Exact recovery in the Ising Blockmodel

Quentin Berthet*, Philippe Rigollet† and Piyush Srivastava‡

University of Cambridge
Massachusetts Institute of Technology
California Institute of Technology

Abstract. We consider the problem associated to recovering the block structure of an Ising model given independent observations on the binary hypercube. This new model, called the Ising blockmodel, is a perturbation of the mean field approximation of the Ising model known as the Curie–Weiss model: the sites are partitioned into two blocks of equal size and the interaction between those of the same block is stronger than across blocks, to account for more order within each block. We study probabilistic, statistical and computational aspects of this model in the high-dimensional case when the number of sites may be much larger than the sample size.

AMS 2000 subject classifications: Primary 62H30; secondary 82B20.
Key words and phrases: Ising blockmodel, Curie-Weiss, stochastic blockmodel, planted partition, spectral partitioning.

1. INTRODUCTION

The past decades have witnessed an explosion of the amount of data collected. Along with this expansion comes the promise of a better understanding of an observed phenomenon by extracting relevant information from this data. Larger datasets not only call for faster methods to process them but also lead us to completely rethink the way data should be modeled. Specifically, these new datasets arise as the agglomeration of a multitude of basic entities and, rather than their average behavior, most of the information is contained in their interactions. *Graphical models* (a.k.a Markov Random Fields) have proved to be a very useful tool to turn raw data into networks that are amenable to clustering or community detection. Specifically, given random variables $\sigma_1, \dots, \sigma_p$, the goal is to output a graph on p nodes, one for each variable, where the edges encode conditional independence between said variables [Lau96]. Graphical models have been successfully employed in a variety of applications such as image analysis [Bes86],

*Supported by an Isaac Newton Trust Early Career Support Scheme and by a Faculty Fellow Award from The Alan Turing Institute.

†Supported by Grants NSF-DMS-1541099, NSF-DMS-1541100, DARPA-BAA-16-46 and a grant from the MIT NEC Corporation.

‡Supported in part by the United States NSF grant CCF-1319745. This work was performed while PS was supported by a post-doctoral fellowship of the Center for the Mathematics of Information at the California Institute of Technology.

natural language processing [MS99] and genetics [LS03, SRN+05] for example.

Originally introduced in the context of statistical physics to explain the observed behavior of various magnetic materials [Isi25], the Ising Model is a graphical model for binary random variables $\sigma_1, \ldots, \sigma_p \in \{-1, 1\}$, hereafter called *spins*. Despite its simplicity, this model has been effective at capturing a large class of physical systems. More recently, this model was proposed to model social interactions such as political affinities, where σ_j may represent the vote of U.S. senator j on a random bill [BEGd08] (see also the data used in [DGH08] for the U.S. House of Representatives). In this context, much effort has been devoted to estimating the underlying structure of the graphical model [BMS08, RWL10, Bre15] under sparsity assumptions. At the same time, the theoretical side of social network analysis has witnessed a lot of activity around the estimation and reconstruction of stochastic blockmodels [HLL83] as a simple but efficient way to capture the notion of *communities* in social networks. These random graph models assume an underlying partition of the nodes, leading to inhomogeneous connection probabilities between nodes. Given the realization of such a graph, the goal is to recover the partition of the nodes. Already in the context of a balanced partition into two communities, this model has revealed interesting threshold phenomena [MNS15, MNS13, Mas14].

In this work, we combine the notions of stochastic blockmodel and that of graphical model by assuming that we observe independent copies of a vector $\sigma = (\sigma_1, \ldots, \sigma_p) \in \{-1, 1\}^p$ distributed according to an Ising model with a block structure analogous to the one arising in the stochastic blockmodel.

Specifically, assume that $p \geq 2$ is an even integer and let $S \subset [p] := \{1, \ldots, p\}$ be a subset of size $|S| = m = p/2$. For any partition $(S, \bar{S})$, where $\bar{S} = [p] \setminus S$ denotes the complement of S, write $i \sim j$ if $(i, j) \in S^2 \cup \bar{S}^2$ and $i \nsim j$ if $(i, j) \in [p]^2 \setminus (S^2 \cup \bar{S}^2)$. Fix $\beta, \alpha \in \mathbb{R}$ and let $\sigma \in \{-1, 1\}^p$ have density $f_{S,\alpha,\beta}$ with respect to the counting measure on $\{-1, 1\}^p$ given by

$$f_{S,\alpha,\beta}(\sigma) = \frac{1}{Z_{\alpha,\beta}} \exp\Big[\frac{\beta}{2p} \sum_{i \sim j} \sigma_i \sigma_j + \frac{\alpha}{2p} \sum_{i \nsim j} \sigma_i \sigma_j\Big], \tag{1.1}$$

where

$$Z_{S,\alpha,\beta} := \sum_{\sigma \in \{-1,1\}^p} \exp\Big[\frac{\beta}{2p} \sum_{i \sim j} \sigma_i \sigma_j + \frac{\alpha}{2p} \sum_{i \nsim j} \sigma_i \sigma_j\Big] \tag{1.2}$$

is a normalizing constant traditionally called *partition function*. Let $\mathbb{P}_{S,\alpha,\beta}$ denote the probability distribution over $\{-1, 1\}^p$ that has density $f_{S,\alpha,\beta}$ with respect to the counting measure on $\{-1, 1\}^p$. We call this model the *Ising Blockmodel* (IBM). We write simply $f_{\alpha,\beta}$ and $\mathbb{P}_{\alpha,\beta}$ to emphasize the dependency on α, β and and simply $\mathbb{P}_S$ for emphasize the dependency on S.

When $\alpha = \beta > 0$, the model (1.1) is the mean field approximation of the (ferromagnetic) Ising model and is called the *Curie-Weiss* model (without external field). It can be readily seen from (1.1) that vectors $\sigma \in \{-1, 1\}^p$ that present a lot of pairs (i, j) with opposite spins (high energy configurations), i.e., $\sigma_i \sigma_j < 0$, receive less probability than vectors where most of the spins agree (low energy configurations). There are however much fewer vectors with low energy in the discrete hypercube and this tension between *energy* and *entropy* is responsible for phase transitions in such systems.

When positive, the parameter $\beta > 0$ is called *inverse temperature* and it controls the strength of interactions, and therefore, the weight given to the energy term. When $\beta \to 0$, the entropy term dominates and $\mathbb{P}_{\beta,\beta}$ tends to the uniform density over $\{-1,1\}^p$. When $\beta \to \infty$, $\mathbb{P}_{\beta,0} \to .5\delta_{\mathbf{1}} + .5\delta_{-\mathbf{1}}$, where δ_x denotes the Dirac point mass at x and $\mathbf{1} = (1,\dots,1) \in \{-1,1\}^p$ denotes the all-ones vector of dimension p, the energy term dominates and it affects the global behavior of the system as follows.

Let $\mu^{\mathrm{CW}} = \sigma^\top \mathbf{1}/p$ denote the *magnetization* of σ. When $\mu^{\mathrm{CW}} \simeq 0$, then σ has a balanced numbers of positive and negative spins (paramagnetic behavior) and when $|\mu^{\mathrm{CW}}| \gg 0$, then σ has a large proportion of spins with a given sign (ferromagnetic behavior). When p is large enough, the Curie-Weiss model is known to obey a phase-transition from ferromagnetic to paramagnetic behavior when the temperature crosses a threshold (see subsection A for details). This striking result indicates that when the temperature decreases (β increases), the model changes from that of a disordered system (no preferred inclination towards -1 or $+1$) to that of an ordered system (a majority of the spins agree to the same sign). This behavior is interesting in the context of modeling social interactions and indicates that if the strength of interactions is large enough ($\beta > 1$) then a partial consensus may be found. Formally, the Curie-Weiss model may also be defined in the anti-ferromagnetic case $\beta < 0$—we abusively call it "inverse temperature" in this case also—to model the fact that negative interactions are encouraged. For such choices of β, the distribution is concentrated around balanced configurations σ that have magnetization close to 0. Moreover, as $\beta \to -\infty$, $\mathbb{P}_{\beta,\beta}$ converges to the uniform distribution on configurations with zero magnetization (assuming that p is even so that such configurations exist for simplicity). As a result, the anti-ferromagnic case arises when no consensus may be found and and the spins are evenly split between positive and negative.

In reality though, a collective behavior may be fragmented into communities and the IBM is meant to reflect this structure. Specifically, since $\beta > \alpha$, the strength β of interactions within the blocks S and $\bar{S}$ is larger than that across blocks S and $\bar{S}$. As will become clear from our analysis, the case where $\alpha < 0$ presents interesting configurations whereby the two blocs S and $\bar{S}$ have polarized behaviors, that is opposite magnetization in each block.

The rest of this paper is organized as follows. In Section 2, we study the probability distributions $\mathbb{P}_{\alpha,\beta}$, for $\alpha < \beta$ and exhibit phase transitions. Next, in Section 3, we consider the problem of recovering the partition $S, \bar{S}$ from n iid samples from $\mathbb{P}_{\alpha,\beta}$.

Finally note that the size p of the system has to be large enough to observe interesting phenomena. In this paper we are also concerned with such high dimensional systems and our results will be valid for large enough p, potentially much larger than the number of observations. In particular, we often consider asymptotic statements as $p \to \infty$. However, in the statistical applications of Section 3 we are interested in understanding the scaling of the number of observations as a function of p. To that end, we keep track of the first order terms in p and only let higher order terms vanish when convenient.

2. PROBABILISTIC ANALYSIS OF THE ISING BLOCKMODEL

We will see in Section 3 that given $\sigma^{(1)}, \dots, \sigma^{(n)}$ that are independent copies of $\sigma \sim \mathbb{P}_{\alpha,\beta}$, the sample covariance matrix $\hat\Sigma$ defined by

$$\hat\Sigma = \frac{1}{n}\sum_{t=1}^{n} \sigma^{(t)}{\sigma^{(t)}}^\top , \tag{2.1}$$

is a sufficient statistic for S. From basic concentration results (see Section 3), it can be shown that this matrix concentrates around the true covariance matrix $\Sigma = \mathbb{E}_{\alpha,\beta}\big[\sigma\sigma^\top\big]$ where $\mathbb{E}_{\alpha,\beta}$ denotes the expectation associated to $\mathbb{P}_{\alpha,\beta}$. Unfortunately, computing Σ directly is quite challenging. Instead, we show that when p is large enough, then $\mathbb{P}_{\alpha,\beta}$ is spiked around specific values, which, in turn, give us a handle of quantities of the form $\mathbb{E}_{\alpha,\beta}[\varphi(\sigma)]$ for some test function φ. Beyond our statistical task, we show phase transitions that are interesting from a probabilistic point of view.

2.1 Free energy

Let $\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}$ denote the *IBM Hamiltonian* (or "energy") defined on $\{-1,1\}^p$ by

$$\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}(\sigma) = -\Big(\frac{\beta}{2p}\sum_{i\sim j}\sigma_i\sigma_j + \frac{\alpha}{2p}\sum_{i\nsim j}\sigma_i\sigma_j\Big) , \tag{2.2}$$

so that

$$f_{\alpha,\beta}(\sigma) = \frac{e^{-\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}(\sigma)}}{Z_{\alpha,\beta}}$$

Akin to the Curie-Weiss model, the density $f_{\alpha,\beta}$ puts uniform weights on configurations that have the same magnetization structure. To make this statement precise, for any $A \subset [p]$ define $\mathbf{1}_A \in \{0,1\}^p$ to be the indicator vector of A and let $\mu_A = \sigma^\top \mathbf{1}_A/|A|$ denote the *local magnetization* of σ on A. It follows from elementary computations that

$$\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}(\sigma) = -\frac{m}{4}\Big(2\alpha\mu_S\mu_{\bar S} + \beta(\mu_S^2 + \mu_{\bar S}^2)\Big) , \tag{2.3}$$

where we recall that $m = p/2$. Moreover, the number of configurations σ with local magnetizations $\mu = (\mu_S, \mu_{\bar S}) \in [-1,1]^2$ is given by

$$\binom{m}{\frac{\mu_S+1}{2}m}\binom{m}{\frac{\mu_{\bar S}+1}{2}m}$$

This quantity can be approximated using Stirling's formula (see Lemma 18): For any $\mu \in (1+\varepsilon, 1-\varepsilon)$, there exists two positive constants $\underline{c}, \bar c$ such that

$$\frac{\underline{c}}{\sqrt{m}}e^{-mh\left(\frac{\mu+1}{2}\right)} \le \binom{m}{\frac{\mu+1}{2}m} \le \frac{\bar c}{\sqrt{m}}e^{mh\left(\frac{\mu+1}{2}\right)} , \qquad \forall\, m \ge 1$$

where $h : [0,1] \to \mathbb{R}$ is the binary entropy function defined by $h(0) = h(1) = 1$ and for any $s \in (0,1)$ by

$$h(s) = -s\log(s) - (1-s)\log(1-s) .$$

Thus, IBM induces a marginal distribution on the local magnetizations that has density

$$\frac{\ell_m}{mZ_{\alpha,\beta}} \exp\Big[-\frac{m}{4} g(\mu_S, \mu_{\bar{S}})\Big]\,, \tag{2.4}$$

where $\underline{c}^2 \le \ell_m \le \bar{c}^2$ and

$$g(\mu_S, \mu_{\bar{S}}) = -2\alpha\mu_S\mu_{\bar{S}} - \beta(\mu_S^2 + \mu_{\bar{S}}^2) - 4h\big(\frac{\mu_S + 1}{2}\big) - 4h\big(\frac{\mu_{\bar{S}} + 1}{2}\big)\,. \tag{2.5}$$

Note that the support of this density is implicitly the set of possible values for pairs local magnetizations of vectors in $\{-1,1\}^p$, that is the set $\mathcal{M}^2$, where

$$\mathcal{M}^2 := \big\{\frac{s^\top \mathbf{1}_{[m]}}{m}, s \in \{-1,1\}^m\big\} \subset [-1,1]^2\,. \tag{2.6}$$

We call the function g the *free energy* of the Ising blockmodel and its structure of minima is known to control the behavior of the system. Indeed, g^* denote the minimum value of g over $\mathcal{M}^2$. It follows from (2.4) that any local magnetization $(\mu_S, \mu_{\bar{S}}) \in \mathcal{M}^2$ such that $g(\mu_S, \mu_{\bar{S}}) > g^*$ has a probability exponentially smaller than any magnetization that minimizes g over $\mathcal{M}^2$. Intuitively, this results in a distribution that is concentrated around its modes. Before quantifying this effect, we study the minima, known as *ground states* of the free energy g, when defined over the continuum $[-1,1]^2$.

