Supplemental Material to SMAC: Simultaneous Mapping and Clustering Using Spectral Decompositions

Chandrajit Bajaj[1], Tingran Gao[3], Zihang He[2], Qixing Huang[1], and Zhenxiao Liang[2]

[1]Department of Computer Science, The University of Texas at Austin
[2]Institute for Interdisciplinary Information Sciences, Tsinghua Univesity
[3]Department of Statistics, The University of Chicago

Abstract

This document contains all the proofs to the main paper on SMAC: Simultaneous Mapping and Clustering Using Spectral Decomposition.

1 Organization of the Supplemental Material

We organize our paper's supplemental material as follows. In Section 2, we present a set of tools for analyzing the stability of leading eigen-vectors of a perturbed data matrix. These stability bounds are expressed with respect to the algebraic constants of the original matrix and the perturbation matrix. In Section 3, we study how these stability bounds shape out under different noise models of the perturbation matrix. We then use these results to analyze the exact recovery conditions of PermSMAC in Section 4.

2 Stability of Eigen-decomposition in the Deterministic Setting

In this Section, we present our framework for analyzing the stability of eigen-decompositions. The framework is based on a few key lemmas regarding the stability of eigenvalues and eigenvectors (Section 2.1). Their proofs are deferred to Sections 2.2- 2.5.

2.1 Key Lemmas

Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$, we are interested in controlling the stability of the leading eigenvectors of a symmetric block matrix $A\otimes(I_m-\frac{1}{m}\mathbf{1}\mathbf{1}^T) \in \mathbb{R}^{nm\times nm}$ under a symmetric block noise matrix $N \in \mathbb{R}^{nm\times nm}$.

To state the eigen-decomposition stability problem. Let $\lambda_i, 1 \leq i \leq n$ and $\boldsymbol{s}_i, 1 \leq i \leq n$ be the eigenvalues and corresponding eignevectors of A, respectively. Similarly, let $\mu_i, 1 \leq i \leq nm$ and $\boldsymbol{u}_i, 1 \leq i \leq nm$ be the eigenvalues and corresponding eigenvectors of $\overline{A} = A \otimes (I_m - \frac{1}{m}\mathbf{1}\mathbf{1}^T) + N$, respectively. We are interested in bounding the difference between $\lambda_i, 1 \leq i \leq k$ and $\mu_{(i-1)m+j}, 1 \leq i \leq k, 1 \leq j \leq m$ as well as the difference between the column space $(\boldsymbol{s}_1, \cdots, \boldsymbol{s}_k) \otimes H_m$ and that of $U_{k(m-1)} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{k(m-1)})$, where $H_m \in \mathbb{R}^{m\times(m-1)}$ is a basis matrix of $I - \frac{1}{m}\mathbf{1}\mathbf{1}^T$. For convenience, we use $S_k = (\boldsymbol{s}_1, \cdots, \boldsymbol{s}_k)$ to denote the leading k eigenvectors of A, and use $\overline{S}_k = (\boldsymbol{s}_{k+1}, \cdots, \boldsymbol{s}_n)$ to collect its remaining eigenvectors. First, we present the following result regarding eigenvalue stability:

Lemma 2.1. *(**Eigenvalue Stability.**) Denote $E_{11} = (S_k \otimes H_m)^T N (S_k \otimes H_m)$. Suppose*

$$\|N\| + \|E_{11}\| < \lambda_k - \lambda_{k+1},$$

then we have for every $1 \leq i \leq k, 1 \leq j \leq m-1, p = (i-1)(m-1)+j$,

$$-\|E_{11}\| \leq \sigma_p - \lambda_i \leq \|E_{11}\| + \frac{\|N\| + \|E_{11}\|}{\lambda_i - \lambda_{k+1}}(\|N\| - \|E_{11}\|). \tag{1}$$

Proof: See Section 2.2. □

Remark 1. Note that (1) is tighter than the well-known Weyl's inequality $|\lambda_1 - \sigma_i| \leq \|N\|$, as

$$\|E_{11}\| + \frac{\|N\| + \|E_{11}\|}{\lambda_i - \lambda_{k+1}}(\|N\| - \|E_{11}\|) < \|E_{11}\| + (\|N\| - \|E_{11}\|) = \|N\|.$$

In particular, as $\|E_{11}\| \ll \|N\|$ in most cases, so our bound is significantly better than the Weyl's inequality when $\|N\| + \|E_{11}\| \ll \lambda_k - \lambda_{k+1}$.

Moreover, our result is also tighter than that of [Eldridge et al., 2018], which introduces a similar stability bound:

$$-\|E_{11}\| \leq \sigma_p - \lambda_i \leq \|E_{11}\| + \frac{\|N\|^2}{\lambda_i - \lambda_{k+1} - \|N\| + \|E_{11}\|}.$$

(1) is tighter as $\lambda_i - \lambda_{k+1} - \|N\| + \|E_{11}\| \leq \lambda_i - \lambda_{k+1}$ and $\|N\|^2 - \|E_{11}\|^2 \leq \|N\|^2$.

Remark 2. To gain additional intuitions on (1), let us consider the special case where $k = m = 1$, and let $\lambda_i(A)$ be the i-th greatest eigenvalues of symmetric matrix A. For the matrix function $A(t) = A + tN$ depending on the parameter t with $0 \leq t \leq 1$, let $\lambda_i(t) = \lambda_i(A(t))$, and $\boldsymbol{u}_i(t)$ corresponds to eigenvalue $\lambda_i(t)$. First of all, it is easy to show that (e.g., through differentiating $A(t)\boldsymbol{u}_1(t) = \lambda_1(t)\boldsymbol{u}_1(t)$ and utilizing $\|\boldsymbol{u}(t)\| = 1$)

$$\lambda_1'(t) = \boldsymbol{u}_1(t)^T A'(t) \boldsymbol{u}_1(t) = \boldsymbol{u}_1(t)^T N \boldsymbol{u}_1(t). \tag{2}$$

Recall that $E_{11} = S_k^T N S_k = \boldsymbol{u}_1(t)^T N \boldsymbol{u}_1(t)$ is a scalar in this special case. Hence the first order expansion of $\lambda_1(t)$ is $\lambda_1(t) \approx \lambda(A) + tE_{11}$. We now proceed to differentiate (2), which gives

$$\lambda_1''(t) = \boldsymbol{u}_1'(t)^T N \boldsymbol{u}_1(t) + \boldsymbol{u}_1(t)^T N \boldsymbol{u}_1'(t) = 2\boldsymbol{u}_1(t)^T N \boldsymbol{u}_1'(t). \tag{3}$$

Note that the expression of $\boldsymbol{u}_1'(t)$ is given by

$$\boldsymbol{u}_1'(t) = \sum_{i=2}^{n} \frac{\boldsymbol{u}_i(t)^T N \boldsymbol{u}_1(t)}{\lambda(t) - \lambda_i(t)} \boldsymbol{u}_i(t).$$

Substituting it into formula (3) yields the second order approximation of $\lambda(t)$ would be

$$\lambda_1(t) \approx \lambda_1(A) + E_{11}t + \sum_{i=2}^{n} \frac{|\boldsymbol{u}_i(t)^T N \boldsymbol{u}_1(t)|^2}{\lambda_1(A) - \lambda_i(A)} t^2,$$

or loosely speaking,

$$\lambda_1(t) \approx \lambda_1(A) + E_{11}t + \sum_{i=2}^{n} \frac{|\boldsymbol{u}_i(t)^T N \boldsymbol{u}_1(t)|^2}{\lambda_1(A) - \lambda_2(A)} t^2,$$

in which

$$\begin{aligned}\sum_{i=2}^{n} |\boldsymbol{u}_i(t)^T N \boldsymbol{u}_1(t)|^2 &= \sum_{i=1}^{n} \left|\boldsymbol{u}_i(t)^T \left(N \boldsymbol{u}_1(t)\right)\right|^2 - \left(\boldsymbol{u}_1(t)^T N \boldsymbol{u}_1(t)\right)^2 \\ &= |N \boldsymbol{u}_1(t)|^2 - E_{11}^2 \\ &\leq \|N\|^2 - \|E_{11}\|^2.\end{aligned}$$

Thus the second-order Taylor expansion of $\lambda(t)$ implies

$$\|E_{11}\| + \frac{\|N\|^2 - \|E_{11}\|^2}{\lambda_1(A) - \lambda_2(A)}$$

is a good approximation for $|\lambda_1(A+N)-\lambda_1(A)|$, which is exactly what we proposed in formula (1).

However, it must be pointed out that this expansion is just an approximation rather than a real bound but it provides a good insight why (1) comes out. Besides, it would also be hard to generalize this expansion idea to $k>1$ since $\|E_{11}\|$ will no longer be a scalar if $k>1$.

To characterize the difference between $S_k \otimes H_m$ and $U_{k(m-1)} := (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{k(m-1)})$, we consider the following decomposition of $U_{k(m-1)}$:

$$U_{k(m-1)} = (S_k \otimes H_m)X + Y,$$

where

$$X \in \mathbb{R}^{k(m-1)\times k(m-1)}, \quad S_k^T Y = 0.$$

In other words, $(S_k \otimes H_m)X$ is the projection of $U_{k(m-1)}$ onto the column space of $S_k \otimes H_m$, and Y is the projection of $U_{k(m-1)}$ onto the dual space $\overline{S}_k \otimes H_m$. Intuitively, we say $U_{k(m-1)}$ is stable if Y is small and X is close to a unitary matrix (which defines all orthogonal basis of a linear space). The following Lemma provides a bound on the difference between X and a unitary matrix:

Lemma 2.2. ***(Controlling** X.)* *Denote* $E_{11} = (S_k \otimes H_m)^T N (S_k \otimes H_m)$. *Suppose* $\lambda_k - \lambda_{k+1} > \|N\| + \|E_{11}\|$. *Then there exists a unitary matrix* $R \in O(k(m-1))$ *so that*

$$\|X - R\| \leq 1 - \sqrt{1 - \left(\frac{\|N\|}{\lambda_k - \lambda_{k+1} - \|E_{11}\|}\right)^2}. \tag{4}$$

In particular,

$$\|X - R\| \leq \frac{\|N\|^2}{(\lambda_k - \lambda_{k+1} - \|E_{11}\|)^2}. \tag{5}$$

Proof: See Appendix 2.3. □

It is easy to derive an upper bound on $\|Y\|$ using Lemma 2.2. However, our analysis requires bounding individual blocks $Y_i, 1 \leq i \leq n$, making such bounds insufficient for our purpose. To this end, we introduce the following expression of Y, from which we will derive block-wise bounds.

Lemma 2.3. ***(Controlling** Y.)* *Denote* $\Sigma_i = \mathrm{diag}(\sigma_{(i-1)(m-1)+1}, \cdots, \sigma_{i(m-1)}), 1 \leq i \leq k$, $\overline{\Lambda} = \mathrm{diag}(\lambda_{k+1}, \cdots, \lambda_n)$, *and* $\overline{S}_k = (\boldsymbol{s}_{k+1}, \cdots, \boldsymbol{s}_n)$. *Let*

$$B_i = (\overline{S}_k \otimes H_m)(\lambda_i I - \overline{\Lambda} \otimes I_{m-1})^{-1} (\overline{S}_k \otimes H_m)^T, \qquad 1 \leq i \leq k. \tag{6}$$

Suppose $\|N\| + \|E_{11}\| < \lambda_k - \lambda_{k+1}$, *then*

$$Y_i := (\boldsymbol{y}_{(i-1)m+1}, \cdots, \boldsymbol{y}_{im}) = \sum_{l=0}^{\infty} \big((I - B_i N)^{-1} B_i\big)^{l+1} N \big(S_k \otimes H_m\big) X_i (\lambda_i I_m - \Sigma_i)^l. \tag{7}$$

Proof: See Appendix 2.4. □

We now present the last lemma which applies (7) to obtain a L^∞-type bound on blocks of Y_i. Specifically, let $E_b = \boldsymbol{e}_b \otimes H_m \in \mathbb{R}^{n(m-1)\times(m-1)}$. The following Lemma provides a way to bound $\|E_b^T Y\|$:

Lemma 2.4. ***(Bounding** L^∞**-norm of** $\boldsymbol{Y}$)* *Given* $i \in \{1, \cdots, k\}$. *Let* B_i *be defined by (6). Suppose there are small constants* $\epsilon_1, \epsilon_2, \epsilon_3, \delta < 1$ *such that the following four conditions are satisfied:*

- $\|B_i\|\|N\| \leq \epsilon_1$.
- $\|B_i\|\|\lambda_i I_m - \Sigma_i\| \leq \epsilon_2$.
- $\exists j_0 \geq 0,\ m_0 \geq 0$, *s.t.*, $\forall\ 0 \leq j \leq j_0,\ i_l \geq 0, 1 \leq l \leq j+1$, *where* $0 \leq \sum_{l=1}^{j+1} i_l \leq m_0$,

$$\max_{1\leq b\leq n} \|E_b^T (B_i N)^{i_1} B_i \cdots (B_i N)^{i_{j+1}} B_i N (S_k \otimes H_m)\| \leq \epsilon_3 \|B_i\|^j \cdot \delta^{\sum_{l=1}^{j+1} i_l + 1}. \tag{8}$$

- $\epsilon_2 + \delta < 1$.

Then

$$\max_{1\le b\le n} \|E_b^T Y_i\| \le \|X_i\| \cdot \left(\frac{\epsilon_1}{1-\epsilon_1-\epsilon_2} \cdot \left((\frac{\epsilon_2}{1-\epsilon_1})^{j_0+1} + (\frac{\epsilon_1}{1-\epsilon_2})^{m_0} \right) + \frac{\delta}{1-\epsilon_2-\delta}\epsilon_3 \right). \tag{9}$$

Proof: See Appendix 2.5. □

Remark 3. As we will see later, the dominant term in (8) is $\|E_1^T BNS\|$, which can be controlled using standard concentration bound. The technical difficulty is how to extend it to high order moments. Later we will show how to achieve this goal by controlling power moments. Note that (8) does not incur a strong bound. For example, in randomized models we consider in this paper, $\|B\|\|N\| = O\Big(\frac{1}{\sqrt{\log(n)}}\Big)$. However, the right-hand side of (8) only decays at a geometric rate.

Remark 4. It turns out the major task in terms of controlling eigen-decomposition stability is to provide bounds on E_{11}, N and the left-hand sides in (8). This is the goal of the next two sections.

