ON RATES OF CONVERGENCE
IN THE CURIE-WEISS-POTTS MODEL
WITH AN EXTERNAL FIELD

Peter Eichelsbacher\textsuperscript{1} and Bastian Martschink\textsuperscript{2}

Abstract: Stein’s Method provides a way of bounding the distance of a probability
distribution to a target distribution. In the present paper we obtain rates of conver-
gence for limit theorems via Stein’s Method of exchangeable pairs in the context of
the Curie-Weiss-Potts model. Our results include Kolmogorov bounds for multivari-
ate normal approximation in the whole domain $\beta \geq 0$ and $h \geq 0$, as well as rates for
non-Gaussian approximations at the extremity of the critical line of the model.

AMS 2000 Subject Classification: Primary 60F05; Secondary 82B20, 82B26.

Key words: Stein’s method, exchangeable pairs, Curie-Weiss Potts models, critical tempera-
ture.

\textsuperscript{1}Ruhr-Universität Bochum, Fakultät für Mathematik, NA 3/66, D-44780 Bochum, Germany,
peter.eichelsbacher@ruhr-uni-bochum.de
\textsuperscript{2}Ruhr-Universität Bochum, Fakultät für Mathematik, NA 3/68, D-44780 Bochum, Germany,
bastian.martschink@rub.de
Both authors have been supported by Deutsche Forschungsgemeinschaft via SFB/TR 12.
1. Introduction

1.1 The Curie-Weiss-Potts model The Curie-Weiss-Potts model is a mean field approximation of the well-known Potts model, a famous model in equilibrium statistical mechanics. It is defined in terms of a mean interaction averaged over all sites in the model, more precisely, by sequences of probability measures of \( n \) spin random variables that may occupy one of \( q \) different states. For \( q = 2 \) the model reduces to the simpler Curie-Weiss model. Two ways in which the Curie-Weiss-Potts model approximates the Potts model are discussed in [18] and [17]. Probability limit theorems for the Curie-Weiss-Potts model are proved first in [13]. One reason of interest in this model is its explicit exhibition of a number of properties of real substances, such as multiple phase transitions, metastable states and others. In comparison to the Curie-Weiss model it has a more intricate phase transition structure because for example at the critical inverse temperature it has not a second-order phase transition like the Curie-Weiss model but a first order transition. In order to carry out the analysis of the model, detailed information about the structure of the set of canonical equilibrium macro-states is required. The probability observing a configuration \( \sigma \in \{1, \ldots, q\}^n \) in an exterior field \( h \) equals

\[
P_{\beta,h,n}(\sigma) = \frac{1}{Z_{\beta,h,n}} \exp \left( \frac{\beta}{2n} \sum_{1 \leq i < j \leq n} \delta_{\sigma_i,\sigma_j} + h \sum_{i=1}^n \delta_{\sigma_i,1} \right) \tag{1.1}\]

where \( \delta \) is the Kronecker symbol, \( \beta := T^{-1} \) is the inverse temperature and \( Z_{\beta,h,n} \) is the normalization constant known as the partition function. More precisely:

\[
Z_{\beta,h,n} = \sum_{\sigma \in \{1, \ldots, q\}^n} \exp \left( \frac{\beta}{2n} \sum_{1 \leq i < j \leq n} \delta_{\sigma_i,\sigma_j} + h \sum_{i=1}^n \delta_{\sigma_i,1} \right).
\]

For \( \beta \) small, the spin random variables are weakly dependent while for \( \beta \) large they are strongly dependent. It was shown in [27] that at \( h = 0 \) the model undergoes a phase transition at the critical inverse temperature

\[
\beta_c = \begin{cases} 
q, & \text{if } q \leq 2 \\
2 \frac{q-1}{q-2} \log(q-1), & \text{if } q > 2;
\end{cases} \tag{1.2}
\]

and that this transition is first order if \( q > 2 \). Our interest is in the limit distribution of the empirical vector of the spin variables

\[
N = (N_1, \ldots, N_q) = \left( \sum_{i=1}^n \delta_{\sigma_i,1}, \ldots, \sum_{i=1}^n \delta_{\sigma_i,q} \right) \tag{1.3}
\]

which counts the number of spins of each color for a configuration \( \sigma \). Note that the normalized empirical vector \( L_n := N/n \) belongs to the set of probability vectors

\[
\mathcal{H} = \{ x \in \mathbb{R}^q : x_1 + \cdots + x_q = 1 \text{ and } x_i \geq 0, \forall i \}. 
\]
For $q > 2$ and $\beta < \beta_c$ $L_n$ satisfies the law of large numbers $P_{\beta,0,n}(L_n \in d\nu) \Rightarrow \delta_{\nu_0}(d\nu)$ as $n \to \infty$, where $\nu_0 = (1/q, \ldots, 1/q) \in \mathbb{R}^q$. For $\beta > \beta_c$ the law of large numbers breaks down and is replaced by the limit

$$P_{\beta,0,n}(L_n \in d\nu) \Rightarrow \frac{1}{q} \sum_{i=1}^{q} \delta_{\nu_i(\beta)}(d\nu),$$

where $\{\nu_i(\beta), i = 1, \ldots, q\}$ are $q$ distinct probability vectors in $\mathbb{R}^q$, distinct from $\nu_0$. The \textit{first order phase transition} is the fact that for $i = 1, \ldots, q$

$$\lim_{\beta \to \beta_c^+} \nu_i(\beta) \neq \nu_0,$$

see [13]. The case of \textit{non-zero} external field $h \neq 0$ was considered in [2] and it turned out that the first-order transition remains on a critical line. The line was computed explicitly in [3], see (1.4).

In the present work we will obtain certain known probabilistic limit theorems for the Curie-Weiss-Potts model, especially for the empirical vector of the spin variables $N$, \textit{but at the same time} we present rates of convergence for all the limit theorems. We consider the fluctuations of the empirical vector $N$ around its typical value outside the critical line and we describe the fluctuations and rates of convergence at a extremity of the critical line. This extends previous results on the Curie-Weiss Potts model with no external field [13] as well as with external field [15]. The method of proof will be an application of Stein’s method of so called exchangeable pairs in the case of multivariate normal approximation as well as the application of Stein’s method in the case of non-Gaussian approximation. Stein’s method will be explained later.

We turn to the description of the set of canonical equilibrium macro-states of the Curie-Weiss-Potts model. These states are solutions of an unconstrained minimization problem involving probability vectors on $\mathbb{R}^q$. The macro-states describe equilibrium configurations of the model in the thermodynamic limit $n \to \infty$. For each $i$ the $i$’th component of an equilibrium macro-state gives the asymptotic relative frequency of spins taking the spin value $i$ with $i \in \{1, \ldots, q\}$. We appeal to the theory of large deviations to define the canonical equilibrium macro-states. Sanov’s theorem states that with respect to the product measures $P_n(\omega) = 1/q^n$ for $\omega \in \{1, \ldots, q\}^n$ the empirical vectors $L_n$ satisfy a large deviations principle (LDP) on $\mathcal{H}$ with speed $n$ and rate function given by the relative entropy

$$I(x) = \sum_{i=1}^{q} x_i \log(q x_i), \ x \in \mathcal{H}.$$ 

We use the formal notation $P_n(L_n \in dx) \approx \exp(-n I(x))$ (precise definition: [9]). The LDP for $L_n$ with respect to $P_{\beta,h,n}$ can be proved as in [11]. Let

$$f_{\beta,h}(x) = \sum_{i=1}^{q} x_i \log(q x_i) - \frac{\beta}{2} \sum_{i=1}^{q} x_i^2 - h x_1, \ x \in \mathcal{H}.$$
Then $P_{\beta,h,n}(L_n \in d\nu) \approx \exp(-nJ_{\beta,h}(\nu))$ with

$$J_{\beta,h}(\nu) := f_{\beta,h}(\nu) - \inf_{x \in \mathcal{H}} f_{\beta,h}(x),$$

see also [8]. Now if $J_{\beta,h}(\nu) > 0$, then $\nu$ has an exponentially small probability of being observed. Hence the corresponding set of canonical equilibrium macro-states are naturally defined by

$$E_{\beta,h} := \{ \nu \in \mathcal{H} : \nu \text{ minimizes } f_{\beta,h}(\nu) \}.$$

Remark that the specific Gibbs free energy for the Curie-Weiss-Potts model is the quantity $\psi(\beta, h)$ defined by the limit

$$-\beta \psi(\beta, h) = \lim_{n \to \infty} \frac{1}{n} \log Z_{\beta,h,n}.$$

From the large deviations result it follows that

$$-\beta \psi(\beta, h) = -\inf_{x \in \mathcal{H}} f_{\beta,h}(x).$$

In the case $h = 0$ and $q > 2$, it is known since [27] (for detailed proofs see [8] Theorem 3.1) that $E_{\beta,0}$ consists of one element for any $0 < \beta < \beta_c$, where $\beta_c$ is the critical inverse temperature given in (1.2). For any $\beta > \beta_c$, the set consists of $q$ elements and at $\beta_c$ it consists of $q + 1$ elements. In the case with an external field $h \geq 0$ the global minimizers of $f_{\beta,h}$ can be described as follows. In [3] the following critical line was computed:

$$h_T := \left\{ (\beta, h) : 0 \leq h < h_0 \text{ and } h = \log(q-1) - \beta \frac{q-2}{2(q-1)} \right\} \quad (1.4)$$

with extremities $(\beta_c, 0)$ and $(\beta_0, h_0)$, where

$$\beta_0 = 4 \frac{q-1}{q} \quad \text{and} \quad h_0 = \log(q-1) - 2 \frac{q-2}{q} \quad ((\beta_0, h_0) \text{ were already determined in [2]}).$$

Now consider the parametrization

$$x_z := \left( \frac{1+z}{2}, \frac{1-z}{2(q-1)}, \ldots, \frac{1-z}{2(q-1)} \right), \quad z \in [-1,1].$$

Depending on the parameters $(\beta, h)$ the function $f_{\beta,h}$ presents one or several global minimizers. The following statement summarizes the results of [27], [8] in the case $h = 0$ and of [3] for $h > 0$.

**Theorem 1.1.** Let $\beta, h \geq 0$.

1. If $h > 0$ and $(\beta, h) \notin h_T$, the function $f_{\beta,h}$ has a unique global minimum point in $\mathcal{H}$. This minimizer is analytic in $\beta$ and $h$ outside of $h_T \cup \{(\beta_0, h_0)\}$.

2. If $h > 0$ and $(\beta, h) \in h_T$, the function $f_{\beta,h}$ has two global minimum points in $\mathcal{H}$. More precisely, for any $z \in (0, (q-2)/q)$, the two global minimum points of $f_{\beta_z,h_z}$ at

$$\beta_z = 2 \frac{q-1}{zq} \log \left( \frac{1+z}{1-z} \right) \quad \text{and} \quad h_z = \log(q-1) - \frac{q-2}{2(q-1)} \beta_z$$

are the points $x_{\pm z}$. 
(3) If \( h = 0 \) and \( \beta < \beta_c \), the unique global minimum point of \( f_{\beta,0} \) is \((1/q, \ldots, 1/q)\).

(4) If \( h = 0 \) and \( \beta > \beta_c \), there are \( q \) global minimum points of \( f_{\beta,0} \), which all equal \( x_z \) up to a permutation of the coordinates for some \( z \in ((q-2)/q, 1) \).

(5) If \( h = 0 \) and \( \beta = \beta_c \), there are \( q + 1 \) global minimum points of \( f_{\beta,0} \): the symmetric one \((1/q, \ldots, 1/q)\) together with the permutations of

\[
\left( \frac{q-1}{q}, \frac{1}{q(q-1)}, \ldots, \frac{1}{q(q-1)} \right).
\]

Interesting enough, the very first results on probabilistic limit theorems ([12] for the Curie-Weiss model and [13] for the Curie-Weiss-Potts model) used the structure of the global minimum points of another function \( G_{\beta,h} \). For the Curie-Weiss-Potts model with \( h = 0 \) this function is given by

\[
G_{\beta,0}(u) := \frac{1}{2} \beta \langle u, u \rangle - \log \sum_{i=1}^{q} e^{\beta u_i}, \quad u \in \mathbb{R}^q.
\]

With convex duality one obtains the alternative representation of the specific Gibbs free energy:

\[
\beta \psi(\beta,0) = \min_{u \in \mathbb{R}^q} G_{\beta,0}(u) + \log q.
\]

Actually \( f_{\beta,0} \) and \( G_{\beta,0} \) have the same global minimum points, see [18] for \( \beta \neq \beta_c \). A proof of this result for any \( \beta > 0 \) can be found in [8, Theorem 3.1]. The main reason to use \( G_{\beta,0} \) instead of \( f_{\beta,0} \) is the usefulness of a representation of the distribution of \( L_n \) in terms of \( G_{\beta,h} \), called *Hubbard-Stratonovich transform* (see [13, Lemma 3.2] and the proof of Lemma 4.5 in this paper). This is a famous tool since the work of Ellis and Newman [12]. For \( \beta > 0 \) and \( h \) real the global minimum points of \( f_{\beta,h} \) coincide with the global minimum points of the function

\[
G_{\beta,h}(u) := \frac{1}{2} \beta \langle u, u \rangle - \log \left( \sum_{i=1}^{q} \exp(\beta u_i + h \delta_{i,1}) \right), \quad u \in \mathbb{R}^q
\]

(1.5)

(for a proof see [14, Theorem B.1]; or apply a general result on minimum points of certain functions related by convex duality, [8, Theorem A.1], see also [26]). Hence we know that all statements of Theorem 1.1 hold true for \( G_{\beta,h} \).

**Corollary 1.2.** The statements in Theorem 1.1 for the global minimum points of \( f_{\beta,h} \) hold true one to one for \( G_{\beta,h} \), defined in (1.5).