2.2 Ground states

Recall that when $\alpha = \beta$, the block structure vanishes and the IBM reduces to the well-known Curie-Weiss model. We gather in Appendix A useful facts about the Curie-Weiss model that we use in the rest of this section.

The following proposition characterizes the ground states of the Ising blockmodel. For any $p \in [1,\infty]$, we denote by $\|\cdot\|_p$ the ℓ_p norm of $\mathbb{R}^2$ and by $\mathcal{B}_p = \{x \in \mathbb{R}^2, : \|x\|_p \le 1\}$ the unit ball with respect to that norm.

PROPOSITION 1. *For any $b \in \mathbb{R}$, let $\pm\tilde{x}(b) \in (-1,1), \tilde{x}(b) \ge 0$ denote the ground state(s) of the Curie-Weiss model with inverse temperature b. The free energy $g_{\alpha,\beta}$ of the IBM defined in (2.5) has the following minima:*
If $\beta + |\alpha| \le 2$, then $g_{\alpha,\beta}$ has a unique minimum at $(0,0)$.
If $\beta + |\alpha| > 2$, then three cases arise:

1. *If $\alpha = 0$, then $g_{\alpha,\beta}$ has four minima at $(\pm\tilde{x}(\beta/2), \pm\tilde{x}(\beta/2))$,*
2. *If $\alpha > 0$, then $g_{\alpha,\beta}$ has two minima at $\tilde{s} = (\tilde{x}(\frac{\beta+\alpha}{2}), \tilde{x}(\frac{\beta+\alpha}{2}))$ and $-\tilde{s}$,*
3. *If $\alpha < 0$, then $g_{\alpha,\beta}$ has two minima at $\tilde{s} = (\tilde{x}(\frac{\beta-\alpha}{2}), -\tilde{x}(\frac{\beta-\alpha}{2}))$ and $-\tilde{s}$.*

In particular, for all values of the parameters α and β, all ground states $(\tilde{x}, \tilde{y})$ satisfy $\tilde{x}^2 = \tilde{y}^2 < 1$.

This result is illustrated in Figure 2, composed of contour plots of the free energy $g_{\alpha,\beta}$ on the square $[-1,1]^2$, for several values of the parameters.

PROOF. Throughout this proof, for any $b \in \mathbb{R}$, we denote by $g_b^{\mathrm{CW}}(x), x \in [-1,1]$, the free energy of the Curie-Weiss model with inverse temperature b. We write $g := g_{\alpha,\beta}$ for simplicity to denote the free energy of the IBM.

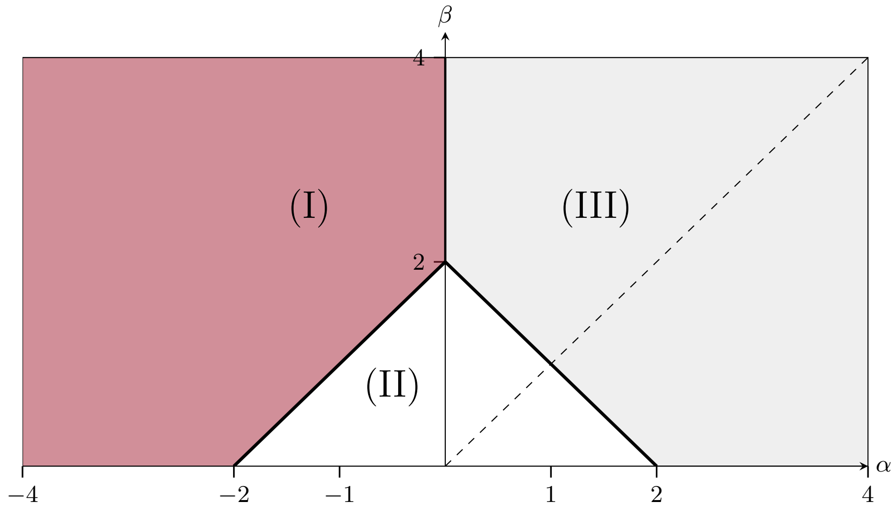

Figure 1: The phase diagram of the Ising block model, with three regions for the parameters α and $\beta > 0$. In region (I), where $\alpha < 0$ and $\beta + |\alpha| > 2$, there are two ground states of the form $(x, -x)$ and $(-x, x), x > 0$. In region (II), where $\beta + |\alpha| < 2$, there is one ground state at $(0, 0)$. In region (III), where $\alpha > 0$ and $\beta + |\alpha| > 2$, there are two ground states of the form (x, x) and $(-x, -x), x > 0$. At the boundary between regions (I) and (III), there are four ground states. The dotted line has equation $\alpha = \beta$, we only consider parameters in the region to its left, where $\beta > \alpha$.

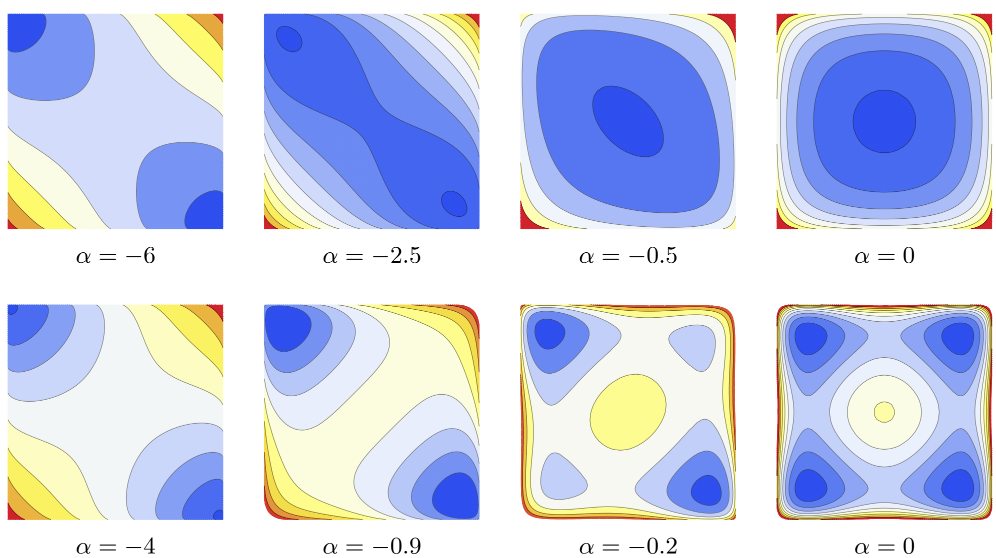

Figure 2: Contour plots of the values of the free energy $g_{\alpha,\beta}$, with higher values in red and lower values in blue, corresponding to ground states. *Top row* : Several choices for $\alpha < 0$, and $\beta = 1.5 < 2$. *Bottom row* : Several choices for $\alpha < 0$, and $\beta = 2.5 > 2$. The same plots with $\alpha > 0$ can be obtained by a $90°$ rotation, by symmetry of the function.

Note that

$$g(x,y) = g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(x) + g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(y) + \alpha(x-y)^2\,. \tag{2.7}$$

We split our analysis according to the sign of α. Note first that if $\alpha = 0$, we have

$$g(x,y) = g^{\mathrm{CW}}_{\frac{\beta}{2}}(x) + g^{\mathrm{CW}}_{\frac{\beta}{2}}(y)\,.$$

It yields that:

- If $\beta \le 2$, then $g^{\mathrm{CW}}_{\frac{\beta}{2}}$ has a unique local minimum at $x = 0$ which implies that g has a unique minimum at $(0,0)$
- If $\beta > 2$, then $g^{\mathrm{CW}}_{\frac{\beta}{2}}$ has exactly two minima at $\tilde{x}(\beta/2)$ and $-\tilde{x}(\beta/2)$, where $\tilde{x}(\beta/2) \in (-1,1)$. It implies that g has four minima at $(\pm\tilde{x}(\beta/2), \pm\tilde{x}(\beta/2))$.

Next, if $\alpha > 0$, in view of (2.7) we have

$$g(x,y) \ge g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(x) + g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(y)$$

with equality iff $x = y$. It implies that:

- If $\alpha + \beta \le 2$, then g has a unique minimum at $(0,0)$
- If $\alpha + \beta > 2$, then g has two minima on $\mathcal{A}$ at $(\tilde{x}(\frac{\beta+\alpha}{2}), \tilde{x}(\frac{\beta+\alpha}{2}))$ and at $(-\tilde{x}(\frac{\beta+\alpha}{2}), -\tilde{x}(\frac{\beta+\alpha}{2}))$.

Finally, note that $(x-y)^2 \le 2x^2 + 2y^2$ with equality iff $x = -y$. Thus, if $\alpha < 0$, in view of (2.7) we have

$$g(x,y) \ge g^{\mathrm{CW}}_{\frac{\beta-\alpha}{2}}(x) + g^{\mathrm{CW}}_{\frac{\beta-\alpha}{2}}(y) \tag{2.8}$$

with equality iff $x = -y$. It implies that

- If $\beta - \alpha \le 2$, then g has a unique minimum at $(0,0)$
- If $\beta - \alpha > 2$, then g has two minima at $(\tilde{x}(\frac{\beta-\alpha}{2}), -\tilde{x}(\frac{\beta-\alpha}{2}))$ and at $(-\tilde{x}(\frac{\beta-\alpha}{2}), \tilde{x}(\frac{\beta-\alpha}{2}))$.

□

Using the localization of the ground states from Lemma 15, we also get the following local and global behaviors of the free energy of the IBM around the ground states.

Lemma 2. *Assume that $\beta + |\alpha| \neq 2$. Denote by $(\tilde{x}, \tilde{y})$ any ground state of Ising blockmodel and recall that $\tilde{x}^2 = \tilde{y}^2$. Then the following holds:*

1. *The Hessian $H_{\alpha,\beta}$ of $g_{\alpha,\beta}$ at $(\tilde{x}, \tilde{y})$ is given by*

$$H_{\alpha,\beta} = -2\begin{pmatrix} \beta & \alpha \\ \alpha & \beta \end{pmatrix} + \frac{4}{1-\tilde{x}^2} I_2\,.$$

In particular $H_{\alpha,\beta}$ has eigenvalues $2(\alpha-\beta) + 4/(1-\tilde{x}^2)$ and $-2(\alpha+\beta) + 4/(1-\tilde{x}^2)$ associated with eigenvectors $(1,-1)$ and $(1,1)$ respectively.

2. *There exists positive constants $\delta = \delta(\beta + |\alpha|)$, $\kappa^2 = \kappa^2(\beta + |\alpha|)$ such that the following holds. For any $(x, y) \in (-1, 1)^2$, we have*

$$g(x,y) \geq g(\tilde{x},\tilde{y}) + \frac{\kappa^2}{2}\Big(\|(x,y) - (\tilde{x},\tilde{y})\|_\infty \wedge \delta\Big)^2 . \tag{2.9}$$

Moreover,

If $\beta + |\alpha| > 2$, we can take $\delta = e^{-(\beta+|\alpha|)}\dfrac{\beta + |\alpha| - 2}{4(\beta + |\alpha|)}$ and $\kappa^2 = 1 - \dfrac{2}{\beta + |\alpha|}$.