2.2 Proof of Lemma 2.1

Denote $\Lambda_k = (\lambda_1, \cdots, \lambda_k)$ and $\overline{\Lambda}_k = (\lambda_{k+1}, \cdots, \lambda_n)$. It is clear that we can decompose A as

$$A = S_k \Lambda_k S_k^T + \overline{S}_k \overline{\Lambda}_k \ \overline{S}_k^T.$$

For convenience we set $p = (i-1)(m-1)+j$ and define $L_m = I_m - \frac{1}{m}\mathbf{1}\mathbf{1}^T$. To control $|\lambda_i - \sigma_{(i-1)(m-1)+j}|, 1 \le i \le k,\ 1 \le j \le m-1$, we consider an eigen-decomposition of $A \otimes L_m + N$ with respect to basis spanned by $S_k \otimes H_m$ and $\overline{S}_k \otimes H_m$. Introduce matrices

$$E_{11} = (S_k \otimes H_m)^T N (S_k \otimes H_m),$$
$$E_{21} = (\overline{S}_k \otimes H_m)^T N (S_k \otimes H_m),$$
$$E_{22} = (\overline{S}_k \otimes H_m)^T N (\overline{S}_k \otimes H_m).$$

Using unitary matrix $S \otimes H_m$ to change the basis for matrices $A \otimes L_m$ and N, respectively, we obtain that

$$(S \otimes H_m)^T (A \otimes L_m)(S \otimes H_m) = \begin{bmatrix} \Lambda_k & 0 \\ 0 & \overline{\Lambda}_k \end{bmatrix} \otimes I_{m-1}$$
$$(S \otimes H_m)^T N (S \otimes H_m) = \begin{bmatrix} E_{11} & E_{21}^T \\ E_{21} & E_{22} \end{bmatrix}$$

Observe that the matrix $A \otimes L_m + N - \mu I$ is congruent to

$$B(\mu) = \begin{bmatrix} \Lambda_k \otimes I_{m-1} + E_{11} - \mu I + E_{21}^T (\mu I - \overline{\Lambda}_k - E_{22})^{-1} E_{21} & 0 \\ 0 & \overline{\Lambda}_k \otimes I_{m-1} + E_{22} - \mu I \end{bmatrix}$$

and has the same inertia as well, which can be verified from the identity

$$\begin{aligned} & (S \otimes H_m)^T (A \otimes L_m + N - \mu I)(S \otimes H_m) \\ = & \begin{bmatrix} \Lambda_k \otimes I_{m-1} + E_{11} - \mu I & E_{21}^T \\ E_{21} & \overline{\Lambda}_k \otimes I_{m-1} + E_{22} - \mu I \end{bmatrix} \\ = & \begin{bmatrix} I & P \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} \Lambda_k \otimes I_{m-1} + E_{11} - \mu I + E_{21}^T (\mu I - \overline{\Lambda}_k \otimes I_{m-1} - E_{22})^{-1} E_{21} & 0 \\ 0 & \overline{\Lambda}_k \otimes I_{m-1} + E_{22} - \mu I \end{bmatrix} \begin{bmatrix} I & 0 \\ P^T & I \end{bmatrix}^{-1}, \end{aligned}$$

in which $P = E_{21}^T (\mu I - \overline{\Lambda}_k \otimes I_{m-1} - E_{22})^{-1}$. Hence by letting $\mu = \sigma_p$ we have $\lambda_p(A \otimes L_m + N - \sigma_p I) = 0$ and then $\lambda_p(B(\sigma_p)) = 0$, and further more

$$\sigma_p = \lambda_p \left(\begin{bmatrix} \Lambda_k \otimes I_{m-1} & 0 \\ 0 & \overline{\Lambda}_k \otimes I_{m-1} + E_{22} \end{bmatrix} + \begin{bmatrix} E_{11} + E_{21}^T (\sigma_p I - \overline{\Lambda}_k \otimes I_{m-1} - E_{22})^{-1} E_{21} & 0 \\ 0 & 0 \end{bmatrix} \right).$$

However, the Weyl's inequality told that

$$\begin{aligned}\sigma_p - \lambda_i = \sigma_p - \lambda_p(\Lambda \otimes I_{m-1}) &\geq \lambda_{\min}\left(\begin{bmatrix} E_{11} + E_{21}^T(\sigma_p I - \overline{\Lambda}_k \otimes I_{m-1} - E_{22})^{-1}E_{21} & 0 \\ 0 & 0\end{bmatrix}\right) \\ &\geq \min\{\lambda_{\min}\big(E_{11} + E_{21}^T(\sigma_p I - \overline{\Lambda}_k \otimes I_{m-1} - E_{22})^{-1}E_{21}\big), 0\} \\ &\geq -\|E_{11}\| \end{aligned} \tag{10}$$

in which we used the fact that $\sigma_p - \overline{\Lambda}_k \otimes I_{m-1} - E_{22}$ is positive definite, and

$$\begin{aligned}\sigma_p - \lambda_i = \sigma_p - \lambda_p(\Lambda \otimes I_{m-1}) &\leq \lambda_{max}\left(\begin{bmatrix} E_{11} + E_{21}^T(\sigma_p I - \overline{\Lambda}_k \otimes I_{m-1} - E_{22})^{-1}E_{21} & 0 \\ 0 & 0\end{bmatrix}\right) \\ &\leq \|E_{11} + E_{21}^T(\sigma_p I - \overline{\Lambda}_k \otimes I_{m-1} - E_{22})^{-1}E_{21}\| \\ &\leq \|E_{11} + E_{21}^T\big((\sigma_p - \lambda_{k+1})I - E_{22}\big)^{-1}E_{21}\| \end{aligned} \tag{11}$$

It remains to bound $E_{21}^T(\sigma_p I - \overline{\Lambda}_k \otimes I_{m-1} - E_{22})^{-1}E_{21}$. Towards this end, we consider an arbitrary value $\mu > \lambda_{max}(N)$. It is clear that

$$\mu I - \begin{bmatrix} E_{11} & E_{21}^T \\ E_{21} & E_{22}\end{bmatrix} \tag{12}$$

is a positive definite matrix, and $\mu I - E_{22}$ is also positive definite. From the identity

$$\mu I - \begin{bmatrix} E_{11} & E_{21}^T \\ E_{21} & E_{22}\end{bmatrix} = \begin{bmatrix} I & Q \\ 0 & I\end{bmatrix}^{-1} \begin{bmatrix} \mu I - E_{11} - E_{21}^T(\mu I - E_{22})^{-1}E_{21} & 0 \\ 0 & \mu I - E_{22}\end{bmatrix} \begin{bmatrix} I & 0 \\ Q^T & I\end{bmatrix}^{-1}$$

in which $Q = E_{21}^T(E_{22} - \mu I)^{-1}$, we can see that $\mu I - E_{11} - E_{21}^T(\mu I - E_{22})^{-1}E_{21}$ is positive definite as well. Applying it to (12), we obtain that

$$\mu I - E_{11} - E_{21}^T(\mu I - E_{22})^{-1}E_{21} \succeq 0,$$

or equivalently,

$$E_{11} + E_{21}^T(\mu I - E_{22})^{-1}E_{21} \preceq \mu I. \tag{13}$$

Similarly, for $\mu < \lambda_{\min}(N)$, we have

$$E_{11} + E_{21}^T(\mu I - E_{22})^{-1}E_{21} \succeq \mu I. \tag{14}$$

Letting $\mu \to \|N\|$ and $\mu \to -\|N\|$ respectively, we obtain that

$$\begin{aligned} E_{21}^T(\|N\|I - E_{22})^{-1}E_{21} &\preceq \|N\|I - E_{11}, \\ E_{21}^T(\|N\|I + E_{22})^{-1}E_{21} &\preceq \|N\|I + E_{11}. \end{aligned}$$

Here the second inequality arise by taking negative of (14).

As thus for any non-negative number α, β, we have

$$E_{11} + E_{21}^T\big(\alpha(\|N\|I - E_{22})^{-1} + \beta(\|N\|I + E_{22})^{-1}\big)^{-1}E_{21} \preceq (\alpha+\beta)\|N\|I + (1+\beta-\alpha)E_{11}.$$

If we in addition have

$$E_{21}^T\big((\sigma_p - \lambda_{k+1})I - E_{22}\big)^{-1}E_{21} \preceq E_{21}^T\big(\alpha(\|N\|I - E_{22})^{-1} + \beta(\|N\|I + E_{22})^{-1}\big)^{-1}E_{21}, \tag{15}$$

for some $\alpha, \beta \geq 0$, then we reach an upper bound

$$\lambda_{\max}\Big(E_{11} + E_{21}^T\big((\sigma_p - \lambda_{k+1})I - E_{22}\big)^{-1}E_{21}\Big) \leq (\alpha+\beta)\|N\| + |1+\beta-\alpha|\|E_{11}\| \tag{16}$$

To ensure inequality (15), we need to show that the following inequality

$$\frac{1}{\sigma_p - \lambda_{k+1} - x} \leq \frac{\alpha}{\|N\| - x} + \frac{\beta}{\|N\| + x}$$

holds for every eigenvalue x of E_{22}, or in a stronger sense, for all $-\|N\| \le x \le \|N\|$. Define

$$\theta = \beta/\alpha, \varepsilon = \|N\|, a = \|E_{11}\|$$

and then the above condition changes to

$$\alpha \ge \frac{\varepsilon^2 - x^2}{(\sigma_p - \lambda_{k+1} - x)\big((1+\theta)\varepsilon + (1-\theta)x\big)}$$

for all $-\varepsilon \le x \le \varepsilon$. Hence we can take α as

$$\begin{aligned}\alpha_\theta &= \max_{-\varepsilon \le x \le \varepsilon} \frac{\varepsilon^2 - x^2}{(\sigma_p - \lambda_{k+1} - x)\big((1+\theta)\varepsilon + (1-\theta)x\big)} \\ &= \frac{\varepsilon}{(\sigma_p - \lambda_{k+1})(1+\theta)} \max_{-1 \le x \le 1} \frac{1 - x^2}{(1 - \frac{\varepsilon}{\sigma_p - \lambda_{k+1}} x)(1 + \frac{1-\theta}{1+\theta} x)}\end{aligned}$$

and substitute it into inequality (16) we obtain

$$\lambda_{\max}\left(E_{11} + E_{21}^T\big((\sigma_p - \lambda_{k+1})I - E_{22}\big)^{-1} E_{21}\right) \le (1+\theta)\alpha_\theta \varepsilon + |1 + (\theta - 1)\alpha_\theta| a.$$

But it is easy to see that $\alpha_\theta \le \alpha_0 \le \frac{2\varepsilon}{\sigma_p - \lambda_{k+1} + \varepsilon} \le \frac{2\varepsilon}{\lambda_k - \lambda_{k+1}} \le 1$, so the absolute value sign can be removed from the inequality above, and substitute the expression of α_θ into it:

$$\lambda_{\max}\left(E_{11} + E_{21}^T\big((\sigma_p - \lambda_{k+1})I - E_{22}\big)^{-1} E_{21}\right) \le (1+\theta)\alpha_\theta \varepsilon + (\theta - 1)\alpha_\theta a + a \tag{17}$$

To simplify our computation, we introduce the following trigonometric function notations:

$$\begin{aligned}\sin\phi &= \frac{1-\theta}{1+\theta}, & -\frac{\pi}{2} \le \phi \le \frac{\pi}{2}; \\ \sin\psi &= \frac{\varepsilon}{\sigma_p - \lambda_{k+1}}, & 0 \le \psi < \frac{\pi}{6}; \\ \sin\omega &= \frac{a}{\varepsilon}, & 0 \le \omega \le \frac{\pi}{2}\end{aligned} \tag{18}$$

Applying proposition 1 under these notations, we can rewrite α_θ as

$$\alpha_\theta = \frac{\sin\psi}{1+\theta} \cdot \frac{1}{\cos^2 \frac{\phi - \psi}{2}}, \tag{19}$$

and substitute it into (17) to obtain

$$\begin{aligned}\lambda_{\max}\left(E_{11} + E_{21}^T\big((\sigma_p - \lambda_{k+1})I - E_{22}\big)^{-1} E_{21}\right) &\le (1+\theta)\alpha_\theta \varepsilon + (\theta - 1)\alpha_\theta a + a \\ &= a + \frac{\varepsilon \sin\psi}{\cos^2 \frac{\phi - \psi}{2}}(1 - \sin\omega \sin\phi)\end{aligned} \tag{20}$$

By optimizing over θ, namely over ϕ, and applying proposition 2, we have

$$\begin{aligned}&\lambda_{\max}\left(E_{11} + E_{21}^T\big((\sigma_p - \lambda_{k+1})I - E_{22}\big)^{-1} E_{21}\right) \\ \le& a + \varepsilon \sin\psi \min_{-\pi/2 \le \phi \le \pi/2} \frac{1 - \sin\omega \sin\phi}{\cos^2 \frac{\phi - \psi}{2}} \\ =& a + \varepsilon \frac{\sin\psi \cos^2\omega}{1 + \sin\omega \sin\psi} \\ =& \|E_{11}\| + \frac{\|N\|^2 - \|E_{11}\|^2}{\sigma_p - \lambda_{k+1} + \|E_{11}\|} \\ \le& \|E_{11}\| + \frac{\|N\|^2 - \|E_{11}\|^2}{\lambda_i - \lambda_{k+1}},\end{aligned} \tag{21}$$

where in the last step we used the fact $\sigma_p - \lambda_i \ge -\|E_{11}\|$ that has been proved before. □

Proposition 1. *Given two real number α, β with $-\frac{\pi}{2} \le \alpha, \beta \le \frac{\pi}{2}$, we have*

$$\max_{-1\le x\le 1} \frac{1-x^2}{(1+x\sin\alpha)(1+x\sin\beta)} = \frac{1}{\cos^2\frac{\alpha+\beta}{2}}$$

Proof. It is trivial when $x = \pm 1$ so we always assume $-1 < x < 1$ in the following. Taking transformation $x = \frac{1-p^2}{1+p^2}, p \in \mathbb{R}\backslash\{0\}$, we have

$$\begin{aligned}
&\frac{(1+x\sin\alpha)(1+x\sin\beta)}{1-x^2}\\
=&\ \frac{1}{4}\left[(1-\sin\alpha)p + (1+\sin\alpha)p^{-1}\right]\left[(1-\sin\beta)p + (1+\sin\beta)p^{-1}\right]\\
=&\ \frac{1}{4}\left[(1-\sin\alpha)(1-\sin\beta)p^2 + 2(1-\sin\alpha\sin\beta) + (1+\sin\alpha)(1+\sin\beta)p^{-2}\right]\\
\ge&\ \frac{1}{2}(1-\sin\alpha\sin\beta+\cos\alpha\cos\beta)\\
=&\ \frac{1}{2}\big(1+\cos(\alpha+\beta)\big)\\
=&\ \cos^2\frac{\alpha+\beta}{2}.
\end{aligned}$$

