 3 instead (37) from Lemma 2. We can then follow the same argument in the proof leading up to (61) where we make the second modification:

$$\begin{aligned}
& \langle U_j H - U_j^*, R - I_d \rangle + \langle U_j (\check{M} - \hat{P} H^\top)^\top \hat{P}, R - I_d \rangle \\
& \leq 2 \left(1 + \frac{16C_0 \sqrt{1+\epsilon} (1 + \sigma \sqrt{d})}{3 \sqrt{np}} \right) \\
& \quad \times \left(\|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} \right) \sqrt{2dh_R} \\
& \quad + \left(\frac{41}{33} + \frac{16 \sqrt{1+\epsilon} C_0 (1 + \sigma \sqrt{d})}{3 \sqrt{np}} \right) \\
& \quad \times 23C_0 \left(\frac{1 + \sigma \sqrt{d}}{\sqrt{np}} \right) \|G_{j2}\|_{\text{op}} \sqrt{2dh_R} \\
& \quad + \left(\frac{41}{33} + \frac{16 \sqrt{1+\epsilon} C_0 (1 + \sigma \sqrt{d})}{3 \sqrt{np}} \right) \\
& \quad \times 26C_0 \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \frac{\sqrt{2dh_R}}{\sqrt{n}} \\
& \quad + \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle \\
& \leq 4 \left(\|F_{j1}\|_{\text{op}} + \|F_{j2}\|_{\text{op}} \right) \sqrt{2dh_R} \\
& \quad + 46C_0 \left(\frac{1 + \sigma \sqrt{d}}{\sqrt{np}} \right) \|G_{j2}\|_{\text{op}} \sqrt{2dh_R} \\
& \quad + 52C_0 \left(\frac{\sqrt{\log n} + \sigma \sqrt{d}}{\sqrt{np}} \right) \frac{\sqrt{2dh_R}}{\sqrt{n}} \\
& \quad + \frac{\sigma}{\sqrt{n}} \frac{1}{np} \left\langle \sum_{k \neq j} A_{jk} W_{jk}, R - I_d \right\rangle,
\end{aligned}$$

where the last inequality holds under the assumption that $\frac{\sqrt{np}}{1+\sigma\sqrt{d}} > \frac{176}{25} C_0 \sqrt{1+\epsilon}$. In the way, we obtain the exact same upper bound as (61). There is no change to the rest of the proof.

E. Helper Lemmas and Proofs

Lemma 10: We have $\lambda_i(\mathbb{E}A \otimes I_d) = (n-1)p$ for all $i \leq d$ and $\lambda_i(\mathbb{E}A \otimes I_d) = -p$ for all $d+1 \leq i \leq nd$.

Proof: Recall that $\mathbb{E}A = pJ_n - pI_n$. The eigenvalues of the matrix $\mathbb{E}A$ are characterized as follows.

$$\begin{aligned}
\lambda_1(\mathbb{E}A) &= (n-1)p, \\
\lambda_2(\mathbb{E}A) &= \dots = \lambda_n(\mathbb{E}A) = -p.
\end{aligned}$$

Therefore the eigenvalues of $\mathbb{E}A \otimes I_d$ are as follows.

$$\begin{aligned}
\lambda_1(\mathbb{E}A \otimes I_d) &= \dots = \lambda_d(\mathbb{E}A \otimes I_d) = (n-1)p, \\
\lambda_{d+1}(\mathbb{E}A \otimes I_d) &= \dots = \lambda_{dn}(\mathbb{E}A \otimes I_d) = -p.
\end{aligned}$$

□

Proof of Lemma 1: The probability of \mathcal{E} is from Lemma 9 of [17] (for $\|(A \otimes J_d) \circ W\|_{\text{op}}$), Theorem 5.2 of [25] (for $\|A - \mathbb{E}A\|_{\text{op}}$), and Bernstein's inequality (for $\max_{j \in [n]} |\sum_{k \neq j} A_{jk} - np|$). Note that

$$\begin{aligned}
& \|X - (\mathbb{E}A \otimes I_d)\|_{\text{op}} \\
& = \|(A \otimes I_d) + \sigma(A \otimes J_d) \circ W - (\mathbb{E}A \otimes I_d)\|_{\text{op}} \\
& \leq \|(A - \mathbb{E}A) \otimes I_d\|_{\text{op}} + \sigma \|(A \otimes J_d) \circ W\|_{\text{op}} \\
& = \|A - \mathbb{E}A\|_{\text{op}} + \sigma \|(A \otimes J_d) \circ W\|_{\text{op}}.
\end{aligned}$$