The *detour* first describing the canonical equilibrium macro-states of the Curie-Weiss-Potts model using large deviation theory and second using convex duality has the following reason. Applying Stein’s method we will *automatically meet* the function \( G_{\beta,h} \) and the limit theorems and the proof of certain rates of convergence will depend on the location of the global minimum points of \( G_{\beta,h} \) (as in [12], [13] and [14] and a lot of other papers). But for \( h > 0 \) only \( f_{\beta,h} \) and its minimizers were completely characterized in the literature, see Theorem 1.1. So we had to argue that we also know the phase diagram of \( G_{\beta,h} \).
1.2 Stein’s method of exchangeable pairs  

Starting with a bound for the distance between univariate random variables and the normal distribution Stein’s method was first published in [23] (1972). Being particularly powerful in the presence of both local dependence and weak global dependence his method proved to be very successful. In [24] Stein introduced his exchangeable pair approach. At the heart of the method is a coupling of a random variable \( W \) with another random variable \( W' \) such that \((W, W')\) is exchangeable, i.e. their joint distribution is symmetric. Central in his approach is the fact that for all antisymmetric measurable functions \( F(x, y) \) we have \( \mathbb{E}[F(W, W')] = 0 \) if the expectation exists. Stein proved further on that a measure of proximity of \( W \) to normality may be provided by the exchangeable pair if \( W' - W \) is sufficiently small. He assumed the property that there is a number \( \lambda > 0 \) such that the expectation of \( W' - W \) with respect to \( W \) satisfies
\[
\mathbb{E}[W' - W \mid W] = -\lambda W.
\]

Heuristically, this condition can be understood as a linear regression condition: if \((W, W')\) were bivariate normal with correlation \( \varrho \), then \( \mathbb{E}(W' \mid W) = \varrho W \) and the condition would be satisfied with \( \lambda = 1 - \varrho \). Stein proved that for any uniformly Lipschitz function \( h \),
\[
\mathbb{E}|h(W) - \mathbb{E}h(Z)| \leq \delta \|h'\| \quad \text{with} \quad Z \text{ denoting a standard normally distributed random variable and}
\]
\[
\delta = 4\mathbb{E}\left|1 - \frac{1}{2\lambda} \mathbb{E}((W' - W)^2 \mid W)\right| + \frac{1}{2\lambda} \mathbb{E}|W - W'|^3.
\]  

(1.6)

Stein’s approach has been successfully applied in many models, see e.g. [24] or [25] and references therein. In [21], the range of application was extended by replacing the linear regression property by a weaker condition assuming that there is also a random variable \( R = R(W) \) such that
\[
\mathbb{E}[W' - W \mid W] = -\lambda W + R.
\]

While the approach has proved successful also in non-normal contexts (see [4], [6] and [10]) it remained restricted to the one-dimensional setting for a long time. The problem was that it was not obvious how to transfer the linearity condition into the multivariate case. However in [5] this issue was finally addressed. They extended the linearity condition to the multivariate setting such that, for all \( i \in \{1, \ldots, d\} \), \( \mathbb{E}[W'_i - W_i \mid W] = -\lambda W_i \) for a fixed number \( \lambda \), where now \( W = (W_1, \ldots, W_d) \) and \( W' = (W'_1, \ldots, W'_d) \) are identically distributed \( d \)-vectors with uncorrelated components. As in the univariate case an extension to the additional remainder term \( R \) would be straightforward. This coupling has the purpose to estimate the distance to the standard multivariate distribution. Applying the linear regression heuristic in the multivariate case leads to a new condition due to [19]:
\[
\mathbb{E}[W' - W \mid W] = -\Lambda W + R
\]  

(1.7)

for an invertible \( d \times d \) matrix \( \Lambda \) and a remainder term \( R = R(W) \). This linearity condition is more natural than the one of [5]. Different exchangeable pairs, obviously, will yield different \( \Lambda \) and \( R \).
Interesting enough the Curie-Weiss-Potts model will be an example to demonstrate the power of the approach in [19]. Constructing an exchangeable pair in the Curie-Weiss-Potts model to obtain an approximate linear regression property (1.7) leads us to the function $G_{\beta,h}$. This will be sketched now. Let $q > 2$, $h = 0$ and $\beta < \beta_c$, and let $x_0$ denote the unique global minimum point of $G_{\beta,0}$, see Theorem [11]. We consider

$$W := \sqrt{n} \left( \frac{N}{n} - x_0 \right) = \sqrt{n} \left( L_n - x_0 \right).$$

We produce a spin collection $\sigma' = (\sigma'_i)_{i \geq 1}$ via a Gibbs sampling procedure: Let $I$ be uniformly distributed over $\{1, \ldots, n\}$ and independent from all other random variables involved. We will now replace the spin $\sigma_i$ by $\sigma'_i$ drawn from the conditional distribution of the $i$'th coordinate given $(\sigma_j)_{j \neq i}$, independently from $\sigma_i$. We define

$$Y_i := (Y_{i,1}, \ldots, Y_{i,q})^t := (\delta_{\sigma_i,1}, \ldots, \delta_{\sigma_i,q})^t$$

and consider

$$W' := W - \frac{Y_i}{\sqrt{n}} + \frac{Y'_i}{\sqrt{n}}.$$

(1.8)

Hence it is not hard to see that $(W, W')$ is an exchangeable pair. This construction will also be evident for all the proofs in this paper. Let $F := \sigma(\sigma_1, \ldots, \sigma_n)$. We obtain

$$E \left[ W'_i - W_i \mid F \right] = \frac{1}{\sqrt{n}} E \left[ Y'_{I,i} - Y_{I,i} \mid F \right]$$

$$= \frac{1}{\sqrt{n}} \frac{1}{n} \sum_{j=1}^n E \left[ Y'_{j,i} - Y_{j,i} \mid \sigma_1, \ldots, \sigma_n \right]$$

$$= - \frac{1}{\sqrt{n}} \frac{1}{n} \sum_{j=1}^n Y_{j,i} + \frac{1}{\sqrt{n}} \frac{1}{n} \sum_{j=1}^n E \left[ \delta_{\sigma'_j,i} \mid \sigma_1, \ldots, \sigma_n \right].$$

Using our construction we obtain

$$E \left[ \delta_{\sigma'_j,i} \mid \sigma_1, \ldots, \sigma_n \right] = E \left[ \delta_{\sigma_j,i} \mid (\sigma_t)_{t \neq j} \right] = P_{\beta,0,n} (\sigma_j = i \mid (\sigma_t)_{t \neq j}).$$

By a straightforward calculation (see Lemma [4,3]) we get that

$$P_{\beta,0,n} (\sigma_j = i \mid (\sigma_t)_{t \neq j}) = \frac{\exp (\beta m_{i,j}(\sigma))}{\sum_{k=1}^q \exp (\beta m_{k,j}(\sigma))}$$

with $m_{i,j}(\sigma) := \frac{1}{n} \sum_{l \neq j} \delta_{\sigma_l,i}$. Using the notion $m_i(\sigma) := \frac{1}{n} \sum_{t=1}^n \delta_{\sigma_t,i}$ one obtains

$$\frac{1}{\sqrt{n}} \frac{1}{n} \sum_{j=1}^n E \left[ \delta_{\sigma'_j,i} \mid (\sigma_t)_{t \neq j} \right] = \frac{1}{\sqrt{n}} \frac{1}{n} \sum_{k=1}^q \exp (\beta m_k(\sigma)) + R_n(i).$$

(1.9)

Moreover by definition of $G_{\beta,h}$ (see Lemma [4,2]) we obtain

$$\frac{\exp (\beta u_m)}{\sum_{k=1}^q \exp (\beta u_k)} = u_m - \frac{1}{\beta} \frac{\partial}{\partial u_m} G_{\beta,0}(u).$$
Summarizing we obtain
\[
E[W'_i - W_i | \mathcal{F}] = -\frac{1}{n} W_i - \frac{x_{0,i}}{\sqrt{n}} + R_n(i) + \frac{1}{\sqrt{n}} \left( m_i(\sigma) - \frac{1}{\beta} \frac{\partial}{\partial u_i} G_{\beta,0}(m(\sigma)) \right)
\] (1.10)
with \( m(\sigma) = (m_1(\sigma), \ldots, m_q(\sigma)) \). Hence using all informations on \( G_{\beta,h} \) (Taylor-expansion and the results on the global minimum points, Theorem 1.1, Corollary 1.2), it seems to be possible to calculate \( \Lambda \) and \( R \) in the regression condition (1.7). Indeed we are able to proof that this condition is satisfied for any \((\beta, h)\).

In Section 2 of the present paper, the limit theorems and the rates of convergence are stated. They include a central limit theorem and a bound on the distance to a multivariate normal distribution for \( L_n \) outside the critical line. When \( G_{\beta,h} \) has several global minimizers, that is when \((\beta, h) \in h_T \) or \( \beta \geq \beta_c \) and \( h = 0 \), the empirical vector \( L_n \) is close to either one or the other of the minimizers. In this case we determine a central limit theorem with conditioning for \( L_n \). Next we describe the fluctuations at the extremity \((\beta_0, h_0)\) of the critical line, again combined with a rate of convergence. In Section 3 we state an abstract nonsingular multivariate normal approximation theorem for smooth test functions from \[19\]. Moreover we present a Kolmogorov bound (nonsmooth test functions) for bounded random vectors \( W' - W \) under exchangeability. Finally we state an abstract non-Gaussian univariate approximation theorem for the Kolmogorov-distance from \[10\]. Section 4 contains some auxiliary results which will be necessary for the proofs given in Section 5.

## 2. Statement of results

Let us fix some notation. From now on we will write random vectors in \( \mathbb{R}^d \) in the form \( w = (w_1, \ldots, w_d)^t \), where \( w_i \) are \( \mathbb{R} \)-valued variables for \( i = 1, \ldots, d \). If a matrix \( \Sigma \) is symmetric, nonnegative definite, we denote by \( \Sigma^{1/2} \) the unique symmetric, nonnegative definite square root of \( \Sigma \). \( Id \) denotes the identity matrix and from now on \( Z \) will denote a random vector having standard multivariate normal distribution. The expectation with respect to the measure \( P_{\beta,h,n} \) will be denoted by \( E := E_{P_{\beta,h,n}} \).

Let \( q > 2 \). We first consider the issue of the fluctuations of the empirical vector \( N \) defined in \[13\] around its typical value. The case of the Curie-Weiss model \( (q = 2) \) was considered in \[22\] and \[12\]. A Berry-Esseen bound was proved in \[10\] (and independently in \[6\]). The Curie-Weiss-Potts model was treated in \[13\] and for non-zero external field in \[15, Theorem 3.1\]. To the best of our knowledge there are no Kolmogorov bounds known.

**Theorem 2.1.** Let \( \beta > 0 \) and \( h \geq 0 \) with \((\beta, h) \neq (\beta_0, h_0)\). Assume that there is a unique minimizer \( x_0 \) of \( G_{\beta,h} \). Let \( W \) be the following random variable:

\[
W := \sqrt{n} \left( \frac{N}{n} - x_0 \right).
\]
If $Z$ has the $q$-dimensional standard normal distribution, we have for every three times differentiable function $g$,
\[
|E g(W) - E g \left( \Sigma^{-1/2} Z \right) | \leq C \cdot n^{-1/2},
\]
for a constant $C$ and $\Sigma := E [W W^t]$. Indeed we obtain that $C = O(q^6)$.

Remark that we compare the distribution of the rescaled vector $N$ with a multivariate normal distribution with covariance matrix $E [W W^t]$. This is an advantage of Stein’s method: for any fixed number of particles/spins $n$, we are able to compare the distribution of $W$ with a distribution with the same $n$-dependent covariance structure.

In order to state our next result we introduce conditions on the function classes $G$ we consider. Following [20], let $\Phi$ denote the standard normal distribution in $\mathbb{R}^q$. We define for $g$:
\begin{align}
    g^+_\delta(x) &= \sup \{ g(x+y) : |y| \leq \delta \}, \\
    g^-_\delta(x) &= \inf \{ g(x+y) : |y| \leq \delta \}, \\
    \tilde{g}(x,\delta) &= g^+_\delta(x) - g^-_\delta(x).
\end{align}

Let $G$ be a class of real measurable functions on $\mathbb{R}^q$ such that
\begin{enumerate}
    \item The functions $g \in G$ are uniformly bounded in absolute value by a constant, which we take to be 1 without loss of generality.
    \item For any $q \times q$ matrix $A$ and any vector $b \in \mathbb{R}^q$, $g(Ax + b) \in G$.
    \item For any $\delta > 0$ and any $g \in G$, $g^+_\delta(x)$ and $g^-_\delta(x)$ are in $G$.
    \item For some constant $a = a(G,q)$, $\sup_{g \in G} \left\{ \int_{\mathbb{R}^q} \tilde{g}(x,\delta) \Phi(dx) \right\} \leq a\delta$. Obviously we may assume $a \geq 1$.
\end{enumerate}

Considering the one dimensional case, we notice that the collection of indicators of all half lines, and indicators of all intervals form classes in $G$ that satisfy these conditions with $a = \sqrt{2/\pi}$ and $a = 2\sqrt{2/\pi}$ respectively. This was shown for example in [20]. In dimension $q \geq 1$ the class of indicators of convex sets is known to be such a class. Using this notation we are able to present an equivalent Theorem to 2.1 for our function classes $G$.

**Theorem 2.2.** Let $\beta > 0$ and $h \geq 0$ with $(\beta, h) \neq (\beta_0, h_0)$. Assume that there is a unique minimizer $x_0$ of $G_{\beta,h}$. Let $W$ and $Z$ be as in 2.1 Then, for all $g \in G$ with $|g| \leq 1$, we have
\[
|E g(W) - E g \left( \Sigma^{-1/2} Z \right) | \leq C \log(n) \cdot n^{-1/2},
\]
for a constant $C$ and $\Sigma := E [W W^t]$.

Letting $G$ be the collection of indicators of lower quadrants the distance above specializes to the Kolmogorov distance.

When the function $G_{\beta,h}$ has several global minimizers, the empirical vector $N/n$ is close to either one or the other of these minima. We determine the conditional fluctuations and a rate of convergence:
Theorem 2.3. Assume that $\beta, h \geq 0$ and that $G_{\beta,h}$ has multiple global minimum points $x_1, \ldots, x_l$ with $l \in \{2, q, q+1\}$ and let $\epsilon > 0$ be smaller than the distance between any two global minimizers of $G_{\beta,h}$. Furthermore, let $W^{(i)} := \sqrt{n} \left( \frac{N_n}{n} - x_i \right)$. Then, if $Z$ has the $q$-dimensional standard normal distribution, under the conditional measure $P_{\beta,h,n} \left( \cdot \mid N_n \in B(x_i, \epsilon) \right)$, we have for every three times differentiable function $g$,

$$\left| \mathbb{E} g(W^{(i)}) - \mathbb{E} \left( \Sigma^{-1/2} Z \right) \right| \leq C \cdot n^{-1/2},$$

for a constant $C$ and $\Sigma := \mathbb{E} \left[ W^{(i)} (W^{(i)})^t \right]$. $B(x_i, \epsilon)$ denotes the open ball of radius $\epsilon$ around $x_i$.

We note that we get a similar result as in Theorem 2.2 for the function classes $G$ in the case of several global minimizers. Finally we will take a look at the extremity $(\beta_0, h_0)$ of the critical line $h_T$. Given a vector $u \in \mathbb{R}^q$, we denote by $u^\perp$ the vector space made of all vectors orthogonal to $u$ in the Euclidean space $\mathbb{R}^q$. Consider the hyperplane

$$M := \left\{ x \in \mathbb{R}^q : \sum_{i=1}^q x_i = 0 \right\},$$

(2.4)

which is parallel to $H$. The fluctuations belong to $M$, since all global minimizers are in $H$. The following result extends [12, Theorem 3.9] which applies to the case of the Curie-Weiss model at the critical inverse temperature. We remind the fact that at $(\beta_0, h_0)$ the function $G_{\beta_0, h_0}$ has the unique minimizer $x = (1/2, 1/2(q-1), \ldots, 1/2(q-1)) \in \mathbb{R}^q$. Now we take $u = (1-q, 1, \ldots, 1) \in M \subset \mathbb{R}^q$ and define a real valued random variable $T$ and a random vector $V \in M \cap u^\perp$ such that

$$N = n x + n^{3/4} T u + n^{1/2} V.$$  

(2.5)

Since $N - n x \in M$, the implicit definition of $T$ and $V$ presents a partition into a vector in (the subspace spanned by) $u$ and $u^\perp$. The main interest is the limiting behaviour of $T$. The new scaling of $W$ is given by

$$\frac{N_j - n \frac{1}{2(q-1)}}{n^{3/4}} = T_j/n^{1/4}, \quad j = 2, \ldots, q,$n

and its possible limit we observe is reminiscent to [12], see also [7]. The following theorem gives a Kolmogorov bound for Theorem 3.7 in [15].

