

# An extension of the angular synchronization problem to the heterogeneous setting\*

Mihai Cucuringu<sup>\*†</sup>, Hemant Tyagi<sup>‡§</sup>

December 29, 2020

## Abstract

Given an undirected measurement graph  $G = ([n], E)$ , the classical angular synchronization problem consists of recovering unknown angles  $\theta_1, \dots, \theta_n$  from a collection of noisy pairwise measurements of the form  $(\theta_i - \theta_j) \bmod 2\pi$ , for each  $\{i, j\} \in E$ . This problem arises in a variety of applications, including computer vision, time synchronization of distributed networks, and ranking from preference relationships. In this paper, we consider a generalization to the setting where there exist  $k$  unknown groups of angles  $\theta_{l,1}, \dots, \theta_{l,n}$ , for  $l = 1, \dots, k$ . For each  $\{i, j\} \in E$ , we are given noisy pairwise measurements of the form  $\theta_{\ell,i} - \theta_{\ell,j}$  for an *unknown*  $\ell \in \{1, 2, \dots, k\}$ . This can be thought of as a natural extension of the angular synchronization problem to the heterogeneous setting of multiple groups of angles, where the measurement graph has an unknown edge-disjoint decomposition  $G = G_1 \cup G_2 \dots \cup G_k$ , where the  $G_i$ 's denote the subgraphs of edges corresponding to each group. We propose a probabilistic generative model for this problem, along with a spectral algorithm for which we provide a detailed theoretical analysis in terms of robustness against both sampling sparsity and noise. The theoretical findings are complemented by a comprehensive set of numerical experiments, showcasing the efficacy of our algorithm under various parameter regimes. Finally, we consider an application of bi-synchronization to the graph realization problem, and provide along the way an iterative graph disentangling procedure that uncovers the subgraphs  $G_i$ ,  $i = 1, \dots, k$  which is of independent interest, as it is shown to improve the final recovery accuracy across all the experiments considered.

**Keywords:** group synchronization, spectral algorithms, matrix perturbation theory, singular value decomposition, random matrix theory.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Angular Synchronization over <math>SO(2)</math></b>	<b>3</b>
<b>3</b>	<b><math>k</math>-synchronization: setup, motivation, &amp; main results</b>	<b>6</b>
3.1	Applications . . . . .	6
3.2	Summary of our main contributions . . . . .	8
<b>4</b>	<b>Bi-synchronization: Synchronization with two groups of angles</b>	<b>9</b>
4.1	Spectral method . . . . .	10
4.2	Erdős-Renyi measurement model . . . . .	10
4.3	Main result . . . . .	11
4.4	Proof of Theorem 1 . . . . .	12

<sup>1</sup>Department of Statistics and Mathematical Institute, University of Oxford, Oxford, UK. Email: mihai.cucuringu@stats.ox.ac.uk

<sup>2</sup>The Alan Turing Institute, London, UK

<sup>3</sup>Inria, Univ. Lille, CNRS, UMR 8524 - Laboratoire Paul Painlevé, F-59000. Email: hemant.tyagi@inria.fr

<sup>4</sup>Authors are listed in alphabetical order.

\*This work was supported by EPSRC grant EP/N510129/1

<b>5</b>	<b><i>k</i>-synchronization: Synchronization with <i>k</i> groups of angles</b>	<b>13</b>
5.1	Erdős-Renyi measurement model . . . . .	14
5.2	Main result . . . . .	14
5.3	Proof of Theorem 2 . . . . .	16
<b>6</b>	<b>Numerical Experiments</b>	<b>17</b>
6.1	Setup I: correlation versus graph sparsity . . . . .	18
6.2	Setup II: fixed gap experiments, correlation versus graph sparsity . . . . .	18
6.3	Comparison with the Normalized Spectral and SDP relaxations . . . . .	19
6.4	Graph disentangling procedure . . . . .	22
<b>7</b>	<b>An application to the graph realization problem</b>	<b>24</b>
7.1	The graph realization problem . . . . .	24
7.2	The ASAP algorithm . . . . .	27
7.3	The multi-graph realization problem . . . . .	27
<b>8</b>	<b>Conclusion and future directions</b>	<b>30</b>
<b>A</b>	<b>Perturbation analysis</b>	<b>34</b>
<b>B</b>	<b>Useful concentration inequalities</b>	<b>35</b>
B.1	Sum of sub-Gaussian random variables . . . . .	35
B.2	Spectral norm of random matrices . . . . .	35
<b>C</b>	<b>Proofs from Section 4</b>	<b>35</b>
C.1	Proof of Proposition 1 . . . . .	35
C.2	Proof of Lemma 1 . . . . .	36
C.3	Proof of Lemma 2 . . . . .	36
C.4	Proof of Lemma 3 . . . . .	37
C.5	Proof of Proposition 2 . . . . .	38
<b>D</b>	<b>Proofs from Section 5</b>	<b>38</b>
D.1	Proof of Lemma 4 . . . . .	38
D.2	Proof of Lemma 5 . . . . .	42
D.3	Proof of Lemma 6 . . . . .	45

# 1 Introduction

Finding group elements from noisy pairwise measurements of their ratios is known as the *group synchronization* problem. For example, the synchronization problem over the special orthogonal group  $SO(d)$  consists of estimating a set of  $n$  unknown  $d \times d$  rotation matrices  $R_1, \dots, R_n \in SO(d)$  from noisy measurements of a subset of the pairwise ratios  $R_i R_j^{-1}$

$$\text{minimize}_{R_1, \dots, R_n \in SO(d)} \sum_{\{i,j\} \in E} w_{ij} \|R_i R_j^{-1} - R_{ij}\|_F^2,$$

where  $\|\cdot\|_F$  denotes the Frobenius norm, and  $w_{ij}$  are non-negative weights denoting the confidence in the noisy pairwise measurements  $R_{ij}$ . Here,  $E$  is the set of pairs for which a ratio of group elements is available, and can be realized as the edge set of an undirected graph  $G = ([n], E)$  with vertices corresponding to the group elements.

Spectral and semidefinite programming (SDP) relaxations for solving an instance of the above synchronization problem were introduced and analyzed by Singer [34] in the context of angular synchronization, over the group  $SO(2)$  of planar rotations. Therein, one is asked to estimate  $n$  unknown angles  $\theta_1, \dots, \theta_n \in [0, 2\pi)$  given  $m$  noisy measurements  $\delta_{ij}$  of their offsets  $\theta_i - \theta_j \bmod 2\pi$ . The difficulty of the problem is amplified on one hand by the amount of noise in the offset measurements, and on the other hand by the fact that

$m \ll \binom{n}{2}$ , i.e., only a very small subset of all possible pairwise offsets are measured. In general, one may consider other groups  $\mathcal{G}$  ( $\mathbb{Z}_2$ ,  $\text{SO}(d)$ ,  $\text{O}(d)$ ) for which there are available noisy measurements  $g_{ij}$  of ratios between group elements

$$g_{ij} = g_i g_j^{-1}; \quad g_i, g_j \in \mathcal{G}; \quad \{i, j\} \in E. \quad (1.1)$$

Whenever the group  $\mathcal{G}$  is compact and has a real or complex representation, one may construct a real or Hermitian matrix (which may also be construed as a matrix of matrices) where the element in position  $\{i, j\}$  is the matrix representation of the measurement  $g_{ij}$  (which can also be a matrix of size  $1 \times 1$ , as is the case for synchronization over  $\mathbb{Z}_2$ ), or the zero matrix if there is no direct measurement for the ratio of  $g_i$  and  $g_j$ . For example, the rotation group  $\text{SO}(3)$  has a real representation using  $3 \times 3$  rotation matrices, and the group  $\text{SO}(2)$  of planar rotations has a complex representation as points on the unit circle.

**Applications** Instances of the group synchronization problem have emerged in a very wide range of applications in recent years, ranking from structural biology to computer vision and ranking systems. To name a few specific examples, synchronization over  $\text{SO}(2)$  plays an important role in the framework for ranking [17], image reconstruction from pairwise intensity differences [42, 43], and also engineering, in a specific divide-and-conquer algorithm for local-to-global sensor network localization [19], also described in detail in Section 7.

In the context of ranking, the angular synchronization paradigm was leveraged in [17] for the purpose of ranking items (resp. players) given a sparse noisy subset of pairwise preference relationships (resp. match outcomes). The approach in [17] starts by compactifying the real line by wrapping it over the upper half of the unit circle, rendering the problem amenable to standard synchronization over the compact group  $\text{SO}(2)$ . The estimated solution allowed for the recovery of the player rankings, after a post-processing step of modding out the best circular permutation. Note that the proposed approach only focused on recovering the individual rankings, and not the magnitude (i.e., strength) of each player, as it was proposed in our recent paper [22]. As detailed in Section 3, applications from the literature on ranking from heterogeneous data, provided part of the motivation for our present work. Concerning synchronization over  $\text{SO}(3)$ , specific applications include the *structure-from-motion* problem in computer vision [3], global alignment of 3D scans in computer graphics [39], structural biology for identifying the 3D structure molecules using NMR spectroscopy [20], and cryo-electron microscopy [33, 36].

**Notation** We denote vectors and matrices in lower case and upper case letters, respectively. For a finite set  $S$ , we denote  $\binom{S}{2}$  to be the collection of all subsets of  $S$  of size two.  $X \sim U[a, b]$  denotes a random variable  $X$  with uniform distribution over  $[a, b]$ . For a graph  $G = (V, E)$ , we denote its binary adjacency matrix by  $A$ , with  $A_{ij} = 1$  iff  $\{i, j\} \in E$ . For a complex-valued vector  $z$ , we denote its complex conjugate transpose by  $z^*$ . For  $M \in \mathbb{C}^{m \times n}$ , we denote its spectral norm by  $\|M\|_2$ .

**Paper outline** Section 2 gives an overview of the angular synchronization problem and the various methodologies proposed for solving it. Section 3 introduces the  $k$ -synchronization problem, its motivation and summary of main results. Section 4 focuses on the bi-synchronization problem (where  $k = 2$ ), and theoretically analyzes the performance of a spectral method under a probabilistic generative model assumption. Section 5 extends these results to the general  $k$ -synchronization setup. Section 6 details the outcomes of a variety of numerical experiments across different parameter regimes and algorithmic approaches, and also introduces an iterative graph disentangling procedure. Section 7 is an application of bi-synchronization to the setting of the graph realization problem. Finally, Section 8 is a summary of results and future research outlook.

## 2 Angular Synchronization over $\text{SO}(2)$

Let  $G = ([n], E)$  be an undirected graph and denote  $\theta_1, \dots, \theta_n \in [0, 2\pi)$  to be a collection of  $n$  unknown angles. The classical angular synchronization problem was introduced by Singer [34] and can be summarized as follows.

*Synchronization over  $SO(2)$ .*

- **Input:**  $\Theta_{ij}$  = noisy version of  $(\theta_i - \theta_j) \bmod 2\pi$ ; for each  $\{i, j\} \in E$ .
- **Goal:** Recover the unknown ground truth angles:  $\theta_1, \dots, \theta_n \in [0, 2\pi)$ .

In order to be able to recover the angles, the graph  $G$  needs to be connected. If there is no noise in the measurements, then one can easily recover the angles uniquely by fixing a root node, sequentially traversing a spanning tree in  $G$ , and summing the offsets modulo  $2\pi$ . Note that the angles will all be uniquely determined up to an additive phase given by the angle corresponding to the root node. Of course, once the measurement graph becomes disconnected, it is no longer possible to recover the angles and relate pairs of angles that belong to different disconnected components.

The problem at hand becomes more interesting and challenging in the noisy and sparse setting – one would ideally aim to have a method which is able to recover the original angles in a stable and robust manner. A sequential propagation approach that iteratively integrates the pairwise offsets across one, or potentially multiple, spanning trees is prone to fail quickly due to accumulation of errors. Instead, one would ideally put forth a method that integrates all the pairwise angle offsets in a globally consistent manner, and does so in a computationally efficient approach, while being robust to the noise and sparsity inherent in real data.

In an attempt to preserve the angle offsets as best as possible, Singer considered the following maximization problem. Consider first the  $n \times n$  measurement matrix  $H$  where for each  $\{i, j\} \in \binom{[n]}{2}$ ; we set

$$H_{ij} = \begin{cases} e^{i\Theta_{ij}} & \text{if } \{i, j\} \in E \\ 0 & \text{if } \{i, j\} \notin E \end{cases}, \quad (2.1)$$

with diagonal entries  $H_{ii} = 1$ . Note that the fact  $\Theta_{ji} = (-\Theta_{ij}) \bmod 2\pi$  renders  $H$  to be a Hermitian matrix. Next, consider the following maximization problem

$$\max_{\theta_1, \dots, \theta_n \in [0, 2\pi)} \sum_{i, j=1}^n e^{-i\theta_i} H_{ij} e^{i\theta_j}, \quad (2.2)$$

which gets incremented by  $+1$  whenever an assignment of angles  $\theta_i$  and  $\theta_j$  perfectly satisfies the given edge constraint  $\Theta_{ij} = \theta_i - \theta_j \bmod 2\pi$  (i.e., for a *good* edge), while the contribution of an incorrect assignment (i.e., of a *bad* edge) will be uniformly distributed over the unit circle in the complex plane. However, (2.2) is non-convex and computationally expensive to solve in practice.

**Spectral relaxation** Singer [34] considered solving instead

$$\max_{z_1, \dots, z_n \in \mathbb{C}; \sum_{i=1}^n |z_i|^2 = n} \sum_{i, j=1}^n z_i^* H_{ij} z_j,$$

where the individual constraints  $z_i = e^{i\theta_i}$  having unit magnitude are replaced by the much weaker single constraint  $\sum_{i=1}^n |z_i|^2 = n$ . We thus arrive at the spectral relaxation of (2.2), given by  $\max_{\|z\|_2^2 = n} z^* H z$ , which can be solved via a simple eigenvector computation, by setting  $z = v_1$ , the top eigenvector of  $H$ . Finally, we obtain the estimated rotation angles  $\hat{\theta}_1, \dots, \hat{\theta}_n$  and their corresponding elements in  $SO(2)$ ,  $\hat{r}_1, \dots, \hat{r}_n$  as

$$\hat{r}_i = e^{i\hat{\theta}_i} = \frac{v_{1,i}}{|v_{1,i}|}, \quad i = 1, 2, \dots, n. \quad (2.3)$$

Note that the estimation of the rotation angles  $\theta_1, \dots, \theta_n$  is up to an additive phase since  $e^{i\phi} v_1$  is also an eigenvector of  $H$  for any  $\phi \in \mathbb{R}$ .

**Semidefinite programming relaxation** As an alternative to the spectral relaxation, [34] also introduced a semidefinite programming (SDP) relaxation of (2.2). One can first write the objective function in (2.2) as

$$\sum_{i,j=1}^n e^{-i\theta_i} H_{ij} e^{i\theta_j} = \text{trace}(H\Upsilon),$$

where  $\Upsilon$  is a rank-1,  $n \times n$  Hermitian matrix with  $\Upsilon_{ij} = e^{i(\theta_i - \theta_j)}$ ,  $\forall i, j = 1, 2, \dots, n$ . After dropping the rank-1 constraint, all the remaining constraints are convex and altogether define the following SDP relaxation for (2.2)

$$\begin{aligned} & \underset{\Upsilon \in \mathbb{C}^{n \times n}}{\text{maximize}} && \text{trace}(H\Upsilon) \\ & \text{subject to} && \Upsilon_{ii} = 1 \quad i = 1, \dots, n \\ & && \Upsilon \succeq 0. \end{aligned} \tag{2.4}$$

This program is similar to the seminal Goemans-Williamson SDP relaxation for the well-known MAX-CUT problem of finding the maximum cut in a weighted graph. The only difference in (2.4) stems from the fact that we optimize over the cone of complex-valued Hermitian positive semidefinite matrices, not just real symmetric matrices. Note that the recovered solution is not necessarily of rank-1, and the final estimate is obtained from the best rank-1 approximation.

**Generalized power method** In recent work [11], Boumal proposed a modified version of the power method, with a provable convergence to the global optimum. This approach bears the advantage of empirically converging faster compared to the previous convex relaxations. The proposed method, dubbed as *generalized power method* (GPM), succeeds in the same noise regime as its predecessors, and enjoys the theoretical guarantees that, under a suitable noise regime, second-order necessary optimality conditions are also sufficient, despite its non-convexity. In a follow-up work [45], Zhong and Boumal established near-optimal bounds for phase synchronization, by proving that the SDP relaxation is tight for a suitable noise regime under a Spiked Gaussian Wishart model for synchronization. The same line of work provided guarantees that GPM converges to its global optimum under a similar noise regime, and also established a linear convergence rate for GPM.

**Message passing algorithms** Finally, we also point out the work of Perry et al. [32], who provide message-passing algorithms for synchronization problems over compact groups such as  $\mathbb{Z}_2$  and  $\text{SO}(2)$ , using tools from representation theory and statistical physics. Under a Gaussian noise model ensemble, the authors identify regimes where the problem is computationally easy, computationally hard, and statistically impossible, thus providing evidence of a statistical-to-computational gap [6].

**Probabilistic models for angular synchronization** Before concluding this section, we mention that the performance of the spectral and SDP relaxation-based estimators was demonstrated by Singer under the following measurement graph and noise model. Consider  $G$  to be generated via the Erdős-Rényi model, where each edge of  $G$  is present independently with probability  $\lambda \in [0, 1]$ . For each  $\{i, j\} \in E$ , the measurements  $\Theta_{ij}$  (with  $i < j$  w.l.o.g) are assumed to be generated independently through the random model

$$\Theta_{ij} = \begin{cases} (\theta_i - \theta_j) \bmod 2\pi & \text{with probability } p \\ \sim U[0, 2\pi) & \text{with probability } (1 - p) \end{cases} . \tag{2.5}$$

It was shown in [34] that the top eigenvector of  $H$  has above random correlation with the ground truth vector, when  $n$  is large enough relative to  $p$ . Additional noise models have been considered subsequently in [4, 32, 31]. For  $z \in \mathbb{C}^n$  a vector with unit modulus complex entries, and  $W \in \mathbb{C}^{n \times n}$  a Hermitian Gaussian Wigner matrix (with i.i.d. complex standard Gaussian entries above its diagonal), the matrix of pairwise measurements is modeled as

$$C = zz^* + \sigma W,$$

and the task becomes to recover the vector  $z$  given  $C$ . The Maximum Likelihood Estimator maximizes  $x^*Cx$  over the parameter space  $|x_1| = |x_2| = \dots = |x_n| = 1$ . Bandeira et al. [4] prove exact recovery under

a similar noise model for  $\mathbb{Z}_2$ , while for the above complex case, they establish that, under suitable noise regimes, the SDP relaxation admits a unique solution of rank 1, revealing the global optimum of (2.2).

### 3 $k$ -synchronization: setup, motivation, & main results

Motivated by real-world applications, in particular instances of the graph realization problem (detailed later in Section 7) and ranking from preference relationships, we now introduce a variation of the above synchronization formulation, dubbed as  $k$ -synchronization, where there exist  $k$  underlying latent sequences of angles.

**k-synchronization** Given the measurement graph  $G = ([n], E)$ , we consider the scenario where  $G$  has an **unknown decomposition** into  $k$  edge-disjoint subgraphs  $G_1 = ([n], E_1), \dots, G_k = ([n], E_k)$  such that  $E_l \cap E_{l'} = \emptyset$ , for  $l \neq l'$ , and  $\cup E_l = E$ .

*k-synchronization over  $SO(2)$ .*

- **Input:**  $\Theta_{ij} =$  noisy version of  $(\theta_{l,i} - \theta_{l,j}) \bmod 2\pi$ ; for each  $\{i, j\} \in E_l$ , for a fixed  $l \in \{1, 2, \dots, k\}$ .
- **Goal:** Recover the unknown ground truth angles:  $\theta_{l,1}, \dots, \theta_{l,n} \in [0, 2\pi)$ , for all  $l \in \{1, 2, \dots, k\}$ .

**Bi-synchronization** ( $k = 2$ ) For simplicity, we detail below the case  $k = 2$ , which we refer to as *bi-synchronization*, and illustrate it in Figure 1. For ease of notation, we shall denote by  $\alpha$  and  $\beta$  the two collections of angles. Say there exists a collection of ground truth angles  $\alpha_1, \dots, \alpha_n$  (respectively  $\beta_1, \dots, \beta_n$ ) for the synchronization problem over the graph  $G_1$  (respectively  $G_2$ ). For each  $\{i, j\} \in E$ , the user is given pairwise information  $\Theta_{ij}$  where

$$\Theta_{ij} = \begin{cases} A_{ij} = \text{noisy version of } (\alpha_i - \alpha_j) \bmod 2\pi; & \text{if } \{i, j\} \in E_1 \\ B_{ij} = \text{noisy version of } (\beta_i - \beta_j) \bmod 2\pi; & \text{if } \{i, j\} \in E_2. \end{cases}$$

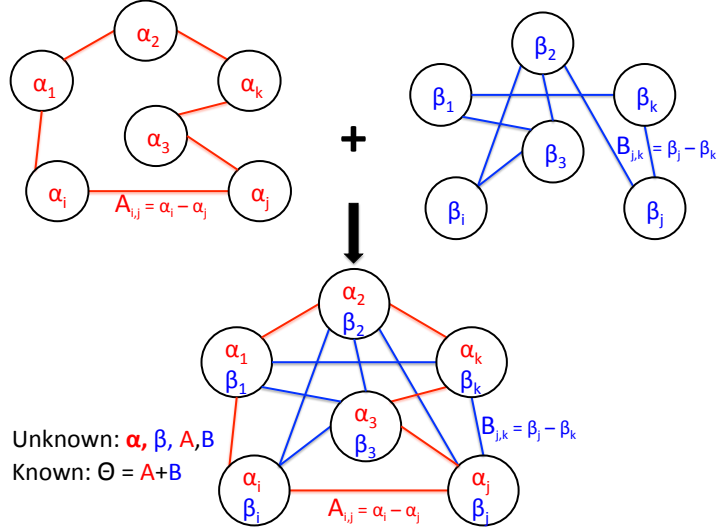
Given the combined set of noisy incomplete measurements  $\Theta_{ij}$  for  $\{i, j\} \in E$ , one is asked to recover an estimate for  $\alpha_1, \dots, \alpha_n$  and  $\beta_1, \dots, \beta_n$ . The motivation for this problem arises in structural biology, where the distance measurements between the pairwise atoms may correspond to different configurations, as is the case of molecules that may have multiple conformations. In ranking systems, the two groups of measurements associated to  $G_1$  and  $G_2$  may correspond to two different judges, whose latent rankings we seek to recover. We detail these applications in Section 3.1. The following remarks are in order.

- As for classical synchronization, we clearly require the sub-graphs  $G_1, G_2$  to be respectively connected in order to be able to recover the  $\alpha_i$ 's and  $\beta_i$ 's.
- In case  $G_1, G_2$  were known, then the problem reduces to the classical synchronization problem over  $G_1, G_2$  respectively.

#### 3.1 Applications

We refer the reader to Sections 6 and 7 for an application to the graph realization problem, and an algorithmic procedure for extracting the individual measurement subgraphs. In this section, we elaborate on several applications that have motivated this work.

**Graph realization problem and alignment of patch embeddings** In Section 7 we consider an application of  $k$ -synchronization to the popular *graph realization problem*, where one is asked to recover a cloud of points in  $\mathbb{R}^D$  from a small subset of noisy pairwise Euclidean distances. In applications of this problem to sensor network localization and structural biology, one may have readily available information



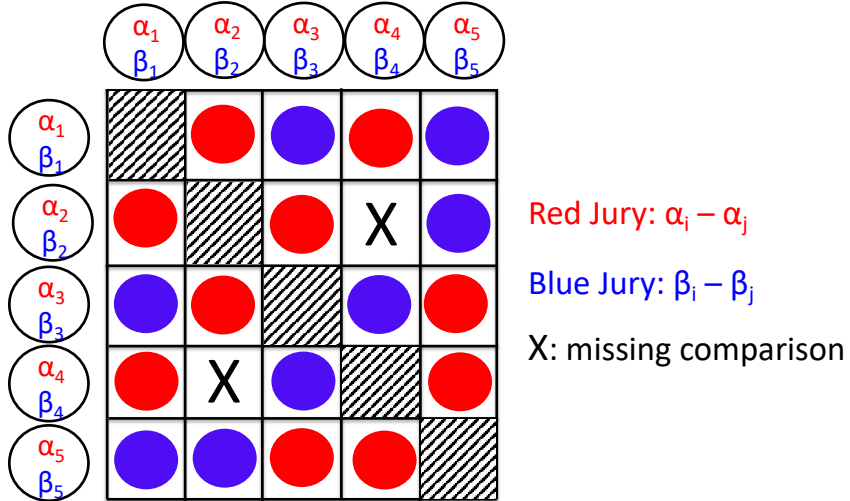
**Figure 1:** An instance of the bi-synchronization problem. The measurement graph (bottom row) available to the user is a union of two edge disjoint graphs (top row) on the same vertex set. The top graphs each capture angle offsets for two different sets of angles: the top left graph encodes noise pairwise offsets between the angles  $\alpha_1, \dots, \alpha_n$ , while the top right graph contains similar information from the angles  $\beta_1, \dots, \beta_n$ .

on the embedding of a subset of nodes, and is left with the task of integrating the given embeddings into a globally consistent structure. We consider the scenario where the point cloud to be recovered may assume one of two possible configurations (denoted as type- $X$  and type- $Y$ ), and thus whenever a pair of subgraph embeddings are aligned (eg, via Procrustes analysis), the resulting alignment measurement may come from either configurations (i.e., the two subgraphs to be aligned are either both of type- $X$  or both of type- $Y$ ). We refer the reader to Section 7 for additional details and results on our application in this setting.

**Extraction of latent rankings in multiple voters systems** Yet another very closely related application arises in the ranking and recommendation systems literature. In many real-world scenarios, it is often the case that there exist multiple rating systems (eg, judges or voters) who provide incomplete and inconsistent rankings or pairwise comparisons between the same set of players or items. For example, in tournaments of  $n$  players, where everyone plays against everybody else in a total of  $\binom{n}{2}$  matches, it may well be the case that, due to the large number of matches, a set of  $k$  judges are distributed matches they are to referee. In other words, the set of all matches are randomly partitioned, and each set in the partition is attributed to a single judge, as shown pictorially in the example in Figure 2 for  $k = 2$ . A similar setting can arise in applications well beyond sports and related competitions, including rating items in terms of preference relations between pairs of items. Such examples arise in many learning and social data analysis problems, including recommendation systems and collaborative-filtering for ranking movies for a user based on the movie rankings provided by other users, and information retrieval where one is interested in combining the results of different search engines.

The problem now becomes to disentangle the intrinsic ranking of all players that corresponds to each judge. When a judge is asked to evaluate a pair of players or items, she or he may base their decision on a specific latent feature which the said judge construes as most relevant for the ranking task. Different judges may use different intrinsic features on which they base their decision. The problem at hand now becomes to uncover the ranking provided by each judge. In light of the pipeline proposed in [17] by a subset of the authors, the  $k$ -synchronization problem and results covered in the present paper allow for the extraction of such latent rankings, given a single measurement matrix of pairwise ranking preferences.