By Weyl's inequality, we have $\max_{i \in [nd]} |\lambda_i(X) - \lambda_i(\mathbb{E}A \otimes I_d)| \leq \|X - (\mathbb{E}A \otimes I_d)\|_{\text{op}}$. Together with Lemma 10, (24) holds under the event \mathcal{E} . For (25), we have

$$\begin{aligned}
\max_j \|X_j\|_{\text{op}} &= \max_j \|(\mathbb{E}A_j \otimes I_d) + (X_j - (\mathbb{E}A_j \otimes I_d))\|_{\text{op}} \\
&\leq \max_j \|(\mathbb{E}A_j \otimes I_d)\|_{\text{op}} + \|X - (\mathbb{E}A \otimes I_d)\|_{\text{op}} \\
&\leq p \sqrt{n} + C_0 (1 + \sigma \sqrt{d}) \sqrt{np}.
\end{aligned}$$

If $\frac{np}{(1+\sigma\sqrt{d})^2} \geq 64C_0^2$ is further assumed, we have $C_0(1 + \sigma \sqrt{d}) \sqrt{np} \leq np/8$. Then (24), together with the fact $\|\Lambda^{-1}\|_{\text{op}} = 1/\lambda_d(X)$ and $\|\Lambda\|_{\text{op}} = \lambda_1(X)$, leads to (26)-(28).

Since $\lambda_d(\mathbb{E}A \otimes I_d) - \lambda_{d+1}(\mathbb{E}A \otimes I_d) = np$ according to Lemma 10, we have $\|X - (\mathbb{E}A \otimes I_d)\|_{\text{op}} \leq (\lambda_d(\mathbb{E}A \otimes I_d) - \lambda_{d+1}(\mathbb{E}A \otimes I_d))/8$. By Lemma 2 of [16], we have

$$\begin{aligned}
\min_{O \in \mathcal{O}_d} \|H - O\|_{\text{op}} &\leq \|UU^\top - U^*U^{*\top}\|_{\text{op}} \\
&\leq \frac{\|X - (\mathbb{E}A \otimes I_d)\|_{\text{op}}}{(1 - \frac{1}{8})np} \leq \frac{8C_0 (1 + \sigma \sqrt{d})}{7 \sqrt{np}},
\end{aligned}$$

and

$$\begin{aligned}
\|H\Lambda - \Lambda H\|_{\text{op}} &\leq 2 \|X - (\mathbb{E}A \otimes I_d)\|_{\text{op}} \\
&\leq 2C_0 (1 + \sigma \sqrt{d}) \sqrt{np},
\end{aligned}$$

and $\|H^{-1}\|_{\text{op}} \leq 4/3$. Note that

$$\begin{aligned}
\|UH - U^*\|_{\text{op}} &= \|UU^\top U^* - U^*U^{*\top}U^*\|_{\text{op}} \\
&\leq \|UU^\top - U^*U^{*\top}\|_{\text{op}} \|U^*\|_{\text{op}} \leq \|UU^\top - U^*U^{*\top}\|_{\text{op}}.
\end{aligned}$$

Hence, (29)-(31) hold. Consider any $j \in [n]$. By (31), we have

$$\begin{aligned}
\|U_j\|_{\text{op}} &\leq \|U_j H\|_{\text{op}} \|H^{-1}\|_{\text{op}} \\
&\leq \left(\|U_j H - U_j^*\|_{\text{op}} + \|U_j^*\|_{\text{op}} \right) \|H^{-1}\|_{\text{op}} \\
&= \frac{4}{3} \left(\|U_j H - U_j^*\|_{\text{op}} + \frac{1}{\sqrt{n}} \right).
\end{aligned}$$

