Normalized Spectral Map Synchronization

Yanyao Shen
UT Austin
Austin, TX 78712
shenyanyao@utexas.edu

Qixing Huang
TTI Chicago and UT Austin
Austin, TX 78712
huangqx@cs.utexas.edu

Nathan Srebro
TTI Chicago
Chicago, IL 60637
nati@ttic.edu

Sujay Sanghavi
UT Austin
Austin, TX 78712
sanghavi@mail.utexas.edu

Abstract

Estimating maps among large collections of objects (e.g., dense correspondences across images and 3D shapes) is a fundamental problem across a wide range of domains. In this paper, we provide theoretical justifications of spectral techniques for the map synchronization problem, i.e., it takes as input a collection of objects and noisy maps estimated between pairs of objects along a connected object graph, and outputs clean maps between all pairs of objects. We show that a simple normalized spectral method (or NormSpecSync) that projects the blocks of the top eigenvectors of a data matrix to the map space, exhibits surprisingly good behavior — NormSpecSync is much more efficient than state-of-the-art convex optimization techniques, yet still admitting similar exact recovery conditions. We demonstrate the usefulness of NormSpecSync on both synthetic and real datasets.

1 Introduction

The problem of establishing maps (e.g., point correspondences or transformations) among a collection of objects is connected with a wide range of scientific problems, including fusing partially overlapped range scans [1], multi-view structure from motion [2], re-assembling fractured objects [3], analyzing and organizing geometric data collections [4] as well as DNA sequencing and modeling [5]. A fundamental problem in this domain is the so-called *map synchronization*, which takes as input noisy maps computed between pairs of objects, and utilizes the natural constraint that composite maps along cycles are identity maps to obtain improved maps.

Despite the importance of map synchronization, the algorithmic advancements on this problem remain limited. Earlier works formulate map synchronization as solving combinatorial optimizations [1, 6, 7, 8]. These formulations are restricted to small-scale problems and are susceptible to local minimums. Recent works establish the connection between the cycle-consistency constraint and the low-rank property of the matrix that stores pairwise maps in blocks; they cast map synchronization as low-rank matrix inference [9, 10, 11]. These techniques exhibit improvements on both the theoretical and practical sides. In particular, they admit exact recovery conditions (i.e., on the underlying maps can be recovered from noisy input maps). Yet due to the limitations of convex optimization, all of these methods do not scale well to large-scale datasets.

In contrast to convex optimizations, we demonstrate that spectral techniques work remarkably well for map synchronization. We focus on the problem of synchronizing permutations and introduce a robust and efficient algorithm that consists of two simple steps. The first step computes the top eigenvectors of a data matrix that encodes the input maps, and the second step rounds each block of

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

the top-eigenvector matrix into a permutation matrix. We show that such a simple algorithm possesses a remarkable denoising ability. In particular, its exact recovery conditions match the state-of-the-art convex optimization techniques. Yet computation-wise, it is much more efficient, and such a property enables us to apply the proposed algorithm on large-scale dataset (e.g., many thousands of objects). Spectral map synchronization has been considered in [12, 13] for input observations between all pairs of objects. In contrast to these techniques, we consider incomplete pair-wise observations, and provide theoretical justifications on a much more practical noise model.

2 Algorithm

In this section, we describe the proposed algorithm for permutation synchronization. We begin with the problem setup in Section 2.1. Then we introduce the algorithmic details in Section 2.2.

2.1 Problem Setup

Suppose we have n objects $S_1, \cdots, S_n$. Each object is represented by m points (e.g., feature points on images and shapes). We consider bijective maps $\phi_{ij} : S_i \to S_j, 1 \leq i, j \leq n$ between pairs of objects. Following the convention, we encode each such map ϕ_{ij} as a permutation matrix $X_{ij} \in \mathcal{P}_m$, where $\mathcal{P}_m$ is the space of permutation matrices of dimension m:

$$\mathcal{P}_m := \{X | X \in [0,1]^{m \times m}, X\mathbf{1}_m = \mathbf{1}_m, X^T\mathbf{1}_m = \mathbf{1}_m\},$$

where $\mathbf{1}_m = (1, \cdots, 1)^T \in \mathbb{R}^m$ is the vector whose elements are 1.

The input permutation synchronization consists of noisy permutations $X_{ij}^{\text{in}} \in \mathcal{G}$ along a connected object graph $\mathcal{G}$. As described in [4, 9], a widely used pipeline to generate such input is to 1) establish the object graph $\mathcal{G}$ by connecting each object and similar objects using object descriptors (e.g., HOG [14] for images) , and 2) apply off-the-shelf pair-wise object matching methods to compute the input pair-wise maps (e.g., SIFTFlow [15] for images and BIM [16] for 3D shapes).

The output consists of improved maps between all of objects

$$X_{ij}, 1 \leq i, j \leq n.$$

2.2 Algorithm

We begin with defining a data matrix $X^{\text{obs}} \in \mathbb{R}^{nm \times nm}$ that encodes the initial pairwise maps in blocks:

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

where $d_i := |\{S_j | (S_i, S_j) \in \mathcal{G}\}|$ is the degree of object S_i in graph $\mathcal{G}$.

Remark 1. Note that the way we encode the data matrix is different from [12, 13] in the sense that we follow the common strategy for handling irregular graphs and use a normalized data matrix.

The proposed algorithm is motivated from the fact that when the input pair-wise maps are correct, the correct maps between all pairs of objects can be recovered from the leading eigenvectors of X^{obs}:

Proposition 2.1. *Suppose there exist latent maps (e.g., the ground-truth maps to one object) $X_i, 1 \leq i \leq n$ so that $X_{ij}^{in} = X_j^T X_i, (i,j) \in \mathcal{G}$. Denote $W \in \mathbb{R}^{nm \times m}$ as the matrix that collects the first m eigenvectors of X^{obs} in its columns. Then the underlying pair-wise maps can be computed from the corresponding matrix blocks of matrix WW^T:*

$$X_j^T X_i = \frac{\sum_{i=1}^n d_i}{\sqrt{d_i d_j}} (WW^T)_{ij}, \quad 1 \leq i, j \leq n. \tag{2}$$

The key insight of the proposed approach is that even when the input maps are noisy (i.e., the blocks of X^{obs} are corrupted), the leading eigenvectors of X^{obs} are still stable under these perturbations (we will analyze this stability property in Section 3). This motivates us to design a simple two-step permutation synchronization approach called NormSpecSync. The first step of NormSpecSync computes the leading eigenvectors of W; the second step of NormSpecSync rounds the induced

Algorithm 1 NormSpecSync

Input: X_{obs} based on (1), $\delta_{\max}$
Initialize W_0: set W_0 as an initial guess for the top-m orthonormal eigenvectors, $k \leftarrow 0$
while $\|W^{(k)} - W^{(k-1)}\| > \delta_{\max}$ **do**
  $W^{(k+1)^+} = X^{\text{obs}} \cdot W^{(k)}$,
  $W^{(k+1)} R^{(k+1)} = W^{(k+1)^+}$, (QR factorization),
  $k \leftarrow k + 1$.
end while
Set $W = W^{(k)}$ and $\overline{X}_{i1}^{\text{spec}} = (WW^T)_{i1}$.
Round each $\overline{X}_{i1}^{\text{spec}}$ into the corresponding X_{i1} by solving (3).
Output: $X_{ij} = X_{j1}^T X_{i1},\ 1 \le i, j \le n$.

matrix blocks (2) into permutations. In the following, we elaborate these two steps and analyze the complexity. Algorithm 1 provides the pseudo-code.

Leading eigenvector computation. Since we only need to compute the leading m eigenvectors of X^{obs}, we propose to use generalized power method. This is justified by the observation that usually there exists a gap between λ_m and λ_{m+1}. In fact, when the input pair-wise maps are correct, it is easy to derive that the leading eigenvectors of X^{obs} are given by:

$$\lambda_1(X^{\text{obs}}) = \cdots = \lambda_m(X^{\text{obs}}) = 1,\ \lambda_{m+1}(X^{\text{obs}}) = \lambda_{n-1}(\mathcal{G}),$$

where $\lambda_{n-1}(\mathcal{G})$ is the second largest eigenvalue of the normalized adjacency matrix of $\mathcal{G}$. As we will see later, the eigen-gap $\lambda_m(X^{\text{obs}}) - \lambda_{m+1}(X^{\text{obs}})$ is still persistent in the presence of corrupted pair-wise maps, due to the stability of eigenvalues under perturbation.

Projection onto $\mathcal{P}_m$. Denote $X_{ij}^{\text{spec}} := \frac{\sum_{i=1}^n d_i}{\sqrt{d_i d_j}} (WW^T)_{ij}$. Since the underlying ground-truth maps $X_{ij}, 1 \le i, j \le n$ obey $X_{ij} = X_{jk}^T X_{ik}, 1 \le i, j \le n$ for any fixed k, we only need to round X_{ik}^{spec} into X_{ik}. Without losing generality, we set $k = 1$ in this paper.

The rounding is done by solving the following constrained optimization problem, which projects X_{i1}^{obs} onto the space of permutations via the Frobenius norm:

$$\begin{aligned} X_{i1} &= \arg\min_{X \in \mathcal{P}_m} \|X - X_{i1}^{\text{obs}}\|_F^2 = \arg\min_{X \in \mathcal{P}_m} \left(\|X\|_F^2 + \|X_{i1}^{\text{obs}}\|_F^2 - 2\langle X, X_{i1}^{\text{obs}} \rangle \right) \\ &= \arg\max_{X \in \mathcal{P}_m} \langle X, X_{i1}^{\text{obs}} \rangle. \end{aligned} \tag{3}$$

The optimization problem described in (3) is the so-called linear assignment problem, which can be solved exactly using the Hungarian algorithm whose complexity is $O(m^3)$ (c.f. [17]). Note that the optimal solution of (3) is invariant under global scaling and shifting of X_{i1}^{obs}, so we omit $\frac{\sum_{i=1}^n d_i}{\sqrt{d_i d_j}}$ and $\frac{1}{m}\mathbf{1}\mathbf{1}^T$ when generating X_{ij}^{obs} (See Algorithm 1).

Time complexity of NormSpecSync. Each step of the generalized power method consists of a matrix-vector multiplication and a QR factorization. The complexity of the matrix-vector multiplication, which leverages of the sparsity in X^{obs}, is $O(n_E \cdot m^2)$, where n_E is the number of edges in $\mathcal{G}$. The complexity of each QR factorization is $O(nm^3)$. As we will analyze laser, generalized power method converges linearly, and setting $\delta_{\max} = 1/n$ provides a sufficiently accurate estimation of the leading eigenvectors. So the total time complexity of the Generalized power method is $O\big((n_E m^2 + nm^3)\log(n)\big)$. The time complexity of the rounding step is $O(nm^3)$. In summary, the total complexity of NormSpecSync is $O\big((n_E m^2 + nm^3)\log(n)\big)$. In comparison, the complexity of the SDP formulation [9], even when it is solved using the fast ADMM method (alternating direction of multiplier method), is at least $O(n^3 m^3 n_{admm}$. So NormSpecSync exhibits significant speedups when compared to SDP formulations.