The proposition follows immediately. □

Proposition 2. *Given ϕ, ψ, ω as define in (18), we claim that*

$$\min_{-\pi/2\le\phi\le\pi/2} \frac{1-\sin\omega\sin\phi}{\cos^2\frac{\phi-\psi}{2}} = \frac{\cos^2\omega}{1+\sin\omega\sin\psi}$$

Proof. Note that

$$\begin{aligned}
&1-\sin\omega\sin\phi\\
=&\ 1-\sin\omega\sin(\phi-\psi+\psi)\\
=&\ 1-\sin\omega\left(\sin\psi\cos(\phi-\psi)+\cos\psi\sin(\phi-\psi)\right)\\
=&\ 1-\sin\omega\left(\sin\psi\left(\cos^2\frac{\phi-\psi}{2}-\sin^2\frac{\phi-\psi}{2}\right)+2\cos\psi\sin\frac{\phi-\psi}{2}\cos\frac{\phi-\psi}{2}\right).
\end{aligned}$$

Define $p = \tan\frac{\phi-\psi}{2}$. p can be taken over interval $[\tan(-\pi/4-\psi/2), \tan(\pi/4-\psi/2)]$ as taking ϕ over $[-\pi/2, \pi/2]$. Some trigonometric calculation shows that $[\tan(-\pi/4-\psi/2), \tan(\pi/4-\psi/2)]$ is equal to $[-\frac{\cos\psi}{1-\sin\psi}, \frac{\cos\psi}{1+\sin\psi}]$. Using the identity we just proved, we have

$$\begin{aligned}
&\frac{1-\sin\omega\sin\phi}{\cos^2\frac{\phi-\psi}{2}}\\
=&p^2+1-\sin\omega\big((1-p^2)\sin\psi+2p\cos\psi\big)\\
=&(1+\sin\omega\sin\psi)p^2-2p\sin\omega\cos\psi+(1-\sin\omega\sin\psi)
\end{aligned}$$

We can see it is a quadratic function about p, so it has a minimum at $p_0 = \frac{\sin\omega\cos\psi}{1+\sin\omega\sin\psi}$. It is clear that $-\frac{\cos\psi}{1-\sin\psi} \le p_0 \le \frac{\cos\psi}{1+\sin\psi}$, hence p_0 can be certainly taken. Further more, this quadratic function has a minimum value $\frac{\cos^2\omega}{1+\sin\omega\sin\psi}$ on this point, which completes our proof. □

2.3 Proof of Lemma 2.2

Also denote $I - \frac{1}{m}\mathbf{1}\mathbf{1}^T$ as L_m. We only prove the first inequality, since the second inequality can be inferred from

$$1-\sqrt{1-u^2} \le u,\ -1 \le u \le 1.$$

Multiply both sides of

$$\big(A \otimes L_m + N\big)\big((S_k \otimes H_m)X + Y\big) = \big((S_k \otimes H_m)X + Y\big)\Sigma$$

by $(\overline{S}_k \otimes L_m)^T$, yielding

$$\begin{aligned}
&(\overline{S}_k \otimes H_m)^T (A \otimes L_m + N)((S_k \otimes H_m)X + Y) = (\overline{S}_k \otimes H_m)^T Y \cdot \Sigma \\
\Leftrightarrow\ &(\overline{S}_k \otimes H_m)^T N((S_k \otimes H_m)X + Y) = (\overline{S}_k \otimes H_m)^T Y \cdot \Sigma - \big((\overline{S}_k\overline{\Lambda}) \otimes H_m\big)^T Y.
\end{aligned} \tag{22}$$

Now we prove the following proposition, which will be used later:

Proposition 3. *Denote* $\Lambda = \operatorname{diag}(\lambda_1, \cdots, \lambda_k)$ *and* $\Sigma = \operatorname{diag}(\sigma_1, \cdots, \sigma_n)$. *Suppose* $\lambda_1 \geq \cdots \lambda_k > \sigma_1 \geq \cdots \geq \sigma_n$, *then for any* $Y \in \mathbb{R}^{n \times k}$, *we have*

$$\|Y\Lambda - \Sigma Y\| \geq (\lambda_k - \sigma_1)\|Y\|.$$

Proof: Without losing generality, we can assume $\sigma_n \geq 0$, since we can always shift $\lambda_i, 1 \leq i \leq k$ and $\sigma_j, 1 \leq j \leq n$ by the same amount without changing the value of $Y\Lambda - \Sigma Y$ and $\lambda_k - \sigma_1$. With this assumption, the proof directly follows from triangle inequality:

$$\|Y\Lambda - \Sigma Y\| \geq \ \|Y\Lambda\| - \|\Sigma Y\| \geq \ \|Y(\lambda_k I_k)\| - \|(\sigma_1 I_n)Y\| = (\lambda_k - \sigma_1)\|Y\|.$$

□

Now let us come back to the proof of Lemma 2.2. Since the columns of $U_{k(m-1)} = (S_k \otimes H_m)X + Y$ are orthogonal, we have

$$\begin{aligned}
\|N\| &\geq \|(\overline{S}_k \otimes H_m)^T N\| \geq \|(\overline{S}_k \otimes H_m)^T N((S_k \otimes H_m)X + Y)\| \\
&= \|(\overline{S}_k \otimes H_m)^T Y \cdot \Sigma - \big(\overline{\Lambda} \otimes I_{m-1}\big) \cdot \big(\overline{S}_k \otimes H_m\big)^T Y\| \\
&\geq \min_{1 \leq i \leq km} |\sigma_i - \lambda_{k+1}| \|\big(\overline{S}_k \otimes H_m\big)^T Y\| \qquad \text{(Applying Proposition 3)} \\
&= \min_{1 \leq i \leq km} |\sigma_i - \lambda_{k+1}| \|Y\| \\
&\geq (\lambda_k - \lambda_{k+1} - \|E_{11}\|)\|Y\|,
\end{aligned} \tag{23}$$

where the last equality is due to Lemma 2.1.

As $V^T V = X^T X + Y^T Y = I_{k(m-1)}$. It follows that

$$\sqrt{1 - \left(\frac{\|N\|}{\lambda_k - \lambda_{k+1} - \|E_{11}\|}\right)^2} \leq \sqrt{1 - \|Y\|^2} = \sigma_{\min}(X) \leq \sigma_{\max}(X) \leq 1.$$

Let $X = V_X \Sigma_X W_X^T$ be the SVD of X. Define $R = V_X W_X^T$. We have

$$\begin{aligned}
\|X - R\| &= \|V_X \Sigma_X W_X^T - V_X W_X^T\| \\
&\leq \|\Sigma_X - I\| \\
&\leq 1 - \sqrt{1 - \left(\frac{\|N\|}{\lambda_k - \lambda_{k+1} - \|E_{11}\|}\right)^2},
\end{aligned}$$

which ends the proof. □

2.4 Proof of Lemma 2.3

First, we have

$$
\begin{aligned}
&(\overline{S}_k \otimes H_m)\big(\lambda_i I - \overline{\Lambda} \otimes I_{m-1} - E_{22}\big)^{-1}(\overline{S}_k \otimes H_m)^T \\
&= (\overline{S}_k \otimes H_m)(\lambda_i I - (\overline{\Lambda} \otimes I_{m-1}))^{-\frac{1}{2}}\Big(I - (\lambda_i I - (\overline{\Lambda} \otimes I_{m-1}))^{-\frac{1}{2}}(\overline{S}_k \otimes H_m)^T N(\overline{S}_k \otimes H_m)(\lambda_i I - (\overline{\Lambda} \otimes I_{m-1}))^{-\frac{1}{2}}\Big)^{-1} \cdot \\
&\quad (\lambda_i I - (\overline{\Lambda} \otimes I_{m-1}))^{-\frac{1}{2}}(\overline{S}_k \otimes H_m)^T \\
&= \sum_{k=0}^{\infty}\Big((\overline{S}_k \otimes H_m)(\lambda_i I - (\overline{\Lambda} \otimes I_{m-1}))^{-1}(\overline{S}_k \otimes H_m)^T N\Big)^k(\overline{S}_k \otimes H_m)(\lambda_i I - (\overline{\Lambda} \otimes I_{m-1}))^{-1}(\overline{S}_k \otimes H_m)^T \\
&= (I - B_i N)^{-1} B_i,
\end{aligned}
$$

in which

$$
B_i = (\overline{S}_k \otimes H_m)(\lambda_i I - \overline{\Lambda} \otimes I_{m-1})^{-1}(\overline{S}_k \otimes H_m)^T
$$

In the above argument, since

$$
\|(\lambda_i I - (\overline{\Lambda} \otimes I_{m-1}))^{-\frac{1}{2}}(\overline{S}_k \otimes H_m)^T N(\overline{S}_k \otimes H_m)(\lambda_i I - (\overline{\Lambda} \otimes I_{m-1}))^{-\frac{1}{2}}\| \le \frac{\|N\|}{\lambda_k - \lambda_{k+1}} < 1,
$$

we can safely apply Taylor expansion.

Now let us consider each column of Y. Denote $p = (i-1)\cdot(m-1)+j$. By solving linear system for X, Y, we have

$$
\begin{aligned}
\boldsymbol{y}_p =&(\overline{S}_k \otimes H_m)\left(\sigma_p I - (\overline{\Lambda} \otimes I_{m-1} + E_{22})\right)^{-1}(\overline{S}_k \otimes H_m)^T N(S_k \otimes H_m)\boldsymbol{x}_p \\
=&(\overline{S}_k \otimes H_m)\left(\lambda_i I - (\overline{\Lambda} \otimes I_{m-1} + E_{22}) - (\lambda_i - \sigma_p)I\right)^{-1}(\overline{S}_k \otimes H_m)^T N(S_k \otimes H_m)\boldsymbol{x}_p \\
=&\sum_{k=0}^{\infty}(\overline{S}_k \otimes H_m)\Big((\lambda_i I - (\overline{\Lambda} \otimes I_{m-1} + E_{22}))^{-1}\Big)^{k+1}(\overline{S}_k \otimes H_m)^T N(S_k \otimes H_m)\boldsymbol{x}_p(\lambda_i - \sigma_p)^k \\
=&\sum_{k=0}^{\infty}\Big((\overline{S}_k \otimes H_m)\big(\lambda_i I - ((\overline{\Lambda} \otimes I_{m-1}) + E_{22})\big)^{-1}(S \otimes H_m)^T\Big)^{k+1} N(\overline{S}_k \otimes H_m)\boldsymbol{x}_p(\lambda_i - \sigma_p)^k \qquad (24) \\
=&\sum_{k=0}^{\infty}\Big((I - B_i N)^{-1}B_i\Big)^{k+1} N(S_k \otimes H_m)\boldsymbol{x}_p(\lambda_i - \sigma_p)^k
\end{aligned}
$$

Like before, from $\|N\| + \|E_{11}\| < \lambda_k - \lambda_{k+1}$ it is easy to check that

$$
\begin{aligned}
&(\lambda_i - \sigma_p)\|(\overline{S}_k \otimes H_m)\big(\lambda_i I - (\overline{\Lambda} \otimes I_{m-1} + E_{22})\big)^{-1}(\overline{S}_k \otimes H_m)^T\| \\
\le&(\lambda_i - \sigma_p)\|\big(\lambda_i I - (\overline{\Lambda} \otimes I_{m-1} + E_{22})\big)^{-1}\| \\
\le&\|E_{11}\| \times \frac{1}{\lambda_i - \lambda_{k+1} - \|N\|} \\
\le&\frac{\|E_{11}\|}{\lambda_k - \lambda_{k+1} - \|N\|} < 1,
\end{aligned}
$$

which shows that the power series above would be surely convergent.

Putting this in the matrix form leads to

$$
Y_{\cdot,i} = \sum_{l=0}^{\infty}\Big((I - B_i N)^{-1}B_i\Big)^{l+1} N(S_k \otimes H_m)X_{\cdot,i}(\lambda_i I - \Sigma_k^{(i)})^l.
$$

□

2.5 Proof of Lemma 2.4

First of all, since $\|B_iN\| \le \|B_i\| \cdot \|N\| < 1$, it is clear that

$$(I - B_iN)^{-1}B_i = \sum_{j\ge 0}(B_iN)^j B_i. \tag{25}$$

Consider the following three terms:

$$\begin{aligned}
\delta_1 &:= \sum_{j\ge j_0+1} ((I - B_iN)^{-1}B_i)^{j+1} N(S_k \otimes H_m)X_i(\lambda_i I_m - \Sigma_i)^j. \\
\delta_2 &:= \sum_{j=0}^{j_0} \sum_{m\ge m_0+1} \sum_{\sum_{l=1}^{j+1} i_l = m} \Big(\prod_{l=1}^{j+1}(B_iN)^{i_l}B\Big) N(S_k \otimes H_m)X_i(\lambda_i I_m - \Sigma_i)^j. \\
\delta_3 &:= \sum_{j=0}^{j_0} \sum_{m=0}^{m_0} \sum_{\sum_{l=1}^{j+1} i_l = m} \Big(\prod_{l=1}^{j+1}(B_iN)^{i_l}B\Big) N(S_k \otimes H_m)X_i(\lambda_i I_m - \Sigma_i)^j.
\end{aligned} \tag{26}$$

It is clear that

$$Y_i = \delta_1 + \delta_2 + \delta_3.$$

Now we bound $E_b^T\delta_k, 1 \le k \le 3$. First of all, we have

$$\begin{aligned}
\|E_b^T\delta_1\| \le \|\delta_1\| &\le \sum_{j\ge j_0+1} \|(I - B_iN)^{-1}B_i\|^{j+1}\|N\|\|S_k \otimes H_m\|\|X_i\|\|\lambda_i I_m - \Sigma_i\|^j \\
&\le \sum_{j\ge j_0+1} (\frac{\|B_i\|}{1 - \|B_i\|\|N\|})^{j+1}\|N\|\|S_k \otimes H_m\|\|X_i\|\|\lambda_i I_m - \Sigma_i\|^j \\
&= \frac{\|B_i\|\|N\|\|X\|}{1 - \|B_i\|\|N\|} \sum_{j\ge j_0+1} (\frac{\|B_i\|\|\lambda_i I_m - \Sigma_i\|}{1 - \|B_i\|\|N\|})^j \\
&= \frac{\epsilon_1}{1 - \epsilon_1 - \epsilon_2} \cdot \Big(\frac{\epsilon_2}{1 - \epsilon_1}\Big)^{j_0+1}\|X_i\|.
\end{aligned} \tag{27}$$