If $\beta + |\alpha| < 2$, we can take $\delta = \sqrt{(2 - (\beta + |\alpha|))/6}$ and $\kappa^2 = 2 - (\beta + |\alpha|)$.

PROOF. Elementary calculus yields directly that

$$H_{\alpha,\beta} = \begin{pmatrix} -2\beta + \frac{4}{1-\tilde{x}^2} & -2\alpha \\ -2\alpha & -2\beta + \frac{4}{1-\tilde{y}^2} \end{pmatrix} .$$

Moreover, it follows from Proposition 1 that all ground states satisfy $\tilde{x}^2 = \tilde{y}^2$. This completes the proof of the first point.

We now turn to the proof of the second point and split the analysis into four cases: (i) $\alpha \geq 0$ and $\beta + \alpha < 2$, (ii) $\alpha \geq 0$ and $\beta + \alpha > 2$, (iii) $\alpha < 0$ and $\beta - \alpha < 2$, (iv) $\alpha < 0$ and $\beta - \alpha > 2$.

Case (i): $\alpha > 0$ and $\beta + \alpha < 2$. Recall that in this case, g has a unique minimum at $(0, 0)$. Therefore, in view of (2.7) and Lemma 15, we have

$$\begin{aligned} g(x,y) - g(0,0) &= g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(0) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(0) + \alpha(x-y)^2 \\ &\geq \frac{1}{2}\big(2 - (\beta + |\alpha|)\big)\big[(|x - 0| \wedge \varepsilon')^2 + (|y - 0| \wedge \varepsilon')^2\big] \\ &\geq \frac{1}{2}\big(2 - (\beta + |\alpha|)\big)\big(\|(x,y) - (0,0)\|_\infty \wedge \varepsilon'\big)^2 . \end{aligned}$$

where $\varepsilon' = \sqrt{(2 - (\beta + |\alpha|))/6}$ which concludes this case.

Case (ii): $\alpha > 0$ and $\beta + \alpha > 2$. Recall that in this case, g has two minima denoted generically by $(\tilde{x}, \tilde{y})$ where $\tilde{x} = \tilde{y}$. Therefore, in view of (2.7) and Lemma 15, we have

$$\begin{aligned} g(x,y) - g(\tilde{x},\tilde{y}) &= g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(\tilde{x}) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(\tilde{y}) + \alpha(x-y)^2 \\ &\geq \frac{1}{2}\big(1 - \frac{2}{\beta + |\alpha|}\big)\big[(|x - 0| \wedge \varepsilon)^2 + (|y - 0| \wedge \varepsilon)^2\big] \\ &\geq \frac{1}{2}\big(1 - \frac{2}{\beta + |\alpha|}\big)\big(\|(x,y) - (0,0)\|_\infty \wedge \varepsilon\big)^2 . \end{aligned}$$

where $\varepsilon = e^{-(\beta+|\alpha|)}\dfrac{\beta + |\alpha| - 2}{4(\beta + |\alpha|)}$ which concludes this case.

Case (iii): $\alpha < 0$ and $\beta - \alpha < 2$. Recall that in this case, g has a unique minimum at $(0, 0)$. Moreover, in view of (2.8) and Lemma 15, it holds

$$\begin{aligned} g(x,y) - g(0,0) &\geq g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(0) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(0) \\ &\geq g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(0) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(0) \\ &\geq \frac{1}{2}\big(2 - (\beta + |\alpha|)\big)\big(\|(x,y) - (0,0)\|_\infty \wedge \varepsilon'\big)^2 . \end{aligned}$$

where in the second inequality, we used the fact that

$$g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(0) = g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(0) = -4h(1/2)\,,$$

and we concluded as in Case (i).

Case (iv): $\alpha < 0$ and $\beta - \alpha > 2$. Recall that in this case, g has two minima denoted generically by $(\tilde{x}, \tilde{y})$ where $\tilde{x} = -\tilde{y}$. Therefore, in view of (2.7) and (2.8), we have

$$g(x,y) - g(\tilde{x},\tilde{y}) \geq g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(\tilde{x}) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(-\tilde{x}) - 4\alpha\tilde{x}^2\,.$$

Next, observe that from the definition (A.1) of the free energy in the Curie-Weiss model, we have

$$-g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(\tilde{x}) - g^{\mathrm{CW}}_{\frac{\beta+\alpha}{2}}(-\tilde{x}) - 4\alpha\tilde{x}^2 = -g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(\tilde{x}) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(-\tilde{x})\,.$$

The above two displays yield

$$\begin{aligned} g(x,y) - g(\tilde{x},\tilde{y}) &\geq g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(x) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(\tilde{x}) + g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(y) - g^{\mathrm{CW}}_{\frac{\beta+|\alpha|}{2}}(-\tilde{x}) \\ &\geq \frac{1}{2}\big(1 - \frac{2}{\beta+|\alpha|}\big)\big(\|(x,y)-(0,0)\|_\infty \wedge \varepsilon\big)^2\,. \end{aligned}$$

where we concluded as in Case (ii).

□

2.3 Concentration

As mentioned above, quantities of the form $\mathbb{E}_{\alpha,\beta}[\varphi(\sigma)]$ cannot in general be computed explicitly in the IBM. Fortunately, it will be sufficient for us to compute quantities of the form $\mathbb{E}_{\alpha,\beta}[\varphi(\mu)]$, where we recall that $\mu = (\mu_S, \mu_{\bar{S}})$ denotes the pair of local magnetizations of a random configuration $\sigma \in \{-1,1\}^p$ drawn according to $\mathbb{P}_{\alpha,\beta}$. While exact computation is still a hard problem, these quantities can be be well approximated using the fact that $\mathbb{P}_{\alpha,\beta}$ is highly concentrated around its ground states for large enough p.

To leverage concentration, we need to consider the "large m" (or equivalently "large p") asymptotic framework. As a result, it will be convenient to write for two sequences a_m, b_m that $a_m \simeq_m b_m$ if $a = (1 + o_m(1))b_m$.

Our main result hinges on the following proposition that compares the distribution of $\mu = (\mu_S, \mu_{\bar{S}}) \in [-1,1]$ to a certain mixture of Gaussians that are centered at the ground states.

THEOREM 3. *Let $\varphi : [-1,1]^2 \to [0,1]$ be any nonnegative bounded continuous test function. Denote by $\tilde{s}$ any ground state and assume that there exists positive constants C, γ, for which $\mathbb{E}\big[\varphi\big(\tilde{s} + \frac{2}{\sqrt{m}}H^{-1/2}Z\big)\big] \geq Cm^{-\gamma}$ where $Z \sim \mathcal{N}_2(0, I_2)$ and $H = H_{\alpha,\beta}$ denotes the Hessian of the free energy $g_{\alpha,\beta}$ at $\tilde{s}$. Then*

$$\mathbb{E}_{\alpha,\beta}[\varphi(\mu)] \simeq_m \frac{1}{|G|}\sum_{\tilde{s}\in G} \mathbb{E}\big[\varphi(\tilde{s} + \frac{2}{\sqrt{m}}H^{-1/2}Z)\big]\,.$$

where $G \subset \{(\pm\tilde{x}, \pm\tilde{x})\}$ denotes the set of ground states of the IBM.

PROOF. Recall that $\mathcal{M}^2$ defined in (2.6) denotes the set of possible values for pairs of local magnetization and observe that

$$\mathbb{E}_{\alpha,\beta}[\varphi(\mu)] = \frac{1}{Z_{\alpha,\beta}} \sum_{\mu \in \mathcal{M}^2} \varphi(\mu) z_m(\mu)\,,$$

where

$$z_m(\mu) := \exp\Big(-\frac{m}{4}\left(-2\alpha\mu_S\mu_{\bar{S}} - \beta(\mu_S^2 + \mu_{\bar{S}}^2)\right)\Big)\binom{m}{\frac{\mu_S+1}{2}m}\binom{m}{\frac{\mu_{\bar{S}}+1}{2}m} \tag{2.10}$$

We split the local magnetization μ according to their ℓ_2 distance to the closest ground state. Let $G \subset [0,1]^2$ denote the set of ground states and fix $\delta := (\rho/\kappa)\sqrt{(\log m)/m}$, where $\rho > 0$ is a constant to be chosen later and κ is defined in Lemma 2. For any ground state $\tilde{s} \in G$, define $\mathcal{V}_{\tilde{s}}$ to be the neighborhood of $\tilde{s}$ defined by

$$\mathcal{V}_{\tilde{s}} = \left\{\mu \in \mathcal{M}^2 \,:\, \|\mu - \tilde{s}\|_\infty \le \delta\right\},$$

where $\delta > 0$ is also defined in Lemma 2. Moreover, define

$$\mathcal{V} = \bigcup_{\tilde{s} \in G} \mathcal{V}_{\tilde{s}}\,,$$

and assume that m is large enough so that (i) the above union is a disjoint one and (ii), there exists a constant $C > 0$ depending on α and β such that for any $(x,y) \in \mathcal{V}$, we have $||x|-1| \wedge ||y|-1| \ge C > 0$. Denote by $g^*_{\alpha,\beta}$ the value of the free energy at any of the ground states.

We first treat the magnetizations outside $\mathcal{V}$. Using Lemma 2 together with Lemma 17, we get

$$\begin{aligned}
0 \le \exp\big(\frac{m}{4} g^*_{\alpha,\beta}\big) \sum_{\mu \notin \mathcal{V}} \varphi(\mu) z_m(\mu) &\le \exp\big(\frac{m}{4} g^*_{\alpha,\beta}\big) \sum_{\mu \notin \mathcal{V}} \exp\Big(-\frac{m}{4} g_{\alpha,\beta}(\mu)\Big) \\
&\le m^2 \exp\big(-\frac{m}{4}\frac{\kappa^2\delta^2}{2}\big) \le m^{2-\frac{\rho^2}{2}} = o_m(m^{-\gamma})\,,
\end{aligned} \tag{2.11}$$

for $\rho > 4\sqrt{8\gamma}$.

Next assume that $\mu \in \mathcal{V}$. Our starting point is the following approximation, that follows from Lemma 18: for any $\mu \in \mathcal{V}$,

$$z_m(\mu) = \frac{1}{\pi m} \frac{\exp\left(-\frac{m}{4} g_{\alpha,\beta}(\mu_S, \mu_{\bar{S}})\right)}{\sqrt{(1-\mu_S^2)(1-\mu_{\bar{S}}^2)}}(1 + o_m(1))\,, \tag{2.12}$$

Define $\mathcal{V}' = \mathcal{V}_{\tilde{s}} - \{\tilde{s}\}$. A Taylor expansion around $\tilde{s}$ gives for nay $u \in \mathcal{V}'$,

$$g_{\alpha,\beta}(\tilde{s} + u) = g_{\alpha,\beta}(\tilde{s}) + \frac{1}{2} u^\top H u + O(\delta^3).$$

where $H = H_{\alpha,\beta}$ denotes the Hessian of $g_{\alpha,\beta}$ at the ground state $\tilde{s}$. The above two displays yield

$$
\begin{aligned}
&\exp\big(\frac{m}{4}g^*_{\alpha,\beta}\big)\sum_{\mu\in\mathcal{V}_{\tilde s}}\varphi(\mu)z_m(\mu)\\
&\quad=\exp\big(\frac{m}{4}g^*_{\alpha,\beta}\big)\sum_{u\in\mathcal{V}'}\varphi(\tilde s+u)z_m(\tilde s+u)\\
&\quad\simeq_m \frac{1}{\pi m(1-\tilde x^2)}\sum_{u\in\mathcal{V}'}\varphi(\tilde s+u)\exp\big(-\frac{m}{8}u^\top Hu\big)\\
&\quad\simeq_m \frac{m}{\pi(1-\tilde x^2)}\int_{\delta\mathcal{B}_\infty}\varphi(\tilde s+x)\exp\big(-\frac{m}{8}x^\top Hx\big)dx\\
&\quad=\frac{1}{\pi(1-\tilde x^2)}\frac{1}{\sqrt{\det H}}\int_{\|H^{-\frac12}z\|_\infty\le\frac{\delta\sqrt{m}}{2}}\varphi\big(\tilde s+\frac{2}{\sqrt m}H^{-1/2}z\big)\exp\big(-\frac{\|z\|_2}{2}\big)dz\\
&\quad\simeq_m \frac{1}{1-\tilde x^2}\frac{2}{\sqrt{\det H}}\Big(\mathbb{E}\big[\varphi\big(\tilde s+\frac{2}{\sqrt m}H^{-1/2}Z\big)\big]-T_m\Big)\,.
\end{aligned}
$$

where $Z\sim\mathcal{N}_2(0,I_2)$ and

$$
T_m=\int_{z\,:\,z^\top H^{-1}z\ge\frac{m\delta^2}{2}}\varphi\big(\tilde s+\frac{2}{\sqrt m}H^{-1/2}z\big)\exp\big(-\frac{\|z\|_2}{2}\big)dz
$$

Here, the third equality replaces the sum by a Riemann integral and in the last one we use the following facts: (i) the set vectors z such that $\|H^{-\frac12}z\|_\infty\le 1$ contains a Euclidean ball of positive radius $r(\alpha,\beta)$ and (ii) $\delta\sqrt{m}\to\infty$. Next, observe that since φ takes values in $[0,1]$, we have

$$
\begin{aligned}
0\le T_m&\le 2\pi\mathbb{P}(Z^\top HZ\ge m/2)\\
&\le 2\pi\mathbb{P}\big(\|Z\|^2-2\ge\frac{m}{2\lambda_{\max}(H)}-2\big)\\
&\le 2\pi\sqrt{e}\exp\big(-\frac{m}{8\lambda_{\max}(H)}\big)=o(m^{-\gamma})
\end{aligned}
\tag{2.13}
$$

for $m\ge 8\lambda_{\max}(H)$ and where we used Lemma 19.