3 Analysis

In this section, we provide an analysis of NormSpecSync under a generalized Erdős-Rényi noise model.

3.1 Noise Model

The noise model we consider is given by two parameters m and p. Specifically, we assume the observation graph $\mathcal{G}$ is fixed. Then independently for each edge $(i,j) \in \mathcal{E}$,

$$X_{ij}^{\text{in}} = \begin{cases} I_m & \text{with probability } p \\ P_{ij} & \text{with probability } 1-p \end{cases} \tag{4}$$

where $P_{ij} \in \mathcal{P}_m$ is a random permutation.

Remark 2. The noise model described above assumes the underlying permutations are identity maps. In fact, one can assume a generalized noise model

$$X_{ij}^{\text{in}} = \begin{cases} X_{j1}^T X_{i1} & \text{with probability } p \\ P_{ij} & \text{with probability } 1-p \end{cases}$$

where $X_{i1}, 1 \le i \le n$ are pre-defined underlying permutations from object S_i to the first object S_1. However, since P_{ij} are independent of X_{i1}. It turns out the model described above is equivalent to

$$X_{j1} X_{ij}^{\text{in}} X_{i1}^T = \begin{cases} I_m & \text{with probability } p \\ P_{ij} & \text{with probability } 1-p \end{cases}$$

Where P_{ij} are independent random permutations. This means it is sufficient to consider the model described in (4).

Remark 3. The fundamental difference between our model and the one proposed in [11] or the ones used in low-rank matrix recovery [18] is that the observation pattern (i.e., $\mathcal{G}$) is fixed, while in other models it also follows a random model. We argue that our assumption is more practical because the observation graph is constructed by comparing object descriptors and it is dependent on the distribution of the input objects. On the other hand, fixing $\mathcal{G}$ significantly complicates the analysis of NormSpecSync, which is the main contribution of this paper.

3.2 Main Theorem

Now we state the main result of the paper.

Theorem 3.1. *Let $d_{\min} := \min_{1\le i\le n} d_i$, $d^{\text{avg}} := \sum_i d_i/n$, and denote ρ as the second top eigenvalue of normalized adjacency matrix of $\mathcal{G}$. Assume $d_{\min} = \Omega(\sqrt{n}\ln^3 n)$, $d^{\text{avg}} = O(d_{\min})$, $\rho < \min\{p, 1/2\}$. Then under the noise model described above, NormSpecSync recovers the underlying pair-wise maps with high probability if*

$$p > C \cdot \frac{\ln^3 n}{d_{\min}/\sqrt{n}}, \tag{5}$$

for some constant C.

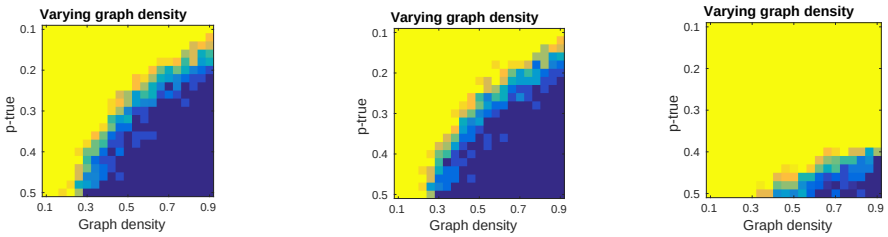
Proof Roadmap. The proof of Theorem 3.1 combines two stability bounds. The first one considers the projection step:

Proposition 3.1. *Consider a permutation matrix $X = (x_{ij}) \in \mathcal{P}_m$ and another matrix $\overline{X} = (\overline{x}_{ij}) \in \mathbb{R}^{m\times m}$. If $\|X - \overline{X}\| < \frac{1}{2}$, then*

$$X = \arg\min_{Y\in\mathcal{P}_m} \|Y - \overline{X}\|_F^2.$$

Proof. The proof is quite straight-forward. In fact,

$$\|X - \overline{X}\|_\infty \le \|X - \overline{X}\| < \frac{1}{2}.$$



(a) NormSpecSync (2.25 seconds) (b) SDP (203.12 seconds) (c) DiffSync(1.07 seconds)

Figure 1: Comparisons between NormSpecSync, SDP[9], DiffSync[13] on the noise model described in Sec. 2.

This means the corresponding element $\overline{x}_{ij}$ of each non-zero element in x_{ij} is dominant in its row and column, i.e.,

$$x_{ij} \neq 0 \;\leftrightarrow\; \overline{x}_{ij} > \max(\max_{k\neq j} \overline{x}_{ik}, \max_{k\neq i} \overline{x}_{kj}),$$

which ends the proof. ■

The second bound concerns the block-wise stability of the leading eigenvectors of X^{obs}:

Lemma 3.1. *Under the assumption of Theorem 3.1, then w.h.p.,*

$$\left\| \frac{\sum_{i=1}^{n} d_i}{\sqrt{d_i d_1}} (WW^T)_{i1} - I_m \right\| < \frac{1}{3}, \quad 1 \leq i \leq n. \tag{6}$$

It is easy to see that we can prove Theorem 3.1 by combing Lemma 3.1 and Prop. 3.1. Yet unlike Prop. 3.1, the proof of Lemma 3.1 is much harder. The major difficulty is that (6) requires controlling each block of the leading eigenvectors, namely, it requires a L^∞ bound, whereas most stability results on eigenvectors are based on the L^2-norm. Due to space constraint, we defer the proof of Lemma 3.1 to Appendix A and the supplemental material. ■

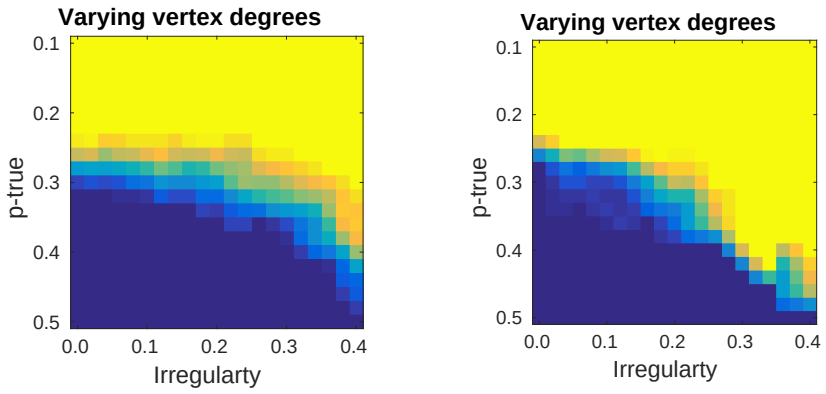
Near-optimality of NormSpecSync. Theorem 3.1 implies that NormSpecSync is near-optimal with respect to the information theoretical bound described in [19]. In fact, when $\mathcal{G}$ is a clique, (5) becomes $p > C \cdot \frac{\ln^3(n)}{\sqrt{n}}$, which aligns with the lower bound in [19] up to a polylogarithmic factor. Following the model described in [19], we can also assume that the observation graph $\mathcal{G}$ is sampled with a density factor q, namely, two objects are connected independently with probability q. In this case, it is easy to see that $d_{\min} > O(nq/\ln n)$ w.h.p., and (5) becomes $p > C \cdot \frac{\ln^4 n}{\sqrt{nq}}$. This bound also stays within a polylogarithmic factor from the lower bound in [19], indicating the near-optimality of NormSpecSync.

4 Experiments

In this section, we perform quantitative evaluations of NormSpecSync on both synthetic and real examples. Experimental results show that NormSpecSync is superior to state-of-the-art map synchronization methods in the literature. We organize the remainder of this section as follows. In Section 4.1, we evaluate NormSpecSync on synthetic examples. Then in section 4.2, we evaluate NormSpecSync on real examples.

4.1 Quantitative Evaluations on Synthetic Examples

We generate synthetic data by following the same procedure described in Section 2. Specifically, each synthetic example is controlled by three parameters $\mathcal{G}$, m, and p. Here $\mathcal{G}$ specifies the input graph; m describes the size of each permutation matrix; p controls the noise level of the input maps. The input maps follow a generalized Erdos-Renyi model, i.e., independently for each edge $(i,j) \in \mathcal{G}$ in the input graph, with probability p the input map $X_{ij}^{in} = I_m$, and otherwise X_{ij}^{in} is a random permutation. To simplify the discussion, we fix $m = 10, n = 200$ and vary the observation graph $\mathcal{G}$ and p to evaluate NormSpecSync and existing algorithms.



(a) NormSpecSync (b) SpecSync

Figure 2: Comparison between NorSpecSync and SpecSync on irregular observation graphs.

Dense graph versus sparse graph. We first study the performance of NormSpecSync with respect to the density of the graph. In this experiment, we control the density of $\mathcal{G}$ by following a standard Erdős-Rényi model with parameter q, namely independently, each edge is connected with probability q. For each pair of fixed p and q, we generate 10 examples. We then apply NormSpecSync and count the ratio that the underlying permutations are recovered. Figure 1(a) illustrates the success rate of NormSpecSync on a grid of samples for p and q. Blue and yellow colors indicate it succeeded and failed on all the examples, respectively, and the colors in between indicate a mixture of success and failure. We can see that NormSpecSync tolerates more noise when the graph becomes denser. This aligns with our theoretical analysis result.

NormSpecSync versus SpecSync. We also compare NormSpecSync with SpecSync [12], and show the advantage of NormSpecSync on irregular observation graphs. To this end, we generate $\mathcal{G}$ using a different model. Specifically, we let the degree of the vertex to be uniformly distribute between $(\frac{1}{2} - q)n$ and $(\frac{1}{2} + q)n$. As illustrated in Figure 2, when q is small, i.e., all the vertices have similar degrees, the performance of NormSpecSync and SpecSync are similar. When q is large, i.e., $\mathcal{G}$ is irregular, NormSpecSync tend to tolerate more noise than SpecSync. This shows the advantage of utilizing a normalized data matrix.

NormSpecSync versus DiffSync. We proceed to compare NormSpecSync with DiffSync [13], which is a permutation synchronization method based on diffusion distances. NormSpecSync and DiffSync exhibit similar computation efficiency. However, NormSpecSync can tolerate significantly more noise than DiffSync, as illustrated in Figure 1(c).

NormSpecSync versus SDP. Finally, we compare NormSpecSync with SDP [9], which formulates permutation synchronization as solving a semidefinite program. As illustrated in Figure 1(b), the exact recovery ability of NormSpecSync and SDP are similar. This aligns with our theoretical analysis result, which shows the near-optimality of NormSpecSync under the noise model of consideration. Yet computationally, NormSpecSync is much more efficient than SDP. The averaged running time for SpecSync is 2.25 second. In contrast, SDP takes 203.12 seconds in average.

4.2 Quantitative Evaluations on Real Examples

In this section, we present quantitative evaluation of NormSpecSync on real datasets.

