Math. Stat. Learn. 1 (2018), 227–256
DOI 10.4171/MSL/6

Mathematical Statistics and Learning

Group synchronization on grids

Emmanuel Abbe, Laurent Massoulié, Andrea Montanari,
Allan Sly and Nikhil Srivastava

Abstract. Group synchronization requires to estimate unknown elements $(\boldsymbol{\theta}_v)_{v\in V}$ of a compact group $\mathfrak{G}$ associated to the vertices of a graph $G = (V, E)$, using noisy observations of the group differences associated to the edges. This model is relevant to a variety of applications ranging from structure from motion in computer vision to graph localization and positioning, to certain families of community detection problems.

We focus on the case in which the graph G is the d-dimensional grid. Since the unknowns $\boldsymbol{\theta}_v$ are only determined up to a global action of the group, we consider the following weak recovery question. Can we determine the group difference $\boldsymbol{\theta}_u^{-1}\boldsymbol{\theta}_v$ between far apart vertices u, v better than by random guessing? We prove that weak recovery is possible (provided the noise is small enough) for $d \geq 3$ and, for certain finite groups, for $d \geq 2$. Vice-versa, for some continuous groups, we prove that weak recovery is impossible for $d = 2$. Finally, for strong enough noise, weak recovery is always impossible.

Mathematics Subject Classification (2010). 62F10.

Keywords. Graphs, group synchronization, weak recovery, community detection.

1. Introduction

In the group synchronization problem, we are given a (countable) graph $G = (V, E)$, a group $\mathfrak{G}$ and, for each edge $(u, v) \in E$, a noisy observation $\boldsymbol{Y}_{u,v}$. The objective is to estimate group elements $(\boldsymbol{\theta}_v)_{v\in V}$ associated to the vertices $v \in V$, under the assumption that the $\boldsymbol{Y}_{u,v}$ are noisy observations of the group difference between the adjacent vertices. Roughly speaking (see below for a precise definition):

$$\boldsymbol{Y}_{uv} = \boldsymbol{\theta}_u^{-1}\boldsymbol{\theta}_v + \text{noise}\,. \tag{1.1}$$

In order for the above to be unambiguous, we will assume that an orientation (u, v) is fixed arbitrarily for each edge.

It is useful to introduce two concrete examples.

Example 1.1. The simplest example is $\mathfrak{G} = \mathbb{Z}_2 = \{(+1, -1), \cdot\}$, the group with elements $(+1, -1)$ and operation given by ordinary multiplication (equivalently, the group of integers modulo 2). For each edge $(u, v) \in E$ we are given $\boldsymbol{Y}_{uv}$, which is a noisy observation of $\boldsymbol{\theta}_u \boldsymbol{\theta}_v = \boldsymbol{\theta}_u^{-1} \boldsymbol{\theta}_v$. For instance we can assume that, for some $p \in [0, 1/2)$,

$$\boldsymbol{Y}_{uv} = \begin{cases} \boldsymbol{\theta}_u \boldsymbol{\theta}_v & \text{with probability } 1 - p, \\ -\boldsymbol{\theta}_u \boldsymbol{\theta}_v & \text{with probability } p. \end{cases} \tag{1.2}$$

with the $(\boldsymbol{Y}_{uv})_{(u,v)\in E}$ conditionally independent given $(\boldsymbol{\theta}_v)_{v\in V}$. In other words $\boldsymbol{Y}_{uv}$ is the output of a binary symmetric channel with flip probability p and input $\boldsymbol{\theta}_u \boldsymbol{\theta}_v$.

We will refer to this case as $\mathbb{Z}_2$ synchronization.

Example 1.2. Consider $\mathfrak{G} = \mathrm{O}(m)$: the group of $m \times m$ orthogonal matrices, with the following noise model. Let $(\boldsymbol{Z}_{uv})_{(u,v)\in E}$ be an i.i.d. collection of matrices with i.i.d. standard normal entries, and define

$$\boldsymbol{Y}_{uv} = \mathcal{P}_{\mathrm{O}(m)}(\boldsymbol{\theta}_u^{-1} \boldsymbol{\theta}_v + \sigma \boldsymbol{Z}_{uv}) \,. \tag{1.3}$$

Here $\mathcal{P}_{\mathrm{O}(m)}$ is the projector for the Frobenius norm $\|\cdot\|_F$ onto the orthogonal group, namely for a matrix $\boldsymbol{M}$ with singular value decomposition $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{T}}$, we set $\mathcal{P}_{\mathrm{O}(m)}(\boldsymbol{M}) = \boldsymbol{U}\boldsymbol{V}^{\mathsf{T}}$.

Group synchronization plays an important role in a variety of applications.

Structure from motion is a central problem in computer vision: given multiple images of an object taken from different points of view (and in presence of noise or occlusions) we want to reconstruct the 3-dimensional structure of the object [9, 22, 31, 40]. A possible intermediate step towards this goal consists in estimating the relative orientation of the object with respect to the camera in each image. This can be formulated as a group synchronization problem over $\mathfrak{G} = \mathrm{SO}(3)$, whereby $\boldsymbol{\theta}_u$ describes the orientation of image u, and pairwise image registration is used to construct the relative rotations $\boldsymbol{Y}_{uv}$.

Graph localization and positioning. Consider a set of nodes with positions

$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d.$$

We want to reconstruct the nodes positions from noisy measurements of the pairwise distances $\|\boldsymbol{x}_u - \boldsymbol{x}_v\|_2$. This question arises in sensor network positioning [14, 29], imaging [10, 37], manifold learning [38], to name only a few applications. It is often the case that measurements are only available for pairs $u, v \in [n]$ that are close enough, e.g. only if $\|\boldsymbol{x}_u - \boldsymbol{x}_v\|_2 \le \rho$ for ρ a certain communication range [18, 34].

Graph localization can be interpreted as a group synchronization problem in multiple ways. First, we can interpret the unknown position $\boldsymbol{x}_v$ as a translation and hence view it as a synchronization problem over the group of translations in d dimensions. Alternatively we can adopt a divide-and-conquer approach following [10]. First, we consider cliques in the graph and find their relative positions. Then we reconstruct the relative orientations of various cliques, which can be formulated as an $\mathrm{SO}(d)$ synchronization problem.

Community detection and the symmetric stochastic block model. The k-groups symmetric stochastic block model is a random graph over n vertices generated as follows [1, 25]. First, partition the vertex set into k subsets of size n/k, uniformly at random. Then connect vertices independently, conditional on the partition. Two vertices are connected with probability p if they belong to the same subset, and with a smaller probability $q < p$ otherwise. Given a realization of this graph, we would like to identify the partition. This problem is in fact closely related to synchronizations over $\mathbb{Z}_k$ (the group of integers modulo k). Extensions of the stochastic block model where edges are endowed with labels have also been considered [20]. In particular the so-called censored block model considered in [36] corresponds precisely to Example 1.1 on an Erdős–Rényi graph.

The literature on group synchronization is fairly recent and rapidly growing. The articles [35, 41] discuss it in a variety of applications and propose several synchronization algorithms, mostly based on spectral methods or semidefinite programming (SDP) relaxations. Theoretical analysis — mostly in the case of random (or complete) graphs G — is developed in [2, 7, 19, 33]. Most of these studies use perturbation theoretic arguments which crucially rely on the fact that the Laplacian (or connection Laplacian, [8]) of the underlying graph has a spectral gap. This paper shows that *nontrivial recovery is possible even in the absence of a spectral gap,* as in the case of grids with $d \geq 3$. A crucial role in our proofs is played by the fact that the pseudoinverse of the graph Laplacian has appropriately bounded trace. The trace of the Laplacian's pseudoinverse is also known as the Kirchoff index of the graph, and can be thought as a measure of how well connected is the graph: a small Kirchoff index corresponds to a well-connected graph. As shown in [5, 7], the Kirchoff index provides a lower bound on the minimum error achievable in a group synchronization problem (see also [3, 4] for related work).

In the present paper we study the case in which G is the d-dimensional grid, $d \geq 1$. Namely, $V = \mathbb{Z}^d$, and –to be definite– we orient edges in the positive direction:

$$E \equiv \big\{(x, y) : y - x \in \{e_1, \dots, e_d\}\big\}, \tag{1.4}$$

where $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ is the ith element of the canonical basis in $\mathbb{R}^d$.

By construction, we can hope to determine the unknowns $(\boldsymbol{\theta}_x)_{x\in\mathbb{Z}^d}$ only up to a global action by a group element. In other words, we cannot distinguish between $(\boldsymbol{\theta}_x)_{x\in\mathbb{Z}^d}$ and $(\boldsymbol{g}\boldsymbol{\theta}_x)_{x\in\mathbb{Z}^d}$ for some $\boldsymbol{g} \in \mathfrak{G}$. We thus ask the following *weak recovery question*:

Is it possible to estimate $\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$ better than random guessing, as $\|x-y\|_2 \to \infty$?

Note that, if we have an estimator $T_{xy}(\boldsymbol{Y})$ of the relative group element $\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$, we can produce an estimate $(\widehat{\boldsymbol{\theta}}_x(\boldsymbol{Y}))_{x\in V}$ of the overall configuration (up to a global action of $\mathfrak{G}$), by assuming $\boldsymbol{\theta}_{x_0} = \boldsymbol{I}_m$, and setting $\widehat{\boldsymbol{\theta}}_x(\boldsymbol{Y}) = T_{x_0,x}(\boldsymbol{Y})$. This point will be discussed further in the next section.

In absence of noise (i.e. if $\boldsymbol{Y}_{uv} = \boldsymbol{\theta}_u^{-1}\boldsymbol{\theta}_v$ exactly), the answer to the above question is always positive: we can multiply the observations $\boldsymbol{Y}_{uv}$'s along any path connecting x to y to reconstruct exactly $\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$. However for any arbitrarily small noise level, errors add up along the path and this simple procedure is equivalent to random guessing for $\|x - y\|_2 \to \infty$. The weak recovery question hence amounts to asking whether we can avoid error propagation.

Focusing on the case of compact matrix groups, we will present the following main results:

Low noise, $d \geq 3$. For sufficiently low noise, we prove that weak recovery is possible for $d \geq 3$ and any group. As mentioned above, this shows that group synchronization is possible even in graphs with vanishing spectral gap.

High noise. Vice-versa, weak recovery is impossible in any dimension at sufficiently high noise (or for $d = 1$ at any positive noise).

Discrete groups. For the special case of $\mathbb{Z}_2$-synchronization, we prove that weak recovery is possible (at low enough noise) for all $d \geq 2$.

Continuous groups, $d = 2$. Vice-versa, for the simplest example of continuous group, SO(2), we prove that weak recovery is impossible for $d = 2$.

The above pattern is completely analogous to the one of phase transitions in spin models within statistical physics [11]. We refer to Section 4 for a discussion of the connection with statistical physics.

The rest of the paper is organized as follows. Section 3 presents formal definitions and statements of our main results. In order to achieve optimal synchronization, it is natural to consider the Bayes posterior of the unknowns $(\boldsymbol{\theta}_v)_{v\in V}$, cf. Section 4. While this does not lead directly to efficient algorithms, it clarifies the connection with statistical physics. Some useful intuition can be developed by considering the case[1] in which $\theta_v \in \mathbb{R}$ and $Y_{uv} = \theta_v - \theta_u + Z_{uv}$ with $(Z_{uv})_{(u,v)\in E}$ i.i.d. noise. This can be treated by elementary methods, cf. Section 5. Finally, Section 6 and 7 prove our positive results (reconstruction is possible) with other proofs deferred to the appendices.

Notations. Throughout the paper we use boldface symbols (e.g. $\boldsymbol{\theta}_x, \boldsymbol{Y}_{xy}$) to denote elements of the group $\mathfrak{G}$, and, occasionally, for matrices with the same dimensions as elements of $\mathfrak{G}$ (i.e. $m \times m$ matrices). We normal symbols for other quantities (including vectors and matrices). Given two vectors $u, v \in \mathbb{R}^n$, we denote by $\langle u, v\rangle = \sum_{i=1}^n u_i v_i$ their scalar product. Analogously, for matrices $A, B \in \mathbb{R}^{m\times n}$, $\langle A, B\rangle = \mathrm{Tr}(AB^{\mathsf{T}})$ is their scalar product.

[1]Strictly speaking, this is not a special case of the problem studied in the rest of the paper, because $\mathfrak{G} = \mathbb{R}$ is not a compact group.

2. Definitions and problem formulation

As mentioned above, $G = (V, E)$ will be the infinite d-dimensional grid, and $\mathfrak{G}$ a compact matrix group. Without substantial loss of generality, we will assume $\mathfrak{G} \subseteq \mathrm{O}(m)$ (the group of $m \times m$ orthogonal matrices). We attach to each vertex $x \in V$ an element $\boldsymbol{\theta}_x \in \mathfrak{G}$ which may be deterministic or random, chosen independently from some distribution.

We are given observations $\boldsymbol{Y} = (\boldsymbol{Y}_{xy})_{(x,y)\in E}$, $\boldsymbol{Y}_{xy} \in \mathfrak{G}$, that are conditionally independent given $\boldsymbol{\theta}$. We assume that observations are unbiased in the following sense:

$$\mathbb{E}\{\boldsymbol{Y}_{xy}|\boldsymbol{\theta}\} = \lambda\, \boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y\,, \tag{2.1}$$

where the parameter λ is a natural measure of the signal-to-noise ratio. Note that, since $\boldsymbol{Y}_{xy} \in \mathrm{O}(m)$, we have, by Jensen's inequality,

$$m = \mathbb{E}\big\{\langle \boldsymbol{Y}_{xy}, \boldsymbol{Y}_{xy}\rangle \big|\boldsymbol{\theta}\big\} \geq \big\langle \mathbb{E}\{\boldsymbol{Y}_{xy}|\boldsymbol{\theta}\}, \mathbb{E}\{\boldsymbol{Y}_{xy}|\boldsymbol{\theta}\}\big\rangle = \lambda^2 m\,. \tag{2.2}$$

Hence we have $|\lambda| \leq 1$, and we can assume, without loss of generality, $\lambda \in [0, 1]$. In particular, if $\lambda = 1$, then Jensen's inequality is satisfied with equality, which corresponds to $\boldsymbol{Y}_{xy} = \boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$ almost surely (i.e. $\lambda = 1$ corresponds to noiseless observations).