**Cluster-synchronization** Finally, we mention a closely related problem, coming from the ranking literature, of extracting partial rankings from pairwise comparison data. Similar to the setup pursued in this paper, there exist two collections of unknown angles grouped into two clusters  $\mathcal{C}_1 = (\alpha_1, \alpha_2, \dots, \alpha_{n_1})$ , and  $\mathcal{C}_2 = (\beta_{n_1+1}, \beta_{n_1+2}, \dots, \beta_{n_1+n_2})$  such that whenever a pair of angles from the same cluster is selected, the user



**Figure 2:** An instance of the **bi-synchronization** problem in the context of ranking from pairwise comparisons with multiple voting systems. The red and blue judges each give a rating for pair of items  $\{i, j\}$ .  $(\alpha_1, \dots, \alpha_n)$  denote the intrinsic strength/skill of the players as construed by the red judge, while  $(\beta_1, \dots, \beta_n)$  capture the strength/skill of the same set of  $n$  players as perceived by the blue judge.

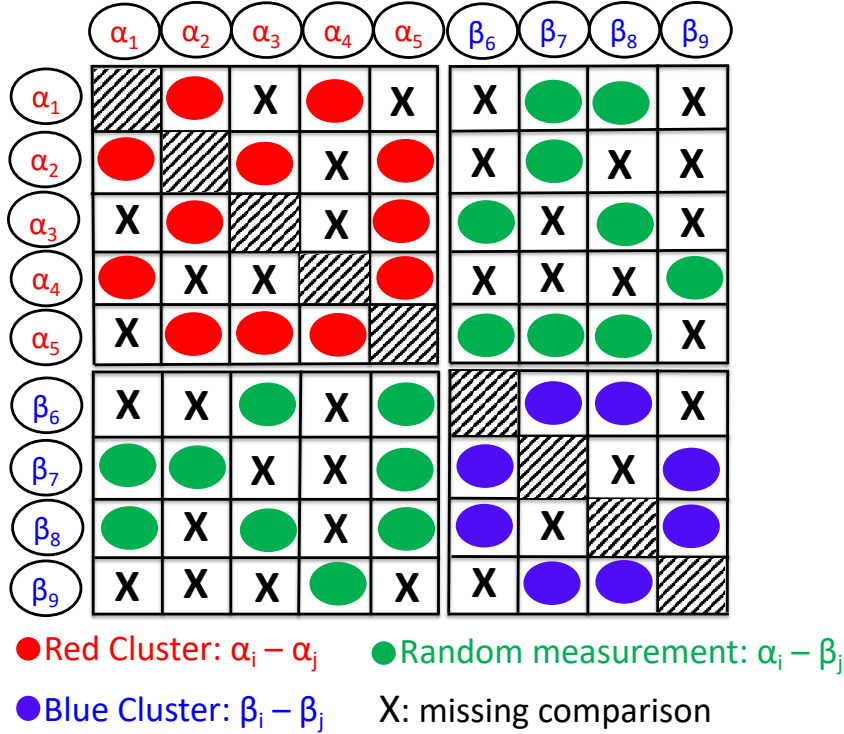
has available a noisy proxy for the angle offset  $\alpha_i - \alpha_j$  or  $\beta_i - \beta_j$ . However, when a pair is selected such that one angle is from  $\mathcal{C}_1$  and the second angle from cluster  $\mathcal{C}_2$ , then the available measurement  $\Theta_{ij}$  is completely random, uniformly distributed in  $[0, 2\pi)$  (in particular,  $\Theta_{ij}$  is **not** a (potentially noisy) proxy of  $\alpha_i - \beta_j$ ). This setup is illustrated in Figure 3, highlighting the cluster structure of the measurement matrix. As yet another instance of this problem, imagine for example two jigsaw puzzles being mixed together, and the goal is to disentangle the two sets of pieces. Or consider an archaeological site where two frescoes are identified, each of which is broken into many pieces which are now mixed together. We expect similar techniques as those studied in the present paper to be applicable in this setting as well, motivated by the fact that the expected measurement matrix has a low-rank structure.

Finally, we stress there is a fundamental difference between the setup in the bi-synchronization and cluster-synchronization problems. In both problems, there exist two groups of angles (of  $\alpha$ -type and  $\beta$ -type); however, in bi-synchronization the two groups must be of the same size  $n_1 = n_2$  and a measurement is either a noisy version between two  $\alpha$ -type angles or two  $\beta$ -type angles), while in cluster-synchronization we allow  $n_1 \neq n_2$  and there also exist measurements across different types of angles. Figures 2 and 3 illustrate pictorially the structural differences between these two problems.

### 3.2 Summary of our main contributions

1. We introduce the  $k$ -synchronization problem and also propose a probabilistic generative model for it. This model is detailed in Section 4 and Section 5.
2. We propose a spectral algorithm (Algorithm 1) which, under a certain approximate orthogonality assumption on the angles, amounts to finding the top  $k$  eigenvectors of the measurement matrix  $H$  given in (2.1). Furthermore, we also provide a detailed theoretical analysis of this algorithm for estimating the unknown angles in each group, under the aforementioned generative model. These results are stated as Theorem 1 for  $k = 2$  and Theorem 2 for the general case.
3. We provide a comprehensive set of numerical experiments in Section 6 that illustrate the performance of our proposed spectral algorithm, under various parameter regimes (noise level, graph sparsity, and number of groups of angles  $k$ ), and also compare it with the corresponding SDP relaxation, and also a normalized spectral version.
4. We consider an application of bi-synchronization to the graph realization problem in Section 7. Along the way, we also provide an iterative graph disentangling procedure, which can be of independent inter-





**Figure 3:** An instance of the **cluster-synchronization** problem, closely related to the *bi-synchronization* problem we study in this paper. The collection of angles is partitioned into two clusters  $\mathcal{C}_1 = (\alpha_1, \alpha_2, \dots, \alpha_{n_1})$ ,  $\mathcal{C}_2 = (\beta_{n_1+1}, \beta_{n_1+2}, \dots, \beta_{n_2})$ , of potentially unequal sizes  $n_1, n_2$  (as is the case in this example, with  $n_1 = 5, n_2 = 4$ ), such that when a pair of angles is chosen from the same cluster, we obtain a (potentially noisy) offset  $\alpha_i - \alpha_j$  or  $\beta_i - \beta_j$ . However, when a pair  $(\alpha_i, \beta_j)$  of angles from two different clusters is compared, the measured offset  $\Theta_{ij}$  is completely random and contains no meaningful information.

est, and provide numerical experiments demonstrating that the proposed iterative scheme consistently improves the final recovery accuracy.

## 4 Bi-synchronization: Synchronization with two groups of angles

We begin by analyzing the case where  $k = 2$  since it is relatively easier as compared to the general case. Recall that we are interested in recovering the two sequences of angles  $\alpha_1, \dots, \alpha_n \in [0, 2\pi)$  and  $\beta_1, \dots, \beta_n \in [0, 2\pi)$ . Denoting  $z_1, z_2 \in \mathbb{C}^n$  to be the entry-wise representations of  $(\alpha_i)_i$  and  $(\beta_i)_i$  on the unit circle where

$$z_{1,i} = \frac{1}{\sqrt{n}} e^{i\alpha_i} \quad \text{and} \quad z_{2,i} = \frac{1}{\sqrt{n}} e^{i\beta_i}; \quad i = 1, \dots, n, \quad (4.1)$$

we will assume that the vectors  $z_1, z_2 \in \mathbb{C}^n$  defined in (4.1) are  $\delta$ -orthogonal for  $\delta \in (0, 1)$ .

**Definition 1.** A collection of vectors  $x_1, x_2, \dots, x_m \in \mathbb{C}^n$  are said to be  $\delta$ -orthogonal for  $\delta \in [0, 1]$  if

$$|\langle x_i, x_j \rangle| \leq \delta \|x_i\|_2 \|x_j\|_2, \quad \forall i \neq j.$$

This is the case, for instance, when the angles  $\alpha_i, \beta_i$  are generated uniformly at random in  $[0, 2\pi)$ , and are i.i.d. Indeed, one can then show using standard large deviation inequalities that the vectors  $z_1, z_2 \in \mathbb{C}^n$  in (4.1) will be approximately orthogonal with high probability. This is stated formally in the form of the following Proposition.

**Proposition 1.** If  $\alpha_i, \beta_i \sim U[0, 2\pi)$  i.i.d for  $i = 1, \dots, n$ , it then follows for any given  $\delta \in (0, 1)$  that  $|\langle z_1, z_2 \rangle| \leq \delta$  holds with probability at least  $1 - 2e \exp(-c\delta^2 n)$ . Here  $c > 0$  is a constant.

Its proof is deferred to Appendix C.1.

## 4.1 Spectral method

Our aim is to recover the  $\alpha_i$ 's and  $\beta_i$ 's in such a way that we preserve as best as possible the available input pairwise measurement. To this end, with the matrix  $H$  as defined in (2.1), one may consider the following objective function

$$\max_{\substack{\alpha_1, \dots, \alpha_n \in [0, 2\pi) \\ \beta_1, \dots, \beta_n \in [0, 2\pi)}} \sum_{\{i, j\} \in E_1} e^{-\iota \alpha_i} H_{ij} e^{\iota \alpha_j} + \sum_{\{i, j\} \in E_2} e^{-\iota \beta_i} H_{ij} e^{\iota \beta_j}. \quad (4.2)$$

Since the optimization problem (4.2) is non-convex and computationally expensive to solve in practice, we consider the following spectral relaxation, while keeping the definition (4.1) in mind

$$\max_{\|w_1\|_2^2 = \|w_2\|_2^2 = 1} \sum_{\{i, j\} \in E_1} w_{1,i}^* H_{ij} w_{1,j} + \sum_{\{i, j\} \in E_2} w_{2,i}^* H_{ij} w_{2,j}.$$

Since we do not know the decomposition of the edge set into  $E_1$  and  $E_2$ , and the ground truths  $z_1, z_2$  are approximately orthogonal by assumption, we consider

$$\max_{\substack{\|w_1\|_2^2 = \|w_2\|_2^2 = 1 \\ w_1 \perp w_2}} \sum_{i, j=1}^n (w_{1,i}^* H_{ij} w_{1,j} + w_{2,i}^* H_{ij} w_{2,j}).$$

Altogether, the spectral relaxation of the non-convex optimization problem in (4.2) is given by

$$\max_{\substack{\|w_1\|_2^2 = \|w_2\|_2^2 = 1 \\ w_1 \perp w_2}} w_1^* H w_1 + w_2^* H w_2, \quad (4.3)$$

whose solution is obtained by setting  $w_1 = v_1$  and  $w_2 = v_2$ , the top two eigenvectors of  $H$ . Finally, we can then obtain the estimates  $\hat{\alpha}_i, \hat{\beta}_i$  for  $i = 1, \dots, n$  as

$$e^{\iota \hat{\alpha}_i} = \frac{v_{1,i}}{|v_{1,i}|}, \quad e^{\iota \hat{\beta}_i} = \frac{v_{2,i}}{|v_{2,i}|}.$$

**Remark 1.** *The solution of (4.3) is clearly unchanged if we replace  $H$  with  $H - cI$  for some scalar  $c$ . Therefore one can set the entries  $H_{ii}$  to the same (arbitrarily chosen) real value.*

## 4.2 Erdős-Rényi measurement model

We are interested in the following random measurement model.

- The graph  $G$  is generated via the Erdős-Rényi random graph model  $G(n, \lambda)$ , where each edge of  $G$  is present with probability  $\lambda \in [0, 1]$ .
- For each  $\{i, j\} \in E$ , we assign it to  $E_1$  with prob.  $q$  and to  $E_2$  with prob.  $1 - q$ . So the graphs  $G_1$  and  $G_2$  are clearly edge disjoint, i.e.,  $E_1 \cap E_2 = \emptyset$ , and moreover,  $E_1 \cup E_2 = E$ .
- Consider  $\{i, j\} \in E$ , and assume  $i < j$  w.l.o.g. Let  $N_{ij} \sim U[0, 2\pi)$  i.i.d. If  $\{i, j\} \in E_1$ , we obtain  $\Theta_{ij} = (\alpha_i - \alpha_j) \bmod 2\pi$  with probability  $\tilde{q}_1$ , and  $\Theta_{ij} = N_{ij}$  with probability  $1 - \tilde{q}_1$ , i.i.d. If  $\{i, j\} \in E_2$ , we obtain  $\Theta_{ij} = (\beta_i - \beta_j) \bmod 2\pi$  with probability  $\tilde{q}_2$ , and  $\Theta_{ij} = N_{ij}$  with probability  $1 - \tilde{q}_2$ , i.i.d.

Let us denote  $p_1 = q\tilde{q}_1$  and  $p_2 = (1 - q)\tilde{q}_2$ . For each  $\{i, j\} \in E$ , with  $i < j$ , we can write the obtained measurement  $\Theta_{ij}$  as

$$\Theta_{ij} = \begin{cases} (\alpha_i - \alpha_j) \bmod 2\pi & \text{w.p } p_1, \\ (\beta_i - \beta_j) \bmod 2\pi & \text{w.p } p_2, \\ N_{ij} \sim U[0, 2\pi) & \text{w.p } \eta = 1 - p_1 - p_2, \end{cases} \quad (4.4)$$

where  $p_1$  (respectively  $p_2$ ) denotes the probability of a correct pairwise measurement in  $G_1$  (respectively  $G_2$ ), and  $1 - p_1 - p_2$  is the probability of getting an incorrect measurement. Moreover, we have  $\Theta_{ji} = (-\Theta_{ij}) \bmod 2\pi$ .

The difficulty of the problem is amplified on one hand by the noise and sparsity of the measurements (dictated by  $\eta$ ,  $\lambda$  respectively), and on the other hand by the ambiguity that stems from the fact that the user does not know a-priori whether a given edge (i.e., the measurement that comes with it) corresponds to a (perhaps noisy) offset  $\alpha_i - \alpha_j$  or an offset  $\beta_i - \beta_j$ . In other words, the graphs  $G_1$  and  $G_2$ , in particular their respective edge sets  $E_1$  and  $E_2$ , are not known to the user a-priori. From now on, we will assume w.l.o.g.<sup>1</sup> that  $p_1 > p_2$ .

We start by mapping the points over the complex unit circle as in (2.1) to construct a Hermitian matrix  $H$ . Due to (4.4), we have for each  $i < j$  that

$$H_{ij} = \begin{cases} e^{i(\alpha_i - \alpha_j)} & \text{w.p. } p_1\lambda, \\ e^{i(\beta_i - \beta_j)} & \text{w.p. } p_2\lambda, \\ e^{iN_{ij}}, N_{ij} \sim U[0, 2\pi) & \text{w.p. } (1 - p_1 - p_2)\lambda, \\ 0 & \text{w.p. } 1 - \lambda. \end{cases} \quad (4.5)$$

In light of Remark 1, we set  $H_{ii} = \lambda(p_1 + p_2)$  for convenience. Note that  $H_{ji} = H_{ij}^*$ , hence we can write  $\mathbb{E}[H]$  as

$$\mathbb{E}[H] = np_1\lambda z_1 z_1^* + np_2\lambda z_2 z_2^*. \quad (4.6)$$

$\mathbb{E}H$  is of rank at most 2, unless  $z_1$  is a scalar multiple of  $z_2$ . Given the above, we can write the available measurement matrix  $H$  as

$$H = \mathbb{E}(H) + R = n\lambda(p_1 z_1 z_1^* + p_2 z_2 z_2^*) + R, \quad (4.7)$$

where  $R$  is a random Hermitian matrix with  $R_{ii} = 0$ , whose elements  $R_{ij} = H_{ij} - \mathbb{E}[H_{ij}]$  are zero-mean independent random variables for  $i \leq j$ . In particular, for each  $i < j$ , we have

$$R_{ij} = \begin{cases} e^{i(\alpha_i - \alpha_j)} - p_1\lambda e^{i(\alpha_i - \alpha_j)} - p_2\lambda e^{i(\beta_i - \beta_j)} & \text{w.p. } p_1\lambda, \\ e^{i(\beta_i - \beta_j)} - p_1\lambda e^{i(\alpha_i - \alpha_j)} - p_2\lambda e^{i(\beta_i - \beta_j)} & \text{w.p. } p_2\lambda, \\ e^{iN_{ij}} - p_1\lambda e^{i(\alpha_i - \alpha_j)} - p_2\lambda e^{i(\beta_i - \beta_j)}, N_{ij} \sim U[0, 2\pi) & \text{w.p. } (1 - p_1 - p_2)\lambda, \\ -p_1\lambda e^{i(\alpha_i - \alpha_j)} - p_2\lambda e^{i(\beta_i - \beta_j)} & \text{w.p. } 1 - \lambda, \end{cases} \quad (4.8)$$

and  $R_{ji} = R_{ij}^*$ .

### 4.3 Main result

Our main theorem for the bi-synchronization problem for the case of approximately orthogonal representations  $z_1, z_2$  is outlined below.

**Theorem 1.** *Let  $z_1, z_2$  in (4.1) be  $\delta$ -orthogonal for  $\delta \in (0, 1)$ , and let  $p_1 > p_2$ . For constants  $\varepsilon \in (0, 1)$  and  $\mu \in [0, 1/2]$ , let*

$$n \geq \frac{72(2 + \varepsilon)^2 C(\lambda, p_1, p_2)}{\mu^2 \lambda^2 \min\{(p_1 - p_2)^2, p_2^2\}},$$

where  $C(\lambda, p_1, p_2) > 0$  depends only on  $\lambda, p_1, p_2$  (see (4.14)). Then with probability at least  $1 - 3n \exp\left(-\frac{8C(\lambda, p_1, p_2)n}{\bar{\sigma}(\lambda, p_1, p_2)^2 c_\varepsilon}\right)$ , the following bounds hold.

$$\begin{aligned} |z_1, v_1|^2 &\geq 1 - \left( \sqrt{\frac{2\mu}{1-\mu}} + \sqrt{1 - \frac{p_1 - p_2}{\sqrt{(p_1 - p_2)^2 + 4p_1 p_2 \delta^2}}} \right)^2, \\ |z_2, v_2|^2 &\geq 1 - \left( \sqrt{\frac{2\mu}{1-\mu}} + \sqrt{1 - \frac{p_1(1-\delta^2) - p_2}{\sqrt{(p_1 - p_2)^2 + 4p_1 p_2 \delta^2}}} \right)^2. \end{aligned}$$

Here,  $c_\varepsilon, \bar{\sigma}(\lambda, p_1, p_2) > 0$  only depend on the indicated parameters (see 4.15).

The following remarks are in order regarding Theorem 1.

<sup>1</sup>Clearly, some separation is needed in order to be able to distinguish the graphs.

1. The above correlation bounds are essentially composed of two terms – one arising due to the perturbation  $R$ , and the other due to the approximate orthogonality of  $z_1, z_2$ . This second term vanishes in the special case  $\delta = 0$ .
2. In order to understand the scaling of the terms involved, it is convenient to replace the terms  $C(\lambda, p_1, p_2), \bar{\sigma}(\lambda, p_1, p_2)$  by their upper bounds  $C(\lambda, p_1, p_2) \lesssim \lambda$  and  $\bar{\sigma}(\lambda, p_1, p_2) \lesssim 1$ . We then have that if

$$\lambda \gtrsim \frac{\log n}{n}, \quad n \gtrsim \frac{1}{\mu^2 \lambda \min\{(p_1 - p_2)^2, p_2^2\}},$$

then the stated correlation bounds hold with probability at least  $1 - \frac{1}{n}$ .

#### 4.4 Proof of Theorem 1

Let us denote  $\tilde{\lambda}_1 \geq \tilde{\lambda}_2$  (resp.  $\tilde{v}_1, \tilde{v}_2$ ) to be the two largest eigenvalues (resp. corresponding eigenvectors) of  $\mathbb{E}[H]$ . Since  $z_1, z_2$  are approximately orthogonal and  $p_1 \neq p_2$ , we would expect, for small enough  $\delta$ , that  $\tilde{\lambda}_1 \approx np_1\lambda$ ,  $\tilde{\lambda}_2 \approx np_2\lambda$  and  $|\langle \tilde{v}_1, z_1 \rangle| \approx 1$ ,  $|\langle \tilde{v}_2, z_2 \rangle| \approx 1$  hold. This is stated precisely in the following Lemma.

**Lemma 1.** *It holds that  $\tilde{\lambda}_2 \leq np_2\lambda$  and  $\tilde{\lambda}_1 \geq np_1\lambda$ . Moreover, for any given  $\delta \in (0, 1)$ , if  $z_1, z_2 \in \mathbb{C}^n$  as defined in (4.1) are  $\delta$ -orthogonal, then the following bounds hold.*

$$\tilde{\lambda}_2 \geq \frac{(np_1\lambda + np_2\lambda) - \sqrt{(np_1\lambda - np_2\lambda)^2 + 4n^2p_1p_2\lambda^2\delta^2}}{2}, \quad (4.9)$$

$$\tilde{\lambda}_1 \leq \frac{(np_1\lambda + np_2\lambda) + \sqrt{(np_1\lambda - np_2\lambda)^2 + 4n^2p_1p_2\lambda^2\delta^2}}{2}, \quad (4.10)$$

$$|\langle z_1, \tilde{v}_1 \rangle|^2 \geq \frac{p_1 - p_2}{\sqrt{(p_1 - p_2)^2 + 4p_1p_2\delta^2}}, \quad |\langle z_2, \tilde{v}_2 \rangle|^2 \geq \frac{p_1(1 - \delta^2) - p_2}{\sqrt{(p_1 - p_2)^2 + 4p_1p_2\delta^2}}. \quad (4.11)$$

*Proof.* See Appendix C.2. □

Let us now denote  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  (resp.  $v_1, v_2, \dots, v_n$ ) to be the eigenvalues (resp. eigenvectors) of  $H$ . As a next step, we would like to quantify the deviation of  $v_1, v_2$  from  $\tilde{v}_1, \tilde{v}_2$  respectively; clearly this is dictated solely by the perturbation matrix  $R$  arising due to noise. The following Lemma states this precisely. Note that the statement of the Lemma is deterministic – it is conditioned on the event that  $\|R\|_2 \leq \Delta$  for a suitably small value of  $\Delta$ .

**Lemma 2.** *If  $\|R\|_2 \leq \Delta$  holds for  $\Delta < \min\{n(p_1 - p_2)\lambda, np_2\lambda\}$ , we then have that*

$$|\langle v_1, \tilde{v}_1 \rangle|^2 \geq 1 - \left( \frac{\Delta}{np_1\lambda - np_2\lambda - \Delta} \right)^2, \quad (4.12)$$

$$|\langle v_2, \tilde{v}_2 \rangle|^2 \geq 1 - \left( \frac{\Delta}{\min\{np_1\lambda - np_2\lambda - \Delta, np_2\lambda - \Delta\}} \right)^2. \quad (4.13)$$

*Proof.* See Appendix C.3. □

We now need to bound the spectral norm of the perturbation matrix  $R$ . The following lemma shows that  $\|R\|_2 \lesssim \sqrt{n}$  holds with high probability. Before proceeding, let us write  $R = \text{Re}(R) + i\text{Im}(R)$  where  $\text{Re}(R), \text{Im}(R) \in \mathbb{R}^{n \times n}$  are respectively symmetric and skew symmetric matrices.

**Lemma 3.** *Denote*

$$\begin{aligned} C(\lambda, p_1, p_2) = & 2p_1\lambda[(1 - p_1\lambda)^2 + (p_2\lambda)^2] + 2p_2\lambda[(1 - p_2\lambda)^2 + (p_1\lambda)^2] \\ & + (1 - p_1 - p_2)\lambda\left[\frac{1}{2} + (p_1\lambda + p_2\lambda)^2\right] + (1 - \lambda)(p_1\lambda + p_2\lambda)^2, \end{aligned} \quad (4.14)$$

and let  $\bar{\sigma}(\lambda, p_1, p_2) > 0$  be such that

$$\max_{i,j} \{\| \text{Re}(R)_{ij} \|_\infty, \| \text{Im}(R)_{ij} \|_\infty\} \leq \bar{\sigma}(\lambda, p_1, p_2). \quad (4.15)$$

Then for any  $\varepsilon \geq 0$ , there exists a universal constant  $c_\varepsilon > 0$  such that

$$\|R\|_2 \leq (2 + \varepsilon)6\sqrt{2C(\lambda, p_1, p_2)n}, \quad (4.16)$$

with probability at least  $1 - 3n \exp\left(-\frac{8C(\lambda, p_1, p_2)n}{\bar{\sigma}(\lambda, p_1, p_2)^2 c_\varepsilon}\right)$ .

*Proof.* See Appendix C.4. □

We now plug in  $\Delta = (2 + \varepsilon)6\sqrt{2C(\lambda, p_1, p_2)n}$  in Lemma 2. For a fixed  $\mu \in [0, 1/2]$ , one can readily verify that

$$n \geq \frac{72(2 + \varepsilon)^2 C(\lambda, p_1, p_2)}{\mu^2 \lambda^2 \min\{(p_1 - p_2)^2, p_2^2\}} \iff \Delta \leq \mu \min\{n\lambda(p_1 - p_2), n\lambda p_2\}. \quad (4.17)$$

Moreover, provided (4.17) holds, the bounds on  $|\langle v_1, \tilde{v}_1 \rangle|^2, |\langle v_2, \tilde{v}_2 \rangle|^2$  in Lemma 2 change to

$$|\langle v_i, \tilde{v}_i \rangle|^2 \geq 1 - \left(\frac{\mu}{1 - \mu}\right)^2; \quad i = 1, 2. \quad (4.18)$$

The statement of Theorem 1 now follows via the following Proposition which also completes the proof.

**Proposition 2.** Consider  $x, y, \bar{x} \in \mathbb{C}^n$  of unit  $\ell_2$  norm. Let  $\varepsilon, \bar{\varepsilon} \in [0, 1]$ , and let

$$|\langle x, y \rangle|^2 \geq 1 - \varepsilon, \quad |\langle \bar{x}, y \rangle|^2 \geq 1 - \bar{\varepsilon}.$$

It then follows that

$$|\langle x, \bar{x} \rangle|^2 \geq 1 - (\sqrt{\varepsilon} + \sqrt{2\bar{\varepsilon}})^2. \quad (4.19)$$

*Proof.* See Appendix C.5. □

## 5 $k$ -synchronization: Synchronization with $k$ groups of angles

We now consider the general  $k$ -synchronization problem involving  $k$  groups of angles, namely  $\theta_{l,1}, \dots, \theta_{l,n}$  for  $l = 1, \dots, k$ . Let  $z_l \in \mathbb{C}^n$  denote the entry-wise representations of  $(\theta_{l,i})_i$  on the unit circle where

$$z_{l,i} = \frac{1}{\sqrt{n}} e^{i\theta_{l,i}}; \quad i = 1, \dots, n. \quad (5.1)$$

As before, we will assume that the representations  $z_l \in \mathbb{C}^n$  in (5.1) are  $\delta$ -orthogonal. If  $k$  is small relative to  $n$ , then this is the case when the angles  $\theta_{l,i}$ ,  $1 \leq l \leq k$ ,  $1 \leq i \leq n$  are generated uniformly at random in  $[0, 2\pi)$  and are i.i.d. Indeed, using Proposition 1 and applying the union bound, it follows that  $z_1, \dots, z_k$  are  $\delta$ -orthogonal with probability at least

$$1 - 2 \binom{k}{2} e^{-c\delta^2 n} \geq 1 - e^{2 \log k} e^{-c\delta^2 n} \geq 1 - e^{-c\delta^2 n/2},$$

if  $n \geq \frac{4 \log k}{c\delta^2}$ .