From (25), (28) and that $C_0 > 7$, we have $\|X - X^{(j)}\|_{\text{op}} \leq 2 \|X_j\|_{\text{op}} \leq (\lambda_d(X) - \lambda_{d+1}(X))/2$. By Davis-Kahan theorem, we have

$$\|U^{(j)}(U^{(j)})^\top - UU^\top\|_{\text{op}}$$

$$\begin{aligned}
 & \leq \frac{2 \| (X - X^{(j)}) U \|_{\text{op}}}{\lambda_d(X) - \lambda_{d+1}(X)} \\
 & \leq \frac{2 (\| X_j U \|_{\text{op}} + \| X_j \|_{\text{op}} \| U_j \|_{\text{op}})}{\lambda_d(X) - \lambda_{d+1}(X)} \\
 & = \frac{2 (\| U_j \Lambda \|_{\text{op}} + \| X_j \|_{\text{op}} \| U_j \|_{\text{op}})}{\lambda_d(X) - \lambda_{d+1}(X)} \\
 & \leq \frac{2 (\| U_j \|_{\text{op}} \| \Lambda \|_{\text{op}} + \| X_j \|_{\text{op}} \| U_j \|_{\text{op}})}{\lambda_d(X) - \lambda_{d+1}(X)} \\
 & \leq 4 \| U_j \|_{\text{op}} \\
 & \leq 6 \left(\| U_j H - U_j^* \|_{\text{op}} + \frac{1}{\sqrt{n}} \right),
 \end{aligned}$$

where the second to last inequality is due to (25), (26), and (28), and the last inequality is due to (33).

Together with (29) and that $C_0 > 7$ we have

$$\begin{aligned}
 \| U^{(j)} (U^{(j)})^\top - U U^\top \|_{\text{op}} & \leq 6 \left(\| UH - U^* \|_{\text{op}} + \frac{1}{\sqrt{n}} \right) \\
 & \leq \frac{48C_0(1 + \sigma\sqrt{d})}{7\sqrt{np}} + \frac{6}{\sqrt{np}} \\
 & \leq \frac{7C_0(1 + \sigma\sqrt{d})}{\sqrt{np}}.
 \end{aligned}$$

Then,

$$\begin{aligned}
 & \| U^{(j)} H^{(j)} - U^* \|_{\text{op}} \\
 & = \| (U^{(j)} (U^{(j)})^\top - U U^\top) U^* + UH - U^* \|_{\text{op}} \\
 & \leq \| U^{(j)} (U^{(j)})^\top - U U^\top \|_{\text{op}} + \| UH - U^* \|_{\text{op}} \\
 & \leq \frac{9C_0(1 + \sigma\sqrt{d})}{\sqrt{np}}.
 \end{aligned}$$

For $\| G_{j1} \|_{\text{op}}$, we have

$$\begin{aligned}
 & \sqrt{n} \| G_{j1} \|_{\text{op}} \\
 & = \left\| \sum_{k \neq j} A_{jk} \Lambda^{-1} - I_d \right\|_{\text{op}} \\
 & \leq \| np \Lambda^{-1} - I_d \|_{\text{op}} + \left\| \left(\sum_{k \neq j} A_{jk} - np \right) \Lambda^{-1} \right\|_{\text{op}} \\
 & \leq \| np - \Lambda \|_{\text{op}} \| \Lambda^{-1} \|_{\text{op}} + \left| \sum_{k \neq j} A_{jk} - np \right| \| \Lambda^{-1} \|_{\text{op}} \\
 & \leq \max_{i \in [d]} |np - \lambda_i(X)| \| \Lambda^{-1} \|_{\text{op}} + \left| \sum_{k \neq j} A_{jk} - np \right| \| \Lambda^{-1} \|_{\text{op}} \\
 & \leq 2C_0 \left(\sqrt{\frac{\log n}{np}} + \frac{\sigma\sqrt{d}}{\sqrt{np}} \right),
 \end{aligned}$$

where the last inequality is due to \mathcal{E} , (24), and (27).