CMU Hotel/House. We first evaluate NormSpecSync on CMU Hotel and CMU House datasets [20]. The CMU Hotel dataset contains 110 images, where each image has 30 marked feature points. In our experiment, we estimate the initial map between a pair of images using RANSAC [21]. We consider two observation graphs: a clique observation graph $\mathcal{G}_{full}$, where we have initial maps computed between all pairs of images, and a sparse observation graph $\mathcal{G}_{sparse}$. $\mathcal{G}_{sparse}$ is constructed to only connect similar images. In this experiment, we connect an edge between two images if the difference in their HOG descriptors [22] is smaller than $\frac{1}{2}$ of the average descriptor differences among all pairs of images. Note that $\mathcal{G}_{sparse}$ shows high variance in terms of vertex

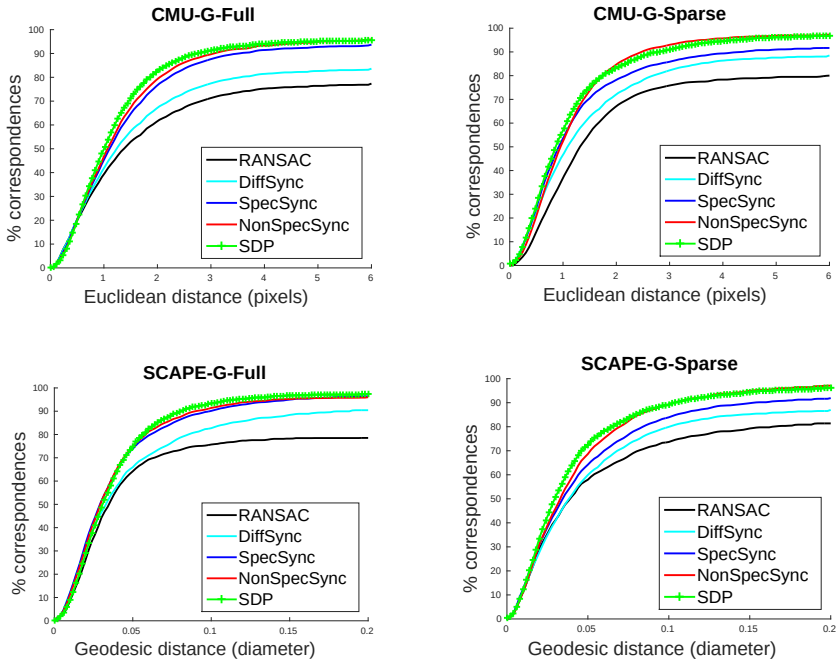


Figure 3: Comparison between NorSpecSync, SpecSync, DiffSync and SDP on CMU Hotel/House and SCAPE. In each dataset, we consider a full observation graph and a sparse observation graph that only connects potentially similar objects.

degree. The CMU House dataset is similar to CMU Hotel, containing 100 images and exhibiting slightly bigger intra-cluster variability than CMU Hotel. We construct the observation graphs and the initial maps in a similar fashion. For quantitative evaluation, we measure the cumulative distribution of distances between the predicted target points and the ground-truth target points.

Figure 3(Left) compares NormSpecSync with the SDP formulation, SpecSync, and DiffSync. On both full and sparse observation graphs, we can see that NormSpecSync, SDP and SpecSync are superior to DiffSync. The performance of NormSpecSync and SpecSync on $\mathcal{G}_{full}$ is similar, while on $\mathcal{G}_{sparse}$, NormSpecSync shows a slight advantage, due to its ability to handle irregular graphs. Moreover, although the performance of NormSpecSync and SDP are similar, SDP is much slower than NormSpecSync. For example, on G_{sparse}, SDP took 1002.4 seconds, while NormSpecSync only took 3.4 seconds.

SCAPE. Next we evaluate NormSpecSync on the SCAPE dataset. SCAPE consists of 71 different poses of a human subject. We uniformly sample 128 points on each model. Again we consider a full observation graph $\mathcal{G}_{full}$ and a sparse observation graph $\mathcal{G}_{sparse}$. $\mathcal{G}_{sparse}$ is constructed in the same way as above, except we use the shape context descriptor [4] for measuring the similarity between 3D models. In addition, the initial maps are computed from blended-intrinsic-map [16], which is the state-of-the-art technique for computing dense correspondences between organic shapes. For quantitative evaluation, we measure the cumulative distribution of geodesic distances between the predicted target points and the ground-truth target points. As illustrated in Figure 3(Right), the relative performance between NormSpecSync and the other three algorithms is similar to CMU Hotel and CMU House. In particular, NormSpecSync shows an advantage over SpecSync on $\mathcal{G}_{sparse}$. Yet in terms of computational efficiency, NormSpecSync is far better than SDP.

5 Conclusions

In this paper, we propose an efficient algorithm named NormSpecSync towards solving the permutation synchronization problem. The algorithm adopts a spectral view of the mapping problem and exhibits surprising behavior both in terms of computation complexity and exact recovery conditions. The theoretical result improves upon existing methods from several aspects, including a fixed obser-

vation graph and a practical noise method. Experimental results demonstrate the usefulness of the proposed approach.

There are multiple opportunities for future research. For example, we would like to extend NormSpecSync to handle the case where input objects only partially overlap with each other. In this scenario, developing and analyzing suitable rounding procedures become subtle. Another example is to extend NormSpecSync for rotation synchronization, e.g., by applying Spectral decomposition and rounding in an iterative manner.

Acknowledgement. We would like to thank the anonymous reviewers for detailed comments on how to improve the paper. The authors would like to thank the support of DMS-1700234, CCF-1302435, CCF-1320175, CCF-1564000, CNS-0954059, IIS-1302662, and IIS-1546500.

A Proof Architecture of Lemma 3.1

In this section, we provide a roadmap for the proof of Lemma 3.1. The detailed proofs are deferred to the supplemental material.

Reformulate the observation matrix. The normalized adjacency matrix $\bar{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ can be decomposed as $\bar{A} = ss^T + V\Lambda V^T$, where the dominant eigenvalue is 1 and corresponding eigenvector is s. We reformulate the observation matrix as $\frac{1}{p}M = \bar{A} \otimes I_m + \tilde{N}$, and it is clear to see that the ground truth result relates to the term $(ss^T) \otimes I_m$, while the noise comes from two terms: $(V\Lambda V^T) \otimes I_m$ and $\tilde{N}$. More specifically, the noise not only comes from the randomness of uncertainty of the measurements, but also from the graph structure, and we use ρ to represent the spectral norm of Λ. When the graph is disconnected or near disconnected, ρ is close to 1 and it is impossible to recover the ground truth.

Bound the spectral norm of $\tilde{N}$. The noise term $\tilde{N}$ consists of random matrices with mean zero in each block. In a complete graph, the spectral norm is bounded by $O(\frac{1}{p\sqrt{n}})$, however, when considering the graph structure, we give a $O(\frac{1}{p\sqrt{d_{\min}}})$ bound.

Measure the block-wise distance between U and $s \otimes I_m$. Let $M = U\Sigma U^T + U_2\Sigma_2 U_2^T$, we want to show the distance between U and $s \otimes \mathbf{1_m}$ is small, where the distance function $dist(\cdot)$ is defined as:

$$dist(U, V) = \min_{R:RR^T=I} \Big\|U - VR\Big\|_B, \tag{7}$$

and this $B-$norm for any matrix X represented in the form $X = [X_1^T, \cdots, X_n^T]^T \in \mathbb{R}^{mn\times m}$ is defined as

$$\|X\|_B = \max_i \|X_i\|_F. \tag{8}$$

More specifically, we bound the distance between U and $s \otimes I_m$ by constructing a series of matrix $\{A_k\}$, and we can show for some $k = O(\log n)$, the distances from $s \otimes A_k$ to both U and $s \otimes I_m$ are small. Therefore, by using the triangle inequality, we can show that U and $s \otimes I_m$ is close.

Sketch proof of Lemma 3.1. Once we are able to show that there exists some rotation matrix R, such that $dist(U, s \otimes I_m)$ is in the order of $o(\frac{1}{\sqrt{n}})$, then it is straightforward to prove Lemma 3.1. Intuitively, this is because the measurements from the eigenvectors is close enough to the ground truth, hence their second moment will still be close. Formally speaking,

$$\begin{aligned}
&\Big\|U_iU_j^T - (s_i \cdot I_m)(s_j \cdot I_m)\Big\| && (9)\\
=&\Big\|U_iRR^TU_j^T - (s_i \cdot I_m)(s_j \cdot I_m)\Big\| && (10)\\
=&\Big\|U_iR(R^TU_j^T - (s_j \cdot I_m)^T) + (U_iR - s_i \cdot I_m)(s_j \cdot I_m)^T\Big\| && (11)\\
\le&\Big\|U_i\Big\| \cdot dist(U, s \otimes I_m) + dist(U, s \otimes I_m) \cdot \Big\|s_j \cdot I_m\Big\| && (12)
\end{aligned}$$

On the other hand, notice that

$$\Big\|\frac{\sum_{i=1}^n d_i}{\sqrt{d_id_j}}U_iU_j^T - I_m\Big\| = \frac{\sum_{i=1}^n d_i}{\sqrt{d_id_j}}\Big\|U_iU_j^T - (s_i \cdot I_m)(s_j \cdot I_m)\Big\|, \tag{13}$$

and we only need to show that (13) is in the order of $o(1)$. The details are included in the supplemental material.

References

[1] D. F. Huber, "Automatic three-dimensional modeling from reality," Tech. Rep., 2002.

[2] D. Crandall, A. Owens, N. Snavely, and D. Huttenlocher, "Discrete-continuous optimization for large-scale structure from motion," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 3001–3008.

[3] Q.-X. Huang, S. Flöry, N. Gelfand, M. Hofer, and H. Pottmann, "Reassembling fractured objects by geometric matching," in *ACM SIGGRAPH 2006 Papers*, 2006, pp. 569–578.

[4] V. G. Kim, W. Li, N. Mitra, S. DiVerdi, and T. Funkhouser, "Exploring collections of 3D models using fuzzy correspondences," *Transactions on Graphics (Proc. of SIGGRAPH 2012)*, vol. 31, no. 4, Aug. 2012.

[5] W. Marande and G. Burger, "Mitochondrial dna as a genomic jigsaw puzzle," *Science*, vol. 318, Jul. 2007.

[6] C. Zach, M. Klopschitz, and M. Pollefeys, "Disambiguating visual relations using loop constraints." in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 1426–1433.

[7] D. Crandall, A. Owens, N. Snavely, and D. Huttenlocher, "Discrete-continuous optimization for large-scale structure from motion," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2011, pp. 3001–3008.

[8] A. Nguyen, M. Ben-Chen, K. Welnicka, Y. Ye, and L. Guibas, "An optimization approach to improving collections of shape maps," in *Eurographics Symposium on Geometry Processing (SGP)*, 2011, pp. 1481–1491.

[9] Q. Huang and L. Guibas, "Consistent shape maps via semidefinite programming," *Computer Graphics Forum, Proc. Eurographics Symposium on Geometry Processing (SGP)*, vol. 32, no. 5, pp. 177–186, 2013.