For the general  $k$ -way synchronization, the objective function analogous to (4.3) is given by

$$\max_{\substack{W \in \mathbb{C}^{n \times k} \\ W^* W = I_{k \times k}}} W^* H W, \quad (5.2)$$

which is maximized when the columns of  $W$  are taken to be the top  $k$  eigenvectors of  $H$ , namely  $v_1, \dots, v_k$ . The estimates  $\hat{\theta}_{l,i}$  are then recovered by normalizing the entries of  $v_{l,i}$ ; the complete recovery procedure is outlined in Algorithm 1. Note that shifting the diagonal entries of  $H$  by a scaled identity matrix does not change the solution of (5.2), thus for convenience, we can set  $H_{ii} = c$  for any arbitrarily chosen  $c \in \mathbb{R}$ .

---

**Algorithm 1** Algorithm for the k-Synchronization problem
 

---

- 1: **Input:** Graph  $G = (V, E)$  and measurements  $\Theta_{ij}$  for  $(i, j) \in E$ .
- 2: Construct the Hermitian matrix  $H \in \mathbb{C}^{n \times n}$  with entries given by ??.
- 3: Compute the top  $k$  eigenvectors  $v_1, v_2, \dots, v_k$  of  $H$ .
- 4: For each  $l = 1, \dots, k$ , obtain estimates  $\hat{\theta}_{l,i}$  for  $i = 1, \dots, n$  where

$$e^{i\hat{\theta}_{l,i}} = \frac{v_l(i)}{|v_l(i)|}. \quad (5.3)$$


---

## 5.1 Erdős-Rényi measurement model

Analogous to the  $k = 2$  case, we consider the graph  $G = (V, E)$  to be generated via the Erdős-Rényi random graph model  $G(n, \lambda)$ , where each edge of  $G$  is present with probability  $\lambda \in [0, 1]$ . Given  $G$ , we then obtain for each  $\{i, j\} \in E$  ( $i < j$  w.l.o.g) the value  $\Theta_{i,j}$  as

$$\Theta_{ij} = \begin{cases} (\theta_{l,i} - \theta_{l,j}) \pmod{2\pi}; & \text{w.p. } p_l \text{ for } l = 1, \dots, k \\ N_{ij} \sim U[0, 2\pi); & \text{w.p. } \eta = 1 - \sum_{l=1}^k p_l, \end{cases} \quad (5.4)$$

where  $p_l$  denotes the probability of a correct pairwise measurement in  $G_l$ , and  $\eta$  is the probability of getting an incorrect measurement. We will assume w.l.o.g that  $p_1 > p_2 > \dots > p_k$ .

We map the points over the complex unit circle as in (2.1) to construct a Hermitian matrix  $H$ . Then due to (5.4), we have for each  $i < j$  that

$$H_{ij} = \begin{cases} e^{i(\theta_{l,i} - \theta_{l,j})} & \text{with probability } p_l \lambda \text{ for } l = 1, \dots, k \\ e^{iN_{ij}}, N_{ij} \sim U[0, 2\pi) & \text{with probability } (1 - \sum_{l=1}^k p_l) \lambda \\ 0 & \text{with probability } 1 - \lambda. \end{cases}$$

Moreover, we set each diagonal entry  $H_{ii} = \lambda(p_1 + p_2 + \dots + p_k)$  for convenience. We can then write  $\mathbb{E}[H]$  and  $H$  as

$$\begin{aligned} \mathbb{E}[H] &= np_1 \lambda z_1 z_1^* + np_2 \lambda z_2 z_2^* + \dots + np_k \lambda z_k z_k^*, \\ H &= \mathbb{E}(H) + R = n\lambda(p_1 z_1 z_1^* + \dots + p_k z_k z_k^*) + R, \end{aligned}$$

where  $R$  is a random Hermitian matrix with  $R_{ii} = 0$ , and whose elements  $R_{ij} = H_{ij} - \mathbb{E}[H_{ij}]$  are zero-mean independent random variables for  $i \leq j$ . In particular, for each  $i < j$ , we have

$$R_{ij} = \begin{cases} e^{i(\theta_{l,i} - \theta_{l,j})} - \sum_{l'=1}^k p_{l'} \lambda e^{i(\theta_{l',i} - \theta_{l',j})} & \text{w.p. } p_l \lambda, \quad l = 1, \dots, k \\ e^{iN_{ij}} - \sum_{l'=1}^k p_{l'} \lambda e^{i(\theta_{l',i} - \theta_{l',j})}, N_{ij} \sim U[0, 2\pi) & \text{w.p. } (1 - \sum_{l'=1}^k p_{l'}) \lambda \\ - \sum_{l'=1}^k p_{l'} \lambda e^{i(\theta_{l',i} - \theta_{l',j})} & \text{w.p. } 1 - \lambda, \end{cases} \quad (5.5)$$

and  $R_{ji} = R_{ij}^*$ . For  $\delta$  sufficiently small, clearly the matrix  $\mathbb{E}[H]$  will be rank  $k$ . An easy application of Gerschgorin's disk theorem reveals that the choice  $\delta < 1/(k-1)$  suffices.

## 5.2 Main result

Before stating our main result, it will be helpful to define additional notation. For  $1 \leq m \leq k$ , denote  $S_m = \sum_{j=1}^m p_j$ . For any  $\delta \in [0, 1]$ , let us define the numbers  $\psi_1(\delta), \dots, \psi_k(\delta)$  as follows.

1.  $\psi_1(\delta) := \frac{p_2(k-1)}{p_1 - p_2} \delta$ .
2.  $\psi_j(\delta) := C_j \sqrt{\psi_{j-1}(\delta)}$  for  $2 \leq j \leq k$  where

$$C_j = \frac{p_{j+1}(k-1)\sqrt{2} + 4\sqrt{2}[2S_{j-1} + \sqrt{2}(j-1)(j-2)(p_1 - p_j) + \frac{(j-1)}{2}(p_2(k-1) - 2p_{j+1})]}{p_j - p_{j+1}}$$

with  $p_{k+1} = 0$ .

Moreover, with the convention  $S_0 = 0$ , let us also define

$$E_j := 4\sqrt{2}S_{j-2} + 8(j-2)(j-3)(p_1 - p_{j-1}) + 4\sqrt{2}p_2(k-1)(j-2) + 4(j-1)(p_1 - p_{j+1}) + p_{j+1}(k-1); \quad 2 \leq j \leq k, \quad (5.6)$$

$$\tilde{E} := 4\sqrt{2}S_{k-1} + 8(k-1)(k-2)(p_1 - p_k) + 4\sqrt{2}p_2(k-1)^2. \quad (5.7)$$

Our main theorem for the  $k$ -synchronization problem for the case of approximately orthogonal representations  $z_1, z_2, \dots, z_k$  is outlined below.

**Theorem 2.** *Let  $z_1, \dots, z_k$  in (5.1) be  $\delta$ -orthogonal for  $\delta \in (0, 1)$ . Assuming  $p_1 > p_2 > \dots > p_k$ , under the notation defined previously, suppose that the following conditions are satisfied for a constant  $\mu \in [0, 1/2]$ .*

1.  $\delta \leq \sqrt{2\psi_j(\delta)} \leq \frac{1}{2}$  for  $1 \leq j \leq k-1$ ;
2.  $\psi_1(\delta) \leq \psi_2(\delta) \leq \dots \leq \psi_{k-1}(\delta)$ ;
3.  $\psi_j(\delta) \leq \mu^2 \left( \frac{p_j - p_{j+1}}{2E_{j+1}} \right)^2$  for  $1 \leq j \leq k-2$ ;
4.  $\psi_{k-1}(\delta) \leq \mu^2 \min \left\{ \left( \frac{p_k}{2\tilde{E}} \right)^2, \left( \frac{p_{k-1} - p_k}{2E_k} \right)^2 \right\}$ ,

with  $E_j, \tilde{E}$  defined in (5.6), (5.7) respectively. Moreover, for a constant  $\varepsilon \in (0, 1/2)$ , let  $n$  satisfy

$$n \geq \frac{288(2 + \varepsilon)^2 C(\lambda, p_1, \dots, p_k)}{\mu^2 \lambda^2 (\min_{1 \leq j \leq k} \{p_j - p_{j+1}\})^2}$$

where  $C(\lambda, p_1, \dots, p_k) > 0$  depends only on  $\lambda, p_1, \dots, p_k$  (see (5.11)). Then with probability at least  $1 - 3n \exp\left(-\frac{8C(\lambda, p_1, \dots, p_k)n}{\bar{\sigma}(\lambda, p_1, \dots, p_k)^2 c_\varepsilon}\right)$ , it holds that

$$|\langle v_j, z_j \rangle|^2 \geq 1 - \left( \sqrt{\psi_j(\delta)} + \frac{\mu}{1 - \mu} \right)^2, \quad \forall j = 1, \dots, k.$$

Here,  $c_\varepsilon, \bar{\sigma}(\lambda, p_1, \dots, p_k) > 0$  only depend on the indicated parameters (see (5.12)).

The following remarks are in order regarding Theorem 2.

1. A big part of the analysis revolves around deriving bounds on the correlation between the top  $k$  eigenvectors of  $\mathbb{E}[H]$  and  $z_j$ 's for each  $j = 1, \dots, k$  (see Lemma 4). When  $k = 2$ , this was relatively easier, as we had closed-form expressions for the top two eigenvalues of  $\mathbb{E}[H]$ . This is not the case in general of course, and thus, we resort to a deflation argument that leads to upper and lower bounds on the top  $k$  eigenvalues of  $\mathbb{E}[H]$ , defined in a recursive manner.
2. The conditions involving  $\psi_j(\delta)$  impose that  $\delta$  be sufficiently small. In the special case  $\delta = 0$ , we obtain  $\psi_j = 0$  for each  $j = 1, \dots, k$ . We believe that the requirement imposed on  $\delta$  in Theorem 2 is pessimistic, but obtaining less stringent conditions appears to be difficult.
3. It is possible to replace the terms  $C(\lambda, p_1, \dots, p_k)$  and  $\bar{\sigma}(\lambda, p_1, \dots, p_k)$  by their respective upper bounds  $C(\lambda, p_1, \dots, p_k) \lesssim \lambda$  and  $\bar{\sigma}(\lambda, p_1, \dots, p_k) \lesssim 1$ . We then have for  $\delta$  small enough that if

$$\lambda \gtrsim \frac{\log n}{n}, \quad n \gtrsim \frac{1}{\mu^2 \lambda (\min_{1 \leq j \leq k} \{p_j - p_{j+1}\})^2},$$

then the stated correlation bounds hold with probability at least  $1 - \frac{1}{n}$ .

### 5.3 Proof of Theorem 2

Let us denote  $\tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_k$  (resp.  $\tilde{v}_1, \dots, \tilde{v}_k$ ) to be the  $k$  largest eigenvalues (resp. corresponding eigenvectors) of  $\mathbb{E}[H]$ . Since  $z_i$ 's are approximately orthogonal, and  $p_i \neq p_j$ , one would expect for small enough  $\delta$  that  $\tilde{\lambda}_i \approx np_i \lambda$  and  $|\langle \tilde{v}_i, z_i \rangle| \approx 1$  holds. This is stated precisely in the following Lemma.

**Lemma 4.** For  $1 \leq m \leq k$ , denote  $S_m = \sum_{j=1}^m p_j$ . Let  $\delta \in [0, 1]$  additionally satisfy the following conditions.

1.  $\delta \leq \sqrt{2\psi_j(\delta)} \leq \frac{1}{2}$  for  $1 \leq j \leq k-1$ .
2.  $\psi_1(\delta) \leq \psi_2(\delta) \leq \dots \leq \psi_{k-1}(\delta)$ .

It then follows for each  $j = 1, \dots, k$  that

$$np_j \lambda - l_j(\delta) \leq \tilde{\lambda}_j \leq np_j \lambda + u_j(\delta), \quad |\langle \tilde{v}_j, z_j \rangle|^2 \geq 1 - \psi_j(\delta), \quad (5.8)$$

where  $l_j(\delta), u_j(\delta)$  are defined as follows.

$$l_j(\delta) := 4\lambda \sqrt{2\psi_{j-1}(\delta)(nS_{j-1} + \sqrt{2}(j-1)(j-2)(np_1 - np_j) + np_2(k-1)(j-1))}, \quad (5.9)$$

$$u_j(\delta) := 4 \sum_{i=1}^{j-1} (np_i - np_{j+1}) \lambda \sqrt{\psi_i(\delta)} + np_{j+1} \lambda (k-1) \delta, \quad (5.10)$$

with  $\psi_0(\cdot), S_0 = 0$ . In particular,  $l_1(\delta) = 0$  and  $u_1(\delta) = np_2 \lambda (k-1) \delta$ .

*Proof.* See Appendix D.1. □

Our next step is to provide a generalization of Lemma 2. Recall that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  (resp.  $v_1, v_2, \dots, v_n$ ) are denoted to be the eigenvalues (resp. eigenvectors) of  $H$ . Our goal is to quantify the deviation of  $v_i$  from  $\tilde{v}_i$  for each  $i = 1, \dots, k$ . Before stating the lemma we need to define some additional notation.

**Lemma 5.** Under the notation and conditions in Lemma 4, let  $\delta$  additionally satisfy the following conditions for  $\mu \in [0, 1/2]$ .

1.  $\psi_j(\delta) \leq \mu^2 \left( \frac{p_j - p_{j+1}}{2E_{j+1}} \right)^2$  for  $1 \leq j \leq k-2$ ,
2. and  $\psi_{k-1}(\delta) \leq \mu^2 \min \left\{ \left( \frac{p_k}{2E} \right)^2, \left( \frac{p_{k-1} - p_k}{2E_k} \right)^2 \right\}$ ,

with  $E_j, \tilde{E}$  defined in (5.6), (5.7) respectively. If  $\|R\|_2 \leq \Delta$  with  $\Delta \leq \mu \frac{n\lambda}{2} \min_{1 \leq j \leq k} \{p_j - p_{j+1}\}$ , it then follows for each  $j = 1, \dots, k$  that  $|\langle \tilde{v}_j, v_j \rangle|^2 \geq 1 - \left( \frac{\mu}{2(1-\mu)} \right)^2$ .

*Proof.* See Appendix D.2. □

Next, we proceed to bound the spectral norm of the perturbation matrix  $R$ . This is shown in the following lemma, the proof of which is similar to that of Lemma 3.

**Lemma 6.** Let  $R = \text{Re}(R) + \iota \text{Im}(R)$ , where  $\text{Re}(R), \text{Im}(R) \in \mathbb{R}^{n \times n}$  are respectively symmetric and skew-symmetric matrices. Denote

$$\begin{aligned} C(\lambda, p_1, \dots, p_k) &= \sum_{l=1}^k 2p_l \lambda \left[ (1 - p_l \lambda)^2 + \left( \sum_{l' \neq l} p_{l'} \lambda \right)^2 \right] + \left( 1 - \sum_{l=1}^k p_l \lambda \right) \lambda \left[ \frac{1}{2} + \left( \sum_{l'=1}^k p_{l'} \lambda \right)^2 \right] \\ &\quad + (1 - \lambda) \left( \sum_{l'=1}^k p_{l'} \lambda \right)^2, \end{aligned} \quad (5.11)$$



and let  $\bar{\sigma}(\lambda, p_1, \dots, p_l) > 0$  be such that

$$\max_{i,j} \{ \|Re(R)_{ij}\|_\infty, \|Im(R)_{ij}\|_\infty \} \leq \bar{\sigma}(\lambda, p_1, \dots, p_k). \quad (5.12)$$

Then for any  $\varepsilon \geq 0$ , there exists a universal constant  $c_\varepsilon > 0$  such that

$$\|R\|_2 \leq (2 + \varepsilon)6\sqrt{2C(\lambda, p_1, \dots, p_k)n}, \quad (5.13)$$

with probability at least  $1 - 3n \exp\left(-\frac{8C(\lambda, p_1, \dots, p_k)n}{\bar{\sigma}(\lambda, p_1, \dots, p_k)^2 c_\varepsilon}\right)$ .

*Proof.* See Appendix D.3. □

Plugging  $\Delta = (2 + \varepsilon)6\sqrt{2C(\lambda, p_1, \dots, p_k)n}$  in Lemma 5, we obtain

$$\Delta \leq \mu \frac{n\lambda}{2} \min_{1 \leq j \leq k} \{p_j - p_{j+1}\} \iff n \geq \frac{288(2 + \varepsilon)^2 C(\lambda, p_1, \dots, p_k)}{\mu^2 \lambda^2 (\min_{1 \leq j \leq k} \{p_j - p_{j+1}\})^2}. \quad (5.14)$$

Hence if  $n$  satisfies the condition in (5.14) for  $\mu \in [0, 1/2]$ , and  $\delta$  satisfies the conditions in Lemmas 4 and 5, it holds with probability at least  $1 - 3n \exp\left(-\frac{8C(\lambda, p_1, \dots, p_k)n}{\bar{\sigma}(\lambda, p_1, \dots, p_k)^2 c_\varepsilon}\right)$  that

$$|\langle \tilde{v}_j, z_j \rangle|^2 \geq 1 - \psi_j(\delta), \quad |\langle \tilde{v}_j, v_j \rangle|^2 \geq 1 - \left(\frac{\mu}{2(1-\mu)}\right)^2; \quad j = 1, \dots, k.$$

Finally, by using Proposition 2 for each  $j = 1, \dots, k$  with  $\varepsilon_j = \psi_j(\delta)$  and  $\bar{\varepsilon}_j = \left(\frac{\mu}{2(1-\mu)}\right)^2$ , we obtain

$$|\langle v_j, z_j \rangle|^2 \geq 1 - \left(\sqrt{\psi_j(\delta)} + \frac{\mu}{1-\mu}\right)^2.$$

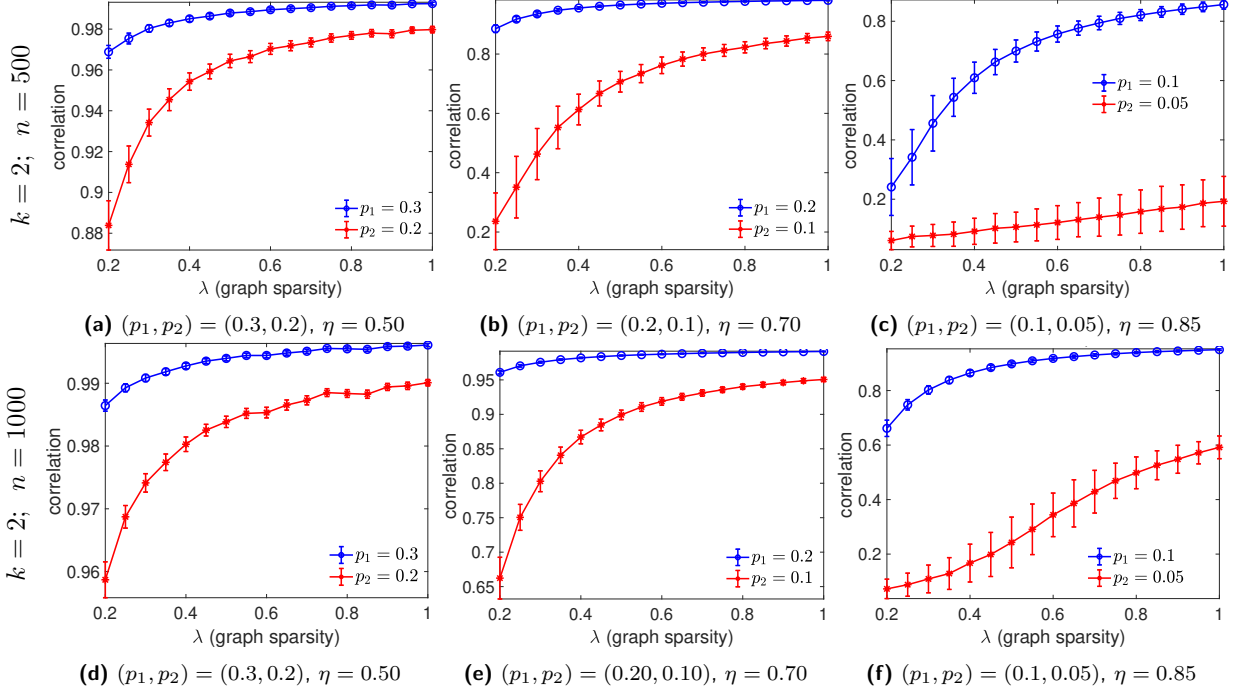
This completes the proof.

## 6 Numerical Experiments

This section details the outcomes of a variety of numerical experiments of our proposed Algorithm 1 for the general setting of  $k$ -synchronization, showcasing its robustness to noise and sampling sparsity. We evaluate performance in a number of settings, where we vary the number of nodes  $n$ , the number of collections of angles  $k$ , and show goodness of recovery as a function of the sampling sparsity of the measurement graph  $G$ . Furthermore, we also experiment with a modified version of Algorithm 1 which employs an additional normalization step, and also compare with the analogous SDP relaxation eq. (2.4) for  $k$ -synchronization, in a subset of the experiments. We measure performance by the correlation between the group of estimated and ground truth angles. More specifically, if  $\theta_i$ , resp.  $\hat{\theta}_i$ , denotes the ground truth, resp. estimated, angles for  $i = 1, \dots, n$ , we compute the correlation as

$$\text{Corr}(\theta, \hat{\theta}) = |\langle z, \hat{z} \rangle|, \quad \text{with} \quad z_i = \frac{1}{\sqrt{n}} e^{i\theta_i}, \quad \text{and} \quad \hat{z}_i = \frac{1}{\sqrt{n}} e^{i\hat{\theta}_i}, \quad i = 1, \dots, n.$$

**Remark 2.** For convenience, one may construe the given measurement graph  $G$  as a union of  $k + 1$  edge-disjoint subgraphs, where  $G_1, G_2, \dots, G_k$  denote the subgraphs of good measurements (with respective edge densities  $p_1\lambda, p_2\lambda, \dots, p_k\lambda$ ), while the subgraph  $W$  contains all the outlier bad measurements (with edge density  $\eta\lambda$ , where we recall that  $\eta = 1 - (p_1 + p_2 + \dots + p_k)$ ).



**Figure 4:** Experimental Setup I: Correlation of recovered angles with the ground truth for  $n = 500$  (top), respectively  $n = 1000$  (bottom), with  $k = 2$  groups of angles, and various noise levels  $\eta \in \{0.50, 0.70, 0.85\}$  as a function of  $\lambda$  (sparsity of the measurement graph). Results are averaged over  $20 \times 20 = 400$  runs (i.e., over 20 random instances of groups of angles, and for each such a group of angles, we consider 20 random instances of the sampling graph according to  $p_1$  and  $p_2$ ).

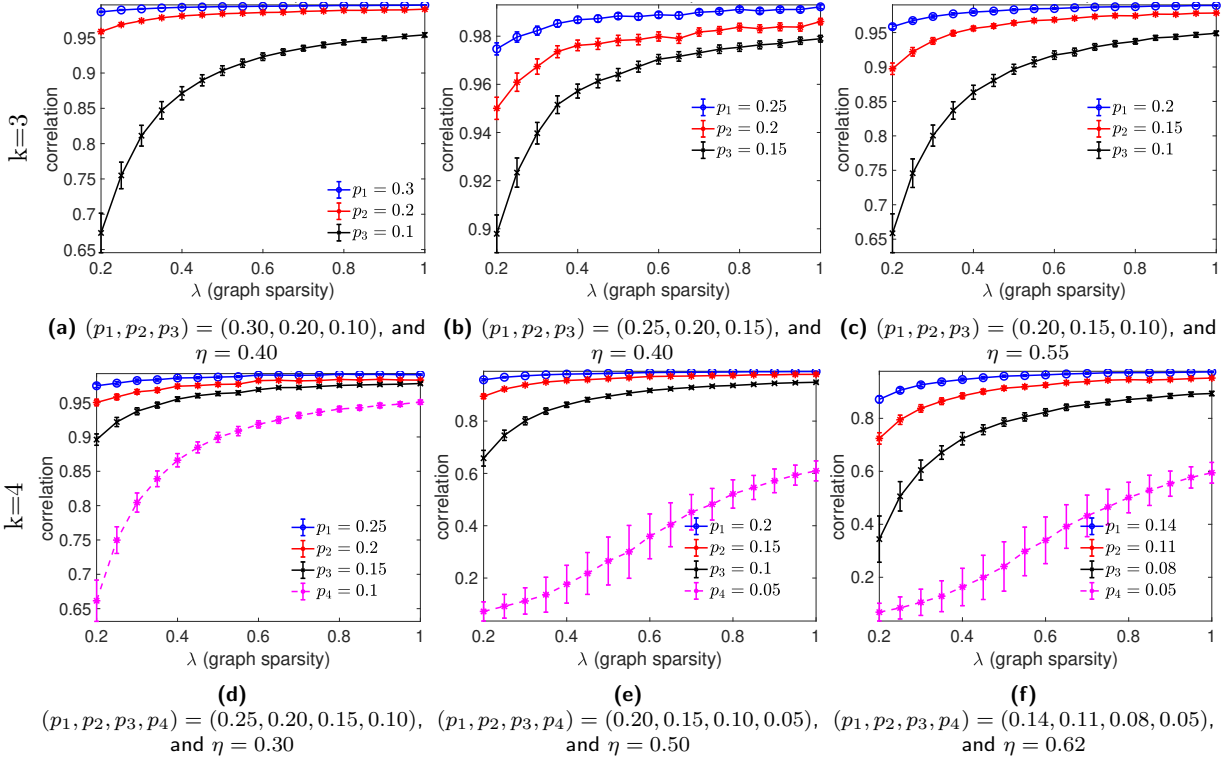
## 6.1 Setup I: correlation versus graph sparsity

The plots in Figure 4 pertain to the setting of bi-synchronization; the top, respectively bottom, row considers the case of  $n = 500$ , respectively  $n = 1000$ , angles for various sparsity and noise levels. As expected, the recovery of the collection of angles whose good measurement graph  $G_1$  is denser (i.e., higher  $p_1$ ) is significantly better than that for which the measurement graph  $G_2$  is sparser (i.e.,  $p_1 > p_2$ ). As the noise level increases, the performance gap between the two recovery error curves becomes wider, as expected. For sparse graphs and large levels of noise (as shown in Figure 4 (c) and (f)), the recovery is significantly worse, especially for the angles corresponding to the sparser measurement graph  $G_2$ .

