

# Group Synchronization on Grids

Emmanuel Abbe\*, Laurent Massoulié†, Andrea Montanari‡, Allan Sly§, Nikhil Srivastava¶

June 26, 2017

## Abstract

Group synchronization requires to estimate unknown elements  $(\theta_v)_{v \in V}$  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  $\theta_v$  are only determined up to a global action of the group, we consider the following weak recovery question. Can we determine the group difference  $\theta_u^{-1}\theta_v$  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  $d \geq 2$ . Viceversa, for some continuous groups, we prove that weak recovery is impossible for  $d = 2$ . Finally, for strong enough noise, weak recovery is always impossible.

## 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  $\mathbf{Y}_{u,v}$ . The objective is to estimate group elements  $(\theta_v)_{v \in V}$  associated to the vertices  $v \in V$ , under the assumption that the  $\mathbf{Y}_{u,v}$  are noisy observations of the group difference between the adjacent vertices. Roughly speaking (see below for a precise definition):

$$\mathbf{Y}_{uv} = \theta_u^{-1}\theta_v + \text{noise} . \tag{1.1}$$

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

---

\*Program in Applied and Computational Mathematics, and EE Department, Princeton University

†Inria, MSR-Inria Joint Centre

‡Department of Electrical Engineering and Department of Statistics, Stanford University

§Department of Mathematics, Princeton University

¶Department of Mathematics, University of California, Berkeley

each edge  $(u, v) \in E$  we are given  $Y_{uv}$  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  $p \in [0, 1/2)$ ,

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

with the  $(\mathbf{Y}_{uv})_{(u,v) \in E}$  conditionally independent given  $(\boldsymbol{\theta}_v)_{v \in V}$ . In other words  $\mathbf{Y}_{uv}$  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  $\mathfrak{G} = O(m)$ : the group of  $m \times m$  orthogonal matrices, with the following noise model. Let  $(\mathbf{Z}_{uv})_{(u,v) \in E}$  be an i.i.d. collection of matrices with i.i.d. standard normal entries, and define

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

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

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 [MMM13, CMG13, OS15, WBS16]. 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  $\mathfrak{G} = SO(3)$ , whereby  $\boldsymbol{\theta}_u$  describes the orientation of image  $u$ , and pairwise image registration is used to construct the relative rotations  $\mathbf{Y}_{uv}$ .

*Graph localization and positioning.* Consider a set of nodes with positions  $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$ . We want to reconstruct the nodes positions from noisy measurements of the pairwise distances  $\|\mathbf{x}_u - \mathbf{x}_v\|_2$ . This question arises in sensor network positioning [HB01, OMK10], imaging [CSC12, SS11], manifold learning [TDSL00], 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  $\|\mathbf{x}_u - \mathbf{x}_v\|_2 \leq \rho$  for  $\rho$  a certain communication range [Sin08, JM13].

Graph localization can be interpreted as a group synchronization problem in multiple ways. First, we can interpret the unknown position  $\mathbf{x}_v$  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 [CSC12]. 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  $SO(d)$  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 [Moo17, Abb17]. 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 [LMX15]. In particular the so-called censored block model considered in [SLKZ15] corresponds precisely to Example 1.1 on an Erdős-Rényi graph.

The literature on group synchronization is fairly recent and rapidly growing. The articles [Sin11, WS13] 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 [ABBS14, BSAB14, JMRT16, PWBM16]. Most of these studies use perturbation theoretic arguments which crucially rely on the fact that the Laplacian (or connection Laplacian, [BSS13]) 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  $d \geq 3$ , whose Laplacian pseudoinverses have appropriately bounded trace rather than norm.

In the present paper we are interested in  $G$  being the  $d$ -dimensional grid<sup>1</sup>,  $d \geq 2$ . Namely,  $V = \mathbb{Z}^d$ , and –to be definite– we orient edges in the positive direction:

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

where  $e_i = (0, \dots, 0, 1, 0, \dots, 0)$  is the  $i$ -th element of the canonical basis in  $\mathbb{R}^d$ . We expect other  $d$ -dimensional graphs (e.g. random geometric graphs) to present a qualitatively similar behavior.

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

*Is it possible to estimate  $\theta_x^{-1}\theta_y$  better than random guessing, as  $\|x - y\|_2 \rightarrow \infty$ ?*

Note that, in absence of noise (i.e. if  $\mathbf{Y}_{uv} = \theta_u^{-1}\theta_v$  exactly), the answer is always positive: we can multiply the observations  $\mathbf{Y}_{uv}$ 's along any path connecting  $x$  to  $y$  to reconstruct exactly  $\theta_x^{-1}\theta_y$ . However for any arbitrarily small noise level, errors add up along the path and this simple procedure is equivalent to random guessing for  $\|x - y\|_2 \rightarrow \infty$ . 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.

**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  $d \geq 2$ . We expect the same to hold generally for other discrete groups.

**Continuous groups,  $d = 2$ .** Vice-versa, for the simplest continuous group,  $\text{SO}(2)$ , 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 [Geo11]. We refer to Section 3 for a discussion of the connection with statistical physics.

---

<sup>1</sup>The case  $d = 1$  is somewhat trivial.

The rest of the paper is organized as follows. Section 2 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 3. 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<sup>2</sup> 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 4. Finally, Section 5 and 6 prove our positive results (reconstruction is possible) with other proofs deferred to the appendices.

**Notations.** Throughout the paper we use boldface symbols (e.g.  $\boldsymbol{\theta}_x, \mathbf{Y}_{xy}$ ) to denote elements of the group  $\mathfrak{G}$ , and normal symbols for other quantities (including vectors and matrices).

## 2 Main results

As mentioned above,  $G = (V, E)$  will be the infinite  $d$ -dimensional grid, and  $\mathfrak{G}$  a compact matrix group. Without loss of generality, we will assume  $\mathfrak{G} \subseteq O(m)$  (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  $\mathbf{Y} = (\mathbf{Y}_{xy})_{(x,y) \in E}$ ,  $\mathbf{Y}_{xy} \in \mathfrak{G}$ , that are conditionally independent given  $\boldsymbol{\theta}$ . We assume that observations are unbiased in the following sense:

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

where the parameter  $\lambda \in [0, 1]$  is a natural measure of the signal-to-noise ratio. In particular,  $\lambda = 1$  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  $\mathbb{E}\{\mathbf{Y}_{xy} | \boldsymbol{\theta}\} = (1 - 2p) \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y$ , and therefore  $\lambda = (1 - 2p)$ .
- For  $O(m)$  synchronization (cf. Example 1.2) we have  $\mathbb{E}\{\mathbf{Y}_{xy} | \boldsymbol{\theta}\} = \lambda(\sigma^2) \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y$  where  $\sigma^2 \mapsto \lambda(\sigma^2)$  is a continuous function on  $[0, \infty)$  with  $\lambda(\sigma^2) \rightarrow 1$  as  $\sigma^2 \rightarrow 0$  and  $\lambda(\sigma^2) \rightarrow 0$  as  $\sigma^2 \rightarrow \infty$  (see Appendix A).

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

$$\mathbb{P}(\mathbf{Y}_{x,y} \in \cdot | \boldsymbol{\theta}) = \mathbb{P}(\mathbf{Y}_{x,y} \in \cdot | \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y) = Q(\cdot | \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y). \quad (2.2)$$

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.