[10] L. Wang and A. Singer, "Exact and stable recovery of rotations for robust synchronization," *CoRR*, vol. abs/1211.2441, 2012.

[11] Y. Chen, L. J. Guibas, and Q. Huang, "Near-optimal joint object matching via convex relaxation," 2014. [Online]. Available: http://arxiv.org/abs/1402.1473

[12] D. Pachauri, R. Kondor, and V. Singh, "Solving the multi-way matching problem by permutation synchronization," in *Advances in Neural Information Processing Systems*, 2013, pp. 1860–1868.

[13] D. Pachauri, R. Kondor, G. Sargur, and V. Singh, "Permutation diffusion maps (pdm) with application to the image association problem in computer vision," in *Advances in Neural Information Processing Systems*, 2014, pp. 541–549.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, ser. CVPR '05, 2005, pp. 886–893.

[15] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, "Sift flow: Dense correspondence across different scenes," in *Proceedings of the 10th European Conference on Computer Vision: Part III*, ser. ECCV '08, 2008, pp. 28–42.

[16] V. G. Kim, Y. Lipman, and T. Funkhouser, "Blended intrinsic maps," in *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4. ACM, 2011, p. 79.

[17] R. Burkard, M. Dell'Amico, and S. Martello, *Assignment Problems*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2009.

[18] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun. 2011. [Online]. Available: http://doi.acm.org/10.1145/1970392.1970395

[19] Y. Chen, C. Suh, and A. J. Goldsmith, "Information recovery from pairwise measurements: A shannon-theoretic approach," *CoRR*, vol. abs/1504.01369, 2015.

[20] T. S. Caetano, L. Cheng, Q. V. Le, and A. J. Smola, "Learning graph matching," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[21] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[22] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, *ACM Trans. Graph.*, vol. 21, no. 4, pp. 807–832, 2002.

A Reformulating the Noise Term in the Model

As is shown in 3.1, each observed block matrix is corrupted by permutation matrices, therefore the noise can not be directly modeled as random matrices with zero mean, and we need to be careful with it. Since without loss of generality, we can assume each block matrix X_i of the underlying ground truth matrix X to be I_m, for convenience, define X^{gt} as:

$$X^{\text{gt}} = A \otimes I_m + \frac{1-p}{p} A \otimes \left(\frac{1}{m} \mathbf{1_m} \cdot \mathbf{1_m}^T \right). \tag{14}$$

Here, A is the adjacency matrix for graph $\mathcal{G}$. p is the probability that we observe I_m correctly, and with probability $1-p$, the observed block matrix is wrong, i.e., we observe a random permutation. Accordingly, the noise model N we propose written in a block-wise form is:

$$N_{ij} = \begin{cases} \frac{1}{p} I_m - I_m - \frac{1-p}{p} \frac{1}{m} \mathbf{1}_m \cdot \mathbf{1}_m^T, & \text{with probability } p, \\ \frac{1}{p} P_{ij} - I_m - \frac{1-p}{p} \frac{1}{m} \mathbf{1}_m \cdot \mathbf{1}_m^T, & \text{with probability } 1-p, \end{cases} \qquad (i,j) \in \mathcal{G}, \tag{15}$$

and for $(i,j) \notin \mathcal{G}$, $N_{ij} = 0$. Intuitively, the bias term $I_m + \frac{1-p}{p} \frac{1}{m} 1 \cdot 1^T$ is added to offset the first moment of the noise matrix and is helpful for our theoretic analysis. The expectation of each N_{ij} is:

$$\begin{aligned} \mathbb{E}[N_{ij}] =& p\mathbb{E}\left[\frac{1}{p} I_m - I_m - \frac{1-p}{p} \frac{1}{m} \mathbf{1}_m \cdot \mathbf{1}_m^T \right] + (1-p)\mathbb{E}\left[\frac{1}{p} P_{ij} - I_m - \frac{1-p}{p} \frac{1}{m} \mathbf{1}_m \cdot \mathbf{1}_m^T \right] && (16) \\ =& I_m + \frac{1-p}{p} \mathbb{E}[P_{ij}] - I_m - \frac{1-p}{p} \frac{1}{m} \mathbf{1}_m \cdot \mathbf{1}_m^T && (17) \\ =& 0. && (18) \end{aligned}$$

We can rewrite our input matrix X^{in} as a ground truth term with an additive noise term:

$$X^{\text{in}} = pX^{\text{gt}} + pN. \tag{19}$$

Accordingly,

$$X^{\text{obs}} = \left(D^{-\frac{1}{2}} \otimes I_m \right) X^{in} \left(D^{-\frac{1}{2}} \otimes I_m \right). \tag{20}$$

Now, we reconsider the adjacency matrix A of graph $\mathcal{G}$. Ideally, we get a noisy observation of the ground truth matrix X (where all blocks are I_m) when graph $\mathcal{G}$ is fully connected. However, in practice, we may encounter problems getting all the pair mapping results, i.e., incomplete observation. Accordingly, only a subset of entries in A is set to be 1, i.e., $A_{ij} = 1$ if X_{ij} is observed, otherwise $A_{ij} = 0$. We denote the normalized adjacency matrix as $\bar{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$, and we will frequently use the dominant eigenvector s of spectral decomposition of $\bar{A}$:

$$\bar{A} = ss^T + V \Lambda V^T, \tag{21}$$

as well as the second largest eigenvalue of $\bar{A}$, and denote it as $\rho = \|\Lambda\|$.

Furthermore, denote M^* as pX^{obs}, combine previous formula in this section, we have

$$\begin{aligned} \frac{1}{p} M^* &= \bar{A} \otimes I_m + \frac{1-p}{p} \bar{A} \otimes \left(\frac{1}{m} \mathbf{1_m} \cdot \mathbf{1_m}^T \right) + \bar{N} && (22) \\ &= \bar{A} \otimes \left(I_m + \frac{1-p}{p} \frac{1}{m} \mathbf{1_m} \cdot \mathbf{1_m}^T \right) + \bar{N}, && (23) \end{aligned}$$

where $\bar{N} := \left(D^{-\frac{1}{2}} \otimes I_m \right) N \left(D^{-\frac{1}{2}} \otimes I_m \right)$. We normalize the right hand side of (22) and define M as follows:

$$\begin{aligned} \frac{1}{p} M &\triangleq \left(I_n \otimes \left(I_m - \frac{1-\sqrt{p}}{m} \mathbf{1_m} \cdot \mathbf{1_m}^T \right) \right) \left(\frac{1}{p} M^* \right) \left(I_n \otimes \left(I_m - \frac{1-\sqrt{p}}{m} \mathbf{1_m} \cdot \mathbf{1_m}^T \right) \right) && (24) \\ &= \bar{A} \otimes I_m + \tilde{N}, && (25) \end{aligned}$$

where

$$\tilde{N} = \left(I_n \otimes \left(I_m - \frac{1-\sqrt{p}}{m} \mathbf{1_m} \cdot \mathbf{1_m}^T \right) \right) \bar{N} \left(I_n \otimes \left(I_m - \frac{1-\sqrt{p}}{m} \mathbf{1_m} \cdot \mathbf{1_m}^T \right) \right). \tag{26}$$

The purpose of our algorithm is to find the spectral decomposition of (22), i.e., the first m eigenvectors of M^*, since we can only get M^* from the data but not M. However, with basic assumptions on A and p (so that the spectral gap is not too small), we can show that both (22) and (24) give us the same top-m eigenvectors. Therefore, we move our target from (22) to (24) when checking the performance of our algorithm. In the noiseless setting, the algorithm's first m eigenvectors would be $s \otimes I_m$ up to a rotation operation, and we would like to show the convergence to this result in the noisy setting, with reasonable assumption on the value of p, ρ and $\|\tilde{N}\|$.

B Consistency of Eigenvectors

$\rho, \|\tilde{N}\|$ are the spectral norm of $\Lambda, \tilde{N}$, respectively. We assume

$$\rho + \|\tilde{N}\| < (1+\rho)/2, \qquad \frac{1}{p}\rho + \|\bar{N}\| < 1. \tag{27}$$

We can interpret ρ as the noise from the adjacency graph, and $\tilde{N}$ as the noise from observation. Obviously, $\rho < 1$, and accordingly we have $\rho + \|\tilde{N}\| < 1$, which is a necessary condition for the eigengap. The second condition in (27) implies that $\rho < p$, which means that the graph needs to be well connected enough to tolerate large observation noise. Intuitively, this constraint is due to the interaction between graph noise and observation noise, which is also necessary.

B.1 Lemma B.1

Lemma B.1. *With assumption (27), the top-m eigenvectors of M and M^* are consistent, i.e., they span the same space.*

Proof. It is obvious to find the following relationship of M^* and M:

$$\frac{1}{p}M = \frac{1}{p}M^*(I_n \otimes (I_m - \frac{1-p}{m}\mathbf{1_m}\cdot\mathbf{1_m}^T)). \tag{28}$$

Hence,

$$\begin{aligned}
\frac{1}{p}(M^* - M) =& \frac{1-p}{m}\frac{1}{p}M^*\left(I_n \otimes \mathbf{1_m}\cdot\mathbf{1_m}^T\right) && (29)\\
=& \frac{1-p}{m}\left(\bar{A}\otimes\left(I_m + \frac{1-p}{p}\frac{1}{m}\mathbf{1_m}\cdot\mathbf{1_m}^T\right)\right)\left(I_n\otimes\mathbf{1_m}\cdot\mathbf{1_m}^T\right) && (30)\\
=& \frac{1-p}{p}\left(ss^T + V\Lambda V^T\right)\otimes\left(\frac{1}{m}\mathbf{1_m}\cdot\mathbf{1_m}^T\right) && (31)\\
=& \frac{1-p}{p}\left(\frac{1}{\sqrt{m}}s\otimes\mathbf{1_m}\cdot\frac{1}{\sqrt{m}}s^T\otimes\mathbf{1_m}^T + \sum_{i=1}^{n-1}\sigma_i(\Lambda)\frac{1}{\sqrt{m}}v_i\otimes\mathbf{1_m}\cdot\frac{1}{\sqrt{m}}v_i^T\otimes\mathbf{1_m}^T\right), && (32)
\end{aligned}$$

where (30) is because of the noise term always has a block row (column) sum of 0. In (31) we use the property of Kronecker product, i.e., $(X_1\otimes X_2)(X_3\otimes X_4) = (X_1X_3)\otimes(X_2X_4)$, as well as the eigenvector decomposition of $\bar{A}$ defined in (21). Meanwhile, let v_i be the ith column of V,

$$\begin{aligned}
\frac{1}{p}M^*(v_i\otimes\mathbf{1_m}) =& \bar{A}\otimes\left(I_m + \frac{1-p}{p}\frac{1}{m}\mathbf{1_m}\cdot\mathbf{1_m}^T\right)(v_i\otimes\mathbf{1_m}) + \bar{N}(v_i\otimes\mathbf{1_m}) && (33)\\
=& \left(\bar{A}v_i\right)\otimes\left(\left(I_m + \frac{1-p}{p}\frac{1}{m}\mathbf{1_m}\cdot\mathbf{1_m}^T\right)\mathbf{1_m}\right) && (34)\\
=& \frac{1}{p}\sigma_i(\Lambda)(v_i\otimes\mathbf{1_m}) && (35)
\end{aligned}$$

Therefore, the difference between M and M^* only includes the change in eigenvalues corresponding to eigenvectors $\frac{1}{\sqrt{m}}v_i\otimes\mathbf{1_m}$ and $\frac{1}{\sqrt{m}}s\otimes\mathbf{1_m}$. On the other hand, since $\rho + \|\tilde{N}\| < 1$, the top-m eigenvalues of $\bar{A}\otimes I_m$ are $[1, 1, \cdots, 1]$, $\sigma_m(M) > 1 - \|\tilde{N}\| > \rho$, hence the top-$m$ eigenvectors of M do not include $v_i\otimes\mathbf{1_m}$. Meanwhile, since the maximum eigenvalue corresponding to $v_i\otimes\mathbf{1_m}$ for (22) is $\frac{1}{p}\rho$, with assumption (27), the top-m eigenvectors of M^* also do not include $v_i\otimes\mathbf{1_m}$. Therefore, the top-m eigenvectors of M and M^* are consistent.