Figure 5 shows similar plots for the case of  $k = 3$ -way (top row) and  $k = 4$ -way (bottom row) synchronization. When looking at the plots corresponding to  $k = 3$ , as expected, the sparsest measurement graph  $G_3$  attains the worst performance (i.e., the black curve corresponding to  $p_3$ ). Furthermore, when the sparsity levels are close to each other, i.e.,  $p_1 - p_2$  and  $p_2 - p_3$  are small, the overall performance decreases, and the standard deviation error bars increase in magnitude. The bottom-row plots of the same Figure 5 exhibit similar behaviour for the case of  $k = 4$ -way synchronization. The bigger the gap between adjacent sparsity levels, the bigger the performance gap between the recovery levels of the corresponding collections of angles. For the sparsest setting and large levels of noise (as shown in Figure 5 (e) and (f)), we observe a significantly worse recovery for the angles corresponding to the sparser measurement graph  $G_4$  (the magenta curve labeled  $p_4$ ).

## 6.2 Setup II: fixed gap experiments, correlation versus graph sparsity

We also consider a second experimental setup, where we fix the gap between consecutive values of the  $p_i$  sampling parameter corresponding to the fraction of *good* measurements from  $G_i$ . We recall the mixture model considered in (4.4) for  $k = 2$ , and the more general setting from (5.4) for any  $k \geq 2$ , where  $p_i \lambda$  denotes the probability of a correct pairwise measurement from subgraph  $G_i$ ,  $\eta$  is the probability of getting an incorrect measurement, and  $\lambda$  the overall sparsity of the measurement graph. Recall that we have



**Figure 5:** Experimental Setup I: Correlation of recovered angles with the ground truth for  $n = 1000$ , with  $k = 3$  (top) and  $k = 4$  (bottom) collections of angles, and various noise levels  $\eta$ , as a function of  $\lambda$  (sparsity of the measurement graph). Results are averaged over  $20 \times 20 = 625$  runs.

assumed w.l.o.g that  $p_1 > p_2 > \dots > p_k$ . In the following experiments, for a given noise level  $\eta$ , we choose  $p_1, \dots, p_k$  such that  $\sum_{l=1}^k p_l = 1 - \eta$  and  $p_{l+1} - p_l = \gamma, \forall l = 1, \dots, k - 1$ . For example, under the complete measurement graph scenario  $\lambda = 1$ , if  $k = 4$ ,  $\eta = 0.20$ , and the gap parameter  $\gamma = 0.05$ , the resulting good-edge probabilities are given by  $p_1 = 0.275$ ,  $p_2 = 0.225$ ,  $p_3 = 0.175$ ,  $p_4 = 0.125$ . The final mixture model is then

$$\Theta_{ij} = \begin{cases} \theta_{l,i} - \theta_{l,j} & \text{with probability } p_l \lambda \text{ for } l = 1, \dots, k \\ \Theta_{ij} \sim U[0, 2\pi] & \text{with probability } (1 - \sum_{l=1}^k p_l) \lambda \\ 0 & \text{with probability } 1 - \lambda. \end{cases} \quad (6.1)$$

Figure 6 reports the outcomes of numerical experiments, for three different values of  $k \in \{2, 3, 4\}$  (indexing the rows), and three values of the graph density parameter  $\lambda \in \{0.2, 0.4, 0.8\}$ .

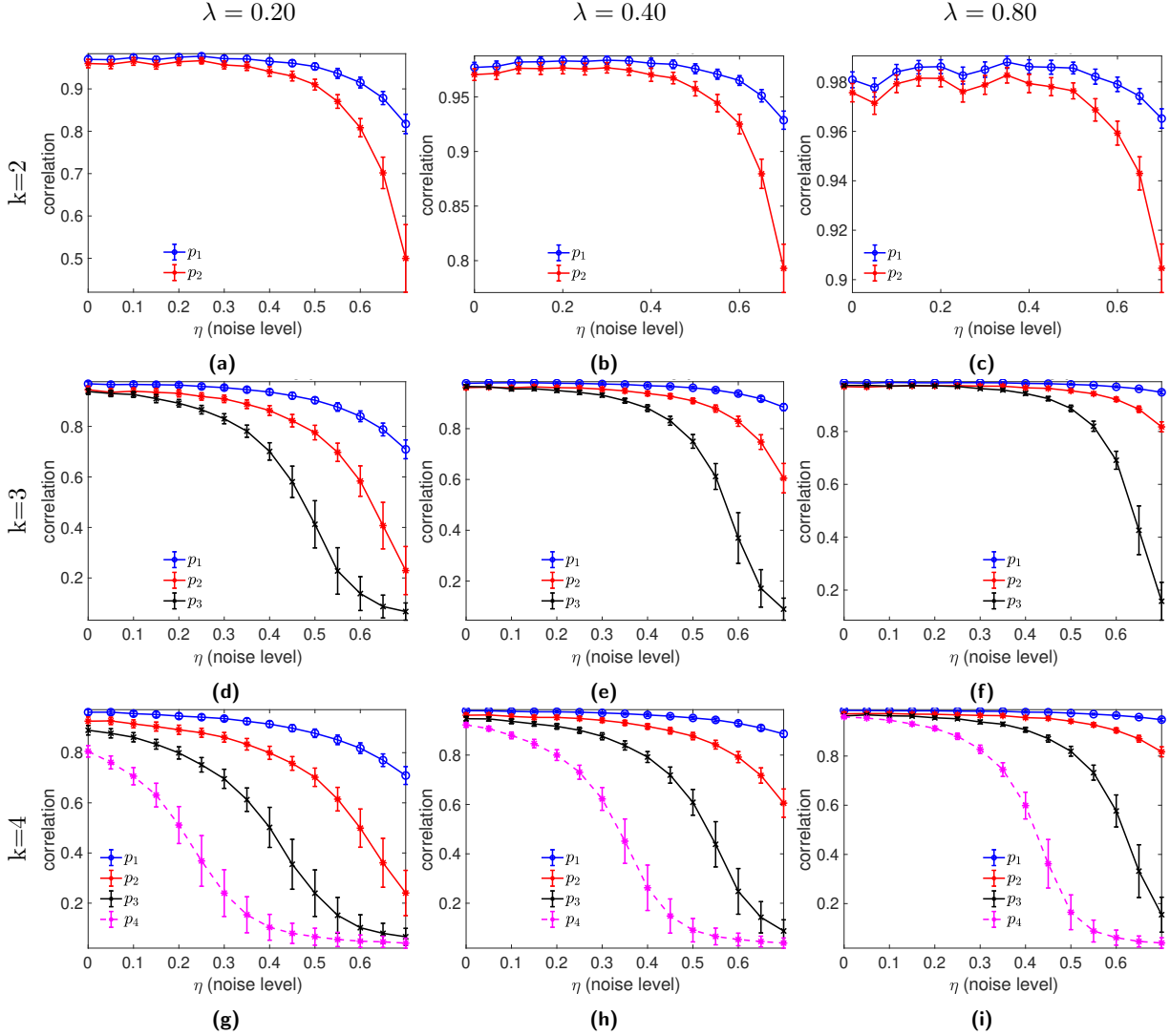
### 6.3 Comparison with the Normalized Spectral and SDP relaxations

This section compares different spectral and SDP relaxations of the heterogeneous angular  $k$ -synchronization problem under the second experimental setup, where we fix the gap probability  $\gamma$ , and compare the accuracy of the various methods as we vary the noise level.

**Normalization of  $H$**  When dealing with measurement graphs exhibiting a skewed degree distribution, a suitable normalisation is typically employed in problems involving spectral methods, such as clustering (see for eg. [29]). To this end, instead of extracting the final estimated angles from the top eigenvectors of  $H$ , the authors of [19] considered the following normalization of  $H$  by the diagonal matrix  $D$  with diagonal elements given by  $D_{ii} = \sum_{j=1}^N |H_{ij}|$ , arriving at

$$R = D^{-1}H. \quad (6.2)$$

Note that  $R$  is similar to the Hermitian matrix  $D^{-1/2}HD^{-1/2}$  via  $R = D^{-1/2}(D^{-1/2}HD^{-1/2})D^{1/2}$ , and thus  $R$  has  $n$  real eigenvalues. This normalization was subsequently employed with success in [17, 22], where

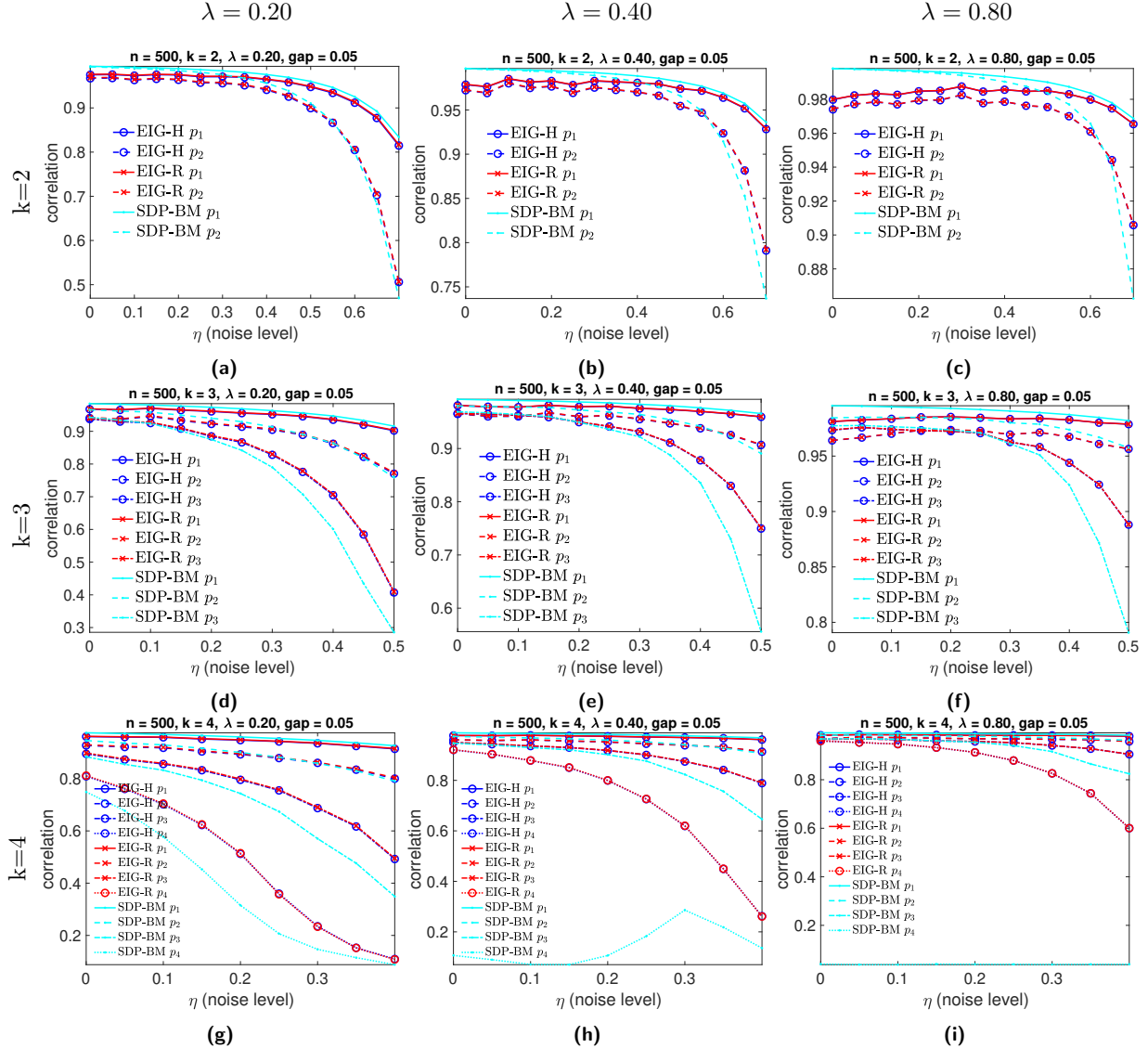


**Figure 6:** Experimental Setup II: Correlation of recovered angles with the ground truth, where we keep constant the gap  $\gamma = 0.05$  between consecutive sampling probabilities  $p_{l+1} - p_l = \gamma, l = 1, \dots, k - 1$ , of the subgraphs  $G_l$ . We also fix the number of angles  $n = 500$ , and vary  $k \in \{2, 3, 4\}$  (indexing the rows), and the overall measurement graph sparsity  $\lambda \in \{0.2, 0.4, 0.8\}$  (indexing the columns). Each plot shows the correlation with the ground truth, as we vary the noise level  $\eta$ . Results are averaged over  $20 \times 20 = 400$  runs.

it was shown to outperform the unnormalized version.

Note that the operator  $R$  was studied in the context of angular synchronization and the graph realization problem [19], and also in [35] which introduced Vector Diffusion Maps for nonlinear dimensionality reduction and explored the interplay with the Connection-Laplacian operator for vector fields over manifolds. We also remark that these Hermitian operators (construed as Hermitian analogues to the usual graph Laplacians) have been successfully used in the directed clustering [18] and ranking [17] literatures, where the net outcome of pairwise matches between players can be encoded in a digraph with a skew-symmetric adjacency matrix. In particular, [17] formulated the ranking problem as an instance of the angular synchronization problem, considered an angular embedding and relied on the top eigenvector of the above matrix operator  $R$  to recover a one-dimensional ordering of the players. Very recently, we have considered the same normalization in [22] in the context of ranking and synchronization over  $\mathbb{R}$  via SVD. In another recent work, [24] proposed a certain deformation of the combinatorial Laplacian, in particular, the *Dilation Laplacian*, which is shown to perform well for ranking in directed networks of pairwise comparisons.

Figure 7 shows the correlation of recovered angles with the ground truth in the second experimental



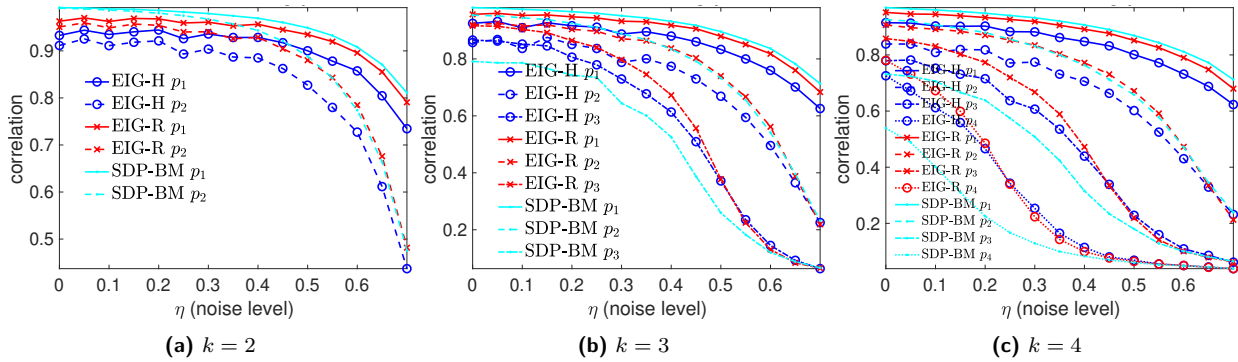
**Figure 7:** Experimental Setup II: Correlation of recovered angles with the ground truth in the second experimental setup where we compared the performance of the two spectral relaxations and the SDP relaxation. EIG-H denotes the relaxation that uses the top eigenvector of matrix  $H$  from (4.5); EIG-R employs the normalized matrix eq. (6.2), and SDP-BM denotes the SDP relaxation eq. (2.4) solved via the Burer-Monteiro approach. We keep constant the gap  $\gamma = 0.05$  between consecutive sampling probabilities  $p_{l+1} - p_l = \gamma, l = 1, \dots, k - 1$ , of the subgraphs  $G_l$ . The number of angles is kept fixed at  $n = 500$ , and we vary  $k \in \{2, 3, 4\}$  (indexing the rows), and the overall measurement graph sparsity  $\lambda \in \{0.2, 0.4, 0.8\}$  (indexing the columns). Each plot shows the correlation with ground truth, as we vary the noise level  $\eta$ . Results are averaged over  $20 \times 20 = 400$  runs.

setup, comparing the performance of the two spectral relaxations and the SDP relaxation. The full set of methods we compare are

- EIG-H: the relaxation that uses the top  $k$  eigenvectors of matrix  $H$ , as in Algorithm 1;
- EIG-R: employs the normalized version of  $H$  as detailed in (6.2),
- SDP-BM: denotes the SDP relaxation (2.4) solved via the Burer-Monteiro approach [9, 12]. Note that after solving the SDP program, we extract angles from the top  $k$  eigenvectors of the computed solution  $\Upsilon$  (an  $n \times n$  complex-valued Hermitian positive-semidefinite matrix), following the same steps outlined in Algorithm 1.

We fixed the number of angles  $n = 500$ , and kept constant the gap  $\gamma = 0.05$  between consecutive sampling probabilities  $p_{l+1} - p_l = \gamma$ ,  $l = 1, \dots, k - 1$ , of the subgraphs  $G_l$ . We varied  $k \in \{2, 3, 4\}$  and the overall measurement graph sparsity  $\lambda \in \{0.2, 0.4, 0.8\}$ , and plotted the correlation with ground truth, as we varied the noise level  $\eta$ . The experiments reveal that the SDP relaxation typically outperforms the spectral relaxations on the angle recovery task, with the exception of the set of angles corresponding to the sparsest measurement subgraph  $G_3$  for the case  $k = 3$ , and  $G_3$  and  $G_4$  for the case  $k = 4$ .

**Numerical experiments for the Barabási–Albert model** Finally, to highlight the benefits of the normalization step, we also compare the two spectral relaxations with the SDP relaxation in the setting where the underlying measurement graph (indicating the presence or absence of an edge) is generated from the *Barabási–Albert* (BA) model [8], which is known to lead to skewed degree distributions. We denote by  $G_{BA}$  the resulting BA network, which can be used to build the relevant measurement matrix. In other words, the  $\Theta$  matrix is given by the Hadamard product between the complete angle offset measurement graph (for e.g., as generated by the model in (6.1) with  $\lambda = 1$ ), and the  $G_{BA}$  graph.



**Figure 8:** Experimental Setup II: Performance comparison in the setting of a Barabási–Albert model, for  $n = 500$ , edge density  $\lambda = 0.20$ , gap  $\gamma = 0.05$ ,  $k \in \{2, 3, 4\}$ , as a function of the noise levels  $\eta$ . Results are averaged over  $20 \times 20$  runs.

## 6.4 Graph disentangling procedure

In this section, we give a procedure for uncovering the latent measurement good graphs  $G_1, G_2, \dots, G_k$  corresponding to the different groups of angles. We refer to this process as “*graph disentangling*”, which is illustrated in Figure 9. The input to this pipeline is given by the initial estimates of the  $k$  groups of angles,  $\hat{\theta}_{l,i}$ , for  $l = 1, \dots, k$ , and  $i = 1, \dots, n$ . In a nutshell, the algorithm we propose considers the **residual matrix** for each recovered solution with respect to the initial input matrix of pairwise measurements  $\Theta$

$$\Psi_{l,ij} = \min\{(\Theta_{ij} - \hat{\Theta}_{l,ij}) \bmod 2\pi, (\hat{\Theta}_{l,ij} - \Theta_{ij}) \bmod 2\pi\}, \quad l = 1, \dots, k, \quad (6.3)$$

where  $\hat{\Theta}_{l,ij} = (\hat{\theta}_{l,i} - \hat{\theta}_{l,j}) \bmod 2\pi$ , and leverages this set of matrices  $\Psi_l, l = 1, \dots, k$  for the purpose of attributing a given edge  $\{i, j\} \in E$  to one of the subgraphs  $G_1, G_2, \dots, G_k$ . More specifically, for each edge  $\{i, j\}$  of the initial measurement graph  $G$ , we consider the set of residuals induced by each solution vector  $\hat{\theta}_l, l = 1, \dots, k$ , and attribute the smallest residual in the list to the corresponding estimated graph  $\hat{G}_l$ . In other words, we first compute the matrix of smallest residual errors

$$\Gamma_{ij} = \min_{l=1, \dots, k} \Psi_{l,ij}, \quad \{i, j\} \in E, \quad (6.4)$$

and then update the residual matrix corresponding to each set of angles  $l = 1, \dots, k$

$$\tilde{\Psi}_{l,ij} = \begin{cases} \Gamma_{ij} & \text{if } \Psi_{l,ij} = \Gamma_{ij} \\ 0 & \text{if otherwise} \end{cases}. \quad (6.5)$$

Note that we have decomposed  $\Gamma_{ij} = \sum_{l=1}^k \tilde{\Psi}_{l,ij}$ , making the implicit assumption that there is a unique minimum in (6.4) (if this is not the case, we break ties arbitrarily and assign the residual to a single matrix

$\tilde{\Psi}_l$ ). The top row of Figure 10 shows the distribution of the nonzero entries for each matrix  $\tilde{\Psi}_l$ ,  $l = 1, 2, 3$ , for an instance of the  $k$ -synchronization problem with  $k = 3$ .

In the remaining part of this procedure, we use the support of the  $k$  matrices  $\tilde{\Psi}^{(1)}, \tilde{\Psi}^{(2)}, \dots, \tilde{\Psi}^{(k)}$  to recover the  $k + 1$  underlying graphs given by

- the  $k$  measurement *good* graphs  $G_1, G_2, \dots, G_k$ , and with adjacency matrices given by  $A_1, A_2, \dots, A_k$ , henceforth denoted as the list  $(G_l, A_l)$ ,  $l = 1, \dots, k$ ;
- the *bad* graph  $W$  (with adjacency matrix  $B$ ) corresponding to the outlier entries.

Denote by  $(\tilde{G}_l, \tilde{A}_l)$  the support graph corresponding to the matrix of residuals  $\tilde{\Psi}^{(l)}$ ,  $l = 1 \dots k$ . Let

$$\Theta^{(l)} = \Theta \odot \tilde{A}_l,$$

where  $\odot$  denotes the entry-wise Hadamard product of two matrices, and note that this also induces the following decomposition of the initial measurement matrix  $\Theta$

$$\Theta = \sum_{l=1}^k \Theta^{(l)}.$$

Recall the ground truth decomposition of the adjacency matrix  $A$  of the measurement graph  $G$

$$A = \sum_{l=1}^k A_l + B,$$

and the fact that  $G$  is comprised of both *good* and *bad* edges. Each good edge  $\{i, j\} \in E_l$  will ideally be placed in its corresponding graph  $\tilde{G}_l$  as it should achieve the smallest residual in (6.4). However, each (*bad*) edge in  $W$  will also be randomly attributed to one of the graphs  $\tilde{G}_l$ . To this end, we perform a second synchronization stage, where we further synchronize individually each  $\Theta^{(l)}$  matrix, recover the solution angles  $\hat{\theta}_{l,i}^{(1)}$  (the superscript is to indicate that this is the first iteration of our entire pipeline described thus far), and then rebuild each denoised pairwise comparison matrix  $\hat{\Theta}^{(l)}$  as

$$\hat{\Theta}_{ij}^{(l)} = (\hat{\theta}_{l,i}^{(1)} - \hat{\theta}_{l,j}^{(1)}) \pmod{2\pi}.$$

Similar to (6.3), we first build the matrix of residuals between  $\hat{\Theta}^{(l)}$  and  $\Theta^{(l)}$ , restricted to the support of  $\tilde{G}_l$  (which only contain edges present in the original measurement graph  $G$ ), and finally classify each edge as being *good* or *bad* depending on the magnitude of its residual. In doing so, for simplicity, we assume knowledge of the noise level  $1 - p_l$ , and classify as *bad* edges the  $1 - p_l$  percent of the edges with the largest residual, as depicted in the histograms in Figure 10. This altogether renders the following decomposition of  $(\tilde{G}_l, \tilde{A}_l)$  into the *bad* and *good* subgraphs with adjacency matrices satisfying

$$\tilde{A}_l = \tilde{A}_l^{good} + \tilde{A}_l^{bad}.$$

Finally, we obtain edge-disjoint estimates  $\hat{A}_1, \dots, \hat{A}_k$  for the set of graph adjacency matrices  $A_1, \dots, A_k$  as

$$\hat{A}_l := \tilde{A}_l^{good},$$

while for the bad graph  $W$ , we pool together all the bad edges inferred from each of the  $k$  solutions, and estimate the adjacency matrix of  $W$  as

$$\hat{B} := \sum_{l=1}^k \tilde{A}_l^{bad}.$$

Figure 9 is a pictorial representation of the above pipeline for  $k = 2$ , that depicts the extraction of the underlying subgraph structures. We consider the case of a complete measurement graph (thus  $\lambda = 1$ ), with  $k = 2$  subgraphs  $G_1$  and  $G_2$  corresponding to two sets of measurement angles  $\theta_1$  and  $\theta_2$ , with good edge probability  $p_1 = 0.45$ , respectively  $p_2 = 0.35$ , for the correct pairwise offsets. Recall that in the bi-synchronization setup, we denoted two such sequences of angles by  $\theta_1 = (\alpha_1, \dots, \alpha_n)$  and  $\theta_2 = (\beta_1, \dots, \beta_n)$ . The subgraph  $W$  of outlier measurements has  $\eta = 1 - (p_1 + p_2) = 0.2$ . These graphs are depicted in the top row of Figure 9. The second row of this figure illustrates the combined graph  $G = G_1 \cup G_2 \cup W$  (with adjacency matrices  $A = A_1 + A_2 + B$ ), which is available to the user. The bottom row shows the disentangled graph estimates as recovered by the above procedure.

**An iterative approach for synchronization** The entire pipeline described thus far can be iterated  $M$  times, and the entire process is summarized in Algorithm 2. The bottom plot of the Figure 10 shows a histogram of the residuals for each of the three sets of estimated angles, after  $M = 20$  iterations of the above process. Note that it is noticeable to the naked eye that the residuals become significantly smaller after the iterative process, for all three sets of angles, hinting on the effectiveness of the proposed procedure. To this end, we furthermore plot in Figure 11 the correlation between the estimated solutions and the ground truth, for each of the first  $M = 20$  iterations of Algorithm 2. We consider three different problem instances  $k = 2$  (with  $p_1 = 0.23, p_2 = 0.15$ ),  $k = 3$  (with  $p_1 = 0.18, p_2 = 0.15, p_3 = 0.12$ ),  $k = 4$  (with  $p_1 = 0.2, p_2 = 0.17, p_3 = 0.13, p_4 = 0.10$ ), and remark that in each problem instance and for each collection of angles, the correlations increase for subsequent iterations, showcasing the efficacy of the iterative procedure.

**Remark 3.** Note that the proposed iterative scheme is of independent interest, and could also be considered in the context of the classical angular synchronization problem, where  $k = 1$  and the goal is uncover the subgraph  $G_1$  of good edges and the subgraph  $W$  of bad edges. Furthermore, re-weighting schemes could also be explored, where a confidence score (at iteration  $r$ ) on the available measurement can be used and based on the residual from the previous iteration  $r - 1$ . To the best of our knowledge, such an iterative re-weighting scheme has not been previously considered in the synchronization literature, and it would be an interesting research direction to analyze how it compares to the standard methods detailed in Section 2.