An estimator is a collection of measurable functions  $T_{uv} : \mathbf{Y} \mapsto T_{u,v}(\mathbf{Y}) \in \mathfrak{G}$  indexed by all vertex pairs  $u, v \in V$  (here  $\mathbf{Y} = (\mathbf{Y}_{xy})_{(x,y) \in E}$  denotes the set of all observations).

---

<sup>2</sup>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.

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

$$\liminf_{\|x-y\| \rightarrow \infty} \left\| \mathbb{P}(\boldsymbol{\theta}_x T_{xy}(\mathbf{Y}) \boldsymbol{\theta}_y^{-1} \in \cdot) - \mathbb{P}_{\text{Haar}}(\cdot) \right\|_{\text{TV}} \geq \varepsilon > 0. \quad (2.3)$$

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

If noise is strong enough, the problem becomes unsolvable.

**Theorem 2.** *Assume that:*

1.  $\mathbb{P}(\mathbf{Y}_{x,y} \in \cdot | \boldsymbol{\theta}) = \mathbb{P}(\mathbf{Y}_{x,y} \in \cdot | \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y)$ .
2.  $\mathbb{P}(\mathbf{Y}_{x,y} \in \cdot | \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y)$  has density  $q(\mathbf{y} | \boldsymbol{\theta}_0)$ ,  $\boldsymbol{\theta}_0 \in \mathfrak{G}$  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

$$\inf_{\mathbf{y}, \boldsymbol{\theta}_0} q(\mathbf{y} | \boldsymbol{\theta}_0) > 1 - p_c(d), \quad (2.4)$$

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  $\inf_{\mathbf{y}, \boldsymbol{\theta}} q(\mathbf{y} | \boldsymbol{\theta}) > 0$  (since  $p_c(d = 1) = 1$ ).

In  $d = 2$  the situation is more complicated. For certain discrete groups the problem is solvable at low enough noise: we consider here the case  $\mathfrak{G} = \mathbb{Z}_2$ , but we expect the same conclusion to hold more generally. A result related to the next one was established in [HM82] using a Peierls argument (Section 3 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 [HM82] does not imply immediately weak recovery.

**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  $p \leq p_*$  then the weak recovery problem is solvable.*

On the contrary, we expect that weak recovery is not possible in  $d = 2$  dimensions, for continuous groups even for very weak noise. This is analogous to the celebrated Mermin-Wagner theorem in statistical mechanics [MW66, Mer67]. For the sake of simplicity, we focus on the case of  $\mathfrak{G} = \text{SO}(2)$  which is isomorphic to  $\text{U}(1)$ , the group of complex variables of unit modulus, with ordinary multiplication. Let  $Z$  a  $\text{U}(1)$ -valued random variable with density  $g$  satisfying

$$g \in C^2, \quad \inf_{s \in [0, 2\pi]} g(e^{is}) > 0. \quad (2.5)$$

We consider observation on the edges corrupted by multiplicative noise

$$\mathbf{Y}_{xy} = \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y Z_{xy}, \quad (2.6)$$

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

**Theorem 4.** *If  $d = 2$  and  $\mathfrak{G} = \text{SO}(2)$  with noise model satisfying (2.5) and (2.6), then the weak recovery problem is not solvable.*

### 3 Bayesian posterior and connection to statistical physics

In this section it is convenient to assume a more general model in which the observations  $\mathbf{Y}_{xy} \in \mathbb{R}^{m \times m}$  are not necessarily elements of the matrix group  $\mathfrak{G}$ . We assume that the conditional distribution of the observations  $\mathbf{Y}_{xy}$  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

$$\frac{d\mathbb{P}}{d\mathbb{P}_\#}(\mathbf{Y}_{x,y}|\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}_0} \exp \left\{ -u(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y; \mathbf{Y}_{xy}) \right\}. \quad (3.1)$$

where  $u : \mathfrak{G} \times \mathbb{R}^{m \times m} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a measurable function bounded below. Applying Bayes formula, we can write the posterior  $\mu_{\mathbf{Y}}(B) = \mathbb{P}(\boldsymbol{\theta} \in B|\mathbf{Y})$  as

$$\mu_{\mathbf{Y}}(d\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}(\mathbf{Y})} \exp \left\{ - \sum_{(x,y) \in E} u(\boldsymbol{\theta}_x^{-1}\boldsymbol{\theta}_y; \mathbf{Y}_{xy}) \right\} \mu_0(d\boldsymbol{\theta}), \quad (3.2)$$

where  $\mu_0(d\boldsymbol{\theta}) = \mu_0(d\boldsymbol{\theta}_1) \cdots \mu_0(d\boldsymbol{\theta}_n)$  is the product Haar measure over the unknowns and  $\mathcal{Z}(\mathbf{Y})$  is a normalization constant. The joint distribution (3.2) takes the form of a Gibbs measure on the graph  $G$ .

**Remark 3.1.** For Eq. (3.2) 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  $E = \{(x, y) \in V \times V : y - x \in \{e_1, \dots, e_d\}\}$ . Then the quantity

$$\sup_{T_{xy}(\cdot)} \left\| \mathbb{P}(\boldsymbol{\theta}_x T_{xy}(\mathbf{Y}) \boldsymbol{\theta}_y^{-1} \in \cdot) - \mathbb{P}_{\text{Haar}}(\cdot) \right\|_{\text{TV}} \quad (3.3)$$

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  $L \rightarrow \infty$  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

$$u(\boldsymbol{\theta}\boldsymbol{\tau}; Y) = u(\boldsymbol{\tau}; \boldsymbol{\theta}^{-1}Y) = u(\boldsymbol{\theta}; Y\boldsymbol{\tau}^{-1}). \quad (3.4)$$

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  $\mathbf{Y}$ , the distribution  $\mu_{\mathbf{Y}}(\cdot)$  of  $\boldsymbol{\theta}$  in (3.2) coincides with that of  $\{\boldsymbol{\theta}_x \boldsymbol{\tau}_x^{-1}\}_{x \in V}$  where  $\boldsymbol{\theta}$  is distributed according to  $\mu_{\tilde{\mathbf{Y}}}(\cdot)$ , and  $\tilde{\mathbf{Y}}_{xy} = \boldsymbol{\tau}_x \mathbf{Y}_{xy} \boldsymbol{\tau}_y^{-1}$ . By taking  $\boldsymbol{\tau}_x = \boldsymbol{\theta}_{0,x}$  for all  $x$ , we can assume that  $\boldsymbol{\theta}_{0,x} = \mathbf{I}_m$  for all  $x$ , which then leads to the  $(\mathbf{Y}_{xy})_{(xy) \in E}$  being i.i.d. with common distribution

$$\mathbf{Y}_{xy} \sim \frac{1}{\mathcal{Z}_1} \exp \left\{ -u(\mathbf{I}_m; \mathbf{Y}_{xy}) \right\} \mathbb{P}_\#(d\mathbf{Y}_{xy}). \quad (3.5)$$

In the jargon of statistical physics, Gibbs measures of the form (3.2) with associated parameters distribution (3.5) are known as spin-glasses on the ‘Nishimori line.’ These were first introduced

for the case  $\boldsymbol{\theta}_x \in \{+1, -1\}$  [Nis81] and subsequently generalized to other groups in [GHLDB85]. Several results about spin glasses on the Nishimori line were derived in [ON93, Nis01] and the connection with Bayesian statistics was emphasized in [Iba99, Mon08]. The weak recovery phase transition corresponds to a paramagnetic-ferromagnetic phase transition in physics language.