Theorem 2.4. For $(\beta, h) = (\beta_0, h_0)$ let $x = (1/2, 1/2(q-1), \ldots, 1/2(q-1))$ be the unique minimizer of $G_{\beta_0, h_0}$ and $u = (1-q, 1, \ldots, 1)$. Furthermore, let $Z_{q,T}$ be a random variable distributed according to the probability measure on $\mathbb{R}$ with the density

$$f_{q,T}(t) := f_{q,T,n} := C \cdot \exp \left( -\frac{1}{4 \mathbb{E}(T^4)} t^4 \right),$$

where
where $T$ is defined in (2.5). Then we obtain for any uniformly Lipschitz function $g : \mathbb{R} \to \mathbb{R}$ that
\[
|\mathbb{E}g(T) - \mathbb{E}g(Z_{q,T})| \leq C \cdot n^{-1/4}.
\]
Moreover we obtain
\[
\sup_{t \in \mathbb{R}} |\mathbb{P}(T \leq t) - F_{q,T}(t)| \leq C \cdot n^{-1/4} \quad \text{(bound for the Kolmogorov-distance)},
\]
where $F_{q,T}$ denotes the distribution function of $f_{q,T}$.

**Remark 2.5.** As we will see in the proof of Theorem 2.4, the density $f_{q,T}$ originally has the form
\[
\exp \left( -\frac{4(q-1)^4}{3E(T\psi(T))} t^4 \right)
\]
(up to a constant) with a function $\psi$ such that $E(T\psi(T)) = \frac{16(q-1)^4}{3}E(T^4)$. From [15, Theorem 3.7] we know that $T$ converges in distribution to the probability measure on $\mathbb{R}$ proportional to $g_q(t) = \exp \left( -\frac{4(q-1)^4}{3} t^4 \right)$.

Hence we conclude that $\lim_{n \to \infty} f_{q,T,n} = g_q$ point-wise and therefore $\frac{16(q-1)^4}{3}E[T^4] \to 1$. Remark, that the rate of convergence of Theorem 2.4 also holds true when we compare the distribution of $T$ with the law on $\mathbb{R}$ with density $g_q$.

Additionally we get a theorem for the random vector $V$, improving Theorem 3.7 in [15].

**Theorem 2.6.** Let $V$ be defined as in (2.5). For $(\beta, h) = (\beta_0, h_0)$ and $\Sigma := E[V \mathbf{V}^t]$ we have that for every three times differentiable function $g$,\[
|\mathbb{E}g(V) - \mathbb{E}g(\Sigma^{-1/2}Z)| \leq C \cdot n^{-1/4},
\]
where $C$ is an absolute constant.

**Remark 2.7.** The proofs of Theorem 2.4 and Theorem 2.6 employed a fourth-order Taylor expansion of $G_{\beta_0, h_0}$, see (6.11) and (6.12) in the Appendix. Without a doubt, the first and the third term in (6.11), as well as in (6.12), gives the order $O(n^{-1/4})$.

### 3. Stein’s method

Let us fix some more notations. The transpose of the inverse of a matrix will be presented in the form $A^{-t} := (A^{-1})^t$. Furthermore we will need the supremum norm, denoted by $\| \cdot \|$ for both functions and matrices. For derivatives of smooth functions $f : \mathbb{R}^d \to \mathbb{R}$, we use the notation $\nabla$ for the gradient operator. For a function $f : \mathbb{R}^d \to \mathbb{R}$, we abbreviate
\[
|f|_1 := \sup_i \| \frac{\partial}{\partial x_i} f \|, \quad |f|_2 := \sup_{i,j} \| \frac{\partial^2}{\partial x_i \partial x_j} f \|,
\]
(3.1)
and so on, if these derivatives exist.

The method of Stein is based on the characterization of the normal distribution that \( Y \in \mathbb{R}^d \), \( d \in \mathbb{N} \), is a centered multivariate normal vector with covariance matrix \( \Sigma \) if and only if

\[
\mathbb{E} \left[ \nabla^t \Sigma \nabla f(Y) - Y^t \nabla f(Y) \right] = 0 \quad \text{for all smooth } f : \mathbb{R}^d \to \mathbb{R}.
\]  

(3.2)

It is well known, see [1] and [16], that for any \( g : \mathbb{R}^d \to \mathbb{R} \) being differentiable with bounded first derivatives, if \( \Sigma \in \mathbb{R}^{d \times d} \) is symmetric and positive definite, there is a solution \( f : \mathbb{R}^d \to \mathbb{R} \) to the equation

\[
\nabla^t \Sigma \nabla f(w) - w^t \nabla f(w) = g(w) - \mathbb{E} g \left( \Sigma^{-1/2} Z \right),
\]

(3.3)

which holds for every \( w \in \mathbb{R}^d \). If, in addition, \( g \) is \( n \) times differentiable, there is a solution \( f \) which is also \( n \) times differentiable and one has for every \( k = 1, \ldots, n \) the bound

\[
\frac{\partial^k f(w)}{\prod_{j=1}^k \partial w_{ij}} \leq \frac{1}{k} \frac{\partial^k g(w)}{\prod_{j=1}^k \partial w_{ij}}
\]

(3.4)

for every \( w \in \mathbb{R}^d \). We will apply Theorem 2.1 in [19]:

**Theorem 3.1.** (Reinert, Röllin: 2009)

Assume that \((W, W')\) is an exchangeable pair of \( \mathbb{R}^d \)-valued random vectors such that

\[
\mathbb{E}[W] = 0, \quad \mathbb{E}[WW^t] = \Sigma,
\]

with \( \Sigma \in \mathbb{R}^{d \times d} \) symmetric and positive definite. If \((W, W')\) satisfies (1.7) for an invertible matrix \( \Lambda \) and a \( \sigma(W) \)-measurable random vector \( R \) and if \( Z \) has \( d \)-dimensional standard normal distribution, we have for every three times differentiable function \( g \),

\[
\left| \mathbb{E} g(W) - \mathbb{E} g \left( \Sigma^{-1/2} Z \right) \right| \leq \frac{|g|^2}{4} A + \frac{|g|^3}{12} B + \left( |g|_1 + \frac{1}{2} d \| \Sigma \|^{1/2} |g|_2 \right) C,
\]

(3.5)

where, with \( \lambda^{(i)} := d \sum_{m=1}^d \left| (\Lambda^{-1})_{m,i} \right| \),

\[
A = \sum_{i,j=1}^d \lambda^{(i)} \sqrt{V \left[ \mathbb{E}[ (W_i' - W_i)(W_j' - W_j) \mid W] \right]},
\]

\[
B = \sum_{i,j,k=1}^d \lambda^{(i)} \mathbb{E} \left[ (W_i' - W_i)(W_j' - W_j)(W_k' - W_k) \right],
\]

(3.6)

\[
C = \sum_{i=1}^d \lambda^{(i)} \sqrt{V[R_i]}.
\]

The advantage of Stein’s method is that the bounds to a multivariate normal distribution reduce to the computation of, or bounds on, low order moments, here bounds on the absolute third moments, on a conditional variance and on the variance of the remainder term. Such variance computations may be difficult, but we will get rates of convergence at the same time.
In the same context as in [19] we can show the following Theorem, presenting bounds in Kolmogorov distance. Our development differs from Reinert and Röllin, as we use the relationship to the bounds in [21]. We obtain a bound of order $\log(n) \cdot n^{-1/2}$ assuming some boundedness, improving Corollary 3.1 in [19].

**Theorem 3.2.** Let $(W, W')$ be an exchangeable pair with $E[WW'] = \Sigma$. Again we assume that $(W', W)$ satisfies (1.7) for an invertible matrix $\Lambda$ and a $\sigma(W)$-measurable random vector $R$ and additionally $|W' - W| \leq A$. Then,

$$
\sup_{g \in \mathcal{G}} \left| E[g(W) - E[g(Z)] \right| \leq C \left[ \log(n) A_1 + (\log(n) \|\Sigma\|^{1/2} + 1) A_2 
+ (1 + \log(n) \sum_{i=1}^q \|W_i\| + a) A^3 A_3 
+ a A^3 \left( \frac{1}{A^2} + A_3 \right) \right]
$$

where, with $\lambda^{(i)} := \sum_{m=1}^q \left| (\Lambda^{-1})_{m,i} \right|$, 

$$
A_1 = \sum_{i,j=1}^q \lambda^{(i)} \sqrt{V \left[ E[ (W'_i - W_i)(W'_j - W_j) \mid W ] \right]}, 
$$

$$
A_2 = \sum_{i=1}^q \lambda^{(i)} \sqrt{E \left[ R_i^2 \right]}, 
$$

$$
A_3 = \sum_{i=1}^q \lambda^{(i)}, 
$$

$C$ denotes a constant and $a > 1$ is taken from the conditions on $\mathcal{G}$.

**Proof.** To prove the Theorem, first we assume that $\Sigma = \text{Id}$. Throughout the proof we write $C$ for universal constants, not necessarily the same at each occurrence. First we consider the multivariate Stein equation deduced from (3.3) with $\Sigma = \text{Id}$ given by

$$
\nabla^t \nabla f(w) - w^t \cdot \nabla f(w) = g(w) - E[g(Z)]. \tag{3.7}
$$

For $g \in \mathcal{G}$ define the following smoothing

$$
g_t(x) = \int_{\mathbb{R}^q} g \left( \sqrt{t} z + \sqrt{1-t} x \right) \Phi(z) dz.
$$

For $g_t$ (3.7) is solved by the function

$$
f_t(x) = -\frac{1}{2} \int_{\frac{1}{t}}^1 \left( g_s(x) - E[g(Z)] \right) \frac{ds}{1-s},
$$
Abbreviating $E$ where see [16]. Again by [16], we have that for $|g| \leq 1$, there exists a constant $C$, depending only on the dimension $q$, such that in the notion (3.1)

$$|f_t|_1 \leq C$$  \hfill (3.8)

$$|f_t|_2 \leq C \log(t^{-1}).$$  \hfill (3.9)

According to [16] (see also [19], Lemma A.1) there is also a constant $C > 0$, depending on $q$, such that for all $t \in (0, 1)$:

$$\sup\{||Eg(W) - Eg(Z)| : g \in \mathcal{G}\} \leq C \cdot \delta_t + a \sqrt{t},$$  \hfill (3.10)

where $a > 1$ is the constant of $\mathcal{G}$ and

$$\delta_t := \sup\{||Eg_t(W) - Eg_t(Z)| : g \in \mathcal{G}\}.$$

Thus, it remains to estimate $\delta_t$. Using exchangeability and the linearity condition we have

$$0 = \frac{1}{2} \mathbb{E}[(W' - W)^t \Lambda^{-t}(\nabla f_t(W') + \nabla f_t(W))]$$

$$= \mathbb{E}[(W' - W)^t \Lambda^{-t}\nabla f_t(W)] + \frac{1}{2} \mathbb{E}[(W' - W)^t \Lambda^{-t}(\nabla f_t(W') - \nabla f_t(W))]$$

$$= \mathbb{E}[R^t \Lambda^{-t}\nabla f_t(W)] - \mathbb{E}[W^t \nabla f_t(W)] + \frac{1}{2} \mathbb{E}[(W' - W)^t \Lambda^{-t}(\nabla f_t(W') - \nabla f_t(W))].$$

Abbreviating $f_j^{(1)} := \frac{\partial}{\partial x_j} f_t$ and $f_{i,j}^{(2)} := \frac{\partial^2}{\partial x_i \partial x_j} f_t$, etc. for a function $f_t$, we obtain

$$\mathbb{E}[W^t \nabla f_t(W)] = \frac{1}{2} \mathbb{E}[(W' - W)^t \Lambda^{-t}(\nabla f_t(W') - \nabla f_t(W))] + \mathbb{E}[R^t \Lambda^{-t}\nabla f_t(W)]$$

$$= \frac{1}{2} \sum_{m,i,j=1}^q (\Lambda^{-1})_{m,i} \mathbb{E}[(W'_i - W_i)(f_j^{(1)}(W') - f_j^{(1)}(W))]$$

$$+ \sum_{m,i,j=1}^q (\Lambda^{-1})_{m,i} \mathbb{E}[R_i f_j^{(1)}(W)].$$

Hence,

$$\mathbb{E}g_t(W) - \mathbb{E}g_t(Z)$$

$$= \sum_{i=1}^q \mathbb{E}[f_{i,i}^{(2)}(W)] - \frac{1}{2} \sum_{m,i,j=1}^q (\Lambda^{-1})_{m,i} \mathbb{E}[(W'_i - W_i)(f_j^{(1)}(W') - f_j^{(1)}(W))]$$

$$- \sum_{m,i,j=1}^q (\Lambda^{-1})_{m,i} \mathbb{E}[R_i f_j^{(1)}(W)]$$

$$= \frac{1}{2} \mathbb{E}[2 \sum_{i=1}^q f_{i,i}^{(2)}(W) - 2 \sum_{m,i,j=1}^q (\Lambda^{-1})_{m,i} R_i W_j f_{j,j}^{(2)}(W)$$

$$- \sum_{m,i,j=1}^q (\Lambda^{-1})_{m,i} (W'_i - W_i)(W'_j - W_j) f_{j,j}^{(2)}(W)]$$
\[ + \mathbb{E}\left[ \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} R_i W_j f^{(2)}_{j,j'} (W) - \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} R_i f^{(1)}_{j} (W) \right] \]
\[ - \frac{1}{2} \mathbb{E}\left[ \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} (W_i' - W_i) (f^{(1)}_{j} (W') - f^{(1)}_{j} (W)) \right] \]
\[ - \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} (W_i' - W_i) (W_j' - W_j) f^{(2)}_{j,j'} (W) \]
\[ =: J_1 + J_2 + J_3. \]

Using (3.9) and the fact that \( \mathbb{E}\left[ (W' - W)(W' - W)^t \right] = 2 \Lambda^{t} - 2 \mathbb{E}[WR^{t}] \) we get
\[ |J_1| = \left| \frac{1}{2} \mathbb{E}\left[ 2 \sum_{i=1}^{q} f^{(2)}_{i,i} (W) - 2 \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} R_i W_j f^{(2)}_{j,j'} (W) \right] \right| \]
\[ \leq C \log(t^{-1}) \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} \mathbb{E}\left[ (2(\Lambda)_{m,i} - 2 R_i W_j - \mathbb{E}\left[(W_i' - W_i)(W_j' - W_j)|W]\right] \right] \]
\[ \leq C \log(t^{-1}) \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} \sqrt{\mathbb{E}\left[ (2(\Lambda)_{m,i} - 2 R_i W_j - \mathbb{E}\left[(W_i' - W_i)(W_j' - W_j)|W]\right]^2 \right]} \]
\[ = C \log(t^{-1}) \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} \sqrt{\mathbb{E}\left[ (W_i' - W_i)(W_j' - W_j)|W]\right]. \]

Additionally, again using (3.8) and (3.9) and the fact that \( \mathbb{E}[WW'] = \Sigma \), we have
\[ |J_2| = \left| \mathbb{E}\left[ \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} R_i W_j f^{(2)}_{j,j'} (W) - \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} R_i f^{(1)}_{j} (W) \right] \right| \]
\[ \leq C \log(t^{-1}) \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} \mathbb{E}\left| R_i W_j \right| - C \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} \mathbb{E}\left| R_i \right| \]
\[ \leq C \log(t^{-1}) \sum_{m,i=1}^{q} (\Lambda^{-1})_{m,i} \sqrt{\mathbb{E}\left| R_i \right|^2}. \]