For (36), consider any $j \in [n]$, we have

$$\begin{aligned}
 & \| U^{(j)} H^{(j)} - U^* \|_{\text{op}, \infty} \\
 & \leq \| UH - U^* \|_{\text{op}, \infty}
 \end{aligned}$$

$$\begin{aligned}
 & + \| (UH - U^*) - (U^{(j)} H^{(j)} - U^*) \|_{\text{op}, \infty} \\
 & \leq \| UH - U^* \|_{\text{op}, \infty} \\
 & + \| (UH - U^*) - (U^{(j)} H^{(j)} - U^*) \|_{\text{op}} \\
 & = \| UH - U^* \|_{\text{op}, \infty} + \| UH - U^{(j)} H^{(j)} \|_{\text{op}} \\
 & = \| UH - U^* \|_{\text{op}, \infty} + \| U U^\top U^* - U^{(j)} (U^{(j)})^\top U^* \|_{\text{op}} \\
 & \leq \| UH - U^* \|_{\text{op}, \infty} + \| U U^\top - U^{(j)} (U^{(j)})^\top \|_{\text{op}} \\
 & \leq \| UH - U^* \|_{\text{op}, \infty} + 6 \left(\| U_j H - U_j^* \|_{\text{op}} + \frac{1}{\sqrt{n}} \right) \\
 & \leq 7 \left(\| UH - U^* \|_{\text{op}, \infty} + \frac{1}{\sqrt{n}} \right),
 \end{aligned}$$

where the second to last inequality is due to (34). \square

Proof of Lemma 5: Consider any $j \in [n]$. Then

$$\begin{aligned}
 & \left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \\
 & \leq \left\| \sum_{k \neq j} (A_{jk} - p) \Delta_k^{(j)} \right\|_{\text{op}} + p \left\| \sum_{k \neq j} \Delta_k^{(j)} \right\|_{\text{op}} \\
 & \leq \left\| \sum_{k \neq j} (A_{jk} - p) \Delta_k^{(j)} \right\|_{\text{op}} + p \sqrt{n} \| \Delta^{(j)} \|_{\text{op}}.
 \end{aligned}$$

Since $\{A_{jk} - p\}_{k \neq j}$ is independent of $\Delta^{(j)}$, we use the matrix Bernstein's inequality (Lemma 11) for the operator norm of $\sum_{k \neq j} (A_{jk} - p) \Delta_k^{(j)}$. For each $k \in [n]$, note that $\mathbb{E}((A_{jk} - p) \Delta_k^{(j)} | \Delta_k^{(j)}) = 0$ and $\| (A_{jk} - p) \Delta_k^{(j)} \|_{\text{op}} \leq \| \Delta_k^{(j)} \|_{\text{op}} \leq \| \Delta^{(j)} \|_{\text{op}, \infty}$. For the matrix variance term, we have

$$\begin{aligned}
 & \max \left\{ \left\| \mathbb{E} \left[\sum_{k \neq j} (A_{jk} - p)^2 \Delta_k^{(j) \top} \Delta_k^{(j)} \middle| \Delta^{(j)} \right] \right\|_{\text{op}}, \right. \\
 & \left. \left\| \mathbb{E} \left[\sum_{k \neq j} (A_{jk} - p)^2 \Delta_k^{(j)} \Delta_k^{(j) \top} \middle| \Delta^{(j)} \right] \right\|_{\text{op}} \right\} \\
 & = p(1 - p) \\
 & \times \max \left\{ \left\| \sum_{k \neq j} \Delta_k^{(j) \top} \Delta_k^{(j)} \right\|_{\text{op}}, \left\| \sum_{k \neq j} \Delta_k^{(j)} \Delta_k^{(j) \top} \right\|_{\text{op}} \right\} \\
 & = p(1 - p) \cdot \max \left\{ \| \Delta^{(j)} \|_{\text{op}}^2, \| (\Delta_1^{(j)}, \dots, \Delta_n^{(j)})^\top \|_{\text{op}}^2 \right\}. \quad (74)
 \end{aligned}$$