In this lemma, we try to convey the message that M^* is a perturbation of M that still guarantees the same top-m eigenvector subspace if assumption (27) is satisfied. Moreover, due to the structure of the problem, the perturbation in this problem can be well explained as a summation of n terms, where each of them is a rank-1 matrix spanned by one of the eigenvectors, thus making our analysis easier. ■

C Convergence Analysis

Lemma C.1. *Alg.1 has linear convergence rate. More specifically, let $M = U\Sigma U^T + U_2\Sigma_2U_2^T$, and $diag(\Sigma) = [\sigma_1, \cdots, \sigma_m]$, $diag(\Sigma_2) = [\sigma_{m+1}, \cdots, \sigma_{mn}]$, the columns of U are the m-dominant eigenvectors, then $\hat{X} = W_iW_i^T$ in Alg.1 converges to UU^T linearly.*

Proof. Let $W_k = UC_k + U_2C_{k,2}$, where $U \in \mathbb{R}^{mn\times m}, C_k \in \mathbb{R}^{m\times m}, U_2 \in \mathbb{R}^{mn\times m(n-1)}, C_{k,2} \in \mathbb{R}^{m(n-1)\times m}$. In the following, we assume that C_k is invertible. First of all,

$$W_k^+ = \left(U\Sigma U^T + U_2\Sigma_2 U_2^T\right)(UC_k + U_2C_{k,2}) = U\Sigma C_k + U_2\Sigma_2 C_{k,2}. \tag{36}$$

Besides, $C_k, C_{k,2}$ should satisfy the following:

$$I_m = W_k^T W_k = C_k^T C_k + C_{k,2}^T C_{k,2}. \tag{37}$$

Also, since the orthogonality of column vectors of U and U_2,

$$U^T W_k^+ = \Sigma C_k, \qquad U_2^T W_k^+ = \Sigma_2 C_{k,2}. \tag{38}$$

According to Alg. 1, let the SVD decomposition of W_k^+ be $U_k\Sigma_k^+ V_k^T$, then U_k is what the QR factorization step gives us. Furthermore,

$$W_{k+1}W_{k+1}^T = W_k^+ V_k \Sigma_k^{+-2} V_k^T W_k^{+T} = W_k^+ \left(W_k^{+T} W_k^+\right)^{-1} W_k^{+T}. \tag{39}$$

Substitute (39) with (36), we have:

$$\begin{aligned} W_{k+1}W_{k+1}^T &= W_k^+ \left(W_k^{+T} W_k^+\right)^{-1} W_k^{+T} && (40)\\ &= (U\Sigma C_k + U_2\Sigma_2 C_{k,2})\left(C_k^T\Sigma^2 C_k + C_{k,2}^T\Sigma_2^2 C_{k,2}\right)^{-1}(U\Sigma C_k + U_2\Sigma_2 C_{k,2})^T && (41)\\ &= (U\Sigma + U_2\Sigma_2 C_{k,3})\left(\Sigma^2 + C_{k,3}^T\Sigma_2^2 C_{k,3}\right)^{-1}(U\Sigma + U_2\Sigma_2 C_{k,3})^T, && (42) \end{aligned}$$

where $C_{k,3} = C_{k,2}C_k^{-1}$.

Proposition C.1. *Let the minimum eigenvalue in Σ be σ_m, and the maximum eigenvalue in Σ_2 be σ_{m+1}, $\sigma_m > \sigma_{m+1}$, we have:*

$$\left\|\Sigma^{-1}X^T\Sigma_2^2 X\Sigma^{-1}\right\|_F \le \frac{\sigma_{m+1}^2}{\sigma_m^2}\left\|X^T X\right\|_F. \tag{43}$$

This is obvious once we check the eigenvalues on both sides.

Multiply both sides of (37) with C_k^{-1}, we have:

$$\begin{aligned} C_k^{-1} &= C_k^T + C_{k,2}^T C_{k,2} C_k^{-1}, && (44)\\ \Rightarrow \quad I &= C_k C_k^T + C_k C_{k,2}^T C_{k,2} C_k^{-1} = C_k C_k^T\left(I + C_{k,3}^T C_{k,3}\right), && (45)\\ \Rightarrow C_{k,3}^T C_{k,3} &= \left(C_k C_k^T\right)^{-1} - I. && (46) \end{aligned}$$

According to (40), we have:

$$U^T W_{k+1}W_{k+1}^T U = C_{k+1}C_{k+1}^T = \Sigma\left(\Sigma^2 + C_{k,3}^T\Sigma_2^2 C_{k,3}\right)^{-1}\Sigma^T. \tag{47}$$

Therefore,

$$\left(C_{k+1}C_{k+1}^T\right)^{-1} = \Sigma^{-1}\left(\Sigma^2 + C_{k,3}^T\Sigma_2^2 C_{k,3}\right)\Sigma^{-1} = I + \Sigma^{-1}C_{k,3}^T\Sigma_2^2 C_{k,3}\Sigma^{-1}. \tag{48}$$

According to Proposition C.1 and (46), we have:

$$\begin{aligned} &\left\|(C_{k+1}C_{k+1}^T)^{-1} - I\right\|_F && (49)\\ =&\left\|\Sigma^{-1}C_{k,3}^T\Sigma_2^2 C_{k,3}\Sigma^{-1}\right\|_F && (50)\\ \le&\frac{\sigma_{m+1}^2}{\sigma_m^2}\left\|C_{k,3}^T C_{k,3}\right\|_F && (51)\\ =&\frac{\sigma_{m+1}^2}{\sigma_m^2}\left\|\left(C_k C_k^T\right)^{-1} - I\right\|_F. && (52) \end{aligned}$$

Therefore, W_k has a linear convergence rate. ■

D Prove Lemma 3.1

D.1 Lemma D.1

Lemma D.1. *The spectral norm of $\bar{N}$ is $O(\frac{\sqrt{\ln n}}{p\sqrt{d_{\min}}})$ with high probability.*

Proof. First we introduce the inequality that is useful in the following content, i.e., the Bernstein inequality in the matrix form.

Proposition D.1. Matrix Bernstein. *Let $\{S_k\}$ be a serious of independent matrices with same dimension $d_1 \times d_2$, satisfying $\mathbb{E}[S_k] = \mathbf{0}, \|S_k\| \leq L$. Let Z be the sum of matrices S_k, i.e., $Z = \sum_k S_k$, and let the "variance" of Z be defined as follow:*

$$v(Z) = max\left\{\left\|\sum_k \mathbb{E}\left[S_k S_k^*\right]\right\|, \left\|\sum_k \mathbb{E}\left[S_k^* S_k\right]\right\|\right\}. \tag{53}$$

Then, we have:

$$\mathbb{E}[\|Z\|] \leq \sqrt{2v(Z)\log(d_1+d_2)} + \frac{1}{3}L\log(d_1+d_2) \tag{54}$$

$$\mathbb{P}(\|Z\| \geq t) \leq (d_1+d_2)\exp\left(\frac{-t^2/2}{v(Z)+Lt/3}\right), \tag{55}$$

for $t \geq 0$.

The $\tilde{N}$ we care about can be written as a summation of $2n$ terms as follows:

$$\tilde{N} = \left(\tilde{N}_{:1} + \tilde{N}_{:2} + \cdots \tilde{N}_{:n}\right) + \left(\tilde{N}_{:1} + \tilde{N}_{:2} + \cdots \tilde{N}_{:n}\right)^T, \tag{56}$$

and $\tilde{N}_{:i} = \mathcal{M}_i(\tilde{N})$, where $\mathcal{M}_i(\cdot)$ operator is a mask operation on a square matrix such that only the first i blocks on the ith column are revealed. It is easy to verify that (56) is correct since the diagonal blocks of $\tilde{N}$ are all zero matrices. The norm of $\tilde{N}$ is bounded by two times the norm of the sum of first n terms on the right hand side of (56). Since the n terms are independent, we can use Bernstein inequality.

We observe that,

$$\mathbb{E}\left[\tilde{N}_{:i}\right] = 0, \qquad \|\tilde{N}_{:i}\| \leq \frac{1+p}{p}\frac{1}{\sqrt{d_i d_{\min}}}, \tag{57}$$

$$\left\|\sum_i \mathbb{E}\left[N_{:i}N_{:i}^T\right]\right\| = \left\|\sum_i \mathbb{E}\left[N_{:i}^T N_{:i}\right]\right\| \leq \max_i\left\{\frac{1}{d_i}\sum_{j\in N(i)}\frac{1}{d_j}\right\}\frac{1-p^2}{p^2} \leq \frac{1}{d_{\min}}\frac{1-p^2}{p^2}. \tag{58}$$

Therefore, as long as $d_{\min} = \Omega(\ln n)$, (55) is dominant by the variance term. Accordingly, $\|\tilde{N}\| = O\left(\frac{\sqrt{\ln n}}{p\sqrt{d_{\min}}}\right)$, w.h.p..