**Example 3.2.** The simplest example is the so-called *random bond Ising model* which is obtained by taking  $\boldsymbol{\theta}_x \in \{+1, -1\}$  and

$$\mu_{\mathbf{Y}}(\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}(\mathbf{Y})} \exp \left\{ \beta \sum_{(x,y) \in E} \mathbf{Y}_{xy} \boldsymbol{\theta}_x \boldsymbol{\theta}_y \right\}, \quad (3.6)$$

where  $\mathbf{Y}_{xy} = +1$  with probability  $1 - p$  and  $\mathbf{Y}_{xy} = -1$  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  $d = 3$ ) [PTPV09, HTPV07].

**Example 3.3.** Take  $\mathfrak{G} = \text{O}(m)$  (the group of orthogonal matrices), and assume

$$\mathbf{Y}_{xy} = \boldsymbol{\theta}_{0,x}^{-1} \boldsymbol{\theta}_{0,y} + \sigma \mathbf{Z}_{xy} \quad (3.7)$$

where  $\mathbf{Z}_{xy}$  is a noise matrix with i.i.d. entries  $(\mathbf{Z}_{xy})_{ij} \sim \mathcal{N}(0, 1)$ . 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 (3.2) takes the form

$$\mu_{\mathbf{Y}}(d\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}(\mathbf{Y})} \exp \left\{ \beta \sum_{(x,y) \in E} \text{Tr}(\boldsymbol{\theta}_x \mathbf{Y}_{xy} \boldsymbol{\theta}_y^{\top}) \right\} \mu_0(d\boldsymbol{\theta}), \quad (3.8)$$

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

**Example 3.4.** In the case  $\mathfrak{G} = \text{SO}(2)$  we can identify  $\boldsymbol{\theta}_x$  with an angle in  $[0, 2\pi)$ , and let

$$\mathbf{Y}_{xy} = \boldsymbol{\theta}_{0,y} - \boldsymbol{\theta}_{0,x} + \mathbf{Z}_{xy}, \quad \text{mod } 2\pi, \quad (3.9)$$

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

The Gibbs measure (3.2) takes the form

$$\mu_{\mathbf{Y}}(d\boldsymbol{\theta}) = \frac{1}{\mathcal{Z}(\mathbf{Y})} \exp \left\{ - \sum_{(x,y) \in E} u(\mathbf{Y}_{xy} - \boldsymbol{\theta}_y + \boldsymbol{\theta}_x) \right\} \mu_0(d\boldsymbol{\theta}). \quad (3.10)$$

For the purpose of analysis we can assume  $\mathbf{Y}_{xy} = \mathbf{Z}_{xy}$ . 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 3.1.** *Consider the Gibbs measure (3.2) on the  $d$ -dimensional grid, with parameters  $\mathbf{Y}_{xy} \in \mathfrak{G}$  distributed according to Eq. (3.5) and satisfying Eq. (2.1). Then, the following hold:*

1. For  $d \geq 3$ , and  $\mathfrak{G} \subseteq \text{O}(m)$  is any compact matrix group, then there exists  $\lambda_{\text{UB}} < 1$  such that the model is in a ferromagnetic phase for any  $\lambda > \lambda_{\text{UB}}$ .
2. For the case of Example 3.2 (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 3.4 (i.e.  $\mathfrak{G} = \text{SO}(2)$ ) 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 3.3 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 3.3, note that we can always project  $\mathbf{Y}_{xy}$  onto the group  $\text{O}(m)$ , hence recovering the setting of Example 1.2. Since weak recovery is possible in the latter, it is also possible in the former.  $\square$

As already pointed out in Section 2, the existence of a ferromagnetic phase for Example 3.2 (i.e.  $\mathfrak{G} = \mathbb{Z}_2$ ) was already obtained in [HM82]. Note however that [HM82] establish existence of a spontaneous magnetization, while here we prove the existence of long range point-to-point correlation, which is equivalent to weak recovery.

## 4 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

$$Y_{x,y} = \theta_y - \theta_x + Z_{x,y}. \quad (4.1)$$

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

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

$$Y = D\theta + Z. \quad (4.2)$$

As usual,  $\theta$  can be determined only up to a global shift. To resolve this ambiguity, it is convenient to assume that  $\theta$  is centered:  $\langle \theta, 1 \rangle = 0$ . Consider the least square estimator  $\hat{\theta}(Y) = D^\dagger Y = DL^\dagger Y$  where  $\dagger$  denotes the pseudoinverse. A standard calculation [Was13, Theorem 13.13] yields the following formula for the mean square error

$$\text{MSE}(L, \sigma^2) \equiv \frac{1}{L^d} \mathbb{E}\{\|\hat{\theta}(Y) - \theta\|_2^2\} \quad (4.3)$$

$$= \frac{\sigma^2}{L^d} \text{Tr}_0((D^\top D)^\dagger) = \frac{\sigma^2}{L^d} \text{Tr}_0(\mathcal{L}^\dagger). \quad (4.4)$$

Here we denoted by  $\mathcal{L} = D^\top D$  the Laplacian of  $G$  and by  $\text{Tr}_0$  the trace on the subspace orthogonal to the all-ones vector. The eigenpairs of the Laplacian are [Moh97]:

$$v(p)_x = \frac{1}{L^{d/2}} e^{i\langle p, x \rangle}, \quad \lambda(p) = \sum_{i=1}^d [2 - 2 \cos(p_i)], \quad (4.5)$$

$$p \in B_L \equiv \left\{ \frac{2\pi}{L} (n_1, \dots, n_d) : n_i \in \{0, \dots, L-1\} \right\}. \quad (4.6)$$

Hence

$$\text{MSE}(L, \sigma^2) = \frac{\sigma^2}{L^d} \sum_{p \in B_L \setminus \{0\}} \frac{1}{\lambda(p)}. \quad (4.7)$$

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

**Fact 4.1.** *The mean square error of least-square estimation within the translation synchronization model of Eq. (4.1) is*

$$\frac{1}{\sigma^2} \text{MSE}(L, \sigma^2) = \begin{cases} \frac{L}{12} + O_L(1) & \text{for } d = 1, \\ \frac{1}{2\pi} \log L + O_L(1) & \text{for } d = 2, \\ C(d) + o_L(1) & \text{for } d \geq 3. \end{cases} \quad (4.8)$$

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  $d \geq 3$ .

## 5 Proof of Theorem 1

Throughout this section we assume a probability distribution  $\mathbb{P}$  over  $\boldsymbol{\theta}, \mathbf{Y}$  satisfying the unbiasedness condition Eq. (2.1). For most of our analysis, we consider general estimators  $T_{uv} : \mathbf{Y} \mapsto T_{uv}(\mathbf{Y}) \in \mathbb{R}^{m \times m}$  whose output is not necessarily in  $\mathfrak{G}$ , and let  $\mathbf{T}_{uv} = T_{uv}(\mathbf{Y})$  (as projecting them into  $\mathfrak{G}$  at the end can only increase their accuracy). Also, we set  $u(n) = (n, \dots, n) \in \mathbb{Z}^d$ , and denote by  $\mathcal{P}_+$  the set of infinite increasing paths in the grid, that start at 0.