---

**Algorithm 2** ITERATIVE SYNCHRONIZATION AND GRAPH DISENTANGLING ALGORITHM

---

- 1: **Input:** Measurement graph  $G = ([n], E)$ , pairwise offset measurements  $\Theta_{ij}$  for  $\{i, j\} \in E$ , number of sets of angles  $k$ , number of iterations  $M$ .  
// STAGE 1: COMPUTE INITIAL ESTIMATES FOR  $\hat{\theta}_{l,i}^{(0)}$
  - 2: Estimate angles  $\hat{\theta}_{l,i}^{(0)}$  from the top  $k$  eigenvectors of the  $H$  as detailed in Algorithm 1.  
// STAGE 2: ITERATIVE PROCEDURE
  - 3: **for**  $r = 1, 2, \dots, M$  **do**
  - 4:   Build the residual matrices  $\Psi_l, l = 1, \dots, k$  as in (6.3).
  - 5:   Use the residual matrices to decompose the input measurement matrix  $\Theta$  into  $\Theta = \sum_{l=1}^k \Theta^{(l)}$ .
  - 6:   Synchronize each individual matrix  $\Theta^{(l)}$ , obtain a denoised estimate  $\hat{\Theta}^{(l)}$  matrix, and angles  $\hat{\theta}_{l,i}^{(r)}$
  - 7:   Consider the residual matrix between  $\hat{\Theta}^{(l)}$  and  $\Theta^{(l)}$ , and classify each existing edge as *good* or *bad*
  - 8:   Producing the set of graph estimates  $\hat{G}_1, \dots, \hat{G}_n$ , and  $W$ .
  - 9: **end for**
  - 10: **Output:** Estimates  $\hat{\theta}_{l,i}^{(r)}$ , for  $l = 1, \dots, k$ , and  $i = 1, \dots, n$ , and subgraphs of *good* edges  $\hat{G}_1^{(r)}, \dots, \hat{G}_k^{(r)}$  and subgraph of *bad* edges  $W^{(r)}$ .
- 

## 7 An application to the graph realization problem

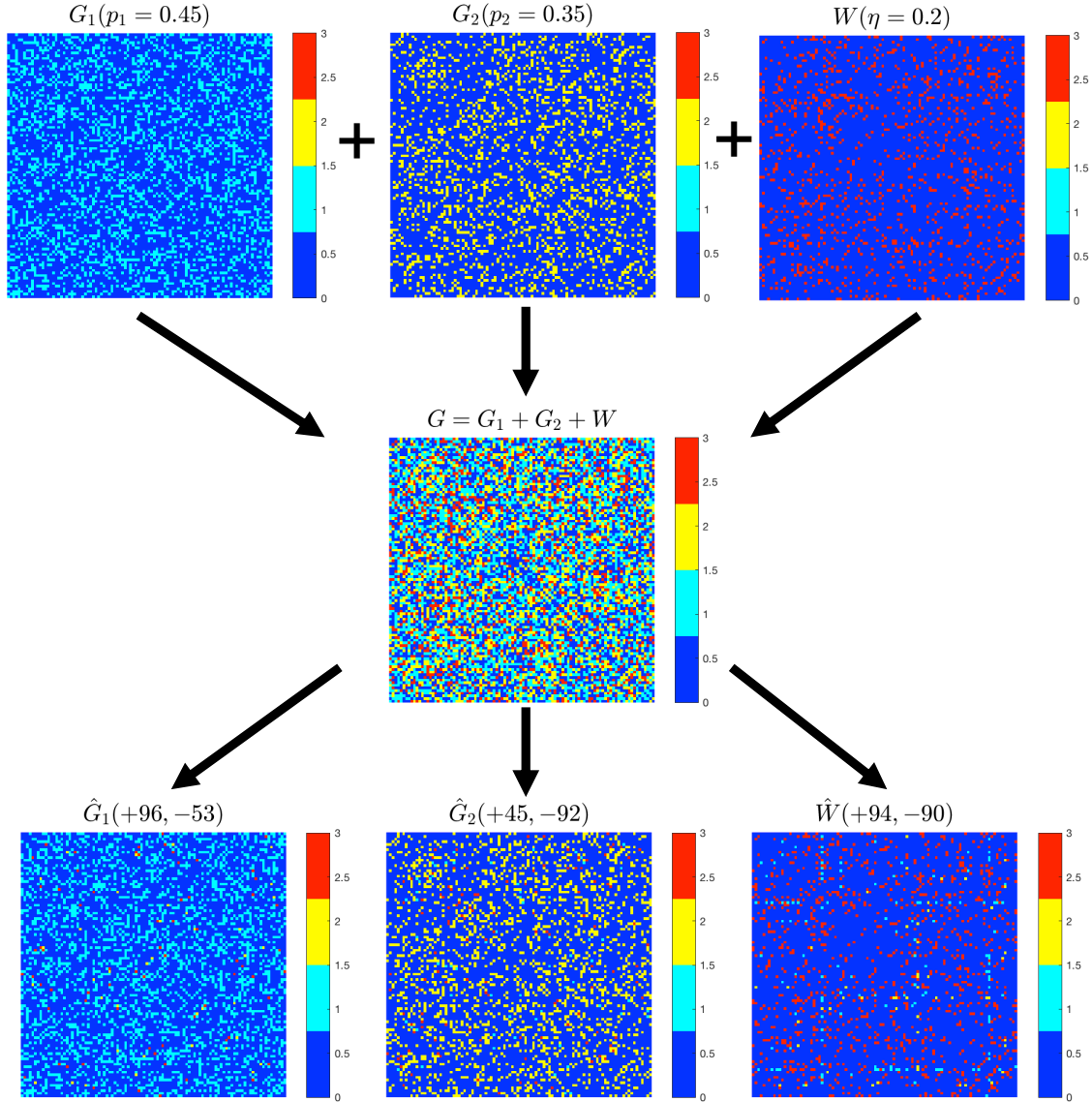
We now turn our attention to an application of bi-synchronization to a particular instance of the graph realization problem, where the graph disentangling procedure will prove useful.

### 7.1 The graph realization problem

In the *graph realization problem* (GRP), one is given a graph  $G = ([n], E)$  together with a non-negative distance measurement  $d_{ij}$  associated with each edge, and is asked to compute a realization of  $G$  in  $\mathbb{R}^d$ . In other words, for any pair of adjacent nodes  $i$  and  $j$ , the distance  $d_{ij} = d_{ji}$  is available, and the goal is to find a  $d$ -dimensional embedding  $p_1, p_2, \dots, p_n \in \mathbb{R}^d$  such that  $\|p_i - p_j\| = d_{ij}$ , for all  $\{i, j\} \in E$ . Due to its practical significance, the GRP has attracted a lot of attention in recent years, across many communities such as wireless sensor networks [10, 38], structural biology [28], dimensionality reduction, Euclidean ball packing and multidimensional scaling (MDS) [16].

In many real world applications, such as sensor networks and structural biology, the given distances  $d_{ij}$  between adjacent nodes are not accurate,  $d_{ij} = \|p_i - p_j\| + \varepsilon_{ij}$  where  $\varepsilon_{ij}$  represents the added noise, and



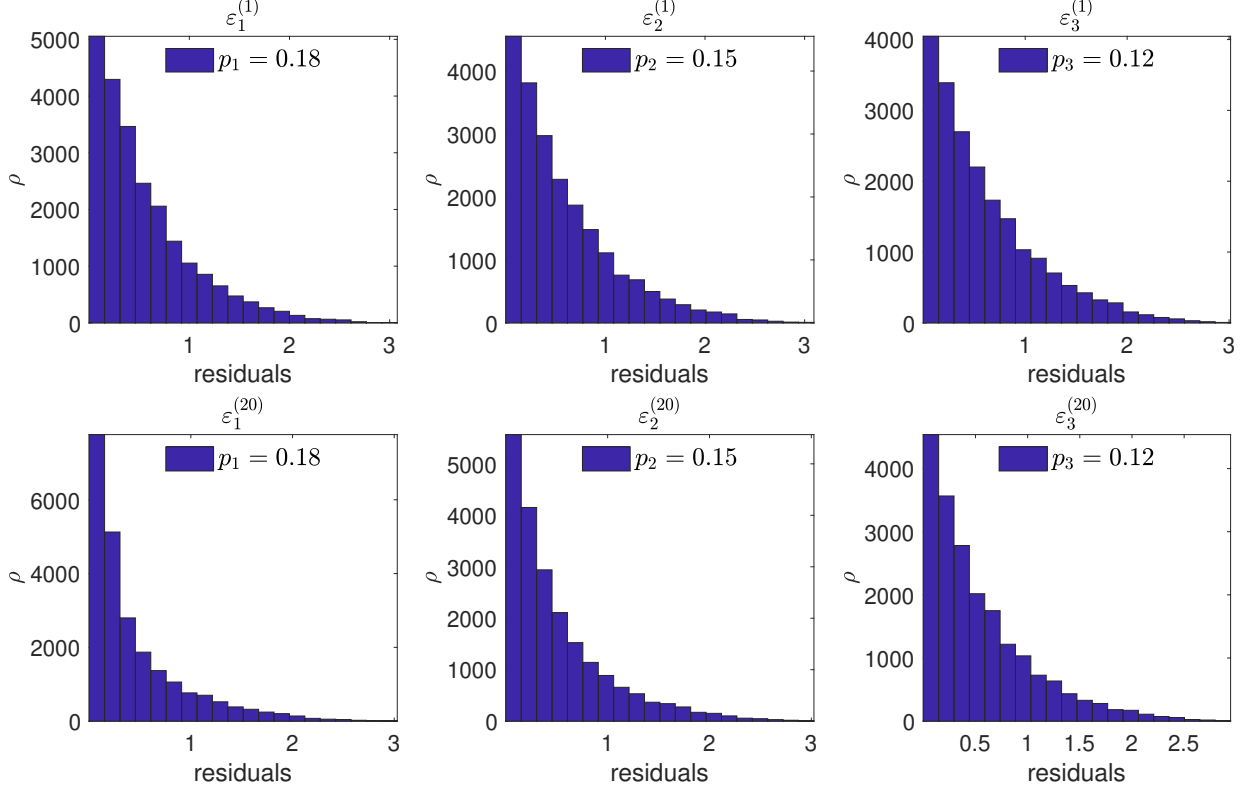


**Figure 9:** Top: ground truth *good* graph  $G_1, G_2$  and the *bad* graph  $W$ , all edge disjoint, of size  $n = 100$ , with corresponding probabilities  $p_1 = 0.45$ ,  $p_2 = 0.35$  and  $\eta = 0.20$ . Middle: the user observes an angle offset matrix  $\Theta$  whose underlying entangled measurement graph  $G = G_1 \cup G_2 \cup W$  is depicted. Bottom: final estimates of  $\hat{G}_1, \hat{G}_2, \hat{W}$  from our recovery pipeline, with a visualization of the classification errors; the titles contain the two types of errors being made: the number of extra edges (+) and the number of missing edges (-) for each individual estimated graph.

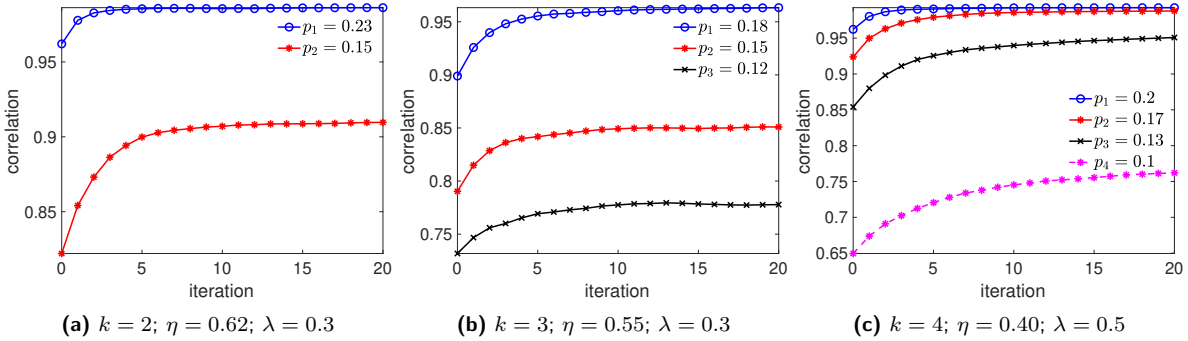
the goal is to find an embedding that realizes all known distances  $d_{ij}$  as best as possible. The 2D-ASAP algorithm proposed in [19] belongs to the group of algorithms that integrate local distance information into a global structure determination. For every sensor, one first identifies globally rigid<sup>2</sup> subgraphs of its 1-hop neighborhood, denoted as patches. Each patch is then separately localized in a coordinate system of its own, using for example either a stress minimization approach [27], or semidefinite programming (SDP).

In the noise-free case, the coordinates of the nodes in each patch must agree with their global positioning up to some unknown rigid motion, that is, up to translation, rotation and possibly reflection. To every patch

<sup>2</sup>A framework is a finite graph  $G$  together with a finite set of vectors in  $d$ -space, where each  $p_i$  corresponds to a node of  $G$ , and the edges of  $G$  correspond to fixed length bars connecting the nodes. Two frameworks with the same graph  $G$  are equivalent if the lengths of their edges are the same. A framework  $G(p)$  is *globally rigid* in  $\mathbb{R}^d$  if all frameworks  $G(q)$  in  $\mathbb{R}^d$  which are  $G(p)$ -equivalent (have all bars the same length as  $G(p)$ ) are congruent to  $G(p)$  (that is, they are related by a rigid transformation). A graph  $G$  is *generically globally rigid* in  $\mathbb{R}^d$  if  $G(p)$  is globally rigid at all generic configurations  $p$  [14, 15]. Only recently it was demonstrated that global rigidity is a generic property in this sense, for graphs in each dimension [15, 26].



**Figure 10:** Histogram of synchronization residuals given by the non-zero entries of the matrix  $\tilde{\Psi}_l$ ,  $l = \{1, 2, 3\}$ , as defined in (6.5), where the  $k = 3$  underlying measurement graphs have edges densities  $(p_1, p_2, p_3) = (0.18, 0.15, 0.12)$ , noise level  $\eta = 0.55$ , and sparsity  $\lambda = 0.3$ . The top (resp. bottom) row plots the non-zero residual entries after one iteration (resp., 20 iterations of Algorithm 2), with columns indexing the three groups of underlying angles. Note that for each column (i.e., group of angles), the residuals become significantly smaller after 20 iterations, for all three groups of angles, showcasing the effectiveness of the iterative procedure.



**Figure 11:** Correlation with ground truth for the first 20 iterations of Algorithm 2, for three problem instances with  $k = 2, 3, 4$  and  $n = 500$ , at various noise levels  $\eta$  and sparsity levels  $\lambda$ . In all three problem instances, and across all subsets of angles within each instance, the iterative procedure leads to an increase in the correlation with ground truth, especially for the groups of angles with the sparsest measurement graph.

there corresponds an element of the Euclidean group  $\text{Euc}(d)$ , and the goal is to estimate the group elements that will properly align all the patches in a globally consistent manner. By finding the optimal alignment of all pairs of patches whose intersection is large enough, we obtain measurements for the ratios of the unknown group elements, which is thus an instance of the group synchronization problem over the Euclidean group  $\text{Euc}(d)$ .

Unlike the generative models analyzed in the earlier sections, the measurement graph  $G$  arising in the graph realization problem is a disc graph, as opposed to an Erdős-Rényi graph. Consider a patch  $P_i$  to be given by a central node  $i$  along with all its neighbors within a radius  $r$ . In order for a pair of patches  $P_i$

and  $P_j$  to be aligned, they need to have enough points in common (typically at least  $d + 1$ , where  $d$  is the ambient dimension), which in turn means that the two patches  $P_i$  and  $P_j$  have their center nodes at most  $2r$  distance apart. Thus, the measurement graph  $G$  is best modeled by a disc graph.

The specific problem instance we consider here is one where there exist two underlying non-congruent embeddings that we seek to recover,  $X$  and  $Y \in \mathbb{R}^{n \times 2}$ . For each such embedding, we have available information on the embedding of patches (subgraph embeddings), but whenever we have available a pairwise alignment between two patches, we do not know a-priori whether this measurement pertains to structure  $X$  or  $Y$ . We further detail this setup later in the section, and next introduce an existing algorithm from the literature on the graph realization problem.

## 7.2 The ASAP algorithm

This section details the steps of the ASAP (As-Synchronized-As-Possible) algorithm, a divide-and-conquer pipeline proposed in [19] for the problem of scalable recoverings of point clouds from a sparse noisy set of pairwise distance information.

The ASAP approach starts by decomposing the given graph  $G$  into overlapping subgraphs (referred to as *patches*), which are then embedded via the method of choice. To every local patch embedding, there corresponds a scaling and an element of the Euclidean group  $\text{Euc}(d)$  of  $d$ -dimensional rigid transformations, and our goal is to estimate the group elements that will properly align all the patches in a globally consistent framework. The local optimal alignments between pairs of overlapping patches (whose intersection should be large enough - preferably tens of even hundreds of nodes) yield noisy measurements for the ratios of the above unknown group elements. Finding group elements from noisy measurements of their ratios is nothing but the group synchronization problem, the central topic of our present paper. For completeness, Table 3 gives an overview of the approach, and Figure 12 shows a schematic view of the pipeline we consider, as detailed in [19].

**Algorithm 3** The ASAP algorithm [19].

INPUT	$G = (V, E)$ , $ V  = n$ , $ E  = m$ , $d$
Choose Patches Embed Patches	1. Break $G$ into $N$ overlapping patches $P_1, \dots, P_N$ . 2. Embed each patch $P_i$ separately via the method of choice (for eg, cMDS).
Step 1 Rotate & Reflect	1. Align all pairs of patches $(P_i, P_j)$ that have enough nodes in common. 2. Estimate their relative rotation and possibly reflection $H_{ij} \in O(d) \subset \mathbb{R}^{d \times d}$ . 3. Build a sparse $dN \times dN$ symmetric matrix $H = (H_{ij})$ where entry $ij$ is itself a matrix in $O(d)$ . 4. Define $\mathcal{H} = D^{-1}H$ , where $D$ is a diagonal matrix with $D_{1+d(i-1), 1+d(i-1)} = \dots = D_{di, di} = \text{deg}(i)$ , $i = 1, \dots, N$ , where $\text{deg}(i)$ is the node degree of patch $P_i$ . 5. Compute the top $d$ eigenvectors $v_i^{\mathcal{H}}$ of $\mathcal{H}$ satisfying $\mathcal{H}v_i^{\mathcal{H}} = \lambda_i^{\mathcal{H}}v_i^{\mathcal{H}}$ , $i = 1, \dots, d$ . 6. Estimate the global reflection and rotation of patch $P_i$ by the orthogonal matrix $\hat{h}_i$ that is closest to $\tilde{H}_i$ in Frobenius norm, where $\tilde{H}_i$ is the submatrix corresponding to the $i^{\text{th}}$ patch in the $dN \times d$ matrix formed by the top $d$ eigenvectors $[v_1^{\mathcal{H}} \dots v_d^{\mathcal{H}}]$ . 7. Update the embedding of patch $P_i$ by applying the orthogonal transformation $\hat{h}_i$ .
Step 2 Translate	Solve $m \times n$ overdetermined system of linear equations for optimal translation in each dimension.
OUTPUT	Estimated coordinates $\hat{x}_1, \dots, \hat{x}_n$

## 7.3 The multi-graph realization problem

The patch alignment problem and the ASAP algorithm detailed above could be construed as aligning pieces of a jigsaw puzzles. Patches could be interpreted as pieces of the puzzle, and the task becomes to assign to each individual piece an element of the Euclidean Group in such a way that when we apply the specific transformation to each individual patch, all the pieces of the puzzle synchronize together. For the case of a

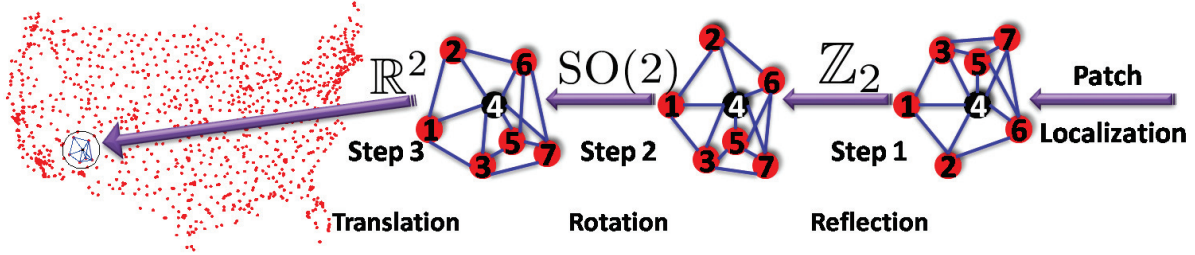


Figure 12: The ASAP recovery process in  $d = 2$  dimensions, considered in [19], for a patch in the US graph.

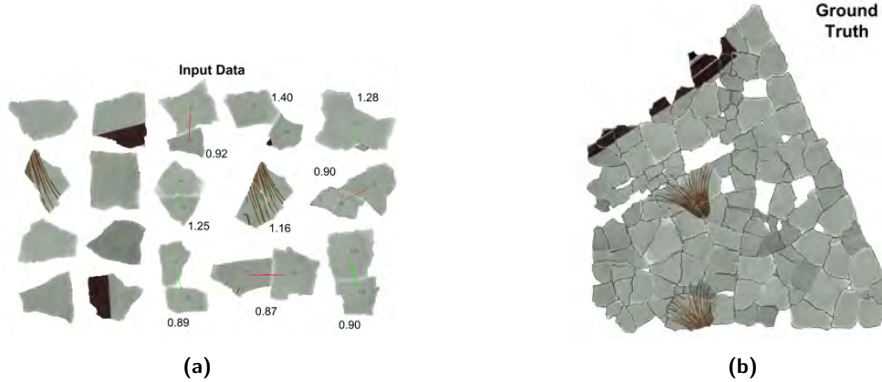


Figure 13: Alignment of frescos, as explored in [37]. The alignment of the pieces can be construed as an instance of the synchronization problem over the Euclidean group of rigid motions.

jigsaw puzzle, only a rotation and translation are needed; a reflection is not required since this is trivially handled for the pieces of the jigsaw puzzle, by simply identifying the front-back of each piece.

Finally, we recall here the *cluster-synchronization* problem detailed in Section 3.1, which can be solved and analyzed with similar tools as those used in the present article. Imagine a instance where one combines two different sets of puzzles, and aims to disentangle the two sets of pieces. In a real-world application, this could arise, for instance, at an archaeological site where two broken froscoes are uncovered, both shattered into pieces which got mixed up, and the goal is to separate out the two froscoes. When one considers a pair of pieces, it is not known whether they both come from same fresco, or from two different ones. A probabilistic model for this setting is one where there are two groups of angles  $\mathcal{C}_1 = (\alpha_1, \alpha_2, \dots, \alpha_{n_1})$ , and  $\mathcal{C}_2 = (\beta_{n_1+1}, \beta_{n_1+2}, \dots, \beta_{n_1+n_2})$ , and an available measurement between two angles  $\alpha_i$  and  $\beta_j$  from two different clusters is completely uniformly distributed in the unit circle, while a measurement between a pair of angles from the same cluster is a noisy version of the offset  $\alpha_i - \alpha_j$  or  $\beta_i - \beta_j$ . This model is reminiscent of the standard stochastic block model, well studied in the clustering and network analysis literatures.

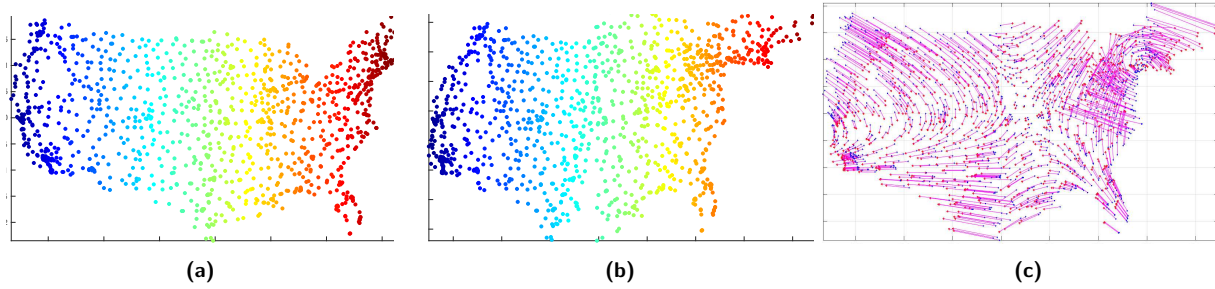
**Setup** We consider the following instance of the graph realization problem, where one aims to recover the coordinates of two clouds of points,  $X$  and  $Y \in \mathbb{R}^{n \times 2}$ , indexed by the same set of nodes. More specifically, the goal is to identify two clouds of points,  $p_1^{(X)}, \dots, p_n^{(X)} \in \mathbb{R}^d$  and  $p_1^{(Y)}, \dots, p_n^{(Y)} \in \mathbb{R}^2$ , given information on subgraph embeddings from either confirmation. In other words, to each node we associate a patch (eg, its one-hop neighborhood), which may assume one of two possible embeddings  $\mathcal{P}_i^{(X)}$  and  $\mathcal{P}_i^{(Y)}$ . Whenever a pair of patches  $\mathcal{P}_i$  and  $\mathcal{P}_j$  are compared and aligned, one of three things may happen.