■

D.2 Lemma D.2

Lemma D.2. *Introduce matrix series $A_{k,i}$, $0 \leq i \leq k$, which are recursively defined as*

$$A_{1,1} = I_m, A_{1,0} = I_m, \tag{59}$$

$$A_{k,0} = \sum_{i=0}^{k-1} A_{k-1,i}B_i, A_{k,i} = A_{k-1,i-1}, 1 \leq i \leq k. \tag{60}$$

where $B_i = \left(s^T \otimes I_m\right)E^i(s \otimes I_m)$. Then,

$$U_k = \sum_{i=0}^{k} E^i(s \otimes A_{k,i}). \tag{61}$$

Also,

$$A_{k+1,0} = U_k^T(s \otimes I_m) = \left(s^T \otimes I_m\right)M^k(s \otimes I_m) \tag{62}$$

Proof. The proof is quite straightforward and use mathematical induction. When $k = 1$, (61) is true because

$$\sum_{i=0}^{1} E^i \left(s \otimes A_{1,i}\right) = E\left(s \otimes I_m\right) + \left(s \otimes I_m\right) = \left(E + ss^T\right)\left(s \otimes I_m\right) = M\left(s \otimes I_m\right) = U_1. \tag{63}$$

Suppose (61) is true when $k \leq j-1, j \geq 2$. Then

$$U_j = \left(\left(ss^T\right) \otimes I_m + E\right) \sum_{i=0}^{j-1} E^i \left(s \otimes A_{j-1,i}\right) \tag{64}$$

$$= \sum_{i=0}^{j-1} \left(\left(ss^T\right) \otimes I_m + E\right) E^i \left(s \otimes A_{j-1,i}\right) \tag{65}$$

$$= \sum_{i=0}^{j-1} \left(E^{i+1}\left(s \otimes A_{j-1,i}\right) + \left(s \otimes I_m\right)\left(s^T \otimes I_m\right) E^i \left(s \otimes I_m\right) A_{j-1,i}\right) \tag{66}$$

$$= \sum_{i=1}^{j} E^i \left(s \otimes A_{j-1,i}\right) + s \otimes A_{j,0} \tag{67}$$

$$= \sum_{i=0}^{j} E^i \left(s \otimes A_{j-1,i}\right). \tag{68}$$

■

D.3 Lemma D.3

Lemma D.3. $\mathbf{A_{k,i}}$ **is close to** $\mathbf{I_m}$. *For all $k \leq 10\log(n)$, we have*

$$\|A_{k,i} - I_m\| \leq \frac{3(k-i)\sqrt{\ln n}}{1 - (\rho + \|\tilde{N}\|)} \|\tilde{N}\|^2, \quad 0 \leq i \leq k. \tag{69}$$

Proof. We prove this by using mathematical induction. When $k = 1$, (69) is trivial because the left side is 0. Now, suppose (69) is true for $k \leq j$, and let us consider $k = j + 1$. It is clear that we only need to show that $A_{j+1,0}$ satisfies (69), as the rest of $A_{j+1,i}$ follow directly according to definition. Notice that $\|\tilde{N}\| = O(\frac{\sqrt{\ln n}}{p\sqrt{d_{\min}}})$, it follows that we can choose sufficient large n such that

$$\frac{3k\sqrt{\ln n}}{1 - (\rho + \|\tilde{N}\|)} \|\tilde{N}\|^2 \leq \frac{1}{2}, \quad 1 \leq k \leq 10\log(n). \tag{70}$$

For $A_{j+1,0}$, we bound the difference between $A_{j+1,0}$ and I_m as follows:

$$\|A_{j+1,0} - I_m\| \leq \|A_{j,0} - I_m\| + \sum_{i=1}^{j} \|A_{j,i}\| \|B_i\| \tag{71}$$

$$\leq \frac{3j\sqrt{\ln n}}{1 - (\rho + \|\tilde{N}\|)} \|\tilde{N}\|^2 + \sum_{i=1}^{j} (1 + \frac{3(j-i)\sqrt{\ln n}}{1 - (\rho + \|\tilde{N}\|)} \|\tilde{N}\|^2) \|B_i\| \tag{72}$$

$$\leq \frac{3j\sqrt{\ln n}}{1 - (\rho + \|\tilde{N}\|)} \|\tilde{N}\|^2 + (1 + \frac{3(j-1)\sqrt{\ln n}}{1 - (\rho + \|\tilde{N}\|)} \|\tilde{N}\|^2) \|B_1\| + \sum_{i=2}^{j} 2\|B_i\| \tag{73}$$

$$\leq \frac{3j\sqrt{\ln n}}{1 - (\rho + \|\tilde{N}\|)} \|\tilde{N}\|^2 + 2p\sqrt{\ln n} \|\tilde{N}\|^2 + \sum_{i=2}^{j} 2(\|\tilde{N}\| + \rho)^{i-2} \|\tilde{N}\|^2 \tag{74}$$

$$\leq \frac{(3j+3)\sqrt{\ln n}}{1 - (\rho + \|\tilde{N}\|)} \|\tilde{N}\|^2. \tag{75}$$

Hence, $A_{j+1,0}$ also satisfies (69). Therefore, the lemma is proved. ■

D.4 Lemma D.4

Lemma D.4. *The spectral norm of B_i is upper bounded as follows:*

$$\|B_1\| \leq \frac{c\sqrt{\ln n}}{p\sqrt{\sum d_i}}, w.h.p. \qquad \|B_i\| \leq \left(\|\tilde{N}\| + \rho\right)^{i-2} \|\tilde{N}\|^2, \quad \forall i \geq 2. \tag{76}$$

Proof. In our problem, according to the definition of N_{ij}, it is easy to find that

$$\mathbb{E}[N_{ij}N_{ij}^T] = \frac{1-p^2}{p^2}\left(I_m - \frac{1}{m}\mathbf{1_m}\cdot\mathbf{1_m}^T\right) \tag{77}$$

Therefore, if we let k be the number of matrices (in the form of N_{ij}) in the summation Z, then, plug in the result to (54) and (55), we have:

$$v(Z) = k\cdot\frac{1-p^2}{p^2},\ \ L = \frac{1+p}{p}. \tag{78}$$

Now, if we let $Z = \sum_{(i,j)\in\mathcal{Z}} N_{ij}$ and $k = |\mathcal{Z}|$. Then,

$$\mathbb{E}[\|Z\|] = \sqrt{2k\frac{1-p^2}{p^2}\log(2m)} + \frac{1+p}{3p}\log(2m), \tag{79}$$

$$\mathbb{P}(\|Z\| \geq t) \leq 2m\cdot\exp(\frac{-t^2/2}{k\cdot\frac{1-p^2}{p^2} + \frac{(1+p)t}{3p}}) \leq \delta. \tag{80}$$

Here, we let the right hand side of the above equation (80) to be less than some value δ, then we have the following restriction for t when solving the inequality on the right:

$$t \geq -\frac{1+p}{3p}\ln(\frac{\delta}{2m}) + \frac{1}{p}\sqrt{\frac{(1+p)^2}{9}\ln^2(\frac{\delta}{2m}) - 2k(1-p^2)\ln(\frac{\delta}{2m})}. \tag{81}$$

If we let δ to be in the scale of $\frac{1}{k}$, then it is not hard to find that the restriction of t in (81) (in order to have a high probability guarantee on $\|Z\|$) is in $O(\sqrt{k}\ln n)$.

According to definition,

$$B_1 = (s^T\otimes I_m)E(s\otimes I_m) = (s^T\otimes I_m)(V\Lambda V^T\otimes I_m + \tilde{N})(s\otimes I_m) = (s^T\otimes I_m)\tilde{N}(s\otimes I_m). \tag{82}$$

Therefore, the spectral norm of B_1 satisfies:

$$\begin{aligned}
\|B_1\| &= \|(s^T\otimes I_m)\tilde{N}(s\otimes I_m)\| && (83)\\
&= \|\sum_{(i,j)\in[n]\times[n]} \tilde{N_{ij}}s_is_j\| && (84)\\
&\leq \|\sum_{(i,j)\in[n]\times[n]} \bar{N_{ij}}s_is_j\| && (85)\\
&\leq \frac{1}{\sum d_i}\|\sum_{(i,j)\in[n]\times[n]} N_{ij}\| && (86)\\
&\leq \frac{c_1\sqrt{\ln n}}{p\sqrt{\sum d_i}} \quad w.h.p. && (87)\\
& && (88)
\end{aligned}$$

On the other hand, the proof of the bound for $\|B_i\|$ is as follows:

$$\begin{aligned}
\|B_i\| \leq& \|(s^T\otimes I_m)E\|\|E\|^{i-2}\|E(s\otimes I_m)\| && (89)\\
\leq& (\|\tilde{N}\| + \|\Lambda\|)^{i-2}\|\tilde{N}(s\otimes I_m)\|^2 && (90)\\
\leq& (\|\tilde{N}\| + \rho)^{i-2}\|\tilde{N}\|^2. && (91)
\end{aligned}$$

■

D.5 Lemma D.5

Lemma D.5. *For each fixed $k \geq 2$,*

$$\|E^k(s\otimes I_m)\|_B \leq \frac{ck\sqrt{m}\sqrt{\ln n}}{pd_{\min}} \tag{92}$$

Proof. Now, we proceed to bound the term $\|E^k(s\otimes I_m)\|_B$, and we start with simple case when k is small (and just consider spectral norm, since there is only a $\sqrt{m}$ difference). After deriving the bound for small k

values, we will use recursion to bound k in general.
$k = 1$. Then,

$$\|(e_1^T \otimes I_m)E(s \otimes I_m)\| \tag{93}$$
$$=\|(e_1^T \otimes I_m)\tilde{N}(s \otimes I_m)\| \tag{94}$$
$$=\|\sum_i \tilde{N}_{1i}s_i\| \tag{95}$$
$$=\frac{s_1}{d_1}\|\sum_i N_{1i}\| \tag{96}$$
$$\leq\frac{c\sqrt{\ln n}}{p\sqrt{\sum d_i}} \tag{97}$$
$$\leq\frac{c\sqrt{\ln n}}{pd_{\min}}. \tag{98}$$

Assume that we have a result in the form of the following:

$$\|(e_1^T \otimes I_m)E^k(s \otimes I_m)\| \leq C_A(k)\sqrt{\ln n} \tag{99}$$

Proposition D.2.

$$\mathbb{P}[\|\sum_{i,j}(V\Lambda V^T)^t_{1i}\tilde{N}_{ij}s_j\| > \frac{c\rho^t\sqrt{\ln n}}{p\sqrt{\sum d_i}}] = O(\frac{1}{n}). \tag{100}$$

This is easy to check using Bernstein inequality and the property that $\sum_i (V\Lambda V^T)^{2t}_{1i} \leq \rho^{2t}$.