Similarly,

$$\begin{aligned}
\|E_b^T\delta_2\| \le \|\delta_2\| &\le \sum_{j=0}^{j_0} \sum_{m\ge m_0+1} \sum_{\sum_{l=1}^{j+1} i_l = m} \Big(\prod_{l=1}^{j+1}(\|B_i\|\|N\|)^{i_l}\|B_i\|\Big)\|N\|\|X_i\|\|\lambda_i I_m - \Sigma_i\|^j \\
&= \sum_{j=0}^{j_0} \sum_{m\ge m_0+1} \sum_{\sum_{l=1}^{j+1} i_l = m} \big(\|B_i\|^{m+j+1}\|N\|^{m+1}\big)\|X_i\|\|\lambda_i I_m - \Sigma_i\|^j \\
&= \sum_{j=0}^{j_0} \sum_{m\ge m_0+1} C_{m+j}^j\big(\|B_i\|^{m+j+1}\|N\|^{m+1}\big)\|X_i\|\|\lambda_i I_m - \Sigma_i\|^j \\
&= \sum_{j=0}^{j_0}(\|B_i\|\|\lambda_i I_m - \Sigma_i\|)^j \sum_{m\ge m_0+1} C_{m+j}^j\big(\|B_i\|\|N\|\big)^{m+1}\|X_i\| \\
&\le \|X_i\| \sum_{m\ge m_0+1} \epsilon_1^{m+1} \sum_{j=0}^{j_0} \epsilon_2^j C_{m+j}^j.
\end{aligned} \tag{28}$$

Since when $-1 < x < 1$,

$$\frac{1}{(1-x)^{j+1}} = \sum_{m=0}^{\infty} x^m C_{m+j}^m.$$

It follows that

$$
\begin{aligned}
\|E_b^T\delta_2\| &\le \|X_i\| \sum_{m\ge m_0+1} \epsilon_1^{m+1} \sum_{j=0}^{j_0} \epsilon_2^j C_{m+j}^j \\
&\le \|X_i\| \sum_{m\ge m_0+1} \epsilon_1^{m+1} \frac{1}{(1-\epsilon_2)^{m+1}} \\
&= \|X_i\| \big(\frac{\epsilon_1}{1-\epsilon_2}\big)^{m_0} \frac{\epsilon_1}{1-\epsilon_1-\epsilon_2}.
\end{aligned} \tag{29}
$$

We use a similar strategy to bound δ_3. In fact,

$$
\begin{aligned}
\|E_b^T\delta_3\| &\le \sum_{j=0}^{j_0} \sum_{m=0}^{m_0} \sum_{\sum_{l=1}^{j+1} i_l = m} \|E_b^T\Big(\prod_{l=1}^{j+1}(B_iN)^{i_l}B_i\Big)N(S_k\otimes H_m)\|\|X_i\|\|\lambda_i I_m - \Sigma_i\|^j \\
&\le \sum_{j=0}^{j_0} \sum_{m=0}^{m_0} \sum_{\sum_{l=1}^{j+1} i_l = m} \|B_i\|^j \epsilon_3 \delta^{m+1} \|X_i\|\|\lambda_i I_m - \Sigma_i\|^j \\
&\le \|X_i\|\epsilon_3 \sum_{j=0}^{j_0} \epsilon_2^j \sum_{m=0}^{m_0} C_{m+j}^j \delta^{m+1} \\
&= \|X_i\|\epsilon_3 \sum_{m=0}^{m_0} \delta^{m+1} \sum_{j=0}^{j_0} C_{m+j}^j \epsilon_2^j \\
&\le \|X_i\|\epsilon_3 \sum_{m=0}^{m_0} \delta^{m+1} \frac{1}{(1-\epsilon_2)^{m+1}} \\
&\le \frac{\delta}{1-\epsilon_2-\delta}\cdot \epsilon_3 \cdot \|X_i\|.
\end{aligned} \tag{30}
$$

Combing (27), (29) and (30), we have

$$
\max_{1\le b\le n} \|E_b^T Y_i\| \le \|X_i\| \cdot \left(\frac{\epsilon_1}{1-\epsilon_1-\epsilon_2}\cdot\Big((\frac{\epsilon_2}{1-\epsilon_1})^{j_0+1} + (\frac{\epsilon_1}{1-\epsilon_2})^{m_0}\Big) + \frac{\delta}{1-\epsilon_2-\delta}\epsilon_3\right),
$$

which ends the proof. □

3 Stability of Eigen-decomposition in Randomized Noise Models

3.1 Key Lemmas

The next Lemma provides various bounds regarding the random matrix model discussed above: Given a constant $c \ge 1$. Suppose $d_{\min} \ge 2c\log(n)$ and $n \ge 10$. Consider a random matrix $N \in \mathbb{R}^{nm\times nm}$ whose blocks are given by

$$
N_{ij} = \begin{cases} \frac{1}{\sqrt{d_i d_j}} X_{ij} & (i,j)\in\mathcal{E} \\ 0 & \text{otherwise} \end{cases}
$$

where $X_{ij}, (i,j)\in\mathcal{E}$ are independent random matrices that satisfy

$$
\mathbb{E}[X_{ij}] = 0, \qquad \|X_{ij}\| \le K_{ij},
$$

where $K_{ij}, (i,j)\in\mathcal{E}$ are positive constants. Let $K = \max_{(i,j)\in\mathcal{E}} K_{ij}$, then

Lemma 3.1. *With probability at least* $1-\frac{m}{n^{2c}}$,

$$
\|N\| \le \frac{CK}{\sqrt{d_{\min}}}
$$

for some absolute constant C.

Lemma 3.2.
Given l positive integer $t_1, \dots, t_l$, and suppose $l \le \frac{\log n}{\log \log n}$, $d_{\min} > \log n$, define a vector $\boldsymbol{\omega}$ as

$$\omega_i = \frac{1}{\sqrt{d_i}} \sum_{j \in \mathcal{N}(i)} \frac{1}{\sqrt{d_j}},$$

and setting a constant $\|B\|_{\boldsymbol{\omega}}$ associated to B and $\boldsymbol{\omega}$ as

$$\|B\|_{\boldsymbol{\omega}} := \|abs(B) \cdot \boldsymbol{\omega}\|_\infty$$

in which $abs(B)$ is obtained by taking absolute values elementwisely from B, then we have

$$\begin{aligned}\|(\boldsymbol{e}_i^T \otimes I_m) \prod_{j=1}^{l} \big((B^{t_j} \otimes I_m)N\big) S_k\|_\infty &\le \Big(\sqrt{\frac{\log n}{d_{\min}}} \cdot CK\Big)^l \cdot \|B\|_{\boldsymbol{\omega}} \cdot \|B\|_\infty^{t-l/2-1} \cdot \|B\|_{\max}^{l/2} \cdot \sqrt{\frac{d_{\min}}{|\mathcal{E}|}} \|U_k\| \\ &\le \left(\sqrt{\frac{\log n}{d_{\min}}} CK\right)^l \|B\|_\infty^{t-l/2} \|B\|_{\max}^{l/2} \sqrt{\frac{d_{\max}}{|\mathcal{E}|}} \|U_k\| \qquad (31)\end{aligned}$$

with probability exceeding $1 - n^{-1/l}$. Recall that in our random model we have $S_k = \boldsymbol{s} \otimes U_k$. Note that we put no special assumption on B here. Namely, B could be different from B_i defined in Section (2.1).

Lemma 3.3. *Given noise matrix N under the model in (3.1), and*

$$K \le \frac{1}{C} \sqrt{\frac{d_{\min}}{\|B_i\| \cdot \|B_i\|_\infty \cdot \log n}},$$

then we have the block-wise bound on Y_i

$$\max_{1 \le b \le n} \|(\boldsymbol{e}_b^T \otimes I_m) Y_i\| \le \frac{\|B_i\|_{\boldsymbol{\omega}}}{\|B_i\|_\infty} \sqrt{\frac{d_{\min}}{|\mathcal{E}|}} \le \sqrt{\frac{d_{\max}}{|\mathcal{E}|}}$$

with probability exceeding $1 - \frac{1}{n^c}$ for some absolute constant C depending on c. Recall that B_i and Y_i were given by

$$B_i := \left(\overline{S}_k (\lambda_i I - \overline{\Lambda})^{-1} \overline{S}_k^T\right) \otimes I_m,$$

$$Y_i := (\boldsymbol{y}_{(i-1)m+1}, \cdots, \boldsymbol{y}_{im}) = \sum_{l=0}^{\infty} \big((I - B_i N)^{-1} B_i\big)^{l+1} N (S_k \otimes I_m) X_i (\lambda_i I_m - \Sigma_i)^l,$$

in Section (2.1) respectively, where $1 \le i \le k$.

3.2 Proof of Lemma 3.1

We provide a bound for $\|N\|$ by giving an estimation of $\mathbb{E}[\mathrm{tr}(N^b)]$ for even positive integers b.

Proposition 4. *Given two positive integer b, w such that $1 \le w \le \frac{b}{2} + 1$, let G be a undirected complete graph of w vertices then the number of cycles of length b in G satisfying the following properties can be bounded above by $C^b b^{b/2}$ (or equivalently $C' b^{b/2+1-w} w!$ for some slightly different absolute constant C') for some absolute constant C:*

- *No self-loop in the cycle;*
- *Each vertex of G appears;*
- *Each edge appearing in the cycle would appear at least two times.*

Proof. Define the following variables associated to w vertices:

- The degrees (not consider the multiplicity of edge) of w vertices $d_1, \dots, d_w$ in the cycle, which is actually the degree of some vertex in the induced undirected graph by the cycle;
- The multi-degrees of w vertices $D_1, \dots, D_w$ in the cycle, which is actually the number of times some vertex appears in the cycle sequence.

We say a leg in the cycle is innovative if the undirected edge of that leg didn't arose before when traversing the cycle, and non-innovative if otherwise. It is clear that

$$d_1 + \cdots + d_w = 2j, \quad D_1 + \cdots + D_w = b, \quad d_t \le D_t.$$

Thus the number of the possibilities of $\{D_t\}$ could be bounded by 2^b, so it suffices to consider only fixed $\{D_t\}$. Define T_p such that T_p/p is the number of vertices with multi-degree p, or equivalently, the number of vertices in the cycle having multi-degree p. It is clear that $\sum_{p=1}^{\infty} T_p = b$.

We try to show that it is possible to reconstruct the cycle by recording a group of arrays. In detail, supposing some valid cycle is $i_1 \to i_2 \to \cdots \to i_b \to i_1$, define an array $\boldsymbol{\alpha}$ of length b such that $\alpha_t := 1$ if $i_t i_{t+1}$ ($i_{b+1} = i_1$ here) is innovative while $\alpha_t := 0$ if otherwise, then the following arrays would be recorded:

- A non-negative integer array $\boldsymbol{c}$ of length w such that $q = c_1 + \cdots + c_w \le b$ (to be modified);
- A bit array $\boldsymbol{v} = v_1 \dots v_b$ of length b, in which there are exactly q ones and $b - q$ zeros. v_t will be forced to be 1 if $\alpha_t = 1$.
- Non-negative integer E, which is in fact an encoding of the cycle using some method we would show in the following.

We define E by

$$E = L + L_{max} M$$

in which $L_{max} := \binom{q}{c_1, \dots, c_w}$, and L and M are non-integer numbers which will be defined next.

To define L, an integer array $\boldsymbol{u}$ would be introduced which is of length q with t appearing exactly c_t times for $1 \le t \le w$. L is simply the encoding of $\boldsymbol{u}$, thus it could be bounded by

$$0 \le L < L_{max} := \binom{q}{c_1, \dots, c_w} \le C_1^b \frac{q^q}{c_1^{c_1} \dots c_w^{c_w}}$$

for some absolute constant C_1 from the Stirling's formula and the condition $q \le b$.

Now we would describe how to calculate the number M from the known cycle and all other information stored. First decode $\boldsymbol{u}$ from L and $c_1, \dots, c_w$. Let $m_1 = i_1 - 1$, $m_{1,max} = w$. Consider $1 \le t \le b - 1$ and suppose $i_1 \to \cdots \to i_t$ is known. If $v_t = 1$ is the l-th one in $\boldsymbol{v}$ ($l \le q$), u_l will be used to indicated the multi-degree of vertex i_{t+1}, thus a integer m_{t+1} such that $0 \le m_{t+1} < T_{u_l}/u_l$ can uniquely determine i_{t+1}. In this case let $m_{t+1,max} = T_{u_l}/u_l$. If $v_t = 0$, from the definition of $\boldsymbol{v}$ the leg $i_t i_{t+1}$ must be a non-innovative leg. Thus an integer m_{t+1} such that $0 \le m_{t+1} < d_{i_t} \le D_{i_t}$ can uniquely determine i_{t+1}. In this case let $m_{t+1,max} = D_{i_t}$. In the end, encode M as

$$M = m_1 + m_2 m_{1,max} + \cdots + m_b m_{b-1,max} \dots m_{1,max}.$$

From the construction of M, clearly we can reconstruct the original cycle from M together with all other stored information $\boldsymbol{D}, \boldsymbol{v}, \boldsymbol{c}, L$. On the contrary, given one cycle, for each valid $\boldsymbol{v}$ we can find such a (unique) $M(\boldsymbol{v})$ and $E(\boldsymbol{v})$ depending on $\boldsymbol{v}$ since $\boldsymbol{u}$ can be completely determined by $\boldsymbol{v}$ and the cycle.