Throughout the proof, we will use repeatedly the following two elementary facts. First, for any two matrices  $\mathbf{A}, \mathbf{B}$ ,  $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F$ . Second, if  $\mathbf{B}$  is an orthogonal matrix, then  $\|\mathbf{AB}\|_F = \|\mathbf{A}\|_F$ .

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

**Lemma 5.1.** *Consider  $d \geq 3$ . Then there exists an estimator  $T = (T_{u,v})$ , and absolute constants  $\lambda_0 < 1, C_0$  such that, for all  $\lambda > \lambda_0$  and all even  $n$ ,*

$$\mathbb{E}\{\boldsymbol{\theta}_0 \mathbf{T}_{0,u(n)} \boldsymbol{\theta}_{u(n)}^{-1}\} = \mathbf{I}_m, \quad (5.1)$$

$$\mathbb{E}\left\{\left\|\boldsymbol{\theta}_0 \mathbf{T}_{0,u(n)} \boldsymbol{\theta}_{u(n)}^{-1} - \mathbf{I}_m\right\|_F^2\right\} \leq C_0 m(1 - \lambda). \quad (5.2)$$

*Proof.* Benjamini, Pemantle and Peres [BPP<sup>+</sup>98] construct a probability measure  $\mu$  over paths in  $\mathcal{P}_+$  satisfying the so called exponential intersection property (EIT). Namely, there exist absolute constants  $\beta_* < 1$ ,  $C_*$  such that

$$(\mu \times \mu)\{(\gamma_1, \gamma_2) \in \mathcal{P}_+ \times \mathcal{P}_+ : |\gamma_1 \cap \gamma_2| \geq k\} \leq C_* \beta_*^k. \quad (5.3)$$

Let  $\mathcal{P}_+(v)$  be the set of increasing paths starting at 0 and ending at  $v \in \mathbb{Z}^d$ . For  $n$  even, we construct a probability measure  $\mu_n$  over  $\mathcal{P}_+(u(n))$  as follows. Define the hyperplane  $H(n) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 3n/2\}$ , and let  $R_n(x_1, x_2, x_3) = (n - x_1, n - x_2, n - x_3)$  denote the reflection with respect to this hyperplane. For  $\gamma \sim \mu$ , let  $\gamma^{(1,n)}$  denote the path obtained by stopping  $\gamma$  when it hits  $H(n)$ , and denote by  $\gamma^{(n)}$  its extension obtained by reflecting the with respect to  $H(n)$ . We let  $\mu_n$  be the probability distribution of  $\gamma^{(n)}$  (note that  $\gamma^{(n)}$  ends at  $u(n)$  by construction). It follows immediately from Eq. (5.3) that  $\mu_n$  satisfies the EIT for some new absolute constants  $C, \beta$ , that are independent of  $n$ :

$$\mu_n \times \mu_n\{(\gamma_1, \gamma_2) \in \mathcal{P}_+(u(n)) \times \mathcal{P}_+(u(n)) : |\gamma_1 \cap \gamma_2| \geq k\} \leq C \beta^k. \quad (5.4)$$

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

$$\mathbf{Y}_\gamma := \mathbf{Y}_{I_1(\gamma)} \mathbf{Y}_{I_2(\gamma)} \cdots \mathbf{Y}_{I_{3n}(\gamma)}, \quad (5.5)$$

$$\mathbf{T}_{0,u} := \frac{1}{\lambda^{3n}} \mathbb{E}_\gamma(\mathbf{Y}_\gamma), \quad (5.6)$$

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

$$\mathbb{E} \mathbf{T}_{0,u(n)} = \boldsymbol{\theta}_0^{-1} \boldsymbol{\theta}_{u(n)}. \quad (5.7)$$

Observe that if two paths  $\gamma_1, \gamma_2$  in  $\mathcal{P}_+(u(n))$  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  $\mathbb{E}_{\gamma_1, \gamma_2}$  expectation with respect to  $\gamma_1, \gamma_2 \sim_{iid} \mu_n$

$$\begin{aligned} \mathbb{E} \{ \mathbf{T}_{0,u} \mathbf{T}_{0,u}^\top \} &= \frac{1}{\lambda^{6n}} \mathbb{E}_{\gamma_1, \gamma_2} \mathbb{E} \mathbf{Y}_{\gamma_1} (\mathbf{Y}_{\gamma_2})^\top \\ &= \frac{1}{\lambda^{6n}} \mathbb{E}_{\gamma_1, \gamma_2} \mathbb{E} \mathbf{Y}_{I_1(\gamma_1)} \cdots (\mathbb{E} \mathbf{Y}_{I_{3n}(\gamma_1)} \mathbf{Y}_{I_{3n}(\gamma_2)}^\top) \mathbf{Y}_{I_{3n-1}(\gamma_2)}^\top \cdots \mathbf{Y}_{I_1(\gamma_2)}^\top \\ &\stackrel{(a)}{=} \frac{1}{\lambda^{6n}} \mathbb{E}_{\gamma_1, \gamma_2} \lambda^{|\gamma_1| + |\gamma_2| - 2|\gamma_1 \cap \gamma_2|} \mathbf{I}_m \\ &= \mathbb{E}_{\gamma_1, \gamma_2} \lambda^{-2|\gamma_1 \cap \gamma_2|} \mathbf{I}_m. \end{aligned}$$

where (a) follows by repeatedly applying the identity  $\mathbf{Y}_e \mathbf{Y}_e^\top = \mathbf{I}_m$  for any edge  $e$ , each time an intersection appears, and taking expectation with respect to  $\mathbf{Y}_{e_1}, \mathbf{Y}_{e_2}$  for not repeated edges. By this last expression, the trace  $\tau$  of  $\mathbb{E} \{ \mathbf{T}_{0,u} \mathbf{T}_{0,u}^\top \}$  reads  $\tau = m \mathbb{E} (\lambda^{-2X})$  where  $X$  is a random variable counting the number of intersections in two paths  $\gamma_1, \gamma_2$  independently drawn from  $\mu_n$ .

Thus for  $\lambda^2 > \beta$ ,

$$m^{-1}\tau = \sum_{x \geq 0} \lambda^{-2x} [\mathbb{P}(X \geq x) - \mathbb{P}(X \geq x+1)] \quad (5.8)$$

$$= 1 + \sum_{x > 0} \mathbb{P}(X \geq x) [\lambda^{-2x} - \lambda^{-2x+2}] \quad (5.9)$$

$$\leq 1 + (1 - \lambda^2) \sum_{x > 0} C (\beta/\lambda^2)^x \quad (5.10)$$

$$= 1 + (1 - \lambda^2) \frac{C\beta}{\lambda^2 - \beta}, \quad (5.11)$$

where the inequality follows from Eq. (5.4). Thus

$$\mathbb{E} \left\{ \|\boldsymbol{\theta}_0 \mathbf{T}_{0,u} \boldsymbol{\theta}_u^{-1} - \mathbf{I}_m\|_F^2 \right\} = \text{Tr} \mathbb{E} \left\{ \boldsymbol{\theta}_0 \mathbf{T}_{0,u} \mathbf{T}_{0,u}^\top \boldsymbol{\theta}_0^\top \right\} - 2 \text{Tr} \mathbb{E} \left\{ \boldsymbol{\theta}_0 \mathbf{T}_{0,u} \boldsymbol{\theta}_u^{-1} \right\} + m \quad (5.12)$$

$$= \tau - m \quad (5.13)$$

$$\leq (1 - \lambda^2) m \frac{C\beta}{\lambda^2 - \beta}, \quad (5.14)$$

where we used Eq. (5.7) together with our previous bound on  $\tau$ . The second statement of the Lemma follows.  $\square$

