Exact recovery in block spin Ising models at the critical line

Matthias Löwe

Fachbereich Mathematik und Informatik, Universität Münster, Orléans-Ring 10, 48149 Münster, Germany
e-mail: maloewe@math.uni-muenster.de

Kristina Schubert

Fakultät für Mathematik, Technische Universität Dortmund, Vogelpothsweg 87, 44227 Dortmund, Germany
e-mail: kristina.schubert@tu-dortmund.de

Abstract: We show how to exactly reconstruct the block structure at the critical line in the so-called Ising block model. This model was recently re-introduced by Berthet, Rigollet and Srivastava in [2]. There the authors show how to exactly reconstruct blocks away from the critical line and they give an upper and a lower bound on the number of observations one needs; thereby they establish a minimax optimal rate (up to constants). Our technique relies on a combination of their methods with fluctuation results obtained in [20]. The latter are extended to the full critical regime. We find that the number of necessary observations depends on whether the interaction parameter between two blocks is positive or negative: In the first case, there are about $N \log N$ observations required to exactly recover the block structure, while in the latter case $\sqrt{N} \log N$ observations suffice.

Keywords and phrases: Block models, Ising model, Curie-Weiss model, fluctuations, critical temperature.

1. Introduction

In a recent paper Berthet, Rigollet and Srivastava rediscovered a block version of the Curie-Weiss-Ising model [2]. This model had been introduced earlier in the statistical physics literature, see e.g. [11], [10], [9], [6]. Extensions of these models are studied in [19] or [21]. The first of these papers uses a very general interaction structure, while the latter investigates the situation in the spirit of social interaction models or statistical physics models on random graphs, see [23], [3], [7], and [18]. A related version of this model has been investigated using the method of moments in [15], [17] and [16].

The article by Berthet et al. is motivated by a considerable amount of articles investigating block models in the recent past, see e.g. [1], [12], [5], [22], [4]. The model is interesting from both a probabilistic and a statistical perspective.

To define the model, one starts by partitioning the set $\{1, \ldots, N\}$ into a set $S \subset \{1, \ldots, N\}$ with $|S| = \frac{N}{2}$ and its complement S^c. To this end we need to

1796

assume that N is even; the model itself can be defined and analyzed for arbitrary block sizes (see [19], where large deviations and Central Limit Theorems are proved for a general block structure), but the statistical questions then become more tricky.

The set S induces a Hamiltonian on the binary hypercube $\{-1, +1\}^N, N \in \mathbb{N}$ given by

$$H_{N,\alpha,\beta,S}(\sigma) := -\frac{\beta}{2N} \sum_{i \sim j} \sigma_i \sigma_j - \frac{\alpha}{2N} \sum_{i \nsim j} \sigma_i \sigma_j, \quad \sigma \in \{-1, +1\}^N. \qquad (1)$$

Here we choose $\beta > 0$ and $0 \le |\alpha| \le \beta$, and we write $i \sim j$, if either $i, j \in S$ or $i, j \in S^c$ and $i \nsim j$, otherwise. The Hamiltonian (1) (or energy function), in turn, induces a Gibbs measure on $\{-1, +1\}^N$ given by:

$$\mu_{N,\alpha,\beta}(\sigma) = \mu_{N,\alpha,\beta,S}(\sigma) := \frac{e^{-H_{N,\alpha,\beta}(\sigma)}}{\sum_{\sigma'} e^{-H_{N,\alpha,\beta}(\sigma')}} =: \frac{e^{-H_{N,\alpha,\beta}(\sigma)}}{Z_{N,\alpha,\beta}},$$

where $\sigma = (\sigma_i)_{i=1}^N \in \{-1, +1\}^N$.

The behavior of the model can sometimes be studied best, when analyzing an order parameter. In this case such an order parameter is given by the vector of block magnetizations, $m := m^N := (m_1^N, m_2^N)$, where

$$m_1 := m_1^N := m_1(\sigma) := \frac{2}{N} \sum_{i \in S} \sigma_i \quad \text{and} \quad m_2 := m_2^N := m_2(\sigma) := \frac{2}{N} \sum_{i \notin S} \sigma_i.$$

Its advantage is that the Hamiltonian can be rewritten as

$$H_{N,\alpha,\beta,S}(\sigma) = -\frac{N}{2} \left(\frac{1}{2}\alpha m_1 m_2 + \beta \frac{1}{4} m_1^2 + \frac{1}{4} \beta m_2^2 \right).$$

To describe and understand the phase transitions in the model recall that without a partitioning (and with $\alpha = \beta$) we would be back in the situation of the so-called Curie-Weiss model at inverse temperature β. This model is defined by the Hamiltonian $H_{CW}(\sigma) = \frac{1}{2N} \sum_{i,j} \sigma_i \sigma_j$ for $\sigma \in \{\pm 1\}^N$ and the corresponding Gibbs measure $\mu_{N,\beta}^{CW}(\sigma) = \frac{e^{-\beta H_{CW}(\sigma)}}{Z_{N,\beta}^{CW}}$. It has been extensively studied, see e.g. [8]. In particular, it has been shown that its equilibrium measures are intrinsically related to the largest solution, $m^+(\beta)$, of the equation

$$z = \tanh(\beta z).$$

More precisely, for all $\beta \ge 0$, in the Curie-Weiss model the random variable $\frac{1}{N} \sum_{i=1}^N \sigma_i$ asymptotically concentrates in the points $m^+(\beta)$ and $-m^+(\beta)$. Note that $m^+(\beta) = 0$, if and only if $\beta \le 1$.

A similar change in the behavior was shown to be true for the block spin Ising model, see [2].

Theorem 1.1 (cf. [2, Proposition 4.1]). *In the above setting assume that $|\alpha| \le \beta$ and denote by $\rho_{N,\alpha,\beta}$ the distribution of m under the Gibbs measure $\mu_{N,\alpha,\beta}$.*

- If $\beta + |\alpha| \le 2$, then $\rho_{N,\alpha,\beta}$ weakly converges to the Dirac measure in $(0,0)$.
- If $\beta + |\alpha| > 2$ and $\alpha = 0$, then $\rho_{N,\alpha,\beta}$ weakly converges to the mixture of Dirac measures $\frac{1}{4} \sum_{s_1,s_2 \in \{-,+\}} \delta_{(s_1 m^+(\beta/2), s_2 m^+(\beta/2))}$.
- If $\beta + \alpha > 2$ and $\alpha > 0$, then $\rho_{N,\alpha,\beta}$ weakly converges to the mixture of Dirac measures: $\frac{1}{2}\big(\delta_{(m^+(\frac{\alpha+\beta}{2}), m^+(\frac{\alpha+\beta}{2}))} + \delta_{(-m^+(\frac{\alpha+\beta}{2}), -m^+(\frac{\alpha+\beta}{2}))}\big)$.
- If $\beta + |\alpha| > 2$ and $\alpha < 0$, then $\rho_{N,\alpha,\beta}$ weakly converges to the mixture of Dirac measures: $\frac{1}{2}\big(\delta_{(m^+(\frac{\beta-\alpha}{2}), -m^+(\frac{\beta-\alpha}{2}))} + \delta_{(-m^+(\frac{\beta-\alpha}{2}), m^+(\frac{\beta-\alpha}{2}))}\big)$.

Theorem 1.1 can be considered as a Law of Large Numbers for m. Central Limit Theorems for m were proved in [20].

They are also very useful for understanding the following reconstruction result, even though the authors in [2] choose a different approach. We will come back to this point when having described the reconstruction mechanism.

In the major part of their work [2] Berthet et al. consider the question, whether with a given number of observations n one can reconstruct S exactly, and, if so, how n relates to N. One of their main findings is

Theorem 1.2 (cf. [2, Corollary 4.6]). *If the parameters α and β satisfy $|\alpha| \le \beta$, $\alpha < \beta$, and $|\alpha| + \beta \ne 2$, then there exist positive constants C_1 and C_2 that depend on α and β such that there is an algorithm that recovers the block structure (S, S^c) exactly with probability $1 - \delta$ whenever*

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

where n denotes the number of observations.

There are two regimes of parameters excluded by Theorem 1.2. The first is $\alpha = \beta$. While one can still prove limit theorems in this case (see [15]), it is rather obvious that it is impossible to reconstruct S in this case: The interaction simply does not differentiate between spins in the same block and spins in different blocks.