Here we would construct a special $\boldsymbol{v}$ and show that $E(\boldsymbol{v})$ is small enough to carry our proof forward. Let $\boldsymbol{v}$ be all blanks except on the position v_t such that $\alpha_t = 1$ (v_t will be forced to set as 1 in such case). As for all other non-innovative legs $i_t i_{t+1}$, we set $v_t := 1$ if $b/D_{i_{t+1}} < D_{i_t}$ while $v_t := 0$ if otherwise. Clearly

$$M < M_{max} := m_{1,max} \dots m_{b,max}.$$

For each degree value p, the term T_p/p will appear c_p times in the sequence $m_{t,max}$ from the definition, and further more we denote by f_p the number of times that the degree p appears in the sequence $m_{t,max}$. By such a construction, we have

$$\begin{aligned}
& E(\boldsymbol{v}) < L_{max} M_{max} \\
& \leq C_1^b \frac{q^q}{c_1^{c_1} \dots c_w^{c_w}} \times w \prod_{\substack{i_t i_{t+1} \\ \text{innovative}}} \frac{T_{D_{i_{t+1}}}}{D_{i_{t+1}}} \prod_{\substack{i_t i_{t+1} \\ \text{non-innovative}}} \min\{D_{i_t}, b/D_{i_{t+1}}\} \\
& = C_1^b \frac{q^q}{c_1^{c_1} \dots c_w^{c_w}} \times w \left(\prod_{p=1}^{w} (T_p/p)^{c_p} \right) \cdot \left(\prod_{p=1}^{w} p^{f_p} \right) \\
& = C_2^b \left(\prod_{p=1}^{w} (q/p)^{c_p} \right) \cdot \left(\prod_{p=1}^{w} (T_p/c_p)^{c_p} \right) \cdot \left(\prod_{p=1}^{w} p^{f_p} \right) \\
& \leq C_3^b \left(\prod_{p=1}^{w} (b/p)^{c_p} \right) \cdot \left(\prod_{p=1}^{w} p^{f_p} \right) \qquad (32) \\
& = C_3^b \prod_{\substack{i_t i_{t+1} \\ \text{innovative}}} \frac{b}{D_{i_{t+1}}} \prod_{\substack{i_t i_{t+1} \\ \text{non-innovative}}} \min\{D_{i_t}, b/D_{i_{t+1}}\} \\
& = C_3^b \prod_{t=1}^{b} \sqrt{\frac{b D_{i_t}}{D_{i_{t+1}}}} \prod_{\substack{i_t i_{t+1} \\ \text{innovative}}} \sqrt{\frac{b}{D_{i_t} D_{i_{t+1}}}} \prod_{\substack{i_t i_{t+1} \\ \text{non-innovative}}} \min\left\{ \sqrt{\frac{b}{D_{i_t} D_{i_{t+1}}}}, \sqrt{\frac{D_{i_t} D_{i_{t+1}}}{b}} \right\} \\
& \leq C_3^b b^{b/2} \prod_{\substack{i_t i_{t+1} \\ \text{innovative}}} \sqrt{\frac{b}{D_{i_t} D_{i_{t+1}}}} \prod_{\substack{i_t i_{t+1} \\ \text{non-innovative}}} \min\left\{ \sqrt{\frac{b}{D_{i_t} D_{i_{t+1}}}}, \sqrt{\frac{D_{i_t} D_{i_{t+1}}}{b}} \right\} \qquad (33) \\
& \leq C_3^b b^{b/2}
\end{aligned}$$

for some slightly different absolute constants C_2, C_3, in which (32) comes from the fact that $(T_p/c_p)^{c_p} \leq e^{T_p/e}$ and $T_1 + \dots T_w = q \leq b$. To explain the last step in the transformation above, consider each edge $\{r, s\}$ in the cycle. It have been given that $\{r, s\}$ will appear at least 2 times in the cycle, supposing it appears exact $k \geq 2$ times, then it will contribute a factor

$$\sqrt{\frac{b}{D_r D_s}} \min\left\{ \sqrt{\frac{b}{D_r D_s}}, \sqrt{\frac{D_r D_s}{b}} \right\}^{k-1} \leq 1$$

to the formula (33).

In this way, we can always encode the cycle into a tuple $(\boldsymbol{D}, \boldsymbol{v}, \boldsymbol{c}, E)$ for some $\boldsymbol{v}$ such that $E < C_3^b b^{b/2}$ for some absolute constant C_3. The total number of such tuples can be clearly bounded by $C^b b^{b/2}$ for some slightly different absolute constant C.

In the end, by Stirling's formula we can write the ratio between $b^{b/2}$ and $b^{b/2+1-w} w!$ as

$$\frac{b^{b/2}}{b^{b/2+1-w} w!} \leq C_4^b \left(\frac{b}{w} \right)^w$$

for some absolute constant C_4. However simple calculus gives that $(b/w)^w \leq e^{w/e} \leq e^{b/e}$. Hence the number of valid cycles can be also bounded by $C' b^{b/2+1-w} w!$ for some slightly different constant C'.

□

Proposition 5. *Suppose $S(n)$ is the set consisting of all permutations on $\{1, \dots, n\}$. Let $(i_1, i_2), \dots, (i_b, i_1)$ be a cycle of length b in the graph. Let $e_1, \dots, e_j$ are the j distinct undirected edges appearing in the cycle, with occurrence times $\alpha_1, \dots, \alpha_j$, and $f_1 \dots, f_w$ are the w distinct vertices appearing in the cycle, with occurrence times $\beta_1, \dots, \beta_w$. Suppose all α are at least 2, then we have*

$$\mathop{\mathbb{E}}_{\sigma \in U_{S(n)}} [N_{\sigma(i_1)\sigma(i_2)} \dots N_{\sigma(i_b)\sigma(i_1)}] \leq \frac{K^{b-2j}}{d_{min}^{b-w+1} (n-1) \dots (n-w+1)}$$

Proof. Define G as a graph such that (s,t) is an edge of G if and only if (f_s, f_t) appears in the cycle $(i_1, \dots, i_b, i_1)$. We have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\sigma \in U_{S(n)}} [N_{\sigma(i_1)\sigma(i_2)} \dots N_{\sigma(i_b)\sigma(i_1)}] \\
\le &\mathop{\mathbb{E}}_{\sigma \in U_{S(n)}} [|N_{\sigma(i_1)\sigma(i_2)} \dots N_{\sigma(i_b)\sigma(i_1)}|] \\
\le &\mathop{\mathbb{E}}_{\sigma \in U_{S(n)}, X} \left[\frac{1}{d^{\beta_1}_{\sigma(f_1)} \dots d^{\beta_w}_{\sigma(f_w)}} |X_{\sigma(e_1)}|^{\alpha_1} \dots |X_{\sigma(e_j)}|^{\alpha_j} \right] \\
\le &K^{\alpha_1 + \dots + \alpha_j - 2j} \mathop{\mathbb{E}}_{\sigma \in U_{S(n)}} \left[\frac{1}{d^{\beta_1}_{\sigma(f_1)} \dots d^{\beta_w}_{\sigma(f_w)}} \mathop{\mathbb{E}}_X [X^2_{\sigma(e_1)} \dots X^2_{\sigma(e_j)}] \right] \\
\le &K^{\alpha_1 + \dots + \alpha_j - 2j} \mathop{\mathbb{E}}_{\sigma \in U_{S(n)}} \left[\frac{1}{d^{\beta_1}_{\sigma(f_1)} \dots d^{\beta_w}_{\sigma(f_w)}} \delta(\sigma; 1, \dots, w) \right],
\end{aligned}$$

in which $\delta(\sigma; p_1, \dots, p_l) := 1$ if corresponding $X_{\sigma(f_{p_s}),\sigma(f_{p_t})}$ are non-zero for all $(p_s, p_t) \in G$, and $\delta(\sigma; p_1, \dots, p_l) := 0$ otherwise. It is easy to see that if $\delta(\sigma; P) = 1$ for some σ then all subset $P' \subseteq P$ also satisfy $\delta(\sigma; P') = 1$. It suffices to bound $\mathbb{E}_{\sigma \in U_{S(n)}} \left[d^{-\beta_1}_{\sigma(f_1)} \dots d^{-\beta_w}_{\sigma(f_w)} \delta(\sigma; 1, \dots, w) \right]$. To use induction to achieve this end, we relax our assumption such that now we have not a cycle but just a connected graph consisting of w vertices $f_1, \dots, f_w$ with degrees $\beta_1, \dots, \beta_w$. We can find a vertex such that the graph is still connected after removing this vertex. Without loss of generality, suppose it is f_w, and one of its adjacent vertices is f_{w-1}. Fixing $\sigma(f_{w-1})$, there are at most $d_{\sigma(f_{w-1})}$ choices for $\sigma(f_w)$ to make $\delta(\sigma; 1, \dots, w)$ not vanish. Thus we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\sigma} \left[\frac{1}{d^{\beta_1}_{\sigma(f_1)} \dots d^{\beta_w}_{\sigma(f_w)}} \delta(\sigma; 1, \dots, w) \right] \\
\le &\mathop{\mathbb{E}}_{\sigma(f_1), \dots, \sigma(f_{w-1})} \left[d^{-\beta_1}_{\sigma(f_1)} \dots d^{-\beta_{w-1}}_{\sigma(f_{w-1})} \delta(\sigma; 1, \dots, w-1) \mathop{\mathbb{E}}_{\sigma(f_w)} \left[d^{-\beta_r}_{\sigma(f_w)} \delta(\sigma; w-1, w) \right] \right] \\
\le &d^{-\beta_w}_{min} \mathop{\mathbb{E}}_{\sigma(f_1), \dots, \sigma(f_{w-1})} \left[\frac{d_{\sigma(f_{w-1})}}{n - w + 1} d^{-\beta_1}_{\sigma(f_1)} \dots d^{-\beta_{w-1}}_{\sigma(f_{w-1})} \delta(\sigma; 1, \dots, w-1) \right] \\
\le &\frac{1}{d^{\beta_r}_{min}(n - w + 1)} \mathop{\mathbb{E}}_{\sigma} \left[d^{-\beta_1}_{\sigma(f_1)} \dots d^{-\beta_{w-2}}_{\sigma(f_{w-2})} d^{-\beta_{w-1}+1}_{\sigma(f_{w-1})} \delta(\sigma; 1, \dots, w-1) \right].
\end{aligned}$$

Observe that the last line has the form of recursion in the sense that we now have a connected graph consisting of $f_1, \dots, f_{w-1}$ with degrees $\beta_1, \dots, \beta_{w-2}, \beta_{w-1} + 1$, hence we obtain that

$$\begin{aligned}
\mathop{\mathbb{E}}_{\sigma} \left[\frac{1}{d^{\beta_1}_{\sigma(f_1)} \dots d^{\beta_w}_{\sigma(f_w)}} \delta(\sigma; 1, \dots, w) \right] &\le \frac{1}{{d_{min}}^{\beta_1 + \dots + \beta_w - w + 1} (n-1) \dots (n - w + 1)} \\
&= \frac{1}{d^{b-w+1}_{min} (n-1) \dots (n - w + 1)}
\end{aligned}$$

□

Proposition 6.

$$\mathbb{E}[\text{tr}(N^b)] \le C^b \frac{n}{d^{b/2}_{min}}$$

for some absolute constant C.

Proof. We have

$$\mathbb{E}[\text{tr}(N^b)] = \sum_{1 \le i_1, \dots, i_b \le n} \mathbb{E}[N_{i_1 i_2} \dots N_{i_b i_1}], \tag{34}$$

which is a sum over all cycles $i_1 \to \cdots \to i_b \to i_1$ of length b. According to the independence of N_{ij}, all terms are zero but those terms in which every $\{i_s, i_{s+1}\}$ appears at least two times. Consider cycles with w distinct vertices and j distinct edges appearing. Let $\mathcal{V}(b,w,j)$ be the set of all such cycles. Using Proposition 4, $|\mathcal{V}(b,w,j)|$ will be bounded by

$$C_1^b b^{b/2+1-w} w! \times \binom{n}{w} = C_1^b b^{b/2+1-w} n(n-1)\dots(n-w+1)$$

where C_1 is an absolute constant. While Proposition 5 gives the expected contribution to sum (34) over cycles of the certain shape, which means

$$\begin{aligned}
&\sum_{(i_1,\dots,i_b)\in\mathcal{V}(b,w,j)} \mathbb{E}[N_{i_1 i_2},\dots N_{i_b i_1}] \\
=& \mathop{\mathbb{E}}_{\sigma\in U_{S(n)}}\left[\sum_{(i_1,\dots,i_b)\in\mathcal{V}(b,w,j)} \mathbb{E}[N_{\sigma(i_1)\sigma(i_2)}\dots N_{\sigma(i_b)\sigma(i_1)}]\right] \\
=& \sum_{(i_1,\dots,i_b)\in\mathcal{V}(b,w,j)} \mathop{\mathbb{E}}_{\sigma\in U_{S(n)}}[N_{\sigma(i_1)\sigma(i_2)}\dots N_{\sigma(i_b)\sigma(i_1)}] \\
\le& C_1^b b^{b/2+1-w} n(n-1)\dots(n-w+1) \times \frac{K^{b-2j}}{d_{min}^{b-w+1}(n-1)\dots(n-w+1)} \\
\le& C_1^b n d_{min}^{-b/2} \times \left(\frac{K^2 b}{d_{min}}\right)^{b/2-w+1} \\
\le& C^b \frac{n}{d_{min}^{b/2}}
\end{aligned}$$

for some absolute constant C in which we used the fact that $j \ge w-1$ and $b = O(\log n), K = O\left(\sqrt{\frac{d_{min}}{\log n}}\right)$.

Theorem 3.1. *Suppose $K = O\left(\sqrt{\frac{d_{min}}{\log n}}\right)$, then*

$$\|N\| \le \frac{C}{\sqrt{d_{min}}}$$

holds with high probability for some absolute constant C.