The two examples given in the introduction fit this general definition:

- For $\mathbb{Z}_2$ synchronization (cf. Example 1.1) we have $\mathbb{E}\{\boldsymbol{Y}_{xy}|\boldsymbol{\theta}\} = (1-2p)\,\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$, and therefore $\lambda = (1 - 2p)$.
- For $\mathrm{O}(m)$ synchronization (cf. Example 1.2) we have $\mathbb{E}\{\boldsymbol{Y}_{xy}|\boldsymbol{\theta}\} = \lambda(\sigma^2)\,\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$ where $\sigma^2 \mapsto \lambda(\sigma^2)$ is a continuous function on $[0,\infty)$ with $\lambda(\sigma^2) \to 1$ as $\sigma^2 \to 0$ and $\lambda(\sigma^2) \to 0$ as $\sigma^2 \to \infty$ (see Appendix A).

A simple mechanism to produce the noisy observations $\boldsymbol{Y}_{xy}$ consists in introducing a probability kernel Q on $\mathfrak{G}$ and stipulate that, for each edge (x, y),

$$\mathbb{P}(\boldsymbol{Y}_{x,y} \in \cdot\,|\boldsymbol{\theta}) = \mathbb{P}(\boldsymbol{Y}_{x,y} \in \cdot\,|\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y) = \mathrm{Q}(\,\cdot\,|\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y)\,. \tag{2.3}$$

In other words, all observations are obtained by passing $\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$ through the same noisy channel. While our results do not necessarily assume this structure, both of the examples given above are of this type.

As mentioned in the introduction, in general we can only hope to estimate the vertex variables $(\boldsymbol{\theta}_x)_{x\in V}$ up to a global action of the group $\mathfrak{G}$. In particular, under the model (2.3), the law of $(\boldsymbol{Y}_{xy})_{(xy)\in E}$ is invariant under the transformation $\boldsymbol{\theta}_x \mapsto \boldsymbol{g}\boldsymbol{\theta}_x$ for a fixed $\boldsymbol{g} \in \mathfrak{G}$. One possible option to remove this unidentifiability issue is to fix the value of $\boldsymbol{\theta}_{x_0}$ at one specific vertex $x_0 \in V$, for instance by stipulating that $\boldsymbol{\theta}_{x_0} = \boldsymbol{I}_m$. An alternative choice is to define the reconstruction problem as the one of estimating the group differences $\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$ for every pair of vertices $x, y \in V$. We will mostly adopt the second formulation, but the discussion below clarifies that the points of views are in fact closely related (especially for transitive graphs).

Formally, an estimator is a collection of measurable functions $T_{uv}\colon \boldsymbol{Y} \mapsto T_{uv}(\boldsymbol{Y}) \in \mathfrak{G}$ indexed by all vertex pairs $u, v \in V$ (here $\boldsymbol{Y} = (\boldsymbol{Y}_{xy})_{(x,y)\in E}$ denotes the set of all observations). In this paper we ask whether there exists an estimator such that $T_{xy}(\boldsymbol{Y})$ is more correlated to $\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$ than random guessing, even when $\|x - y\| \to \infty$. In order to formalize this idea, it is useful to consider two extreme cases (here and below $\mathbb{P}_X$ denotes the law of the random variable X):

- $T_{xy}(\boldsymbol{Y})$ is "random guessing", i.e. independent of the vertex variables $(\boldsymbol{\theta}_x)_{x\in V}$. In this case
$$\big(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y, T_{xy}(\boldsymbol{Y})\big) \sim \mathbb{P}_{\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y} \times \mathbb{P}_{T_{xy}(\boldsymbol{Y})}$$
for every $x \neq y$.
- $T_{xy}(\boldsymbol{Y})$ is a perfect estimator, i.e. $T_{xy}(\boldsymbol{Y}) = \boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$. In this case
$$(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y)^{-1} T_{xy}(\boldsymbol{Y}) = \boldsymbol{I}_m$$
with probability one.

Motivated by these remarks, we introduce the following definition.

Definition 2.1. We say that the weak recovery problem is solvable for the probability distribution $\mathbb{P}$ over $(\boldsymbol{\theta}, \boldsymbol{Y})$ defined above if there exists an estimator T, and $\varepsilon > 0$, such that

$$\liminf_{\|x-y\|\to\infty} \big\|\mathbb{P}_{(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y, T_{xy}(\boldsymbol{Y}))} - \mathbb{P}_{\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y} \times \mathbb{P}_{T_{xy}(\boldsymbol{Y})}\big\|_{\mathrm{TV}} \geq \varepsilon > 0\,. \tag{2.4}$$

Slightly different versions of this definition are also potentially interesting. The next lemma clarifies the relation with mutual information. Given two random variables U, W, we denote by $\mathrm{I}(U; W)$ their mutual information, namely

$$\mathrm{I}(U; W) \equiv \mathbb{E}\Big\{\log \frac{\mathrm{d}\mathbb{P}_{U,W}}{\mathrm{d}(\mathbb{P}_U \times \mathbb{P}_W)}(U, W)\Big\}\,, \tag{2.5}$$

where $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$ denoted the Radon–Nikodym derivative of μ with respect ν.

The next lemma connects the definition of weak recovery given above with the asymptotic behavior of the mutual information $\mathrm{I}(T_{xy}(\boldsymbol{Y}); \boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y)$, and with another measure of dependency expressed in terms of total variation distance. Its proof is presented in Appendix B.

Lemma 2.2. *If the weak recovery problem is solvable, then there exists an estimator T, and $\varepsilon' > 0$, such that*

$$\liminf_{\|x-y\|\to\infty} \mathrm{I}\big(T_{xy}(\boldsymbol{Y}); \boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y\big) \geq \varepsilon' > 0\,. \tag{2.6}$$

Vice-versa, assume that $(\boldsymbol{\theta}_x)_{x\in V} \sim_{\rm iid} \mathbb{P}_{\rm Haar}$ *(with* $\mathbb{P}_{\rm Haar}$ *denoting the Haar measure on* $\mathfrak{G}$*). If*

$$\liminf_{\|x-y\|\to\infty} \left\|\mathbb{P}\big(\boldsymbol{\theta}_x T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1} \in \cdot\big) - \mathbb{P}_{\rm Haar}\big(\,\cdot\,\big)\right\|_{\rm TV} \ge \varepsilon'' > 0\,, \tag{2.7}$$

then the weak recovery problem is solvable.

As mentioned above, it is instructive to consider a slightly different formulation of the weak recovery question, whereby we try to estimate the hidden configuration $(\boldsymbol{\theta}_x)_{x\in V}$. Formally, for each $x \in V$, we let $\widehat{\boldsymbol{\theta}}_x : \boldsymbol{Y} \mapsto \widehat{\boldsymbol{\theta}}_x(\boldsymbol{Y})$ be an estimator of the corresponding group element $\boldsymbol{\theta}_x$. Consider a finite box in $\mathbb{Z}^d$, namely the subset of vertices $V_L = V \cap [-L, L]^d$, of cardinality $|V_L| = (2L+1)^d$. We then evaluate the performance of the estimator $\widehat{\boldsymbol{\theta}}$ by considering the overlap

$$R_L(\widehat{\boldsymbol{\theta}}) = \mathbb{E}\sup_{\boldsymbol{Q}\in\mathfrak{G}} \left\{\frac{1}{|V_L|}\sum_{x\in V_L} \langle \boldsymbol{\theta}_x, \boldsymbol{Q}\widehat{\boldsymbol{\theta}}_x(\boldsymbol{Y})\rangle\right\}. \tag{2.8}$$

Note that the supremum over $\boldsymbol{Q}$ is introduced in order to remove the global invariance. For instance, in the case of $\mathbb{Z}_2$-synchronization, $R_L(\widehat{\boldsymbol{\theta}})$ measures the expected fraction of agreements minus disagreements, up to a global flip.

Definition 2.3. Assume $(\boldsymbol{\theta}_x)_{x\in V}$ to be independent with $\mathbb{E}\{\boldsymbol{\theta}_x\} = \boldsymbol{0}$. We say that the estimator $\widehat{\boldsymbol{\theta}}$ *achieves positive overlap* for the synchronization problem if

$$\lim\inf_{L\to\infty} R_L(\widehat{\boldsymbol{\theta}}) \ge \varepsilon > 0\,. \tag{2.9}$$

While at a first look this formulation might appear very different from the one of Definition 2.1, the two points of view are closely related. Indeed, if weak recovery is possible — in the sense of Definition 2.1 — then the estimators $\big(T_{xy}(\boldsymbol{Y})\big)_{x,y\in V}$ provide an overall estimate of the configuration $(\boldsymbol{\theta}_x)_{x\in V}$, always up to a global group action. The estimate is simply given by

$$\widehat{\boldsymbol{\theta}}_x(\boldsymbol{Y}) = T_{x_0,x}(\boldsymbol{Y})\,, \tag{2.10}$$

where $x_0 \in V$ is an arbitrarily chosen vertex. This amounts to assuming that $\boldsymbol{\theta}_{x_0} = \boldsymbol{I}_m$. For instance, we can set $x_0 = 0$.

In all of our results, when we prove weak recovery, the same proof implies that the estimator $\widehat{\boldsymbol{\theta}}_x(\boldsymbol{Y}) = T_{0,x}(\boldsymbol{Y})$ achieves positive overlap in the sense given in the last definition.

3. Main results

Our first result establishes that the problem is solvable if noise is small enough in $d \ge 3$ dimensions.

Theorem 1. *If $d \geq 3$, then there exists $\lambda_{\text{UB}} \in (0,1)$ such that, if $\lambda > \lambda_{\text{UB}}$ then the weak recovery problem is solvable. Further, the corresponding estimator achieves positive overlap.*

If noise is strong enough, the problem becomes unsolvable.

Theorem 2. *Assume that:*

(1) $\mathbb{P}(\boldsymbol{Y}_{x,y} \in \cdot\,|\boldsymbol{\theta}) = \mathbb{P}(\boldsymbol{Y}_{x,y} \in \cdot\,|\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y)$.

(2) $\mathbb{P}(\boldsymbol{Y}_{x,y} \in \cdot\,|\boldsymbol{\theta}_0)$ *has density* $q(\,\cdot\,|\boldsymbol{\theta}_0)$, $\boldsymbol{\theta}_0 \in \mathfrak{G}$, *with respect to the Haar probability measure.*

Let $p_c(d) \in (0,1]$ the critical threshold for percolation on the d-dimensional grid. If

$$\inf_{\boldsymbol{y},\boldsymbol{\theta}_0} q(\boldsymbol{y}|\boldsymbol{\theta}_0) > 1 - p_c(d)\,, \tag{3.1}$$

then the weak recovery problem is not solvable.

In particular, for $d = 1$, the recovery is not solvable as soon as the noise is strictly positive $\inf_{\boldsymbol{y},\boldsymbol{\theta}} q(\boldsymbol{y}|\boldsymbol{\theta}) > 0$ *(since $p_c(d=1) = 1$).*

Note that for the case of a finite group $\mathfrak{G}$, denoting by $Q(\,\cdot\,|\boldsymbol{\theta}_0)$ the probability mass function on $\mathfrak{G}$ induced by $q(\,\cdot\,|\boldsymbol{\theta}_0)$, we have that condition (3.1) reads

$$\min_{\boldsymbol{y},\boldsymbol{\theta}_0} Q(\boldsymbol{y}|\boldsymbol{\theta}_0) > (1 - p_c(d))/|\mathfrak{G}|.$$

In particular, the percolation threshold in $d = 2$ dimensions is [13]

$$p_c(d = 2) = 1/2,$$

and therefore for $\mathbb{Z}_2$ synchronization with $d = 2$, the weak recovery problem is impossible if the flip probability is $p \in (1/4, 3/4)$.

The appearance of the percolation threshold $p_c(d)$ is quite interesting and can be explained intuitively. Let $p \equiv 1 - \inf_{\boldsymbol{y},\boldsymbol{\theta}_0} q(\boldsymbol{y}|\boldsymbol{\theta}_0)$. Under the assumptions of this theorem, the noisy observations $\boldsymbol{Y}_{x,y}$ can be characterized as follows. For each edge $(x, y) \in E$, draw an independent Bernoulli random variable $U_{xy} \in \{0,1\}$ with success probability p. If $U_{xy} = 0$, then let $\boldsymbol{Y}_{xy}$ be pure noise (i.e. Haar distributed). Otherwise, let $\boldsymbol{Y}_{xy} \sim q_*(\,\cdot\,|\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y)$ for a suitably constructed kernel q_* (in other words, this observation might contain some information about $\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$). This determines a percolation subgraph of the grid (namely the graph induced by edges with $U_{xy} = 1$), which can be referred to as the "information graph". It is intuitively clear that, for two vertices u, v, the relative group element $\boldsymbol{\theta}_u^{-1}\boldsymbol{\theta}_v$ can be estimated in a non-trivial way only if u and v are connected by a path in the information graph. This happens with probability bounded away from zero as $\|u - v\| \to \infty$, only if $p > p_c(d)$.