Another obvious question left open by Theorem 1.2 is, what happens at the critical line $|\alpha| + \beta = 2$. The purpose of this note is to fill this gap. We will show:

Theorem 1.3. *If the parameters α and β satisfy $\alpha < \beta$, $\alpha \ne 0$, and $|\alpha| + \beta = 2$, then the following holds true.*

- *If $\alpha > 0$, there exists a positive constant C_3 such that there is an algorithm that recovers the block structure (S, S^c) exactly with probability $1 - \delta$ whenever*
$$n \ge C_3 N \log(N/\delta).$$

- *If $\alpha < 0$, there exists a positive constant C_4 such that there is an algorithm that recovers the block structure (S, S^c) exactly with probability $1 - \delta$ whenever*
$$n \ge C_4 \sqrt{N} \log(N/\delta).$$

Here, n denotes the number of observations.

Remark 1.4. *Note that the number n of necessary observations indicates that the phase one enters at the phase transition point $\alpha > 0$ and $\alpha + \beta = 2$ is of a different nature than the phase one enters at $\alpha < 0$ and $|\alpha| + \beta = 2$ – even though both points still belong to the high temperature regime according to Theorem 1.1. We exclude the case $\alpha = 0$, because it is already covered by Theorem 1.2.*

Remark 1.5. *Following the proof in [2] one can see that the constants C_1 and C_3 in the above Theorems 1.2 and 1.3 depend on α and β and explode when α tends to β. We will not elaborate on this point.*

The rest of this note is devoted to the proof of Theorem 1.3. To this end we will quickly recap the general reconstruction approach from [2] in Section 2. Section 3 will generalize a result on the critical fluctuations of $m_1 + m_2$ and $m_1 - m_2$ from [20]. In particular, we will treat the case of negative α which was omitted in [20]. These two ingredients will yield the proof of Theorem 1.3, which will be given in Section 4.

2. The strategy for block recovery

While part 2 of Theorem 1.2 could, in principle, be shown using a large deviations estimate, the first part needs some more sophisticated arguments. In this section we will recall how the approach proposed in [2] by Berthet et al. works. We will be a bit brief and especially refer the reader to [2] for proofs.

Given n observations $\sigma^{(1)}, \ldots, \sigma^{(n)}$ the log-likelihood function for S is given by

$$\mathcal{L}(S) = -n \log Z_{N,\alpha,\beta}(S) - \sum_{k=1}^{n} H_{N,\alpha,\beta,S}(\sigma^{(k)}).$$

Since it is easily seen that $Z_{N,\alpha,\beta}$, as a sum over all configurations σ, is independent of S, because we know its size, maximizing $\mathcal{L}(S)$ amounts to minimizing $\sum_{k=1}^{n} H_{N,\alpha,\beta,S}(\sigma^{(k)})$. On the other hand, taking the $N \times N$ matrix Q with elements $Q_{ij} = \frac{\beta}{N}$, if $i \sim j$ and $Q_{ij} = \frac{\alpha}{N}$, otherwise, one readily sees that $H_{N,\alpha,\beta,S}(\sigma) = -\frac{1}{2}\text{Tr}(\sigma\sigma^T Q)$ (where the upper index T indicates transposition). Note that Q depends on S.

Thus finding the maximum likelihood estimator for S amounts to maximizing $\frac{1}{2}\text{Tr}(\hat{\Sigma}Q)$ or $\text{Tr}(\hat{\Sigma}Q)$, where $\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n} \sigma^{(k)}\sigma^{(k)^T}$ is the empirical covariance matrix of the observations.

Here, we maximize over all matrices Q that induce a bisection of $\{1, \ldots N\}$ into sets S and S^c of equal size. More precisely, the $N \times N$ matrix $Q = (Q_{ij})_{1 \leq i,j \leq N}$ has to satisfy the following properties: Q is symmetric, $Q_{ij} \in \{\frac{\beta}{N}, \frac{\alpha}{N}\}$, the diagonal entries are equal to $\frac{\beta}{N}$, $Q_{ij} = Q_{jk} = \frac{\beta}{N}$ implies $Q_{ik} = \frac{\beta}{N}$ for all $i, j, k \in \{1, \ldots N\}$ and for each $i = 1, \ldots, N$ we have $|\{j : Q_{ij} = \frac{\beta}{N}\}| = \frac{N}{2}$.

Now, for each fixed σ the function $\text{Tr}(\sigma\sigma^T Q)$ is maximized by the same set S (associated to the matrix Q) no matter, what α and β are, as long as $\alpha < \beta$. The same holds true for n fixed observations $\sigma^{(1)}, \ldots, \sigma^{(n)}$ and $\text{Tr}(\hat{\Sigma}Q)$.

Indeed, it is an easy matter to check that

$$\frac{2n}{N}\mathrm{Tr}(\hat{\Sigma}Q) = \sum_{k=1}^{n} \alpha m_1(\sigma^{(k)})m_2(\sigma^{(k)}) + \frac{\beta}{2}((m_1(\sigma^{(k)}))^2 + (m_2(\sigma^{(k)}))^2).$$

As $\sigma^{(1)},\ldots,\sigma^{(n)}$ are fixed so is $\frac{2}{N}\sum_{i=1}^{N}\sigma_i^{(k)} =: c_k$. This means we have that

$$m_2(\sigma^{(k)}) = c_k - m_1(\sigma^{(k)})$$

for all k and constants c_k depending on k (of course), but not depending on the partitioning (S, S^c). Thus given our n observations $\sigma^{(1)},\ldots,\sigma^{(n)}$ we compute

$$\frac{2n}{N}\mathrm{Tr}(\hat{\Sigma}Q)$$
$$= \alpha\sum_{k=1}^{n}(m_1(\sigma^{(k)}))(c_k - m_1(\sigma^{(k)})) + \frac{\beta}{2}\sum_{k=1}^{n}(m_1(\sigma^{(k)}))^2 + (c_k - m_1(\sigma^{(k)}))^2).$$

At first glance the right hand side may appear rather tricky, because even though the observations $\sigma^{(1)},\ldots,\sigma^{(n)}$ are taken independently, the $m_1(\sigma^{(k)})$ are not, when S is the free variable. However, multiplying out the products (resp. the square) and rearranging the terms we see that

$$\frac{2n}{N}\mathrm{Tr}(\hat{\Sigma}Q) = K + (\beta - \alpha)\sum_{k=1}^{n}(m_1(\sigma^{(k)}))^2 - (\beta - \alpha)\sum_{k=1}^{n}c_k m_1(\sigma^{(k)})$$

for some constant K depending on β and $\sigma^{(1)},\ldots,\sigma^{(n)}$ but not on the choice of S. This reveals that the minimal and maximal points of $\frac{2n}{N}\mathrm{Tr}(\hat{\Sigma}Q)$ as a function of S do not depend on α and β as long as $\beta \neq \alpha$. Only, whether they are maxima or minima depends on whether $\alpha < \beta$ or $\alpha > \beta$. Hence, as long as we keep $\alpha < \beta$ we can choose any values for them we like.

We can therefore also set $\beta = N$ and $\alpha = -N$. This transforms our optimization problem into

$$\max_{R \in \mathcal{R}} \mathrm{Tr}(\hat{\Sigma}R) \quad \text{with} \quad \mathcal{R} = \{R = rr^T : r \in \{\pm 1\}^N : \sum_{i=1}^{N} r_i = 0\} \qquad (2)$$

(cf. (3.2) in [2]). Obviously each $R = rr^T \in \mathcal{R}$ again induces a bisection of $\{1,\ldots,N\}$ into two sets S and S^c of equal size, where $i \sim j$, if $r_i = r_j$. Moreover, there is a one-to-one correspondence between the set of valid matrices Q and $\mathcal{R}$; e.g. given some matrix Q with the aforementioned properties, the corresponding matrix $R = rr^T \in \mathcal{R}$ is given by $r_i = 1$, if $Q_{1i} = \frac{\beta}{N}$ and $r_i = -1$ otherwise. Thus each $R \in \mathcal{R}$ is an estimator for the unknown blocks.

In [2] the authors now proceed in two steps. First the σ are centered, i.e. $\sigma \in \{\pm 1\}^N$ is replaced by $\overline{\sigma} := \Pi\sigma$ with $\Pi := \mathrm{Id}_N - \frac{1}{N}\mathbb{1}_N$, where $\mathbb{1}_N$ is the $N \times N$ matrix with all elements equal to 1. Since for all $R \in \mathcal{R}$, we have that

$\mathrm{Tr}[\hat{\Sigma}R] = \mathrm{Tr}[\hat{\Gamma}R]$, where $\hat{\Gamma} = \Pi\hat{\Sigma}\Pi$ the likelihood function remains the same over $\mathcal{R}$ when we replace $\hat{\Sigma}$ by $\hat{\Gamma}$. The decisive step in [2] is then to embed the optimization problem (2) into a larger class of optimization problems. Hence, instead of solving (2) the authors look for solutions of

$$\max_{R \in \mathcal{E}^+} \mathrm{Tr}(\hat{\Gamma}R) \tag{3}$$

with

$\mathcal{E}^+ := \{R : R$ is a positive semidefinite, symmetric $N \times N$ matrix
$\qquad$ with 1 on the diagonal$\}$.