Proof. Setting $b = 4\lceil \log n \rceil$ which is an even number, we write the constant in Proposition (6) as C_1 here, then the Proposition 6 together with the Markov's inequality gives

$$\begin{aligned}
\Pr\left[\|N\| > \frac{3C_1}{\sqrt{d_{min}}}\right] &= \Pr\left[\|N\|^b > \frac{3^b C_1^b}{d_{min}^{b/2}}\right] \\
&\le \frac{n}{3^b} < \frac{1}{n^4}.
\end{aligned}$$

Hence $\|N\| \le \frac{C}{\sqrt{d_{min}}}$ with high probability in which $C = 3C_1$ is an absolute constant. □

3.3 Proof of Lemma 3.2

Proof. First we have

$$\|B\|_{\boldsymbol{\omega}} = \|abs(B)\cdot\boldsymbol{\omega}\|_\infty \le \|\boldsymbol{\omega}\|_\infty \|B\|_\infty \le \sqrt{\frac{d_{\max}}{d_{\min}}}\|B\|_\infty$$

Hence it suffices to prove (31) by showing that

$$\|(\boldsymbol{e}_1^T \otimes I_m)\prod_{j=1}^{l}\big((B^{t_j}\otimes I_m)N\big)S_k\|_\infty \le \left(\sqrt{\frac{\log n}{d_{\min}}}\cdot CK\right)^l \cdot \|B\|_{\boldsymbol{\omega}} \cdot \|B\|_\infty^{t-l/2-1} \cdot \|B\|_{\max}^{l/2} \cdot \sqrt{\frac{d_{\min}}{|\mathcal{E}|}}\|U_k\| \quad (35)$$

holds with probability exceeding $1 - \frac{1}{n^c}$. Denote by $a_{pq}^{(j)}$ the element of matrix B^{t_j} in p-th row and q-th column. In this way, we can expand the left side of (35) as

$$\left\| \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}} a_{1\alpha_1}^{(1)} N_{\alpha_1\beta_1} a_{\beta_1\alpha_2}^{(2)} \dots a_{\beta_{l-1}\alpha_l}^{(l)} N_{\alpha_l\beta_l} (s_{\beta_l} U_k) \right\| \tag{36}$$

in which we used the fact that $S_k = \boldsymbol{s} \otimes U_k$. Here the summation is taken over all possible integer arrays $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ which are both of length l. Recall that $s_{\beta_l} = \sqrt{\frac{d_{\beta_l}}{|\mathcal{E}|}}$ and $N_{p,q} = \frac{1}{\sqrt{d_p d_q}} X_{p,q}$ if $(p,q) \in \mathcal{E}$ while $N_{p,q} = 0$ otherwise. We thereby denote by $\mathcal{H}$ as the set consisting of all pairs $(\boldsymbol{\alpha},\boldsymbol{\beta})$ such that $(\alpha_j, \beta_j) \in \mathcal{E}$ for all j. Hence it suffices to prove that

$$\begin{aligned} &\frac{1}{\sqrt{|\mathcal{E}|}} \left\| \sum_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{H}} \frac{a_{1,\alpha_1}^{(1)} a_{\beta_1,\alpha_2}^{(2)} \dots a_{\beta_{l-1},\alpha_l}^{(l)}}{\sqrt{d_{\alpha_1} d_{\beta_1} \dots d_{\alpha_{l-1}\beta_{l-1}} d_{\alpha_l}}} \cdot X_{\alpha_1\beta_1} \dots X_{\alpha_l\beta_l} U_k \right\| \\ \le & \Big(\sqrt{\frac{\log n}{d_{\min}}} \cdot CK \Big)^l \cdot \|B\|_{\boldsymbol{\omega}} \cdot \|B\|_\infty^{t-l/2-1} \cdot \|B\|_{\max}^{l/2} \cdot \sqrt{\frac{d_{\min}}{|\mathcal{E}|}} \|U_k\|, \end{aligned} \tag{37}$$

or in a stronger sense,

$$\begin{aligned} &\left\| \sum_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{H}} \frac{a_{1,\alpha_1}^{(1)} a_{\beta_1,\alpha_2}^{(2)} \dots a_{\beta_{l-1},\alpha_l}^{(l)}}{\sqrt{d_{\alpha_1} d_{\beta_1} \dots d_{\alpha_{l-1}} d_{\beta_{l-1}} d_{\alpha_l}}} \cdot X_{\alpha_1\beta_1} \dots X_{\alpha_l\beta_l} \right\| \\ \le & C^l (\log n)^{l/2} \cdot \|B\|_{\boldsymbol{\omega}} \cdot \|B\|_\infty^{t-l/2-1} \cdot \|B\|_{\max}^{l/2}. \end{aligned} \tag{38}$$

since $\|X_{pq}\| \le K$ for all p, q. To this end, we employ the power moment method, which needs us to show

$$\begin{aligned} &\mathbb{E}\left[\left\| \left(\sum_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathcal{H}} \prod_{r=1}^{2k} \frac{a_{1,\alpha_1}^{(1)} a_{\beta_1,\alpha_2}^{(2)} \dots a_{\beta_{l-1},\alpha_l}^{(l)}}{\sqrt{d_{\alpha_1} d_{\beta_1} \dots d_{\alpha_{l-1}} d_{\beta_{l-1}} d_{\alpha_l}}} \cdot X_{\alpha_1^{(r)}\beta_1^{(r)}} \dots X_{\alpha_l^{(r)}\beta_l^{(r)}} \right)^{2k} \right\| \right] \\ \le & (\log n)^{kl} \cdot \|B\|_{\boldsymbol{\omega}}^{2k} \cdot \|B\|_\infty^{2kt-kl-2k} \cdot \|B\|_{\max}^{kl}. \end{aligned} \tag{39}$$

holds for all integer k such that $kl \le O(\log n)$. The left side of (39) can be furthermore relaxed and expanded to

$$\sum_{(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})\in\mathcal{H}^{2k}} \prod_{r=1}^{2k} \frac{|a_{1,\alpha_1^{(r)}}^{(1)} a_{\beta_1^{(r)},\alpha_2^{(r)}}^{(2)} \dots a_{\beta_{l-1}^{(r)},\alpha_l^{(r)}}^{(l)}|}{\sqrt{d_{\alpha_1^{(r)}} d_{\beta_1^{(r)}} \dots d_{\alpha_{l-1}^{(r)}} d_{\beta_{l-1}^{(r)}} d_{\alpha_l^{(r)}}}} \cdot \left| \mathbb{E}\left[\prod_{r=1}^{2k} X_{\alpha_1^{(r)}\beta_1^{(r)}} \dots X_{\alpha_l^{(r)}\beta_l^{(r)}} \right] \right| \tag{40}$$

where $\hat{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}^{(1)}, \dots, \boldsymbol{\alpha}^{(2k)})$, $\hat{\boldsymbol{\beta}} = (\boldsymbol{\beta}^{(1)}, \dots, \boldsymbol{\beta}^{(2k)})$. Here $\mathcal{H}^{2k}$ is defined such that $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) \in \mathcal{H}^{2k}$ if and only if $(\boldsymbol{\alpha}^{(r)}, \boldsymbol{\beta}^{(r)}) \in \mathcal{H}$. An important observation is that $\mathbb{E}[X_{\alpha_1\beta_1} \dots X_{\alpha_l\beta_l}]$ is non-zero if and only if all unordered pairs $\{\alpha_j, \beta_j\}$ will appear at least two times among these $2k$ pairs since all X_{pq} are independent for distinct $\{p,q\}$ and $\mathbb{E}[X_{pq}] = 0$. Let $\mathcal{J}_{2k} \subseteq \mathcal{H}^{2k}$ be the set consisting of all $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ such that every $(\alpha_j^{(r)}, \beta_j^{(r)})$ would appear at least two times among these $2kl$ unordered pairs, which means it suffices to prove that

$$\sum_{(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})\in\mathcal{J}_{2k}} \prod_{r=1}^{2k} \frac{|a_{1,\alpha_1^{(r)}}^{(1)} a_{\beta_1^{(r)},\alpha_2^{(r)}}^{(2)} \dots a_{\beta_{l-1}^{(r)},\alpha_l^{(r)}}^{(l)}|}{\sqrt{d_{\alpha_1^{(r)}} d_{\beta_1^{(r)}} \dots d_{\alpha_{l-1}^{(r)}} d_{\beta_{l-1}^{(r)}} d_{\alpha_l^{(r)}}}} \le (\log n)^{kl} \cdot \|B\|_{\boldsymbol{\omega}} \cdot \|B\|_\infty^{2kt-kl-2k} \cdot \|B\|_{\max}^{2kl}. \tag{41}$$

To estimate the sum in the left side of (41), we introduce the concept of summing graph.

Definition 1. *A summing graph $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_1, \mathcal{E}_2)$ is formally defined as following:*

1. *$\mathcal{V}_s$ contains $1 + n$ nodes, in which one node is special and called the root node while the remaining n nodes are used to represent n variables that we are summing over. In fact, the root node represents the fixed index 1 in $a_{1,\alpha_1^{(r)}}^{(1)}$ in (41).*

2. *$\mathcal{E}_1$ is a undirected edge set on vertex set $\mathcal{V}_s$. If $(u, v) \in \mathcal{E}_1$, then the variables represented by u, v are adjacent in $\mathcal{G}$.*

3. *$\mathcal{E}_2$ is an undirected labeled edge set on vertex set $\mathcal{V}_s$ and $\mathcal{E}_2$ can contain multiple edges. Each edge in $\mathcal{E}_2$ has the form (p, q, n) where p, q is the end node while n is an integer label.*

A summing graph $\mathcal{G}_s$ combined with a vector θ can induce a sum $\Sigma(\mathcal{G}_s, \boldsymbol{\theta})$ as defined below:

$$\Sigma(\mathcal{G}_s, \boldsymbol{\theta}) = \sum_{\boldsymbol{v} \in \mathcal{U}(\mathcal{G}_s)} \prod_{(p,q,n) \in \mathcal{E}_2} [B^n]_{v_p, v_q} / F(\boldsymbol{v}, \boldsymbol{\theta})$$

in which we define

$$\mathcal{U}(\mathcal{G}_s) = \{\boldsymbol{v} \in \{1, \ldots, n\}^{|\mathcal{V}_s|} : \text{for every } (p, q) \in \mathcal{E}_1,\ v_p, v_q \text{ are adjacent in } \mathcal{G}\},$$

$$F(\mathcal{G}_s, \boldsymbol{v}, \boldsymbol{\theta}) = \prod_{p \in \mathcal{V}_s} \frac{1}{d_{v_i}^{\theta_p/2}} \prod_{p \notin \mathcal{V}_s} \frac{1}{d_{\min}^{\theta_p/2}}.$$

Return to the original problem. For some $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \in \mathcal{J}_{2k}$, we can divide all $2kl$ terms of form $\{\alpha_j^{(r)}, \beta_j^{(r)}\}$ into a number of groups such that each group includes the same unordered pairs. By the definition of $\mathcal{J}_{2k}$ each group has a size of at least 2. Without loss of generality we could assume all groups are of size 2 or 3. For example, if some $\{p, q\}$ occurs 7 times among $2kl$ unordered pairs, we can divide these 7 terms into 3 groups of size 2, 2, 3 respectively. Let the number of groups with size 2 be $kl - 3u$, then the number of groups with size 3 would be $2u$ since there are $2kl$ terms in total. To obtain identical relations between ordered pairs $(\alpha_j^{(r)}, \beta_j^{(r)})$, we can use a bit string of length $2kl$ to indicate whether the corresponding $(\alpha_j^{(r)}, \beta_j^{(r)})$ has the same order as its first appearance. Fixing u, the number of possible groupings would be

$$\frac{(2kl)!}{2^{kl-3u} 6^{2u} (kl - 3u)!(2u)!} = \frac{(2kl)!}{2^{kl} 4.5^u (kl - 3u)!(2u)!}.$$

For some grouping configuration with $kl - u$ groups together with a $2kl$ bit string like mentioned above, a summing group $\mathcal{G}_1$ could be constructed as below:

- $\mathcal{V}(\mathcal{G}_1)$ has $2(kl - u) + 1$ nodes. In particular, one of them represents 1, which is the root node, while each of others represents exactly one element in $kl - u$ groups. More precisely, each node can be regarded as a set of variables of form $\alpha_i^{(r)}$ or $\beta_i^{(r)}$ so that all variables in the same node are forced to have the same value.

- The edge joining nodes representing $\alpha_i^{(r)}, \beta_i^{(r)}$ would be in $\mathcal{E}_1(\mathcal{G}_1)$ for all possible i, r. Thus each node except the root node is associated with exact one edge in $\mathcal{E}_1(\mathcal{G})$.

- Starting from an empty $\mathcal{E}_2(\mathcal{G}_1)$, add a labeled edge $(\overline{\beta_i^{(r)}}, \overline{\alpha_{i+1}^{(r)}}, i+1)$ for all possible $\beta_i^{(r)}, \alpha_{i+1}^{(r)}$, and add a labeled edge $(\boldsymbol{root}, \alpha_1^{(r)}, 1)$ for all possible $\alpha_1^{(r)}$. Recall that $\mathcal{E}_2(\mathcal{G}_1)$ is a unordered edge set containing multiple edges.

- $\boldsymbol{\theta}_1$ is defined as an integer vector of the same length as the number of nodes in $\mathcal{G}_1$. Let $\theta_{1,0}$ be 0 for root node 0, and $\theta_{1,p}$ be the number of times that variable represented by node p appears in $\alpha_i^{(r)}$ and $\beta_i^{(r)}$ except $\beta_l^{(r)}$. It is clear that $\theta_{1,p} \geq 2$ for node p where p is not the root node and does not contain variable of form $\beta_l^{(r)}$. Hence there are at most $2k$ indices p such that $\theta_p = 1$.

In this way, we can easily verify that

$$\sum_{(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \in \mathcal{J}_{2k}} \prod_{r=1}^{2k} \frac{\left|a^{(1)}_{1, \alpha_1^{(r)}} a^{(2)}_{\beta_1^{(r)}, \alpha_2^{(r)}} \cdots a^{(l)}_{\beta_{l-1}^{(r)}, \alpha_l^{(r)}}\right|}{\sqrt{d_{\alpha_1^{(r)}} d_{\beta_1^{(r)}} \cdots d_{\alpha_{l-1}^{(r)}} d_{\beta_{l-1}^{(r)}} d_{\alpha_l^{(r)}}}} \leq \sum_{\mathcal{G}_1} \Sigma(\mathcal{G}_1, \boldsymbol{\theta}_1)$$

in which $\mathcal{G}_1$ is taken over all possible configurations of grouping.

Next, we will provide an estimation on $\Sigma(\mathcal{G}_1, \boldsymbol{\theta}_1)$ by induction where $\mathcal{G}_1$ contains $2(kl - u) + 1$ nodes.