- (with probability  $p_1$ ) Both subgraph embeddings are of type- $X$ , and we align  $\mathcal{P}_i^{(X)}$  and  $\mathcal{P}_j^{(X)}$ , with  $\{i, j\} \in E_1$ ;
- (with probability  $p_2$ ) both subgraph embeddings are of type- $Y$ , and we align  $\mathcal{P}_i^{(Y)}$  and  $\mathcal{P}_j^{(Y)}$  with  $\{i, j\} \in E_2$ ;
- (with probability  $\eta = 1 - p_1 - p_2$ ) one embedding is of type- $X$  and the other one of type- $Y$ , for example  $\mathcal{P}_i^{(X)}$  and  $\mathcal{P}_j^{(Y)}$  (with  $\{i, j\} \in E_W$ , the edge set of the bad subgraph) are aligned, and the resulting

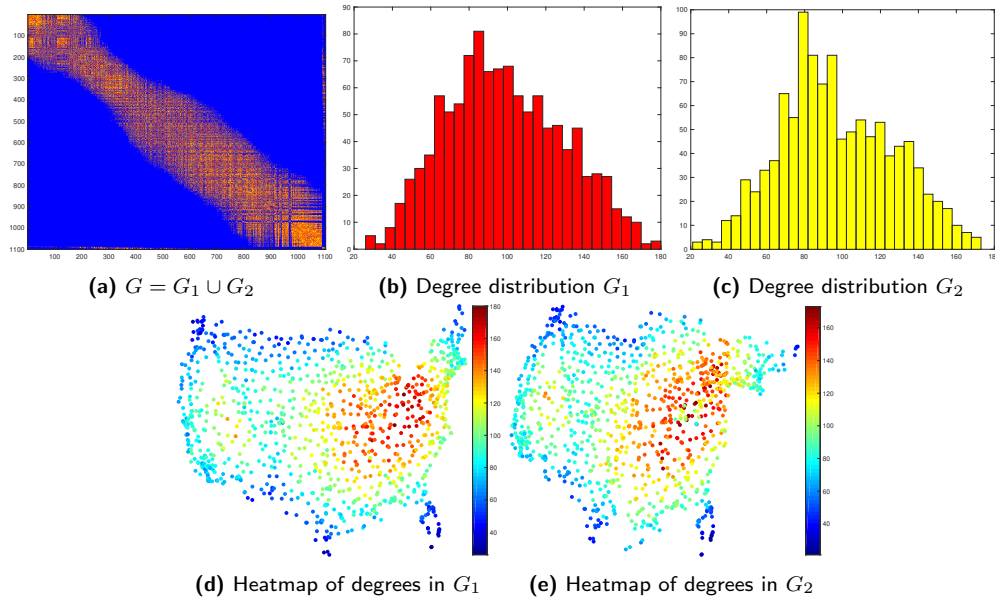
outlier measurement is a random number uniformly distributed in the unit circle, and thus completely uninformative.

The above is nothing but an instance of the bi-synchronization problem considered in this paper. The task then becomes to identify the two subgraphs  $G_1$  and  $G_2$ , which can ultimately facilitate the recovery of the two global embeddings  $X$  and  $Y$ .

In our experiments, we let the point cloud  $Y$  be to a non-rigid transformation of point cloud  $X$ , and the goal is to recover both embeddings. Figure 14 shows the two embeddings  $X$  (a) and  $Y$  (b), while (c) shows the Procrustes alignment of two point clouds, together with the displacement error bars, clearly showing the two embeddings are not congruent. Both sets of patches take their 2D coordinates from their corresponding ground truth embedding. For example, in the noiseless case, the 2D embedding of patch  $P_i^{(X)}$  is given by the coordinates within  $X$  of the nodes contained in  $P_i^{(X)}$ . When noise is added, indexed by  $\sigma$ , we perturb both 2-D coordinates by additive Gaussian noise with mean 0 and variance  $\sigma^2$ .

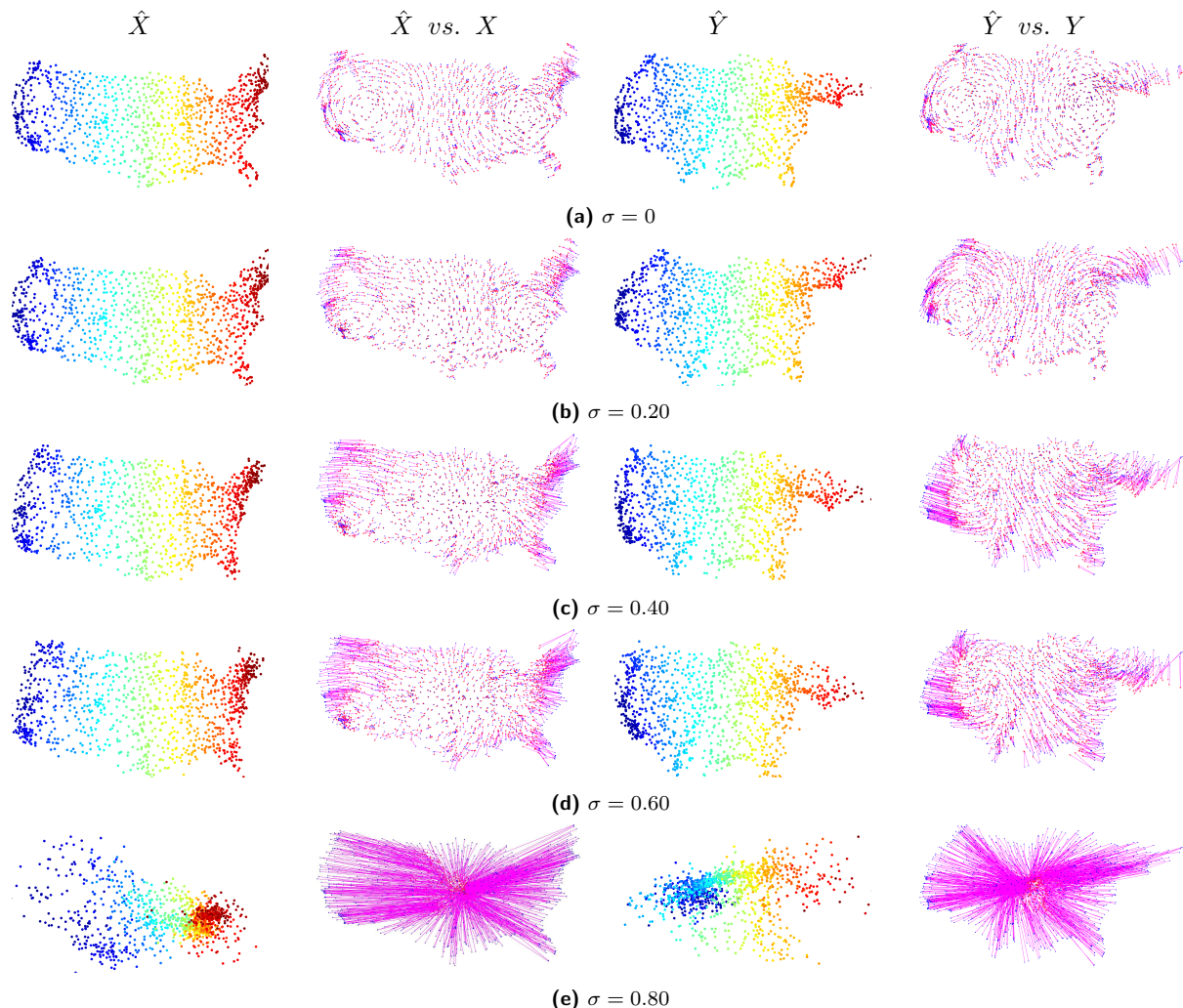


**Figure 14:** Clean data, two non-congruent embeddings of the US graph: (a) Original US map (b) Non-rigid transformation of the original data (shear transformation, followed by a rotation and scaling of the Midwest and East Coast). (c) Procrustes alignment of (a) and (b) showing the displacements between the two embeddings.



**Figure 15:** (a) Adjacency matrix of the measurement graph  $G$ , highlighting the decomposition  $G = G_1$  (red) +  $G_2$  (yellow). Plots (b) and (c) show the histogram of degrees in  $G_1$  and  $G_2$ . The embeddings of  $X$  and  $Y$  are colored by their corresponding degree, i.e., each node  $i$  in the embedding (that lies at the center of patch  $P_i$ ) is colored by the number of other patches it overlaps with.

The top left plot of Figure 15 shows the adjacency matrix of the measurement graph  $G$ , highlighting the decomposition  $G = G_1 \cup G_2$ , with edges of  $G_1$ , respectively  $G_2$ , are colored in red, respectively yellow. The rightmost plots in the top row of the same Figure 15 show the histogram of degrees in the patch graphs  $G_1$  and  $G_2$  motivating the normalized version  $D^{-1}H$ .



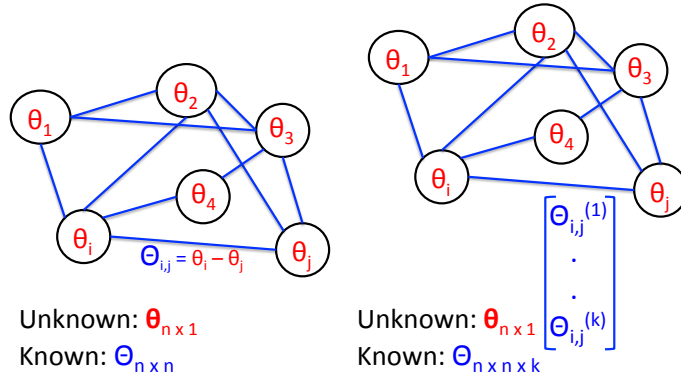
**Figure 16:** Recovery of the two estimated embeddings  $\hat{X}$  and  $\hat{Y}$  shown in the first and third columns, where the coloring is given by the longitude. The second column shows the Procrustes alignment between the recovered embedding  $\hat{X}$  (red) and the ground truth  $X$  (blue), together with the displacement error bars. The fourth column shows a similar visualization for  $\hat{Y}$  and the ground truth  $Y$ . We vary the noise level  $\sigma$  used in the perturbations of the local patch embeddings in both ensembles, across the following set of values  $\sigma \in \{0, 0.2, 0.4, 0.6, 0.8\}$ .

## 8 Conclusion and future directions

This paper considers an extension of the classical synchronization problem (of recovering a group of angles given a subset of noisy pairwise angle offsets) to the setting when *multiple* latent groups of angles exist. We propose probabilistic generative models for  $k$ -synchronization, along with spectral algorithms which we theoretically analyze, under a suitably defined notion of  $\delta$ -orthogonality. There are various further extensions one could pursue.

**List-synchronization** An interesting direction, ongoing work similar in spirit to the setup explored in the present paper, is that of *list-synchronization*, where there exists a single latent group of angles,  $\theta_1, \dots, \theta_n$ , and for each available edge  $\{i, j\}$  in the measurement graph  $G$ , one has available not only a single measurement proxy for the offset  $\theta_i - \theta_j$ , but rather a list of  $k$  possible offsets, of which only a single one is correct (or approximately correct), while the remaining  $k - 1$  elements of the edge list are outliers, as depicted in Figure 17. This setting arises for instance in the GRP problem, where in the high-noise regime, one may consider multiple candidates for the pairwise angle offset, when aligning pairs of overlapping patches via the method of choice (see for example Section 6 within [19]). This problem is also related to the MAX-2-LIN mod  $L$

problem [2, 5], for maximizing the number of satisfied linear equations mod  $L$  with exactly two variables in each equation, also known to be an instance of unique games [25].



**Figure 17:** Left: classical angular synchronization; Right: an instance of the  $k$ -list synchronization problem, where the goal is to recover a single group of  $n$  angles, from a small subset of pairwise noisy offsets  $\theta_i - \theta_j$ ; however, there are now multiple candidates for a given pairwise offset, only one of which is correct (or approximately correct).

**SDP relaxation** Extending the theoretical results obtained in this paper for the spectral relaxations to the setting of an SDP relaxation is another direction of interest. The advantage of the SDP relaxation stems from the fact that one can explicitly enforce the unit magnitude constraints in the formulation, by adding ones on the main diagonal of the Gram matrix in (2.4). A potentially interesting approach, albeit computationally expensive, is to consider a SDP which optimizes over  $k$  complex-valued Hermitian positive semidefinite matrices  $\Upsilon^{(1)}, \Upsilon^{(2)}, \dots, \Upsilon^{(k)}$ , while enforcing the unit magnitude diagonal constraints individually for each such matrix.

**Sparse regime of the measurement graph** Another direction worth exploring pertains to improving existing algorithms for the sparse regimes. For example, in graph clustering problems, it is well known that spectral methods underperform in the very sparse regime, where edges exist with probability  $\Theta(\frac{1}{n})$ . Therein, graph regularization techniques have been proven to be effective both on the theoretical and experimental fronts; for eg., see regularization in the sparse regime for the signed clustering problem [21, 13, 1]. Hence, a natural question to ask is what would be a corresponding graph regularization technique for very sparse measurement graphs in the group synchronization setting. For example, would a simple local augmentation step be effective to this end, where for each 2-path  $v_i, v_j, v_k$ , with edges  $\{i, j\}$  and  $\{j, k\}$ , one can add an estimated measurement for edge  $\{i, k\}$  stemming from the fact that summing up over triangles yields  $0 \bmod 2\pi$ . Alternatively, for each missing edge  $\{i, k\}$ , one could consider all open triangles  $i, j', k$  (with edges  $\{i, j'\}$  and  $\{j', k\}$ ), and each such triangle could cast a vote via the above procedure for the missing edge.

**Cluster-synchronization** Finally, we recall here the setup of cluster-synchronization, detailed in the last paragraph of Section 3.1, on the recovery of two groups of angles (of potentially different sizes  $n_1, n_2$ ) from a noisy subset of pairwise offsets captured in a matrix of size  $(n_1 + n_2) \times (n_1 + n_2)$ . Therein, the noise model imparts a cluster structure on the measurement matrix, as depicted in Figure 3. Proposing spectral and SDP algorithms for the recovery of both the latent clusters and the groups of angles themselves, is an interesting future direction worth pursuing, well motivated by applications within the ranking and recommender systems literature.

Additional research directions include extending the  $k$ -synchronization pipeline to compact groups beyond  $\text{SO}(2)$ , and exploring the iterative (potentially re-weighted) scheme in the context of classical group synchronization ( $k = 1$ ), as detailed in Remark 3 and motivated by Figure 11.

## References

- [1] A. A. AMINI, A. CHEN, P. J. BICKEL, AND E. LEVINA, *Pseudo-likelihood methods for community detection in large sparse networks*, The Annals of Statistics, 41 (2013), p. 2097–2122.

- [2] G. ANDERSSON, L. ENGBRETTSEN, AND J. HÅSTAD, *A new way of using semidefinite programming with applications to linear equations mod  $p$* , Journal of Algorithms, 39 (2001), pp. 162–204.
- [3] M. ARIE-NACHIMSON, S. Z. KOVALSKY, I. KEMELMACHER-SHLIZERMAN, A. SINGER, AND R. BASRI, *Global motion estimation from point matches*, in 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission, 2012, pp. 81–88.
- [4] A. S. BANDEIRA, N. BOUMAL, AND A. SINGER, *Tightness of the maximum likelihood semidefinite relaxation for angular synchronization*, Mathematical Programming, 163 (2017), pp. 145–167.
- [5] A. S. BANDEIRA, Y. CHEN, R. R. LEDERMAN, AND A. SINGER, *Non-unique games over compact groups and orientation estimation in cryo-EM*, Inverse Problems, 36 (2020), p. 064002.
- [6] A. S. BANDEIRA, A. PERRY, AND A. S. WEIN, *Notes on computational-to-statistical gaps: predictions using statistical physics*, 2018.
- [7] A. S. BANDEIRA AND R. VAN HANDEL, *Sharp nonasymptotic bounds on the norm of random matrices with independent entries*, Ann. Probab., 44 (2016), pp. 2479–2506.
- [8] A.-L. BARABASI AND R. ALBERT, *Emergence of scaling in random networks*, Science, 286 (1999), pp. 509–512.
- [9] A. I. BARVINOK, *Problems of distance geometry and convex properties of quadratic maps*, Discrete & Computational Geometry, 13 (1995), pp. 189–202.
- [10] P. BISWAS, T. C. LIAN, T. C. WANG, AND Y. YE, *Semidefinite programming based algorithms for sensor network localization*, ACM Transactions on Sensor Networks, 2 (2006), pp. 188–220.
- [11] N. BOUMAL, *Nonconvex phase synchronization*, SIAM Journal on Optimization, 26 (2016), pp. 2355–2377.
- [12] S. BURER AND R. MONTEIRO, *Local minima and convergence in low-rank semidefinite programming*, Mathematical Programming, 103 (2005), pp. 427–444.
- [13] K. CHAUDHURI, F. CHUNG, AND A. TSIATAS, *Spectral clustering of graphs with general degrees in the extended planted partition model*, in 25th Annual Conference on Learning Theory, vol. 23 of Proceedings of Machine Learning Research, Edinburgh, Scotland, 2012, JMLR Workshop and Conference Proceedings, pp. 35.1–35.23.
- [14] R. CONNELLY, *On generic global rigidity*, Applied Geometry and Discrete Mathematics, 4 (1991), pp. 147–155.
- [15] ———, *Generic global rigidity*, Discrete Comput. Geom, 33 (2005), pp. 549–563.
- [16] T. F. COX AND M. A. A. COX, *Multidimensional Scaling*, Monographs on Statistics and Applied Probability 88, Chapman & Hall/CRC, Boca Raton, FL, 2001.
- [17] M. CUCURINGU, *Sync-Rank: Robust Ranking, Constrained Ranking and Rank Aggregation via Eigenvector and Semidefinite Programming Synchronization*, IEEE Transactions on Network Science and Engineering, 3 (2016), pp. 58–79.
- [18] M. CUCURINGU, H. LI, H. SUN, AND L. ZANETTI, *Hermitian matrices for clustering directed graphs: insights and applications*, AISTATS 2020.
- [19] M. CUCURINGU, Y. LIPMAN, AND A. SINGER, *Sensor network localization by eigenvector synchronization over the Euclidean group*, ACM Trans. Sen. Netw., 8 (2012), pp. 19:1–19:42.
- [20] M. CUCURINGU, A. SINGER, AND D. COWBURN, *Eigenvector synchronization, graph rigidity and the molecule problem*, Information and Inference, 1 (2012), pp. 21–67.



- [21] M. CUCURINGU, A. V. SINGH, D. SULEM, AND H. TYAGI, *Regularized spectral methods for clustering signed networks*, arXiv:2011.01737, (2020).
- [22] D’ASPREMONT, A., CUCURINGU, M., AND TYAGI, H., *Ranking and synchronization from pairwise measurements via SVD*, accepted to Journal of Machine Learning Research, arXiv:1906.02746, (2020).
- [23] C. DAVIS AND W. M. KAHAN, *The rotation of eigenvectors by a perturbation. iii*, SIAM Journal on Numerical Analysis, 7 (1970), pp. 1–46.
- [24] M. FANUEL AND J. SUYKENS, *Deformed laplacians and spectral ranking in directed networks*, Applied and Computational Harmonic Analysis, (2017).
- [25] U. FEIGE AND L. LOVÁSZ, *Two-prover one-round proof systems: Their power and their problems (extended abstract)*, in Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing, STOC ’92, New York, NY, USA, 1992, Association for Computing Machinery, p. 733–744.
- [26] S. J. GORTLER, A. D. HEALY, AND D. P. THURSTON, *Characterizing generic global rigidity*, AMERICAN JOURNAL OF MATHEMATICS, 4 (2010), p. 897.
- [27] C. GOTSMAN AND Y. KOREN, *Distributed graph layout for sensor networks*, in Proc. Intl. Symp. Graph Drawing, 2004, pp. 273–284.
- [28] B. HENDRICKSON, *The molecule problem: Exploiting structure in global optimization*, SIAM J Optimization, 5 (1995), pp. 835–857.
- [29] J. KUNEGIS, S. SCHMIDT, A. LOMMATZSCH, J. LERNER, E. W. D. LUCA, AND S. ALBAYRAK, *Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization*, SIAM, 2010, pp. 559–570.
- [30] R.-C. LI, *Relative perturbation theory: Ii. eigenspace and singular subspace variations*, SIAM Journal on Matrix Analysis and Applications, 20 (1998), pp. 471–492.
- [31] A. PERRY, A. S. WEIN, A. S. BANDEIRA, AND A. MOITRA, *Optimality and sub-optimality of PCA for spiked random matrices and synchronization*, arXiv preprint arXiv:1609.05573, (2016).
- [32] ———, *Message-Passing Algorithms for Synchronization Problems over Compact Groups*, Communications on Pure and Applied Mathematics, 71 (2018), pp. 2275–2322.
- [33] Y. SHKOLNISKY AND A. SINGER, *Viewing direction estimation in cryo-em using synchronization*, SIAM journal on imaging sciences, 5 (2012), pp. 1088–1110.
- [34] A. SINGER, *Angular synchronization by eigenvectors and semidefinite programming*, Appl. Comput. Harmon. Anal., 30 (2011), pp. 20–36.
- [35] A. SINGER AND H. T. WU, *Vector diffusion maps and the Connection Laplacian*, Communications on Pure and Applied Mathematics, (2012).
- [36] A. SINGER, Z. ZHAO, Y. SHKOLNISKY, AND R. HADANI, *Viewing angle classification of cryo-electron microscopy images using eigenvectors*, SIAM Journal on Imaging Sciences, 4 (2011), pp. 723–759.
- [37] E. SIZIKOVA AND T. FUNKHOUSER, *Wall painting reconstruction using a genetic algorithm*, in Proceedings of the 14th Eurographics Workshop on Graphics and Cultural Heritage, GCH ’16, Eurographics Association, 2016, p. 83–91.
- [38] M. TUBAISHAT AND S. MADRIA, *Sensor networks: An overview*, in IEEE Potentials, vol. 22, 2003, pp. 20–23.
- [39] T. TZENEVA, *Global alignment of multiple 3-d scans using eigenvector synchronization*, 2011.
- [40] R. VERSHYNIN, *Introduction to the non-asymptotic analysis of random matrices*, Cambridge University Press, 2012, pp. 210–268.

- [41] H. WEYL, *Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung)*, Mathematische Annalen, 71 (1912), pp. 441–479.
- [42] S. YU, *Angular embedding: A robust quadratic criterion*, IEEE Trans. Pattern Anal. Mach. Intell., 34 (2012), pp. 158–173.
- [43] S. X. YU, *Angular embedding: From jarring intensity differences to perceived luminance*, in 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20–25 June 2009, Miami, Florida, USA, IEEE Computer Society, 2009, pp. 2302–2309.
- [44] Y. YU, T. WANG, AND R. J. SAMWORTH, *A useful variant of the davis–kahan theorem for statisticians*, Biometrika, 102 (2015), pp. 315–323.
- [45] Y. ZHONG AND N. BOUMAL, *Near-optimal bounds for phase synchronization*, SIAM Journal on Optimization, 28 (2018), pp. 989–1016.

## A Perturbation analysis

Let  $A \in \mathbb{C}^{n \times n}$  be Hermitian with eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  and corresponding eigenvectors  $v_1, v_2, \dots, v_n \in \mathbb{C}^n$ . Let  $\tilde{A} = A + W$  be a perturbed version of  $A$ , with the perturbation matrix  $W \in \mathbb{C}^{n \times n}$  being Hermitian. Let us denote the eigenvalues of  $\tilde{A}$  and  $W$  by  $\tilde{\lambda}_1 \geq \dots \geq \tilde{\lambda}_n$  and  $\epsilon_1 \geq \epsilon_2 \geq \dots \geq \epsilon_n$  respectively.

To begin with, we would like to quantify the perturbation of the eigenvalues of  $\tilde{A}$  with respect to the eigenvalues of  $A$ . Weyl’s inequality [41] is a very useful result in this regard.

**Theorem 3** (Weyl’s Inequality [41]). *For each  $i = 1, \dots, n$ , it holds that*

$$\lambda_i + \epsilon_n \leq \tilde{\lambda}_i \leq \lambda_i + \epsilon_1. \quad (\text{A.1})$$

*In particular, this implies that  $\tilde{\lambda}_i \in [\lambda_i - \|W\|_2, \lambda_i + \|W\|_2]$ .*

One can also quantify the perturbation of the subspace spanned by eigenvectors of  $A$ , this was established by Davis and Kahan [23]. Before introducing the theorem, we need some definitions. Let  $U, \tilde{U} \in \mathbb{C}^{n \times k}$  (for  $k \leq n$ ) have orthonormal columns respectively and let  $\sigma_1 \geq \dots \geq \sigma_k$  denote the singular values of  $U^* \tilde{U}$ . Then, the  $k$  principal angles between  $U, \tilde{U}$  are defined as  $\theta_i := \cos^{-1}(\sigma_i)$  for  $1 \leq i \leq k$ , with each  $\theta_i \in [0, \pi/2]$ . It is usual to denote  $\Theta(U, \tilde{U}) \in \mathbb{R}^{k \times k}$  to be a diagonal matrix with  $\theta_i$  as its entries (arranged in decreasing order), and to define  $\sin \Theta(U, \tilde{U}) \in \mathbb{C}^{k \times k}$  entrywise. Denoting  $\|\cdot\|$  to be any unitarily invariant norm (Frobenius, spectral, etc.), the following relation holds (see for eg., [30, Lemma 2.2] and ensuing discussion).

$$\|\sin \Theta(U, \tilde{U})\| = \|(I - \tilde{U} \tilde{U}^*) U_1\|.$$

With the above notation in mind, we now introduce a version of the Davis-Kahan theorem taken from [44, Theorem 1] (see also [30, Theorem 3.2]).

**Theorem 4** (Davis-Kahan). *Fix  $1 \leq r \leq s \leq n$ , let  $d = s - r + 1$ , and let  $V = (v_r, \dots, v_s) \in \mathbb{C}^{n \times d}$  and  $\tilde{V} = (\tilde{v}_r, \dots, \tilde{v}_s) \in \mathbb{C}^{n \times d}$ . Write*

$$\varepsilon = \inf \left\{ |\hat{\lambda} - \lambda| : \lambda \in [\lambda_s, \lambda_r], \hat{\lambda} \in (-\infty, \tilde{\lambda}_{s+1}] \cup [\tilde{\lambda}_{r-1}, \infty) \right\},$$

*where we define  $\tilde{\lambda}_0 = \infty$  and  $\tilde{\lambda}_{n+1} = -\infty$  and assume that  $\varepsilon > 0$ . Then*

$$\|\sin \Theta(U, \tilde{U})\| \leq \frac{\|W\|}{\delta}.$$

For instance, if  $r = s = j$ , we obtain the following useful result

$$\sin \Theta(\tilde{v}_j, v_j) = \|(I - v_j v_j^*) \tilde{v}_j\|_2 \leq \frac{\|W\|_2}{\min \left\{ |\tilde{\lambda}_{j-1} - \lambda_j|, |\tilde{\lambda}_{j+1} - \lambda_j| \right\}}. \quad (\text{A.2})$$

## B Useful concentration inequalities

### B.1 Sum of sub-Gaussian random variables

Recall the following Hoeffding-type inequality for sums of independent sub-Gaussian random variables.

**Proposition 3.** [40, Proposition 5.10] *Let  $X_1, \dots, X_n$  be independent centered sub-Gaussian random variables and let  $K = \max_i \|X_i\|_{\psi_2}$ . Then for every  $\mathbf{a} \in \mathbb{R}^n$ , and every  $t \geq 0$ , we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^n a_i X_i\right| \geq t\right) \leq e \cdot \exp\left(-\frac{c't^2}{K^2\|\mathbf{a}\|_2^2}\right), \quad (\text{B.1})$$

where  $c' > 0$  is an absolute constant.

### B.2 Spectral norm of random matrices

We will make use of the following result for bounding the spectral norm of symmetric matrices with independent, centered random variables.