While Theorems 1 and 2 provide a complete qualitative picture for $d \geq 3$, for $d = 2$ the situation is more complicated. First of all, we show that for certain discrete groups the problem is solvable at low enough noise. We consider here the case $\mathfrak{G} = \mathbb{Z}_2$, but it would be interesting to generalize this result to other finite groups.

Theorem 3. *Consider $d = 2$, and $\mathfrak{G} = \mathbb{Z}_2$, with uniform flip probability p. Then there exists $p_* \in (0, 1)$ such that, if $p \le p_*$ then the weak recovery problem is solvable. Further, the corresponding estimator achieves positive overlap.*

On the contrary, for continuous groups, we expect weak recovery not to be possible in $d = 2$ dimensions, even for very weak noise. This is analogous to the celebrated Mermin–Wagner theorem in statistical mechanics [21, 26]. For the sake of simplicity, we focus on the case of $\mathfrak{G} = \mathrm{SO}(2)$ which is isomorphic to U(1), the group of complex variables of unit modulus, with ordinary multiplication. Let Z a U(1)-valued random variable with density g satisfying

$$g \in C^2, \qquad \inf_{s \in [0,2\pi]} g(e^{is}) > 0\,. \tag{3.2}$$

We consider observation on the edges corrupted by multiplicative noise

$$\boldsymbol{Y}_{xy} = \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y Z_{xy}\,, \tag{3.3}$$

where $(Z_{xy})_{(x,y) \in E} \sim_{\mathrm{iid}} g$.

Theorem 4. *If $d = 2$ and $\mathfrak{G} = \mathrm{SO}(2)$ with noise model satisfying (3.2) and (3.3), then the weak recovery problem is not solvable.*

Remark 3.1. A result related to Theorem 3 was established in [15] using a Peierls argument (Section 4 outlines the connection with the statistical physics formulation). We present here an independent proof that also provides an efficient recovery algorithm. Also notice that the result of [15] does not imply immediately weak recovery.

Remark 3.2. While we state and prove these results for regular grids, we expect similar results to hold for more general graphs with finite-dimensional structure. A natural class of graphs to investigate possible generalizations is given by random geometric graphs well above the connectivity threshold.

Some of the proof techniques used here should generalize to such graphs. For instance, Theorem 1 is proved via a second moment calculation that relies on the recurrence properties of suitable ensembles of random paths in the d-dimensional grid. These recurrence properties should remain unchanged for related ensembles of paths in random geometric graphs.

Similarly, the recursive argument to prove Theorem 3 uses the decomposition of a grid of linear size L into $(L/\ell)^2$ blocks of linear side ℓ. This is repeated recursively, down to some large size ℓ_0. Such a recursive construction can be repeated (within a suitable approximation error) for random geometric graphs as well.

We leave the investigation of these generalizations to future work.

Remark 3.3. Some of our results are limited to specific groups: Theorem 3 applies to $\mathbb{Z}_2$ and Theorem 4 to SO(2). It is natural to ask whether they generalize to arbitrary finite groups (the former) and to arbitrary compact Lie groups (the latter).

4. Bayesian posterior and connection to statistical physics

In this section it is convenient to assume a more general model in which the observations $\boldsymbol{Y}_{xy} \in \mathbb{R}^{m\times m}$ are not necessarily elements of the matrix group $\mathfrak{G}$. We assume that the conditional distribution of the observations $\boldsymbol{Y}_{xy}$ given the unknowns $\boldsymbol{\theta}_x$ is absolutely continuous with respect to a reference measure $\mathbb{P}_\#$ (independent of $\boldsymbol{\theta}$). In practice, we will take $\mathbb{P}_\#$ to be either the Haar measure on $\mathfrak{G}$, or the Lebesgue measure on $\mathbb{R}^{m\times m}$. We denote the corresponding density by

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}_\#}(\boldsymbol{Y}_{x,y}|\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}_0}\exp\big\{-u(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y;\boldsymbol{Y}_{xy})\big\}, \tag{4.1}$$

where $u\colon \mathfrak{G}\times\mathbb{R}^{m\times m}\to\mathbb{R}\cup\{+\infty\}$ is a measurable function bounded below. Assuming the prior distribution of $(\boldsymbol{\theta}_x)_{x\in V}$ to be i.i.d. Haar, we can use Bayes formula to write the posterior $\mu_{\boldsymbol{Y}}(B) = \mathbb{P}(\boldsymbol{\theta}\in B|\boldsymbol{Y})$ as

$$\mu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}(\boldsymbol{Y})}\exp\Big\{-\sum_{(x,y)\in E} u(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y;\boldsymbol{Y}_{xy})\Big\}\,\mu_0(\mathrm{d}\boldsymbol{\theta}), \tag{4.2}$$

where $\mu_0(\mathrm{d}\boldsymbol{\theta}) = \mu_0(\mathrm{d}\boldsymbol{\theta}_1)\cdots\mu_0(\mathrm{d}\boldsymbol{\theta}_n)$ is the product Haar measure over the unknowns and $\mathcal{Z}(\boldsymbol{Y})$ is a normalization constant. The joint distribution (4.2) takes the form of a Gibbs measure on the graph G.

Remark 4.1. For Eq. (4.2) to make sense, the graph G needs to be finite. However, the Bayesian interpretation implies immediately that quantities of interest have a well defined limit over increasing sequences of graphs. In particular, we can take G to be the finite grid with vertex set

$$V = \{-L,\dots,L\}^d,$$

and edges

$$E = \big\{(x,y)\in V\times V : y-x\in\{e_1,\dots,e_d\}\big\}.$$

Then the quantity

$$\sup_{T_{xy}(\cdot)}\big\|\mathbb{P}_{(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y,T_{xy}(\boldsymbol{Y}))} - \mathbb{P}_{\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y}\times\mathbb{P}_{T_{xy}(\boldsymbol{Y})}\big\|_{\mathrm{TV}} \tag{4.3}$$

is obviously non-decreasing in L (because larger L corresponds to a larger class of estimators) and hence admits a well defined limit. We will refer succinctly to this $L\to\infty$ limit as the model on "the d-dimensional grid".

In the rest of this section, it will be useful to distinguish between the arguments of the posterior density (that we will keep denoting by $(\boldsymbol{\theta}_x)_{x\in V}$), and the true unknowns that we will denote by $(\boldsymbol{\theta}_{0,x})_{x\in V}$. We further assume that the function u satisfies

$$u(\boldsymbol{\theta}\boldsymbol{\tau}; Y) = u(\boldsymbol{\tau}; \boldsymbol{\theta}^{-1}Y) = u(\boldsymbol{\theta}; Y\boldsymbol{\tau}^{-1}), \tag{4.4}$$

for any $\boldsymbol{\theta}$, $\boldsymbol{\tau} \in \mathfrak{G}$ and any $Y \in \mathbb{R}^{m\times m}$. This condition is verified by all of our examples. Thanks to this symmetry, for any $\{\boldsymbol{\tau}_x\}_{x\in V}$ and any $\boldsymbol{Y}$, the distribution $\mu_{\boldsymbol{Y}}(\,\cdot\,)$ of $\boldsymbol{\theta}$ in (4.2) coincides with that of $\{\boldsymbol{\theta}_x\boldsymbol{\tau}_x^{-1}\}_{x\in V}$ where $\boldsymbol{\theta}$ is distributed according to $\mu_{\widetilde{\boldsymbol{Y}}}(\cdot)$, and $\widetilde{\boldsymbol{Y}}_{xy} = \boldsymbol{\tau}_x\boldsymbol{Y}_{xy}\boldsymbol{\tau}_y^{-1}$. By taking $\boldsymbol{\tau}_x = \boldsymbol{\theta}_{0,x}$ for all x, we can assume that $\boldsymbol{\theta}_{0,x} = \boldsymbol{I}_m$ for all x, which then leads to the $(\boldsymbol{Y}_{xy})_{(xy)\in E}$ being i.i.d. with common distribution

$$\boldsymbol{Y}_{xy} \sim \frac{1}{\mathcal{Z}_1}\exp\big\{-u(\boldsymbol{I}_m; \boldsymbol{Y}_{xy})\big\}\mathbb{P}_{\#}(\mathrm{d}\boldsymbol{Y}_{xy})\,. \tag{4.5}$$

In the jargon of statistical physics, Gibbs measures of the form (4.2) with associated parameters distribution (4.5) are known as spin-glasses on the “Nishimori line.” These were first introduced for the case $\boldsymbol{\theta}_x \in \{+1, -1\}$ [27] and subsequently generalized to other groups in [12]. Several results about spin glasses on the Nishimori line were derived in [28, 30] and the connection with Bayesian statistics was emphasized in [17, 24]. The weak recovery phase transition corresponds to a paramagnetic-ferromagnetic phase transition in physics language.

Example 4.2. The simplest example is the so-called *random bond Ising model* which is obtained by taking $\boldsymbol{\theta}_x \in \{+1, -1\}$ and

$$\mu_{\boldsymbol{Y}}(\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}(\boldsymbol{Y})}\exp\Bigg\{\beta\sum_{(x,y)\in E}\boldsymbol{Y}_{xy}\boldsymbol{\theta}_x\boldsymbol{\theta}_y\Bigg\}\,, \tag{4.6}$$

where $\boldsymbol{Y}_{xy} = +1$ with probability $1-p$ and $\boldsymbol{Y}_{xy} = -1$ with probability p. The Nishimori line is given by the condition $\beta = (1/2)\log((1-p)/p)$. It is easy to see that this is equivalent to the Bayes posterior for the $\mathbb{Z}_2$ synchronization model of Example 1.1, if we take $\boldsymbol{\theta}_{0,x} = +1$.

This model has attracted considerable interest within statistical physics. In particular, high-precision numerical estimates of the phase transition location yield $p_c \approx 0.1092$ (in $d = 2$) and $p_c \approx 0.233$ (in $d = 3$) [16, 32].

Example 4.3. Take $\mathfrak{G} = \mathrm{O}(m)$ (the group of orthogonal matrices), and assume

$$\boldsymbol{Y}_{xy} = \boldsymbol{\theta}_{0,x}^{-1}\boldsymbol{\theta}_{0,y} + \sigma\,\boldsymbol{Z}_{xy} \tag{4.7}$$

where $\boldsymbol{Z}_{xy}$ is a noise matrix with i.i.d. entries $(\boldsymbol{Z}_{xy})_{ij} \sim \mathsf{N}(0,1)$. This model is analogous to the one of Example 1.2, although we do not project observations onto the orthogonal group.

After a simple calculation, the Gibbs measure (4.2) takes the form

$$\mu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}(\boldsymbol{Y})} \exp\Big\{\beta \sum_{(x,y)\in E} \mathrm{Tr}\big(\boldsymbol{\theta}_x \boldsymbol{Y}_{xy} \boldsymbol{\theta}_y^{\mathsf{T}}\big)\Big\}\, \mu_0(\mathrm{d}\boldsymbol{\theta})\,, \tag{4.8}$$

where $\beta = 1/\sigma^2$. By the symmetry under $\mathrm{O}(m)$ rotations, for the purpose of analysis we can assume $\boldsymbol{Y}_{xy} = \boldsymbol{I}_m + \sigma\, \boldsymbol{Z}_{xy}$ which is the usual setting in physics.

Example 4.4. In the case $\mathfrak{G} = \mathrm{SO}(2)$ we can identify $\boldsymbol{\theta}_x$ with an angle in $[0, 2\pi)$, and let

$$\boldsymbol{Y}_{xy} = \boldsymbol{\theta}_{0,y} - \boldsymbol{\theta}_{0,x} + \boldsymbol{Z}_{xy}, \qquad \mathrm{mod}\ 2\pi\,, \tag{4.9}$$

where $\boldsymbol{Z}_{xy}$ is noise with density proportional to $\exp(-u(z))$ for $u(z)$ a periodic function bounded below.

The Gibbs measure (4.2) takes the form

$$\mu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}(\boldsymbol{Y})} \exp\Big\{-\sum_{(x,y)\in E} u\big(\boldsymbol{Y}_{xy} - \boldsymbol{\theta}_y + \boldsymbol{\theta}_x\big)\Big\}\, \mu_0(\mathrm{d}\boldsymbol{\theta})\,. \tag{4.10}$$

For the purpose of analysis we can assume $\boldsymbol{Y}_{xy} = \boldsymbol{Z}_{xy}$. This is known as the "XY model" in physics.

Our results have direct implications on these models that we summarize in the following statement.

Corollary 4.1. *Consider the Gibbs measure (4.2) on the d-dimensional grid, with parameters $\boldsymbol{Y}_{xy} \in \mathfrak{G}$ distributed according to Eq. (4.5) and satisfying Eq. (2.1). Then, the following hold:*

(1) *For $d \geq 3$, and $\mathfrak{G} \subseteq \mathrm{O}(m)$ is any compact matrix group, then there exists $\lambda_{\mathrm{UB}} < 1$ such that the model is in a ferromagnetic phase for any $\lambda > \lambda_{\mathrm{UB}}$.*

(2) *For the case of Example 4.2 (i.e. $\mathfrak{G} = \mathbb{Z}_2$) and $d \geq 2$, there exists $p_* \in (0,1)$ such that the model is in a ferromagnetic phase for any $p \leq p_*$.*

(3) *For the case of Example 4.4 (i.e. $\mathfrak{G} = \mathrm{SO}(2)$) and $d = 2$ the model is not in a ferromagnetic phase provided $z \mapsto u(z)$ is bounded.*

(4) *For any group $\mathfrak{G}$, $d \geq 2$, there exists a constant $c(d)$ such that, if $\|u\|_\infty \leq c(d)$ then the model is not in a ferromagnetic phase.*

Furthermore point 1 applies to Example 4.3 as well.