The estimation of \( J_3 \) is a bit more involved. We have
\[ |2J_3| = \left| \sum_{m,i,j=1}^{q} \mathbb{E}\left[ (\Lambda^{-1})_{m,i} (W_i' - W_i) (f^{(1)}_{j} (W') - f^{(1)}_{j} (W)) \right] \right| + (\Lambda^{-1})_{m,i} (W_i' - W_i) (W_j' - W_j) f^{(2)}_{j,j'} (W)] \]
\[
\leq \sum_{m,i,j=1}^{q} (\Lambda^{-1})_{m,i} \mathbb{E} \left[ (W'_i - W_i) (f_j^{(1)}(W') - f_j^{(1)}(W) - (W'_j - W_j) f_{j,j}^{(2)}(W)) \right] \\
\leq \sum_{m,i=1}^{q} (\Lambda^{-1})_{m,i} \mathbb{E} \left[ A^3 \sum_{j,k=1}^{q} \int_{0}^{1} (1 - \tau) f_{j,j,k}^{(3)}(W' - \tau(W' - W)) d\tau \right].
\]

We abbreviate \( M := W' - \tau(W' - W) \). It is important to notice that we can use (3.7) to obtain for \( w \in \mathbb{R}^q \)

\[
\sum_{j,k=1}^{q} f_{j,j,k}^{(3)}(w) + \sum_{j,k=1}^{q} w_j f_{j,j,k}^{(2)}(w) + \sum_{k=1}^{q} f_k^{(1)}(w) = \sum_{k=1}^{q} g_k^{(1)}(w).
\]

Additionally we notice that

\[
\int_{0}^{1} (1 - \tau) d\tau = \frac{1}{2} \quad \text{and} \quad \left| \int_{0}^{1} (1 - \tau) M_i d\tau \right| \leq |W'_i| + |W_i|.
\]

Thus,

\[
|2J_3| \leq \sum_{m,i=1}^{q} (\Lambda^{-1})_{m,i} A^3 \left| \mathbb{E} \left[ \int_{0}^{1} (1 - \tau) [ - \sum_{k=1}^{q} f_k^{(1)}(M) - \sum_{j,k=1}^{q} M_j f_{j,j,k}^{(2)}(M) + \sum_{k=1}^{q} g_k^{(1)}(M)] d\tau \right] \right|
\]

\[
\leq \sum_{m,i=1}^{q} (\Lambda^{-1})_{m,i} \left( A^3 C + A^3 \sum_{j=1}^{q} C \log(t^{-1}) \mathbb{E} [ |W'_j| + |W_j| ] \right.
\]

\[
+ \left. \left| A^3 \sum_{k=1}^{q} \mathbb{E} \left[ \int_{0}^{1} (1 - \tau) g_k^{(1)}(M) d\tau \right] \right| \right) =: T
\]

\[
\leq \sum_{m,i=1}^{q} (\Lambda^{-1})_{m,i} \left[ A^3 C \left( 1 + \log(t^{-1}) \sum_{j=1}^{q} \mathbb{E} |W_j| \right) + T \right].
\]

By partial integration we obtain

\[
\sum_{k=1}^{q} g_k^{(1)}(w) = \sum_{k=1}^{q} \sqrt{1-t} \int_{\mathbb{R}^q} g \left( \sqrt{t}z + \sqrt{1-t}w \right) \Phi_k^{(1)}(z) dz.
\]

Keeping in mind that \( \sum_{k=1}^{q} \int_{\mathbb{R}^q} \Phi_k^{(1)}(z) dz = 0 \) and using the definitions (2.1), (2.2) and (2.3) we obtain

\[
T \leq \frac{\sqrt{1-t}}{\sqrt{t}} A^3 \sum_{k=1}^{q} \mathbb{E} \left[ \int_{\mathbb{R}^q} (1 - \tau) g(\sqrt{t}z + \sqrt{1-t}M) \Phi_k^{(1)}(z) dz d\tau \right].
\]
For general $\Sigma$ we can standardize $\delta$ with 

$$\delta = \text{Id}$$

for the $\Sigma = ((\Sigma - 1/2) - 1)\leq B =: \sup x \geq 2t$ for the choice above, then by enlarging $\Sigma$ case, but with the standardized variables.

Thus, combining the estimates of $g$ for the $\Sigma = Id$ if $t > 1$ for the choice above, then by enlarging $C$ as necessary, the theorem is trivial. For general $\Sigma$ we can standardize $Y = \Sigma^{-1/2}W$. With the conditions of $G$ we have that $g(\Sigma^{-1/2}x) \in G$. Hence the bounds (3.8) and (3.9) can be applied. The proof now continues as for the $\Sigma = Id$ case, but with the standardized variables.

$$\begin{align*}
&\frac{1}{\sqrt{t}} A^3 \sum_{k=1}^q \mathbb{E}\left[ \int_0^1 \int_{\mathbb{R}^q} (1 - \tau) |g(\sqrt{t}z + \sqrt{1 - tM}) - g(\sqrt{1 - tM})\Phi_k^{(1)}(z)dzd\tau| \right] \\
&\leq \frac{1 - t}{2\sqrt{t}} A^3 \sum_{k=1}^q \mathbb{E}\left[ \int_{\mathbb{R}^q} |g(\sqrt{1 - tW} - g(\sqrt{1 - tW})\Phi_k^{(1)}(z)|dz \\
&\leq \frac{1 - t}{2\sqrt{t}} A^3 \sum_{k=1}^q \mathbb{E}\left[ \int_{\mathbb{R}^q} \tilde{g}(W, A, t, z) - \tilde{g}(Z, A, t, z)\Phi_k^{(1)}(z)dz \\
&\leq \frac{1 - t}{2\sqrt{t}} A^3 \sum_{k=1}^q \mathbb{E}\left[ \int_{\mathbb{R}^q} \tilde{g}(Z, A, t, z)\Phi_k^{(1)}(z)dz \\
&=: B_1 + B_2.
\end{align*}$$

$\delta := \sup \{ |\mathbb{E}[g(W)] - \mathbb{E}[g(Z)] : g \in G \}$ we obtain

$$B_1 \leq \frac{\sqrt{1 - t}}{2\sqrt{t}} A^3 \delta \leq \frac{C \delta}{\sqrt{t}} A^3.$$ 

Furthermore by using the conditions established for the function classes $G$

$$B_2 \leq \frac{\sqrt{1 - t}}{2\sqrt{t}} A^3 \cdot a \left( \sqrt{1 - t}A + \sum_{k=1}^q \int_{\mathbb{R}^q} \sqrt{t}|z||\Phi_k^{(1)}(z)dz \right)$$

$$= a \frac{\sqrt{1 - t}}{2\sqrt{t}} A^4 + aC \frac{\sqrt{1 - t}}{2} A^3$$

$$\leq aCA^3 \left( \frac{A}{\sqrt{t}} + 1 \right).$$

Thus, combining the estimates of $J_1, J_2$ and $J_3$ with (3.10) we have

$$\delta \leq C \left[ \log(t^{-1})A_1 + \log(t^{-1})A_2 + A_3^3 A_3 \left( 1 + \log(t^{-1}) \sum_{i=1}^q \mathbb{E}|W_i| \right) \right]$$

$$+ \frac{\delta}{\sqrt{t}} A^3 A_3 + aA^3 A_3 \left( \frac{A}{\sqrt{t}} + 1 \right) + a\sqrt{t}.$$ 

Setting $\sqrt{t} = 2CA^3 A_4$, provided it is less than 1, simple manipulations yield the result for $\Sigma = Id$. If $t > 1$ for the choice above, then by enlarging $C$ as necessary, the theorem is trivial. For general $\Sigma$ we can standardize $Y = \Sigma^{-1/2}W$. With the conditions of $G$ we have that $g(\Sigma^{-1/2}x) \in G$. Hence the bounds (3.8) and (3.9) can be applied. The proof now continues as for the $\Sigma = Id$ case, but with the standardized variables. \qed
In [10], as well as in [6], Stein’s method of exchangeable pairs was introduced for non-normal distributional approximations. They consider a class of densities on $\mathbb{R}$ which was originally introduced in [25]. Let $p$ be a regular, strictly positive density on the interval $I = [a, b]$. We suppose that this density has a derivative $p'$ that is also regular on $I$ with countable many sign changes. Furthermore $p'$ should be continuous at the sign changes and $\int_I p(x) \log(p(x)) dx < \infty$. Additionally we assume that
\[
\psi(x) := \frac{p'(x)}{p(x)}
\]
(3.11)
is regular. A density $p$ fulfilling these conditions will be called nice. Now a random variable $Z$ is distributed according to $p$ if and only if
\[
\mathbb{E} [f'(Z) + \psi(Z)f(Z)] = 0
\]
for a suitable class of functions. The corresponding Stein-identity is
\[
f'(x) + \psi(x)f(x) = g(x) - P(g),
\]
(3.12)
where $g$ is a measurable function for which $\int_I |g(x)| p(x) dx < \infty$, $P(x) := \int_x^- \int_{\infty}^- p(y) dy$ and $P(g) := \int_I g(y)p(y) dy$. For the proof of Theorem 2.4 we will apply Theorem 2.4 and Theorem 2.5 in [10]:

**Theorem 3.3.** Let $p$ be a nice density. Let $p_W$ be a probability distribution such that a random variable $Z_W$ is distributed according to $p_W$ if and only if $\mathbb{E}([W\psi(W)]f'(Z_W)+\psi(Z_W)f(Z_W)) = 0$ for a suitably chosen class of functions and with $\psi$ as in (3.11). Let $(W, W')$ be an exchangeable pair of real-valued random variables such that
\[
\mathbb{E} [W' - W | W] = \lambda \psi(W) - R(W)
\]
(3.13)
for some random variable $R = R(W)$, $0 < \lambda < 1$ and $\psi$ as in (3.11).

(1): Let us assume that for any absolutely continuous function $g$ the solution $f_g$ of (3.12) satisfies
\[
\|f_g\| \leq c_1 \|g'\|, \quad \|f_g'\| \leq c_2 \|g'\| \quad \text{and} \quad \|f_g''\| \leq c_3 \|g'\|.
\]
Then for any uniformly Lipschitz function $g$, we obtain $|\mathbb{E} [g(W)] - \mathbb{E} [g(Z_W)]| \leq \delta \|g'\|$ with
\[
\delta := \frac{c_2}{2\lambda} \left( \sqrt{\mathbb{E} ((W - W')^2|W])} \right)^{1/2} + \frac{c_4}{4\lambda} \mathbb{E} |W - W'|^3 + \frac{c_1 + c_2 \sqrt{\mathbb{E} (W^2)}}{\lambda} \sqrt{\mathbb{E} (R^2)}.
\]
(3.14)

(2): Let us assume that for any function $g(x) := 1_{\{x \leq z\}}(x), z \in \mathbb{R}$, the solution $f_z$ of (3.12) satisfies
\[
|f_z(x)| \leq d_1, \quad |f_z'(x)| \leq d_2 \quad \text{and} \quad |f_z'(x) - f_z'(y)| \leq d_3
\]
and
\[
|\psi(x) f_z(x)'| = \left| \left( \frac{p'(x)}{p(x)} f_z(x)' \right) \right| \leq d_4
\]
(3.15)
for all real \( x \) and \( y \), where \( d_1, d_2, d_3 \) and \( d_4 \) are constants. Then we obtain for any \( A > 0 \)
\[
\sup_{t \in \mathbb{R}} \left| P(W \leq t) - \int_{-\infty}^{t} p_W(t) \, dt \right| \leq \frac{d_2}{2\lambda} \left( \sqrt{\mathbb{E}[(W' - W)^2|W|]} \right)^{1/2} \\
+ (d_1 + d_2 \sqrt{\mathbb{E}(W^2)} + \frac{3}{2} A) \sqrt{\frac{\mathbb{E}(R^2)}{\lambda}} + \frac{1}{\lambda} \left( \frac{d_4 A^3}{4} \right) \\
+ \frac{3A}{2} \mathbb{E}(|\psi(W)|) + \frac{d_3}{2\lambda} \mathbb{E}((W - W')^2 1_{|W - W'| \geq A}).
\]

4. Auxiliary results

Let us fix the convenient that for \( k, t \in \{1, \ldots, q\} \) we will always write \( \sum_{k \neq m} \) instead of \( \sum_{k=1}^{q} \) and so on. For the proof of the multivariate normal approximations we will apply Theorem 3.1 and Theorem 3.2, respectively. In the introduction we already presented the construction of the exchangeable pair for \( W \), the rescaled empirical spin vector of the Curie-Weiss-Potts model. We have already seen in (1.10), that this construction will lead us to \( G_{\beta,h} \), defined in (1.5). We collect some further results on this function. First we state a result on the structure of the minimizers of \( G_{\beta,h} \), determined in several papers, collected in [15]:

**Proposition 4.1.** Let \( \beta, h \geq 0 \) and let \( x \) be a global minimum of \( G_{\beta,h} \). Then:

1. The vector \( x \) has the coordinate \( \min(x_i) \) repeated \( q - 1 \) times at least.
2. If \( h > 0 \), then \( x_1 > x_i \), for all \( i \in \{2, \ldots, q\} \).
3. The inequality \( \min(x_i) > 0 \) holds.
4. For any \( q \geq 3 \) and any \((\beta, h)\), or \( q = 2 \) and \((\beta, h) \neq (\beta_c, 0)\), where \( \beta_c \) denotes the critical temperature, one has \( \min(x_i) < 1/\beta \).

An important identity is the following simple statement:

**Lemma 4.2.** For \( u \in \mathbb{R}^q \), we obtain
\[
\frac{\exp(\beta u_m + h\delta_{m,1})}{\sum_{k=1}^{q} \exp(\beta u_k + h\delta_{k,1})} = u_m - \frac{1}{\beta} \frac{\partial}{\partial u_m} G_{\beta,h}(u).
\]

**Proof.** Direct calculation yields:
\[
\frac{\partial}{\partial u_m} G_{\beta,h}(u) = \frac{\partial}{\partial u_m} \left( \frac{\beta}{2} \langle u, u \rangle - \log \left( \sum_{k=1}^{q} \exp(\beta u_k + h\delta_{k,1}) \right) \right) = \beta u_m - \beta \frac{\exp(\beta u_m + h\delta_{m,1})}{\sum_{k=1}^{q} \exp(\beta u_k + h\delta_{k,1})}.
\]
Rearranging the equality gives the result. \(\square\)
Lemma 4.3. For arbitrary $i \in \{1, \ldots, q\}$ we have

\[
P_{\beta, h, n}(\sigma_j = i \mid (\sigma_t)_{t \neq j}) = \begin{cases} \frac{\sum_{k=1}^{\sigma} \exp(\beta m_{i,k}(\sigma))}{\sum_{k=1}^{\sigma} \exp(\beta m_{j,k}(\sigma) + h \delta_{k,1})}, & i \in \{2, \ldots, q\}; \\
\frac{\sum_{k=1}^{\sigma} \exp(\beta m_{i,k}(\sigma) + h)}{\sum_{k=1}^{\sigma} \exp(\beta m_{j,k}(\sigma) + h \delta_{k,1})}, & i = 1. \end{cases}
\]

Proof. For $x_1, \ldots, x_n \in \{1, \ldots, q\}$ we have:

\[
P_{\beta, h, n}(\sigma_j = i \mid (\sigma_t)_{t \neq j}) = \frac{P_{\beta, h, n}(\{\sigma_j = i\} \cap \{(\sigma_t)_{t \neq j}\})}{P_{\beta, h, n}(\{(\sigma_t)_{t \neq j}\})}.
\]

For any fixed $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ we obtain

\[
\frac{P_{\beta, h, n}(\sigma_1 = x_1, \ldots, \sigma_j = i, \ldots, \sigma_n = x_n)}{P_{\beta, h, n}(\sigma_1 = x_1, \ldots, \sigma_{j-1} = x_{j-1}, \sigma_{j+1} = x_{j+1}, \ldots, \sigma_n = x_n)}
= \frac{Z_{\beta, h, n}^{-1} \exp \left( \frac{\beta}{2n} \sum_{t \neq j} x_{j,t} + \frac{\beta}{2n} \sum_{l \neq j} \delta_{x_{j,t}} + \frac{\beta}{n} \sum_{l \neq j} \delta_{x_{t,1}} + h \sum_{l \neq j} \delta_{x_{t,1}} + h \delta_{t,1} \right)}{
\sum_{k=1}^{q} Z_{\beta, h, n}^{-1} \exp \left( \frac{\beta}{2n} \sum_{l \neq j} x_{j,l} + \frac{\beta}{2n} \sum_{l \neq j} \delta_{x_{j,l}} + \frac{\beta}{n} \sum_{l \neq j} \delta_{x_{l,k}} + h \sum_{l \neq j} \delta_{x_{t,1}} + h \delta_{k,1} \right)}.
\]

Cancelling equivalent terms in numerator and denominator and finally distinguishing between $i = 1$ and $i \neq 1$ yields the result.