The question is, of course, when a solution of (3) also provides a solution to (2). This is, where the authors in [2] spend a considerable amount of work to show that:

Theorem 2.1 (cf. [2, Theorem 3.3]). *The semidefinite programming problem* (3) *has a unique maximum at* $R^* \in \mathcal{E}^+$ *with probability* $1 - \delta$ *whenever*

$$n > C_{\alpha,\beta} \frac{\log(4N/\delta)}{Z - Z'}(1 + o_p(1)).$$

Here $Z := \mathbb{E}(\sigma_i\sigma_j)$, *if* $i \sim j$, $Z' := \mathbb{E}(\sigma_i\sigma_j)$, *if* $i \not\sim j$, *and* $C_{\alpha,\beta}$ *is a constant depending on* α *and* β.

Moreover, this unique solution of (3) *is of the form* $R^* = r_S r_S^T$, *where* $r_S := \mathbb{1}_S - \mathbb{1}_{S^c}$ *for a set* S *with cardinality* $\frac{N}{2}$. *Thus* $R^* \in \mathcal{R}$. *In particular, the semidefinite programming solution, if it exists, recovers exactly the block structure* (S, S^c).

3. A limit theorem

Theorem 2.1 obviously asks for an estimate of $Z - Z'$. As a matter of fact, in [2] the authors show by a comparison of the distribution of (m_1, m_2) with a Gaussian distribution that for $\alpha \leq 0$ and $|\alpha| + \beta > 2$, the difference $Z - Z'$ is of constant order in N, while it is of order $\frac{1}{N}$, in all other cases, whenever $|\alpha| + \beta \neq 2$. Indeed, for $\alpha > 0$ and $\alpha + \beta < 2$ this also easily follows from the Central Limit Theorem 1.2 for the vector m in [20] and it would follow from corresponding Central Limit Theorems also in the other cases when $|\alpha| + \beta \neq 2$, if such were proven (they are quite likely true, because similar Central Limit Theorems hold for the vector $\frac{1}{N}\sum_{i=1}^{N}\sigma_i$ in the Curie-Weiss model, see e.g. [14]).

For $|\alpha| + \beta = 2$ there is no such estimate for $Z - Z'$ in [2]. Indeed, in view of Theorem 3.1 below the main tool in this reference does not work, because at least for $\alpha > 0$, and $\alpha + \beta = 2$ the vector $\sqrt{N}(m_1, m_2)$ is not asymptotically Gaussian. Moreover, considering Theorem 6 and Remark 7 in [16] one may wonder, whether an exact reconstruction of (S, S^c) on the basis of the correlations between the

spins is possible at all. There, the authors show that for $\alpha > 0$ and $\alpha + \beta = 2$ asymptotically:

$$\mathbb{E}(\sigma_i \sigma_j) = \sqrt{\frac{12}{N}} \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})} \qquad \text{for } i \neq j$$

independent of whether the sites i and j belong to the same block, or to different blocks. Hence, asymptotically the correlations $\mathbb{E}(\sigma_i \sigma_j)$ at $\alpha + \beta = 2$ do not depend on whether $i \sim j$ or $i \not\sim j$, nor do they depend on α and β.

In the sequel we generalize a theorem from [20] to analyze the fluctuations of (m_1, m_2) for all $|\alpha| < \beta$ with $|\alpha| + \beta = 2$. This well help us to prove Theorem 1.3. Indeed, at $|\alpha| + \beta = 2$ the following holds:

Theorem 3.1. *For the parameters of the block spin Ising model α and β assume that $|\alpha| < \beta$ and that $|\alpha| + \beta = 2$.*

(a) *Then, if $\alpha > 0$ on a scale $\sqrt{N}/2$ the difference between m_1 and m_2, i.e. $\tilde{m}_1 - \tilde{m}_2 := \frac{\sqrt{N}}{2}(m_1 - m_2)$, is asymptotically Gaussian with mean 0 and variance $\frac{2}{2-(\beta-\alpha)}$.*

If $\alpha < 0$, $\frac{N^{\frac{1}{4}}}{2}(m_1 - m_2)$ converges in distribution to a probability measure ρ on $\mathbb{R}$, which is absolutely continuous with Lebesgue-density

$$g(x) = \exp\left(-\frac{1}{12}x^4\right)/K,$$

where K is a normalizing constant to make ρ a probability measure.

(b) *For the overall spin $m_1 + m_2$, we have the following: If $\alpha > 0$ there is a non-standard Central Limit Theorem, i.e. $\frac{N^{\frac{1}{4}}}{2}(m_1 + m_2)$ converges in distribution to the probability measure ρ given in part (a) of the theorem. If $\alpha < 0$, there is a Gaussian Central Limit Theorem, i.e. $\frac{\sqrt{N}}{2}(m_1 + m_2)$ converges in distribution to a Gaussian with mean zero and variance $-\frac{1}{\alpha}$.*

Remark 3.2. *Theorem 3.1 also shows the difference between $|\alpha| + \beta = 2, \alpha > 0$ and $|\alpha| + \beta = 2, \alpha < 0$. Indeed, according to Theorem 1.1 for $|\alpha| + \beta$ slightly larger than 2, the block magnetizations m_1 and m_2 are close together when $\alpha > 0$, while they move apart for $\alpha < 0$.*

Moreover, notice that the probability measure ρ in Theorem 3.1 is the same as the limiting distribution of the appropriately rescaled magnetization in the Curie-Weiss model, see [8, Theorem V.9.5]

Proof. The proof of Theorem 3.1 makes use of the so-called Hubbard-Stratonovich transform.

Introduce the random variables

$$w_1 := w_1(\sigma) := \frac{1}{N} \sum_{i=1}^{N} \sigma_i = \frac{1}{N}\left(\sum_{i \in S} \sigma_i + \sum_{i \notin S} \sigma_i\right)$$

and

$$w_2 := w_2(\sigma) := \frac{1}{N}\left(\sum_{i \in S} \sigma_i - \sum_{i \notin S} \sigma_i\right)$$

together with the vectors $\tilde{w} = (\tilde{w}_1, \tilde{w}_2)$ and $\hat{w} = (\hat{w}_1, \hat{w}_2)$ consisting of the components

$$\tilde{w}_1 := \sqrt{N}w_1 \qquad \text{and} \qquad \tilde{w}_2 := N^{1/4}w_2,$$

as well as

$$\hat{w}_1 := N^{1/4}w_1 \qquad \text{and} \qquad \hat{w}_2 := \sqrt{N}w_2.$$

Note that

$$\begin{aligned}
H_{N,\alpha,\beta,S}(\sigma) &= -\frac{N}{2}\left(2\alpha\frac{1}{4}m_1 m_2 + \beta\frac{1}{4}m_1^2 + \beta\frac{1}{4}m_2^2\right) \\
&= -\frac{N}{8}\left(2\alpha(w_1+w_2)(w_1-w_2) + \beta\left((w_1+w_2)^2 + (w_1-w_2)^2\right)\right) \\
&= -\frac{N}{4}(\beta(w_1{}^2 + w_2{}^2) + \alpha(w_1{}^2 - w_2{}^2)) \\
&= -\frac{1}{4}(\sqrt{N}(\alpha+\beta)\hat{w}_1{}^2 + (\beta-\alpha)\hat{w}_2{}^2) \\
&= -\frac{1}{4}(\sqrt{N}\kappa_\alpha \hat{w}_1{}^2 + \eta_\alpha \hat{w}_2{}^2),
\end{aligned}$$

where we have set

$$\kappa_\alpha := \alpha + \beta = \begin{cases} 2 & \text{if } \alpha > 0 \\ 2(1+\alpha) & \text{if } \alpha < 0 \end{cases}$$

and

$$\eta_\alpha := \beta - \alpha = \begin{cases} 2 & \text{if } \alpha < 0 \\ 2(1-\alpha) & \text{if } \alpha > 0 \end{cases}.$$

Note that $\kappa_\alpha > 0$ as well as $\eta_\alpha > 0$, if $|\alpha| < \beta$ and $|\alpha| + \beta = 2$.

Similarly,

$$H_{N,\alpha,\beta,S}(\sigma) = -\frac{1}{4}(\kappa_\alpha \tilde{w}_1{}^2 + \sqrt{N}\eta_\alpha \tilde{w}_2{}^2).$$

We now begin with the case of $\alpha < 0$. Then $\kappa_\alpha < 2$, which will make a difference, as we will see.