Let's consider general $k = O(\log n)$. Denote $E_{2:n,2:n}$ as the block matrix that removes the first row block and column block from E, denote E^i_{11} as the block submatrix of E^i at position $(1, 1)$ (instead of E_{11} to the i-th power). Also, let $e_{j-1,n-1}$ be the $(n-1)\times 1$ vector where the $(j-1)$-th entry is 1 and all other entries are 0. We have

$$\|(e_1^T \otimes I_m)E^k(s \otimes I_m)\|_F \tag{101}$$
$$=\|(e_1^T \otimes I_m)E^{k-1}\tilde{N}(s \otimes I_m)\|_F \tag{102}$$
$$=\|(E^{k-1}_{11})(e_1^T \otimes I_m)\tilde{N}(s \otimes I_m) + \sum_{i=1}^{k-1}(E^{k-1-i}_{11})\sum_{j=2}^{n} E_{1j}((e^T_{j-1,n-1} \otimes I_m)E^{i-1}_{2:n,2:n}\tilde{N}_{2:n}(s \otimes I_m))\|_F \tag{103}$$
$$\leq\|E^{k-1}_{11}\|_F\|(e_1^T \otimes I_m)\tilde{N}(s \otimes I_m)\| + \sum_{i=1}^{k-1}\|(E^{k-1-i}_{11})\|_F\|\sum_{j=2}^{n} E_{1j}((e^T_{j-1,n-1} \otimes I_m)E^{i-1}_{2:n,2:n}\tilde{N}_{2:n}(s \otimes I_m))\|. \tag{104}$$
$$\leq\sqrt{m}(\rho + \|\tilde{N}\|)^{k-1}\|(e_1^T \otimes I_m)\tilde{N}(s \otimes I_m)\| \tag{105}$$
$$+\sum_{i=1}^{k-1}\sqrt{m}(\rho + \|\tilde{N}\|)^{k-1-i}\|\sum_{j=2}^{n}\tilde{N}_{1j}((e^T_{j-1,n-1} \otimes I_m)E^{i-1}_{2:n,2:n}\tilde{N}_{2:n}(s \otimes I_m))\| \tag{106}$$
$$+\sum_{i=1}^{k-1}\sqrt{m}(\rho + \|\tilde{N}\|)^{k-1-i}\|\sum_{j=2}^{n}(V\Lambda V^T)_{1j}((e^T_{j-1,n-1} \otimes I_m)E^{i-1}_{2:n,2:n}\tilde{N}_{2:n}(s \otimes I_m))\|. \tag{107}$$

In the above equations, we have used the following properties:

$$\|E^i_{11}\|^2_F = \sum_{j=1}^{m}\sigma^2_j(E^i_{11}) \leq m \cdot \|E^i_{11}\|^2 \leq m \cdot \|E^i\|^2 \leq m(\rho + \|\tilde{N}\|)^{2i}, \tag{108}$$

To simplify the above equation (106), we define:

$$\mathcal{N}(i, \phi, k) \triangleq \sum_{j\in\phi}\tilde{N}_{kj}((e_j^T \otimes I_m)E^{i-1}_{\phi,\phi}\tilde{N}_{\phi}(s \otimes I_m)) \tag{109}$$
$$=\sum_{j\in\phi}\tilde{N}_{kj}((e_j^T \otimes I_m)E^{i-1}_{\phi,\phi}\tilde{N}_{\phi,\phi}(s_\phi \otimes I_m)) + \sum_{j\in\phi}\tilde{N}_{kj}((e_j^T \otimes I_m)E^{i-1}_{\phi,\phi}\tilde{N}_{\phi,\bar{\phi}}(s_{\bar{\phi}} \otimes I_m)), \tag{110}$$

where ϕ is a subset of $[n]$ and $n - |\phi| \leq k = O(\log n)$.

For the first term in (110), we treat the part $((e_j^T \otimes I_m)E_{\phi,\phi}^{i-1}\tilde{N}_\phi(s_\phi \otimes I_m)$ as fixed matrices, then $\mathcal{N}(i,\phi,k)$ is a sum of independent random variables. Using matrix Bernstein inequality, we have

$$v(\mathcal{N}(i,\phi,k)) = \|\mathbb{E}[\sum_{j\in\phi}(s_\phi^T \otimes I_m)\tilde{N}_\phi E_\phi^{i-1}(e_j \otimes I_m)\tilde{N}_{kj}^T\tilde{N}_{kj}(e_j^T \otimes I_m)E_\phi^{i-1}\tilde{N}_\phi(s_\phi \otimes I_m)]\| \tag{111}$$

$$= \|\sum_{j\in\phi}\frac{1-p^2}{p^2}\frac{1}{d_k d_j}(s_\phi^T \otimes I_m)\tilde{N}_\phi E_\phi^{i-1}(e_j \otimes I_m)(e_j^T \otimes I_m)E_\phi^{i-1}\tilde{N}_\phi(s_\phi \otimes I_m)\| \tag{112}$$

$$\leq \frac{1}{p^2}\|\tilde{N}\|^2(\rho + \|\tilde{N}\|)^{2i-2}\frac{1}{d_k d_{\min}} \tag{113}$$

and each term is bounded by $\frac{1+p}{p}\frac{1}{\sqrt{d_k d_j}}C_A(i)\sqrt{\ln n}$.

For the second term in (110), since $\bar{\phi}$ includes $O(\ln n)$ terms, we can separately bound the parts in it:

$$\|\sum_{j\in\phi}\tilde{N}_{kj}((e_j^T \otimes I_m)E_{\phi,\phi}^{i-1}\tilde{N}_{\phi,\bar{\phi}}(s_{\bar{\phi}} \otimes I_m))\| \tag{114}$$

$$=\|\sum_{j,l\in\phi,r\in\bar{\phi}}\tilde{N}_{kj}((e_j^T \otimes I_m)E_{\phi,\phi}^{i-1}(e_l \otimes I_m)\tilde{N}_{lr}(s_r \otimes I_m))\| \tag{115}$$

$$\leq\|\sum_{j\in\phi}\tilde{N}_{kj}\| \cdot (\rho + \|\tilde{N}\|)^{i-1} \cdot \|\sum_{l\in\phi,r\in\bar{\phi}}\tilde{N}_{lr}s_r\| \tag{116}$$

$$\leq\|\tilde{N}\|(\rho + \|\tilde{N}\|)^{i-1}\frac{c\sqrt{\ln n}}{pd_{\min}} \tag{117}$$

Therefore, combining the bounds for the first and second term in (110) using (113), (117):

$$\|\mathcal{N}(i,\phi,k)\| \leq \frac{1}{p}(\rho + \|\tilde{N}\|)^{i-1}\|\tilde{N}\|\frac{c\sqrt{\ln n}}{d_{\min}}, \tag{118}$$

On the other hand, in order to simplify (107), define:

$$\Lambda(i,1,[n]/1) \triangleq \sum_{j\in[n]/1}(V\Lambda V^T)_{1j}(e_j^T \otimes I_m)E_{2:n,2:n}^{i-1}\tilde{N}_{2:n}(s \otimes I_m) \tag{119}$$

$$= \sum_{j\in[n]/1}(V\Lambda V^T)_{1j}[(E_{2:n}^{i-1})_{jj}(e_j^T \otimes I_m)\tilde{N}_{2:n}(s \otimes I_m) \tag{120}$$

$$+ \sum_{i'=1}^{i-1}(E_{2:n}^{i-1-i'})_{jj}\sum_{j'=2,j'\neq j}^{n}E_{jj'}(e_{j'}^T \otimes I_m)E_{2:n/j',2:n/j'}^{i'-1}\tilde{N}_{2:n/j'}(s \otimes I_m)], \tag{121}$$

where the second parameter 1 represents the number of $(V\Lambda V^T)$ terms in the expansion. Accordingly,

$$\|\Lambda(i,1,[n]/1)\| \tag{122}$$

$$\leq\|(E_{2:n}^{i-1})_{jj}\| \cdot \|\sum_{j=2}^{n}\sum_{k\neq j}(V\Lambda V^T)_{1j}\tilde{N}_{jk}s_k\| \tag{123}$$

$$+ \sum_{i'=1}^{i-1}\|(E_{2:n}^{i-1-i'})_{jj}\| \cdot \|\sum_{j=2}^{n}\sum_{j'}(V\Lambda V^T)_{1j}\tilde{N}_{jj'}(e_{j'}^T \otimes I_m)E_{2:n/j'}^{i'-1}\tilde{N}_{2:n/j'}(s \otimes I_m)\| \tag{124}$$

$$+ \sum_{i'=1}^{i-1}\|(E_{2:n}^{i-1-i'})_{jj}\| \cdot \|\sum_{j=2}^{n}\sum_{j'}(V\Lambda V^T)_{1j}(V\Lambda V^T)_{jj'}(e_{j'}^T \otimes I_m)E_{2:n/j'}^{i'-1}\tilde{N}_{2:n/j'}(s \otimes I_m)\| \tag{125}$$

Notice that for (124), we can separate it into two terms to maintain the independence of $\tilde{N}_{jj'}$:

$$\|\sum_{j=2}^{n}\sum_{j'}(V\Lambda V^T)_{1j}^t\tilde{N}_{jj'}(e_{j'}^T \otimes I_m)E_{2:n/j'}^{i'-1}\tilde{N}_{2:n/j'}(s \otimes I_m)\| \tag{126}$$

$$\leq\|\sum_{j=2}^{n}\sum_{j'>j}(V\Lambda V^T)_{1j}^t\tilde{N}_{jj'}(e_{j'}^T \otimes I_m)E_{2:n/j'}^{i'-1}\tilde{N}_{2:n/j'}(s \otimes I_m)\| \tag{127}$$

$$+\|\sum_{j=2}^{n}\sum_{j'<j}(V\Lambda V^T)_{1j}^t\tilde{N}_{jj'}(e_{j'}^T \otimes I_m)E_{2:n/j'}^{i'-1}\tilde{N}_{2:n/j'}(s \otimes I_m)\| \tag{128}$$

For each term above, we can treat it as a sum of independent random variables. Using similar arguments for bounding $\mathcal{N}$, we get the following bound:

$$\|\sum_{j=2}^{n}\sum_{j'}(V\Lambda V^T)^t_{1j}\tilde{N}_{jj'}(e_{j'}^T\otimes I_m)E^{i'-1}_{2:n/j'}\tilde{N}_{2:n/j'}(s_{2:n/j'}\otimes I_m)\| \tag{129}$$

$$\leq \frac{\rho^t}{p}(\rho+\|\tilde{N}\|)^{i'-1}\|\tilde{N}\|\frac{c\sqrt{\ln n}}{d_{\min}}. \tag{130}$$

Recursively, we have:

$$\|\Lambda(i,t,\phi)\| \leq (\rho+\|\tilde{N}\|)^{i-1}\frac{c\rho^t\sqrt{\ln n}}{p\sqrt{\sum d_i}} \tag{131}$$

$$+\sum_{i'=1}^{i-1}(\rho+\|\tilde{N}\|)^{i-1-i'}\frac{\rho^t}{p}(\rho+\tilde{N})^{i'-1}\|\tilde{N}\|\frac{c\sqrt{\ln n}}{d_{\min}} \tag{132}$$

$$+\sum_{i'=1}^{i-1}(\rho+\|\tilde{N}\|)^{i-1-i'}\cdot\|\Lambda(i',t+1,\phi/j')\|. \tag{133}$$

This gives the bound for Λ:

$$\|\Lambda(k,t,\phi)\| \leq (2\rho+2\|\tilde{N}\|)^{k-1}\frac{c\rho^t\sqrt{\ln n}}{pd_{\min}}. \tag{134}$$

On the other hand,

$$C_A(k) \leq (\rho+\|\tilde{N}\|)^{k-1}C_A(1) \tag{135}$$

$$+\sum_{i=1}^{k-1}(\rho+\|\tilde{N}\|)^{k-1-i}(\rho+\|\tilde{N}\|)^{i-1}\frac{\sqrt{\ln n}\|\tilde{N}\|}{pd_{\min}} \tag{136}$$

$$+\sum_{i=1}^{k-1}(\rho+\|\tilde{N}\|)^{k-1-i}\|\Lambda(i,1,\phi)\|/\sqrt{\ln n}. \tag{137}$$

In the base case when $k=1$, we have:

$$C_A(1) \leq \frac{c}{pd_{\min}}. \tag{138}$$

After recursion steps based on (137), we have:

$$\|E^k(s\otimes I_m)\|_B \leq k(2\rho+2\|\tilde{N}\|)^{k-1}\frac{c\sqrt{m}\sqrt{\ln n}}{pd_{\min}} \leq \frac{ck\sqrt{m}\sqrt{\ln n}}{pd_{\min}} \tag{139}$$

with high probability, where we require $\rho+\|\tilde{N}\|\leq 1/2$.