**Theorem 5** ([7, Corollary 3.12, Remark 3.13]). *Let  $X$  be an  $n \times n$  symmetric matrix whose entries  $X_{ij}$  ( $i \leq j$ ) are independent, centered random variables. There exists, for any  $0 < \varepsilon \leq 1/2$ , a universal constant  $c_\varepsilon$  such that for every  $t \geq 0$ ,*

$$\mathbb{P}(\|X\|_2 \geq (1 + \varepsilon)2\sqrt{2}\tilde{\sigma} + t) \leq n \exp\left(-\frac{t^2}{c_\varepsilon \tilde{\sigma}_*^2}\right), \quad (\text{B.2})$$

where

$$\tilde{\sigma} := \max_i \sqrt{\sum_j \mathbb{E}[X_{ij}^2]}, \quad \tilde{\sigma}_* := \max_{i,j} \|X_{ij}\|_\infty.$$

Note that it suffices to employ upper bound estimates on  $\tilde{\sigma}, \tilde{\sigma}_*$  in (B.2).

## C Proofs from Section 4

### C.1 Proof of Proposition 1

Recall that for a random variable  $X$ , its sub-Gaussian norm  $\|X\|_{\psi_2}$  is defined as

$$\|X\|_{\psi_2} := \sup_{p \geq 1} \frac{(\mathbb{E}|X|^p)^{1/p}}{\sqrt{p}}. \quad (\text{C.1})$$

Moreover,  $X$  is a sub-Gaussian random variable if  $\|X\|_{\psi_2}$  is finite. For instance, a bounded random variable  $X$  with  $|X| \leq M$  is sub-Gaussian with  $\|X\|_{\psi_2} \leq M$  (see [40, Example 5.8]).

*Proof of Proposition 1.* Observe that

$$\langle z_1, z_2 \rangle = \frac{1}{n} \sum_{i=1}^n \exp(t(\beta_i - \alpha_i)) = \frac{1}{n} \left[ \sum_{i=1}^n \underbrace{\cos(\beta_i - \alpha_i)}_{R_i} + t \sum_{i=1}^n \underbrace{\sin(\beta_i - \alpha_i)}_{X_i} \right]. \quad (\text{C.2})$$

$(R_i)_i$  are zero-mean iid sub-Gaussian random variables with  $\|R_i\|_{\psi_2} \leq 1$ . Using standard concentration inequalities for sums of sub-Gaussian random variables (see Appendix B.1), there exists a constant  $c' > 0$  such that

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n R_i\right| > t\right) \leq e \cdot \exp(-c't^2 n), \quad (\text{C.3})$$

for  $t \geq 0$ . An identical statement holds for the random variables  $(X_i)_i$ . Using the union bound, we then have with probability at least  $1 - 2e \exp(-c't^2 n)$ , that  $|\langle z_1, z_2 \rangle| \leq 2t$  holds. Plugging  $t = \delta/2$  yields the stated bound.  $\square$

## C.2 Proof of Lemma 1

1. (Bounds on  $\tilde{\lambda}_1, \tilde{\lambda}_2$ ) For this part, we note<sup>3</sup> that

$$\begin{aligned} \mathbb{E}[H]^2 &= n^2 p_1^2 \lambda^2 z_1 z_1^* + n^2 p_2^2 \lambda^2 z_2 z_2^* + 2n^2 p_1 p_2 \lambda^2 z_1 (z_1^* z_2) z_2^* \\ \Rightarrow \text{tr}(\mathbb{E}[H]^2) &= n^2 p_1^2 \lambda^2 + n^2 p_2^2 \lambda^2 + 2n^2 p_1 p_2 \lambda^2 |\langle z_1, z_2 \rangle|^2 = \tilde{\lambda}_1^2 + \tilde{\lambda}_2^2, \end{aligned} \quad (\text{C.4})$$

where we used the linearity of the trace operator. Moreover, we also have that  $\text{tr}(\mathbb{E}[H]) = \tilde{\lambda}_1 + \tilde{\lambda}_2 = np_1 \lambda + np_2 \lambda$ . Using this with (C.4), we readily obtain  $\tilde{\lambda}_1 \tilde{\lambda}_2 = n^2 p_1 p_2 \lambda^2 (1 - |\langle z_1, z_2 \rangle|^2)$  leading to

$$\begin{aligned} (np_1 \lambda + np_2 \lambda - \tilde{\lambda}_2) \tilde{\lambda}_2 &= n^2 p_1 p_2 \lambda^2 (1 - |\langle z_1, z_2 \rangle|^2) \\ \Leftrightarrow \tilde{\lambda}_2^2 - \tilde{\lambda}_2 (np_1 \lambda + np_2 \lambda) + n^2 p_1 p_2 \lambda^2 (1 - |\langle z_1, z_2 \rangle|^2) &= 0 \end{aligned} \quad (\text{C.5})$$

$$\begin{aligned} \Leftrightarrow \tilde{\lambda}_2 &= \frac{np_1 \lambda + np_2 \lambda - \sqrt{(np_1 \lambda + np_2 \lambda)^2 - 4n^2 p_1 p_2 \lambda^2 (1 - |\langle z_1, z_2 \rangle|^2)}}{2} \\ &= \frac{np_1 \lambda + np_2 \lambda - \sqrt{(np_1 \lambda - np_2 \lambda)^2 + 4n^2 p_1 p_2 \lambda^2 |\langle z_1, z_2 \rangle|^2}}{2}, \end{aligned} \quad (\text{C.6})$$

$$\text{and } \tilde{\lambda}_1 = \frac{np_1 \lambda + np_2 \lambda + \sqrt{(np_1 \lambda - np_2 \lambda)^2 + 4n^2 p_1 p_2 \lambda^2 |\langle z_1, z_2 \rangle|^2}}{2}. \quad (\text{C.7})$$

Note that we neglect the other root in (C.5) since  $\tilde{\lambda}_1 \geq \tilde{\lambda}_2$ . In fact,  $\tilde{\lambda}_1 > \tilde{\lambda}_2$  since  $p_1 > p_2$ . The bounds  $\tilde{\lambda}_2 \leq np_2 \lambda$  and  $\tilde{\lambda}_1 \geq np_1 \lambda$  follow from the fact that  $|\langle z_1, z_2 \rangle| \geq 0$ . The bounds in (4.9), (4.10) hold if  $|\langle z_1, z_2 \rangle| \leq \delta$  holds.

2. (Bounds on  $|\langle z_1, \tilde{v}_1 \rangle|^2, |\langle z_2, \tilde{v}_2 \rangle|^2$ ) To bound  $|\langle z_1, \tilde{v}_1 \rangle|^2$ , we note that

$$\begin{aligned} z_1^* \mathbb{E}[H] z_1 &= np_1 \lambda + np_2 \lambda |\langle z_1, z_2 \rangle|^2 \\ \Rightarrow \tilde{\lambda}_1 |\langle z_1, \tilde{v}_1 \rangle|^2 + \tilde{\lambda}_2 |\langle z_1, \tilde{v}_2 \rangle|^2 &= np_1 \lambda + np_2 \lambda |\langle z_1, z_2 \rangle|^2 \\ \Rightarrow \tilde{\lambda}_1 |\langle z_1, \tilde{v}_1 \rangle|^2 + \tilde{\lambda}_2 (1 - |\langle z_1, \tilde{v}_1 \rangle|^2) &= np_1 \lambda + np_2 \lambda |\langle z_1, z_2 \rangle|^2 \end{aligned} \quad (\text{C.8})$$

$$\begin{aligned} \Rightarrow (\tilde{\lambda}_1 - \tilde{\lambda}_2) |\langle z_1, \tilde{v}_1 \rangle|^2 &= np_1 \lambda - \tilde{\lambda}_2 + np_2 \lambda |\langle z_1, z_2 \rangle|^2 \\ &\geq np_1 \lambda - np_2 \lambda \end{aligned} \quad (\text{C.9})$$

$$\Rightarrow |\langle z_1, \tilde{v}_1 \rangle|^2 \geq \frac{p_1 - p_2}{\sqrt{(p_1 - p_2)^2 + 4p_1 p_2 \delta^2}}. \quad (\text{C.10})$$

In (C.8), we used the fact  $1 = \|z_1\|_2^2 = |\langle z_1, \tilde{v}_1 \rangle|^2 + |\langle z_1, \tilde{v}_2 \rangle|^2$ , since  $z_1$  lies in  $\text{span}(\tilde{v}_1, \tilde{v}_2)$ . In (C.9), we used  $|\langle z_1, z_2 \rangle|^2 \geq 0$  and also  $\tilde{\lambda}_2 \leq np_2 \lambda$  (shown earlier). In (C.10), we used the bound  $\tilde{\lambda}_1 - \tilde{\lambda}_2 \leq n\lambda \sqrt{(p_1 - p_2)^2 + 4p_1 p_2 \delta^2}$ , which follows easily from (4.10) and (4.9). Finally, the bound on  $|\langle z_2, \tilde{v}_2 \rangle|^2$  follows in a similar fashion, but by considering  $z_2^* \mathbb{E}[H] z_2$  instead.

## C.3 Proof of Lemma 2

Recall from (4.7) that  $H = \mathbb{E}(H) + R = (n\lambda p_1 z_1 z_1^* + n\lambda p_2 z_2 z_2^*) + R$ . As a consequence of Weyl's inequality for symmetric perturbation of symmetric matrices (see Theorem 3 in Appendix A), we have

$$\lambda_i \in [\tilde{\lambda}_i - \Delta, \tilde{\lambda}_i + \Delta]; \quad i = 1, \dots, n. \quad (\text{C.11})$$

In order to obtain the stated bounds, we will now invoke the Davis-Kahan theorem (see Theorem 4 in Appendix A).

<sup>3</sup>This argument for computing the expressions for  $\tilde{\lambda}_1, \tilde{\lambda}_2$  is based on a post by Davide Giraudo on Mathematics Stack Exchange, question 112186 (<https://math.stackexchange.com/questions/112186/eigenvalues-of-a-sum-of-rank-one-matrices>)

1. **(Bound in (4.12))** From Theorem 4, we see that if  $|\tilde{\lambda}_1 - \lambda_2| > 0$ , then it holds that

$$\|(I - v_1 v_1^*) \tilde{v}_1\|_2 \leq \frac{\|R\|_2}{|\tilde{\lambda}_1 - \lambda_2|} \leq \frac{\Delta}{|\tilde{\lambda}_1 - \lambda_2|}. \quad (\text{C.12})$$

Using (C.11) and the bound  $\tilde{\lambda}_2 \leq np_2\lambda$  from Lemma 1, we obtain  $\lambda_2 \leq np_2\lambda + \Delta$ . Furthermore, using the bound  $\tilde{\lambda}_1 \geq np_1\lambda$  from Lemma 1, we obtain  $\tilde{\lambda}_1 - \lambda_2 \geq np_1\lambda - np_2\lambda - \Delta$ . Therefore if  $\Delta < np_1\lambda - np_2\lambda$ , it follows from (C.12) that

$$\|(I - v_1 v_1^*) \tilde{v}_1\|_2 \leq \frac{\Delta}{np_1\lambda - np_2\lambda - \Delta}.$$

Since  $\|(I - v_1 v_1^*) \tilde{v}_1\|_2^2 = 1 - |\langle \tilde{v}_1, v_1 \rangle|^2$ , we thus obtain the stated bound in (4.12).

2. **(Bound in (4.13))** We now apply Theorem 4 to bound  $\|(I - v_2 v_2^*) \tilde{v}_2\|_2$ . If

$$\min \left\{ |\lambda_1 - \tilde{\lambda}_2|, |\lambda_3 - \tilde{\lambda}_2| \right\} > 0,$$

it follows from Theorem 4 that

$$\|(I - v_2 v_2^*) \tilde{v}_2\|_2 \leq \frac{\|R\|_2}{\min \left\{ |\lambda_1 - \tilde{\lambda}_2|, |\lambda_3 - \tilde{\lambda}_2| \right\}}. \quad (\text{C.13})$$

We saw earlier that  $\lambda_1 - \tilde{\lambda}_2 \geq np_1\lambda - np_2\lambda - \Delta > 0$  if  $\Delta < np_1\lambda - np_2\lambda$ . Also, since  $\lambda_3 \in [-\Delta, \Delta]$  (from (C.11)), it follows that  $\tilde{\lambda}_2 - \lambda_3 \geq np_2\lambda - \Delta > 0$  if  $\Delta < np_2\lambda$ . Plugging these observations in (C.13), and using the fact  $\|(I - v_2 v_2^*) \tilde{v}_2\|_2^2 = 1 - |\langle \tilde{v}_2, v_2 \rangle|^2$ , the stated bound in (4.13) follows readily.

## C.4 Proof of Lemma 3

Since  $\|R\|_2 \leq \|\text{Re}(R)\|_2 + \|\text{Im}(R)\|_2$ , therefore we will bound the terms  $\|\text{Re}(R)\|_2, \|\text{Im}(R)\|_2$  individually w.h.p, which will then imply a bound on  $\|R\|_2$ .

1. **(Bounding  $\|\text{Re}(R)\|_2$ .)** The entries of  $\text{Re}(R)$  above and on the diagonal are independent centered random variables. Therefore we can bound  $\|\text{Re}(R)\|_2$  using a result of Bandeira et al. [7, Corollary 3.12, Remark 3.13] for bounding the spectral norm of such symmetric random matrices (stated as Theorem 5 in Appendix B.2). To do so, it suffices to know upper bounds on the quantities

$$\tilde{\sigma} := \max_i \sqrt{\sum_j \mathbb{E}[\text{Re}(R)_{ij}^2]}, \quad \tilde{\sigma}_* := \max_{i,j} \|\text{Re}(R)_{ij}\|_\infty. \quad (\text{C.14})$$

Starting with  $\tilde{\sigma}$ , we note that

$$\begin{aligned} \mathbb{E}[\text{Re}(R)_{ij}^2] &= p_1\lambda[\cos(\alpha_i - \alpha_j) - p_1\lambda \cos(\alpha_i - \alpha_j) - p_2\lambda \cos(\beta_i - \beta_j)]^2 \\ &\quad + p_2\lambda[\cos(\beta_i - \beta_j) - p_1\lambda \cos(\alpha_i - \alpha_j) - p_2\lambda \cos(\beta_i - \beta_j)]^2 \\ &\quad + (1 - p_1 - p_2)\lambda(\mathbb{E}[\cos N_{ij} - p_1\lambda \cos(\alpha_i - \alpha_j) - p_2\lambda \cos(\beta_i - \beta_j)]^2) \\ &\quad + (1 - \lambda)[p_1\lambda \cos(\alpha_i - \alpha_j) + p_2\lambda \cos(\beta_i - \beta_j)]^2 \\ &\leq 2p_1\lambda[(1 - p_1\lambda)^2 + (p_2\lambda)^2] + 2p_2\lambda[(1 - p_2\lambda)^2 + (p_1\lambda)^2] \\ &\quad + (1 - p_1 - p_2)\lambda \left[ \frac{1}{2} + (p_1\lambda + p_2\lambda)^2 \right] + (1 - \lambda)[p_1\lambda + p_2\lambda]^2 \\ &= C(\lambda, p_1, p_2). \end{aligned}$$

Hence we obtain  $\tilde{\sigma} \leq \sqrt{C(\lambda, p_1, p_2)n}$ . Using Theorem 5 with  $t = 2\sqrt{2}\sqrt{C(\lambda, p_1, p_2)n}$ , and with (4.15) in mind, we then obtain for any  $\varepsilon \geq 0$ , that there exists a universal constant  $c_\varepsilon > 0$  such that

$$\mathbb{P} \left( \left\{ \|\text{Re}(R)\|_2 \geq (2 + \varepsilon)2\sqrt{2}\sqrt{C(\lambda, p_1, p_2)n} \right\} \right) \leq n \exp \left( -\frac{8C(\lambda, p_1, p_2)n}{\tilde{\sigma}(\lambda, p_1, p_2)^2 c_\varepsilon} \right). \quad (\text{C.15})$$

2. (**Bounding  $\|\text{Im}(R)\|_2$ .)** We can write  $\text{Im}(R) = W - W^T$ , where  $W_{ij} = 0$  for  $i \geq j$ . The random variables  $(W_{ij})_{i < j}$  are independent and zero-mean. Since  $\|\text{Im}(R)\|_2 \leq 2\|W\|_2$ , therefore we will focus on bounding  $\|W\|_2$ . Now, while  $W$  is not symmetric, one can easily verify that for the following symmetric matrix

$$\widetilde{W} = \begin{bmatrix} 0 & W \\ W^T & 0 \end{bmatrix} \in \mathbb{R}^{2n \times 2n},$$

we have  $\|\widetilde{W}\|_2 = \|W\|_2$ . Since  $\widetilde{W}$  has independent entries on and above the main diagonal, we can again use the result in [7, Corollary 3.12, Remark 3.13] to bound its spectral norm. To this end, we need to bound the quantities

$$\tilde{\sigma} := \max_i \sqrt{\sum_j \mathbb{E}[\widetilde{W}_{ij}^2]}, \quad \tilde{\sigma}_* := \max_{i,j} \|\widetilde{W}_{ij}\|_\infty.$$

In an identical manner as before, one can readily verify that  $\tilde{\sigma} \leq \sqrt{C(\lambda, p_1, p_2)n}$ . Moreover, it holds that  $\tilde{\sigma}_* \leq \bar{\sigma}(\lambda, p_1, p_2)$ . Applying Theorem 5 on  $\widetilde{W}$  with  $t = 2\sqrt{2}\sqrt{C(\lambda, p_1, p_2)n}$ , and with the above observations in mind, we finally obtain for any  $\varepsilon \geq 0$  that there exists a universal constant  $c_\varepsilon > 0$  such that

$$\mathbb{P}\left(\|\text{Im}(R)\|_2 \geq (2 + \varepsilon)4\sqrt{2}\sqrt{C(\lambda, p_1, p_2)n}\right) \leq 2n \exp\left(-\frac{16C(\lambda, p_1, p_2)n}{\bar{\sigma}(\lambda, p_1, p_2)^2 c_\varepsilon}\right). \quad (\text{C.16})$$

Finally, the statement in the Lemma follows by applying the union bound to (C.15), (C.16).

## C.5 Proof of Proposition 2

We know that  $1 - |\langle x, \bar{x} \rangle|^2 = \|(I - \bar{x}\bar{x}^*)x\|_2^2$ . Moreover,

$$\|(I - \bar{x}\bar{x}^*)x\|_2 \leq \|x - yy^*x\|_2 + \|(yy^* - \bar{x}\bar{x}^*)x\|_2 \leq \sqrt{\varepsilon} + \|yy^* - \bar{x}\bar{x}^*\|_2. \quad (\text{C.17})$$

By observing that

$$\|yy^* - \bar{x}\bar{x}^*\|_2 \leq \|yy^* - \bar{x}\bar{x}^*\|_F = \sqrt{2 - 2|\langle \bar{x}, y \rangle|^2} \leq \sqrt{2\varepsilon},$$

and plugging this in (C.17), we obtain the stated bound.

## D Proofs from Section 5

### D.1 Proof of Lemma 4

- Let us consider  $j = 1$ . Starting with the definition of  $\tilde{\lambda}_1$ , we obtain

$$\tilde{\lambda}_1 := \max_{\|x\|_2=1} x^* \mathbb{E}[H]x \geq z_1^* \mathbb{E}[H]z_1 = np_1\lambda + \sum_{j=2}^k np_j\lambda |\langle z_1, z_j \rangle|^2 \geq np_1\lambda. \quad (\text{D.1})$$

In order to lower bound  $|\langle z_1, \tilde{v}_1 \rangle|^2$ , let us first note that

$$\tilde{\lambda}_1 = \tilde{v}^* \mathbb{E}[H]\tilde{v} = np_1\lambda |\langle z_1, \tilde{v}_1 \rangle|^2 + \sum_{j=2}^k np_j\lambda |\langle z_j, \tilde{v}_1 \rangle|^2.$$

Therefore,  $\tilde{\lambda}_1$  can be bounded as

$$\tilde{\lambda}_1 \leq np_1\lambda |\langle z_1, \tilde{v}_1 \rangle|^2 + \sum_{j=2}^k np_j\lambda |\langle z_j, \tilde{v}_1 \rangle|^2 = (np_1\lambda - np_2\lambda) |\langle z_1, \tilde{v}_1 \rangle|^2 + np_2\lambda \sum_{j=1}^k |\langle z_j, \tilde{v}_1 \rangle|^2. \quad (\text{D.2})$$

We will now upper bound  $\sum_{j=1}^k |\langle z_j, \tilde{v}_1 \rangle|^2$  as follows. Denoting  $Z = [z_1 \ z_2 \ \dots \ z_k] \in \mathbb{C}^{n \times k}$ , clearly

$$\sum_{j=1}^k |\langle z_j, \tilde{v}_1 \rangle|^2 = \sum_{j=1}^k \tilde{v}_1^* z_j z_j^* \tilde{v}_1 = \tilde{v}_1^* Z Z^* \tilde{v}_1 \leq \|Z\|_2^2 \leq 1 + (k-1)\delta, \quad (\text{D.3})$$

where in the last step we used<sup>4</sup> Gershgorin's theorem along with the  $\delta$ -orthonormality of  $z_i$ 's. Using (D.3) in (D.2) along with (D.1), we get

$$\begin{aligned} np_1\lambda &\leq (np_1\lambda - np_2\lambda)|\langle z_1, \tilde{v}_1 \rangle|^2 + np_2\lambda(1 + (k-1)\delta) \\ \iff |\langle z_1, \tilde{v}_1 \rangle|^2 &\geq \frac{np_1\lambda - np_2\lambda - np_2\lambda(k-1)\delta}{np_1\lambda - np_2\lambda} \\ &= 1 - \frac{p_2(k-1)\delta}{p_1 - p_2} = 1 - \psi(\delta). \end{aligned} \quad (\text{D.4})$$

Note that,  $|\langle z_1, \tilde{v}_1 \rangle|^2 > 1/2$  if  $\delta < \frac{p_1 - p_2}{2p_2(k-1)}$ . Finally, we obtain from (D.4) the bound

$$\tilde{\lambda}_1 \leq (np_1\lambda - np_2\lambda) + np_2\lambda(1 + (k-1)\delta) = np_1\lambda + \underbrace{np_2\lambda(k-1)\delta}_{u_1(\delta)}.$$

2. Let us now handle  $\tilde{\lambda}_m, \tilde{v}_m$  for  $m > 1$ . Assume that for  $1 \leq j \leq m-1$ ,  $\tilde{\lambda}_j, \tilde{v}_j$  satisfy

$$\tilde{\lambda}_j \leq np_j\lambda + u_j(\delta), \quad |\langle \tilde{v}_j, z_j \rangle|^2 \geq 1 - \psi_j(\delta), \quad (\text{D.5})$$

where  $\delta \in [0, 1]$  satisfies

$$\delta \leq \sqrt{2\psi_j(\delta)} \leq 1/2; \quad \psi_1(\delta) \leq \dots \leq \psi_{m-1}(\delta). \quad (\text{D.6})$$

We will show that (D.5), (D.6) together imply the bounds

$$np_m\lambda - l_m(\delta) \leq \tilde{\lambda}_m \leq np_m\lambda + u_m(\delta), \quad |\langle \tilde{v}_m, z_m \rangle|^2 \geq 1 - \psi_m(\delta).$$

To begin with, it is easy to see that

$$\begin{aligned} \tilde{\lambda}_m &= \tilde{v}_m^* \mathbb{E}[H] \tilde{v}_m = \max_{\|x\|_2=1} x^* (\mathbb{E}[H] - \sum_{j=1}^{m-1} \tilde{\lambda}_j \tilde{v}_j \tilde{v}_j^*) x \\ &\geq z_m^* (\mathbb{E}[H] - \sum_{j=1}^{m-1} \tilde{\lambda}_j \tilde{v}_j \tilde{v}_j^*) z_m \\ &= z_m^* \mathbb{E}[H] z_m - \sum_{j=1}^{m-1} \tilde{\lambda}_j |\langle \tilde{v}_j, z_m \rangle|^2. \end{aligned} \quad (\text{D.7})$$

One can also verify that  $z_m^* \mathbb{E}[H] z_m \geq np_m\lambda$ . Moreover,  $\tilde{\lambda}_j \leq np_j\lambda + u_j(\delta)$  for  $j = 1, \dots, m-1$  by assumption. In order to upper bound  $|\langle \tilde{v}_j, z_m \rangle|^2$ , let us first note that  $|\langle \tilde{v}_j, z_m \rangle|^2 = 1 - \|z_m - \tilde{v}_j \tilde{v}_j^* z_m\|_2^2$ . We can lower bound  $\|z_m - \tilde{v}_j \tilde{v}_j^* z_m\|_2$  as follows.

$$\begin{aligned} \|z_m - \tilde{v}_j \tilde{v}_j^* z_m\|_2 &\geq \|z_m - z_j z_j^* z_m\|_2 - \|z_j z_j^* z_m - \tilde{v}_j \tilde{v}_j^* z_m\|_2 \\ &\geq \sqrt{1 - \delta^2} - \sqrt{2\psi_j(\delta)} \\ &\geq 1 - \delta - \sqrt{2\psi_j(\delta)} \quad (\text{Since } \sqrt{1 - \delta^2} \geq 1 - \delta) \\ &\geq 1 - 2\sqrt{2\psi_j(\delta)} \quad \left( \text{Since } \delta \leq \sqrt{2\psi_j(\delta)} \right) \\ &\geq 0 \quad (\text{Since } \sqrt{2\psi_j(\delta)} \leq 1/2). \end{aligned}$$

<sup>4</sup> $\|Z\|_2^2 = \lambda_{\max}(Z^*Z)$  where  $\lambda_{\max}(Z^*Z)$  denotes the largest eigenvalue of  $Z^*Z$ . The magnitude of each off-diagonal entry of  $Z^*Z$  is bounded by  $\delta$ , while each diagonal entry is equal to 1. Hence  $\lambda_{\max}(Z^*Z) \leq 1 + (k-1)\delta$  by Gershgorin's theorem.