Our strategy can be summarized as "linearizing" the Hamiltonian by tilting it with a suitable Gaussian random variable (a similar technique was used e.g. in [13]). To this end, let $\mathcal{N}(0, C)$ be a 2-dimensional Gaussian distribution with expectation 0 and covariance matrix C,

$$C = \begin{pmatrix} \frac{2}{\kappa_\alpha} & 0 \\ 0 & \frac{1}{\sqrt{N}} \end{pmatrix}.$$

Denote by $\overline{\rho}_{N,\alpha,\beta} := \mu_{N,\alpha,\beta} \circ (\tilde{w})^{-1}$ the distribution of $\tilde{w}$ under the Gibbs measure and by $\chi_{N,\alpha,\beta} := \overline{\rho}_{N,\alpha,\beta} * \mathcal{N}(0, C)$. Let us compute the Lebesgue density of $\chi_{N,\alpha,\beta}$: Let A be a measurable subset of $\mathbb{R}^2$. Then with

$$K_1 := \frac{1}{2\pi\sqrt{\det C}}, \quad K_2 := \frac{1}{2\pi\sqrt{\det C}Z_{N,\alpha,\beta}} \quad \text{and} \quad K_3 := K_2 2^N$$

we obtain

$$\chi_{N,\alpha,\beta}(A) = \sum_{\sigma \in \{-1,1\}^N} \mathcal{N}(0,C)(A - \tilde{w})\mu_{N,\alpha,\beta}(\sigma)$$

$$= K_1 \sum_{\sigma \in \{-1,1\}^N} \int_{A-\tilde{w}} \exp\left(-\frac{1}{2}\left(\frac{\kappa_\alpha}{2}x^2 + \sqrt{N}y^2\right)\right)\mu_{N,\alpha,\beta}(\sigma)dxdy$$

$$= K_1 \sum_{\sigma \in \{-1,1\}^N} \int_A \exp\left(-\frac{1}{2}\left(\frac{\kappa_\alpha}{2}(x - \tilde{w}_1)^2 + \sqrt{N}(y - \tilde{w}_2)^2\right)\right)\mu_{N,\alpha,\beta}(\sigma)dxdy$$

$$= K_2 \sum_{\sigma \in \{-1,1\}^N} \int_A \exp\left(-\frac{1}{2}\left(\frac{\kappa_\alpha}{2}(x - \tilde{w}_1)^2 + \sqrt{N}(y - \tilde{w}_2)^2\right)\right.$$
$$\left. + \frac{1}{4}\left(\kappa_\alpha\tilde{w}_1^2 + 2\sqrt{N}\tilde{w}_2^2\right)\right)dxdy$$

because $\eta_\alpha = 2$ when $\alpha < 0$. Thus

$$\chi_{N,\alpha,\beta}(A)$$

$$= K_2 \sum_{\sigma \in \{-1,1\}^N} \int_A \exp\left(-\frac{1}{4}\kappa_\alpha x^2 - \frac{\sqrt{N}}{2}y^2 + \frac{1}{2}\kappa_\alpha x\tilde{w}_1 + \sqrt{N}y\tilde{w}_2\right)dxdy$$

$$= K_3 \int_A \exp\left(-\frac{1}{4}\kappa_\alpha x^2 - \frac{\sqrt{N}}{2}y^2\right)\exp\left(\frac{N}{2}\log\cosh\left(\frac{1}{\sqrt{N}}\frac{\kappa_\alpha}{2}x + N^{-\frac{1}{4}}y\right)\right.$$

$$\left. + \frac{N}{2}\log\cosh\left(\frac{\kappa_\alpha}{2\sqrt{N}}x - \frac{1}{N^{\frac{1}{4}}}y\right)\right)dxdy. \tag{4}$$

Note that in the above expression, both the integral term and the constant K_3 depend on N. However, it suffices to consider the convergence of the integral, because once the convergence of the integral is shown for all measurable subsets $A \subset \mathbb{R}^2$ and in particular for $A = \mathbb{R}^2$, this implies the convergence of K_3 to a constant.

Introduce as short-hand for the negative exponent in (4):

$$\Phi(x,y) := \frac{1}{4}\kappa_\alpha x^2 + \frac{\sqrt{N}}{2}y^2$$
$$- \frac{N}{2}\left(\log\cosh\left(\frac{\kappa_\alpha}{2\sqrt{N}}x + \frac{1}{N^{\frac{1}{4}}}y\right) + \log\cosh\left(\frac{\kappa_\alpha}{2\sqrt{N}}x - \frac{1}{N^{\frac{1}{4}}}y\right)\right). \tag{5}$$

Then by Taylor expansion of $\log\cosh(z) = \frac{1}{2}z^2 - \frac{1}{12}z^4 + \mathcal{O}(z^6)$ up to fourth order we see that the y^2-terms cancel, and so do the xy-terms and we arrive at

$$\Phi(x,y) = \frac{x^2}{2}\left(\frac{\kappa_\alpha}{2} - \frac{\kappa_\alpha^2}{4}\right) + \frac{y^4}{12} + \mathcal{O}(N^{-\frac{1}{2}}),$$

where the constant in the $\mathcal{O}(N^{-\frac{1}{2}})$-term depends on x and y. Therefore we obtain for the density of the convolution

$$\chi_{N,\alpha,\beta}(A) = K_3 \int_A \exp\left[-\frac{1}{2}x^2\left(\frac{\kappa_\alpha}{2} - \frac{\kappa_\alpha^2}{4}\right) - \frac{1}{12}y^4 + \mathcal{O}(N^{-\frac{1}{2}})\right] dx\,dy$$

and the convergence in the $\mathcal{O}(N^{-\frac{1}{2}})$-term is uniform on compact subsets of $\mathbb{R}^2$. From here it becomes immediately clear, that the cases $\kappa_\alpha = 2$ and $\eta_\alpha = 2$ have to be treated separately.

We next show how to extend the convergence to non-compact sets. This is done in the spirit of [20] or [13].

To this end for a Borel set $A \subset \mathbb{R}^2$ we split the integral

$$\int_A \exp\left(-\Phi(x,y)\right)dxdy = \int_{A \cap B(0,R)} \exp\left(-\Phi(x,y)\right)dxdy \qquad (6)$$

$$+ \int_{A \cap B(0,R)^c \cap B_{r,N}} \exp\left(-\Phi(x,y)\right)dxdy + \int_{A \cap B_{r,N}^c} \exp\left(-\Phi(x,y)\right)dxdy.$$

Here, for any number $R > 0$ we denote by $B(0,R)$ the (2-dimensional) ball centered in 0 with radius R and by $B_{r,N}$ we denote the box

$$B_{r,N} := (-r\sqrt{N}, r\sqrt{N}) \times (-rN^{1/4}, rN^{1/4}). \qquad (7)$$

Later R will be sent to ∞ and r will be taken very small. Note that we will however first take the limit $N \to \infty$ and then $R \to \infty$. The summands in (6) will be called inner region, intermediate region, and outer region, respectively. They will be treated separately.

As noted earlier for fixed $R > 0$

$$\lim_{N \to \infty} \int_{A \cap B(0,R)} \exp\left(-\Phi(x,y)\right)dxdy$$

$$= \int_{A \cap B(0,R)} \exp\left[-\frac{1}{2}x^2\left(\frac{\kappa_\alpha}{2} - \frac{\kappa\alpha^2}{4}\right) - \frac{1}{12}y^4\right] dx\,dy.$$

This already finishes the treatment of the inner region.

For the outer region $A \cap B_{r,N}^c$, which we will further split up in two regions, we change scales and we write

$$\Phi(x,y) = N\tilde{\Phi}\left(\frac{x}{\sqrt{N}}, \frac{y}{N^{\frac{1}{4}}}\right),$$

where obviously

$$\tilde{\Phi}(x,y) := \frac{1}{4}\kappa_\alpha x^2 + \frac{1}{2}y^2 - \frac{1}{2}\log\cosh\left(\frac{\kappa_\alpha}{2}x + y\right) - \frac{1}{2}\log\cosh\left(\frac{\kappa_\alpha}{2}x - y\right).$$

We denote the free energy of the Curie-Weiss model at inverse temperature β by $f_{CW}(\beta)$. Note that $f_{CW}(\beta)$ is non-negative and given by

$$f_{CW}(\beta) = -\frac{\beta}{2}(m^+(\beta)^2 + \log(\cosh(\beta m^+(\beta)))) = \max_t \left[-\frac{1}{2\beta}t^2 + \log\cosh(t) \right].$$

Then,

$$\log\cosh(x) \leq \frac{1}{4}x^2 + \max_t \left[-\frac{1}{4}t^2 + \log\cosh(t) \right] = \frac{1}{4}x^2 + f_{CW}(2).$$