Since the same calculation holds for all ground states in G, and because the sets $\mathcal{V}_{\tilde s},\ \tilde s\in G$ are disjoint, we get that

$$
\exp\big(\frac{m}{4}g^*_{\alpha,\beta}\big)\sum_{\mu\in\mathcal{V}}\varphi(\mu)z_m(\mu)\simeq_m\frac{1}{1-\tilde x^2}\frac{2}{\sqrt{\det H}}\sum_{\tilde s\in G}\mathbb{E}\big[\varphi\big(\tilde s+\frac{2}{\sqrt m}H^{-1/2}Z\big)\big]\,.
$$

Together with (2.11), the above display yields

$$
\sum_{\mu\in\mathcal{M}^2}\varphi(\mu)z_m(\mu)\simeq_m\frac{2e^{-\frac{m}{4}g^*_{\alpha,\beta}}}{(1-\tilde x^2)\sqrt{\det H}}\sum_{\tilde s\in G}\mathbb{E}\big[\varphi(\tilde s+\frac{2}{\sqrt m}H^{-1/2}Z)\big]\,,
$$

In particular, this expression yields for $\varphi\equiv 1$,

$$
Z_{\alpha,\beta}\simeq_m\frac{2|G|e^{-\frac{m}{4}g^*_{\alpha,\beta}}}{(1-\tilde x^2)\sqrt{\det H}}\,.
$$

The above two displays yield the desired result. □

2.4 Covariance

The covariance matrix $\Sigma = \mathbb{E}_{\alpha,\beta}[\sigma\sigma^\top]$ captures the block structure of IBM and thus plays a major role in the statistical applications of Section 3. Moreover, the coefficients of Σ can be expressed explicitely in terms of the local magnetization μ_S and $\mu_{\bar S}$.

LEMMA 4. *Let $\Sigma = \mathbb{E}_{\alpha,\beta}[\sigma\sigma^\top]$ denote the covariance matrix of a random configuration $\sigma \sim \mathbb{P}_{\alpha,\beta}$. For any $i \neq j \in [p]$, it holds*

$$\begin{aligned} \Delta &:= \Sigma_{ij} = \frac{m}{2(m-1)}\mathbb{E}[\mu_S^2 + \mu_{\bar S}^2] - \frac{1}{m-1} && \text{if } i \sim j \\ \Omega &:= \Sigma_{ij} = \mathbb{E}[\mu_S \mu_{\bar S}] && \text{if } i \nsim j\,. \end{aligned}$$

PROOF. In this proof, we rely on symmetry of the problem: all the spins σ_i in a given block, S or $\bar S$ have the same marginal distribution. Fix $i \neq j$.

If $i \sim j$, for example if $i, j \in S$, we have by linearity of expectation.

$$\Sigma_{ij} = \mathbb{E}[\sigma_i\sigma_j] = \frac{1}{m(m-1)}\big(\mathbb{E}\sum_{(i,j)\in S^2}\sigma_i\sigma_j - m\big) = \frac{m}{m-1}\mathbb{E}[\mu_S^2] - \frac{1}{m-1}\,.$$

Since μ_S and $\mu_{\bar S}$ are identically distributed, we obtain the desired result.

For any $i \nsim j$ we have

$$\Sigma_{ij} = \mathbb{E}[\sigma_i\sigma_j] = \frac{1}{m^2}\mathbb{E}\sum_{(i,j) S\times \bar S}\sigma_i\sigma_j = \mathbb{E}[\mu_S\mu_{\bar S}]\,,.$$

□

Unlike many models in the statistical literature, computing Σ exactly is difficult in the IBM. In particular, it is not immediately clear from Lemma 4 that $\Delta > \Omega$, while this should be intuitively true since $\beta > \alpha$ and therefore the spin interactions are stronger within blocks than across blocks. It turns out that this simple fact can be checked by other means (see Lemma 12) for any $m \geq 2$. In the rest of this subsection, we use asymptotic approximations as $m \to \infty$ to prove effective upper and lower bound on the gap $\Delta - \Omega$.

PROPOSITION 5. *Let Δ and Ω be defined as in Lemma 4 and recall that G denotes the set of ground states of the IBM. Then*

$$\Delta - \Omega \simeq_m \frac{1}{2|G|}\sum_{(\tilde x,\tilde y)\in G}(\tilde x - \tilde y)^2 + \frac{1}{m}\Big(\frac{(\beta-\alpha)(1-\tilde x^2)^2}{2-(\beta-\alpha)(1-\tilde x^2)}\Big)\,.$$

In particular,

- *If $\beta + |\alpha| < 2$, then $\Delta - \Omega \simeq_m \frac{1}{m}\big(\frac{\beta-\alpha}{2-(\beta-\alpha)}\big)$.*
- *If $\beta + |\alpha| > 2$, then three cases arise:*
    1. *if $\alpha = 0$, then $\Delta - \Omega \simeq_m \tilde x^2$,*
    2. *if $\alpha > 0$, then $\Delta - \Omega \simeq_m \frac{1}{m}\big(\frac{(\beta-\alpha)(1-\tilde x^2)^2}{2-(\beta-\alpha)(1-\tilde x^2)}\big) > 0$*

3. if $\alpha < 0$*, then* $\Delta - \Omega \simeq_m 2\tilde{x}^2$ *.*

PROOF. It follows from Lemma 4 that

$$\Delta = \frac{m}{m-1}\mathbb{E}_{\alpha,\beta}[\varphi(\mu)] - \frac{1}{m-1}$$

where $\varphi(\mu) = \|\mu\|_2^2/2$. Therefore, using Theorem 3, we get that for $Z \sim \mathcal{N}_2(0, I_2)$,

$$\begin{aligned}\Delta &\simeq_m \Big(1+\frac{1}{m}\Big)\frac{1}{2|G|}\sum_{\tilde{s}\in G}\|\tilde{s}\|_2^2 + \frac{2}{m}\mathbb{E}\|H^{-1/2}Z\|_2^2 - \frac{1}{m}\\ &= \Big(1+\frac{1}{m}\Big)\frac{1}{2|G|}\sum_{(\tilde{x},\tilde{y})\in G}(\tilde{x}^2+\tilde{y}^2) + \frac{2}{m}\,\mathbf{Tr}(H^{-1}) - \frac{1}{m}\,.\end{aligned}$$

Using the same argument, we get that

$$\Omega \simeq_m \frac{1}{|G|}\sum_{(\tilde{x},\tilde{y})\in G}\tilde{x}\tilde{y} + \frac{4}{m}e_1^\top H^{-1}e_2\,,$$

where $e_1 = (1,0)^\top$ and $e_2 = (0,1)^\top$ are the vectors of the canonical basis of $\mathbb{R}^2$. Therefore

$$\Delta - \Omega \simeq_m \frac{1}{2|G|}\sum_{\tilde{s}\in G}(\tilde{x}-\tilde{y})^2 + \frac{2}{m}v^\top H^{-1}v - \frac{1}{m}(1-\tilde{x}^2)$$

where $v = (1,-1)$. Lemma 2 implies that v is an eigenvector of H and thus of H^{-1} and

$$v^\top H^{-1}v = \frac{1}{\alpha-\beta+2/(1-\tilde{x}^2)}\,.$$

This completes the first part of the proof and it remains only to check the different cases.

- If $\beta + |\alpha| < 2$, then $\tilde{x} = \tilde{y} = 0$ is the unique ground state, which yields the result by substitution.
- If $\beta + |\alpha| > 2$, and
  1. if $\alpha = 0$, then $|G| = 4$ and there are two ground states $(\tilde{x}, -\tilde{x})$ and $(-\tilde{x}, \tilde{x})$ for which $(\tilde{x} - \tilde{y})$ does not vanish. The term in $1/m$ is negligible;
  2. if $\alpha > 0$, then for both ground states $(\tilde{x} - \tilde{y})^2 = 0$ so that

     $$\Delta - \Omega \simeq_m \frac{1}{m}\Big(\frac{(\beta-\alpha)(1-\tilde{x}^2)^2}{2-(\beta-\alpha)(1-\tilde{x}^2)}\Big)$$

     The fact that this quantity is positive, follows from (A.5) with $\gamma = 0$.
  3. if $\alpha < 0$, then there are two ground states $(\tilde{x}, -\tilde{x})$ and $(-\tilde{x}, \tilde{x})$ and we can conclude as in the case $\alpha = 0$ but gain a factor of 2 because all the ground states contribute to the constant term.

□

It follows from proposition 5 that if $\beta + |\alpha| \neq 2$ then the covariance matrix Σ takes two values that are separated by a term of order at least $1/m$ and even sometimes of order 1. In the next section, we leverage this information to derive statistical results.

3. CLUSTERING IN THE ISING BLOCKMODEL

In this section, we focus on the following clustering task: given n i.i.d observations drawn from $\mathbb{P}_{\alpha,\beta}$, recover the partition $(S, \bar{S})$. To that end, we build upon the probabilisitic analysis of the IBM that was carried out in the previous section in order to study the properties of an efficient clustering algorithm together with the fundamental limitations associated to this task.

3.1 Maximum likelihood estimation

Fix a sample size $n \geq 1$. Given n independent copies $\sigma^{(1)}, \dots, \sigma^{(n)}$ of $\sigma \sim \mathbb{P}_{\alpha,\beta}$, the log-likelihood is given by

$$\mathcal{L}_n(S) = \sum_{t=1}^{n} \log\big(\mathbb{P}_{\alpha,\beta}(\sigma^{(t)})\big) = -n \log Z_{\alpha,\beta} - \sum_{t=1}^{m} \mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}(\sigma^{(t)}) \,.$$

where $Z_{\alpha,\beta}$ is the partition function defined in (1.2) and $\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}$ is the IBM Hamiltonian defined in (2.2). While both $Z_{\alpha,\beta}$ and $\mathcal{H}^{\mathrm{IBM}}_{\alpha,\beta}$ could depend on the choice of the block S, it turns out that $Z_{\alpha,\beta}$ is constant over choices of S such that $|S| = m = p/2$.