In the second inequality, we used  $\|z_m - z_j z_j^* z_m\|_2 = \sqrt{1 - |\langle z_j, z_m \rangle|^2} \geq \sqrt{1 - \delta^2}$ . For each  $j = 1, \dots, m-1$ , this leads to

$$|\langle \tilde{v}_j, z_m \rangle|^2 \leq 1 - \left(1 - 2\sqrt{2\psi_j(\delta)}\right)^2 = \left(2\sqrt{2\psi_j(\delta)}\right) \left(2 - 2\sqrt{2\psi_j(\delta)}\right) \leq 4\sqrt{2\psi_j(\delta)}.$$

Using these bounds in (D.7), we can lower bound  $\tilde{\lambda}_m$  as

$$\tilde{\lambda}_m \geq np_m \lambda - 4 \sum_{j=1}^{m-1} (np_j \lambda + u_j(\delta)) \sqrt{2\psi_j(\delta)}. \quad (\text{D.8})$$

Let us now proceed to upper bound  $\tilde{\lambda}_m$ . To this end, we first note that

$$\begin{aligned} \tilde{\lambda}_m &= \tilde{v}_m^* \mathbb{E}[H] \tilde{v}_m = \sum_{i=1}^{m-1} np_i \lambda |\langle z_i, \tilde{v}_m \rangle|^2 + np_m \lambda |\langle z_m, \tilde{v}_m \rangle|^2 + \sum_{j=m+1}^k np_j \lambda |\langle z_j, \tilde{v}_m \rangle|^2 \\ &\leq \sum_{i=1}^{m-1} np_i \lambda |\langle z_i, \tilde{v}_m \rangle|^2 + np_m \lambda |\langle z_m, \tilde{v}_m \rangle|^2 + np_{m+1} \lambda \sum_{j=m+1}^k |\langle z_j, \tilde{v}_m \rangle|^2 \\ &\leq \sum_{i=1}^{m-1} (np_i \lambda - np_{m+1} \lambda) |\langle z_i, \tilde{v}_m \rangle|^2 + (np_m \lambda - np_{m+1} \lambda) |\langle z_m, \tilde{v}_m \rangle|^2 \\ &\quad + np_{m+1} \lambda (1 + (k-1)\delta). \end{aligned} \quad (\text{D.9})$$

We can upper bound  $|\langle z_i, \tilde{v}_m \rangle|^2$  for each  $i = 1, \dots, m-1$  by noting that  $|\langle z_i, \tilde{v}_m \rangle|^2 = 1 - \|z_i - \tilde{v}_m \tilde{v}_m^* z_i\|_2^2$ , and also

$$\begin{aligned} \|z_i - \tilde{v}_m \tilde{v}_m^* z_i\|_2 &\geq \|\tilde{v}_i \tilde{v}_i^* z_i - \tilde{v}_m \tilde{v}_m^* z_i\|_2 - \|z_i - \tilde{v}_i \tilde{v}_i^* z_i\|_2 \\ &= \sqrt{z_i^* (\tilde{v}_i \tilde{v}_i^* + \tilde{v}_m \tilde{v}_m^*) z_i} - \|z_i - \tilde{v}_i \tilde{v}_i^* z_i\|_2 \\ &= \left( \underbrace{|\langle z_i, \tilde{v}_i \rangle|^2}_{\geq 1 - \psi_i(\delta)} + \underbrace{|\langle z_i, \tilde{v}_m \rangle|^2}_{\geq 0} \right)^{1/2} - \underbrace{\left(1 - |\langle z_i, \tilde{v}_i \rangle|^2\right)^{1/2}}_{\leq \psi_i(\delta)} \\ &\geq \sqrt{1 - \psi_i(\delta)} - \sqrt{\psi_i(\delta)} \\ &\geq 1 - \psi_i(\delta) - \sqrt{\psi_i(\delta)} \quad (\text{Since } 1 - \psi_i(\delta) \in [0, 1]) \\ &\geq 1 - 2\sqrt{\psi_i(\delta)} \quad (\text{Since } \psi_i(\delta) \leq \sqrt{\psi_i(\delta)}) \\ &\geq 0 \quad (\text{Since } \sqrt{2\psi_i(\delta)} \leq 1/2). \end{aligned} \quad (\text{D.10})$$

Using (D.10) we thus obtain

$$|\langle z_i, \tilde{v}_m \rangle|^2 = 1 - (1 - 2\sqrt{\psi_i(\delta)})^2 = 2\sqrt{\psi_i(\delta)}(2 - 2\sqrt{\psi_i(\delta)}) \leq 4\sqrt{\psi_i(\delta)}. \quad (\text{D.11})$$

Plugging (D.11) in (D.9) then leads to

$$\tilde{\lambda}_m \leq 4 \sum_{i=1}^{m-1} (np_i \lambda - np_{m+1} \lambda) \sqrt{\psi_i(\delta)} + (np_m \lambda - np_{m+1} \lambda) |\langle z_m, \tilde{v}_m \rangle|^2 + np_{m+1} \lambda (1 + (k-1)\delta). \quad (\text{D.12})$$



Using (D.8),(D.12) and by re-arranging terms, we obtain

$$\begin{aligned}
|\langle z_m, \tilde{v}_m \rangle|^2 &\geq \frac{np_m\lambda - np_{m+1}\lambda(1 + (k-1)\delta) - 4\sum_{i=1}^{m-1}[(np_i\lambda + u_i(\delta))\sqrt{2\psi_i(\delta)} + (np_i\lambda - np_{m+1}\lambda)\sqrt{\psi_i(\delta)}]}{np_m\lambda - np_{m+1}\lambda} \\
&= 1 - \left( \frac{p_{m+1}(k-1)\delta + 4\sum_{i=1}^{m-1} \left[ \left( p_i + \frac{u_i(\delta)}{n\lambda} \right) \sqrt{2\psi_i(\delta)} + (p_i - p_{m+1})\sqrt{\psi_i(\delta)} \right]}{p_m - p_{m+1}} \right) \\
&\geq 1 - \left( \frac{p_{m+1}(k-1)\delta + 4\sqrt{2\psi_{m-1}(\delta)}\sum_{i=1}^{m-1} \left[ 2p_i + \frac{u_i(\delta)}{n\lambda} - p_{m+1} \right]}{p_m - p_{m+1}} \right), \tag{D.13}
\end{aligned}$$

where in the last inequality, we used our assumption  $\psi_1(\delta) \leq \dots \leq \psi_{m-1}(\delta)$ . Now from the definition of  $u_i(\delta)$ , we have for each  $i = 1, \dots, m-1$  that

$$\begin{aligned}
\frac{u_i(\delta)}{n\lambda} &= 4\sum_{j=1}^{i-1} (p_j - p_{i+1})\sqrt{\psi_j(\delta)} + p_{i+1}(k-1)\delta \\
&\leq \sqrt{2}\sum_{j=1}^{i-1} (p_j - p_{i+1}) + \frac{1}{2}p_{i+1}(k-1) \quad \left( \text{since } \delta \leq \sqrt{2\psi_j(\delta)} \leq 1/2 \right) \\
&\leq \sqrt{2}(i-1)(p_1 - p_m) + \frac{1}{2}p_2(k-1). \tag{D.14}
\end{aligned}$$

Using (D.14) in (D.13), we obtain

$$\begin{aligned}
|\langle z_m, \tilde{v}_m \rangle|^2 &\geq 1 - \left( \frac{p_{m+1}(k-1)\delta + 4\sqrt{2\psi_{m-1}(\delta)}\sum_{i=1}^{m-1} \left[ 2p_i + \sqrt{2}(i-1)(p_1 - p_m) + \frac{p_2(k-1)}{2} - p_{m+1} \right]}{p_m - p_{m+1}} \right) \\
&\geq 1 - \left( \frac{p_{m+1}(k-1)\delta + 4\sqrt{2\psi_{m-1}(\delta)}[2S_{m-1} + \sqrt{2}(m-1)(m-2)(p_1 - p_m) + \frac{(m-1)}{2}(p_2(k-1) - 2p_{m+1})]}{p_m - p_{m+1}} \right) \\
&\geq 1 - \left( \frac{p_{m+1}(k-1)\sqrt{2} + 4\sqrt{2}[2S_{m-1} + \sqrt{2}(m-1)(m-2)(p_1 - p_m) + \frac{(m-1)}{2}(p_2(k-1) - 2p_{m+1})]}{p_m - p_{m+1}} \right) \sqrt{\psi_{m-1}(\delta)} \\
&= 1 - \underbrace{C_m \sqrt{\psi_{m-1}(\delta)}}_{\psi_m(\delta)},
\end{aligned}$$

where in the penultimate step, we used  $\delta \leq \sqrt{2\psi_{m-1}(\delta)}$ . To conclude, we now establish the bounds on  $\tilde{\lambda}_m$  stated in the Lemma. The upper bound follows from (D.12) by using  $|\langle z_m, \tilde{v}_m \rangle|^2 \leq 1$ . Indeed,

$$\begin{aligned}
\tilde{\lambda}_m &\leq 4\sum_{i=1}^{m-1} (np_i\lambda - np_{m+1}\lambda)\sqrt{\psi_i(\delta)} + (np_m\lambda - np_{m+1}\lambda) + np_{m+1}\lambda(1 + (k-1)\delta) \\
&= np_m\lambda + \left( 4\sum_{i=1}^{m-1} (np_i\lambda - np_{m+1}\lambda)\sqrt{\psi_i(\delta)} + np_{m+1}\lambda(k-1)\delta \right) \\
&= np_m\lambda + u_m(\delta).
\end{aligned}$$

To lower bound  $\tilde{\lambda}_m$ , we start with (D.8) and the assumption  $\psi_1(\delta) \leq \dots \leq \psi_{m-1}(\delta)$ . This leads to

$$\begin{aligned}\tilde{\lambda}_m &\geq np_m\lambda - 4\sqrt{2\psi_{m-1}(\delta)} \left( \sum_{j=1}^{m-1} (np_j\lambda + u_j(\delta)) \right) \\ &\geq np_m\lambda - 4\sqrt{2\psi_{m-1}(\delta)} (n\lambda S_{m-1} + \sum_{j=1}^{m-1} u_j(\delta)).\end{aligned}$$

Using the definition of  $u_j(\delta)$ , we then obtain

$$\begin{aligned}\tilde{\lambda}_m &\geq np_m\lambda - 4\sqrt{2\psi_{m-1}(\delta)} \left( n\lambda S_{m-1} + \sum_{j=1}^{m-1} \left( 2 \sum_{i=1}^{j-1} (np_i - np_{j+1})\lambda\sqrt{\psi_i(\delta)} + np_{j+1}\lambda(k-1)\delta \right) \right) \\ &\geq np_m\lambda - 4\sqrt{2\psi_{m-1}(\delta)} \left( n\lambda S_{m-1} + \sum_{j=1}^{m-1} \left( \sqrt{2}(j-1)(np_1 - np_{j+1})\lambda + np_{j+1}\lambda(k-1) \right) \right) \\ &\geq np_m\lambda - 4\sqrt{2\psi_{m-1}(\delta)} \left( n\lambda S_{m-1} + \sqrt{2}(m-1)(m-2)(np_1 - np_m)\lambda + np_2\lambda(k-1)(m-1) \right) \\ &= np_m\lambda - l_m(\delta).\end{aligned}$$

## D.2 Proof of Lemma 5

Recall that  $H = \mathbb{E}(H) + R = \sum_{i=1}^k n\lambda p_i z_i z_i^* + R$ . Also recall that via Weyl's inequality (see Theorem 3 in Appendix A), it holds

$$\lambda_i \in [\tilde{\lambda}_i - \Delta, \tilde{\lambda}_i + \Delta]; \quad i = 1, \dots, n. \quad (\text{D.15})$$

As in Lemma 2, we will obtain the bounds via an application of the Davis-Kahan theorem (see Theorem 4 in Appendix A). From now, we will assume that  $\delta$  satisfies the conditions stated in Lemma 4. Before proceeding to the analysis, it will be helpful to use the following upper bound on  $u_j(\delta)$  for each  $j = 2, \dots, k$ .

$$\begin{aligned}u_j(\delta) &= 4 \sum_{i=1}^{j-1} (np_i - np_{j+1})\lambda\sqrt{\psi_i(\delta)} + np_{j+1}\lambda(k-1)\delta \\ &\leq 4(j-1)(np_1 - np_{j+1})\lambda\sqrt{\psi_{j-1}(\delta)} + np_{j+1}\lambda(k-1)\delta \\ &\leq n\lambda\sqrt{\psi_{j-1}(\delta)}[4(j-1)(p_1 - p_{j+1}) + p_{j+1}(k-1)].\end{aligned} \quad (\text{D.16})$$

1. **(Bounding  $|\langle \tilde{v}_1, v_1 \rangle|^2$ )** From Theorem 4, we observe that

$$\|(I - v_1 v_1^*)\tilde{v}_1\|_2 \leq \frac{\Delta}{|\tilde{\lambda}_1 - \lambda_2|}, \quad (\text{D.17})$$

if  $|\tilde{\lambda}_1 - \lambda_2| > 0$ . Since  $\tilde{\lambda}_1 \geq np_1\lambda$  (see Lemma 4), and  $\lambda_2 \leq \tilde{\lambda}_2 + \Delta \leq np_2\lambda + u_2(\delta) + \Delta$  (from (D.15) and Lemma 4), it follows that

$$\begin{aligned}\tilde{\lambda}_1 - \lambda_2 &\geq \tilde{\lambda}_1 - \tilde{\lambda}_2 - \Delta \\ &\geq n\lambda(p_1 - p_2) - u_2(\delta) - \Delta \\ &\geq n\lambda(p_1 - p_2) - n\lambda\sqrt{\psi_1(\delta)}(4(p_1 - p_3) + p_3(k-1)) - \Delta.\end{aligned}$$

Hence for  $\mu \in [0, 1/2]$ , we have that  $\tilde{\lambda}_1 - \lambda_2 \geq (1 - \mu)n\lambda(p_1 - p_2) > 0$  if

$$\Delta \leq \mu \frac{n\lambda(p_1 - p_2)}{2}, \quad \psi_1(\delta) \leq \mu^2 \left( \frac{p_1 - p_2}{8(p_1 - p_3) + 2p_3(k-1)} \right)^2. \quad (\text{D.18})$$

In particular, it then follows from (D.17), that

$$\|(I - v_1 v_1^*) \tilde{v}_1\|_2 \leq \frac{\mu}{2(1 - \mu)}.$$

Finally, the stated bound on  $|\langle \tilde{v}_1, v_1 \rangle|^2$  follows from the identity  $\|(I - v_1 v_1^*) \tilde{v}_1\|_2^2 = 1 - |\langle \tilde{v}_1, v_1 \rangle|^2$ .

2. **(Bounding  $|\langle \tilde{v}_j, v_j \rangle|^2$  for  $1 < j < k$ )** From Theorem 4, we observe that

$$\|(I - v_j v_j^*) \tilde{v}_j\|_2 \leq \frac{\Delta}{\min \{ |\tilde{\lambda}_j - \lambda_{j+1}|, |\tilde{\lambda}_j - \lambda_{j-1}| \}}, \quad (\text{D.19})$$

if  $|\tilde{\lambda}_j - \lambda_{j+1}|, |\tilde{\lambda}_j - \lambda_{j-1}| > 0$ .

Let us first establish conditions under which  $|\tilde{\lambda}_j - \lambda_{j-1}| > 0$  holds. Since  $\lambda_{j-1} \geq \tilde{\lambda}_{j-1} - \Delta$ , it follows that

$$\begin{aligned} \lambda_{j-1} - \tilde{\lambda}_j &\geq \tilde{\lambda}_{j-1} - \tilde{\lambda}_j - \Delta \\ &\geq n\lambda(p_{j-1} - p_j) - l_{j-1}(\delta) - u_j(\delta) - \Delta \quad (\text{From Lemma 4}) \\ &= n\lambda(p_{j-1} - p_j) - 4n\lambda\sqrt{2\psi_{j-2}(\delta)(S_{j-2} + \sqrt{2}(j-2)(j-3)(p_1 - p_{j-1}) + p_2(k-1)(j-2))} \\ &\quad - n\lambda\sqrt{\psi_{j-1}(\delta)[4(j-1)(p_1 - p_{j+1}) + p_{j+1}(k-1)]} - \Delta \quad (\text{From Lemma 4, (D.16)}) \\ &\geq n\lambda(p_{j-1} - p_j) - n\lambda\sqrt{\psi_{j-1}(\delta)E_j} - \Delta \quad (\text{Since } \psi_{j-2}(\delta) \leq \psi_{j-1}(\delta)), \end{aligned}$$

where, Hence for  $\mu \in [0, 1/2]$ , we have that  $\lambda_{j-1} - \tilde{\lambda}_j \geq (1 - \mu)n\lambda(p_{j-1} - p_j) > 0$  if

$$\Delta \leq \mu \frac{n\lambda(p_{j-1} - p_j)}{2}, \quad \psi_{j-1}(\delta) \leq \mu^2 \left( \frac{p_{j-1} - p_j}{2E_j} \right)^2. \quad (\text{D.20})$$

Now let us establish conditions under which  $|\tilde{\lambda}_j - \lambda_{j+1}| > 0$  holds. Since  $\lambda_{j+1} \geq \tilde{\lambda}_{j+1} + \Delta$ , it follows that

$$\begin{aligned} \tilde{\lambda}_j - \lambda_{j+1} &\geq \tilde{\lambda}_j - \tilde{\lambda}_{j+1} - \Delta \\ &\geq n\lambda(p_j - p_{j+1}) - l_j(\delta) - u_{j+1}(\delta) - \Delta \quad (\text{From Lemma 4}) \\ &= n\lambda(p_j - p_{j+1}) - 4n\lambda\sqrt{2\psi_{j-1}(\delta)(S_{j-1} + \sqrt{2}(j-1)(j-2)(p_1 - p_j) + p_2(k-1)(j-1))} \\ &\quad - n\lambda\sqrt{\psi_j(\delta)[4j(p_1 - p_{j+2}) + p_{j+2}(k-1)]} - \Delta \quad (\text{From Lemma 4, (D.16)}) \\ &\geq n\lambda(p_j - p_{j+1}) - n\lambda\sqrt{\psi_j(\delta)(4\sqrt{2}S_{j-1} + 8(j-1)(j-2)(p_1 - p_j) + 4\sqrt{2}p_2(k-1)(j-1))} \\ &\quad + 4j(p_1 - p_{j+2}) + p_{j+2}(k-1) - \Delta \quad (\text{Since } \psi_{j-1}(\delta) \leq \psi_j(\delta)) \\ &= n\lambda(p_j - p_{j+1}) - n\lambda\sqrt{\psi_j(\delta)E_{j+1}} - \Delta \end{aligned}$$

Hence for  $\mu \in [0, 1/2]$  we have that  $\tilde{\lambda}_j - \lambda_{j+1} \geq (1 - \mu)n\lambda(p_j - p_{j+1}) > 0$  if

$$\Delta \leq \mu \frac{n\lambda(p_j - p_{j+1})}{2}, \quad \psi_j(\delta) \leq \mu^2 \left( \frac{p_j - p_{j+1}}{2E_{j+1}} \right)^2. \quad (\text{D.21})$$

Therefore from (D.20), (D.21), it follows that

$$\lambda_{j-1} - \tilde{\lambda}_j, \tilde{\lambda}_j - \lambda_{j+1} \geq n\lambda(1 - \mu) \min \{ p_{j-1} - p_j, p_j - p_{j+1} \} > 0$$

holds if

$$\Delta < \mu \frac{n\lambda}{2} \min \{ (p_j - p_{j+1}), (p_{j-1} - p_j) \}; \quad \psi_{j-1}(\delta) < \mu^2 \left( \frac{p_{j-1} - p_j}{2E_j} \right)^2; \quad \psi_j(\delta) < \mu^2 \left( \frac{p_j - p_{j+1}}{2E_{j+1}} \right)^2.$$

In particular, it then follows from (D.19) that

$$\begin{aligned} \|(I - v_j v_j^*) \tilde{v}_j\|_2 &\leq \frac{\Delta}{\min \left\{ n\lambda(p_{j-1} - p_j) - n\lambda\sqrt{\psi_{j-1}(\delta)}E_j - \Delta, n\lambda(p_j - p_{j+1}) - n\lambda\sqrt{\psi_j(\delta)}E_{j+1} - \Delta \right\}} \\ &\leq \frac{\mu}{2(1-\mu)}. \end{aligned}$$

The bound on  $|\langle \tilde{v}_j, v_j \rangle|^2$  now follows from the identity  $\|(I - v_j v_j^*) \tilde{v}_j\|_2^2 = 1 - |\langle \tilde{v}_j, v_j \rangle|^2$ .

3. **(Bounding  $|\langle \tilde{v}_k, v_k \rangle|^2$ )** From Theorem 4, we observe that

$$\|(I - v_k v_k^*) \tilde{v}_k\|_2 \leq \frac{\Delta}{\min \left\{ |\tilde{\lambda}_k - \lambda_{k+1}|, |\tilde{\lambda}_k - \lambda_{k-1}| \right\}}, \quad (\text{D.22})$$

if  $|\tilde{\lambda}_k - \lambda_{k+1}|, |\tilde{\lambda}_k - \lambda_{k-1}| > 0$ .

Let us first see when  $|\tilde{\lambda}_k - \lambda_{k-1}| > 0$  holds. By proceeding in an identical manner as before (for showing when  $\lambda_{j-1} - \tilde{\lambda}_j > 0$  holds for  $1 < j < k$ ), we readily obtain  $\lambda_{k-1} - \tilde{\lambda}_k \geq n\lambda(p_{k-1} - p_k) - n\lambda\sqrt{\psi_{k-1}(\delta)}E_k - \Delta$ , where  $E_k$  is as defined in (5.6) by plugging  $j = k$ . Hence for  $\mu \in [0, 1/2]$ , we have that  $\lambda_{k-1} - \tilde{\lambda}_k \geq (1 - \mu)n\lambda(p_{k-1} - p_k) > 0$  holds if

$$\Delta \leq \mu \frac{n\lambda(p_{k-1} - p_k)}{2}, \quad \psi_{k-1}(\delta) \leq \mu^2 \left( \frac{p_{k-1} - p_k}{2E_k} \right)^2. \quad (\text{D.23})$$

Now let us establish conditions under which  $|\tilde{\lambda}_k - \lambda_{k+1}| > 0$  holds. Since  $\lambda_{k+1} \leq \tilde{\lambda}_{k+1} + \Delta = \Delta$ , it follows that

$$\begin{aligned} \tilde{\lambda}_k - \lambda_{k+1} &\geq \tilde{\lambda}_k - \Delta \\ &\geq n\lambda p_k - l_k(\delta) - \Delta \quad (\text{From Lemma 4}) \\ &= n\lambda p_k - 4n\lambda\sqrt{2\psi_{k-1}(\delta)}(S_{k-1} + \sqrt{2}(k-1)(k-2)(p_1 - p_k) \\ &\quad + p_2(k-1)^2) - \Delta \quad (\text{From Lemma 4, (D.16)}) \\ &= n\lambda p_k - n\lambda\sqrt{\psi_{k-1}(\delta)}\tilde{E} - \Delta. \end{aligned}$$

Finally, for  $\mu \in [0, 1/2]$ , we see that  $\tilde{\lambda}_k - \lambda_{k+1} \geq n\lambda(1 - \mu)p_k > 0$  holds if

$$\Delta \leq \mu \frac{n\lambda p_k}{2}, \quad \psi_{k-1}(\delta) \leq \mu^2 \left( \frac{p_k}{2\tilde{E}} \right)^2. \quad (\text{D.24})$$

Therefore from (D.23),(D.24), it follows that

$$\lambda_{k-1} - \tilde{\lambda}_k, \tilde{\lambda}_k - \lambda_{k+1} \geq (1 - \mu)n\lambda \min \{p_{k-1} - p_k, p_k\} > 0,$$

provided that

$$\Delta \leq \mu \frac{n\lambda}{2} \min \{p_k, (p_{k-1} - p_k)\}, \quad \psi_{k-1}(\delta) \leq \mu^2 \min \left\{ \left( \frac{p_{k-1} - p_k}{2E_k} \right)^2, \left( \frac{p_k}{2\tilde{E}} \right)^2 \right\}.$$

In particular, it then follows from (D.22) that

$$\begin{aligned} \|(I - v_k v_k^*) \tilde{v}_k\|_2 &\leq \frac{\Delta}{\min \left\{ n\lambda(p_{k-1} - p_k) - n\lambda\sqrt{\psi_{k-1}(\delta)}E_k - \Delta, n\lambda p_k - n\lambda\sqrt{\psi_{k-1}(\delta)}\tilde{E} - \Delta \right\}} \\ &\leq \frac{\mu}{2(1-\mu)}. \end{aligned}$$

Finally, the bound on  $|\langle \tilde{v}_k, v_k \rangle|^2$  is obtained in the same manner as explained for previous cases.

### D.3 Proof of Lemma 6

As the proof follows the same arguments as that in the proof of Lemma 3, we only point out the main differences.

1. (**Bounding**  $\|\mathbf{Re}(R)\|_2$ .) Recall the definition of  $\tilde{\sigma}$ ,  $\tilde{\sigma}_*$  as in (C.14). From (5.12), we have that  $\tilde{\sigma}_* \leq \bar{\sigma}(\lambda, p_1, \dots, p_k)$ . Moreover, note that

$$\begin{aligned}
\mathbb{E}[\mathbf{Re}(R)_{ij}^2] &= \underbrace{\sum_{l=1}^k p_l \lambda [\cos(\theta_{l,i} - \theta_{l,j}) - \sum_{l'=1}^k p_{l'} \lambda \cos(\theta_{l',i} - \theta_{l',j})]^2}_{P_1} \\
&\quad + \underbrace{(1 - \sum_{l=1}^k p_l) \lambda \mathbb{E}[\cos N_{ij} - \sum_{l'=1}^k p_{l'} \lambda \cos(\theta_{l',i} - \theta_{l',j})]^2}_{P_2} \\
&\quad + \underbrace{(1 - \lambda) [\sum_{l=1}^k p_l \lambda \cos(\theta_{l,i} - \theta_{l,j})]^2}_{P_3} \\
&= P_1 + P_2 + P_3. \tag{D.25}
\end{aligned}$$

One can easily verify the following inequalities.

$$\begin{aligned}
P_1 &\leq \sum_{l=1}^k 2p_l \lambda [(1 - p_l \lambda)^2 + (\sum_{l' \neq l} p_{l'} \lambda)^2], \\
P_2 &\leq (1 - \sum_{l=1}^k p_l) \lambda \left[ \frac{1}{2} + (\sum_{l'=1}^k p_{l'} \lambda)^2 \right] \text{ and } P_3 \leq (1 - \lambda) (\sum_{l'=1}^k p_{l'} \lambda)^2.
\end{aligned}$$

Plugging these in (D.25) readily leads to  $\tilde{\sigma} \leq \sqrt{C(\lambda, p_1, \dots, p_k)n}$ . Therefore, using Theorem 5 with  $t = 2\sqrt{2}\sqrt{C(\lambda, p_1, \dots, p_k)n}$ , we obtain for constants  $\varepsilon \geq 0, c_\varepsilon > 0$  that

$$\mathbb{P} \left( \left\{ \|\mathbf{Re}(R)\|_2 \geq (2 + \varepsilon) 2\sqrt{2}\sqrt{C(\lambda, p_1, \dots, p_k)n} \right\} \right) \leq n \exp \left( -\frac{8C(\lambda, p_1, \dots, p_k)n}{\bar{\sigma}(\lambda, p_1, \dots, p_k)^2 c_\varepsilon} \right). \tag{D.26}$$

2. (**Bounding**  $\|\mathbf{Im}(R)\|_2$ .) By proceeding analogously as in the proof of Lemma 3, we obtain for constants  $\varepsilon \geq 0, c_\varepsilon > 0$  that

$$\mathbb{P} \left( \left\{ \|\mathbf{Im}(R)\|_2 \geq (2 + \varepsilon) 4\sqrt{2}\sqrt{C(\lambda, p_1, \dots, p_k)n} \right\} \right) \leq 2n \exp \left( -\frac{16C(\lambda, p_1, \dots, p_k)n}{\bar{\sigma}(\lambda, p_1, \dots, p_k)^2 c_\varepsilon} \right). \tag{D.27}$$

Finally, the statement in the Lemma follows by applying the union bound to (D.26), (D.27).