Hence,

$$\tilde{\Phi}(x,y) = \frac{1}{4}\kappa_\alpha x^2 + \frac{1}{2}y^2 - \frac{1}{2}\log\cosh\left(\frac{\kappa_\alpha}{2}x + y\right) - \frac{1}{2}\log\cosh\left(\frac{\kappa_\alpha}{2}x - y\right)$$

$$\geq \frac{1}{4}\kappa_\alpha x^2 + \frac{1}{2}y^2 - \frac{1}{2}\frac{(\frac{\kappa_\alpha x}{2} + y)^2}{4} - \frac{1}{2}\frac{(\frac{\kappa_\alpha x}{2} - y)^2}{4} - f_{CW}(2)$$

$$= x^2\left(\frac{1}{4}\kappa_\alpha - \frac{1}{16}\kappa_\alpha^2\right) + \frac{y^2}{4} - f_{CW}(2).$$

Setting

$$r_0 := 4\sqrt{\frac{f_{CW}(2)}{\min(\kappa_\alpha - \frac{\kappa_\alpha^2}{4}, 1)}},$$

and, in accordance with (7),

$$B_{r_0,N} := (-r_0\sqrt{N}, r_0\sqrt{N}) \times (-r_0 N^{1/4}, r_0 N^{1/4}),$$

we assume that r is sufficiently small such that $r < r_0$, i.e. $B_{r,N} \subset B_{r_0,N}$, and we split up the outer region

$$A \cap B_{r,N}^c = (A \cap B_{r_0,N}^c) \cup (A \cap (B_{r_0,N} \setminus B_{r,N})).$$

Then, if $(x,y) \notin B_{r_0,N}$ we have $\frac{x^2}{N} + \frac{y^2}{\sqrt{N}} \geq r_0^2$ and hence we know that

$$-\Phi(x,y)$$
$$= -N\tilde{\Phi}\left(\frac{x}{\sqrt{N}}, \frac{y}{N^{\frac{1}{4}}}\right) \leq -N\left(\frac{x^2}{N}\left(\frac{1}{4}\kappa_\alpha - \frac{1}{16}\kappa_\alpha^2\right) + \frac{y^2}{4\sqrt{N}}\right) + Nf_{CW}(2)$$
$$= -\frac{N}{16}\left(\frac{x^2}{N}\left(\kappa_\alpha - \frac{1}{4}\kappa_\alpha^2\right) + \frac{y^2}{\sqrt{N}}\right) - \frac{3N}{4}\left(\frac{x^2}{N}\left(\frac{1}{4}\kappa_\alpha - \frac{1}{16}\kappa_\alpha^2\right) + \frac{y^2}{4\sqrt{N}}\right)$$
$$\quad + Nf_{CW}(2)$$
$$\leq -\frac{3N}{4}\left(\frac{x^2}{N}\left(\frac{1}{4}\kappa_\alpha - \frac{1}{16}\kappa_\alpha^2\right) + \frac{y^2}{4\sqrt{N}}\right)$$
$$\leq -NK_\alpha\left(\frac{x^2}{N} + \frac{y^2}{\sqrt{N}}\right)$$

for some appropriate constant $K_\alpha > 0$ depending on α.

Thus

$$\int_{A \cap B^c_{r_0,N}} \exp\left(-\Phi(x,y)\right) dxdy = \int_{A \cap B^c_{r_0,N}} \exp\left(-N\tilde{\Phi}\left(\frac{x}{\sqrt{N}}, \frac{y}{N^{\frac{1}{4}}}\right)\right) dxdy$$

$$\leq \int_{A \cap B^c_{r_0,N}} \exp\left(-NK_\alpha\left(\frac{x^2}{N} + \frac{y^2}{\sqrt{N}}\right)\right) dxdy$$

$$\leq e^{-NK_\alpha r_0^2/2} \int_{\mathbb{R}^2} \exp\left(-\frac{NK_\alpha}{2}\left(\frac{x^2}{N} + \frac{y^2}{\sqrt{N}}\right)\right) dxdy$$

$$\leq K'_\alpha e^{-NK_\alpha r_0^2/2} \tag{8}$$

for another constant $K'_\alpha > 0$.

Next we want to extend (8) to arbitrarily small $r \in (0, r_0)$. To this end we claim

$$\inf_{x \in B_{r_0,N} \setminus B_{r,N}} \tilde{\Phi}\left(\frac{x}{\sqrt{N}}, \frac{y}{N^{\frac{1}{4}}}\right) > 0, \quad \text{for all } 0 < r < r_0. \tag{9}$$

By rescaling it suffices to consider $\tilde{\Phi}$ on the set $(-r_0, r_0) \times (-r_0, r_0) \setminus (-r, r) \times (-r, r)$.

Obviously, $\tilde{\Phi}(0, 0) = 0$. We want to show that $(0, 0)$ is the unique minimum of $\tilde{\Phi}$.

Indeed, $\tilde{\Phi}$ gets minimal if $\nabla\tilde{\Phi} = 0$ and the latter is given by

$$\nabla\tilde{\Phi}(x, y) = \begin{pmatrix} \frac{1}{2}\kappa_\alpha x - \frac{\kappa_\alpha}{4}\tanh(\frac{\kappa_\alpha x}{2} + y) - \frac{\kappa_\alpha}{4}\tanh(\frac{\kappa_\alpha x}{2} - y) \\ y - \frac{1}{2}\tanh(\frac{\kappa_\alpha x}{2} + y) + \frac{1}{2}\tanh(\frac{\kappa_\alpha x}{2} - y) \end{pmatrix}.$$

Solving $\nabla\tilde{\Phi} = 0$ thus is equivalent to solving

$$G(x, y) := \begin{pmatrix} \frac{1}{2}\tanh(\frac{\kappa_\alpha x}{2} + y) + \frac{1}{2}\tanh(\frac{\kappa_\alpha x}{2} - y) \\ \frac{1}{2}\tanh(\frac{\kappa_\alpha x}{2} + y) - \frac{1}{2}\tanh(\frac{\kappa_\alpha x}{2} - y) \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Obviously, $(0, 0)$ is a solution. To see that $(0, 0)$ is indeed the only solution, we assume that $G(x, y) = (x, y)$ for some (x, y). Note that $x = 0$ immediately implies $y = 0$. If $y = 0$, we have $\tanh(\frac{\kappa_\alpha}{2}x) = x$, which has only the solution $x = 0$, since $\frac{\kappa_\alpha}{2} \leq 1$.

Hence, we can assume that $x \neq 0$ and $y \neq 0$. By the symmetry of tanh, the first equation induced by $G(x, y) = (x, y)$ reads

$$\frac{\tanh(y + \frac{\kappa_\alpha x}{2}) - \tanh(y - \frac{\kappa_\alpha x}{2})}{\kappa_\alpha x} = \frac{2}{\kappa_\alpha}.$$

With $g(z) := (\tanh(x))' = 1 - (\tanh(z))^2$ and the mean value theorem we have

$$\frac{\tanh(y + \frac{\kappa_\alpha x}{2}) - \tanh(y - \frac{\kappa_\alpha x}{2})}{\kappa_\alpha x} = g(x_0) \stackrel{!}{=} \frac{2}{\kappa_\alpha} = \frac{1}{1 + \alpha} > 1, \tag{10}$$

for some x_0 between $y + \frac{\kappa_\alpha x}{2}$ and $y - \frac{\kappa_\alpha x}{2}$ (recall $\alpha < 0$ for the last inequality). Since $0 \leq g(z) \leq 1$ for all $z \in \mathbb{R}$ the in equality in (10) can never be true, and

there is no solution to $G(x, y) = (x, y)$ with $x \neq 0, y \neq 0$. This, in turn implies that on $\mathbb{R}^2$ the function $\tilde{\Phi}$ has a unique minimum at $(0, 0)$. This proves exactly the claim in (9) or in other words for all $r > 0$ there is a constant $\xi(r, r_0)$ such that

$$\tilde{\Phi}\left(\frac{x}{\sqrt{N}}, \frac{y}{N^{1/4}}\right) > \xi(r, r_0)$$

for all $(x, y) \in B_{r_0, N} \setminus B_{r, N}$. From here we get that

$$\int_{A \cap (B_{r_0, N} \setminus B_{r, N})} \exp\left(-\Phi(x, y)\right) dx dy \leq \int_{B_{r_0, N}} e^{-N\xi(r, r_0)} dx dy \leq 2r_0^2 N e^{-N\xi(r, r_0)}$$
$$\leq e^{-N\xi(r, r_0)/2} \tag{11}$$

for N sufficiently large. By (8) and (11) we have

$$\lim_{N \to \infty} \int_{A \cap B_{r, N}^c} \exp\left(-\Phi(x, y)\right) dx dy = 0$$

for $r > 0$ sufficiently small.