Proof. These statements are merely a translation of Theorems 1, 2, 3, 4 for the case in which channel observations take values in $\mathfrak{G}$. For the case in Example 4.3, note that we can always project $\boldsymbol{Y}_{xy}$ onto the group $\mathrm{O}(m)$, hence recovering the setting of Example 1.2. Since weak recovery is possible in the latter, it is also possible in the former. □

As already pointed out in Section 3, the existence of a ferromagnetic phase for Example 4.2 (i.e. $\mathfrak{G} = \mathbb{Z}_2$) was already obtained in [15].

5. A toy example

It is instructive to consider a simple example in which $\mathfrak{G} = \mathbb{R}$ is the group of translations on the real line. This case does not fit the framework of the rest of this paper, but presents the same dichotomy between $d = 2$ and $d \geq 3$ and can be solved by elementary methods.

Throughout this section, we adopt additive notation, and hence the observation on edge (x, y) takes the form

$$Y_{x,y} = \theta_y - \theta_x + Z_{x,y} \,, \tag{5.1}$$

where $\{Z_{xy}\}_{(x,y)\in E}$ are i.i.d. random variables with mean 0 and variance σ^2.

To simplify our treatment, we assume the graph to be the discrete torus, with vertex set $V = \{1, 2, \dots, L\}^d$ and edges $E = \{(x, x + e_j) : \quad x \in V,\ j \in \{1, \dots, d\}\}$ (where we identify $L + 1$ with 1). Denoting by D the difference operator on G, the observation can be written as

$$Y = D\theta + Z \,. \tag{5.2}$$

As usual, θ can be determined only up to a global shift. To resolve this ambiguity, it is convenient to assume that θ is centered: $\langle\theta, 1\rangle = 0$. Consider the least square estimator $\widehat{\theta}(Y) = D^{\dagger}Y = L^{\dagger}D^{\mathsf{T}}Y$ where $\dagger$ denotes the pseudoinverse. A standard elementary calculation [39, Theorem 14.13] yields the following formula for the mean square error

$$\mathrm{MSE}(L, \sigma^2) \equiv \frac{1}{L^d}\mathbb{E}\{\|\boldsymbol{\theta}(Y) - \theta\|_2^2\} \tag{5.3}$$

$$= \frac{\sigma^2}{L^d}\,\mathrm{Tr}_0\left((D^{\mathsf{T}}D)^{\dagger}\right) = \frac{\sigma^2}{L^d}\,\mathrm{Tr}_0\left(\mathcal{L}^{\dagger}\right). \tag{5.4}$$

Here we denoted by $\mathcal{L} = D^{\mathsf{T}}D$ the Laplacian of G and by Tr_0 the trace on the subspace orthogonal to the all-ones vector. The eigenpairs of the Laplacian are [23]:

$$v(p)_x = \frac{1}{L^{d/2}}\,e^{i\langle p,x\rangle}\,, \quad \lambda(p) = \sum_{i=1}^{d}[2 - 2\cos(p)]\,, \tag{5.5}$$

$$p \in B_L \equiv \left\{\frac{2\pi}{L}(n_1, \dots, n_d) : n_i \in \{0, \dots, L-1\}\right\}. \tag{5.6}$$

Hence

$$\mathrm{MSE}(L, \sigma^2) = \frac{\sigma^2}{L^d}\sum_{p\in B_L\setminus\{0\}}\frac{1}{\lambda(p)}\,. \tag{5.7}$$

For large L, the sum can be estimated by approximating it via Riemann integrals to yield the following fact.

Fact 5.1. *The mean square error of least-square estimation within the translation synchronization model of Eq.* (5.1) *is*

$$\frac{1}{\sigma^2}\,\mathrm{MSE}(L,\sigma^2) = \begin{cases} \frac{L}{12} + O_L(1), & \textit{for } d = 1, \\ \frac{1}{2\pi}\log L + O_L(1), & \textit{for } d = 2, \\ C(d) + o_L(1), & \textit{for } d \geq 3, \end{cases} \tag{5.8}$$

where $C(d)$ is a dimension dependent constant.

We observe that this qualitative behavior is the same that we obtain for continuous compact groups, cf. Theorem 1 and Theorem 4: the weak recovery problem is solvable only for $d \geq 3$.

6. Proof of Theorem 1

Throughout this section we assume a probability distribution $\mathbb{P}$ over $\boldsymbol{\theta}, \boldsymbol{Y}$ satisfying the unbiasedness condition Eq. (2.1). For most of our analysis, we consider general estimators $T_{uv}\colon \boldsymbol{Y} \mapsto T_{uv}(\boldsymbol{Y}) \in \mathbb{R}^{m\times m}$ whose output is not necessarily in $\mathfrak{G}$, and let $\boldsymbol{T}_{uv} = T_{uv}(\boldsymbol{Y})$ (as projecting them into $\mathfrak{G}$ at the end can only increase their accuracy).

Before presenting the actual proof, it is useful to discuss the basic intuition. Consider first the case in which we are given noiseless observations. Namely, for each edge in the grid $(u,v) \in E$, we observe $\boldsymbol{Y}_{uv} = \boldsymbol{\theta}_u^{-1}\boldsymbol{\theta}_v$ (note that in this case $\lambda = 1$). We can then construct an straightforward estimate $T_{xy}(\boldsymbol{Y})$: choose an arbitrary directed path γ from x to y, and take the product of the $\boldsymbol{Y}_{uv}$'s along the path (ordered as the edges of the path itself). It is immediate to see that in this case $T_{xy}(\boldsymbol{Y}) = \boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$. However, if observations are noisy, the error of this estimator grows exponentially with the path length. For instance, in the case of $\mathbb{Z}_2$ synchronization, cf. Example 1.1, we would get $T_{xy}(\boldsymbol{Y}) \in \{+1,-1\}$ with $\mathbb{E}\{T_{xy}(\boldsymbol{Y})\} = (1-2p)^{|\gamma|}\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$.

In order to overcome this difficulty, we average over the paths between x and y to reduce the variance of the estimator. It turns out however that averaging uniformly over paths is not sufficient to obtain a tight result and only allows to establish weak recovery for $d \geq 4$. The key feature of the underlying measure over paths is the exponential intersection property. A construction by Benjamini, Pemantle, and Peres [6] provides such a measure for $d = 3$.

In order to apply our construction to our problem, we proceed in two steps. First we consider pair of vertices that differ by a vector on the diagonal of the three-dimensional subgrid: $y - x = (n, n, n, 0, \dots, 0)$. In a second step, we piece together such estimators, to cover all vertex pairs.

Turning to the actual proof, we set $u(n) = (n, \dots, n) \in \mathbb{Z}^d$. We will use repeatedly two elementary facts in linear algebra. First, for any two matrices $\boldsymbol{A}$, $\boldsymbol{B}$,

$$\|\boldsymbol{A}\boldsymbol{B}\|_F \le \|\boldsymbol{A}\|_F \|\boldsymbol{B}\|_F .$$

Second, if $\boldsymbol{B}$ is an orthogonal matrix, then

$$\|\boldsymbol{A}\boldsymbol{B}\|_F = \|\boldsymbol{A}\|_F .$$

As mentioned above, we start by defining the estimator $T_{x,y}(\boldsymbol{Y})$ for $x = 0$, $y = u(n)$, and will then generalize it to other pairs x, y.

Lemma 6.1. *Consider $d \ge 3$. Then there exists an estimator $T = (T_{u,v})$ such that for any $\varepsilon > 0$ there exists $\lambda_0(\varepsilon_0)$, such that, for all $\lambda > \lambda_0$ and all even n,*

$$\mathbb{P}\Big\{\Big\|\boldsymbol{\theta}_0 \boldsymbol{T}_{0,u(n)} \boldsymbol{\theta}_{v(j,n)}^{-1} - \boldsymbol{I}_m\Big\|_F \ge \varepsilon\Big\} \le \varepsilon_0 . \tag{6.1}$$

In fact, it is sufficient to take $\lambda(\varepsilon) = \sqrt{1 - \varepsilon_0^3/(C_1 m)}$, for C_1 an absolute constant.

Proof. We consider the case of $d = 3$, as larger dimensions follow from this case. Denote by $\mathcal{P}_+$ the set of infinite increasing paths in the grid that start at 0. Benjamini, Pemantle and Peres [6] construct a probability measure μ over paths in $\mathcal{P}_+$ satisfying the so-called exponential intersection property (EIT). Namely, there exist absolute constants $\beta_* < 1$, C_* such that

$$(\mu \times \mu)\big\{(\gamma_1, \gamma_2) \in \mathcal{P}_+ \times \mathcal{P}_+ : |\gamma_1 \cap \gamma_2| \ge k\big\} \le C_* \beta_*^k. \tag{6.2}$$

We first need to extend this property from infinite increasing paths starting at 0, to paths starting at 0 and ending at (n, n, n). We use a symmetrization argument. Assume for simplicity that $n/3 \in \mathbb{Z}$. The idea is to take a random walk that follows the distribution of the random walk from [6], stop the walk once it hits the hyperplane

$$H(n) = \big\{x \in \mathbb{R}^3 : \langle x - (n/2, n/2, n/2), (1,1,1)\rangle = 0\big\}$$

(note that this is reached in $3n/2$ steps), and reflect the walk from the hyperplane for the next $3n/2$ steps to end in (n, n, n). The second segment of the walk is thus a deterministic reflection of the first segment. Note however that the reflected segment may no longer stay in the grid $\mathbb{Z}^3$; in fact, the reflection is given by

$$(x, y, z) \mapsto \big(x+n-2/3(x+y+z), y+n-2/3(x+y+z), z+n-2/3(x+y+z)\big),$$

and, for instance, $(1, 0, 0)$ is reflected to $(n + 1/3, n - 2/3, n - 2/3)$. We therefore restrict ourselves next to walk that stay in the grid under reflection.

For this purpose, consider the supper-grid $3\mathbb{Z}^3 = \{(3x, 3y, 3z) : x, y, z \in \mathbb{Z}\}$. For a path γ in $3\mathbb{Z}^3$, we denote by $\mathrm{ext}(\gamma)$ the extension of γ to a path in $\mathbb{Z}^3$ obtained by connecting vertices of the supper-grid with the shortest paths (of length 3) in $\mathbb{Z}^3$.

Let $\mathcal{P}_+^{(3)}$ be the set of infinite increasing paths in $3\mathbb{Z}^3$ starting at 0. Let $\mu^{(3)}$ be the distribution from [6] that has EIT on $\mathcal{P}_+^{(3)}$, with the same parameters as in (6.2). For $\Gamma \sim \mu^{(3)}$, let $\Gamma^{(-)}$ be the restriction of $\text{ext}(\Gamma)$ on the first $3n/2$ steps, $\Gamma^{(+)}$ the reflection of $\Gamma^{(-)}$ from the hyperplane $H(n)$ and let $\Gamma_n = (\Gamma^{(-)}, \Gamma^{(+)})$. Denote by $\mathcal{P}_n$ the set of paths in $\mathbb{Z}^3$ that have the form of Γ_n, i.e. starting at 0, length $3n$, increasing in the first $3n/2$ steps, symmetric with respect to the hyperplane $H(n)$, and thus ending in (n, n, n).

We will work with the probability measure μ_n of Γ_n. Since the reflection part can only double the overlap of two paths, and since the extension from the supper-grid to the grid can only triple the overlap of two paths, it follows from (6.2) that there exists $C, \beta < 1$, both independent of n, such that

$$\mu_n \times \mu_n\big\{(\gamma_1, \gamma_2) \in \mathcal{P}_n \times \mathcal{P}_n : |\gamma_1 \cap \gamma_2| \geq k\big\} \leq C\beta^k. \tag{6.3}$$

For a path $\gamma \in \mathcal{P}_n$, denote the ordered sequence of directed edges in γ by $I_1(\gamma), \dots, I_{3n}(\gamma)$, where $I_j(\gamma) \in E$, $j \in [3n]$, and define

$$\boldsymbol{Y}_\gamma := \boldsymbol{Y}_{I_1(\gamma)} \boldsymbol{Y}_{I_2(\gamma)} \cdots \boldsymbol{Y}_{I_{3n}(\gamma)}\,, \tag{6.4}$$

$$\boldsymbol{T}_{0,u} := \frac{1}{\lambda^{3n}} \mathrm{E}_\gamma(\boldsymbol{Y}_\gamma)\,, \tag{6.5}$$

where E_γ denotes expectation with respect to μ_n. Note that by the assumption (2.1) we have $\mathbb{E}\boldsymbol{Y}_\gamma = \lambda^{3n}\boldsymbol{\theta}_0^{-1}\boldsymbol{\theta}_{u(n)}$ for any $\gamma \in \mathcal{P}_n$, and therefore

$$\mathbb{E}\boldsymbol{T}_{0,u(n)} = \boldsymbol{\theta}_0^{-1}\boldsymbol{\theta}_{u(n)}\,. \tag{6.6}$$