In the case $h = 0$ in [13] Proposition 2.2 it is proved that the Hessian $D^2 G_{\beta,0}(x_0)$ of $G_{\beta,0}$ is positive definite if $x_0$ is a global minimum point, and hence invertible. In [26] Lemma 2 it is stated that $D^2 G_{\beta,h}(x_0)$ is positive definite for any $\beta > 0$ and $h \geq 0$, if $x_0$ is a global minimum point. However this result is not correct. The non-degeneracy of $G_{\beta,h}$ at its minimum points for any $(\beta, h) \neq (\beta_0, h_0)$ is stated next and will be proved in the Appendix.

Lemma 4.4. For all $q > 2$ let $x_s \in \mathbb{R}^q$ denote a global minimum point of $G_{\beta,h}$. Then, if $(\beta, h) \neq (\beta_0, h_0)$, $\beta$ never takes one of the values $\{\frac{1}{x_{s,q}}, \frac{1}{x_{s,1}x_{s,q}}\}$ (implying that $D^2 G_{\beta,h}(x_s)$ is positive definite for any $(\beta, h) \neq (\beta_0, h_0)$).

For the rescaled empirical spin vector of the Curie-Weiss-Potts model, appearing in Theorems 2.1, 2.2 and 2.3, we can bound higher order moments as follows:
Lemma 4.5. For \( W = (W_1, \ldots, W_q) \) as in Theorems 2.1, 2.2 and 2.3 (the \( W^{(i)} \)) we obtain that for any \( l \in \mathbb{N} \) and \( j \in \{1, \ldots, q\} \)

\[
\mathbb{E}|W_j^l| \leq \text{const.}(l), \quad \mathbb{E}|(W_j^{(i)})^l| \leq \text{const.}(l).
\]

Proof. We consider a well known transformation, sometimes called the Hubbard-Stratonovich transformation, expressing the distribution of \( L_n \) in the Curie-Weiss-Potts model in terms of \( G_{\beta,h} \). For \( \beta > 0 \) we pick a random vector \( Y \) in a way that \( L(Y) \) equals a \( q \)-dimensional centered Gaussian vector with covariance matrix \( \beta^{-1}I_q \) and \( Y \) is chosen to be independent from \( N \). \( I_q \) denotes the \( q \times q \) identity matrix. According to a simple adaption of Lemma 3.2 in [13], for any point \( m \in \mathbb{R}^q \) and \( \gamma \in \mathbb{R} \) and any \( n \in \mathbb{N} \) we have

\[
\mathcal{L} \left( \frac{Y}{n^{1/2-\gamma}} + \frac{n(N/n - m)}{n^{1-\gamma}} \right) = \exp \left[ -nG_{\beta,h} \left( m + \frac{y}{n^\gamma} \right) \right] dy \left( \int_{\mathbb{R}^q} \exp \left[ -nG_{\beta,h} \left( m + \frac{y}{n^\gamma} \right) \right] dy \right)^{-1}.
\]

Lemma 3.2 in [13] presented this identity only for \( h = 0 \). The calculations for any \( h \neq 0 \) are omitted. Applying this Lemma for \( \gamma = \frac{1}{2} \) and \( m = x_0 \) (or any other minimum point of \( G_{\beta,h} \)) does not change the finiteness of any of the moments of the \( W_i \). Thus, the new measure has the density

\[
\exp \left[ -nG_{\beta,h} \left( x_0 + \frac{y}{n^{1/2}} \right) \right] dy \left( \int_{\mathbb{R}^q} \exp \left[ -nG_{\beta,h} \left( x_0 + \frac{y}{n^{1/2}} \right) \right] dy \right)^{-1}.
\]

Using second order multivariate Taylor expansion of \( G_{\beta,h} \) and the fact that \( x_0 \) is a global minimum point of \( G_{\beta,h} \) we see that the density of the new measure with respect to Lebesgue measure is given by \( \text{const.} \exp \left[ -\frac{1}{2} \langle y, D^2G_{\beta,h}(x_0) y \rangle \right] \) (up to negligible terms). With Lemma 4.4 we know that for any \( (\beta, h) \neq (\beta_0, h_0) \) the Hessian is positive definite, if \( x_0 \) is a global minimum point. This fact combined with the transformation of integrals yields that a measure with this density has moments of any finite order. □

For the random variables \( T \) and \( V \) in Theorems 2.4 and 2.6, we can bound higher order moments as well:

Lemma 4.6. Consider the extremity \( (\beta, h) = (\beta_0, h_0) \). For \( T \) and \( V \) as in (2.5) we obtain that for any \( l \in \mathbb{N} \) and \( j \in \{1, \ldots, q\} \)

\[
\mathbb{E}|V_j^l| \leq \text{const.}(l), \quad \mathbb{E}|T_j^l| \leq \text{const.}(l).
\]

Proof. Remark that with \( V \in \mathcal{M} \cap u^\perp \) we obtain \( V_1 = 0 \). Hence \( W_1 = (1 - q)n^{1/4}T \). Therefore

\[
V = \frac{1}{n^{1/2}}(N - nx - n^{3/4}Tu) = W - n^{1/4}Tu = (0, W_2 + \frac{1}{(q-1)}W_1, \ldots, W_q + \frac{1}{(q-1)}W_1).
\]
Since $\bar{W} := (W_1, \ldots, W_q) \in \mathcal{M}$, we have $W_1 = -\sum_{k=2}^{q} W_k$. We try to check that $\bar{V} := (V_2, \ldots, V_q)$ has finite moments. Thus it suffices to check, that $(W_2, \ldots, W_q)$ has finite moments. Now we define $G_{q-1}(x), x \in \mathbb{R}^{q-1}$, to be the restriction of $G_{\beta_0, h_0}$ on the last $q-1$ coordinates (the first coordinate will be fixed to $1/2$ in the sequel). Again we apply the Hubbard-Stratonovich transformation introduced in the proof of Lemma 4.3. We choose a $q-1$ dimensional Gaussian vector $Y$ with covariance matrix $\beta_0^{-1}\text{Id}_{q-1}$ and independent of $\bar{W}$. With $x = (1/2(q-1), \ldots, 1/2(q-1)) \in \mathbb{R}^{q-1}$ we have that the law of $Y + \bar{W}$ has the density

$$
\exp \left[ -nG_{q-1} \left( x + \frac{y}{n^{1/2}} \right) \right] \frac{dy}{\int_{\mathbb{R}^{q-1}} \exp \left[ -nG_{q-1} \left( x + \frac{y}{n^{1/2}} \right) \right] dy}^{-1}.
$$

Using second order multivariate Taylor expansion of $G_{q-1}$ and the fact that $(\nabla G_{q-1})(x) = 0$ ($(1/2, x) \in \mathbb{R}^{q}$ is a global minimum point of $G_{\beta_0, h_0}$), we see that the density of the new measure with respect to Lebesgue measure is given by const. $\exp \left[ -\frac{1}{2} \langle y, D^2G_{q-1}(x) y \rangle \right]$ (up to negligible terms). Using the formulas for the second partial derivatives of $G_{\beta_0, h_0}$, see Remark 6.3 in the Appendix, we obtain that

$$D^2G_{q-1}(x) = \frac{4}{q^2} (\mathbb{I}_{q-1} + \text{Id}_{q-1}(q-1)(q-2)),
$$

where $\mathbb{I}_{q-1}$ denotes the $(q-1) \times (q-1)$ matrix with all entries equal to 1. It is an immediate computation that

$$
det(D^2G_{q-1}(x)) = \left( \frac{4(q-1)(q-2)}{q^2} \right)^{q-2} \left( \frac{4(q-1)(q-2)}{q^2} + \frac{4(q-1)}{q^2} \right),
$$

which shows the invertibility of $D^2G_{q-1}(x)$ for any $q \geq 3$. Thus $D^2G_{q-1}(x)$ is positive definite. This fact combined with the transformation of integrals yields that a measure with this density has moments of any finite order.

For $(1-q)T = n^{-1/4}W_1$ we apply the Hubbard-Stratonovich transform with $\gamma = 1/4$. Take a Gaussian random variable with expectation zero and variance $\beta_0^{-1}$, independent of $W_1$. The distribution of $n^{-1/4}Y + T$ has a density proportional to $\exp(-nG_1(c_q1/2 + y/n^{1/4}))$ with some constant $c_q$ only depending on $q$ and $G_1$ being the restriction of $G_{\beta_0, h_0}$ to the first component. A fourth order Taylor expansion similar to (6.4) will give $G_1(x + t) = G_1(x) + \frac{1}{4!}G_1^{(4)}(x + \alpha t)t^4$ for some $\alpha \in (0, 1)$. Hence we conclude with a Lebesgue-density given by const. $\exp(-y^4)$, a measure, which has moments of any finite order. We omit the details. \hfill $\square$

5. PROOFS OF THE THEOREMS

Proof of Theorem 2.1: Our goal is to apply Theorem 3.1. First, given $W$, we construct a coupling $W'$ and will calculate $\Lambda$ and $R$ to get the approximative regression identity (1.7). We will first of all deal with the case $h = 0$. Hence by Theorem 1.1 we have $\beta < \beta_c$ and
\( x_0 = (1/q, \ldots, 1/q) \) being the unique minimum point of \( G_{\beta, 0} \). By the construction given in the introduction, applying Lemma 4.3 and Lemma 4.2 we obtained for any \( i = 1, \ldots, q \):

\[
 \mathbb{E}[W'_i - W_i | \mathcal{F}] = - \frac{1}{n} W_i - \frac{x_{0,i}}{\sqrt{n}} + R_n^{(1)}(i) + \frac{1}{\sqrt{n}} \left( m_i(\sigma) - \frac{1}{\beta} \frac{\partial}{\partial u_i} G_{\beta, 0}(m(\sigma)) \right)
\]

\[
= - \frac{1}{\sqrt{n}} \frac{1}{\beta} \frac{\partial}{\partial u_i} G_{\beta, 0}(m(\sigma)) + R_n^{(1)}(i)
\]

(5.1)

with

\[
R_n^{(1)}(i) := \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left[ \frac{\exp(\beta m_{i,j}(\sigma))}{\sqrt{n}} - \frac{\exp(\beta m_i(\sigma))}{\sqrt{n}} \right] = \frac{1}{\sqrt{n}} \frac{\partial}{\partial u_i} G_{\beta, 0}(m(\sigma))
\]

(5.2)

where \( m_i(\sigma) \) and \( m_{i,j}(\sigma) \) are defined as in (4.1). We have used

\[
m_i(\sigma) - x_{0,i} = \frac{W_i}{\sqrt{n}}
\]

(5.3)

Now we apply (6.2) (see Appendix) to the first summand in (5.1). Since \( x_0 \) is a global minimum of \( G_{\beta, 0} \) we have \( \left( \frac{\partial}{\partial u_i} G_{\beta, 0} \right)(x_0) = 0 \). Hence the first summand in (5.1) is equal to

\[
- \frac{1}{\beta n} \left( \frac{\partial^2}{\partial u_i^2} G_{\beta, 0} \right)(x_0) W_i = - \frac{1}{\beta n} \sum_{k \neq i} \left( \frac{\partial^2}{\partial u_i \partial u_k} G_{\beta, 0} \right)(x_0) W_k + R_n^{(2)}(i)
\]

with

\[
R_n^{(2)}(i) := \mathcal{O} \left( \frac{1}{\sqrt{n}} \left( \frac{W_i}{\sqrt{n}} \right)^2 \right) = \sum_{k \neq i} \mathcal{O} \left( \frac{1}{\sqrt{n}} \frac{W_i W_k}{\sqrt{n}} \right) = \sum_{k \neq i} \mathcal{O} \left( \frac{1}{\sqrt{n}} \frac{W_k W_i}{\sqrt{n}} \right).
\]

(5.4)

Summarizing with \( R(i) := R_n^{(1)}(i) + R_n^{(2)}(i) \) we have

\[
\mathbb{E}[W'_i - W_i | \mathcal{F}] = - \frac{1}{\beta n} \left( \frac{\partial^2}{\partial u_i^2} G_{\beta, 0} \right)(x_0) W_i - \frac{1}{\beta n} \sum_{k \neq i} \left( \frac{\partial^2}{\partial u_i \partial u_k} G_{\beta, 0} \right)(x_0) W_k + R(i)
\]

\[
= - \frac{1}{\beta n} \langle [D^2 G_{\beta, 0}(x_0)]_i, W \rangle + R(i),
\]

where \( \langle \cdot, \cdot \rangle \) denotes the Euclidean scalar-product and \( [D^2 G_{\beta, 0}(x_0)]_i \) the \( i \)-th row of the matrix \( D^2 G_{\beta, 0}(x_0) \). We obtain

\[
\mathbb{E}[W' - W | \mathcal{F}] = - \frac{1}{\beta n} [D^2 G_{\beta, 0}(x_0)] W + R(W)
\]

(5.5)

with \( R(W) = (R(1), \ldots, R(q)) \). We define \( \Lambda = \frac{1}{\beta n} [D^2 G_{\beta, 0}(x_0)] \). With [13, Proposition 2.2], \( D^2 G_{\beta, 0}(\nu) \) is positive definite for any \( \beta > 0 \) and any global minimum point \( \nu \) and therefore \( \Lambda \) is invertible (alternatively one easily sees that \( \Lambda \) is a matrix of the form given in Lemma 6.3 and the determinant is \( \frac{1}{n^q} (1 - \beta/q)^{a-2} (1 - \beta/q) \) which is non-zero because \( \beta_c < q \) with \( \beta_c \) given in (1.2), and therefore \( \beta \neq q \), see Lemma 6.1). Hence (1.7) is fulfilled and we are able to apply Theorem 3.1. In order to calculate the bound given there we need to estimate \( \lambda^{(i)} \) as well as the order of the terms \( A, B \) and \( C \). Note that often in an application of Theorem 3.1 it
might be tedious to calculate \( \Lambda \) (and \( \Sigma \)) and it is not clear whether the calculations have been carried out correctly. In Remark 5.1 we will point out, that there is a nice heuristic in the Curie-Weiss-Potts model expecting \( \Lambda \) as it comes out.