Claim:

$$\Sigma(\mathcal{G}_1, \boldsymbol{\theta}_1) \leq \|B\|_{\boldsymbol{\omega}}^{2k} \cdot \|B\|_{\infty}^{2kt - kl - 2k} \cdot \|B\|_{\max}^{kl} \cdot d_{\min}^{-(l-1)k - u}.$$

Proposition 7. *We can remove some edges from $\mathcal{E}_2(\mathcal{G}_1)$ to obtain a graph $\mathcal{G}_2$ satisfying*

- *$\mathcal{G}_2$ is a tree with edge set $\mathcal{E}_1(\mathcal{G}_2) \cup \mathcal{E}_2(\mathcal{G}_2)$.*
- *$\mathcal{E}_2(\mathcal{G}_2)$ does not contain multiple edges any more.*
- *Except the root node, each node in $\mathcal{G}_2$ is associated with exactly one edge in $\mathcal{E}_1$.*

Define the difference between two graphs $\mathcal{G}$ and $\mathcal{G}'$ as

$$\Delta(\mathcal{G}', \mathcal{G}) := \sum_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}')\setminus\mathcal{E}_2(\mathcal{G})} n$$

for graph $\mathcal{G} \subseteq \mathcal{G}'$, then we have

$$\Sigma(\mathcal{G}_1, \boldsymbol{\theta}_1) \leq \|B\|_{\max}^{\Delta(\mathcal{G}_1,\mathcal{G}_2)} \Sigma(\mathcal{G}_2, \boldsymbol{\theta}_2).$$

Since there are $2kl$ edges in $\mathcal{E}_2(\mathcal{G}_1)$ but only $kl - u$ edges in $\mathcal{E}_2(\mathcal{G}_2)$, we have

$$\Delta(\mathcal{G}_1, \mathcal{G}_2) \geq kl + u.$$

In the following, we will construct a sequence of summing graphs $\mathcal{G}_2, \ldots, \mathcal{G}_m$ in a inductive way, where $\mathcal{G}_m$ is a summing graph with only one node, the root node. Specifically, we impose an inductive assumption which holds for $\mathcal{G}_1$ and $\mathcal{G}_2$: $\mathcal{G}_z$ is a tree over edge set $\mathcal{E}_1(\mathcal{G}_z) \cup \mathcal{E}_2(\mathcal{G}_z)$ for $z = 2, \ldots, m$.

Suppose $\mathcal{G}_z (z \geq 2)$ is given and has more than one nodes, then we can choose from tree $\mathcal{G}_z$ a leaf node π_z which is not the root node. By inductive assumption there exists some τ_z with $(\pi_z, \tau_z) \in \mathcal{E}_1(\mathcal{G}_z)$ and some ν_z with $(\tau_z, \nu_z) \in \mathcal{E}_2(\mathcal{G}_z)$, and $\mathcal{G}_{z+1}$ is still a tree where $\mathcal{G}_{z+1}$ is obtained by removing $\{\pi_z, \tau_z\}$ and their associated edges from $\mathcal{G}_z$. There are two possible bounds:

1. $\theta_{z,\tau_z} \geq 2$. In this case, let $\boldsymbol{\theta}_{z+1}$ be the same as $\boldsymbol{\theta}_z$ on all components except $\theta_{z+1,\tau_z} = \theta_{z,\tau_z} - 2$. Then we have

$$\begin{aligned}
\Sigma(\mathcal{G}_z, \boldsymbol{\theta}_z) &= \sum_{\boldsymbol{v}\in\mathcal{U}(\mathcal{G}_z)} \prod_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}_z)} \left|[B^n]_{v_p,v_q}\right| \Big/ F(\mathcal{G}_z, \boldsymbol{v}, \boldsymbol{\theta}_z) \\
&\leq \sum_{\boldsymbol{v}\in\mathcal{U}(\mathcal{G}_{z+1})} \sum_{v_{\tau_z}} \sum_{v_{\pi_z}\in\mathcal{N}(v_{\tau_z})} \prod_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}_{z+1})} \left|[B^n]_{v_p,v_q}\right| \left|[B^{n(\tau_z,\nu_z)}]_{v_{\tau_z},v_{\nu_z}}\right| \Big/ \Big(F(\mathcal{G}_{z+1}, \boldsymbol{v}, \boldsymbol{\theta}_{z+1}) d_{v_{\tau_z}}\Big) \\
&\leq \sum_{\boldsymbol{v}\in\mathcal{U}(\mathcal{G}_{z+1})} \frac{\prod_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}_{z+1})} \left|[B^n]_{v_p,v_q}\right|}{F(\mathcal{G}_{z+1}, \boldsymbol{v}, \boldsymbol{\theta}_{z+1})} \sum_{v_{\tau_z}} \sum_{v_{\pi_z}\in\mathcal{N}(v_{\tau_z})} \frac{1}{d_{\tau_z}} \left|[B^{n(\tau_z,\nu_z)}]_{v_{\pi_z},v_{\nu_z}}\right| \\
&= \sum_{\boldsymbol{v}\in\mathcal{U}(\mathcal{G}_{z+1})} \frac{\prod_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}_{z+1})} \left|[B^n]_{v_p,v_q}\right|}{F(\mathcal{G}_{z+1}, \boldsymbol{v}, \boldsymbol{\theta}_{z+1})} \sum_{v_{\tau_z}} \left|[B^{n(\tau_z,\nu_z)}]_{v_{\pi_z},v_{\nu_z}}\right| \\
&\leq \sum_{\boldsymbol{v}\in\mathcal{U}(\mathcal{G}_{z+1})} \frac{\prod_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}_{z+1})} \left|[B^n]_{v_p,v_q}\right|}{F(\mathcal{G}_{z+1}, \boldsymbol{v}, \boldsymbol{\theta}_{z+1})} \|B\|_\infty^{\Delta(\mathcal{G}_z,\mathcal{G}_{z+1})} \\
&= \Sigma(\mathcal{G}_{z+1}, \boldsymbol{v}, \boldsymbol{\theta}_{z+1}) \|B\|_\infty^{\Delta(\mathcal{G}_z,\mathcal{G}_{z+1})}
\end{aligned}$$

2. It is trivial that $\theta_{z,\pi_z} \geq 1$ and $\theta_{z,\tau_z} \geq 1$. So we can always let $\boldsymbol{\theta}_{z+1}$ be the same as $\boldsymbol{\theta}_z$ on all components

except $\theta_{z+1,\pi_z} = \theta_{z,\pi_z} - 1$ and $\theta_{z+1,\tau_z} = \theta_{z,\tau_z} - 1$. Then we have

$$
\begin{aligned}
\Sigma(\mathcal{G}_z, \boldsymbol{\theta}_z) &= \sum_{\boldsymbol{v}\in\mathcal{U}(\mathcal{G}_z)} \prod_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}_z)} \left|[B^n]_{v_p,v_q}\right| \Big/ F(\mathcal{G}_z, \boldsymbol{v}, \boldsymbol{\theta_z}) \\
&\leq \sum_{\boldsymbol{v}\in\mathcal{U}(\mathcal{G}_{z+1})} \sum_{v_{\tau_z}} \sum_{v_{\pi_z}\in\mathcal{N}(v_{\tau_z})} \prod_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}_{z+1})} \left|[B^n]_{v_p,v_q}\right| \left|[B^{n(\tau_z,\nu_z)}]_{v_{\tau_z},v_{\nu_z}}\right| \Big/ \left(F(\mathcal{G}_{z+1}, \boldsymbol{v}, \boldsymbol{\theta}_{z+1})\sqrt{d_{\tau_z} d_{\pi_z}}\right) \\
&\leq \sum_{\boldsymbol{v}\in\mathcal{U}(\mathcal{G}_{z+1})} \frac{\prod_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}_{z+1})} \left|[B^n]_{v_p,v_q}\right|}{F(\mathcal{G}_{z+1}, \boldsymbol{v}, \boldsymbol{\theta}_{z+1})} \sum_{v_{\tau_z}} \sum_{v_{\pi_z}\in\mathcal{N}(v_{\tau_z})} \frac{1}{\sqrt{d_{\tau_z} d_{\pi_z}}} \left|[B^{n(\tau_z,\nu_z)}]_{v_{\pi_z},v_{\tau_z}}\right| \\
&= \sum_{\boldsymbol{v}\in\mathcal{U}(\mathcal{G}_{z+1})} \frac{\prod_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}_{z+1})} \left|[B^n]_{v_p,v_q}\right|}{F(\mathcal{G}_{z+1}, \boldsymbol{v}, \boldsymbol{\theta}_{z+1})} \sum_{v_{\tau_z}} \omega_{\tau_z} \left|[B^{n(\tau_z,\nu_z)}]_{v_{\pi_z},v_{\nu_z}}\right| \\
&\leq \sum_{\boldsymbol{v}\in\mathcal{U}(\mathcal{G}_{z+1})} \frac{\prod_{(p,q,n)\in\mathcal{E}_2(\mathcal{G}_{z+1})} \left|[B^n]_{v_p,v_q}\right|}{F(\mathcal{G}_{z+1}, \boldsymbol{v}, \boldsymbol{\theta}_{z+1})} \|B\|_\infty^{\Delta(\mathcal{G}_z,\mathcal{G}_{z+1})-1} \|B\|_{\boldsymbol{\omega}} \\
&= \Sigma(\mathcal{G}_{z+1}, \boldsymbol{\theta}_{z+1}) \|B\|_\infty^{\Delta(\mathcal{G}_z,\mathcal{G}_{z+1})-1} \|B\|_{\boldsymbol{\omega}}
\end{aligned}
$$

Since at first $\theta_{2,p} \geq 1$ for all non-root nodes p and $\theta_{z+1,p} < \theta_{z,p}$ happens only if node p is removed from $\mathcal{G}_z$, the second bound always works. Besides, we have bounds

$$
\Sigma(\mathcal{G}_z, \boldsymbol{\theta}_z) \leq \Sigma(\mathcal{G}_{z+1}, \boldsymbol{\theta}_{z+1}) \|B\|_\infty^{\Delta(\mathcal{G}_z,\mathcal{G}_{z+1})-1} \min\{\|B\|_\infty, \|B\|_{\boldsymbol{\omega}}\} \tag{42}
$$

if $\theta_{z,\tau_z} \geq 2$.

$$
\Sigma(\mathcal{G}_m, \boldsymbol{\theta}_m) = d_{\min}{}^{-\sum_i \theta_{m,i}/2}.
$$

Now let's calculate $\sum_i \theta_{m,i}$. Initially we have

$$
\sum_i \theta_{2,i} = \sum_i \theta_{1,i} = 2k(2l-1).
$$

The inductive steps give $\sum_i \theta_{z+1,i} = \sum_i \theta_{z,i} - 2$ and there are $kl - u$ inductive steps in total. Hence

$$
\sum_i \theta_{m,i} = 2k(2l-1) - 2(kl-u) = 2kl - 2k + 2u,
$$

which means $\Sigma(\mathcal{G}_m, \boldsymbol{\theta}_m) = d_{\min}^{-kl+k+u}$.

Note that there are at most $2k$ nodes p with $\theta_{2,p} = 1$, which means the bound (42) holds for all inducative steps except $2k$ steps. In this way, together with the fact $\|B\|_{\max} \leq \|B\|_\infty$ and $\Delta(\mathcal{G}_1, \mathcal{G}_m) = t$, $\Sigma(\mathcal{G}_1, \boldsymbol{\theta}_1)$ could be expressed as

$$
\begin{aligned}
\Sigma(\mathcal{G}_1, \boldsymbol{\theta}_1) &\leq \left(\frac{\|B\|_{\boldsymbol{\omega}}}{\|B\|_\infty}\right)^{2k} \|B\|_{\max}^{\Delta(\mathcal{G}_1,\mathcal{G}_2)} \|B\|_\infty^{\Delta(\mathcal{G}_2,\mathcal{G}_m)} \Sigma(\mathcal{G}_m, \boldsymbol{\theta}_m) \\
&\leq \left(\frac{\|B\|_{\boldsymbol{\omega}}}{\|B\|_\infty}\right)^{2k} \|B\|_{\max}^{kl} \|B\|_\infty^{\Delta(\mathcal{G}_1,\mathcal{G}_m)-kl} d_{\min}^{-kl+k+u} \\
&\leq \|B\|_{\boldsymbol{\omega}}^{2k} \|B\|_{\max}^{kl} \|B\|_\infty^{t-kl-2k} d_{\min}^{-k(l-1)+u}
\end{aligned}
$$

For $\mathcal{G}_m$ which contains a single root node, we have

$$
\Sigma(\mathcal{G}_m, \boldsymbol{\theta}_m) = d_{\min}{}^{-\sum_i \theta_{m,i}/2} = d_{\min}{}^{-k(l-1)-u},
$$

where the last equation comes from the fact $\sum_i \theta_{1,i} = 2k(2l-1)$ and $\sum_i \theta_{z+1,i} = \sum_i \theta_{z,i} - 2$ if and only an edge in $\mathcal{E}_2(\mathcal{G}_z)$ was removed, but there are exact $kl - u$ edges in $\mathcal{E}_2(\mathcal{G}_1)$ so that

$$
\sum_i \theta_{m,i} = \sum_i \theta_{1,i} - 2(kl-u) = 2kl - 2k + 2u.
$$

Note that the third cases would appear at most $2k$ times and the second cases appear exactly kl times, which means

$$\Sigma(\mathcal{G}_1, \boldsymbol{\theta}_1) \le \|B\|_{\max}^{kl} \|B\|_{\boldsymbol{\omega}}^{2k} \|B\|_{\infty}^{\Delta(\mathcal{G}_1,\mathcal{G}_m)-kl-2k} \Sigma(\mathcal{G}_m, \boldsymbol{\theta}_m)$$

Return to calculate (41). We have

$$\begin{aligned}
&\sum_{(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})\in\mathcal{J}_{2k}} \prod_{r=1}^{2k} \frac{\left|a^{(1)}_{1,\alpha_1^{(r)}} a^{(2)}_{\beta_1^{(r)},\alpha_2^{(r)}} \cdots a^{(l)}_{\beta_{l-1}^{(r)},\alpha_l^{(r)}}\right|}{\sqrt{d_{\alpha_1^{(r)}} d_{\beta_1^{(r)}} \cdots d_{\alpha_{l-1}^{(r)}} d_{\beta_{l-1}^{(r)}} d_{\alpha_l^{(r)}}}} \\
\le &\sum_{\mathcal{G}_1} \Sigma(\mathcal{G}_1, \boldsymbol{\theta}_1) \\
= &\sum_{u\ge 0} \sum_{|\mathcal{V}(\mathcal{G}_1)|=kl-u+1} \Sigma(\mathcal{G}_1, \boldsymbol{\theta}_1) \\
\le &\|B\|_{\boldsymbol{\omega}}^{2k} \|B\|_{\max}^{kl} \|B\|_{\infty}^{t-kl-2k} d_{\min}^{-k(l-1)} \sum_{u\ge 0} \frac{(2kl)!}{2^{kl} 4.5^u (kl-3u)!(2u)!} d_{\min}^{-u}
\end{aligned}$$

For the last line above, we have

$$\frac{(2kl)!}{(kl-3u)!(2u)! d_{\min}^u} \le \frac{(2kl)!}{(kl-3u)!(2u)!(kl)^u} \le (Ckl)^{kl}$$

for some absolute constant C. Thus

$$\sum_{(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})\in\mathcal{J}_{2k}} \prod_{r=1}^{2k} \frac{\left|a^{(1)}_{1,\alpha_1^{(r)}} a^{(2)}_{\beta_1^{(r)},\alpha_2^{(r)}} \cdots a^{(l)}_{\beta_{l-1}^{(r)},\alpha_l^{(r)}}\right|}{\sqrt{d_{\alpha_1^{(r)}} d_{\beta_1^{(r)}} \cdots d_{\alpha_{l-1}^{(r)}} d_{\beta_{l-1}^{(r)}} d_{\alpha_l^{(r)}}}} \le (C\log n)^{kl} \|B\|_{\boldsymbol{\omega}}^{2k} \|B\|_{\max}^{kl} \|B\|_{\infty}^{t-kl-2k} d_{\min}^{-k(l-1)}$$

for some absolute constant C.