LEMMA 6. *The partition function $Z_{\alpha,\beta} = Z_{\alpha,\beta}(S)$ defined in (1.2) is such that $Z_{\alpha,\beta}(S) = Z_{\alpha,\beta}([m])$ for all S of size $|S| = m$. This statement remains true even if $m \neq p/2$.*

PROOF. Fix $S \subset [p]$ such that $|S| = m$ and denote by $\pi : [p] \to [p]$ any bijection that maps $[m]$ to S. By (1.2) and (2.3), it holds

$$\begin{aligned}
Z_{\alpha,\beta}(S) &= \sum_{\sigma \in \{-1,1\}^p} \exp\Big[\frac{1}{4m}\Big(2\alpha(\sigma^\top \mathbf{1}_S)(\sigma^\top \mathbf{1}_{\bar{S}}) - \beta\big((\sigma^\top \mathbf{1}_S)^2 + (\sigma^\top \mathbf{1}_{\bar{S}})^2\big)\Big)\Big] \\
&= \sum_{\substack{\tau=\pi(\sigma) \\ \sigma \in \{-1,1\}^p}} \exp\Big[\frac{1}{4m}\Big(2\alpha(\tau^\top \mathbf{1}_S)(\tau^\top \mathbf{1}_{\bar{S}}) - \beta\big((\tau^\top \mathbf{1}_S)^2 + (\tau^\top \mathbf{1}_{\bar{S}})^2\big)\Big)\Big]
\end{aligned}$$

since π is a bijection. Moreover, $\tau^\top \mathbf{1}_S = \pi(\sigma)^\top \mathbf{1}_S = \sigma^\top \mathbf{1}_{[m]}$ and $\tau^\top \mathbf{1}_{\bar{S}} = \sigma^\top \mathbf{1}_{\overline{[m]}}$. Hence $Z_{\alpha,\beta}(S) = Z_{\alpha,\beta}([m])$. $\square$

Because of the above lemma, we simply write $Z_{\alpha,\beta} = Z_{\alpha,\beta}(S)$ to emphasize the fact that the partition function does not depend on S. It turns out that the log-likelihood is a simple function of S. Indeed, define the matrix $Q = Q_S \in \mathbb{R}^{p \times p}$ such that $Q_{ij} = \frac{\beta}{p}$ for $i \sim j$ and $Q_{ij} = \frac{\alpha}{p}$ for $i \nsim j$. Observe that (2.3) can be written as

$$\mathcal{H}_{\alpha,\beta}(\sigma) = \frac{1}{2}\sigma^\top Q \sigma = \frac{1}{2}\mathbf{Tr}(\sigma\sigma^\top Q) \,.$$

This in turns implies

$$\mathcal{L}_n(S) = -n \log Z_{\alpha,\beta} + \frac{n}{2}\mathbf{Tr}[\hat{\Sigma} Q] \,,$$

where $\hat{\Sigma}$ denotes the empirical covariance matrix defined in (2.1). Since $\alpha < \beta$, it is not hard to see that the likelihood maximization problem $\max_{S \subset [p], |S|=m} \mathcal{L}_n(S)$ is equivalent to

$$\max_{V \in \mathcal{P}} \mathbf{Tr}[\hat{\Sigma} V] \,, \qquad \mathcal{P} = \{vv^\top : v \in \{-1,1\}^p, v^\top \mathbf{1}_{[p]} = 0\} \,. \tag{3.1}$$

In particular, estimating the blocks $(S, \bar{S})$ amounts to estimating $v_S v_S^\top \in \mathcal{P}$, where $v_S = \mathbf{1}_S - \mathbf{1}_{\bar{S}} \in \{-1,1\}^p$. Note that $v_S v_S^\top = v_{\bar{S}} v_{\bar{S}}^\top$.For an adjacency matrix A, the optimization problem $\max_{V \in \mathcal{P}} \mathbf{Tr}[AV]$ is a special case of the *Minimum Bisection* problem and it is known to be NP-hard in general [GJS76]. To overcome this limitation, various approximation algorithms were suggested over the years, culminating with a poly-logarithmic approximation algorithm [FK02]. Unfortunately, such approximations are not directly useful in the context of maximum likelihood estimation. Nevertheless, the maximum likelihood estimation problem at hand is not worst case, but rather a random problem. It can be viewed as a variant of the planted partition model (aka stochastic blockmodel) introduced in [DF89]. Indeed the block structure of Σ unveiled in Lemma 4 can be viewed as similar to the adjacency matrix of a weighted graph with a small bisection. Moreover, $\hat{\Sigma}$ can be viewed as the matrix Σ *planted* in some noise. Here, unlike the original planted partition problem, the noise is correlated and therefore requires a different analysis. In random matrix terminology, the observed matrix in the stochastic block model is of Wigner type, whereas in the IBM, it is of Wishart type. It is therefore not surprising that we can use the same methodology in both cases. In particular, we will use the semidefinite relaxation [GW95] to the MAXCUT problem that was already employed in the planted partition model [ABH16, HWX16].

It can actually be impractical to use directly the matrix $\hat{\Sigma}$ in the above relaxations, and we apply a pre-preprocessing that amounts to a centering procedure, which simplifies our analysis. Given $\sigma \in \{-1,1\}^p$, define its centered version $\bar{\sigma}$ by

$$\bar{\sigma} = \sigma - \frac{\mathbf{1}_{[p]}^\top \sigma}{p} \mathbf{1}_{[p]} = P\sigma \,,$$

where $P = I_p - \frac{1}{p}\mathbf{1}_{[p]}\mathbf{1}_{[p]}^\top$ is the projector onto the subspace orthogonal to $\mathbf{1}_{[p]}$. Moreover, let $\Gamma = P\Sigma P$ and $\hat{\Gamma} = P\hat{\Sigma}P$ respectively denote the covariance and empirical covariance matrices of the vector $\bar{\sigma}$.

Note that for all $V \in \mathcal{P}$, we have that $\mathbf{Tr}[\hat{\Gamma}V] = \mathbf{Tr}[\hat{\Sigma}V]$ since $V\mathbf{1}_{[p]}\mathbf{1}_{[p]}^\top = 0$, so that $PVP = V$. It implies that the likelihood function is unchanged over $\mathcal{P}$ when substituting $\hat{\Sigma}$ by $\hat{\Gamma}$. Moreover, $\mathbb{E}[\hat{\Gamma}] = \Gamma$ and the spectral decomposition of Γ is given by

$$\Gamma = (1-\Delta)P + p\frac{\Delta - \Omega}{2} u_S u_S^\top \,, \tag{3.2}$$

where $u_S = v_S/\sqrt{p}$ is a unit vector. Therefore the matrix Γ has leading eigenvalue $(1-\Delta) + p(\Delta - \Omega)/2$ with associated unit eigenvector u_S. Moreover, its eigengap is $p(\Delta - \Omega)/2$. It is well known in matrix perturbation theory that the eigengap plays a key role in the stability of the spectral decomposition of Γ when observed with noise.

3.2 Exact recovery via semidefinite programming

In this subsection, we consider the following semi-definite programming (SDP) relaxation of the optimization problem (3.1):

$$\max_{V \in \mathcal{E}} \mathbf{Tr}[\hat{\Gamma}V] \,, \qquad \mathcal{E} = \left\{ V \in \mathcal{S}_p : \mathbf{diag}(V) = \mathbf{1}_{[p]}, V \succeq 0 \right\} , \tag{3.3}$$

where $\mathcal{S}_p$ denotes the set of $p \times p$ symmetric real matrices. The set $\mathcal{E}$ is the set of correlation matrices, and it is known as the *elliptope*. We recall the definition of the vector $v_S = \mathbf{1}_S - \mathbf{1}_{\bar{S}} \in \{-1, 1\}^p$ and note that $v_S v_S^\top \in \mathcal{P} \subset \mathcal{E}$. Moreover, we denote by $\hat{V}^{\mathsf{SDP}}$ any solution to the the above program. Our goal is to show that (3.3) has a unique solution given by $\hat{V}^{\mathsf{SDP}} = v_S v_S^\top$, i.e., the SDP relaxation is tight. While the dual certificate approach of [ABH16] could be used in this case (see also [HWX16]) we employ a slightly different proof technique, more geometric, that we find to be more transparent. This approach is motivated by the idea that the relaxation is tight in the population case, suggesting that it might be the case as well when $\hat{\Gamma}$ is close to Γ.

Recall that for any $X_0 \in \mathcal{E}$, the normal cone to $\mathcal{E}$ at X_0 is denoted by $\mathcal{N}_{\mathcal{E}}(X_0)$ and defined by

$$\mathcal{N}_{\mathcal{E}}(X_0) = \big\{C \in \mathcal{S}_p \,:\, \mathbf{Tr}(CX) \le \mathbf{Tr}(CX_0)\,,\, \forall X \in \mathcal{E}\big\}\,.$$

It is the cone of matrices $C \in \mathcal{S}_p$ such that $\max_{X\in\mathcal{E}} \mathbf{Tr}(CX) = \mathbf{Tr}(CX_0)$. Therefore, $v_S v_S^\top$ is a solution of (3.3), i.e., the SDP relaxation is tight, whenever $\hat{\Gamma} \in \mathcal{N}_{\mathcal{E}}(v_S v_S^\top)$. The normal cone can be described using the following Laplacian operator. For any matrix $C \in \mathcal{S}_p$, define

$$L_S(C) := \mathbf{diag}(Cv_S v_S^\top) - C,$$

and observe that $L_S(C)v_S = 0$. Indeed, since $v_S \in \{-1, 1\}^p$, it holds,

$$\mathbf{diag}(Cv_S v_S^\top)v_S = \mathbf{diag}(Cv_S \mathbf{1}_{[p]}^\top)\mathbf{1}_{[p]} = Cv_S\,.$$

PROPOSITION 7. *For any matrix $C \in \mathcal{S}_p$, the following are equivalent*

1. $C \in \mathcal{N}_{\mathcal{E}_p}(v_S v_S^\top)\,.$
2. $L_S(C) = \mathbf{diag}(Cv_S v_S^\top) - C \succeq 0\,,$

Moreover, if $L_S(C) \succeq 0$ has only one eigenvalue equal to 0, then $v_S v_S^\top$ is the unique maximizer of $\mathbf{Tr}(CV)$ over $V \in \mathcal{E}$.

PROOF. It is known [LP96] that the normal cone $\mathcal{N}_{\mathcal{E}}(v_S v_S^\top)$ is given by

$$\mathcal{N}_{\mathcal{E}}(v_S v_S^\top) = \big\{C \in \mathcal{S}_p \,:\, C = D - M, D \text{ diagonal, }, M \succeq 0, v_S^\top M v_S = 0\big\}\,,$$

where $M \succeq 0$ denotes that M is a symmetric, semidefinite positive matrix. We are going to make use of the following facts. First for any diagonal matrix D and any $V \in \mathcal{E}$, it holds $\mathbf{diag}(DV) = D$. Second, taking $V = v_S v_S^\top$, we get

$$L_S(C)v_S v_S^\top = \mathbf{diag}(Cv_S v_S^\top)v_S v_S^\top - Cv_S v_S^\top = \mathbf{diag}(Cv_S v_S^\top) - Cv_S v_S^\top\,,$$

so that

$$\mathbf{diag}(L_S(C)v_S v_S^\top) = 0\,. \tag{3.4}$$

2. $\Rightarrow$ 1. Let $C \in v_S^\top$ be such that $L_S(C) \succeq 0$. By definition, we have $C = \mathbf{diag}(Cv_S v_S^\top) - L_S(C)$ and it remains to check that $v_S^\top L_S(C)v_S = 0$, which follows readily from (3.4) with $V = v_S v_S^\top$.

1. $\Rightarrow$ 2. Let $C = D - M \in \mathcal{N}_{\mathcal{E}_p}(v_S v_S^\top)$ where D is diagonal and $M \succeq 0$, $v_S^\top M v_S = 0$, which implies that $Mv_S = 0$. It yields, $Cv_S v_S^\top = Dv_S v_S^\top$ and

$\mathbf{diag}(Cv_Sv_S^\top) = \mathbf{diag}(Dv_Sv_S^\top) = D$ so that the decomposition is necessarily $D = \mathbf{diag}(Cv_Sv_S^\top)$ and $M = L_S(C) = \mathbf{diag}(Cv_Sv_S^\top) - C$. In particular, $L_S(C) \succeq 0$.

Thus, if $L_S(C) \succeq 0$ then $v_Sv_S^\top$ is a maximizer of $\mathbf{Tr}(CV)$ over $V \in \mathcal{E}$. To prove uniqueness, recall that for any maximizer $V \in \mathcal{E}$, we have $\mathbf{Tr}(CV) = \mathbf{Tr}(Cv_Sv_S^\top)$. Plugging $C = \mathbf{diag}(Cv_Sv_S^\top) - L_S(C)$ and using (3.4) yields

$$\begin{aligned}\mathbf{Tr}(\mathbf{diag}(Cv_Sv_S^\top)V) - \mathbf{Tr}(L_S(C)V) &= \mathbf{Tr}(\mathbf{diag}(Cv_Sv_S^\top)v_Sv_S^\top)\\ &= \mathbf{Tr}(\mathbf{diag}(Cv_Sv_S^\top))\,.\end{aligned}$$

Recall that $\mathbf{Tr}(\mathbf{diag}(Cv_Sv_S^\top)V) = \mathbf{Tr}(\mathbf{diag}(Cv_Sv_S^\top))$ so that the above display yields $\mathbf{Tr}(L_S(C)V) = 0$. Since $V \succeq 0$ and the kernel of the semidefinite positive matrix $L_S(C)$ is spanned by v_S, we have that $V = v_Sv_S^\top$. □

It follows from Proposition 7 that if $L_S(\hat\Gamma) \succeq 0$ and has only one eigenvalue equal to zero, then $v_Sv_S^\top$ is the solution to (3.3). In particular, in this case, the SDP allows exact recovery of the block structure $(S, \bar S)$. Observe that the conditions of Proposition 7 hold if $\hat\Gamma$ is replaced by the population matrix Γ. Indeed, using (3.2), we obtain

$$\begin{aligned}L_S(\Gamma) &= \big(1 - \Delta + p\frac{\Delta - \Omega}{2}\big)I_p - (1-\Delta)P - p\frac{\Delta-\Omega}{2}u_Su_S^\top\\ &= (1-\Delta)\frac{\mathbf{1}_{[p]}}{\sqrt{p}}\frac{\mathbf{1}_{[p]}^\top}{\sqrt{p}} - p\frac{\Delta-\Omega}{2}u_Su_S^\top + p\frac{\Delta-\Omega}{2}I_p\,,\end{aligned}$$

where we used the fact that $I_p - P$ is the projector onto the linear span of $\mathbf{1}_{[p]}$. Therefore, the eigenvalues of $L_S(\Gamma)$ are 0, $1 - \Delta + p(\Delta - \Omega)/2$, both with multiplicity 1 and $p(\Delta-\Omega)/2$ with multiplicity $p-1$. In particular, for $p \ge 2$, $L_S(\Gamma) \succeq 0$ and it has only one eigenvalue equal to zero.