Obviously we have \( \lambda^{(i)} = O(n) \). We continue by estimating \( C \) in Theorem 3.1. First we consider \( R_n^{(1)}(i) \) defined in (5.2):

\[
|R_n^{(1)}(i)| \leq \frac{1}{\sqrt{n}} \frac{1}{n} \sum_{j=1}^{n} \left| \frac{\exp(\beta m_{i,j}(\sigma))}{\sum_{k=1}^{q} \exp(\beta m_k(\sigma))} - \frac{\exp(\beta_m(\sigma))}{\sum_{k=1}^{q} \exp(\beta_m(\sigma))} \right|
\]
\[
\leq \frac{1}{\sqrt{n}} \frac{1}{n} \sum_{j=1}^{n} \sum_{k \neq i} \left| \exp(\beta m_{i,j}(\sigma) + \beta m_k(\sigma)) - \exp(\beta m_i(\sigma) + \beta m_k(\sigma)) \right|.
\]

Using the inequality
\[
|\exp(\alpha x) - \exp(\alpha y)| \leq \frac{|\alpha|}{2}(\exp(\alpha x) + \exp(\alpha y))|x-y|, \text{ for all } \alpha, x, y \in \mathbb{R},
\]
we obtain
\[
|R_n^{(1)}(i)| \leq \frac{1}{\sqrt{n}} \frac{1}{n} \beta e^{2\beta} \sum_{j=1}^{n} \sum_{k \neq i} |m_{i,j}(\sigma) + m_k(\sigma) - m_i(\sigma) - m_{k,j}(\sigma)|.
\]

Consider the first summand \( j = 1 \). In case \( \sigma_1 = i \), we have for all \( k \neq i \) that \( m_{k,1}(\sigma) = m_k(\sigma) \), and therefore
\[
\sum_{k \neq i} |m_{i,1}(\sigma) + m_k(\sigma) - m_i(\sigma) - m_{k,1}(\sigma)| = (q-1)\frac{\delta_{\sigma_1,i}}{n}.
\]

If \( \sigma_1 \neq i \), then there is a \( t \neq i \) with \( m_{t,1}(\sigma) \neq m_i(\sigma) \) and for all \( k \neq t \): \( m_{k,1}(\sigma) = m_k(\sigma) \). By similar observation we have
\[
\sum_{k \neq i} |m_{i,1}(\sigma) + m_k(\sigma) - m_i(\sigma) - m_{k,1}(\sigma)| \leq (q-1)\frac{\delta_{\sigma_1,t}}{n}.
\]

The same observation can be made for any other \( j \in \{1, \ldots, n\} \). With \( |\delta_{\sigma,j,i}| \leq 1 \) we get
\[
|R_n^{(1)}(i)| \leq \frac{1}{\sqrt{n}} \frac{1}{n} (q-1)\beta e^{2\beta} = O(n^{-3/2}).
\]

Since \( W \in \mathcal{M} \), see (2.4), we get \( \sum_{k \neq i} W_k = -W_i \). By Lemma 4.3 we know that \( \mathbb{E}[W_i^2] \leq \text{const.}(2) \) and therefore we obtain that \( \mathbb{E}[R_n^{(2)}(i)] \) in (5.4) is \( O(n^{-3/2}) \). Thus the Cauchy-Schwartz inequality yields \( \mathbb{E}[R(i)^2] = O(n^{-3}) \) for all \( i \in \{1, \ldots, q\} \). We have
\[
C = \sum_{i=1}^{q} \lambda^{(i)} \sqrt{\mathbb{E}[R_i^2]} = O(n^{-1/2}).
\]
The next thing we notice is that $|W'_i - W_i| = \frac{1}{\sqrt{n}}|Y'_i - Y_i| \leq \frac{1}{\sqrt{n}}$ for all $i$. Thus we easily obtain the bound $B = O(n^{-1/2})$. It remains to calculate and to estimate the conditional variance in $A$. This is a bit more involved. We have:

$$
\mathbb{E}[(W'_i - W_i)(W'_j - W_j) \mid \mathcal{F}] = \frac{1}{n^3} \sum_{t,k=1}^{n} Y_{k,i} Y_{t,j} + \frac{1}{n^3} \sum_{t,k=1}^{n} \mathbb{E}[Y'_{k,i} Y'_{t,j} \mid \mathcal{F}]
$$

$$
- \frac{2}{n^3} \sum_{t,k=1}^{n} Y_{k,i} \mathbb{E}[Y'_{t,j} \mid \mathcal{F}] =: A_1 + A_2 + A_3.
$$

Hence we have to bound the variances of these terms. By definition $\mathbb{V}[A_1] = \frac{1}{n^2} \mathbb{V}[m_i(\sigma) m_j(\sigma)]$. Now

$$
\mathbb{V}[m_i(\sigma) m_j(\sigma)] = \mathbb{V}(W_i W_j / n + W_i \sqrt{n} x_{0,j} + W_j \sqrt{n} x_{0,i}) 
$$

$$
\leq \text{const. max}(\frac{1}{n^2} \mathbb{V}(W_i W_j) - \frac{1}{n} \mathbb{V}(W_i)) \leq \text{const.} \left( \mathbb{E}[W_i^2 W_j^2] + n \mathbb{E}[W_i^2] \right)
$$

We make use of Lemma 4.5 to obtain $\mathbb{V}[m_i(\sigma) m_j(\sigma)] = O(1/n)$ and hence $\mathbb{V}[A_1] = O(n^{-3})$. Using a conditional version of Jensen’s inequality we have

$$
\mathbb{V}[A_2] \leq \mathbb{E}\left[ \mathbb{V}[\frac{1}{n^3} \sum_{t,k=1}^{n} Y'_{k,i} Y'_{t,j} \mid \mathcal{F}] \right] = \mathbb{E}\left[ \mathbb{V}[\frac{1}{n^3} \sum_{t,k=1}^{n} Y_{k,i} Y_{t,j} \mid \mathcal{F}] \right] = \mathbb{V}(\frac{1}{n^3} \sum_{t,k=1}^{n} Y_{k,i} Y_{t,j}).
$$

Hence $\mathbb{V}[A_2] = O(n^{-3})$. With Lemma 4.3 we get

$$
-A_3/2 = \frac{1}{n^3} \sum_{t,k=1}^{n} Y_{k,i} \mathbb{E}[Y'_{t,j} \mid \mathcal{F}] = \frac{1}{n^3} \sum_{t,k=1}^{n} Y_{k,i} \sum_{q=1}^{q} \exp(\beta m_{j,t}(\sigma)) \sum_{q=1}^{q} \exp(\beta m_{t,j}(\sigma))
$$

$$
= \frac{1}{n^3} \sum_{t,k=1}^{n} Y_{k,i} \left( \sum_{q=1}^{q} \exp(\beta m_{j,t}(\sigma)) \sum_{q=1}^{q} \exp(\beta m_{j,t}(\sigma)) \right) + \frac{1}{n^2} \sum_{k=1}^{n} Y_{k,i} \sum_{q=1}^{q} \exp(\beta m_{j,t}(\sigma))
$$

$$
= : M_1 + M_2.
$$

By using the same estimations as for $\hat{R}^{(1)}(i)$ we obtain

$$
M_1 \leq \frac{1}{n^3} \sum_{t,k=1}^{n} Y_{k,i} (q-1)\beta e^{2\beta} = \frac{1}{n} (q-1) \beta e^{2\beta} \left( \frac{W_i}{\sqrt{n}} + x_{0,i} \right).
$$

Hence $\mathbb{V}(M_1) = O(n^{-3})$ by Lemma 4.3. We obtain

$$
M_2 = \frac{1}{n} m_i(\sigma) (m_j(\sigma) - \frac{1}{\beta} \frac{\partial}{\partial u_j} G_{\beta,0}(m(\sigma)))
$$

$$
= \frac{1}{n} m_i(\sigma) m_j(\sigma) - \frac{1}{\beta n} m_i(\sigma) \left( \frac{\partial^2}{\partial u_j^2} G_{\beta,0}(x_0)(m_j(\sigma) - x_{0,j})
$$

$$
+ \sum_{k \neq j} \left( \frac{\partial^2}{\partial u_j u_k} G_{\beta,0}(x_0)(m_k(\sigma) - x_{0,k}) + \sqrt{n} R^{(2)}_{\beta}(j) \right),
$$

$$
\mathbb{V}(M_2) = O(n^{-3})
$$

Hence $\mathbb{V}(M_1) + \mathbb{V}(M_2) = O(n^{-3})$. Thus $\mathbb{V}(A_2) = O(n^{-3})$ and

$$
\mathbb{V}(B) = \frac{1}{n^2} \mathbb{V}(W_i W_j) = \frac{1}{n^2} (\mathbb{E}[W_i^2 W_j^2] + n \mathbb{E}[W_i^2]).
$$

We get

$$
\mathbb{V}(A) = O(n^{-3}).
$$

Hence $\mathbb{V}(A) = O(n^{-3})$.
Lemma 4.5, it follows that the maximal variance of all the sums in the representation of $O$ are of order $A_h = \beta_n h$. This completes the proof for $h = 0$. Hence $\Lambda$ has the form (6.3) and according to Lemma 6.2 we have

$$\det(\Lambda) = \frac{1}{n^q} (1 - \beta x_{0,q})^{q-2} (1 - q \beta x_{0,1} x_{0,q}).$$
So if $\beta \notin \{\frac{1}{q_0,1}, \frac{1}{q_0,1,0,0}\}$, the matrix $\Lambda$ is invertible. With Lemma 4.4 we get that $\Lambda$ is invertible for all $(\beta, h) \neq (\beta_0, h_0)$ and hence we are able to apply Theorem 3.1. The bound of $R_n^{(1)}(i, h)$ is $e^h$ times the bound of $R_n^{(1)}(i)$ implying the same order of $C$. The proof of bounding $B$ is unchanged. Bounding $A$ needs once more the bound $R_n^{(1)}(i, h)$ and hence the proof is almost the same as in the case $h = 0$. For the $q$-dependence we note that $\mathbb{E}[|W_i - W_j||W_k - W_t|]$ is independent of $q$ and $\lambda^{(i)} = \mathcal{O}(q)$. Thus $B = \mathcal{O}(q^4)$ by summing three times to $q$. Previous estimations show that $R_n^{(1)} = \mathcal{O}(q)$ and because we have two sums from 1 to $q$ for $W_k$ and $W_t$ we have that $R_n^{(2)} = \mathcal{O}(q^2)$. Thus $\mathbb{E}[R(i)^2] = \mathcal{O}(q^4)$. Hence $C = \mathcal{O}(q^4)$ by summing over $q$. Next we consider $A_1, A_2$ and $A_3$ taken from the proof. While $A_1$ and $A_2$ are independent of $q$, $A_3$ depends on $q$ via $R_n^{(1)}$ and $R_n^{(2)}$. Summing two times over $q$, we obtain $A = \mathcal{O}(q^5)$. Since the brackets in the bound of Reinert and Röllin still contain the parameter $q$ and since $||\Sigma||^{1/2} = \mathcal{O}(q)$, the constant $C$ of our theorem satisfies $C = \mathcal{O}(q^6)$. \hfill \Box

**Proof of Theorem 2.2.** Since the first part of the proof follows the lines of the proof of 2.1 we notice that Theorem 3.2 can be applied. Thus it remains to estimate the bound given there. For the first expression in the bound we notice that $A_1$ is the same expression as the A-term we just calculated for the proof of 2.1. Hence, $\log(n)A_1 = \mathcal{O}(\log(n)n^{-1/2})$. With Lemma 4.5 and the estimation for the C-term in 2.1 we obtain that the second expression is $\mathcal{O}(\log(n)n^{-1/2})$. For the third expression we notice that $a > 1$ is a constant and that $A_3 = \mathcal{O}(n)$. Using again Lemma 4.5 combined with the fact that $A = \frac{1}{\sqrt{n}}$ yields that the third expression is also $\mathcal{O}(\log(n)n^{-1/2})$. Likewise we obtain that the fourth expression is $\mathcal{O}(n^{-1/2})$. Combining these estimation yields the result. \hfill \Box

**Remark 5.1 (Heuristics).** By definition of $G_{\beta,h}$, (1.5), the Hessian of $G_{\beta,h}$ fulfills $D^2G_{\beta,h}(x) = \beta \text{Id} - \beta^2 D^2\Phi(x)$, where $\Phi$ is the log-moment generating function of the single-spin distribution in the Curie-Weiss-Potts model and $x$ is any minimum point. Hence $D^2\Phi$ is the covariance structure of the single-spins, which is

$$D^2\Phi(x) = -\frac{1}{\beta^2}(D^2G_{\beta,h}(x) - \beta \text{Id}). \quad (5.7)$$

We know from Stein’s method that if $(W, W')$ is exchangeable and (1.7) is satisfied with $R = 0$ we have

$$\frac{1}{2} \mathbb{E}[(W' - W)(W' - W)'] = \Sigma \Lambda'. $$

On the one hand in the Curie-Weiss-Potts model we have $\Sigma = [D^2G_{\beta,h}(x)]^{-1} - \beta^{-1}\text{Id}$. On the other hand the left hand side describes the empirical covariance structure of the single-spins:

$$\frac{1}{2} \mathbb{E}[(W_i' - W_i)(W_j' - W_j)] = \frac{1}{2n} \mathbb{E}(Y_{I,i}' - Y_{I,i})(Y_{I,j}' - Y_{I,j}).$$

Therefore with (5.7), heuristically

$$\frac{1}{2} \mathbb{E}[(W' - W)(W' - W)'] \approx \frac{1}{n} \frac{1}{\beta^2} (-D^2G_{\beta,h}(x) + \beta \text{Id}) = ([D^2G_{\beta,h}(x)]^{-1} - \beta^{-1}\text{Id}) \Lambda' .$$
If we now choose \( \Lambda = \Lambda' = \frac{1}{\beta n} D^2 G_{\beta, h}(x) \), the right hand identity is fulfilled.

Proof of Theorem 2.3. The proof uses the fact that the conditional joint distribution of the \((\sigma_i)_i\), conditioned on the event \( \{ \frac{N}{n} \in B(x, \epsilon) \} \), is given by

\[
P_{\beta, h, n, \epsilon}(\sigma) = \frac{1}{Z_{\beta, h, n, \epsilon}} \exp \left( \frac{\beta}{2n} \sum_{1 \leq i \leq j \leq n} \delta_{\sigma_i, \sigma_j} + h \sum_{i=1}^{n} \delta_{\sigma_i, 1} \right) 1_{B(x, \epsilon)}(N/n),
\]

where \( Z_{\beta, h, n, \epsilon} \) denotes a normalization. Thus we are able to start with any minimum point \( x_0 \) and follow the lines of the proof of Theorem 2.1.