■

D.6 Lemma D.6

As we already know, $M = U\Sigma U^T + U_2\Sigma_2 U_2^T$. Now, define $F_k = \Sigma^k U^T(s\otimes I_m)$, we have:

Lemma D.6. *Let F_k be defined as above. Then, for $k \geq \frac{2}{\log(\frac{2}{1+\rho})}$:*

$$\|U_2\Sigma_2^k U_2^T\| \leq \frac{1}{n^2}, \qquad \|F_k^T F_k - I_m\| \leq \frac{(6k+4)\sqrt{\ln n}}{1-(\rho+\|\tilde{N}\|)}\|\tilde{N}\|^2. \tag{140}$$

Proof. According to the eigenvalue stability and assumption (27), we have the following:

$$\sigma_{m+1} < \rho+\|\bar{N}\| < (1+\rho)/2. \tag{141}$$

Taking $k = \frac{2}{\log(\frac{2}{1+\rho})}\log n = O(\log n)$,

$$\|U_2\Sigma_2^k U_2^T\| = \sigma_{m+1}^k(M) \leq \frac{1}{n^2}. \tag{142}$$

Based on the bound of the noises, we will show that F_k is close to I_m.

$$\begin{aligned}
&\|F_k^T F_k - I_m\| && (143)\\
=&\|F_k^T U^T U F_k - I_m\| && (144)\\
=&\|(M^k(s\otimes I_m) - U_2\Sigma_2^k U_2^T(s\otimes I_m))^T(M^k(s\otimes I_m) - U_2\Sigma_2^k U_2^T(s\otimes I_m)) - I_m\| && (145)\\
\le&\|(s^T\otimes I_m)M^{2k}(s\otimes I_m) - I_m\| + 2\|U_2\Sigma_2^k U_2^T(s\otimes I_m)\|\|U_k\| + \|U_2\Sigma_2^k U_2^T(s\otimes I_m)\|^2 && (146)\\
\le&\|A_{2k+1,0} - I_m\| + \frac{3}{n^2} \qquad (\text{ since } \|U_k\| \le \|A_{2k+1,0}\|^{\frac{1}{2}}) && (147)\\
\le&\frac{(6k+4)\sqrt{\ln n}}{1-(\rho+\|\tilde{N}\|)}\|\tilde{N}\|^2. && (148)
\end{aligned}$$

■

The spectral norm bound described above indicates the singular values of F_k fall in

$$\left[1 - \frac{(3k+2)\sqrt{\ln n}}{1-(\rho+\|\tilde{N}\|)}\|\tilde{N}\|, 1 + \frac{(3k+2)\sqrt{\ln n}}{1-(\rho+\|\tilde{N}\|)}\|\tilde{N}\|\right].$$

This means that the singular values of F_k^{-1} fall in

$$\left[1 - \frac{4k\ln n}{1-(\rho+\|\tilde{N}\|)}\|\tilde{N}\|, 1 + \frac{4k\ln n}{1-(\rho+\|\tilde{N}\|)}\|\tilde{N}\|\right],$$

(which implicitly assumes that $\frac{\sqrt{\ln n}\|\tilde{N}\|^2}{1-(\rho+\|\tilde{N}\|)} < \frac{k-2}{4k(3k+2)}$ and is true for large n). Thus, there exists a rotation matrix R, such that

$$\|F_k^{-1} - R\| \le \frac{4k\sqrt{\ln n}}{1-(\rho+\|\tilde{N}\|)}\|\tilde{N}\|^2. \tag{149}$$

D.7 Proof of Lemma 3.1

Lemma D.7. *With assumption (27), we can conclude that*

$$dist(U, s\otimes I_m) \le \frac{c_1\sqrt{\ln^3 n}}{p^2 d_{\min}^{3/2}} + \frac{c_2}{n^2} + \frac{c_3\sqrt{\ln^5 n}}{p\sqrt{n d_{\min}}} = O(\frac{\sqrt{\ln^5 n}}{p\sqrt{n d_{\min}}}). \tag{150}$$

with high probability, where c_1, c_2, c_3 are global constants.

Proof.

$$dist(U, s\otimes I_m) \le dist(U, s\otimes A_{k,0}) + dist(s\otimes A_{k,0}, s\otimes I_m). \tag{151}$$

In the following, we will bound the two terms on the right hand side of (151) separately. First, let R be the minimizer in the rotation matrix class described in (149). Then,

$$\begin{aligned}
&dist(U, s\otimes A_{k,0}) && (152)\\
\le&\|U - (s\otimes A_{k,0})R\|_B && (153)\\
=&\|M^k(s\otimes I_m)F_k^{-1} - U_2\Sigma_2^k U_2^T(s\otimes I_m)F_k^{-1} - (s\otimes A_{k,0})R\|_B && (154)\\
=&\|M^k(s\otimes I_m)(F_k^{-1} - R) - U_2\Sigma_2^k U_2^T(s\otimes I_m)F_k^{-1} + (M^k(s\otimes I_m) - (s\otimes A_{k,0}))R\|_B && (155)\\
\le&\|M^k(s\otimes I_m)\|_B\|F_k^{-1} - R\| + \|U_2\Sigma_2^k U_2^T(s\otimes I_m)\|_B\|F_k^{-1}\| + \|(M^k(s\otimes I_m) - (s\otimes A_{k,0}))R\|_B. && (156)
\end{aligned}$$

We bound each term in (156).

$$\begin{aligned}
&\|M^k(s\otimes I_m)\|_B && (157)\\
\le&\|M^k(s\otimes I_m)\|_F && (158)\\
\le&\sqrt{m}\|M^k(s\otimes I_m)\| && (159)\\
=&\sqrt{m}\|A_{2k+1,0}\|^{\frac{1}{2}} && (160)\\
\le&\sqrt{2m}. && (161)
\end{aligned}$$

$$\|U_2\Sigma_2^k U_2^T(s\otimes I_m)\|_B \le \frac{\sqrt{m}}{n^{10k\cdot\ln\frac{1}{\rho+\|N\|}}} \le \frac{\sqrt{m}}{n^2}. \tag{162}$$

$$\|(M^k(s\otimes I_m)-(s\otimes A_{k,0}))R\|_B \tag{163}$$
$$=\|\sum_{i=1}^{k}(E^i(s\otimes I_m)A_{k,i})\|_B \tag{164}$$
$$\le 2\sum_{i=1}^{k}\frac{ck\sqrt{m}\sqrt{\ln n}}{pd_{\min}} \tag{165}$$
$$\le \frac{c\sqrt{m}\sqrt{\ln^5 n}}{pd_{\min}}. \tag{166}$$

Finally, we have

$$\|U-(s\otimes A_{k,0})R\|_B \le \frac{4k\sqrt{2m}\sqrt{\ln n}}{1-(\rho+\|\tilde{N}\|)}\|\tilde{N}\|^2 + \frac{2\sqrt{m}}{n^2} + \frac{c\sqrt{m}\sqrt{\ln^5 n}}{pd_{\min}}. \tag{167}$$

On the other hand,

$$dist(s\otimes A_{k,0}, s\otimes I_m) \tag{168}$$
$$\le \max(s)dist(A_{k,0}, I_m) \tag{169}$$
$$\le \max(s)\|A_{k,0}, I_m\|_F \tag{170}$$
$$\le \frac{1}{\sqrt{d_{\min}}}\sqrt{m}\|A_{k,0}-I_m\| \tag{171}$$
$$\le \frac{1}{\sqrt{d_{\min}}}\frac{3k\sqrt{m}\sqrt{\ln n}}{1-(\rho+\|\tilde{N}\|)}\|\tilde{N}\|^2. \tag{172}$$

Therefore, combining (167) and (172), $dist(U, s\otimes I_m)$ is bounded by:

$$dist(U, s\otimes I_m) \le \frac{c_1\sqrt{\ln^3 n}}{p^2 d_{\min}^{3/2}} + \frac{c_2}{n^2} + \frac{c_3\sqrt{\ln^5 n}}{pd_{\min}} = O(\frac{\sqrt{\ln^5 n}}{pd_{\min}}). \tag{173}$$

with high probability, where c_1, c_2, c_3 are some global constants. Therefore, the distance is bounded by $O(\frac{\sqrt{\ln^5 n}}{pd_{\min}})$. ■

Now, we are able to show that there exists some rotation matrix R, such that $dist(U, s\otimes I_m)$ is in the order of $o(\frac{1}{\sqrt{n}})$, then it is straightforward to prove Lemma 3.1. This is because:

$$\|U_iU_j^T - (s_i\cdot I_m)(s_j\cdot I_m)\| \tag{174}$$
$$=\|U_iRR^TU_j^T - (s_i\cdot I_m)(s_j\cdot I_m)\| \tag{175}$$
$$=\|U_iR(R^TU_j^T - (s_j\cdot I_m)^T) + (U_iR - s_i\cdot I_m)(s_j\cdot I_m)^T\| \tag{176}$$
$$\le \|U_i\|\cdot dist(U, s\otimes I_m) + dist(U, s\otimes I_m)\cdot\|s_j\cdot I_m\| \tag{177}$$

Therefore,

$$\|\frac{\sum_{i=1}^n d_i}{\sqrt{d_id_j}}U_iU_j^T - I_m\| \tag{178}$$
$$\le dist(U, s\otimes I_m)\max(s_i, s_j)\frac{\sum_{i=1}^n d_i}{\sqrt{d_id_j}} \tag{179}$$
$$\le dist(U, s\otimes I_m)\frac{\sqrt{n}\sqrt{d^{\text{avg}}}}{\sqrt{d_{\min}}}. \tag{180}$$

If we assume that $d^{\text{avg}} = O(d_{\min})$, for $d_{\min} = \Omega(\sqrt{n}\ln^3 n)$, we need:

$$p \ge C\cdot\frac{(\ln n)^3}{d_{\min}/\sqrt{n}}. \tag{181}$$

Without this assumption on the average degree, we need a stronger assumption on the minimum degree, i.e., $d_{\min} = \Omega((nd^{\text{avg}})^{1/3}\ln^2 n)$, and we need:

$$p \ge C\cdot\frac{(\ln n)^3(nd^{\text{avg}})^{1/2}}{d_{\min}^{3/2}}. \tag{182}$$