Now we turn to the intermediate region $A \cap B(0, R)^c \cap B_{r, N}$: Recalling the definition of $\tilde{\Phi}$ we want to expand the term $\log \cosh\left(\frac{\kappa_\alpha}{2\sqrt{N}} x + \frac{1}{N^{\frac{1}{4}}} y\right)$ resp. $\log \cosh\left(\frac{\kappa_\alpha}{2\sqrt{N}} x - \frac{1}{N^{\frac{1}{4}}} y\right)$ for $x, y \in B_{r, N} \setminus B(0, R)$, where we recall the definition of $B_{r, N}$ given in (7). Note that $x, y \in B_{r, N}$ implies $\frac{\kappa_\alpha}{2\sqrt{N}} x \pm \frac{1}{N^{\frac{1}{4}}} y \in (-2r, 2r)$.

For $z \in (-2r, 2r)$ we use the Taylor approximation and estimate the remainder using the Lagrange form

$$\left|\log \cosh(z) - \frac{z^2}{2} + \frac{1}{12} z^4\right| \leq Cr z^6 \text{ resp. } -\log \cosh(z) \geq -\frac{z^2}{2} + \frac{1}{12} z^4 - Cr z^6$$

for some positive constant $C > 0$. Applying this to $\tilde{\Phi}$ and $x, y \in B_{r, N} \setminus B(0, R)$ and using that the terms with odd powers of x (and y) cancel, this gives

$$N\tilde{\Phi}\left(\frac{x}{\sqrt{N}}, \frac{y}{N^{1/4}}\right) \geq \frac{x^2 \kappa_\alpha}{4} + \frac{\sqrt{N} y^2}{2} - \frac{N}{4}\left(\frac{\kappa_\alpha x}{2\sqrt{N}} + \frac{y}{N^{1/4}}\right)^2$$
$$- \frac{N}{4}\left(\frac{\kappa_\alpha x}{2\sqrt{N}} - \frac{y}{N^{1/4}}\right)^2 + \frac{N}{24}\left[\left(\frac{\kappa_\alpha x}{2\sqrt{N}} + \frac{y}{N^{1/4}}\right)^4 + \left(\frac{\kappa_\alpha x}{2\sqrt{N}} - \frac{y}{N^{1/4}}\right)^4\right]$$
$$- Cr\frac{N}{2}\left[\left(\frac{\kappa_\alpha x}{2\sqrt{N}} + \frac{y}{N^{1/4}}\right)^6 + \left(\frac{\kappa_\alpha x}{2\sqrt{N}} - \frac{y}{N^{1/4}}\right)^6\right]$$
$$\geq \frac{x^2}{2}\left(\frac{\kappa_\alpha}{2} - \left(\frac{\kappa_\alpha}{2}\right)^2\right) + \frac{y^4}{12} + \frac{N}{2}\frac{\kappa_\alpha^2 x^2}{4N}\frac{y^2}{\sqrt{N}}$$
$$- CrN\left[\frac{\kappa_\alpha^6 x^6}{64N^3} + 15\frac{\kappa_\alpha^4 x^4}{16N^2}\frac{y^2}{\sqrt{N}} + 15\frac{\kappa_\alpha^2 x^2}{4N}\frac{y^4}{N} + \frac{y^6}{N^{3/2}}\right]$$
$$\geq \frac{x^2}{2}\left[\frac{\kappa_\alpha}{2} - \left(\frac{\kappa_\alpha}{2}\right)^2\right] + \frac{y^4}{12} - CrN\left[\frac{\kappa_\alpha^6 x^6}{64N^3} + \frac{15\kappa_\alpha^4 x^4 y^2}{16N^2\sqrt{N}} + \frac{15\kappa_\alpha^2 x^2 y^4}{4N^2} + \frac{y^6}{N^{3/2}}\right].$$

Now note that for $x, y \in B_{r,N}$ we have

$$CrN \frac{\kappa_\alpha^6 x^6}{64N^3} \leq C \frac{1}{64} \kappa_\alpha^6 x^2 r^5, \tag{12}$$

$$15 CrN \frac{\kappa_\alpha^4 x^4}{16N^2} \frac{y^2}{\sqrt{N}} \leq C \frac{15}{16} \kappa_\alpha^4 x^2 r^5, \tag{13}$$

$$15 CrN \frac{\kappa_\alpha^2 x^2}{4N} \frac{y^4}{N} \leq C \frac{15}{4} \kappa_\alpha^2 x^2 r^5, \tag{14}$$

and

$$CrN \frac{y^6}{N^{3/2}} \leq C y^4 r^3. \tag{15}$$

Note that the coefficients of x^2 resp. of y^4 on the right hand sides of equations (12)–(15) become arbitrarily small, if we take r small enough.

On the set $A_{R,N}^r := A \cap B(0, R)^c \cap B_{r,N}$ we can therefore bound

$$N\tilde{\Phi}\left(\frac{x}{\sqrt{N}}, \frac{y}{N^{1/4}}\right)$$
$$\geq \frac{x^2}{2}\left(\frac{\kappa_\alpha}{2} - \left(\frac{\kappa_\alpha}{2}\right)^2 - C\frac{1}{32}\kappa_\alpha^6 r^5 - C\frac{15}{8}\kappa_\alpha^4 r^5 - \frac{15}{2}C\kappa_\alpha^2 r^5\right) + \frac{y^4}{12}(1 - 12Cr^3).$$

Since $\kappa_\alpha < 2$, there exists some $\varepsilon > 0$ such that for r sufficiently small we have

$$\frac{\kappa_\alpha}{2} - \left(\frac{\kappa_\alpha}{2}\right)^2 - C\frac{1}{32}\kappa_\alpha^6 r^5 - C\frac{15}{8}\kappa_\alpha^4 r^5 - \frac{15}{2}C\kappa_\alpha^2 r^5 > \varepsilon$$

as well as

$$1 - 12Cr^3 > \varepsilon.$$

Hence for such a small value of r and $N \in \mathbb{N}$ we get

$$\int_{A_{R,N}^r} \exp\left(-\Phi(x, y)\right) dx\, dy \leq \int_{A_{R,N}^r} \exp\left(-\frac{\varepsilon}{2}x^2 - \frac{\varepsilon}{12}y^4\right) dx\, dy. \tag{16}$$

Note that the integrand on the left hand side of (16) implicitly depends on N while the integrand on the right hand side does not depend on N and is integrable over $\mathbb{R}^2$, i.e. the estimate in (16) carries over if we take $N \to \infty$ on both sides

$$\lim_{N\to\infty} \int_{A_{R,N}^r} \exp\left(-\Phi(x, y)\right) dx\, dy \leq \int_{A\cap B(0,R)^c} \exp\left(-\frac{\varepsilon}{2}x^2 - \frac{\varepsilon}{12}y^4\right) dx\, dy. \tag{17}$$

Further, as again the integrand on the right hand side of (17) is integrable over all of $\mathbb{R}^2$, the right hand side of (17) vanishes for $R \to \infty$, i.e.

$$\lim_{R\to\infty} \lim_{N\to\infty} \int_{A_{R,N}^r} \exp\left(-\Phi(x, y)\right) dx\, dy = 0.$$

Combining the three considerations above shows that the only contribution to $\chi_{N,\alpha,\beta}$ comes from the inner region

$$\lim_{R\to\infty} \lim_{N\to\infty} \int_A \exp\left(-\Phi(x,y)\right) dx dy = \int_A \exp\left[-\frac{x^2}{2}\left(\frac{\kappa_\alpha}{2} - \frac{\kappa_\alpha^2}{4}\right) - \frac{1}{12}y^4\right] dx\, dy.$$

Thus in particular, the distribution of $\tilde{w}_2 = \frac{N^{1/4}}{2}(m_1 - m_2)$ convoluted with an independent Gaussian distribution with mean 0 and variance $\frac{1}{\sqrt{N}}$ converges to a probability measure ρ on $\mathbb{R}$ with density proportional to $\exp(-\frac{1}{12}x^4)$. Since the $\mathcal{N}(0, \frac{1}{\sqrt{N}})$-distribution converges to 0, as $N \to \infty$, this implies that $\frac{N^{1/4}}{2}(m_1 - m_2)$ converges to ρ in distribution, as well. Moreover, the above calculation yields that the distribution of $\tilde{w}_1 = \frac{\sqrt{N}}{2}(m_1 + m_2)$ convoluted with an independent Gaussian $\mathcal{N}(0, \frac{2}{\kappa_\alpha})$ converges to a Gaussian variable with mean zero and variance $\left(\frac{\kappa_\alpha}{2} - \frac{\kappa_\alpha^2}{4}\right)^{-1}$. In terms of characteristic functions this means

$$\lim_{n\to\infty} \mathbb{E}(e^{it\tilde{w}_1}) e^{-\frac{t^2}{2}\frac{2}{\kappa_\alpha}} = e^{-\frac{t^2}{2}\left(\frac{\kappa_\alpha}{2} - \frac{\kappa_\alpha^2}{4}\right)^{-1}}$$

resp.

$$\lim_{n\to\infty} \mathbb{E}(e^{it\tilde{w}_1}) = e^{-\frac{t^2}{2}\left(\frac{1}{\frac{\kappa_\alpha}{2} - \frac{\kappa_\alpha^2}{4}} - \frac{2}{\kappa_\alpha}\right)} = e^{-\frac{t^2}{2}\left(\frac{1}{1 - \frac{\kappa_\alpha}{2}}\right)} = e^{-\frac{t^2}{2}\left(-\frac{1}{\alpha}\right)}.$$

Hence, $\frac{\sqrt{N}}{2}(m_1 + m_2)$ converges in distribution to a Gaussian distribution with mean zero and variance $-\frac{1}{\alpha}$.