\( \square \)

Proof of Theorem 2.4. We will apply Theorem 3.3. Obviously the density \( p \) is nice. Note that the logarithmic derivative is \( \psi(t) = \frac{p'(t)}{p(t)} = -\frac{16(q-1)^d}{3} t^3 \). The solutions \( f_g \) of the corresponding Stein equations (3.12) - with respect to absolutely continuous test functions \( g \) and with respect to \( g(x) = 1_{\{x \leq t\}}(x), z \in \mathbb{R} \), respectively - fulfill all boundedness assumption of Theorem 3.3. This was proved in [10] Lemma 2.2. By definition of \( T \), see (2.5), we have

\[
T = \frac{1}{(1-q)n^{3/4}}(N_1 - nx_1 - \sqrt{n} V_1) = \frac{1}{(1-q)n^{3/4}} \left( \sum_{i=1}^{n} Y_{i,1} - nx_1 - \sqrt{n} V_1 \right).
\]

We make use of the choice \( V \in \mathcal{M} \cap u^\perp \). With \( V \in \mathcal{M} \) we have \( \sum_{i=1}^{q} V_i = 0 \) and with \( \langle V, u \rangle = V_1(1-q) + \sum_{i=2}^{q} V_i = 0 \) we obtain

\[
\sum_{i=2}^{q} V_i = -V_1 = 0.
\]

Constructing an exchangeable pair \((T, T')\) is just the same as in the introduction: \( T' \) is a random variable being the same as \( T \) except that we pick an index \( I \) uniformly and exchange \( Y_{i,1} \) with \( Y'_{i,1} \) (for \( I = i \) distributed according to the conditional distribution of \( Y_{i,1} \) given \( (Y_{j,1})_{j \neq i} \), independently of \( Y_{i,1} \)). Now we calculate \( \mathbb{E}[T' - T] \) with \( \mathcal{F} = \sigma(\sigma_1, \ldots, \sigma_q) \).

\[
\mathbb{E}[T' - T \mid \mathcal{F}] = \frac{1}{(1-q)n^{3/4}} \sum_{i=1}^{n} \mathbb{E} \left[ Y'_{i,1} - Y_{i,1} \mid \mathcal{F} \right]
\]

\[
= \frac{1}{(1-q)n^{3/4}} \sum_{i=1}^{n} \mathbb{E}[Y'_{i,1} \mid \mathcal{F}] - \frac{1}{n} T - \frac{x_1}{(1-q)n^{3/4}}.
\]

With Lemma 4.2 we obtain

\[
\frac{1}{(1-q)n^{3/4}} \sum_{i=1}^{n} \mathbb{E}[Y'_{i,1} \mid \mathcal{F}] = \frac{1}{(1-q)n^{3/4}}(m_1(\sigma) - \frac{1}{\beta_0} \frac{\partial}{\partial x_1} G_{\beta_0, h_0}(m(\sigma))) + \frac{1}{(1-q)n^{1/4}} R_n^{(1)}(i, h_0).
\]

Hence using \( m_1(\sigma) = x_1 + \frac{(1-q)T}{n^{1/4}} \) and defining \( \tilde{R} := \frac{1}{(q-1)n^{1/4}} R_n^{(1)}(i, h_0) \) we have

\[
\mathbb{E}[T' - T \mid \mathcal{F}] = -\frac{1}{(1-q)\beta_0 n^{3/4}} \frac{\partial}{\partial x_1} G_{\beta_0, h_0} \left( x + \frac{T u}{n^{1/4}} + \frac{V}{n^{1/2}} \right) - \tilde{R}.
\]
A quite tedious fourth-order Taylor expansion of $G_{\beta_0,h_0}$ at $x + \frac{T u}{n^{1/4}} + \frac{V}{n^{1/2}}$ is affiliated in the Appendix, see (6.11), which leads to
\[
\frac{\partial}{\partial x_1} G_{\beta_0,h_0} \left( x + \frac{T u}{n^{1/4}} + \frac{V}{n^{1/2}} \right) = - \frac{16(q-1)^4}{3q n^{3/4}} T^3 \\
+ \mathcal{O} \left( \sum_{j=2}^q \frac{V_j^2}{n} \right) + \mathcal{O} \left( \frac{T^4}{n^{7/4}} \right) + \mathcal{O} \left( \frac{1}{n^{3/4}} f(V/\sqrt{n}, T/n^{1/4}) \right)
\]
with $f(v,t)$ given in (6.7). Hence we obtain $\mathbb{E}[T' - T | F] = \lambda \psi(T) - R$ with
\[
\lambda := \frac{1}{q(q-1)\beta_0 n^{3/2}}
\]
and
\[-R := \mathcal{O} \left( \sum_{j=2}^q \frac{V_j^2}{n^{7/4}} \right) + \mathcal{O} \left( \frac{T^4}{n^{7/4}} \right) + \mathcal{O} \left( \frac{1}{n^{3/4}} f(V/\sqrt{n}, T/n^{1/4}) \right) - \tilde{R}.
\]
Now $0 < \lambda < 1$ for all $n \in \mathbb{N}$ and thus we can apply Theorem 3.3. The moments of $T$ and $V$ are finite, see Lemma 4.6. With $V_1 = 0$ we get $T = \frac{1}{(1-q)n^{1/4}} W_1$. Now we are able to compute the expressions of the bound in Theorem 3.3. We have
\[
\mathbb{E}[(T' - T)^2 | T] = \frac{1}{(1-q)^2 n^{1/2}} \mathbb{E}[(W'_1 - W_1)^2 | W].
\]
Reproducing the proof of Theorem 2.1 we get $\mathbb{V} \{ \mathbb{E}[(W'_1 - W_1)^2 | F] \} = \mathcal{O}(n^{-7/2})$, using $\mathbb{E}[W^l] \leq n^{l/4} \mathbb{E}[T^l] = \mathcal{O}(n^{l/4})$, $l \in \mathbb{N}$. Thus
\[
\frac{1}{2A} \mathbb{V} \{ \mathbb{E}[(T' - T)^2 | T] \}^{1/2} = \mathcal{O}(n^{-1/4}).
\]
Moreover $\mathbb{E}[T' - T]^3 = \frac{1}{(1-q)^3 n^{3/2}} \mathbb{E}[Y'_1 - Y_1] = \mathcal{O}(n^{-9/4})$ and therefore $\frac{3\lambda}{A} \mathbb{E}[T' - T]^3 = \mathcal{O}(n^{-3/4})$. From the proof of Theorem 2.1 we know that $|R^{(1)}(i,h_0)| = \mathcal{O}(n^{-3/2})$, so $\tilde{R} = \mathcal{O}(n^{-7/4})$. Remark that by (6.7) we see that the expectation of $\mathcal{O} \left( \frac{1}{n^{3/4}} f(V/\sqrt{n}, T/n^{1/4}) \right)$ is of order $\mathcal{O}(n^{-2})$. Summarizing we obtain $\sqrt{\mathbb{E}[R^2]} = \mathcal{O}(n^{-7/4})$, hence
\[
\frac{c_1+c_2 \sqrt{\mathbb{E}[T^2]}}{\lambda} \sqrt{\mathbb{E}[R^2]} = \mathcal{O}(n^{-1/4}).
\]
Hence the $\delta$ in (3.14) is of order $\mathcal{O}(n^{-1/4})$. We obtain the same rate of convergence in the Kolmogorov distance, using $|T' - T| \leq \frac{\text{const}}{n^{3/4}} =: A$. The order of the first two summands in (3.16) is $\mathcal{O}(n^{-1/4})$. The third term in (3.16) is of order $\mathcal{O}(n^{-3/4})$ and finally
\[
\frac{3A}{2} \mathbb{E} \{ \psi(T) \} \leq \frac{\text{const}}{n^{3/4}} \mathbb{E}[T^3] = \mathcal{O}(n^{-3/4}),
\]
which completes the proof.

Proof of Theorem 2.6. Since $V_1 = 0$, by the continuous mapping theorem it suffices to show that the random vector $\tilde{V} := (V_2, \ldots, V_q)$ converges towards the $(q-1)$-dimensional centered Gaussian vector with covariance matrix
We will apply Theorem 3.1. We see for any \( q \)

\[
\begin{pmatrix}
q - 2 & -1 & \ldots & -1 \\
-1 & q - 2 & -1 & \ldots & -1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
-1 & \ldots & -1 & q - 2
\end{pmatrix}
\]

\[\frac{q}{2(q-1)^2(q-2)}\]

We will apply Theorem 3.1. We see for any \( q \)

\[
\mathbb{E}[V_i' - V_i \mid \mathcal{F}] = \mathbb{E}[W_i' - W_i \mid \mathcal{F}] = -\frac{1}{n\sqrt{n}} \sum_{j=1}^{n} Y_{ji} + \frac{1}{n\sqrt{n}} \sum_{j=1}^{n} \mathbb{E}[Y_{ji}' \mid \mathcal{F}].
\]

With Lemma 4.2 and \( R_n^{(1)}(i, h_0) \) defined as in (5.2) we obtain

\[
\mathbb{E}[V_i' - V_i \mid \mathcal{F}] = -\frac{1}{n\sqrt{n}} \sum_{j=1}^{n} Y_{ji} + \frac{1}{\beta \sqrt{n}} \frac{\partial}{\partial x_i} G_{\beta, h_0}(m) + R_n^{(1)}(i, h_0)
\]

By the forth-order Taylor expansion of \( G_{\beta, h_0}(m) \), see (6.6), (6.7), (6.9) and (6.12) in the Appendix, we obtain for any \( i \in \{2, \ldots, q\} \)

\[
\mathbb{E}[V_i' - V_i \mid V] = -\frac{4}{\beta n q^2} (1, \ldots, 1, (q^2 - 3q + 3), 1 \ldots, 1, \bar{V}) + R_i,
\]

where \((q-1)(q-2)\) is the \( i \)’th entry of the vector in \( \mathbb{R}^{q-1} \) and

\[
R_i := \mathcal{O}(\sum_{l=1}^{q} \frac{V_i T}{n^{5/4}}) + \mathcal{O}\left(\frac{1}{\sqrt{n}} A_2(i, V/\sqrt{n}, T/n^{1/4})\right) + \mathcal{O}(\frac{T^3}{n^{5/4}}) + \mathcal{O}(\frac{T^4}{n^{3/2}}) + R_n^{(1)}(i, h_0)
\]

(see (6.7) for the definition of \( A_2 \)). Using \( \sum_{k=2, k \neq i}^{q} V_k = -V_i \) we get

\[
<(1, \ldots, 1, (q^2 - 3q + 3), 1 \ldots, 1, \bar{V}) = ((q-1)(q-2))V_i.
\]

Thus the linearity condition of Theorem 3.1 is satisfied with \( \Lambda = \frac{1}{n} \frac{(q-2)}{q} \text{Id}_{q-1 \times q-1} \) and \( R = (R_2, \ldots, R_q) \). With \( q - 2 > 0 \) for \( q \geq 3 \) we get the invertibility of \( \Lambda \) and \( \lambda^{(i)} = \mathcal{O}(n) \). From the proof of Theorem 2.1 we see that

\[
\mathbb{E}[(V_i' - V_i)(V_j' - V_j) \mid \mathcal{F}] = \mathbb{E}[(W_i' - W_i)(W_j' - W_j) \mid \mathcal{F}] = \mathcal{O}(n^{-3/2})
\]

and thus \( \Lambda \) in Theorem 3.1 is of order \( \mathcal{O}(n^{-1/2}) \). Moreover \( |V_i' - V_i| \leq \frac{1}{\sqrt{n}} \) for all \( i \) and thus \( B \) in Theorem 3.1 is of order \( \mathcal{O}(n^{-1/2}) \). It remains to calculate \( C \) in Theorem 3.1. From the proof of Theorem 2.1 we know that \( \mathbb{E}(R_n^{(1)}(i, h_0)) = \mathcal{O}(n^{-3/2}) \). Using bounded moments of \( V_i \) and \( T \) we obtain that the expectations of the first and third term of \( R_i \) are \( \mathcal{O}(n^{-5/4}) \). With (6.7) we further get that the expectations of the second and fourth term of \( R_i \) are \( \mathcal{O}(n^{-3/2}) \) and \( \mathcal{O}(n^{-5/2}) \), respectively. By the Cauchy-Schwartz inequality we get \( \sqrt{\mathbb{E}[R_i^2]} = \mathcal{O}(n^{-5/4}) \), hence \( C = \mathcal{O}(n^{-1/4}) \), which completes the proof. \( \square \)
Remark 5.2. In Remark 5.1 we gave a heuristic, that the matrix $\Lambda$ in the regression condition (1.7) should be expected to be $\frac{1}{\beta n} D^2 G_{\beta,h}(x)$. Our heuristic is confirmed in the proof of Theorem 2.6 since we can rewrite (5.9) as

$$
E\left[ \tilde{V}' - \tilde{V} | \tilde{V} \right] = -\frac{1}{\beta_0 n} D^2 G_{q-1}(x) \tilde{V} + R
$$

where $D^2 G_{q-1}(x)$ denotes the upper $(q-1) \times (q-1)$ part of $D^2 G_{\beta_0,h_0}(x)$. The limiting covariance matrix $\Sigma$ of $\tilde{V}$ is given by $[D^2 G_{q-1}(x)]^{-1} - \beta_0^{-1} I_{q-1}$.

Remark 5.3. These rates of convergence remain still valid if we change our probability measure $P_{\beta,h,n}$ to

$$
P_{\beta,h,n}(\sigma) = \frac{1}{Z_{\beta,h,n}} \exp \left( \frac{\beta}{2} \sum_{1 \leq i \leq j \leq n} \delta_{\sigma_i,\sigma_j} + \sum_{j=1}^{n} \sum_{i=1}^{q} \delta_{\sigma_j,h_i} \right)
$$

for $\beta \in \mathbb{R}^+$ and $h \in \mathbb{R}^q$. This measure and the characteristics of the corresponding function

$$
G_{\beta,h}(u) := \frac{\beta}{2} \langle u, u \rangle - \log \left( \sum_{i=1}^{q} \exp (\beta u_i + h_i) \right)
$$

were studied in [26]. First of all we note that (1.1) is the same model with $h_1 = h$ and $h_i = 0$ for $i \in \{2, \ldots, q\}$. Based on the results of [13] and [26] and following the same procedures as above our results can easily be extended to the case that $h_i \neq 0$ for $i \in \{2, \ldots, q\}$. We omit these extensions here.

6. Appendix

For the proofs of Theorems 2.1, 2.2 and 2.3 and for Lemma 4.4 and Lemma 4.5 we need a multivariate second-order Taylor-expansion of $G_{\beta,h}$ defined in (1.5), for every $(\beta, h) \neq (\beta_0, h_0)$. Let us denote by $D^2 G_{\beta,h}(x)$ the Hessian matrix $\{ \partial^2 G_{\beta,h}(x) / \partial u_i \partial u_j, i, j = 1, \ldots, q \}$ of $G_{\beta,h}$ at $x$. We obtain

$$
G_{\beta,h}(u) = G_{\beta,h}(x) + \sum_{k=1}^{q} \frac{\partial}{\partial u_k} G_{\beta,h}(x)(u_k - x_k) + \frac{1}{2} \langle (u - x), D^2 G_{\beta,h}(x) \cdot (u - x) \rangle \quad (6.1)
$$

$$
+ \frac{1}{6} \sum_{t,k,j=1}^{q} \tilde{R}_{t,k,j}(u_t - x_t)(u_k - x_k)(u_j - x_j)
$$

with $\| \tilde{R}_{t,k,j} \| \leq \| \frac{\partial^3}{\partial u_k \partial u_t \partial u_j} G_{\beta,h} \|$. For any fixed $m \in \{1, \ldots, q\}$ and any $x, u \in \mathbb{R}^q$ it follows that

$$
\frac{\partial}{\partial u_m} G_{\beta,h}(u) = \frac{\partial}{\partial u_m} G_{\beta,h}(x) + \frac{\partial^2}{\partial^2 u_m} G_{\beta,h}(x)(u_m - x_m)
$$

$$
+ \sum_{k \neq m} \frac{\partial^2}{\partial u_k \partial u_m} G_{\beta,h}(x)(u_k - x_k) + \sum_{k=1}^{q} \mathcal{O}((u_m - x_m)(u_k - x_k))
$$

(6.2)
and type (6.3). The following Lemma is some Linear Algebra for a matrix of the form (6.3).