To obtain the two different tail bounds in the lemma statement, we will show two different upper bounds on (74). For the first tail bound (44) of Lemma 5, note that $\| \Delta^{(j)} \|_{\text{op}}^2 \leq \sqrt{n} \| \Delta^{(j)} \|_{\text{op}} \| \Delta^{(j)} \|_{\text{op}, \infty}$. We have

$$\begin{aligned}
 p(1 - p) \| \Delta^{(j)} \|_{\text{op}}^2 & \leq p \sqrt{n} \| \Delta^{(j)} \|_{\text{op}, \infty} \| \Delta^{(j)} \|_{\text{op}} \\
 & \leq p \sqrt{n} \| \Delta^{(j)} \|_{\text{op}, \infty} \| \Delta^{(j)} \|_{\text{op}}.
 \end{aligned}$$

On the other hand,

$$p(1 - p) \| (\Delta_1^{(j)}, \dots, \Delta_n^{(j)})^\top \|_{\text{op}}^2$$

$$\begin{aligned}
&\leq p\sqrt{n} \left\| (\Delta_1^{(j)}, \dots, \Delta_n^{(j)})^\top \right\|_{\text{op}} \|\Delta^{(j)}\|_{\text{op},\infty} \\
&\leq p\sqrt{n} \left\| (\Delta_1^{(j)\top}, \dots, \Delta_n^{(j)\top}) \right\|_{\text{F}} \|\Delta^{(j)}\|_{\text{op},\infty} \\
&= p\sqrt{n} \|\Delta^{(j)}\|_{\text{F}} \|\Delta^{(j)}\|_{\text{op},\infty} \\
&\leq p\sqrt{nd} \|\Delta^{(j)}\|_{\text{op}} \|\Delta^{(j)}\|_{\text{op},\infty}.
\end{aligned}$$

Then (74) can be upper bounded by $p\sqrt{nd} \|\Delta^{(j)}\|_{\text{op}} \|\Delta^{(j)}\|_{\text{op},\infty}$.

Invoking Lemma 11, we have

$$\begin{aligned}
&\mathbb{P} \left(\left\| \sum_{k \neq j} (A_{jk} - p) \Delta_k^{(j)} \right\|_{\text{op}} \geq t \mid \Delta^{(j)} \right) \leq 2d \times \\
&\exp \left(-\frac{t^2/2}{p\sqrt{nd} \|\Delta^{(j)}\|_{\text{op}} \|\Delta^{(j)}\|_{\text{op},\infty} + \|\Delta^{(j)}\|_{\text{op},\infty} t/3} \right).
\end{aligned}$$

Then, by triangle inequality,

$$\begin{aligned}
&\mathbb{P} \left(\left\| \sum_{k \neq j} A_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq p\sqrt{n} \|\Delta^{(j)}\|_{\text{op}} + t \mid \Delta^{(j)} \right) \leq 2d \\
&\times \exp \left(-\frac{t^2/2}{p\sqrt{nd} \|\Delta^{(j)}\|_{\text{op}} \|\Delta^{(j)}\|_{\text{op},\infty} + \|\Delta^{(j)}\|_{\text{op},\infty} t/3} \right).
\end{aligned}$$

For the second tail bound (45) of Lemma 5, one can see that

$$\begin{aligned}
&p(1-p) \cdot \max \left\{ \|\Delta^{(j)}\|_{\text{op}}^2, \left\| (\Delta_1^{(j)\top}, \dots, \Delta_n^{(j)\top}) \right\|_{\text{op}}^2 \right\} \\
&\leq np \|\Delta^{(j)}\|_{\text{op},\infty}^2.
\end{aligned}$$

We then follow the same steps as shown above to obtain (45). \square

Proof of Lemma 6: We follow the proof of Lemma 5. We first have

$$\begin{aligned}
&\left\| \sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}} \\
&\leq \left\| \sum_{k \neq j} (A_{jk} - p) (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}} + p \left\| \sum_{k \neq j} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}} \\
&\leq \left\| \sum_{k \neq j} (A_{jk} - p) (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}} + p \|\Delta^{(j)}\|_{\text{op}}^2.
\end{aligned}$$