To analyze the second situation, i.e. $\alpha > 0$ and $\alpha + \beta = 2$, we now convolute the distribution of $\hat{w}$ with respect to $\mu_{N,\alpha,\beta,S}$ with a 2-dimensional normal distribution $\mathcal{N}(0, C)$, where this time

$$C := \begin{pmatrix} \frac{1}{\sqrt{N}} & 0 \\ 0 & \frac{2}{\beta - \alpha} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{N}} & 0 \\ 0 & \frac{2}{\eta_\alpha} \end{pmatrix}$$

(recall that $\beta - \alpha > 0$). Calling this convolution $\hat{\chi}_{N,\alpha,\beta}$, i.e. setting $\hat{\chi}_{N,\alpha,\beta} := \mu_{N,\alpha,\beta}(\hat{w})^{-1} * \mathcal{N}(0, C)$, we can compute its density as above (recall that $\kappa_\alpha = 2$ in this case). For a 2-dimensional Borel set A we obtain

$$\hat{\chi}_{N,\alpha,\beta}(A) = \sum_{\sigma \in \{-1,1\}^N} \mathcal{N}(0, C)(A - \hat{w})\mu_{N,\alpha,\beta}(\sigma)$$

$$= K_2 \sum_{\sigma \in \{-1,1\}^N} \int_A \exp\left(-\frac{1}{2}\left(\sqrt{N}(x - \hat{w}_1)^2 + \frac{\eta_\alpha}{2}(y - \hat{w}_2)^2\right)\right.$$

$$\left. + \frac{1}{4}\left(\sqrt{N}2\hat{w}_1^2 + \eta_\alpha \hat{w}_2^2\right)\right) dx dy$$

$$= K_2 \sum_{\sigma \in \{-1,1\}^N} \int_A \exp\left(-\frac{\sqrt{N}}{2}x^2 + \sqrt{N}x\hat{w}_1 - \frac{\eta_\alpha}{4}y^2 + \frac{\eta_\alpha}{2}y\hat{w}_2\right) dx dy$$

$$= K_3 \int_A \exp\left(-\frac{\sqrt{N}}{2}x^2 - \frac{\eta_\alpha}{4}y^2\right) \exp\left(\frac{N}{2}\log\cosh\left(\frac{x}{N^{1/4}} + \frac{\eta_\alpha}{2\sqrt{N}}y\right)\right.$$

$$\left. +\frac{N}{2}\log\cosh\left(\frac{x}{N^{1/4}} - \frac{\eta_\alpha}{2\sqrt{N}}y\right)\right) dx\,dy, \tag{18}$$

where again

$$K_2 := \frac{1}{2\pi\sqrt{\det C}Z_{N,\alpha,\beta}}, \qquad K_3 := 2^N \frac{1}{2\pi\sqrt{\det C}Z_{N,\alpha,\beta}}.$$

(but this time C is different).

Expanding $\log\cosh$ again up to fourth order we see that now the x^2-terms in the exponent cancel and we obtain

$$-\frac{\sqrt{N}x^2}{2} - \frac{\eta_\alpha y^2}{4} + \frac{N}{2}\log\cosh\left[\frac{x}{N^{1/4}} + \frac{\eta_\alpha y}{2\sqrt{N}}\right] + \frac{N}{2}\log\cosh\left[\frac{x}{N^{1/4}} - \frac{\eta_\alpha y}{2\sqrt{N}}\right]$$

$$= -\frac{1}{12}x^4 - \frac{1}{2}y^2\left(\frac{\eta_\alpha}{2} - \left(\frac{\eta_\alpha}{2}\right)^2\right) + \mathcal{O}(N^{-\frac{1}{2}})$$

with a $\mathcal{O}(N^{-\frac{1}{2}})$-term that depends on x and y but is uniformly bounded for x and y taken from a compact subset of $\mathbb{R}^2$.

For non-compact sets A we note that the integrand in (18) equals the respective integrand in the case $\alpha < 0$, if we interchange the roles of x and y and replace η_α by κ_α (and use $\log\cosh(z) = \log\cosh(-z)$), i.e. using the notation $A' := \{(y,x) : (x,y) \in A\}$ we have

$$\hat{\chi}_{N,\alpha,\beta}(A) = K_3 \int_{A'} \exp\left(-\frac{\sqrt{N}}{2}y^2 - \frac{\eta_\alpha}{4}x^2\right) \exp\left(\frac{N}{2}\log\cosh\left(\frac{y}{N^{1/4}} + \frac{\eta_\alpha x}{2\sqrt{N}}\right)\right.$$

$$\left. +\frac{N}{2}\log\cosh\left(\frac{\eta_\alpha x}{2\sqrt{N}} - \frac{y}{N^{1/4}}\right)\right) dx\,dy.$$

Again dividing the regime of integration into an inner, an intermediate and an outer region, we already saw in the case $\alpha < 0$ (note that in the case $\alpha < 0$ we did not use the explicit form of κ_α in that part of the proof and hence we may replace κ_α by η_α)

$$\lim_{N\to\infty} \hat{\chi}_{N,\alpha,\beta}(A) = \int_A \exp\left(-\frac{1}{12}x^4 - \frac{1}{2}y^2\left(\frac{\eta_\alpha}{2} - \left(\frac{\eta_\alpha}{2}\right)^2\right)\right) dx\,dy.$$

In particular, $\hat{w}_2 = \frac{\sqrt{N}}{2}(m_1 - m_2)$ converges in distribution to a $\mathcal{N}(0,\tilde{\sigma}^2)$-distribution, where $\tilde{\sigma}^2 = \frac{1}{\frac{\eta_\alpha}{2} - \left(\frac{\eta_\alpha}{2}\right)^2} = \frac{1}{\frac{\beta-\alpha}{2} - \left(\frac{\beta-\alpha}{2}\right)^2}$.

This weak convergence is equivalent to the convergence of the characteristic functions.

Computing the characteristic functions of the Gaussian distribution involved in the above proof, we have therefore shown that the characteristic function of $\hat{w}_2$ in the point $t \in \mathbb{R}$ satisfies

$$\lim_{N \to \infty} \mathbb{E}(e^{it\hat{w}_2}) e^{-\frac{1}{2}(\frac{2}{\beta - \alpha})t^2} = e^{-\frac{1}{2}t^2[\frac{\beta-\alpha}{2} - (\frac{\beta-\alpha}{2})^2]^{-1}}.$$

Therefore,

$$\lim_{N \to \infty} \mathbb{E}(e^{it\hat{w}_2}) = e^{\frac{1}{2}(\frac{2}{\beta-\alpha})t^2} e^{-\frac{1}{2}t^2[\frac{\beta-\alpha}{2} - (\frac{\beta-\alpha}{2})^2]^{-1}} = e^{-\frac{1}{2}t^2 \left(\frac{1}{1 - \frac{\beta-\alpha}{2}} \right)}.$$

On the level of weak convergence, or convergence in distribution, this in turn implies that

$$\hat{w}_2 \xrightarrow{N \to \infty} \mathcal{N}(0, \sigma^2), \quad \sigma^2 = \frac{1}{1 - \frac{\beta - \alpha}{2}}$$

in distribution. Similarly, $\hat{w}_1 = \frac{N^{\frac{1}{4}}}{2}(m_1 + m_2)$ convoluted with an independent Gaussian distribution with mean 0 and variance $\frac{1}{\sqrt{N}}$ converges to a probability measure ρ on $\mathbb{R}$ with density proportional to $\exp(-\frac{1}{12}x^4)$. Since the $\mathcal{N}(0, \frac{1}{\sqrt{N}})$-distribution converges to 0, as $N \to \infty$, this implies that $\frac{N^{1/4}}{2}(m_1 - m_2)$ converges to ρ in distribution, as well. □

4. Proof of Theorem 1.3

We are now ready to give the proof of our central result, Theorem 1.3, which consists of a combination of the previous two sections.