Observe that if two paths γ_1, γ_2 in $\mathcal{P}_n$ intersect in an edge e then they must intersect in the same position since the paths are increasing, i.e. we must have $e = I_k(\gamma_1) = I_k(\gamma_2)$ for some k. Writing for simplicity $u = u(n)$, and denoting by $\mathrm{E}_{\gamma_1,\gamma_2}$ expectation with respect to $\gamma_1, \gamma_2 \sim_{\text{iid}} \mu_n$

$$\begin{aligned}
\mathbb{E}\left\{\boldsymbol{T}_{0,u}\boldsymbol{T}_{0,u}^{\mathsf{T}}\right\} &= \frac{1}{\lambda^{6n}}\, \mathrm{E}_{\gamma_1,\gamma_2}\mathbb{E}\boldsymbol{Y}_{\gamma_1}(\boldsymbol{Y}_{\gamma_2})^{\mathsf{T}} \\
&= \frac{1}{\lambda^{6n}}\, \mathrm{E}_{\gamma_1,\gamma_2}\mathbb{E}\boldsymbol{Y}_{I_1(\gamma_1)} \dots \big(\mathbb{E}\boldsymbol{Y}_{I_{3n}(\gamma_1)}\boldsymbol{Y}_{I_{3n}(\gamma_2)}^{\mathsf{T}}\big)\boldsymbol{Y}_{I_{3n-1}(\gamma_2)}^{\mathsf{T}} \dots \boldsymbol{Y}_{I_1(\gamma_2)}^{\mathsf{T}} \\
&\overset{(a)}{=} \frac{1}{\lambda^{6n}} \mathrm{E}_{\gamma_1,\gamma_2}\lambda^{|\gamma_1|+|\gamma_2|-2|\gamma_1\cap\gamma_2|}\boldsymbol{I}_m \\
&= \mathrm{E}_{\gamma_1,\gamma_2}\lambda^{-2|\gamma_1\cap\gamma_2|}\,\boldsymbol{I}_m\,,
\end{aligned}$$

where (a) follows by repeatedly applying the identity $\boldsymbol{Y}_e\boldsymbol{Y}_e^{\mathsf{T}} = I_m$ for any edge e, each time an intersection appears, and taking expectation with respect to $\boldsymbol{Y}_{e_1}, \boldsymbol{Y}_{e_2}$ for not repeated edges. By this last expression, the trace τ of $\mathbb{E}\left\{\boldsymbol{T}_{0,u}\boldsymbol{T}_{0,u}^{\mathsf{T}}\right\}$ reads $\tau = m\,\mathbb{E}\left(\lambda^{-2X}\right)$ where X is a random variable counting the number of intersections in two paths γ_1, γ_2 independently drawn from μ_n. Thus for $\lambda^2 > \beta$,

$$m^{-1}\tau = \sum_{x\geq 0} \lambda^{-2x}\big[\mathbb{P}(X \geq x) - \mathbb{P}(X \geq x+1)\big] \tag{6.7}$$

$$= 1 + \sum_{x>0} \mathbb{P}(X \geq x)\big[\lambda^{-2x} - \lambda^{-2x+2}\big] \tag{6.8}$$

$$\leq 1 + (1-\lambda^2)\sum_{x>0} C\,(\beta/\lambda^2)^x \tag{6.9}$$

$$= 1 + (1-\lambda^2)\frac{C\beta}{\lambda^2-\beta}, \tag{6.10}$$

where the inequality follows from Eq. (6.3). Thus

$$\mathbb{E}\big\{\big\|\boldsymbol{\theta}_0 \boldsymbol{T}_{0,u}\boldsymbol{\theta}_u^{-1} - \boldsymbol{I}_m\big\|_F^2\big\}$$

$$= \mathrm{Tr}\,\mathbb{E}\big\{\boldsymbol{\theta}_0 \boldsymbol{T}_{0,u}\boldsymbol{T}_{0,u}^{\mathsf{T}}\boldsymbol{\theta}_0^{\mathsf{T}}\big\} - 2\,\mathrm{Tr}\,\mathbb{E}\big\{\boldsymbol{\theta}_0 \boldsymbol{T}_{0,u}\boldsymbol{\theta}_u^{-1}\big\} + m \tag{6.11}$$

$$= \tau - m \tag{6.12}$$

$$\leq (1-\lambda^2)m\frac{C\beta}{\lambda^2-\beta}, \tag{6.13}$$

where we used Eq. (6.6) together with our previous bound on τ.

$$\mathbb{E}\big\{\boldsymbol{\theta}_0 \boldsymbol{T}_{0,u(n)}\boldsymbol{\theta}_{u(n)}^{-1}\big\} = \boldsymbol{I}_m\,, \tag{6.14}$$

$$\mathbb{E}\big\{\big\|\boldsymbol{\theta}_0 \boldsymbol{T}_{0,u(n)}\boldsymbol{\theta}_{u(n)}^{-1} - \boldsymbol{I}_m\big\|_F^2\big\} \leq C_0\, m(1-\lambda^2)\,. \tag{6.15}$$

We therefore conclude that there exist absolute constants $\lambda_0 < 1$, C_0 such that, for all $\lambda > \lambda_0$

$$\mathbb{E}\big\{\big\|\boldsymbol{\theta}_0 \boldsymbol{T}_{0,u(n)}\boldsymbol{\theta}_{u(n)}^{-1} - \boldsymbol{I}_m\big\|_F^2\big\} \leq C_0\, m(1-\lambda^2)\,. \tag{6.16}$$

The claim follows from this bound and Eq. (6.6), using Markov inequality. □

We next treat arbitrary pairs of vertices x, y by "stitching together" estimators constructed in the last lemma. For this task, we use the following linear algebra remark.

Lemma 6.2. *For any two matrices $\boldsymbol{X}_1$, $\boldsymbol{X}_2 \in \mathbb{R}^{m\times m}$, we have*

$$1 + \|\boldsymbol{X}_1\boldsymbol{X}_2 - \boldsymbol{I}\|_F \leq \big(1 + \|\boldsymbol{X}_1 - \boldsymbol{I}\|_F\big)\big(1 + \|\boldsymbol{X}_2 - \boldsymbol{I}\|_F\big), \tag{6.17}$$

Proof. Letting $\boldsymbol{X}_i = \boldsymbol{I} + \boldsymbol{A}_i$, this follows form

$$\big\|\boldsymbol{A}_1 + \boldsymbol{A}_2 + \boldsymbol{A}_1\boldsymbol{A}_2\big\|_F \leq \big\|\boldsymbol{A}_1\big\|_F + \big\|\boldsymbol{A}_2\big\|_F + \big\|\boldsymbol{A}_1\boldsymbol{A}_2\big\|_F \tag{6.18}$$

$$\leq \big\|\boldsymbol{A}_1\big\|_F + \big\|\boldsymbol{A}_2\big\|_F + \big\|\boldsymbol{A}_1\big\|_F\big\|\boldsymbol{A}_2\big\|_F\,. \tag{6.19}$$

□

We can now prove our main result, that is a strengthening of Theorem 1.

Theorem 5. *Consider any $d \geq 3$ and fix $\varepsilon > 0$. Then there exists an estimator $T = (T_{u,v})_{u,v\in V}$, and a constant $\lambda_d(\varepsilon) < 1$, such that, for all $\lambda > \lambda_d$ and all n,*

$$\mathbb{P}\big\{\big\|\boldsymbol{\theta}_x\boldsymbol{T}_{x,y}\boldsymbol{\theta}_y^{-1} - \boldsymbol{I}_m\big\|_F \geq \varepsilon\big\} \leq \varepsilon\,. \tag{6.20}$$

Proof. Without loss of generality, assume $x = 0$.

First consider the case $y = v(n) = (n, 0, 0, \dots, 0)$. For n even and let $w(n) \equiv (n/2, n/2, n/2, 0, \dots, 0)$. Let $(\boldsymbol{T}^{(*)}_{x,y})_{x,y\in V}$ be the estimator of Lemma 6.1 (where we use only the observations on the subgraph induced by the hyperplane $\{x \in \mathbb{Z}^d : x_4 = \cdots = x_d = 0\}$). Define

$$\boldsymbol{T}^{(\#)}_{0,v(n)} = \boldsymbol{T}^{(*)}_{0,w(n)}\boldsymbol{T}^{(*)}_{w(n),v(n)}\,. \tag{6.21}$$

From the inequality (6.17), we get

$$\begin{aligned}&\|\boldsymbol{\theta}_0\boldsymbol{T}^{(\#)}_{0,v(n)}\boldsymbol{\theta}^{-1}_{v(n)} - \boldsymbol{I}\|_F\\ &\quad\leq \big(1 + \|\boldsymbol{\theta}_0\boldsymbol{T}^{(*)}_{0,w(n)}\boldsymbol{\theta}^{-1}_{w(n)} - \boldsymbol{I}\|_F\big)\big(1 + \|\boldsymbol{\theta}_{w(n)}\boldsymbol{T}^{(*)}_{w(n),v(n)}\boldsymbol{\theta}^{-1}_{v(n)} - \boldsymbol{I}\|_F\big) - 1\,.\end{aligned}$$

and hence the claim follows from Lemma 6.1 using union bound (using $\varepsilon_0 = \varepsilon/3$). For n odd, the same claim follows by setting $\boldsymbol{T}^{(\#)}_{0,v(n)} = \boldsymbol{T}^{(\#)}_{0,v(n-1)}\boldsymbol{Y}_{v(n-1),v(n)}$. Finally, by symmetry, the same result holds for $y = v(j,n) = n\,e_j$ along any of the coordinate axes.

Next consider the case of general y. For $j \in \{0, \dots, d\}$ we define

$$w(j) \equiv (y_1, \dots, y_j, 0, \dots, 0)$$

(in particular, $w(0) = 0$ and $w(d) = y$) and let

$$\boldsymbol{T}_{0,y} = \boldsymbol{T}^{(\#)}_{w(0),w(1)}\boldsymbol{T}^{(\#)}_{w(1),w(2)}\cdots\boldsymbol{T}^{(\#)}_{w(d-1),w(d)}\,. \tag{6.22}$$

By the argument above, for all $\lambda > \lambda_*(\varepsilon_0)$ we have

$$\mathbb{P}\big(\max_{1\leq j\leq d}\big\|\boldsymbol{\theta}_{w(j-1)}\boldsymbol{T}^{(\#)}_{w(j-1),w(j)}\boldsymbol{\theta}^{-1}_{w(j)} - \boldsymbol{I}_m\big\|_F \geq \varepsilon_0\big) \leq d\,\varepsilon_0\,. \tag{6.23}$$

By repeated application of Inequality (6.17), on the complement of the event in the right-hand side, one has

$$\|\boldsymbol{\theta}_0\boldsymbol{T}_{0,y}\boldsymbol{\theta}_y^{-1} - \boldsymbol{I}_m\| \leq \prod_{j=1}^{d}\big(\|\boldsymbol{\theta}_{w(j-1)}\boldsymbol{T}^{(\#)}_{w(j-1),w(j)}\boldsymbol{\theta}^{-1}_{w(j)} - \boldsymbol{I}_m\| + 1\big) - 1 \leq \varepsilon_0 2^d\,.$$

The claim follows by taking $\varepsilon_0 = \varepsilon/(2^d)$ and by setting $\lambda_d(\varepsilon) = \lambda_*(\varepsilon/2^d)$. □

7. Proof of Theorem 3

We give a multi-scale scheme to reconstruct the unknowns $\boldsymbol{\theta} = (\boldsymbol{\theta}_x)_{x\in\mathbb{Z}_2}$ when $d = 2$ although our approach could easily be generalized to any $d \geq 3$. Without loss of generality we will consider pairs of vertices u, v in the positive quadrant. For $k \geq 0$ let $\ell_k = 2^{10k(k+1)}$. We partition the lattice $\mathbb{Z}^2$ into blocks of side-length ℓ_k as follows,

$$B_u^{(k)} = \left\{(x_1, x_2) \in \mathbb{Z}^2 : u_i = \lceil x_i/\ell_k \rceil\right\} \tag{7.1}$$

Let $\mathcal{B}^{(k)}$ be the set of blocks at level k and let $D_{u,k}$ denote the unique block in $\mathcal{B}^{(k)}$ containing u. For each block $B \in \mathcal{B}^{(k)}$ we will define synchronization random variables $W_B^{(k)} \in \{-1, 1\}$ that are measurable with respect to $\{\boldsymbol{Y}_{xy}\}_{x,y\in B}$. Our estimate for $\boldsymbol{\theta}_u\boldsymbol{\theta}_v^{-1}$ is $\prod_{k\geq 0} W_{D_{u,k}}^{(k)} W_{D_{v,k}}^{(k)}$. For some large enough $k_\star$ we have that $D_{u,k_\star} = D_{v,k_\star}$ and so $W_{D_{u,k}}^{(k)} W_{D_{v,k}}^{(k)} = 1$ for all $k \geq k_\star$. The product of synchronization variables at u up to level k will be denoted as

$$\widetilde{W}_u^{(k)} = \prod_{\ell=1}^{k} W_{D_{u,\ell}}^{(\ell)}. \tag{7.2}$$

We say that two blocks $B, B' \in \mathcal{B}^{(k)}$ are adjacent (denoted $B \sim B'$) if there exist $x \in B, x \in B'$ such that $(x, x') \in E$. In this case there are exactly ℓ_k such pairs. We say that $B \sim B'$ is an honest edge if the following event holds

$$\mathcal{A}^{(k)}(B, B') = \left\{ \sum_{x\in B, x'\in B'} \boldsymbol{Y}_{xx'}\boldsymbol{\theta}_x\boldsymbol{\theta}_{x'} \geq \frac{9}{10}\ell_k \right\}. \tag{7.3}$$

This condition will mean that edges between vertices along the cut will be informative as we try to synchronize them.