□

3.3.1 A Lower Bound on $\|B\|_\infty$

In the simple case, it is clear that $B = L^\dagger$, $BL = I - \boldsymbol{s}\boldsymbol{s}^T$, $B\boldsymbol{s} = 0$. In the following I will explore the relationship between $\|B\|_\infty$ and $d_{\max}/d_{\min}$. Without loss of generality, suppose $d_1 = d_{\max}$ and $d_2 = d_{\min}$, then consider the vector which is a linear combination of the first column of L denoted as $\boldsymbol{col}_1(L)$ and $\boldsymbol{s}$:

$$\boldsymbol{v} = \boldsymbol{col}_1(L) - \frac{1 - 1/\sqrt{d_1 d_2}}{\sqrt{d_1} + \sqrt{d_2}} \boldsymbol{s}.$$

In this way, it can be easily verified that

$$\|\boldsymbol{v}\|_\infty = \frac{\sqrt{d_2} + 1/\sqrt{d_2}}{\sqrt{d_1} + \sqrt{d_2}}$$

and

$$B\boldsymbol{v} = B\boldsymbol{col}_1(L) = \boldsymbol{col}_1(I - \boldsymbol{s}\boldsymbol{s}^T),$$

so

$$\|B\boldsymbol{v}\|_\infty \ge 1 - \frac{d_1}{|\mathcal{E}|} \ge \frac{1}{2}$$

since $|\mathcal{E}| \ge 2d_1$. Hence we have

$$\|B\|_\infty \ge \frac{\sqrt{d_1} + \sqrt{d_2}}{2(\sqrt{d_2} + 1/\sqrt{d_2})} \ge \frac{1}{4}\sqrt{\frac{d_1}{d_2}} \ge \frac{1}{4}\sqrt{\frac{d_{\max}}{d_{\min}}}.$$

3.4 Proof of Lemma 3.3

Suppose

$$K \leq C_1 \sqrt{\frac{d_{\min}}{\|B_i\| \cdot \|B_i\|_\infty \cdot \log n}},$$

the inequality (31) turns into

$$\|(\boldsymbol{e}_b^T \otimes I_m) \prod_{j=1}^{l} ((B_i^{t_j} \otimes I_m)N)S_k\|_\infty \leq C^l \|B_i\|_\infty^{t-l} \sqrt{\frac{d_{\min}}{|\mathcal{E}|}} \|U_k\|.$$

Thus we have

$$\|(\boldsymbol{e}_b^T \otimes I_m)(B_i N)^{i_1} B_i \ldots (B_i N)^{i_{j+1}} B_i N (S_k \otimes I_m)\| \leq \sqrt{\frac{d_{\min}}{|\mathcal{E}|}} \|U_k\| \cdot \|B_i\|_\infty^j (C_1 C)^{\sum_{l=1}^{j+1} i_l + 1}.$$

On the other hand, by choosing small enough absolute constant C_1, we have

$$\|B_i\|\|N\| \leq \|B_i\| \frac{C_2 K}{\sqrt{d_{\min}}} \leq \frac{1}{10},$$

$$\|B_i\|\|\lambda_i I_m - \Sigma_i\| \leq \|B_i\|\|N\| \leq \frac{1}{10}.$$

Thus applying Lemma 2.4 provides the desired conclusion.

4 Proof of Exact Recovery Conditions Under the Full Setting

4.1 Proof of Theorem 3.1

We prove Theorem 3.1 by establishing the concentration under the un-normalized data, i.e., which is identical to set $d_i = nt$. To begin with, let us rewrite the data matrix under the proposed model of mapping and observation graph. Since we assume the observation graph and the input pair-wise maps are generated through independent procedures, it follows that

$$\overline{X}_{ij} = \frac{1}{nt} \cdot \begin{cases} I_m - \frac{1}{m}\mathbf{1}\mathbf{1}^T & \text{with probability } \eta_{ij} t \\ U_{P_m} - \frac{1}{m}\mathbf{1}\mathbf{1}^T & \text{with probability } (1-\eta_{ij})t \\ 0 & \text{with probability } (1-t) \end{cases} \tag{43}$$

Here $\eta_{ij}, 1 \leq i, j \leq n$ form a matrix $(p-q)(I_k \otimes (\mathbf{1}\mathbf{1}^T)) + q\mathbf{1}\mathbf{1}^T$.

(43) gives rise to

$$E[\overline{X}_{ij}] = \frac{\eta_{ij}}{n}(I_m - \frac{1}{m}\mathbf{1}\mathbf{1}^T), \tag{44}$$

and

$$N_{ij} := \overline{X}_{ij} - E[\overline{X}_{ij}] = \frac{1}{nt} \cdot \begin{cases} (1-\eta_{ij}t)(I_m - \frac{1}{m}\mathbf{1}\mathbf{1}^T) & \text{with probability } \eta_{ij} t \\ U_{P_m} - \frac{1}{m}\mathbf{1}\mathbf{1}^T - \eta_{ij}t(I_m - \frac{1}{m}\mathbf{1}\mathbf{1}^T) & \text{with probability } (1-\eta_{ij})t \\ -\eta_{ij}t(I_m - \frac{1}{m}\mathbf{1}\mathbf{1}^T) & \text{with probability } (1-t) \end{cases} \tag{45}$$

It is obvious that $N_{ij}\mathbf{1} = 0$ and $N_{ij}^T\mathbf{1} = 0$. Moreover,

$$\|N_{ij}\| \leq \frac{1 + \eta_{ij}t}{nt} \leq \frac{2}{nt}.$$

Decompose $\overline{X} = E[\overline{X}] + N$, it follows that

$$E[\overline{X}] = \frac{1}{n}\big((p-q)(I_k \otimes (\mathbf{1}\mathbf{1}^T)) + q(\mathbf{1}\mathbf{1}^T)\big) \otimes (I_m - \frac{1}{m}\mathbf{1}\mathbf{1}^T). \tag{46}$$

Following the convention of notation, let

$$A = \frac{1}{n}\big((p-q)(I_k \otimes (\mathbf{1}\mathbf{1}^T)) + q(\mathbf{1}\mathbf{1}^T)\big).$$

It is easy to check that the rank of A is k, and its top k eigenvalues are given by

$$\lambda_1(A) = q + \frac{p-q}{k},\ \lambda_i(A) = \frac{p-q}{k}, 2 \leq i \leq k. \tag{47}$$

Let $(\frac{1}{\sqrt{k}}\mathbf{1}, H_k)$ be an orthonormal basis for $\mathbb{R}^k$, then it is easy to see that the corresponding top k eigenvectors of A are given by

$$S_k = \sqrt{\frac{k}{n}}\mathbf{1} \otimes \big(\frac{1}{\sqrt{k}}\mathbf{1}, H_k\big).$$

Moreover,

$$\overline{S}_k \overline{S}_k^T = (I_{n_0} - \frac{1}{n_0}\mathbf{1}\mathbf{1}^T) \otimes I_k.$$

To apply Lemma 3.3, it is easy to see that

$$B_1 = \frac{1}{q + \frac{p-q}{k}}(I_{n_0} - \frac{1}{n_0}\mathbf{1}\mathbf{1}^T) \otimes I_k,$$

and

$$B_i = \frac{k}{p-q}(I_{n_0} - \frac{1}{n_0}\mathbf{1}\mathbf{1}^T) \otimes I_k, \quad 2 \leq i \leq k.$$

Denote

$$T = (I_{n_0} - \frac{1}{n_0}\mathbf{1}\mathbf{1}^T) \otimes I_k.$$

It is easy to check that

$$\|T\|_{\boldsymbol{\omega}} = 2, \quad \|T\|_{\infty} = 2, \quad \|T\|_{\max} = 1.$$

Applying Lemma 3.3, we obtain the following stability bound on the top $k(m-1)$ eigen-vectors of $\overline{X}$:

Lemma 4.1. *Let $\overline{U} = (\overline{U}_1^T, \cdots, \overline{U}_n^T)^T$ be the top $k(m-1)$ eigen-vectors of $\overline{X}$. Then there exists a rotation matrix $\overline{R} \in O(k(m-1))$ and a universal constant c, so that when*

$$p - q \geq ck\sqrt{\frac{\log(n)}{nt}},$$

we have w.h.p,

$$\max_{1 \leq i \leq n} \|U_i - (\boldsymbol{e}_i^T S_k \otimes H_m) \cdot \overline{R}\| \leq \frac{1}{6} \cdot \frac{1}{\sqrt{n}}. \tag{48}$$

Complete the proof of Theorem 3.1. Since the spectral norm bounds the difference between the corresponding rows, it follows from Lemma 4.1 that (1) the distance between the corresponding elements between each pair of objects in the embedding space is upper bounded by 1/3, (2) $d_{\text{intra}} < 1/3$, and (3) $d_{\text{inter}} > 2/3$. This means both the intra-cluster maps and the underlying clusters can be recovered, which ends the proof.

4.2 Proof of Theorem 3.2

We prove a stronger recovery condition for inter-cluster maps. Note that inter-cluster map recovery solves the following linear assignment:

$$X_{st} = \underset{X \in \mathcal{P}_m}{\text{argmax}}\langle X, C_{st}\rangle, \quad C_{st} = \sum_{(i,j)\in\mathcal{E}, i \in c_s, j \in c_t} X_{j i_t} X_{ij}^{in} X_{i_s i}. \tag{49}$$

We prove a stronger exact recovery condition as follows. To begin with, we define the minimum number of inter-cluster edges between one pair of clusters as.

$$N_{\text{inter}} = \min_{1 \leq s < t \leq k} N_{st}, \qquad N_{st} := |\{(i,j)|(i,j) \in \mathcal{E}, i \in c_s, j \in c_t\}|. \tag{50}$$

Lemma 4.2. *Given an absolute constant* $c_{\text{inter}} > 0$*. Suppose the intra-cluster rate* q *and* N_{inter} *satisfy the following constraint:*

$$q \geq \sqrt{\frac{c_{\text{inter}} \log(n)}{N_{\text{inter}}}}.$$

Then we have with probability at least $1 - \frac{m^2k^2}{n^{\frac{c_{\text{inter}}}{8}}}$*,*

$$\min_{1 \leq a \leq m} C_{st}(a,a) > \max_{1 \leq a \neq b \leq m} C_{st}(a,b), \qquad 1 \leq s \neq t \leq k. \tag{51}$$

Proof: First of all, it is easy to check that

$$E[C_{st}(a,b)] = \begin{cases} \frac{1}{m}(1-q)+q & a=b \\ \frac{1}{m}(1-q) & a \neq b \end{cases}$$

We apply union bounds by showing that with probability at least $1 - \frac{m^2k^2}{n^{\frac{c_{\text{inter}}}{8}}}$, we have

$$\min_{1 \leq s \neq t \leq k, 1 \leq a \leq m} C_{st}(a,a) > \frac{1}{m}(1-q) + \frac{q}{2}, \tag{52}$$

$$\max_{1 \leq s \neq t \leq k, 1 \leq a \neq b \leq m} C_{st}(a,b) < \frac{1}{m}(1-q) + \frac{q}{2}. \tag{53}$$

Note that each diagonal element $X_{ji_t} X_{ij}^{in} X_{i_s i}(a,a)$ is a random Bernoulli random variable with probability $\frac{1-q}{m} + q$, we can apply lower Chernoff bound to obtain a lower tail bound on $C_{st}(a,a)$, which is

$$Pr[C_{st}(a,a) \leq N_{st}(\frac{1-q}{m} + q - \frac{q}{2})] \leq \exp(-\frac{N_{st}q^2}{8(q+\frac{1-q}{m})}) \leq \exp(-\frac{N_{\text{inter}}q^2}{8(q+\frac{1-q}{m})}). \tag{54}$$

Similarly, each off-diagonal element $X_{ji_t} X_{ij}^{in} X_{i_s i}(a,b)$ is a random Bernoulli random variable with probability $\frac{1-q}{m}$, we can apply upper Chernoff bound to obtain a upper tail bound on $C_{st}(a,b)$, which is

$$Pr[C_{st}(a,b) \leq N_{st}(\frac{1-q}{m} + \frac{q}{2})] \leq \exp(-\frac{N_{st}q^2}{8(\frac{q}{6}+\frac{1-q}{m})}) \leq \exp(-\frac{N_{\text{inter}}q^2}{8(\frac{q}{6}+\frac{1-q}{m})}). \tag{55}$$

Since $q = \sqrt{c_{\text{inter}} \frac{\log(n)}{N_{\text{inter}}}}$. It follows that combing (54) and (55) lead to

$$Pr[C_{st}(a,a) \leq N_{st}(\frac{1-q}{m} + q - \frac{q}{2})] \leq \exp(-\frac{c_{\text{inter}}\log(n)}{8}) \leq \frac{1}{n^{\frac{c_{\text{inter}}}{8}}} \tag{56}$$

$$Pr[C_{st}(a,b) \geq N_{st}(\frac{1-q}{m} + \frac{q}{2})] \leq \frac{1}{n^{\frac{c_{\text{inter}}}{8}}} \tag{57}$$

Applying union bounds (56) and (57), we have that the inter-cluster maps can be recovered with probability at least $1 - \frac{m^2k^2}{n^{\frac{c_{\text{inter}}}{8}}}$. □

Since the observation graph is generated from the Erdős-Rényi model $G(n,t)$. It is easy to check that the number of inter-cluster edges between a pair of clusters concentrates at $[\frac{n^2t}{2k^2}, \frac{2n^2t}{k^2}]$ with overwhelming probability (for example using Chernoff bound), which ends the proof. □

References

[Eldridge et al., 2018] Eldridge, J., Belkin, M., and Wang, Y. (2018). Unperturbed: spectral analysis beyond davis-kahan. volume abs/1706.06516. 1