**Lemma 5.2.** *Consider any  $d \geq 3$  and fix  $\varepsilon > 0$ . For  $n \in \mathbb{N}$ ,  $j \in \{1, \dots, d\}$ , let  $v(j, n) \equiv n e_j$ . Then there exists an estimator  $T = (T_{u,v})_{u,v \in V}$ , and a constant  $\lambda(\varepsilon) < 1$ , such that, for all  $\lambda > \lambda(\varepsilon)$  and all  $n$ ,*

$$\mathbb{P} \left\{ \|\boldsymbol{\theta}_0 \mathbf{T}_{0,v(j,n)} \boldsymbol{\theta}_{v(j,n)}^{-1} - \mathbf{I}_m\|_F \geq \varepsilon \right\} \leq \varepsilon. \quad (5.15)$$

*Proof.* Without loss of generality, assume  $j = 1$ , and set for simplicity  $v(n) = v(j, n)$ . Consider first the case of  $n$  even and let  $w(n) \equiv (n/2, n/2, n/2, 0, \dots, 0)$ . Let  $(\mathbf{T}_{x,y}^{(*)})_{x,y \in V}$  be the estimator of Lemma 5.1 (where we use only the observations on the subgraph induced by the hyperplane  $\{x \in \mathbb{Z}^d : x_4 = \dots = x_d = 0\}$ ). Define

$$\mathbf{T}_{0,v(n)} = \mathbf{T}_{0,w(n)}^{(*)} \mathbf{T}_{w(n),v(n)}^{(*)}. \quad (5.16)$$

From the inequality

$$1 + \|\mathbf{X}_1 \mathbf{X}_2 - \mathbf{I}\|_F \leq (1 + \|\mathbf{X}_1 - \mathbf{I}\|_F)(1 + \|\mathbf{X}_2 - \mathbf{I}\|_F), \quad (5.17)$$

we get

$$\|\boldsymbol{\theta}_0 \mathbf{T}_{0,v(n)} \boldsymbol{\theta}_{v(n)}^{-1} - \mathbf{I}\|_F \leq \left(1 + \|\boldsymbol{\theta}_0 \mathbf{T}_{0,w(n)}^{(*)} \boldsymbol{\theta}_{w(n)}^{-1} - \mathbf{I}\|_F\right) \left(1 + \|\boldsymbol{\theta}_{w(n)} \mathbf{T}_{w(n),v(n)}^{(*)} \boldsymbol{\theta}_{v(n)}^{-1} - \mathbf{I}\|_F\right) - 1.$$

By Lemma 5.1 and Markov's inequality, the probability that one of the Frobenius norms in the right-hand side exceeds  $\delta > 0$  is at most  $C_0 m (1 - \lambda) / \delta^2$ . Thus with probability at least  $1 - 2C_0 m (1 - \lambda) / \delta^2$ , one has

$$\|\boldsymbol{\theta}_0 \mathbf{T}_{0,v(n)} \boldsymbol{\theta}_{v(n)}^{-1} - \mathbf{I}\|_F \leq (1 + \delta)^2 - 1.$$

The right-hand side is at most  $3\delta$  for  $\delta \leq 1$ . The announced result follows for the choice  $\lambda(\varepsilon) = 1 - \varepsilon^3 / (18mC_0)$ .  $\square$

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  $n$ ,*

$$\mathbb{P}\left\{\|\boldsymbol{\theta}_x \mathbf{T}_{x,y} \boldsymbol{\theta}_y^{-1} - \mathbf{I}_m\|_F \geq \varepsilon\right\} \leq \varepsilon. \quad (5.18)$$

*Proof.* Without loss of generality, assume  $x = 0$ . Further, for  $j \in \{0, \dots, d\}$  define  $w(j) \equiv (y_1, \dots, y_j, 0, \dots, 0)$ . In particular,  $w(0) = 0$  and  $w(d) = y$ . Let  $(\mathbf{T}_{x,y}^{(\#)})$  be the estimator of Lemma 5.2, and define

$$\mathbf{T}_{0,y} = \mathbf{T}_{w(0),w(1)}^{(\#)} \mathbf{T}_{w(1),w(2)}^{(\#)} \cdots \mathbf{T}_{w(d-1),w(d)}^{(\#)}. \quad (5.19)$$

By Lemma 5.2, for all  $\lambda > \lambda(\varepsilon_0)$  we have

$$\mathbb{P}\left(\max_{1 \leq j \leq d} \|\boldsymbol{\theta}_{w(j-1)} \mathbf{T}_{w(j-1),w(j)}^{(\#)} \boldsymbol{\theta}_{w(j)}^{-1} - \mathbf{I}_m\|_F \geq \varepsilon_0\right) \leq d\varepsilon_0. \quad (5.20)$$

By repeated application of Inequality (5.17), on the complement of the event in the right-hand side, one has

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

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

## 6 Proof of Theorem 3

We give a multi-scale scheme to reconstruct the unknowns  $\boldsymbol{\theta} = (\boldsymbol{\theta}_x)_{x \in \mathbb{Z}^2}$ . 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  $\ell_k$  as follows,

$$B_u^{(k)} = \{(x_1, x_2) \in \mathbb{Z}^2 : u_i = \lceil x_i / \ell_k \rceil\} \quad (6.1)$$

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  $\{\mathbf{Y}_{xy}\}_{x,y \in B}$ . 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  $W_{D_{u,k_\star}}^{(k_\star)} W_{D_{v,k_\star}}^{(k_\star)} = 1$  for all  $k \geq k_\star$ . The product of synchronization variables at  $u$  up to level  $k$  will be denoted as

$$\tilde{W}_u^{(k)} = \prod_{\ell=1}^k W_{D_{u,\ell}}^{(\ell)}. \quad (6.2)$$

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  $\ell_k$  such pairs. We say that  $B \sim B'$  is an honest edge if the following event holds

$$\mathcal{A}^{(k)}(B, B') = \left\{ \sum_{x \in B, x' \in B'} \mathbf{Y}_{xx'} \boldsymbol{\theta}_x \boldsymbol{\theta}_{x'} \geq \frac{9}{10} \ell_k \right\}. \quad (6.3)$$

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

$$\left| \left\{ B_i \in \mathcal{B}^{(k-1)} : B_i \subset B, B_i \notin \mathcal{G}^{(k-1)} \right\} \right| \leq 1. \quad (6.4)$$

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

$$\bigcap_{\substack{B_1, B_2 \in \mathcal{B}^{(k-1)} \\ B_1, B_2 \subset B, B_1 \sim B_2}} \mathcal{A}^{(k-1)}(B_1, B_2). \quad (6.5)$$