Extending this result to $L_S(\hat\Gamma)$ yields the following theorem, as illustrated in FIgure 3. Let $C_{\alpha,\beta} > 0$ be a positive constant such that $\Delta - \Omega > C_{\alpha,\beta}/p$. Note that such a constant $C_{\alpha,\beta}$ is guaranteed to exist in view of Proposition 5.

Theorem 8. *The SDP relaxation (3.3) has unique maximum at $V = v_Sv_S^\top$ with probability $1-\delta$ whenever*

$$n > 16\big(3 + \frac{2}{C_{\alpha,\beta}}\big)\frac{\log(4p/\delta)}{\Delta - \Omega}(1 + o_p(1))\,.$$

In particular, the SDP relaxation recovers exactly the block structure $(S, \bar S)$.

Proof. Recall that $L_S(\hat\Gamma)v_S = 0$ and any $C \in \mathcal{S}_p$, denote by $\lambda_2[C]$ its second smallest eigenvalue. Our goal is to show that $\lambda_2[L_S(\hat\Gamma)] > 0$. To that end, observe that

$$L_S(\hat\Gamma) = L_S(\Gamma) + \mathbf{diag}\left((\hat\Gamma - \Gamma)v_Sv_S^\top\right) + \Gamma - \hat\Gamma\,.$$

Therefore, using from Weyl's inequality and the fact $\lambda_2[L_S(\Gamma)] = p(\Delta-\Omega)/2$, we get

$$\lambda_2[L_S(\hat\Gamma)] \ge p\frac{\Delta-\Omega}{2} - \|\,\mathbf{diag}\left((\hat\Gamma-\Gamma)v_Sv_S^\top\right)\|_{\mathrm{op}} - \|\hat\Gamma - \Gamma\|_{\mathrm{op}}\,, \tag{3.5}$$

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm. Therefore, it is sufficient to upper bound the above operator norms. This is ensured by the following Lemma.

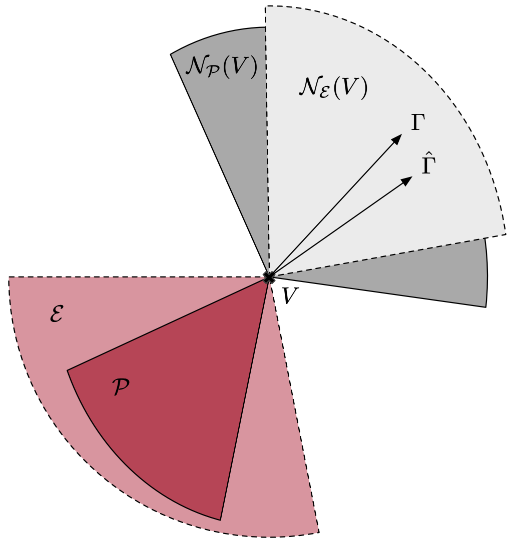

Figure 3: The main idea of our analysis of this convex relaxation. The true value of the parameter $V = v_S v_S^\top$ is the unique solution of both the maximum likelihood problem on $\mathcal{P}$ and of the convex relaxation on $\mathcal{E}$, as Γ belongs to both normal cones at V. The relaxation is therefore tight in the population case. We show that when the sample size is large enough, the sample matrix $\hat{\Gamma}$ is close enough to Γ and also in both normal cones.

LEMMA 9. *Fix $\delta > 0$ and define*

$$\mathcal{R}_{n,p}(\delta) = 2p \max\Big(\sqrt{\frac{(1+2/C_{\alpha,\beta})(\Delta-\Omega)\log(4p/\delta)}{n}}, \frac{(6+4/C_{\alpha,\beta})\log(p/\delta)}{n}\Big).$$

With probability $1-\delta$, it holds simultaneously that

$$\|\hat{\Gamma} - \Gamma\|_{\mathrm{op}} \le \mathcal{R}_{n,p}(\delta)(1+o_p(1)). \tag{3.6}$$

and

$$\|\,\mathbf{diag}\big((\hat{\Gamma}-\Gamma)v_S v_S^\top\big)\|_{\mathrm{op}} \le \mathcal{R}_{n,p}(\delta)(1+o_p(1)). \tag{3.7}$$

PROOF. To prove (3.6), we use a Matrix Bernstein inequality from [Tro15]. To that end, note that

$$\hat{\Gamma} - \Gamma = \frac{1}{n}\sum_{t=1}^{n} M_t\,,$$

where $M_1, \dots, M_n$ are i.i.d random matrices given by $M_t = (\bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top} - \Gamma)$, $t = 1, \dots, n$. We have

$$\|M_t\|_{\mathrm{op}} \le \|\bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top}\|_{\mathrm{op}} + \|\Gamma\|_{\mathrm{op}} \le p + \|\Gamma\|_{\mathrm{op}}.$$

Furthermore, we have that

$$\begin{aligned}\mathbb{E}[M_t^2] &= \mathbb{E}[\|\bar{\sigma}^{(t)}\|^2\bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top} - \bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top}\Gamma - \Gamma\bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top} + \Gamma^2] \\ &= p\mathbb{E}[\bar{\sigma}^{(t)}\bar{\sigma}^{(t)\top}] - \Gamma^2 - \Gamma^2 + \Gamma^2 \preceq p\Gamma.\end{aligned}$$

As a consequence, $\sum_{t=1}^n \mathbb{E}[M_t^2] \preceq p\Gamma$. By Theorem 1.6.2 in [Tro15], this yields

$$\mathbb{P}\big(\|\hat\Gamma - \Gamma\|_{\mathrm{op}} > t\big) \le 2p \exp\Big(-\frac{nt^2}{2p\|\Gamma\|_{\mathrm{op}} + 2(p + \|\Gamma\|_{\mathrm{op}})t}\Big)\,. \tag{3.8}$$

We have $\|\hat\Gamma - \Gamma\|_{\mathrm{op}} \le t$ with probability $1-\delta$ for any t such that

$$\log(2p/\delta) \le \frac{nt^2}{2p\|\Gamma\|_{\mathrm{op}} + 2(p + \|\Gamma\|_{\mathrm{op}})t}\,.$$

This holds for all

$$t \le \max\Big(\sqrt{\frac{4p\|\Gamma\|_{\mathrm{op}}\log(2p/\delta)}{n}}\,,\ \frac{4(p + \|\Gamma\|_{\mathrm{op}})\log(2p/\delta)}{n}\Big)\,.$$

To conclude the proof of (3.6), observe that

$$\|\Gamma\|_{\mathrm{op}} = p\frac{\Delta - \Omega}{2} + 1 - \Delta \le \big(1 + \frac{1}{C_{\alpha,\beta}}\big)(\Delta - \Omega)p\,,$$

where $C_{\alpha,\beta} > 0$ is defined immediately before the statement of Theorem 8.

We now turn to the proof of (3.7). Recall that $v_S \in \{-1,1\}^p$ so that the ith diagonal element is given by

$$\mathbf{diag}\,\big((\hat\Gamma - \Gamma)v_S v_S^\top\big)_{ii} = e_i^\top(\hat\Gamma - \Gamma)v_S\,,$$

where e_i denotes the ith vector of the canonical basis of $\mathbb{R}^p$. Hence,

$$\|\,\mathbf{diag}\,\big((\hat\Gamma - \Gamma)v_S v_S^\top\big)\|_{\mathrm{op}} = \max_{i\in[p]}\big|\,\mathbf{diag}\,\big((\hat\Gamma - \Gamma)v_S v_S^\top\big)_{ii}\big| = \max_{i\in[p]}|e_i^\top(\hat\Gamma - \Gamma)v_S|\,.$$

We bound the right hand-side of the above inequality by noting that

$$e_i^\top(\hat\Gamma - \Gamma)v_S = \frac{m}{n}\sum_{t=1}^n \big(\bar\sigma_i^{(t)}(\mu_S^{(t)} - \mu_{\bar S}^{(t)}) - \mathbb{E}[\bar\sigma_i^{(t)}(\mu_S^{(t)} - \mu_{\bar S}^{(t)})]\big)\,,$$

where $\mu_S^{(t)} = \mathbf{1}_S^\top\bar\sigma^{(t)}/m \in [-1,1]$ and $\mu_{\bar S}^{(t)}$ is defined analogously. The random variables $\bar\sigma_i^{(t)}(\mu_S^{(t)} - \mu_{\bar S}^{(t)}) - \mathbb{E}[\bar\sigma_i^{(t)}(\mu_S^{(t)} - \mu_{\bar S}^{(t)})]$ are centered, i.i.d., and are bounded in absolute value by 2 for all $t \in [n]$. Moreover, it follows from Lemma 4 that the variance of these random variables is bounded by

$$\mathbb{E}[(\mu_S^{(t)} - \mu_{\bar S}^{(t)})^2] \le 2(\Delta - \Omega) + \frac{4}{p} =: \nu^2\,.$$

By a one-dimensional Bernstein inequality, and a union bound over p terms, we have therefore that

$$\mathbb{P}\big(\max_{i\in[p]}|e_i^\top(\hat\Gamma - \Gamma)v_S| > \frac{pt}{n}\big) \le 2p\exp\Big(-\frac{t^2/2}{n\nu^2 + 2t/3}\Big)\,.$$

which yields

$$\max_{i\in[p]}|e_i^\top(\hat\Gamma - \Gamma)v_S| \le p\max\Big(\sqrt{\frac{2\nu^2\log(2p/\delta)}{n}}, \frac{4\log(2p/\delta)}{3n}\Big)\,,$$

with probability $1-\delta$. It completes the proof of (3.7). □

To conclude the proof of Theorem 8, note that for the prescribed choice of n, we have

$$2\mathcal{R}_{n,p}(\delta)(1+o_p(1)) < p\frac{\Delta-\Omega}{2}$$

and it follows from (3.5) that $\lambda_2[L_S(\hat{\Gamma})] > 0$. □

REMARK 10. *We have not attempted to optimize the constant $16(3+2/C_{\alpha,\beta})$ that appears in Theorem 8 and it is arguably suboptimal. One way to see how it can be reduced at least by a factor 2 is by noting that the factor p in the right-hand side of (3.8) is in fact superfluous thus resulting in a extra logarithmic factor in (3.6). This is because, akin to the stochastic blockmodel analysis [ABH16], the matrix deviation inequality from [Tro15] is too coarse for this problem. The extra factor p may be removed using the concentration results of Section 2.3 but at the cost of a much longer argument. Indeed, using Theorem 3, we can establish the concentration of local magnetization around the ground states and conditionally on these magnetizations, the configurations are uniformly distributed. These conditional distributions can be shown to exhibit sub-Gaussian concentration so that $\sigma^\top u$ and thus $\bar{\sigma}^\top u$ are sub-Gaussian with constant variance proxy for any unit vector $u \in \mathbb{R}^p$. This result can yield a bound for $\|\hat{\Gamma}-\Gamma\|_{\mathrm{op}}$ using an ε-net argument that is standard in covariance matrix estimation. With this in mind, we could get an upper bound in (3.6) that is negligible with respect to $\mathcal{R}_{n,p}$ thereby removing a factor 2. Nevertheless, in absence of a tight control of the constant $C_{\alpha,\beta}$, exact constants are hopeless and beyond the scope of this paper.*

Combined with Proposition 5 that quantifies the gap $\Delta-\Omega$ in terms of p, Theorem 8 readily yields the following corollary.

COROLLARY 11. *There exists positive constants C_1 and C_2 that depend on α and β such that the following holds. The SDP relaxation (3.3) recovers the block structure $(S, \bar{S})$ exactly with probability $1-\delta$ whenever*

1. $n \geq C_1 p \log(p/\delta)$ *if* $\beta + |\alpha| < 2$ *or* $\alpha > 0$
2. $n \geq C_2 \log(p/\delta)$ *otherwise.*

In particular, if $\beta - \alpha > 2, \alpha \leq 0$ a number of observations that is logarithmic in the dimension p is sufficient to recover the blocks exactly.

These results suggest that there is a sharp phase transition in sample complexity for this problem, depending on the value of the parameters α and β. We address this question further in Section 4. The last subsection shows that these rates are, in fact, optimal.