Next we recursively define the set of good level k blocks $\mathcal{G}^{(k)}$. A block $B \in \mathcal{B}^{(k)}$ is *good* if

- There is at most one bad $(k-1)$-level sub-block of B, that is

$$\left|\left\{B_i \in \mathcal{B}^{(k-1)} : B_i \subset B, B_i \notin \mathcal{G}^{(k-1)}\right\}\right| \leq 1. \tag{7.4}$$

- All level $k-1$ sub-block edges are honest,

$$\bigcap_{\substack{B_1, B_2 \in \mathcal{B}^{(k-1)} \\ B_1, B_2 \subset B,\, B_1 \sim B_2}} \mathcal{A}^{(k-1)}(B_1, B_2). \tag{7.5}$$

Claim 7.1. *There exists $p_\star > 0$ such that, if $0 < p < p_\star$ then for all $B \in \mathcal{B}^{(k)}$*

$$\mathbb{P}(B \in \mathcal{G}^{(k)}) \geq 1 - 2^{-200k-200}. \tag{7.6}$$

Proof. We will establish (7.6) inductively. Note that blocks at level 0 are good. First we estimate the probability that the honest edge condition holds. Assuming that $p_\star \leq \frac{1}{40}$,

$$\begin{aligned}\mathbb{P}\big(\mathcal{A}^{(k-1)}(B_1, B_2)\big) &= \mathbb{P}\Big(\mathrm{Bin}(\ell_{k-1}, 1-p) \geq \frac{9}{10}\ell_{k-1}\Big) \\ &\geq \mathbb{P}\Big(\mathrm{Bin}\Big(\ell_{k-1}, \frac{39}{40}\Big) \geq \frac{9}{10}\ell_{k-1}\Big) \geq 1 - \exp\big(-\kappa 2^{10k(k-1)}\big)\end{aligned}$$

for some $\kappa > 0$. Thus

$$\mathbb{P}\big(\mathcal{A}^{(k-1)}(B_1, B_2)\big) \geq 1 - 2^{-400k-800} \tag{7.7}$$

for all sufficiently large k. By taking $p_\star$ small enough equation (7.7) holds for small k as well and thus for all k. Hence, since there are 2^{40k} level $k-1$ sub-blocks in each level k block we have that,

$$\mathbb{P}\Bigg(\bigcap_{\substack{B_1, B_2 \in \mathcal{B}^{(k-1)} \\ B_1, B_2 \subset B}} \mathcal{A}^{(k)}(B_1, B_2)\Bigg) \geq 1 - 2^{40k+1} \cdot 2^{-400(k-1)-800} \geq 1 - 2^{-200k-201}. \tag{7.8}$$

Since there are no bad sub-blocks at level 0 this implies (7.6) for $k = 1$. For some $k \geq 2$, assume inductively that equation (7.6) holds up to $k-1$. Then, since the event that blocks are good are independent, for $B \in \mathcal{B}^{(k)}$,

$$\begin{aligned}\mathbb{P}\Big(\big|\{B' \in \mathcal{B}^{(k-1)} : B' \subset B, B' \notin \mathcal{G}^{(k-1)}\}\big| \geq 2\Big) \\ &= \mathbb{P}\big(\mathrm{Bin}\big(2^{40k}, 2^{-200(k-1)-200}\big) \geq 2\big) \\ &\leq \binom{2^{40k}}{2}\big(2^{-200k}\big)^2 \leq 2^{-320k} \leq 2^{-200k-240}.\end{aligned}$$

Combining with equation (7.8) we have that

$$\mathbb{P}(B \in \mathcal{G}^{(k)}) \geq 1 - 2^{-200k-200},$$

as required. □

Next we describe how to inductively construct the synchronization variables $W_B^{(k)}$ in a $k+1$ block B^*. For $B_1 \sim B_2$ k-level sub-blocks of B^* we let

$$\boldsymbol{Y}_{B_1,B_2} = \mathrm{sign}\Bigg(\sum_{B_1 \ni x \sim y \in B_2} \widetilde{W}_x^{(k)} \widetilde{W}_y^{(k)} \boldsymbol{Y}_{xy}\Bigg).$$

We assign the $W_B^{(k)}$ as follows:

1. A quartet is a collection of 4 sub-blocks $B_1 \sim B_2 \sim B_3 \sim B_4 \sim B_1$ that form a square of side length $2\ell_k$. A quartet is incoherent if $\prod_{i=1}^4 \boldsymbol{Y}_{B_i,B_{i+1}} = -1$ where we take $B_5 = B_1$. Let $\mathcal{I}_{B*}^{(k)}$ be the set of sub-blocks of B^* that appear in no incoherent quartets. It is possible for $\mathcal{I}_{B*}^{(k)}$ to be disconnected, in that case take $\mathcal{I}_{B*}^{(k)}$ to be the largest component.

2. If possible, assign $W_B^{(k)}$ for all $B \in \mathcal{I}_{B*}^{(k)}$ such that for all adjacent sub-blocks $B_1, B_2 \in \mathcal{I}_{B*}^{(k)}$ we have that

$$W_{B_1}^{(k)} W_{B_2}^{(k)} = \boldsymbol{Y}_{B_1,B_2} \tag{7.9}$$

Denote the event that such an assignment is possible as $\mathcal{H}_{B*}^{(k+1)}$. If such an assignment is not possible set all the $W_B^{(k)} = 1$. Set $W_B^{(k)} = 1$ for all $B \in (\mathcal{I}_{B*}^{(k)})^c$.

In the following we will write $\mathcal{I} = \mathcal{I}^{(k)} = \mathcal{I}_{B*}^{(k)}$ omitting arguments when clear from the context. Note that on the event $\mathcal{H}_{B*}^{(k+1)}$, the $W_B^{(k)}$ can be found efficiently by assigning the variables iteratively to satisfy equation (7.9).

Claim 7.2. *For $k \geq 1$, if $B \in \mathcal{G}^{(k)}$ is good then the following hold:*

(1) *$\mathcal{H}_B^{(k)}$ holds.*

(2) *There exists a random variable $S_B^{(k)} \in \{-1, 1\}$ such that if $x \in B$ and on the event*

$$\bigcap_{j=0}^{k-1} \left\{ \{D_{x,j} \in \mathcal{G}^{(j)}\} \cap \{D_{x,j} \in \mathcal{I}^{(j)}\} \right\} \tag{7.10}$$

we have that

$$\boldsymbol{\theta}_x = S_B^{(k)} \widetilde{W}_x^{(k)}. \tag{7.11}$$

(3) *Furthermore, for any $B' \in \mathcal{G}^{(k)}$ with $B' \sim B$,*

$$\sum_{x \in B \cap \partial B'} S_B^{(k)} \widetilde{W}_x^{(k)} \boldsymbol{\theta}_x \geq (1 - 2^{-8} + 2^{-10k}) \ell_k. \tag{7.12}$$

(Here $\partial B' \equiv \{x \in \mathbb{Z}^2 : \operatorname{dist}(x, B') = 1\}$.)

Note that we do not (and cannot) construct $S_B^{(k)}$ and observe that it is used in the analysis but not the construction. It accounts for the fact that we can only hope to recover the $\boldsymbol{\theta}_u$ up to a global multiplicative shift.

Proof of Claim 7.2. We proceed inductively. In the base case when $k = 0$ for $x = B \in \mathcal{G}^{(0)}$ we may set $S_x^{(0)} = \boldsymbol{\theta}_x$. With the convention that an empty product is 1 we have that $\widetilde{W}_x^{(0)} = 1$ and so

$$\boldsymbol{\theta}_x = S_x^{(0)} \widetilde{W}_x^{(0)}.$$

Now we assume the claim holds for all $k' < k$ and consider a good block $B \in \mathcal{G}^{(k)}$.

1. For any good $(k-1)$-level sub-blocks, $B_1 \sim B_2$ in B

$$\boldsymbol{Y}_{B_1,B_2} = \operatorname{sign}\Big(\sum_{B_1 \ni x \sim y \in B_2} \widetilde{W}_x^{(k-1)} \widetilde{W}_y^{(k-1)} \boldsymbol{Y}_{xy}\Big) \tag{7.13}$$
$$= \operatorname{sign}\Big(S_{B_1}^{(k-1)} S_{B_2}^{(k-1)} \sum_{B_1 \ni x \sim y \in B_2} S_{B_1}^{(k-1)} \widetilde{W}_x^{(k-1)} S_{B_2}^{(k-1)} \widetilde{W}_y^{(k-1)} \boldsymbol{Y}_{xy}\Big).$$

Our inductive hypothesis implies that there are at most $2^{-8}\ell_{k-1}$ vertices x in this sum with $S_B^{(k-1)} \widetilde{W}_x^{(k-1)} \neq \boldsymbol{\theta}_x$, thus

$$\sum_{B_1 \ni x \sim y \in B_2} S_{B_1}^{(k-1)} \widetilde{W}_x^{(k-1)} S_{B_2}^{(k-1)} \widetilde{W}_y^{(k-1)} \boldsymbol{Y}_{xy} \geq \sum_{B_1 \ni x \sim y \in B_2} \boldsymbol{\theta}_x \boldsymbol{\theta}_y \boldsymbol{Y}_{xy} - 4 \cdot 2^{-8} \ell_{k-1}, \tag{7.14}$$

and so since $\mathcal{A}^{(k-1)}(B_1, B_2)$ holds,

$$\sum_{B_1 \ni x \sim y \in B_2} S_{B_1}^{(k-1)} \widetilde{W}_x^{(k-1)} S_{B_2}^{(k-1)} \widetilde{W}_y^{(k-1)} \boldsymbol{Y}_{xy} \geq \Big(\frac{9}{10} - 4 \cdot 2^{-8}\Big)\ell_{k-1} > 0. \tag{7.15}$$

Combining with equation (7.13) we have that

$$\boldsymbol{Y}_{B_1,B_2} = \operatorname{sign}\Big(S_{B_1}^{(k-1)} S_{B_2}^{(k-1)}\Big). \tag{7.16}$$

It follows that every quartet of good sub-blocks is coherent. If all of the $(k-1)$-level quartets of sub-blocks of B are coherent then there are exactly two assignments of $W_{B_i}^{(k-1)}$ (related by a multiplicative factor of -1) satisfying $W_{B_1}^{(k-1)} W_{B_2}^{(k-1)} = \boldsymbol{Y}_{B_1,B_2}$. If there is one or more incoherent quartet, this must include the single bad sub-block. The sub-blocks in $\mathcal{I}$ are good and there exist two assignments satisfying $W_{B_1}^{(k)} W_{B_2}^{(k)} = \boldsymbol{Y}_{B_1,B_2}$ for all $B_1, B_2 \in \mathcal{I}$, which are

$$W_{B_i}^{(k-1)} \equiv S_{B_i}^{(k-1)} \quad \text{or} \quad W_{B_i}^{(k-1)} \equiv -S_{B_i}^{(k-1)}. \tag{7.17}$$

In either case the procedure will construct $W_{B_i}^{(k)}$ satisfying (7.17) on $\mathcal{I}$ and $\mathcal{H}_B^{(k)}$ holds. We set $S_B^{(k)}$ so that

$$S_B^{(k)} W_{B_i}^{(k-1)} \equiv S_{B_i}^{(k-1)}.$$

2. To verify condition (7.11) we see that for $x \in B_i$,

$$S_B^{(k)} \widetilde{W}_x^{(k)} = S_B^{(k)} W_{B_i}^{(k)} \widetilde{W}_x^{(k-1)} = S_{B_i}^{(k-1)} \widetilde{W}_x^{(k-1)} = \boldsymbol{\theta}_x,$$

where the last equality used the inductive hypothesis.

3. It remains to check the condition on the boundary of B adjacent to some good block B'. Since any sub-block in $\mathcal{I}^c$ must be in a quartet with a bad sub-block, there are at most 3 on any side of B. Thus, summing over sub-blocks B_i of B we have that

$$\begin{aligned}\sum_{x\in B\cap\partial B'} S_B^{(k)}\widetilde{W}_x^{(k)}\boldsymbol{\theta}_x &= \sum_{B_i:B_i\sim B'}\ \sum_{x\in B_i\cap\partial B'} S_B^{(k)}\widetilde{W}_x^{(k)}\boldsymbol{\theta}_x\\ &\geq \sum_{B_i\in\mathcal{I}:B_i\sim B'}\ \sum_{B_i\ni x\sim y\in B'} S_B^{(k)}\widetilde{W}_x^{(k)}\boldsymbol{\theta}_x - 3\ell_{k-1}\\ &\geq \left(1-2^{-8}+2^{-10(k-1)}\right)\ell_{k-1}\left(2^{20k}-3\right) - 3\ell_{k-1}\\ &\geq \left(1-2^{-8}+2^{-10k}\right)\ell_k,\end{aligned}$$

which establishes (7.12). □

By the proceeding claim, if u and v are in the same k-level block on the event

$$\mathcal{J}_{uv}^{(k)} = \bigcap_{j=0}^{k-1}\left\{\{D_{u,j}, D_{v,j}\in\mathcal{G}^{(k)}\}\cap\{D_{u,j}, D_{v,j}\in\mathcal{I}\}\right\}$$

we have that

$$\widetilde{W}_u^{(k-1)}\widetilde{W}_v^{(k-1)} = \boldsymbol{\theta}_u S_B^{(k)}\boldsymbol{\theta}_v S_B^{(k)} = \boldsymbol{\theta}_u\boldsymbol{\theta}_v\,. \tag{7.18}$$

so $\widetilde{W}_u^{(k-1)}\widetilde{W}_v^{(k-1)}$ correctly recovers $\boldsymbol{\theta}_u\boldsymbol{\theta}_v$. A sufficient condition for $D_{u,k}\in\mathcal{G}^{(k)}\cap\mathcal{I}$ is that $D_{u,k}$ and the 8 sub-blocks surrounding it are all good. Thus

$$\mathbb{P}\left(\mathcal{J}_{uv}^{(k)}\right) \geq 1 - \sum_{k'\geq 1} 18\,\mathbb{P}\left(D_{u,k}\in\mathcal{G}^{(k)}\right) \geq 1 - 18\sum_{k'\geq 1} 2^{-200k-200} \geq \frac{9}{10}.$$

Thus

$$\mathbb{P}\left(\widetilde{W}_u^{(k-1)}\widetilde{W}_v^{(k-1)} = \boldsymbol{\theta}_u\boldsymbol{\theta}_v\right) \geq \frac{8}{10}$$

and so the success probability of recovery is at least $\frac{8}{10} > \frac{1}{2}$ independent of the distance between u and v which completes the proof of Theorem 3.