Note that for each $k \in [n]$, $\mathbb{E} \left((A_{jk} - p) (\Delta_k^{(j)})^\top \Delta_k^{(j)} \mid \Delta^{(j)} \right) = 0$ and $\left\| (A_{jk} - p) (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}} \leq \left\| \Delta_k^{(j)} \right\|_{\text{op}}^2 \leq \|\Delta^{(j)}\|_{\text{op},\infty}^2$. In addition,

$$\begin{aligned}
&\left\| \mathbb{E} \left[\sum_{k \neq j} (A_{jk} - p)^2 (\Delta_k^{(j)})^\top \Delta_k^{(j)} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \mid \Delta^{(j)} \right] \right\|_{\text{op}} \\
&= p(1-p) \left\| \sum_{k \neq j} (\Delta_k^{(j)})^\top \Delta_k^{(j)} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}} \\
&\leq p(1-p) \left\| \begin{bmatrix} \Delta_1^{(j)\top} & \dots & \Delta_n^{(j)\top} \end{bmatrix} \right\|_{\text{op}}
\end{aligned}$$

$$\begin{aligned}
&\times \begin{bmatrix} \Delta_1^{(j)} \Delta_1^{(j)\top} & & \\ & \dots & \\ & & \Delta_n^{(j)} \Delta_n^{(j)\top} \end{bmatrix} \begin{bmatrix} \Delta_1^{(j)} \\ \vdots \\ \Delta_n^{(j)} \end{bmatrix} \Big\|_{\text{op}} \\
&\leq p(1-p) \left\| \begin{bmatrix} \Delta_1^{(j)} \Delta_1^{(j)\top} & & \\ & \dots & \\ & & \Delta_n^{(j)} \Delta_n^{(j)\top} \end{bmatrix} \right\|_{\text{op}} \|\Delta^{(j)}\|_{\text{op}}^2 \\
&= p(1-p) \left(\max_{k \in [n]} \left\| \Delta_k^{(j)} \Delta_k^{(j)\top} \right\|_{\text{op}} \right) \|\Delta^{(j)}\|_{\text{op}}^2 \\
&\leq p(1-p) \|\Delta^{(j)}\|_{\text{op},\infty}^2 \|\Delta^{(j)}\|_{\text{op}}^2 \leq p \|\Delta^{(j)}\|_{\text{op},\infty}^2 \|\Delta^{(j)}\|_{\text{op}}^2.
\end{aligned}$$

Hence, by Lemma 11,

$$\begin{aligned}
&\mathbb{P} \left(\left\| \sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}} \geq p \|\Delta^{(j)}\|_{\text{op}}^2 + t \mid \Delta^{(j)} \right) \\
&\leq 2d \exp \left(-\frac{t^2/2}{p \|\Delta^{(j)}\|_{\text{op},\infty}^2 \|\Delta^{(j)}\|_{\text{op}}^2 + \|\Delta^{(j)}\|_{\text{op},\infty}^2 t/3} \right).
\end{aligned}$$

\square

Proof of Lemma 7: Consider any $j \in [n]$. Note that $\{A_{jk}\}_{k \neq j}$, $\{W_{jk}\}_{k \neq j}$, and $\{\Delta_k^{(j)}\}_{k \neq j}$ are mutually independent of each other. By Lemma 12, we have

$$\begin{aligned}
&\sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)} \Big| \{A_{jk}\}_{k \neq j}, \Delta^{(j)} \\
&\stackrel{d}{=} E \left(\sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right)^{\frac{1}{2}} \Big| \{A_{jk}\}_{k \neq j}, \Delta^{(j)}
\end{aligned}$$

where $E \sim \mathcal{MN}(0, I_d, I_d)$ and is independent of $\{A_{jk}\}_{k \neq j}$ and $\{\Delta_k^{(j)}\}_{k \neq j}$. Since $\sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)}$ is symmetric and positive semi-definite, its square root is well-defined. By Lemma 8, there exists some constant $c > 0$, such that for any $t \geq 4\sqrt{d} \left\| \sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}}^{\frac{1}{2}}$, we have