3.3 Information theoretic limitations

In this section, we present lower bounds on the sample size needed to recover the partition $(S, \bar{S})$ and compare them to the upper bounds of Theorem 8. In the sequel, we write $\hat{S} \asymp S$ if either $(\hat{S}, \bar{\hat{S}}) = (S, \bar{S})$ or $(\hat{S}, \bar{\hat{S}}) = (\bar{S}, S)$ to indicate that the two partitions are the same. We write $\hat{S} \not\asymp S$ to indicate that the two partitions are different.

For any balanced partition $(S, \bar{S})$, consider a "neighborhood" $\mathcal{T}_S$ of $(S, \bar{S})$ composed of balanced partitions such that for all $(T, \bar{T}) \in \mathcal{T}_S$, we have $\rho(S,T) = 1$

and $\rho(\bar{S}, \bar{T}) = 1$. We first compute the Kullback–Leibler divergence between the distributions $\mathbb{P}_S$ and $\mathbb{P}_T$.

LEMMA 12. *For any positive β, $\alpha < \beta$, and $T \in \mathcal{T}_S$, it holds that*

$$\mathsf{KL}(\mathbb{P}_T, \mathbb{P}_S) = \frac{p-2}{p}(\beta - \alpha)(\Delta - \Omega)\,.$$

PROOF. By definition of the divergence and of the distributions, we have that

$$\begin{aligned}\mathsf{KL}(\mathbb{P}_T, \mathbb{P}_S) &= \mathbb{E}_T\Big[\log\Big(\frac{\mathbb{P}_T}{\mathbb{P}_S}(\sigma)\Big)\Big] \\ &= \mathbb{E}_T\big[\mathbf{Tr}[(Q_T - Q_S)\sigma\sigma^\top]\big] \\ &= \mathbf{Tr}[(Q_T - Q_S)\Sigma_T]\end{aligned}$$

Note that most of the coefficients of $Q_T - Q_S$ are equal to 0. In fact, noting $\{s\} = S \cap \bar{T}$ and $\{t\} = \bar{S} \cap T$, we have

$$(Q_T - Q_S)_{ij} = \frac{\alpha - \beta}{p} \quad \text{if} \quad \begin{cases} i \in S \setminus \{s\}\,,\ j = s \\ i = s\,,\ j \in S \setminus \{s\} \\ i \in \bar{S} \setminus \{t\}\,,\ j = t \\ i = t\,,\ j \in \bar{S} \setminus \{t\} \end{cases}$$

and

$$(Q_T - Q_S)_{ij} = \frac{\beta - \alpha}{p} \quad \text{if} \quad \begin{cases} i \in S \setminus \{s\}\,,\ j = t \\ i = s\,,\ j \in \bar{S} \setminus \{t\} \\ i \in \bar{S} \setminus \{t\}\,,\ j = s \\ i = t\,,\ j \in S \setminus \{s\}\,, \end{cases}$$

and 0 otherwise. There are therefore $p-2$ coefficients of each sign. Furthermore, whenever $(Q_T - Q_S)_{ij} = (\alpha - \beta)/p$, we have $(\Sigma_T)_{ij} = \Omega$, and whenever $(Q_T - Q_S)_{ij} = (\beta - \alpha)/p$, we have $(\Sigma_T)_{ij} = \Delta$. Computing $\mathbf{Tr}[(Q_T - Q_S)\Sigma_T]$ explicitly yields the desired result. $\square$

From this lemma, we derive the following lower bound.

THEOREM 13. *For $\gamma \in (0, 3/5)$ and $p \geq 6$ and*

$$n \leq \frac{\gamma \log(p/4)}{(\beta - \alpha)(\Delta - \Omega)}.$$

We have

$$\inf_{\hat{S}} \max_{S \in \mathcal{S}} \mathbb{P}_S^{\otimes n}\big((\hat{S}, \bar{\hat{S}}) \not\asymp (S, \bar{S})\big) \geq \frac{p-2}{p}\big(1 - \gamma - \sqrt{\gamma}\big) > 0\,,$$

where the infimum is taken over all estimators of S. Note that the right-hand side of the above inequality goes to 1 as $p \to \infty$ and $\gamma \to 0$.

PROOF. First, note that by Lemma 12, for any $T \in \mathcal{T}_S$, it holds $|\mathcal{T}_S| = (p/2-1)^2$ so that

$$\mathsf{KL}(\mathbb{P}_T^{\otimes n}, \mathbb{P}_S^{\otimes n}) = n\mathsf{KL}(\mathbb{P}_T, \mathbb{P}_S) \le n(\beta-\alpha)(\Delta-\Omega) \le \gamma \log(p/4) \le \frac{\gamma}{2}\log|\mathcal{T}_S|\,.$$

Thus Theorem 2.5 in [Tsy09] yields

$$\begin{aligned}\inf_{\hat S}\max_{S\in\mathcal{P}} \mathbb{P}_S^{\otimes n}(\hat S \not\asymp S) &\ge \frac{\sqrt{|\mathcal{T}_S|}}{1+\sqrt{|\mathcal{T}_S|}}\Big(1-\gamma-\sqrt{\frac{\gamma}{\log(|\mathcal{T}_S|)}}\Big)\\ &\ge \frac{p-2}{p}\big(1-\gamma-\sqrt{\gamma}\big) > 0\,,\end{aligned}$$

for $\gamma \in (0, 3/5)$. □

The lower bound of Theorem 13 matches the upper bounds of Theorem 8 up to numerical constant. This indicates that the SDP relaxation studied in the paper is rate optimal: the sample complexity stated in Corollary 11 has optimal dependence on the dimension p. Note that past work on exact recovery in the stochastic blockmodel [ABH16, HWX16] was able to show that SDP was also optimal with respect to constants. We do not pursue this questions in the present paper.

4. CONCLUSION AND OPEN PROBLEMS

This paper introduced the Ising block model (IBM) for large binary random vectors with an underlying cluster structure. In this model, we studied the sample complexity of recovering exactly the clusters. Unsurprisingly, this paper bears similarities with the stochastic blockmodel but also differences. For example, in the stochastic blockmodel one is given only one observation of the graph. In the IBM, given one realization $\sigma^{(1)} \in \{-1,1\}^p$, the maximum likelihood estimator is the trivial clustering that assigns $i \in [p]$ to a cluster according to the sign of $\sigma_i^{(1)}$, up to a trivial reassignment to keep the partition balanced.

Below is a summary of our main findings:

1. The model exhibits three phases depending on the values taken by two parameters.
2. In one phase, where the two clusters tend to have opposite behavior, the sample complexity is logarithmic in the dimension; in the other two, it is near linear. These sample complexities are proved to be optimal in an information theoretic sense.
3. Akin to the stochastic blockmodel, the optimal sample complexity is achieved using the natural semidefinite relaxation to the MAXCUT problem.

Many questions regarding this model remain open. The first and most natural is the determination of exact constants. Theorem 13 suggests that there exists a universal constant $C^\star$ such that the optimal sample complexity is

$$\frac{C^\star \log(p)}{\beta-\alpha}(\Delta-\Omega)(1+o_p(1))\,.$$

Throughout this paper, we have only kept loosely track of the correct dependency of the constants as function of the constants (α, β). We have shown that the optimal sample complexity, is a product of $\log(p)/(\Delta - \Omega)$ and of a constant term that only becomes arbitrarily large when α is very close to β, with a divergence of order $(\beta - \alpha)^{-1}$. In the spirit of exact thresholds for the stochastic blockmodel [Mas14, MNS15, ABH16], we find that proving existence of the constant $C^\star$ and computing it worthy of investigation but is beyond the scope of the present paper.

Finally, in view of the simple spectral decomposition (3.2) of Γ, one may wonder about the behavior of the a simple method that consists in computing the leading eigenvector of $\hat{\Gamma}$ and clustering according to the sign of its entries. Such a method is the basis of the approach in denser graph models in [McS01] or [AKS98]. The results of such an approach are easily implementable as follows.

Let $\hat{u}$ denote a leading unit eigenvectors of $\hat{\Gamma}$ and consider the following estimate for the partition $(S, \bar{S})$:

$$\hat{S} \asymp \{i \in [p] \mid \hat{u}_i > 0\}\,. \tag{4.1}$$

It follows from the Perron-Frobenius theorem that $\hat{S} \asymp S$ whenever $\operatorname{sign}(\hat{\Gamma}) = \operatorname{sign}(\Gamma)$. This allows for perfect recovery of S, but only holds with high probability when n is of order $\log(p)/(\Delta - \Omega)^2$, which is suboptimal. It is however possible to obtain partial recovery guarantees for the spectral recovery. In order to state our result, for any two partitions $(S, \bar{S})$, $(T, \bar{T})$ define

$$|S \diamond T| = \min\Big(|S \triangle T|, |S \triangle \bar{T}|\Big)$$

where $\triangle$ denotes the symmetric difference.

Proposition 14. *Fix $\delta \in (0,1)$ and let $\hat{S} \subset [p]$ be defined in (4.1). Then, there exits a constant $\gamma_{\alpha,\beta} > 0$ such that with probability $1 - \delta$,*

$$\frac{1}{p}|S \diamond \hat{S}| \le \gamma_{\alpha,\beta} \frac{\log(4p/\delta)}{n(\Delta - \Omega)}\,.$$

Proof. Let $\hat{u}$ denote the leading unit eigenvector of $\hat{\Gamma}$ and let $\hat{v} = \sqrt{p}\hat{u}$. Recall that $v_S = \mathbf{1}_S - \mathbf{1}_{\bar{S}}$ and observe that

$$\begin{aligned} |S \diamond \hat{S}| &= \min\Big(\sum_{i=1}^p \mathbb{I}(\hat{v}_i \cdot (v_S)_i \le 0), \sum_{i=1}^p \mathbb{I}(\hat{v}_i \cdot (v_S)_i \ge 0)\Big) \\ &\le \min\big(\|\hat{v} - v_S\|^2, \|\hat{v} + v_S\|^2\big) = p \min\big(\|\hat{u} - u_S\|^2, \|\hat{u} + u_S\|^2\big)\,, \end{aligned}$$

where in the inequality, we used the fact that $v_S \in \{-1, 1\}^p$ so that

$$\mathbb{I}(\hat{v}_i \cdot (v_S)_i \le 0) \le |\hat{v}_i - (v_S)_i| \mathbb{I}(\hat{v}_i \cdot (v_S)_i \le 0) \le |\hat{v}_i - (v_S)_i|^2\,.$$

Using a variant of the Davis-Kahan lemma (see, e.g, [WBS16]), we get

$$\frac{1}{p}|S \diamond \hat{S}| \le \frac{\|\hat{\Gamma} - \Gamma\|_{\text{op}}^2}{(\lambda_1(\Gamma) - \lambda_2(\Gamma))^2}\,,$$

and the result follows readily from (3.6) and the fact that the eigengap of Γ is given by $p(\Delta - \Omega)/2$. □

In terms of exact recovery, this result is quite weak as it only gives guarantees for a sample complexity of the order of $p\log(p/\delta)/(\Delta-\Omega)$, which is suboptimal by a factor of p. Moreover, for the bound of Proposition 14 to be non-trivial, one already needs the sample size to be of the same order as the one required for exact recovery by semi-definite programming. Nevertheless Proposition 14 raises the question of the optimal rates of estimation of S with respect to the metric $|S\diamond\hat{S}|/p$. While partial recovery is beyond the scope of this paper, it would be interesting to establish the optimal rate.

APPENDIX A: FACTS ABOUT THE CURIE-WEISS MODEL

We begin by stating some well known facts about the Curie-Weiss model. These results are standard in the statistical physics literature and the interested reader can find more details in [FV16, Ell06] for example. However, the precise behavior of the free energy that we need for our subsequent analysis does not seem to be readily available in the literature so we prove below a lemma that suits our purposes.

Recall that the Curie-Weiss model is a special case of the Ising block model when $\alpha=\beta=b$. In this case, the free energy takes the form:

$$g_b^{\mathrm{CW}}(\mu)=-2b\mu^2-4h\Big(\frac{\mu+1}{2}\Big) \tag{A.1}$$

where we recall that $\mu=\sigma^\top\mathbf{1}/p$ is the global magnetization of σ. The minima $x\in(-1,1)$ of g are called ground states and satisfy the first order optimality condition, also known as *mean field equation*

$$\log\Big(\frac{1+x}{1-x}\Big)=2bx\,.$$

If $b\le 1$, then the unique solution to the mean field equation is $x=0$. Moreover, g_b^{CW} is increasing on $[0,1]$.

If $b>1$, then the mean field equation has two solutions $\tilde{x}>0$ and $-\tilde{x}$ in $(-1,1)$. In any case, these solutions are global minima that are also the only local minima of g_b^{CW}. In particular, when $b>1$, g_b^{CW} is monotone decreasing in the interval $(0,\tilde{x})$ and monotone increasing in the interval $(\tilde{x},1)$.