**Claim 6.1.** *There exists  $p_\star > 0$  such that, if  $0 < p < p_\star$  then for all  $B \in \mathcal{B}^{(k)}$*

$$\mathbb{P}(B \in \mathcal{G}^{(k)}) \geq 1 - 2^{-200k-200}. \quad (6.6)$$

*Proof.* We will establish (6.6) inductively. Note that blocks at level 0 are good. First we estimate the probability that the honest edge condition holds. Assuming that  $p_\star \leq \frac{1}{40}$ ,

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

for some  $\kappa > 0$ . Thus

$$\mathbb{P}\left(\mathcal{A}^{(k-1)}(B_1, B_2)\right) \geq 1 - 2^{-400k-800} \quad (6.7)$$

for all sufficiently large  $k$ . By taking  $p_\star$  small enough equation (6.7) 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,

$$\mathbb{P}\left(\bigcap_{\substack{B_1, B_2 \in \mathcal{B}^{(k-1)} \\ B_1, B_2 \subset B}} \mathcal{A}^{(k-1)}(B_1, B_2)\right) \geq 1 - 2^{40k+1} \cdot 2^{-400(k-1)-800} \geq 1 - 2^{-200k-201}. \quad (6.8)$$

Since there are no bad sub-blocks at level 0 this implies (6.6) for  $k=1$ . For some  $k \geq 2$ , assume inductively that equation (6.6) holds up to  $k-1$ . Then, since the event that blocks are good are independent, for  $B \in \mathcal{B}^{(k)}$ ,

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

Combining with equation (6.8) 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

$$\mathbf{Y}_{B_1, B_2} = \text{sign} \left( \sum_{B_1 \ni x \sim y \in B_2} \tilde{W}_x^{(k)} \tilde{W}_y^{(k)} \mathbf{Y}_{xy} \right)$$

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  $\prod_{i=1}^4 \mathbf{Y}_{B_i, B_{i+1}} = -1$  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

$$W_{B_1}^{(k)} W_{B_2}^{(k)} = \mathbf{Y}_{B_1, B_2} \quad (6.9)$$

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 (6.9).

**Claim 6.2.** *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

$$\bigcap_{j=0}^{k-1} \left\{ \{D_{x,j} \in \mathcal{G}^{(j)}\} \cap \{D_{x,j} \in \mathcal{I}^{(j)}\} \right\} \quad (6.10)$$

we have that

$$\boldsymbol{\theta}_x = S_B^{(k)} \tilde{W}_x^{(k)}. \quad (6.11)$$

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

$$\sum_{x \in B \cap \partial B'} S_B^{(k)} \tilde{W}_x^{(k)} \boldsymbol{\theta}_x \geq (1 - 2^{-8} + 2^{-10k}) \ell_k. \quad (6.12)$$

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

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 6.2.* 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  $\tilde{W}_x^{(0)} = 1$  and so

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

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  $B$

$$\begin{aligned} \mathbf{Y}_{B_1, B_2} &= \text{sign} \left( \sum_{B_1 \ni x \sim y \in B_2} \tilde{W}_x^{(k-1)} \tilde{W}_y^{(k-1)} \mathbf{Y}_{xy} \right) \\ &= \text{sign} \left( S_{B_1}^{(k-1)} S_{B_2}^{(k-1)} \sum_{B_1 \ni x \sim y \in B_2} S_{B_1}^{(k-1)} \tilde{W}_x^{(k-1)} S_{B_2}^{(k-1)} \tilde{W}_y^{(k-1)} \mathbf{Y}_{xy} \right) \end{aligned} \quad (6.13)$$

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

$$\sum_{B_1 \ni x \sim y \in B_2} S_{B_1}^{(k-1)} \tilde{W}_x^{(k-1)} S_{B_2}^{(k-1)} \tilde{W}_y^{(k-1)} \mathbf{Y}_{xy} \geq \sum_{B_1 \ni x \sim y \in B_2} \boldsymbol{\theta}_x \boldsymbol{\theta}_y \mathbf{Y}_{xy} - 4 \cdot 2^{-8} \ell_{k-1}, \quad (6.14)$$

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

$$\sum_{B_1 \ni x \sim y \in B_2} S_{B_1}^{(k-1)} \tilde{W}_x^{(k-1)} S_{B_2}^{(k-1)} \tilde{W}_y^{(k-1)} \mathbf{Y}_{xy} \geq \left( \frac{9}{10} - 4 \cdot 2^{-8} \right) \ell_{k-1} > 0. \quad (6.15)$$

Combining with equation (6.13) we have that

$$\mathbf{Y}_{B_1, B_2} = \text{sign} \left( S_{B_1}^{(k-1)} S_{B_2}^{(k-1)} \right). \quad (6.16)$$

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  $W_{B_1}^{(k-1)} W_{B_2}^{(k-1)} = \mathbf{Y}_{B_1, B_2}$ . 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  $W_{B_1}^{(k)} W_{B_2}^{(k)} = \mathbf{Y}_{B_1, B_2}$  for all  $B_1, B_2 \in \mathcal{I}$  which are,

$$W_{B_i}^{(k-1)} \equiv S_{B_i}^{(k-1)} \quad \text{or} \quad W_{B_i}^{(k-1)} \equiv -S_{B_i}^{(k-1)}. \quad (6.17)$$

In either case the procedure will construct  $W_{B_i}^{(k)}$  satisfying (6.17) 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 (6.11) we see that for  $x \in B_i$ ,

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

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

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

which establishes (6.12).  $\square$

By the preceding 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

$$\tilde{W}_u^{(k-1)} \tilde{W}_v^{(k-1)} = \boldsymbol{\theta}_u S_B^{(k)} \boldsymbol{\theta}_v S_B^{(k)} = \boldsymbol{\theta}_u \boldsymbol{\theta}_v. \tag{6.18}$$

so  $\tilde{W}_u^{(k-1)} \tilde{W}_v^{(k-1)}$  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

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

Thus

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

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.

## Acknowledgements

This work was partially supported by the following grants: NSF CAREER Award CCF-1552131, NSF CSOI CCF-0939370 (E.A.); NSF CCF-1319979, NSF DMS-1613091 (A.M); 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–  $\mathbb{E}\{\mathbf{Y}_{xy} | \boldsymbol{\theta}\} = \lambda(\sigma^2) \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y$ . Fixing for simplicity  $x = 1, y = 2$  and dropping the indices  $x, y$  unless necessary, we have  $\mathbf{Y} = \tilde{\mathbf{U}} \tilde{\mathbf{V}}^\top$  where  $\tilde{\mathbf{X}} = \boldsymbol{\theta}_1^{-1} \boldsymbol{\theta}_2 + \sigma \mathbf{Z}$  has singular value decomposition  $\tilde{\mathbf{X}} = \tilde{\mathbf{U}} \boldsymbol{\Sigma} \tilde{\mathbf{V}}^\top$ .