Remark 7.1. This estimator in fact leads to a global estimator for all the $\boldsymbol{\theta}$ which, up to a global sign flip, will correctly reconstruct a $\frac{9}{10}$ fraction of the θ_u. The argument readily generalizes to higher dimensions for discrete groups. An interesting question is whether the multi-scale approach could be generalized to continuous groups when $d \geq 3$ or $d \geq 4$. This seems more subtle as discreteness was used to rule out errors that gradually accumulate over different scales.

Acknowledgements. This work was partially supported by the following grants: NSF CAREERAward CCF-1552131, NSF CSOI CCF-0939370 (E.A.); NSF CCF-1319979, NSF DMS-1613091 (A.M); NSF DMS-1352013 (A.S.); NSF CCF-1553751 and a Sloan Research Fellowship (N.S). We thank the American Institute of Mathematics (San Jose, CA) where part of this work was carried out.

A. Proof of Eq. (2.1) for O(m) synchronization

Here we prove the remark that, under the model of Example 1.2,

$$\mathbb{E}\{\boldsymbol{Y}_{xy}|\boldsymbol{\theta}\} = \lambda(\sigma^2)\,\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y .$$

Fixing for simplicity $x = 1$, $y = 2$ and dropping the indices x, y unless necessary, we have $\boldsymbol{Y} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{V}}^{\mathsf{T}}$, where $\widetilde{\boldsymbol{X}} = \boldsymbol{\theta}_1^{-1}\boldsymbol{\theta}_2 + \sigma \boldsymbol{Z}$ has singular value decomposition $\widetilde{\boldsymbol{X}} = \widetilde{\boldsymbol{U}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{V}}^{\mathsf{T}}$.

Let $\boldsymbol{X} = \boldsymbol{\theta}_1\widetilde{\boldsymbol{X}}\boldsymbol{\theta}_2^{-1} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{T}}$. Our claim is equivalent to $\mathbb{E}\{\boldsymbol{U}\boldsymbol{V}^{\mathsf{T}}|\boldsymbol{\theta}\} = \lambda(\sigma^2)\boldsymbol{I}$. By rotational invariance of the Gaussian distribution, we have $\boldsymbol{X} = \boldsymbol{I} + \sigma\widetilde{\boldsymbol{G}}$ for $(G_{ij})_{1\le i,j\le m} \sim_{\text{iid}} \mathsf{N}(0,1)$ or, equivalently, $\boldsymbol{X} = \boldsymbol{Q}^{\mathsf{T}}(\boldsymbol{I} + \sigma\boldsymbol{G})\boldsymbol{Q}$ for any $\boldsymbol{Q}$ in $\mathrm{O}(m)$. Using the last representation,

$$\boldsymbol{E} = \mathbb{E}\{\boldsymbol{U}\boldsymbol{V}^{\mathsf{T}}|\boldsymbol{\theta}\} = \mathbb{E}\{\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{U}\boldsymbol{V}^{\mathsf{T}}\boldsymbol{Q}|\boldsymbol{\theta}\}$$

for $\boldsymbol{I} + \sigma\boldsymbol{G} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{T}}$. This implies that $\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{E}\boldsymbol{Q} = \boldsymbol{E}$ for any orthogonal matrix $\boldsymbol{Q}$, which can hold only if $\boldsymbol{E} = \lambda\boldsymbol{I}$ for some scalar λ.

Continuity and the limit values of $\lambda(\sigma^2)$ are straightforward.

B. Proof of Lemma 2.2

The first statement is a direct consequence of Pinsker inequality. Indeed, for any two random variables X, Y, denoting by

$$\mathrm{I}(U;W) = D\big(\mathbb{P}_{U,W}\big\|\mathbb{P}_U \times \mathbb{P}_W\big) \ge 2\,\big\|\mathbb{P}_{U,W} - \mathbb{P}_U \times \mathbb{P}_W\big\|_{\text{TV}}^2 . \tag{B.1}$$

The claim follows by applying this inequality to $U = T_{xy}(\boldsymbol{Y})$, $W = \boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$.

In order to prove the second claim, note that, if $(\boldsymbol{\theta}_x)_{x\in V} \sim_{\text{iid}} \mathbb{P}_{\text{Haar}}$, we have

$$\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y \sim \mathbb{P}_{\text{Haar}}$$

for any $x \ne y$, and therefore, under the product measure, $\mathbb{P}_{\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y} \times \mathbb{P}_{T_{xy}(\boldsymbol{Y})}$ we obtain $T_{xy}(\boldsymbol{Y})(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y)^{-1} \sim \mathbb{P}_{\text{Haar}}$. Therefore,

$$\big\|\mathbb{P}_{(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y, T_{xy}(\boldsymbol{Y}))} - \mathbb{P}_{\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y} \times \mathbb{P}_{T_{xy}(\boldsymbol{Y})}\big\|_{\text{TV}} \ge \big\|\mathbb{P}(T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1}\boldsymbol{\theta}_x \in \,\cdot\,) - \mathbb{P}_{\text{Haar}}(\,\cdot\,)\big\|_{\text{TV}} . \tag{B.2}$$

However, both $\boldsymbol{\theta}_y^{-1}\boldsymbol{\theta}_x$ and $T_{xy}(\boldsymbol{Y})$ remains unchanged under the transformation $\boldsymbol{\theta}_v \mapsto \boldsymbol{R}\boldsymbol{\theta}_v$ for some $\boldsymbol{R} \in \mathfrak{G}$. Therefore for any event A, letting $\boldsymbol{R} \sim \mathbb{P}_{\text{Haar}}$, we have

$$\begin{aligned}
\mathbb{P}\big(\boldsymbol{\theta}_x T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1} \in A\big) &= \mathbb{P}\big(T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1}\boldsymbol{\theta}_x \in \boldsymbol{\theta}_x^{-1}A\boldsymbol{\theta}_x\big) && \text{(B.3)}\\
&= \mathbb{E}_{\boldsymbol{R}}\mathbb{P}\big(T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1}\boldsymbol{\theta}_x \in \boldsymbol{R}^{-1}A\boldsymbol{R}\big)\\
&= \mathbb{E}F_A\big(T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1}\boldsymbol{\theta}_x\big)\,, && \text{(B.4)}
\end{aligned}$$

where we used Fubini's theorem, and introduced the function

$$F_A(\boldsymbol{Z}) \equiv \mathbb{P}_{\boldsymbol{R}}(\boldsymbol{Z} \in \boldsymbol{R}^{-1} A \boldsymbol{R}).$$

We therefore have,

$$\begin{aligned}
&\left\| \mathbb{P}\big(\boldsymbol{\theta}_x T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1} \in \cdot\big) - \mathbb{P}_{\text{Haar}}(\,\cdot\,) \right\|_{\text{TV}} \\
&\qquad = \sup_A \left| \mathbb{E} F_A(T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1}\boldsymbol{\theta}_x) - \mathbb{E}_{\text{Haar}} F(\boldsymbol{Z}) \right| && \text{(B.5)} \\
&\qquad \le \left\| \mathbb{P}(T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1}\boldsymbol{\theta}_x \in \cdot) - \mathbb{P}_{\text{Haar}}(\cdot) \right\|_{\text{TV}} && \text{(B.6)} \\
&\qquad \le \left\| \mathbb{P}_{(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y, T_{xy}(\boldsymbol{Y}))} - \mathbb{P}_{\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y} \times \mathbb{P}_{T_{xy}(\boldsymbol{Y})} \right\|_{\text{TV}}, && \text{(B.7)}
\end{aligned}$$

which proves the claim. □

C. Proof of Theorem 2

Let $p \equiv 1 - \inf_{\boldsymbol{y},\boldsymbol{\theta}_0} q(\boldsymbol{y}|\boldsymbol{\theta}_0)$. We can write the conditional probability density $q(\boldsymbol{y}|\boldsymbol{\theta}_0)$ (with respect to the Haar measure) as

$$q(\boldsymbol{y}|\boldsymbol{\theta}_0) = (1-p) + p\, q_*(\boldsymbol{y}|\boldsymbol{\theta})\,. \tag{C.1}$$

Hence observations $(\boldsymbol{Y}_{xy})_{(x,y)\in E}$ can be generated as follows. First draw independent random variables $(U_{xy})_{(x,y)\in E} \sim_{\text{iid}} \text{Bernoulli}(p)$. Then, for each $(x,y) \in E$ such that $U_{xy} = 1$, draw an independent observation

$$\boldsymbol{Y}_{xy} \sim q_*(\cdot\,|\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y).$$

For $(x,y) \in E$ such that $U_{xy} = 0$, draw $\boldsymbol{Y}_{xy}$ according to the Haar measure.

To upper bound the total variation distance in Eq (2.4) we consider the easier problem in which instead of $\boldsymbol{Y}$, we are given all the Bernoulli variables $U = (U_{xy})_{(x,y)\in E}$ and, for each $(x,y) \in E$ such that $U_{xy} = 1$ we are given the group difference $\boldsymbol{D}_{xy} = \boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$. Denoting by

$$\boldsymbol{D} = \{\boldsymbol{D}_{xy}\}_{(x,y)\in E, U_{xy}=1},$$

we then have

$$\begin{aligned}
&\left\| \mathbb{P}\big(\boldsymbol{\theta}_x T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1} \in \cdot\big) - \mathbb{P}_{\text{Haar}}(\,\cdot\,) \right\|_{\text{TV}} \\
&\qquad \le \sup_{\widetilde{T}_{xy}} \left\| \mathbb{P}\big(\boldsymbol{\theta}_x T_{xy}(U;\boldsymbol{D})\boldsymbol{\theta}_y^{-1} \in \cdot\big) - \mathbb{P}_{\text{Haar}}(\,\cdot\,) \right\|_{\text{TV}}. && \text{(C.2)}
\end{aligned}$$

Consider the percolation process defined by the variables U (whereby edge $(x,y) \in E$ is open if $U_{xy} = 1$), and denote by $x \sim_U y$ the event that x and y

are in the same percolation cluster. If x and y are not in the same percolation cluster, then the conditional distribution of $\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y$ conditional on $U; \boldsymbol{D}$ is uniformly on $\mathfrak{G}$. This implies that

$$\left\|\mathbb{P}\big(\boldsymbol{\theta}_x T_{xy}(\boldsymbol{Y})\boldsymbol{\theta}_y^{-1} \in \ \cdot\ \big) - \mathbb{P}_{\text{Haar}}\big(\ \cdot\ \big)\right\|_{\text{TV}} \le \mathbb{P}(x \sim_U y)\,. \tag{C.3}$$

For $p < p_c(d)$, the right hand side goes to 0 as $\|x - y\| \to \infty$ [13], which yields the claim. □

D. Proof of Theorem 4

For $s \in \mathbb{R}$ and $Z \sim g(\cdot)$, we define

$$\begin{aligned}\psi(s) = \|\mathbb{P}\big(Ze^{is} \in \cdot\big) - \mathbb{P}\big(Z \in \cdot\big)\|^2_{L^2(g)} \\ = \int_0^{2\pi}\left(\frac{g(e^{i(t-s)})}{g(e^{i(t)})} - 1\right)^2 g(e^{i(t)})\,\mathrm{d}t \\ = \int_0^{2\pi}\left(\frac{g(e^{i(t-s)})}{g(e^{i(t)})}\right)^2 g(e^{i(t)})\,\mathrm{d}t - 1\,.\end{aligned} \tag{D.1}$$

Note that $\psi(s)$ is twice differentiable, nonnegative and that $\psi(0) = 0$ so $\psi'(s) = 0$ and for some $\kappa = \kappa(g) > 0$,

$$|\psi(s)| \le \kappa|s|^2. \tag{D.2}$$

Let $u, v \in \mathbb{Z}^2$ with $L = \|u - v\|_2$, and define the function $h\colon \mathbb{Z}^2 \to \mathbb{R}$ by

$$h(x) = 1 - \frac{\log\big(1 + \min(\|x - u\|_2; L)\big)}{\log(L + 1)}\,. \tag{D.3}$$

Note that $h(u) = 1$, $h(v) = 0$ and $h(x) = 0$ for $\|x - u\|_2 \ge L$. Fix $\boldsymbol{\theta} \in \mathrm{U}(1)^{\mathbb{Z}^2}$, $s \in [0, 2\pi)$ and define $\boldsymbol{\theta}^{(s)}$ by letting $\boldsymbol{\theta}_x^{(s)} = e^{ish(x)}\,\boldsymbol{\theta}_x$. Let $\mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{Y} \in \cdot)$ be the the conditional distribution of the observations when the hidden variables are $\boldsymbol{\theta} = (\boldsymbol{\theta}_x)_{x\in V}$, and $\mathbb{P}_{\boldsymbol{\theta}^{(s)}}(\boldsymbol{Y} \in \cdot)$, when the hidden variables are $\boldsymbol{\theta}^{(s)} = (\boldsymbol{\theta}_x^{(s)})_{x\in V}$. These two measures are different because they correspond to data generated with different values of the underlying hidden variables. Also, they depend on L because $h(\cdot)$, and hence $\boldsymbol{\theta}^{(s)}$ does.