$$\begin{aligned}
&\mathbb{P} \left(\left\| \sum_{k \neq j} A_{jk} W_{jk} \Delta_k^{(j)} \right\|_{\text{op}} \geq t \mid \{A_{jk}\}_{k \neq j}, \Delta^{(j)} \right) \\
&= \mathbb{P} \left(\left\| E \left(\sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right)^{\frac{1}{2}} \right\|_{\text{op}} \geq t \mid \{A_{jk}\}_{k \neq j}, \Delta^{(j)} \right) \\
&\leq \mathbb{P} \left(\|E\|_{\text{op}} \left\| \left(\sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right)^{\frac{1}{2}} \right\|_{\text{op}} \geq t \mid \{A_{jk}\}_{k \neq j}, \Delta^{(j)} \right) \\
&= \mathbb{P} \left(\|E\|_{\text{op}} \left\| \sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}}^{\frac{1}{2}} \geq t \mid \{A_{jk}\}_{k \neq j}, \Delta^{(j)} \right) \\
&\leq 2 \exp \left(-\frac{ct^2}{\left\| \sum_{k \neq j} A_{jk} (\Delta_k^{(j)})^\top \Delta_k^{(j)} \right\|_{\text{op}}} \right).
\end{aligned}$$

\square

Lemma 11 (Theorem 1.6 of [26]) Consider a finite sequence $\{S_k\}_{k=1}^n$ of independent random square matrices with dimension $d \times d$. Assume that each random matrix satisfies

$$\mathbb{E}S_k = 0 \text{ and } \|S_k - \mathbb{E}S_k\|_{\text{op}} \leq L, \forall k \in [n].$$

Define

$$V := \max \left\{ \left\| \sum_k \mathbb{E}[S_k S_k^\top] \right\|_{\text{op}}, \left\| \sum_k \mathbb{E}[S_k^\top S_k] \right\|_{\text{op}} \right\}.$$

Then

$$\mathbb{P} \left(\left\| \sum_k S_k \right\|_{\text{op}} \geq t \right) \leq 2d \cdot \exp \left(-\frac{t^2/2}{V + Lt/3} \right),$$

for any $t \geq 0$.

Lemma 12: Consider any $W, W', \Delta \in \mathbb{R}^{d \times d}$ such that $W, W' \stackrel{\text{iid}}{\sim} \mathcal{MN}(0, I_d, I_d)$ and Δ is a fixed matrix. We have

$$W\Delta \sim \mathcal{MN}(0, I_d, \Delta^\top \Delta)$$

and

$$W + W' \sim \mathcal{MN}(0, I_d, \Sigma_1 + \Sigma_2).$$

Proof: The first result in the lemma is a property of the matrix normal distribution. To prove the second statement, note that $Z \sim \mathcal{MN}(0, \Sigma_0, \Sigma_1)$ is equivalent to $\text{vec}(Z) \sim \mathcal{N}(0, \Sigma_1 \otimes \Sigma_0)$. Since $\text{vec}(W) \sim \mathcal{N}(0, \Sigma_1 \otimes I_d)$ and $\text{vec}(W') \sim \mathcal{N}(0, \Sigma_2 \otimes I_d)$, we have $\text{vec}(W) + \text{vec}(W') \sim \mathcal{N}(0, (\Sigma_1 + \Sigma_2) \otimes I_d)$. \square