Let  $\mathbf{X} = \boldsymbol{\theta}_1 \tilde{\mathbf{X}} \boldsymbol{\theta}_2^{-1} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top$ . Our claim is equivalent to  $\mathbb{E}\{\mathbf{U} \mathbf{V}^\top | \boldsymbol{\theta}\} = \lambda(\sigma^2) \mathbf{I}$ . By rotational invariance of the Gaussian distribution, we have  $\mathbf{X} = \mathbf{I} + \sigma \tilde{\mathbf{G}}$  for  $(G_{ij})_{1 \leq i, j \leq m} \sim_{iid} \mathcal{N}(0, 1)$  or –equivalently–  $\mathbf{X} = \mathbf{Q}^\top (\mathbf{I} + \sigma \mathbf{G}) \mathbf{Q}$  for any  $\mathbf{Q}$  in  $O(m)$ . Using the last representation,  $\mathbf{E} = \mathbb{E}\{\mathbf{U} \mathbf{V}^\top | \boldsymbol{\theta}\} = \mathbb{E}\{\mathbf{Q}^\top \mathbf{U} \mathbf{V}^\top \mathbf{Q} | \boldsymbol{\theta}\}$  for  $\mathbf{I} + \sigma \mathbf{G} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top$ . This implies that  $\mathbf{Q}^\top \mathbf{E} \mathbf{Q} = \mathbf{E}$  for any orthogonal matrix  $\mathbf{Q}$ , which can hold only if  $\mathbf{E} = \lambda \mathbf{I}$  for some scalar  $\lambda$ .

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

## B Proof of Theorem 2

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

$$q(\mathbf{y} | \boldsymbol{\theta}_0) = (1 - p) + p q_*(\mathbf{y} | \boldsymbol{\theta}). \quad (\text{B.1})$$

Hence observations  $(\mathbf{Y}_{xy})_{(x,y) \in E}$  can be generated as follows. First draw independent random variables  $(U_{xy})_{(x,y) \in E} \sim_{iid} \text{Bernoulli}(p)$ . Then, for each  $(x, y) \in E$  such that  $U_{xy} = 1$ , draw an independent observation  $\mathbf{Y}_{xy} \sim q_*(\cdot | \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y)$ . For  $(x, y) \in E$  such that  $U_{xy} = 0$ , draw  $\mathbf{Y}_{xy}$  according to the Haar measure.

To upper bound the total variation distance in Eq (2.3) we consider the easier problem in which instead of  $\mathbf{Y}$ , 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  $\mathbf{D}_{xy} = \boldsymbol{\theta}_x^{-1} \boldsymbol{\theta}_y$ . Denoting by  $\mathbf{D} = \{\mathbf{D}_{xy}\}_{(x,y) \in E, U_{xy}=1}$ , we then have

$$\left\| \mathbb{P}(\boldsymbol{\theta}_x T_{xy}(\mathbf{Y}) \boldsymbol{\theta}_y^{-1} \in \cdot) - \mathbb{P}_{\text{Haar}}(\cdot) \right\|_{\text{TV}} \leq \sup_{\tilde{T}_{xy}} \left\| \mathbb{P}(\boldsymbol{\theta}_x T_{xy}(U; \mathbf{D}) \boldsymbol{\theta}_y^{-1} \in \cdot) - \mathbb{P}_{\text{Haar}}(\cdot) \right\|_{\text{TV}}. \quad (\text{B.2})$$

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  $U; \mathbf{D}$  is uniformly on  $\mathfrak{G}$ . This implies that

$$\left\| \mathbb{P}(\boldsymbol{\theta}_x T_{xy}(\mathbf{Y}) \boldsymbol{\theta}_y^{-1} \in \cdot) - \mathbb{P}_{\text{Haar}}(\cdot) \right\|_{\text{TV}} \leq \mathbb{P}(x \sim_U y). \quad (\text{B.3})$$

For  $p < p_c(d)$ , the right hand side goes to 0 as  $\|x - y\| \rightarrow \infty$  [Gri89], which yields the claim.

## C Proof of Theorem 4

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

$$\begin{aligned} \psi(s) &= \left\| \mathbb{P}(Z e^{is} \in \cdot) - \mathbb{P}(Z \in \cdot) \right\|_{L^2(g)}^2 = \int_0^{2\pi} \left( \frac{g(e^{i(t-s)})}{g(e^{it})} - 1 \right)^2 g(e^{it}) dt \\ &= \int_0^{2\pi} \left( \frac{g(e^{i(t-s)})}{g(e^{it})} \right)^2 g(e^{it}) dt - 1. \end{aligned} \quad (\text{C.1})$$

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

$$|\psi(s)| \leq \kappa |s|^2. \quad (\text{C.2})$$

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

$$h(x) = 1 - \frac{\log(1 + \min(\|x - u\|_2; L))}{\log(L + 1)}. \quad (\text{C.3})$$

Note that  $h(u) = 1$ ,  $h(v) = 0$  and  $h(x) = 0$  for  $\|x - u\|_2 \geq L$ . Fix  $\boldsymbol{\theta} \in \text{U}(1)^{\mathbb{Z}^2}$ ,  $s \in [0, 2\pi)$  and define  $\boldsymbol{\theta}^{(s)}$  by letting  $\boldsymbol{\theta}_x^{(s)} = e^{is} \boldsymbol{\theta}_x$ . Denote by  $\mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} \in \cdot)$  the conditional distribution of the observations given hidden variables  $\boldsymbol{\theta}$ . We then have, for a constant  $C$ ,

$$\left\| \mathbb{P}_{\boldsymbol{\theta}^{(s)}}(\mathbf{Y} \in \cdot) - \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} \in \cdot) \right\|_{\text{TV}}^2 \leq \left\| \mathbb{P}_{\boldsymbol{\theta}^{(s)}}(\mathbf{Y} \in \cdot) - \mathbb{P}_{\boldsymbol{\theta}}(\mathbf{Y} \in \cdot) \right\|_{L^2}^2 \quad (\text{C.4})$$

$$= \prod_{(x,y) \in E} \left( 1 + \left\| \mathbb{P}_{\nu_s}(\mathbf{Y}_{xy} \in \cdot) - \mathbb{P}_{\nu_0}(\mathbf{Y}_{xy} \in \cdot) \right\|_{L^2}^2 \right) - 1 \quad (\text{C.5})$$

$$= \prod_{(x,y) \in E} (1 + \psi(s(h(x) - h(y)))) - 1 \quad (\text{C.6})$$

$$\leq \prod_{(x,y) \in E} (1 + \kappa s^2 |h(x) - h(y)|^2) - 1 \quad (\text{C.7})$$

$$\leq \prod_{\substack{(x,y) \in E \\ \|x-u\| \leq L}} \left\{ 1 + \frac{C}{(1 + \|x - u\|^2) \log^2 L} \right\} - 1 \quad (\text{C.8})$$

$$= O(1/\log L). \quad (\text{C.9})$$

Taking expectation over  $s$  uniformly random in  $[0, 2\pi)$  (denoted by  $\mathbb{E}_s$ ), we have, for any measurable set  $B$ ,

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

and therefore

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

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

$$\left| \mathbb{E}_s \mathbb{P}(\boldsymbol{\theta}_u T_{uv}(\mathbf{Y}) \boldsymbol{\theta}_v^{-1} \in e^{is} B) - \mathbb{P}(\boldsymbol{\theta}_u T_{uv}(\mathbf{Y}) \boldsymbol{\theta}_v^{-1} \in B) \right| = O(\log(1/L)). \quad (\text{C.12})$$

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

$$\left| \mathbb{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_u T_{uv}(\mathbf{Y}) \boldsymbol{\theta}_v^{-1} \in B) - \mathbb{P}_{\text{Haar}}(B) \right| = O(\log(1/L)). \quad (\text{C.13})$$

This proves the impossibility of weak recovery.