We then have, for a constant C,

$$
\begin{aligned}
\big\|\mathbb{P}_{\boldsymbol{\theta}^{(s)}}\big(\boldsymbol{Y} \in \cdot\big) - \mathbb{P}_{\boldsymbol{\theta}}\big(\boldsymbol{Y} \in \cdot\big)\big\|_{\mathrm{TV}}^2 & \\
&\le \big\|\mathbb{P}_{\boldsymbol{\theta}^{(s)}}\big(\boldsymbol{Y} \in \cdot\big) - \mathbb{P}_{\boldsymbol{\theta}}\big(\boldsymbol{Y} \in \cdot\big)\big\|_{L^2(\mathbb{P}_{\boldsymbol{\theta}})}^2 && \text{(D.4)}\\
&\overset{(a)}{=} \prod_{(x,y)\in E}\Big(1 + \big\|\mathbb{P}_{\boldsymbol{\theta}^{(s)}}\big(\boldsymbol{Y}_{xy} \in \cdot\big) - \mathbb{P}_{\boldsymbol{\theta}}\big(\boldsymbol{Y}_{xy} \in \cdot\big)\big\|_{L^2}^2\Big) - 1 && \text{(D.5)}\\
&= \prod_{(x,y)\in E}\big(1 + \psi\big(s\big(h(x) - h(y)\big)\big)\big) - 1 && \text{(D.6)}\\
&\le \prod_{(x,y)\in E}\big(1 + \kappa\, s^2\, |h(x) - h(y)|^2\big)\big) - 1 && \text{(D.7)}\\
&\le \prod_{\substack{(x,y)\in E\\ \|x-u\|\le L}}\Big\{1 + \frac{C}{\big(1 + \|x-u\|^2\big)\log^2 L}\Big\} - 1 && \text{(D.8)}\\
&= O(1/\log L), && \text{(D.9)}
\end{aligned}
$$

where (a) follows because both $\mathbb{P}_{\boldsymbol{\theta}}$ and $\mathbb{P}_{\boldsymbol{\theta}^{(s)}}$ are product measures. Taking expectation over s uniformly random in $[0, 2\pi)$ (denoted by E_s), we have, for any measurable set B,

$$
\mathrm{E}_s\mathbb{P}_{\boldsymbol{\theta}^{(s)}}\big(\boldsymbol{\theta}_u T_{uv}(\boldsymbol{Y})\boldsymbol{\theta}_v \in B\big) = \mathrm{E}_s\mathbb{P}_{\boldsymbol{\theta}^{(s)}}\big(\boldsymbol{\theta}_u^{(s)} T_{uv}(\boldsymbol{Y})\boldsymbol{\theta}_v^{(s),-1} \in e^{is}\, B\big)\,, \qquad \text{(D.10)}
$$

and therefore

$$
\big|\mathrm{E}_s\mathbb{P}_{\boldsymbol{\theta}^{(s)}}\big(\boldsymbol{\theta}_u^{(s)} T_{uv}(\boldsymbol{Y})\boldsymbol{\theta}_v^{(s),-1} \in e^{is}\, B\big) - \mathbb{P}_{\boldsymbol{\theta}}\big(\boldsymbol{\theta}_u T_{uv}(\boldsymbol{Y})\boldsymbol{\theta}_v^{-1} \in B\big)\big| = O\big(\log(1/L)\big)\,. \qquad \text{(D.11)}
$$

We next take expectation with respect to $(\boldsymbol{\theta}_x)_{x\in\mathbb{Z}^2}$ i.i.d. uniform in U(1). Note that under this distribution, also $(\boldsymbol{\theta}_x)_{x\in\mathbb{Z}^2}$ are i.i.d. uniform in U(1). Letting $\mathbb{P}(\,\cdot\,) = \mathbb{E}\mathbb{P}_{\boldsymbol{\theta}}(\,\cdot\,)$, we have

$$
\big|\mathrm{E}_s\mathbb{P}\big(\boldsymbol{\theta}_u T_{uv}(\boldsymbol{Y})\boldsymbol{\theta}_v^{-1} \in e^{is}\, B\big) - \mathbb{P}\big(\boldsymbol{\theta}_u T_{uv}(\boldsymbol{Y})\boldsymbol{\theta}_v^{-1} \in B\big)\big| = O\big(\log(1/L)\big)\,. \qquad \text{(D.12)}
$$

For any fixed B, $\boldsymbol{\xi}$, $\mathrm{P}_s(\boldsymbol{\xi} \in e^{is} B) = \mathbb{P}_{\mathrm{Haar}}(B)$ and hence we get

$$
\big|\mathbb{P}_{\boldsymbol{\theta}}\big(\boldsymbol{\theta}_u T_{uv}(\boldsymbol{Y})\boldsymbol{\theta}_v^{-1} \in B\big) - \mathbb{P}_{\mathrm{Haar}}(B)\big| = O\big(\log(1/L)\big)\,. \qquad \text{(D.13)}
$$

This proves the impossibility of weak recovery. □

References

[1] E. Abbe, Community detection and stochastic block models: recent developments, 2017. arXiv:1703.10146

[2] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer, Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery, *IEEE Trans. Network Sci. Eng.*, **1** (2014), no. 1, 10–22. MR 3349181

[3] N. Boumal and X. Cheng, Concentration of the kirchhoff index for Erdős–Rényi graphs, *Systems Control Lett.*, **74** (2014), 74–80. Zbl 1300.93151 MR 3278761

[4] P. Barooah and J. P. Hespanha, Estimation on graphs from relative measurements, *IEEE Control Systems*, **27** (2007), no. 4, 57–74.

[5] N. Boumal, On intrinsic Cramér-Rao bounds for Riemannian submanifolds and quotient manifolds, *IEEE Trans. Signal Process.*, **61** (2013), no. 7, 1809–1821. Zbl 06286153 MR 3038392

[6] I. Benjamini, R. Pemantle, and Y. Peres, Unpredictable paths and percolation, *Ann. Probab.*, **26** (1998), no. 3, 1198–1211. Zbl 0937.60070 MR 1634419

[7] N. Boumal, A. Singer, P.-A. Absil, and V. D. Blondel, Cramér–Rao bounds for synchronization of rotations, *Inf. Inference*, **3** (2014), no. 1, 1–39. Zbl 1308.94041 MR 3311447

[8] A. S. Bandeira, A. Singer, and D. A. Spielman, A Cheeger inequality for the graph connection Laplacian, *SIAM J. Matrix Anal. Appl.*, **34** (2013), no. 4, 1611–1630. Zbl 1287.05081 MR 3138103

[9] A. Chatterjee and V. Madhav Govindu, Efficient and robust large-scale rotation averaging, in *Proc. IEEE Intl. Conf. Comp. Vision (Sydney, Australia, 2013)*, 521–528, IEEE, 2013. DOI: 10.1109/ICCV.2013.70

[10] M. Cucuringu, A. Singer, and D. Cowburn, Eigenvector synchronization, graph rigidity and the molecule problem, *Inf. Inference*, **1** (2012), no. 1, 21–67. Zbl 1278.05231 MR 3311440

[11] H.-O. Georgii, *Gibbs measures and phase transitions*, second edition, De Gruyter Studies in Mathematics, 9, Walter de Gruyter & Co., Berlin, 2011. Zbl 1225.60001 MR 2807681

[12] A. Georges, D. Hansel, P. Le Doussal, and J.-P. Bouchaud, Exact properties of spin glasses. II. Nishimori's line: new results and physical implications, *J. Physique*, **46** (1985), no. 11, 1827–1836.

[13] G. Grimmett, *Percolation*, Springer, 1989. Zbl 0691.60089 MR 995460

[14] J. Hightower and G. Borriello, Location systems for ubiquitous computing, *Computer*, **34** (2001), no. 8, 57–66.

[15] T Horiguchi and T Morita, Existence of the ferromagnetic phase in competing random-bond Ising models on the square lattice, *J. Phys. A*, **15** (1982), no. 11, 3551–3560. MR 679113

[16] M. Hasenbusch, F. Parisen Toldin, A. Pelissetto, and E. Vicari, Critical behavior of the three-dimensional $\pm J$ Ising model at the paramagnetic-ferromagnetic transition line, *Phys. Rev. B*, **76** (2007), no. 9, 094402.

[17] Y. Iba, The Nishimori line and Bayesian statistics, *J. Phys. A*, **32** (1999), no. 21, 3875–3888. Zbl 0936.82014 MR 1696814

[18] A. Javanmard and A. Montanari, Localization from incomplete noisy distance measurements, *Found. Comput. Math.*, **13** (2013), no. 3, 297–345. Zbl 1269.05098 MR 3047003

[19] A. Javanmard, A. Montanari, and F. Ricci-Tersenghi, Phase transitions in semidefinite relaxations, *Proc. Natl. Acad. Sci. USA*, **113** (2016), no. 16, E2218–E2223. Zbl 1359.62188 MR 3494080

[20] M. Lelarge, L. Massoulié, and J. Xu, Reconstruction in the labelled stochastic block model, *IEEE Trans. Network Sci. Eng.*, **2** (2015), no. 4, 152–163. MR 3453283

[21] N. D. Mermin, Absence of ordering in certain classical systems, *J. Math. Phys*, **8** (1967), no. 5, 1061–1064.

[22] P. Moulon, P. Monasse, and R. Marlet, Global fusion of relative motions for robust, accurate and scalable structure from motion, in *Proc. IEEE Intl. Conf. Comp. Vision (Sydney, Australia, 2013)*, 3248–3255, IEEE, 2013. DOI: 10.1109/ICCV.2013.403

[23] B. Mohar, Some applications of laplace eigenvalues of graphs, in *Graph symmetry (Montreal, PQ, 1996)*, 225–275, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 497, Kluwer Acad. Publ., Dordrecht, 1997. Zbl 0883.05096 MR 1468791

[24] A. Montanari, Estimating random variables from random sparse observations, *Eur. Trans. Telecommun.*, **19** (2008), no. 4, 385–403.

[25] C. Moore, The computer science and physics of community detection: landscapes, phase transitions, and hardness, , *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, (2017), no. 121, 26–61. MR 3699594

[26] N. D. Mermin and H. Wagner, Absence of ferromagnetism or antiferromagnetism in one-or two-dimensional isotropic Heisenberg models, *Phys. Rev. Lett.*, **17** (1966), no. 22, 1133.

[27] H. Nishimori, Internal energy, specific heat and correlation function of the bond-random Ising model, *Progr. Theor. Exp. Phys.*, **66** (1981), no. 4, 1169–1181.

[28] H. Nishimori, *Statistical physics of spin glasses and information processing. An introduction*, Translated from the 1999 Japanese original, International Series of Monographs on Physics, 111, Oxford University Press, New York, 2001. Zbl 1103.82002 MR 2250384

[29] S. Oh, A. Montanari, and A. Karbasi, Sensor network localization from local connectivity: Performance analysis for the mds-map algorithm, in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, 1–5, IEEE, 2010. DOI: 10.1109/ITWKSPS.2010.5503144

[30] Y Ozeki and H. Nishimori, Phase diagram of gauge glasses, *J. Phys. A*, **26** (1993), no. 14, 3399–3429. MR 1236577

[31] O. Ozyesil and A. Singer, Robust camera location estimation by convex programming, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR, Boston, MA, 2015)*, 2674–2683, IEEE, 2015. DOI: 10.1109/CVPR.2015.7298883

[32] F. Parisen Toldin, A. Pelissetto, and E. Vicari, Strong-disorder paramagnetic-ferromagnetic fixed point in the square-lattice $\pm J$ Ising model, *J. Stat. Phys.*, **135** (2009), no. 5-6, 1039–1061. Zbl 1172.82322 MR 2548603

[33] A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, Message-passing algorithms for synchronization problems over compact groups, 2016. arXiv:1610.04583

[34] A. Singer, A remark on global positioning from local distances, *Proc. Natl. Acad. Sci. USA*, **105** (2008), no. 28, 9507–9511. Zbl 1205.86043 MR 2430205

[35] A. Singer, Angular synchronization by eigenvectors and semidefinite programming, *Appl. Comput. Harmon. Anal.*, **30** (2011), no. 1, 20–36. Zbl 1206.90116 MR 2737931

[36] A. Saade, M. Lelarge, F. Krzakala, and Zdeborová, Spectral detection in the censored block model, in *2015 IEEE International Symposium on Information Theory (ISIT, Hong Kong)*, 1184–1188, IEEE, 2015. DOI: 10.1109/ISIT.2015.7282642

[37] A. Singer and Y. Shkolnisky, Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming, *SIAM J. Imaging Sci.*, **4** (2011), no. 2, 543–572. Zbl 1216.92045 MR 2810897

[38] J. B. Tenenbaum, V. De Silva, and J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science*, **290** (2000), no. 5500, 2319–2323.

[39] L. Wasserman, *All of statistics. A concise course in statistical inference*, Springer Texts in Statistics. Springer-Verlag, New York, 2004. Zbl 1053.62005 MR 2055670

[40] K. Wilson, D. Bindel, and N. Snavely, When is rotations averaging hard?, *Computer Vision (ECCV, Amsterdam, 2016)*, 255–270, Springer, 2016.

[41] L. Wang and A. Singer, Exact and stable recovery of rotations for robust synchronization, *Inf. Inference*, **2** (2013), no. 2, 145–193. Zbl 1309.65070 MR 3311446

Received 26 June, 2017; revised 25 April, 2018

E. Abbe, Department of Mathematics, Princeton University,
Princeton, NJ 08544, USA

E-mail: eabbe@princeton.edu

L. Massoulié, MSR-Inria Joint Centre,
91128 Palaiseau, France

E-mail: laurent.massoulie@inria.fr

A. Montanari, Department of Electrical Engineering and Statistics, Stanford University,
Palo Alto, CA 94305, USA

E-mail: montanari@stanford.edu

A. Sly, Department of Mathematics, Princeton University,
Princeton, NJ 08544, USA

E-mail: asly@math.princeton.edu

N. Srivastava, Department of Mathematics, University of California – Berkeley,
Berkeley, CA 94720, USA

E-mail: nikhil@math.berkeley.edu