Proof of Theorem 1.3. In view of Theorem 2.1, all we need to do is to estimate $Z - Z'$ in the critical case.

Let us first assume that $\alpha > 0$. Then, Theorem 3.1 shows that $\frac{1}{\sqrt{N}}(\sum_{i \in S} \sigma_i - \sum_{j \notin S} \sigma_j)$ converges to a normal distribution with mean 0 and variance $\frac{2}{2 - \beta + \alpha}$. In particular

$$\mathbb{V}\left(\frac{1}{\sqrt{N}} \left(\sum_{i \in S} \sigma_i - \sum_{j \notin S} \sigma_j \right) \right) \to \frac{2}{2 - \beta + \alpha}.$$

But,

$$\mathbb{E} \frac{1}{\sqrt{N}} \left(\sum_{i \in S} \sigma_i - \sum_{j \notin S} \sigma_j \right) = 0,$$

hence

$$\mathbb{V}\left(\frac{1}{\sqrt{N}}\left(\sum_{i\in S}\sigma_i - \sum_{j\notin S}\sigma_j\right)\right)$$

$$= \frac{1}{N}\left(\sum_{i,j\in S}\mathbb{E}(\sigma_i\sigma_j) + \sum_{i,j\notin S}\mathbb{E}(\sigma_i\sigma_j) - 2\sum_{k\in S,l\notin S}\mathbb{E}(\sigma_k\sigma_l)\right)$$

$$= \frac{1}{N}\left(N + 2\frac{N}{2}\left(\frac{N}{2}-1\right)Z - 2\frac{N^2}{4}Z'\right),$$

where we recall that $Z := \mathbb{E}(\sigma_i\sigma_j)$, if $i \sim j$ and $Z' := \mathbb{E}(\sigma_i\sigma_j)$, if $i \nsim j$. Thus

$$\left(\frac{N}{2}-1\right)Z - \frac{N}{2}Z' \to \frac{2}{2-\beta+\alpha} - 1 = \frac{\beta-\alpha}{2-\beta+\alpha}. \tag{19}$$

Note that we excluded the case $\alpha = 0, \beta = 2$ in which case the right hand side would explode.

Note that Z is of order $\frac{1}{\sqrt{N}}$. On the one hand this follows from Theorem 6 and the remark from the paper by Kirsch and Toth [16] cited at the beginning of Section 3 and on the other hand it also follows from Theorem 3.1 directly: As noted in Theorem 3.1 (b) $\frac{1}{N^{3/4}}\sum_i\sigma_i$ converges to a non-degenerate limit distribution. Since $Z \geq |Z'|$ as a consequence of $\beta \geq |\alpha|$ ($\alpha = 1$, i.e. $\alpha = \beta$ being excluded, of course) this limit theorem implies that $\frac{N(N-1)Z}{N^{3/2}} \leq c$ for some constant c which shows that c is at most of order $\frac{1}{\sqrt{N}}$. But then (19) shows that

$$Z - Z' \sim C/N \qquad \text{for some constant } C > 0 \tag{20}$$

where $\sim$ indicates asymptotic equivalence.

Let us now turn to the other case where $\alpha < 0$ and still $\beta + |\alpha| = 2$. Then according to Theorem 3.1 the difference of the spins $\frac{1}{N^{3/4}}(\sum_{i\in S}\sigma_i - \sum_{j\notin S}\sigma_j)$ converges to a distribution ρ on $\mathbb{R}$ with density proportional to $\exp(-\frac{1}{12}x^4)$. In particular, ρ has finite variance

$$\tau := \sqrt{12}\frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})} \sim 1.17$$

Then

$$\mathbb{V}\left(\frac{1}{N^{3/4}}\left(\sum_{i\in S}\sigma_i - \sum_{j\notin S}\sigma_j\right)\right) \to \tau.$$

Taking into account that again $\mathbb{E}(\sum_{i\in S}\sigma_i - \sum_{j\notin S}\sigma_j) = 0$ and following the same calculations as above we obtain:

$$\frac{1}{\sqrt{N}} + \left(\frac{\sqrt{N}}{2} - \frac{1}{\sqrt{N}}\right)Z - \frac{\sqrt{N}}{2}Z' \to \tau. \tag{21}$$

The observation that this time $\frac{1}{\sqrt{N}}\sum_i \sigma_i$ converges to a normal distribution yields that Z is of order $\frac{1}{N}$, hence (21) implies

$$Z - Z' \sim \tilde{C}/\sqrt{N} \qquad \text{for some (other) constant } \tilde{C} > 0. \qquad (22)$$

Combining (20) and (22), respectively, with the reconstruction result Theorem 2.1 from [2] implies the assertion of Theorem 1.3. $\qquad\square$

Acknowledgment

References

[1] A. A. Amini and E. Levina. On semidefinite relaxations for the block model. *Ann. Statist.*, 46(1):149–179, 2018. MR3766949

[2] Q. Berthet, P. Rigollet, and P. Srivastava. Exact recovery in the ising block-model. *The Annals of Statistics*, 47(4):1805–1834, 2019. MR3953436

[3] A. Bovier and V. Gayrard. Rigorous bounds on the storage capacity of the dilute Hopfield model. *J. Statist. Phys.*, 69(3-4):597–627, 1992. MR1193852

[4] G. 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. MR3388257

[5] G. Bresler, E. Mossel, and A. Sly. Reconstruction of Markov random fields from samples: some observations and algorithms. *SIAM J. Comput.*, 42(2):563–578, 2013. MR3037003

[6] F. Collet. Macroscopic limit of a bipartite Curie-Weiss model: a dynamical approach. *J. Stat. Phys.*, 157(6):1301–1319, 2014. MR3277768

[7] R. Cont and M. Löwe. Social distance, heterogeneity and social interactions. *J. Math. Econom.*, 46(4):572–590, 2010. MR2674141

[8] R. S. Ellis. *Entropy, large deviations, and statistical mechanics.* Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1985 original. MR2189669

[9] M. Fedele and P. Contucci. Scaling limits for multi-species statistical mechanics mean-field models. *J. Stat. Phys.*, 144(6):1186–1205, 2011. MR2841921

[10] I. Gallo, A. Barra, and P. Contucci. Parameter evaluation of a simple mean-field model of social interaction. *Math. Models Methods Appl. Sci.*, 19(suppl.):1427–1439, 2009. MR2554157

[11] I. Gallo and P. Contucci. Bipartite mean field spin systems. Existence and solution. *Math. Phys. Electron. J.*, 14:Paper 1, 21, 2008. MR2407199

[12] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving optimal misclassification proportion in stochastic block models. *J. Mach. Learn. Res.*, 18:Paper No. 60, 45, 2017. MR3687603

[13] B. Gentz and M. Löwe. The fluctuations of the overlap in the Hopfield model with finitely many patterns at the critical temperature. *Probab. Theory Related Fields*, 115(3):357–381, 1999. MR1725405

[14] W. Kirsch. A survey on the method of moments. https://www.fernuni-hagen.de/stochastik/downloads/momente.pdf, 2015.

[15] W. Kirsch and G. Toth. Two groups in a Curie-Weiss model. *preprint*, arXiv:1712.08477, 2017.

[16] W. Kirsch and G. Toth. Critical regime in a Curie-Weiss model with two groups and heterogeneous coupling. *preprint*, arXiv:1807.05020, 2018.

[17] W. Kirsch and G. Toth. Two groups in a Curie-Weiss model with heterogeneous coupling. *preprint*, arXiv:1806.06708, 2018.

[18] H. Knöpfel and M. Löwe. Zur Meinungsbildung in einer heterogenen Bevölkerung—ein neuer Zugang zum Hopfield Modell. *Math. Semesterber.*, 56(1):15–38, 2009. MR2480940

[19] H. Knöpfel, M. Löwe, K. Schubert, and A. Sinulis. Fluctuation results for general block spin Ising models. *preprint*, arXiv:1902.02080, 2019. MR4081224

[20] M. Löwe and K. Schubert. Fluctuations for block spin ising models. *Electron. Commun. Probab.*, 23:12 pp., 2018. MR3852267

[21] M. Löwe, K. Schubert, and F. Vermet. Multi-group binary choice with social interaction and a random communication structure – a random graph approach. *preprint*, arXiv:1904.11890, 2019.

[22] E. Mossel, J. Neeman, and A. Sly. Belief propagation, robust reconstruction and optimal recovery of block models. *Ann. Appl. Probab.*, 26(4):2211–2256, 2016. MR3543895

[23] A. A. Opoku, K. Owusu Edusei, and R. K. Ansah. A conditional Curie-Weiss model for stylized multi-group binary choice with social interaction. *J. Stat. Phys.*, 171(1):106–126, 2018. MR3773853