## References

- [Abb17] Emmanuel Abbe, *Community detection and stochastic block models: recent developments*, arXiv:1703.10146 (2017).

- [ABBS14] Emmanuel Abbe, Afonso S Bandeira, Annina Bracher, and Amit Singer, *Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery*, IEEE Transactions on Network Science and Engineering **1** (2014), no. 1, 10–22.
- [BPP<sup>+</sup>98] Itai Benjamini, Robin Pemantle, Yuval Peres, et al., *Unpredictable paths and percolation*, The Annals of Probability **26** (1998), no. 3, 1198–1211.
- [BSAB14] Nicolas Boumal, Amit Singer, P-A Absil, and Vincent D Blondel, *Cramér–rao bounds for synchronization of rotations*, Information and Inference **3** (2014), no. 1, 1–39.
- [BSS13] Afonso S Bandeira, Amit Singer, and Daniel A Spielman, *A cheeger inequality for the graph connection laplacian*, SIAM Journal on Matrix Analysis and Applications **34** (2013), no. 4, 1611–1630.
- [CMG13] Avishek Chatterjee and Venu Madhav Govindu, *Efficient and robust large-scale rotation averaging*, Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 521–528.
- [CSC12] Mihai Cucuringu, Amit Singer, and David Cowburn, *Eigenvector synchronization, graph rigidity and the molecule problem*, Information and Inference **1** (2012), no. 1, 21–67.
- [Geo11] Hans-Otto Georgii, *Gibbs measures and phase transitions*, vol. 9, Walter de Gruyter, 2011.
- [GHLDB85] Antoine Georges, David Hansel, Pierre Le Doussal, and Jean-Philippe Bouchaud, *Exact properties of spin glasses. ii. nishimori’s line: new results and physical implications*, Journal de Physique **46** (1985), no. 11, 1827–1836.
- [Gri89] Geoffrey Grimmett, *Percolation?*, Springer, 1989.
- [HB01] Jeffrey Hightower and Gaetano Borriello, *Location systems for ubiquitous computing*, Computer **34** (2001), no. 8, 57–66.
- [HM82] T Horiguchi and T Morita, *Existence of the ferromagnetic phase in a random-bond Ising model on the square lattice*, Journal of Physics A: Mathematical and General **15** (1982), no. 2, L75.
- [HTPV07] Martin Hasenbusch, Francesco Parisen Toldin, Andrea Pelissetto, and Ettore Vicari, *Critical behavior of the three-dimensional  $\pm J$  Ising model at the paramagnetic-ferromagnetic transition line*, Physical Review B **76** (2007), no. 9, 094402.
- [Iba99] Yukito Iba, *The nishimori line and bayesian statistics*, Journal of Physics A: Mathematical and General **32** (1999), no. 21, 3875.
- [JM13] Adel Javanmard and Andrea Montanari, *Localization from incomplete noisy distance measurements*, Foundations of Computational Mathematics **13** (2013), no. 3, 297–345.

- [JMRT16] Adel Javanmard, Andrea Montanari, and Federico Ricci-Tersenghi, *Phase transitions in semidefinite relaxations*, Proceedings of the National Academy of Sciences **113** (2016), no. 16, E2218–E2223.
- [LMX15] Marc Lelarge, Laurent Massoulié, and Jiaming Xu, *Reconstruction in the labelled stochastic block model*, IEEE Transactions on Network Science and Engineering **2** (2015), no. 4, 152163.
- [Mer67] N David Mermin, *Absence of ordering in certain classical systems*, Journal of Mathematical Physics **8** (1967), no. 5, 1061–1064.
- [MMM13] Pierre Moulon, Pascal Monasse, and Renaud Marlet, *Global fusion of relative motions for robust, accurate and scalable structure from motion*, Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 3248–3255.
- [Moh97] Bojan Mohar, *Some applications of laplace eigenvalues of graphs*, Graph symmetry, Springer, 1997, pp. 225–275.
- [Mon08] Andrea Montanari, *Estimating random variables from random sparse observations*, European Transactions on Telecommunications **19** (2008), no. 4, 385–403.
- [Moo17] Cristopher Moore, *The computer science and physics of community detection: landscapes, phase transitions, and hardness*, arXiv:1702.00467 (2017).
- [MW66] N David Mermin and Herbert Wagner, *Absence of ferromagnetism or antiferromagnetism in one-or two-dimensional isotropic heisenberg models*, Physical Review Letters **17** (1966), no. 22, 1133.
- [Nis81] Hidetoshi Nishimori, *Internal energy, specific heat and correlation function of the bond-random Ising model*, Progress of Theoretical Physics **66** (1981), no. 4, 1169–1181.
- [Nis01] ———, *Statistical physics of spin glasses and information processing: an introduction*, vol. 111, Clarendon Press, 2001.
- [OMK10] Sewoong Oh, Andrea Montanari, and Amin Karbasi, *Sensor network localization from local connectivity: Performance analysis for the mds-map algorithm*, Information Theory (ITW 2010, Cairo), 2010 IEEE Information Theory Workshop on, IEEE, 2010, pp. 1–5.
- [ON93] Y Ozeki and Hidetoshi Nishimori, *Phase diagram of gauge glasses*, Journal of Physics A: Mathematical and General **26** (1993), no. 14, 3399.
- [OS15] Onur Ozyesil and Amit Singer, *Robust camera location estimation by convex programming*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2674–2683.
- [PTPV09] Francesco Parisen Toldin, Andrea Pelissetto, and Ettore Vicari, *Strong-disorder paramagnetic-ferromagnetic fixed point in the square-lattice  $\pm J$  Ising model*, Journal of Statistical Physics **135** (2009), no. 5, 1039–1061.

- [PWBM16] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra, *Message-passing algorithms for synchronization problems over compact groups*, arXiv:1610.04583 (2016).
- [Sin08] Amit Singer, *A remark on global positioning from local distances*, Proceedings of the National Academy of Sciences **105** (2008), no. 28, 9507–9511.
- [Sin11] ———, *Angular synchronization by eigenvectors and semidefinite programming*, Applied and computational harmonic analysis **30** (2011), no. 1, 20–36.
- [SLKZ15] Alaa Saade, Marc Lelarge, Florent Krzakala, and Zdeborová, *Spectral detection in the censored block model*, IEEE International Symposium on Information Theory (ISIT), 2015, p. 11841188.
- [SS11] Amit Singer and Yoel Shkolnisky, *Three-dimensional structure determination from common lines in cryo-em by eigenvectors and semidefinite programming*, SIAM journal on imaging sciences **4** (2011), no. 2, 543–572.
- [TDSL00] Joshua B Tenenbaum, Vin De Silva, and John C Langford, *A global geometric framework for nonlinear dimensionality reduction*, science **290** (2000), no. 5500, 2319–2323.
- [Was13] Larry Wasserman, *All of statistics: a concise course in statistical inference*, Springer Science & Business Media, 2013.
- [WBS16] Kyle Wilson, David Bindel, and Noah Snavely, *When is rotations averaging hard?*, European Conference on Computer Vision, Springer, 2016, pp. 255–270.
- [WS13] Lanhui Wang and Amit Singer, *Exact and stable recovery of rotations for robust synchronization*, Information and Inference (2013), iat005.