REFERENCES

- [1] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *IEEE Trans. Med. Imag.*, vol. 21, no. 11, pp. 1421–1439, Nov. 2002.
- [2] A. C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 26–33.
- [3] M. Gao, Z. Lahner, J. Thunberg, D. Cremers, and F. Bernard, "Isometric multi-shape matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 14183–14193.
- [4] E. Maset, F. Arrigoni, and A. Fusiello, "Practical and efficient multi-view matching," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4568–4576.
- [5] S. Agarwal et al., "Building Rome in a day," *Commun. ACM*, vol. 54, no. 10, pp. 105–112, Oct. 2011.
- [6] C. Gao and A. Y. Zhang, "Iterative algorithm for discrete structure recovery," *Ann. Statist.*, vol. 50, no. 2, pp. 1066–1094, Apr. 2022.
- [7] X. Zhou, M. Zhu, and K. Daniilidis, "Multi-image matching via fast alternating minimization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4032–4040.
- [8] Y. Chen and E. J. Candès, "The projected power method: An efficient algorithm for joint alignment from pairwise differences," *Commun. Pure Appl. Math.*, vol. 71, no. 8, pp. 1648–1714, Aug. 2018.
- [9] Q.-X. Huang and L. Guibas, "Consistent shape maps via semidefinite programming," *Comput. Graph. Forum*, vol. 32, no. 5, pp. 177–186, 2013.
- [10] Y. Chen, L. Guibas, and Q. Huang, "Near-optimal joint object matching via convex relaxation," in *Proc. Int. Conf. Mach. Learn.*, Feb. 2014, pp. 100–108.
- [11] D. Pachauri, R. Kondor, and V. Singh, "Solving the multi-way matching problem by permutation synchronization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013.
- [12] S. Ling, "Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods," *Appl. Comput. Harmon. Anal.*, vol. 60, pp. 20–52, Sep. 2022.
- [13] Y. Shen, Q. Huang, N. Srebro, and S. Sanghavi, "Normalized spectral map synchronization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016.
- [14] H. W. Kuhn, "The Hungarian method for the assignment problem," *Nav. Res. Logistics Quart.*, vol. 2, nos. 1–2, pp. 83–97, Mar. 1955.
- [15] D. Bean, P. J. Bickel, N. El Karoui, and B. Yu, "Optimal M-estimation in high-dimensional regression," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 36, pp. 14563–14568, Sep. 2013.
- [16] E. Abbe, J. Fan, K. Wang, and Y. Zhong, "Entrywise eigenvector analysis of random matrices with low expected rank," *Ann. Statist.*, vol. 48, no. 3, p. 1452, Jun. 2020.
- [17] A. Y. Zhang, "Exact minimax optimality of spectral methods in phase synchronization and orthogonal group synchronization," *Ann. Statist.*, vol. 52, no. 5, pp. 2112–2138, Oct. 2024.
- [18] Y. Chen, Y. Chi, J. Fan, and C. Ma, "Spectral methods for data science: A statistical perspective," *Found. Trends Mach. Learn.*, vol. 14, no. 5, pp. 566–806, 2021.
- [19] M. Inaba, N. Katoh, and H. Imai, "Applications of weighted Voronoi diagrams and randomization to variance-based K-clustering: (Extended abstract)," in *Proc. 10th Annu. Symp. Comput. Geometry*, 1994, pp. 332–339.
- [20] D. Aloise, A. Deshpande, P. Hansen, and P. Popat, "NP-hardness of Euclidean sum-of-squares clustering," *Mach. Learn.*, vol. 75, pp. 245–248, Jan. 2009.
- [21] M. Mahajan, P. Nimbhorkara, and K. Varadarajan, "The planar K-means problem is NP-hard," *Theor. Comput. Sci.*, vol. 442, pp. 13–21, Jul. 2012.
- [22] A. Kumar, Y. Sabharwal, and S. Sen, "Linear-time approximation schemes for clustering problems in any dimensions," *J. ACM*, vol. 57, no. 2, pp. 1–32, Jan. 2010.
- [23] A. Y. Zhang and H. Y. Zhou, "Leave-one-out singular subspace perturbation analysis for spectral clustering," *Ann. Statist.*, vol. 52, no. 5, pp. 2004–2033, Oct. 2024.
- [24] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science*, vol. 47. Cambridge, U.K.: Cambridge Univ. Press, 2018.
- [25] J. Lei and A. Rinaldo, "Consistency of spectral clustering in stochastic block models," *Ann. Statist.*, vol. 43, no. 1, pp. 215–237, Feb. 2015.
- [26] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Found. Comput. Math.*, vol. 12, no. 4, pp. 389–434, Aug. 2012.

Duc Nguyen received the B.S. degree in computer science from Tufts University in 2019 and the Ph.D. degree in computer science from the University of Pennsylvania in 2024.

Anderson Ye Zhang received the B.S. degree in statistics from Zhejiang University, Hangzhou, China, in 2013, and the Ph.D. degree in statistics and data science from Yale University in 2018. He was a William H. Kruskal Instructor with the Department of Statistics, University of Chicago, in 2018. Since 2019, he has been an Assistant Professor of statistics and data science with the University of Pennsylvania.