The following lemma is a refinement of these well-known facts that quantifies the curvature of g_b^{CW} around its minima.

LEMMA 15. *Fix $b>1$ in the Curie-Weiss model and denote by $\tilde{x}>0$ and $-\tilde{x}$ the two ground states. Then it holds:*

$$1-\frac{2b}{2b^2+b-1}<\tilde{x}^2<1-e^{-2b}\,.$$

Moreover, for any $x\in(0,1)$, it holds

$$g_b^{\mathrm{CW}}(x)\ge g_b^{\mathrm{CW}}(\tilde{x})+\frac{b-1}{2b}(|x-\tilde{x}|\wedge\varepsilon)^2, \tag{A.2}$$

and

$$g_b^{\mathrm{CW}}(x)\ge g_b^{\mathrm{CW}}(-\tilde{x})+\frac{b-1}{2b}(|x+\tilde{x}|\wedge\varepsilon)^2, \tag{A.3}$$

where $\varepsilon = \frac{e^{-2b}}{4}\big(1-\frac{1}{b}\big)$.

Fix $b \le 1$ *in the Curie-Weiss model and recall that* $\tilde{x}=0$ *is the unique ground state. Then for any* $x \in (-1,1)$ *it holds*

$$g_b^{\rm CW}(x) \ge g_b^{\rm CW}(0) + (1-b)(x\wedge\varepsilon')^2 \,. \tag{A.4}$$

where

$$\varepsilon' = \sqrt{\frac{1-b}{3}}\,.$$

Proof. Observe that for $x>0$, we have

$$2b\tilde{x} = \log\big(\frac{1+\tilde{x}}{1-\tilde{x}}\big) < \frac{2\tilde{x}}{1-\tilde{x}^2} - \gamma\tilde{x}^3\,, \quad \forall \gamma \le 1\,. \tag{A.5}$$

Taking $\gamma = 0$ implies that $\tilde{x} > \sqrt{1-1/b}$. Plugging this into (A.5) with $\gamma = 1$ yields

$$2b\tilde{x} < \frac{2\tilde{x}}{1-\tilde{x}^2} - \tilde{x}\big(1-\frac{1}{b}\big)\,.$$

Solving for $\tilde{x}$ once again yields

$$\frac{2}{1-\tilde{x}^2} > 2b + 1 - \frac{1}{b} \tag{A.6}$$

Or equivalently that

$$\tilde{x}^2 > 1 - \frac{2b}{2b^2+b-1}\,.$$

Moreover, the mean field equation yields

$$2b > 2b\tilde{x} = \log\big(\frac{1+\tilde{x}}{1-\tilde{x}}\big) > -\log(1-\tilde{x})$$

so that

$$\tilde{x} < 1 - e^{-2b} \tag{A.7}$$

which readily yields the desired upper bound on $\tilde{x}^2$.

We conclude this proof by showing that $g_b^{\rm CW}$ is at least quadratic in a neighborhood of its minima when $b \neq 1$. To that end, observe first that the second and third derivatives of g are given respectively by

$$\frac{\partial^2}{\partial x^2} g_b^{\rm CW}(x) = -4b + \frac{4}{1-x^2}\,, \qquad \frac{\partial^3}{\partial x^3} g_b^{\rm CW}(x) = -\frac{8x}{(1-x^2)^2}\,,$$

First assume that $b>1$. A Taylor expansion of $g_b^{\rm CW}$ around $\tilde{x}$ together with (A.6) and (A.7) yields that for any $\varepsilon \in (0,1)$ and x such that

$$|x-\tilde{x}| \le \varepsilon := \frac{e^{-2b}}{2} \wedge \big(1-\frac{1}{b}\big)\,,$$

$$\begin{aligned}
g_b^{\rm CW}(x) &\ge g_b^{\rm CW}(\tilde{x}) + \big(1-\frac{1}{b}\big)(x-\tilde{x})^2 - \frac{4}{3(1-(\tilde{x}+\varepsilon)^2)^2}|x-\tilde{x}|^3 \\
&\ge g_b^{\rm CW}(\tilde{x}) + \big(1-\frac{1}{b}\big)(x-\tilde{x})^2 - \frac{4}{3(1-\tilde{x}-\varepsilon)}|x-\tilde{x}|^3 \\
&\ge g_b^{\rm CW}(\tilde{x}) + \big(1-\frac{1}{b}\big)(x-\tilde{x})^2 - \frac{4\varepsilon}{3(e^{-2b}-\varepsilon)}(x-\tilde{x})^2 \\
&\ge g_b^{\rm CW}(\tilde{x}) + \frac{1}{2}\big(1-\frac{1}{b}\big)(x-\tilde{x})^2\,.
\end{aligned}$$

Now, using the fact that $g_b^{\rm CW}$ is monotone decreasing on $(0, \tilde{x} - \varepsilon)$ and monotone increasing in $(\tilde{x} + \varepsilon, 1)$, we obtain the claim in (A.2). The lower bound (A.3) follows by symmetry.

Next, assume that $b < 1$. A Taylor expansion of $g_b^{\rm CW}$ around 0 yields that for any x such that $|x| < \varepsilon', \varepsilon' \in (0,1)$,

$$\begin{aligned} g_b^{\rm CW}(x) &> g_b^{\rm CW}(0) + \big[2(1-b) - \frac{4\varepsilon^2}{3(1-\varepsilon^2)^2}\big]x^2 \\ &\geq g_b^{\rm CW}(0) + (1-b)x^2 \end{aligned}$$

for

$$\varepsilon' \leq \sqrt{\frac{1-b}{3}}\,.$$

Using the fact that $g_b^{\rm CW}$ is monotone decreasing on $[1, -\varepsilon)$ and monotone increasing on $(\varepsilon, 1]$ yields (A.4). □

REMARK 16. *When $b = 1$, the Hessian of $g_b^{\rm CW}$ vanishes at 0. In this case, $g_b^{\rm CW}$ is not lower bounded by a quadratic term.*

APPENDIX B: INEQUALITIES

B.1 Bounds on binomial coefficients

We will need the following well known information theoretic estimate. Recall that the binary entropy function $h : [0,1] \to \mathbb{R}$ is defined by $h(0) = h(1) = 0$ and for any $s \in (0,1)$ by

$$h(s) = -s\log(s) - (1-s)\log(1-s)\,.$$

LEMMA 17. *Let m be a positive integer and let $\gamma \in [0,1]$ be such that γm is an integer. Then*

$$\binom{m}{\gamma m} \leq \exp(mh(\gamma))\,.$$

PROOF. Let $X \sim \mathsf{Bin}(n, \gamma)$ be a binomial random variable. Then

$$1 \geq \mathbb{P}(X = \gamma m) = \binom{m}{\gamma m}\gamma^{\gamma m}(1-\gamma)^{(1-\gamma)m} = \binom{m}{\gamma m}\exp(-mh(\gamma))\,.$$

□

The following sharper estimate follows from the Stirling approximation of $n!$ developed in [Rob55].

LEMMA 18. *Let $\varepsilon > 0$, m a positive integer let $\gamma \in [\varepsilon, 1-\varepsilon]$ be such that γm is an integer. We then have*

$$\exp\left(-\frac{1}{12\varepsilon^2 m}\right) \leq \sqrt{2\pi m\gamma(1-\gamma)}\exp(mh(\gamma))\binom{m}{\gamma m} \leq \exp\left(\frac{1}{12m}\right)\,.$$

PROOF. It follows from [Rob55]) that for any positive integer n,

$$1 \le \exp\left(\frac{1}{12n+1}\right) \le \frac{n!}{\sqrt{2\pi n}(n/e)^n} \le \exp\left(\frac{1}{12n}\right).$$

Applying this to

$$\binom{m}{\gamma m} = \frac{m!}{(\gamma m)!((1-\gamma)m)!}$$

yields the desired bounds. □

B.2 Tail bound for χ^2 distribution

We recall here a well known tail bound for χ^2 distributions (see, [LM00, Lemma 1]).

LEMMA 19. *Let $Z \sim \mathcal{N}_2(0, I_2)$ be a bivariate standard Gaussian vector. Then, for any $t \ge 2$, it holds*

$$\mathbb{P}(\|Z\|_2^2 - 2 \ge 2) \le \exp(-t/4)\,.$$

REFERENCES

[ABH16] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, Jan 2016.

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, SODA '98, pages 594–598, Philadelphia, PA, USA, 1998. Society for Industrial and Applied Mathematics.

[BEGd08] Onureena Banerjee, Laurent El Ghaoui, and Alexandre d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *J. Mach. Learn. Res.*, 9:485–516, June 2008.

[Bes86] Julian Besag. On the statistical analysis of dirty pictures. *J. Roy. Statist. Soc. Ser. B*, 48(3):259–302, 1986.

[BMS08] Guy Bresler, Elchanan Mossel, and Allan Sly. Reconstruction of Markov random fields from samples: some observations and algorithms. In *Approximation, randomization and combinatorial optimization*, volume 5171 of *Lecture Notes in Comput. Sci.*, pages 343–356. Springer, Berlin, 2008.

[Bre15] Guy Bresler. Efficiently learning Ising models on arbitrary graphs [extended abstract]. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 771–782. ACM, New York, 2015.

[DF89] M. E. Dyer and A. M. Frieze. The solution of some random np-hard problems in polynomial expected time. *J. Algorithms*, 10(4):451–489, December 1989.

[DGH08] Persi Diaconis, Sharad Goel, and Susan Holmes. Horseshoes in multidimensional scaling and local kernel methods. *Ann. Appl. Stat.*, 2(3):777–807, 09 2008.

[Ell06] Richard S. Ellis. *Entropy, large deviations, and statistical mechanics.* Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1985 original.

[FK02] Uriel Feige and Robert Krauthgamer. A polylogarithmic approximation of the minimum bisection. *SIAM J. Comput.*, 31(4):1090–1118 (electronic), 2002.

[FV16] Sacha Friedli and Yvan Velenik. *Statistical Mechanics of Lattice Systems: a Concrete Mathematical Introduction.* Cambridge University Press, 2016.

[GJS76] M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified NP-complete graph problems. *Theoret. Comput. Sci.*, 1(3):237–267, 1976.

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42(6):1115–1145, 1995.

[HLL83] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109 – 137, 1983.

[HWX16] B. Hajek, Y. Wu, and J. Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016.

[Isi25] E. Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift fur Physik*, 31:253–258, February 1925.

[Lau96] Steffen L. Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series.* The Clarendon Press, Oxford University Press, New York, 1996. Oxford Science Publications.

[LM00] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5):1302–1338, 2000.

[LP96] M. Laurent and S. Poljak. On the facial structure of the set of correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 17(3):530–547, 1996.

[LS03] Steffen L. Lauritzen and Nuala A. Sheehan. Graphical models for genetic analyses. *Statist. Sci.*, 18(4):489–514, 2003.

[Mas14] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 694–703. ACM, 2014.

[McS01] Frank McSherry. Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)*, pages 529–537. IEEE Computer Soc., Los Alamitos, CA, 2001.

[MNS13] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *arXiv preprint arXiv:1311.4115*, 2013.

[MNS15] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3):431–461, 2015.

[MS99] Christopher D. Manning and Hinrich Schütze. *Foundations of statistical natural language processing.* MIT Press, Cambridge, MA, 1999.

[Rob55] Herbert Robbins. A remark on Stirling's formula. *Amer. Math. Monthly*, 62:26–29, 1955.

[RWL10] Pradeep Ravikumar, Martin J. Wainwright, and John D. Lafferty. High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *Ann. Statist.*, 38(3):1287–1319, 2010.

[SRN+05] Paola Sebastiani, Marco F Ramoni, Vikki Nolan, Clinton T Baldwin, and Martin H Steinberg. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. *Nature genetics*, 37(4):435–440, 2005.

[Tro15] Joel A. Tropp. An introduction to matrix concentration inequalities, 2015.

[Tsy09] Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[WBS16] T. Wang, Q. Berthet, and R. J. Samworth. Statistical and computational trade-offs in Estimation of Sparse Pincipal Components. *Ann. Statist.*, 44:1896–1930, 2016.

Quentin Berthet
Department of Pure Mathematics
and Mathematical Statistics,
University of Cambridge
Wilbeforce Road
Cambridge, CB3 0WB, UK
(q.berthet@statslab.cam.ac.uk)

Philippe Rigollet
Department of Mathematics
Massachusetts Institute of Technology
77 Massachusetts Avenue,
Cambridge, MA 02139-4307, USA
(rigollet@math.mit.edu)

Piyush Srivastava
CMS Department
California Institute of Technology
1200 E. California Blvd.,
Pasadena, CA 91125-50007, USA
(piyushs@caltech.edu)