According to Proposition 4.1 we know that for any minimizer \( x \)

Proof. According to Proposition 4.1 we know that for any minimizer \( x \) of \( G_{\beta,h} \) we have \( x_{0,i} = x_{0,k} \) for all \( i, k \in \{2, \ldots, q\} \) and \( x_{0,1} \geq x_{0,k} \) for all \( k \in \{2, \ldots, q\} \) and \( \sum_{i=1}^{q} x_{0,i} = 1 \). Notice that the equation \( \nabla G_{\beta,0}(x_0) = 0 \) implies

\[
x_{0,1} = \frac{\exp(\beta x_{0,1} + h\delta_{1,1})}{\exp(\beta x_{0,1} + h) + (q - 1) \exp(\beta x_{0,q})},
\]

\[
x_{0,q} = \frac{\exp(\beta x_{0,q})}{\exp(\beta x_{0,1} + h) + (q - 1) \exp(\beta x_{0,q})}.
\]

Now we can calculate

\[
\frac{\partial^2}{\partial u_1^2} G_{\beta,h}(x_0) = \beta - \beta^2(q - 1) \frac{\exp(\beta(x_{0,1} + x_{0,q}) + h)}{(\exp(\beta x_{0,1} + h) + (q - 1) \exp(\beta x_{0,q}))^2} = \beta^2 x_{0,1} x_{0,q}.
\]

and

\[
\frac{\partial^2}{\partial u_1 \partial u_q} G_{\beta,h}(x_0) = \frac{\beta^2}{\exp(\beta x_{0,1} + h) + (q - 1) \exp(\beta x_{0,q})^2} = \beta^2 x_{0,1} x_{0,q}.
\]

Moreover

\[
\frac{\partial^2}{\partial u_q^2} G_{\beta,h}(x_0) = \beta - \beta^2 \frac{\exp(\beta x_{0,q} + h)(\exp(\beta x_{0,1} + h) + (q - 2) \exp(\beta x_{0,q}))}{(\exp(\beta x_{0,1} + h) + (q - 1) \exp(\beta x_{0,q}))^2} = \beta - \beta^2 x_{0,1} x_{0,q} + (q - 2) x_{0,q}^2
\]

and

\[
\frac{\partial^2}{\partial u_2 \partial u_q} G_{\beta,h}(x_0) = \frac{\beta^2}{(\exp(\beta x_{0,1} + h) + (q - 1) \exp(\beta x_{0,q}))^2} = \beta^2 x_{0,q}^2.
\]

\( \Box \)

With Lemma 6.1 we get, that the Hessian of \( G_{\beta,h} \) at a global minimum point is a matrix of type (6.3). The following Lemma is some Linear Algebra for a matrix of the form (6.3).
Lemma 6.2. For any \( a, b, c, d \in \mathbb{R} \) consider the following matrix:

\[
\Lambda := \begin{pmatrix}
  a & b & \ldots & \ldots & \ldots & b \\
  b & d & c & \ldots & \ldots & c \\
  b & c & d & c & \ldots & c \\
  \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
  b & c & \ldots & \ldots & c & d
\end{pmatrix} \in \mathbb{R}^{q \times q}.
\]

(6.3)

Then \( \det(\Lambda) = (d - c)^{q-2}(a(d + (q - 2)c) - (q - 1)b^2) \).

Proof. We applied the formula due to Laplace. \( \square \)

Remark 6.3. At the extremity \((\beta_0, h_0) = (4\frac{q-1}{q}, \log(q - 1) - 2\frac{q-2}{q})\) of the critical line, \( x = (1/2, 1/2(q - 1), \ldots, 1/2(q - 1)) \) is the unique minimum point of \( G_{\beta_0, h_0} \). Remark that

\[
\exp(\beta_0 \cdot x_1 + h_0) = (q - 1) \exp(2/q), \quad \exp(\beta_0 \cdot x_q) = \exp(2/q).
\]

With Lemma 6.1 we obtain

\[
\frac{\partial^2 G_{\beta_0, h_0}(x)}{\partial^2 x_1} = \frac{\partial^2 G_{\beta_0, h_0}(x)}{\partial x_1 \partial x_q} = \frac{4(q - 1)}{q^2}
\]

and \( \frac{\partial^2 G_{\beta_0, h_0}}{\partial^2 x_q}(x) = \frac{4(q^2 - 3q + 3)}{q^2} \) and \( \frac{\partial^2 G_{\beta_0, h_0}}{\partial x_q \partial x_{q+1}}(x) = \frac{4}{q^2} \). Thus \( a = b \) in (6.3).

For the proofs of the results at criticality, Theorems 2.4, 2.6 and Lemma 4.6, we need a multivariate fourth-order Taylor-expansion of \( G_{\beta_0, h_0} \) (defined in (1.5)). We fix the notation \( G := G_{\beta_0, h_0} \) for the following calculations. We know that \( x = (1/2, 1/2(q - 1), \ldots, 1/2(q - 1)) \) is the unique minimum point of \( G \). Let \( u = (1 - q, 1, \ldots, 1) \in M \subset \mathbb{R}^q, v \in M \cap u^+ \) and \( t \in \mathbb{R} \).

For any \( p \in \mathbb{N} \) and \( z \in \mathbb{R}^q \) let us fix the notation

\[
R_{i_1, \ldots, i_p}(z) := \left( \frac{\partial^p G}{\partial x_{i_1} \cdots \partial x_{i_p}} \right)(z).
\]

A second-order Taylor-expansion yields

\[
G(x + tu + v) = G(x + tu) + \frac{1}{2} \langle v, (D^2G)(x + tu) v \rangle + \frac{1}{6} \sum_{j,k,l=1}^{q} R_{j,k,l}(x + tu + \gamma v) v_j v_k v_l
\]

for some \( \gamma \in (0, 1) \), since \( \langle (\nabla G)(x + tu), v \rangle = 0 \): the last \( q - 1 \) coordinates of \( x + tu \) are equal and with Lemma 4.2 the last \( q - 1 \) coordinates of the gradient \((\nabla G)(x + tu)\) are equal, and hence it is orthogonal to \( v \). A fourth-order Taylor-expansion for \( G(x + tu) \) yields

\[
G(x + tu) = G(x) + \frac{1}{24} \sum_{j,k,l,m=1}^{q} R_{j,k,l,m}(x + \widetilde{\gamma} tu) t^4 u_j u_k u_l u_m
\]

(6.4)

for some \( \widetilde{\gamma} \in (0, 1) \). To see (6.4) notice that the first-order term is zero since \( x \) is a global minimizer of \( G \). The second-order term is zero, since we know from Lemma 4.4 that \( D^2G_{\beta, h}(x) \)
is positive definite if and only if \((\beta, h) \neq (\beta_0, h_0)\). Hence the third-order term is zero yielding the identity (6.4). Summarizing we obtain:

\[
G(x + tu + v) = G(x) + \frac{1}{2} \langle v, (D^2G)(x + tu) v \rangle + \frac{1}{6} \sum_{j,k,l=1}^{q} R_{j,k,l}(x + tu + \gamma v) v_j v_k v_l + \frac{1}{24} \sum_{j,k,l,m=1}^{q} R_{j,k,l,m}(x + \gamma tu) t^4 u_j u_k u_l u_m. \tag{6.5}
\]

With \(y := x + tu + v\) we will calculate \(\frac{\partial}{\partial y_i} G(y)\) for \(i \in \{1, \ldots, q\}\) using (6.5). The derivative of the first summand in (6.5) is zero since \(x\) is the global minimizer of \(G\). With

\[
\frac{1}{2} \langle v, (D^2G)(x + tu) v \rangle = \frac{1}{2} R_{i,i}(x + tu)(y_i - x_i - tu_i)^2 + \sum_{k \neq i} R_{i,k}(x + tu)(y_k - x_k - tu_k)(y_i - x_i - tu_i) + \sum_{k, l \neq i} R_{l,k}(x + tu)(y_k - x_k - tu_k)(y_l - x_l - tu_l)
\]

we obtain

\[
A_1(i) := \frac{\partial}{\partial y_i} \left( \frac{1}{2} \langle v, (D^2G)(x + tu) v \rangle \right) = R_{i,i}(x + tu)v_i + \sum_{k \neq i} R_{i,k}(x + tu)v_k.
\]

With Lemma 6.1 we obtain \(R_{1,2} = R_{1,3} = \cdots = R_{1,q}\) and since \(v \in u^\perp\) we have \(A_1(1) = 0\). With a second-order Taylor expansion for \(R_{i,k}(x + tu)\) we get for \(t\) small

\[
A_1(i) = \langle R_{i,:}, v \rangle + O_q \left( \sum_{l=1}^{q} v_l t \right).
\]

Here \(O_q\) is the notation that the constant does depend on \(q\). This is because all \(R_{i_1,\ldots,i_p}(x)\) only depend on \(q\), since \(x\) only depends on \(q\). The second partial derivatives of \(G\) were listed in Remark 6.3 hence we end up with

\[
A_1(1) = 0, \quad A_1(i) = \frac{4(q^3 - 3q + 3)}{q^2} v_i + \frac{4}{q^2} \sum_{k=2}^{q} v_k + O_q \left( \sum_{l=1}^{q} v_l t \right), \quad i \geq 2 \tag{6.6}
\]

for small \(t\). The last formula can even be simplified since \(\sum_{k=2}^{q} v_k = 0\) using \(v \in u^\perp\). For reasons of application we will not use this simplification. The partial derivative \(\frac{\partial}{\partial y_i}\) of the third term in (6.5) is

\[
A_2(i) := \frac{1}{2} R_{i,i,i}(x + tu + \gamma v)v_i^2 + \sum_{k \neq i}^q R_{i,i,k}(x + tu + \gamma v)v_i v_k + \frac{1}{2} \sum_{j, k \neq i}^q R_{i,j,k}(x + tu + \gamma v)v_k v_j.
\]

Using Taylor for \(R_{i,j,k}(x + tu + \gamma v)\) we obtain for small \(t\) and small \(v\)

\[
A_2(i) := A_2(i, v, t) := \frac{1}{2} R_{i,i,i}(x)v_i^2 + \sum_{k \neq i}^q R_{i,i,k}(x) v_i v_k + \frac{1}{2} \sum_{j, k \neq i}^q R_{i,j,k}(x) v_k v_j + O_q(f(v, t)) \tag{6.7}
\]
with
\[ O_q(f(v, t)) = O_q\left((t + \sum_{k=2}^{q} c_k(q) v_k)(v_i^2 + v_j \sum_{k \neq i} v_k + \sum_{k,j \neq i} v_k v_j)\right). \]

Here \( c_k(q) \) is denoted (a constant depending on \( q \)) just to see that the relation \( \sum_{k=2}^{q} v_k = 0 \) cannot be applied. In our application the order (in \( n \)) of the \( v_k \)'s will not depend on \( k \), and \( v_2 \) will be smaller in order than \( t \). In this situation we have \( O_q(f(v, t)) = O_q(t v_2^2) \). We will calculate \( A_2(1) \) with the help of the third derivatives which are \( R_{1,1,1}(x) = R_{1,1,k}(x) = 0 \) and
\[ R_{1,j,j}(x) = \frac{16(q-1)(q-2)}{q^3}, \quad R_{1,j,k}(x) = -\frac{16(q-1)}{q^3}. \]

Therefore the first two summands in (6.7) are zero and the third term is, using \( \sum_{k=2}^{q} v_k = 0 \),
\[ \frac{1}{2} \sum_{j,k=2}^{q} R_{1,j,k}(x) v_j v_k = \frac{1}{2} \sum_{j=2}^{q} \frac{16(q-1)(q-2)}{q^3} v_j^2 - \frac{1}{2} \sum_{j,k=2,j \neq k}^{q} \frac{16(q-1)}{q^3} v_j v_k \]
\[ = \frac{8(q-1)}{q^3} \sum_{j=2}^{q} v_j ((q-1)v_j - \sum_{k=2,k \neq j}^{q} v_k) \quad (6.8) \]
\[ = \frac{8(q-1)^2}{q^3} \sum_{j=2}^{q} v_j^2. \]

Finally, the partial derivative \( \frac{\partial}{\partial q} \) of the fourth term in (6.5) is
\[ A_3(i) := A_3(i, t) := \frac{1}{6} R_{i,i,i,i}(x) t^3 u_i^3 + \frac{1}{2} \sum_{k \neq i} R_{i,i,i,k}(x) t^3 u_i^2 u_k + \frac{1}{2} \sum_{k,j \neq i} R_{i,i,j,k}(x) t^3 u_i u_k u_j \]
\[ + \frac{1}{6} \sum_{j,k,l \neq i} R_{i,j,k,l}(x) t^3 u_j u_k u_l + O_q(t^4), \quad (6.9) \]
where we applied a second-order Taylor expansion for \( R_{i,j,k,l}(x + \tilde{\gamma} t u) \). Again we calculate \( A_3(1) \), using the fourth derivatives
\[ R_{1,1,1,1}(x) = \frac{32(q-1)^4}{q^4}, R_{1,1,1,k}(x) = \frac{32(q-1)^3}{q^4}, \quad R_{1,1,k,k}(x) = R_{1,1,j,k}(x) = \frac{32(q-1)^2}{q^4} \]
and
\[ R_{1,k,k,k}(x) = \frac{32(q-1)(2q^2 - 10q + 11)}{q^4}, \quad R_{1,j,j,k}(x) = -\frac{32(q-1)(2q-5)}{q^4} \]
and \( R_{1,j,j,k} = \frac{96(q-1)}{q^4} \). We obtain
\[ \frac{1}{6} R_{1,1,1,1}(x) t^3 (1-q)^3 = -\frac{16(q-1)^7 t^3}{3q^4}, \quad \frac{1}{2} \sum_{k \neq i} R_{1,1,1,k}(x) t^3 (1-q)^2 = -\frac{16(q-1)^6 t^3}{q^4}, \]
and
\[ \frac{1}{2} \sum_{k,j \neq i} R_{1,1,j,k}(x) t^3 (1-q) = -\frac{16(q-1)^5 t^3}{q^4}. \]
Moreover we have
\[\frac{1}{6} \sum_{j,k,l \neq i} R_{1,j,k,l}(x) t^3 = \frac{16(q - 1)^2(2q^2 - 10q + 11)t^3}{3q^4} - \frac{16(q - 1)^2(q - 2)(2q - 5)t^3}{q^4} + \frac{16(q - 1)^2(q - 2)(q - 3)t^3}{q^4}.\]

Hence
\[A_3(1) = -\frac{16(q - 1)^4}{3q} t^3 + O_q(t^4).\] (6.10)

We summarize that the first partial derivative of \(G(y)\) in (6.5) satisfies
\[\frac{\partial}{\partial y_1} G(x + tu + v) = \frac{8(q - 1)^2}{q^3} \sum_{j=2}^{q} v_j^2 - 16(q - 1)^4 \frac{3}{q} t^3 + O_q(f(v, t)) + O_q(t^4),\] (6.11)
using the notation of (6.7). The \(i\)'th partial derivative for \(i \in \{2, \ldots, q\}\) is given by
\[\frac{\partial}{\partial y_i} G(x + tu + v) = A_1(i) + A_2(i, v, t) + A_3(i, t)\] (6.12)
with \(A_j(i)\) defined in (6.6), (6.7) and (6.9).

**Proof of Lemma 4.4.** For the proof we will use the following alternative parametrization of the minimum points of \(G_{\beta, h}\) given by permutations of
\[x_s = \left(\frac{1 + (q - 1)s(\beta, h)}{q}, \frac{1 - s(\beta, h)}{q}, \ldots, \frac{1 - s(\beta, h)}{q}\right), \quad s(\beta, h) \in [0, 1].\]

It is important to notice, see for example [3], that \(s(\beta, h)\) is positive, well-defined and strictly increasing in \(\beta\) on an open interval containing \([\beta_c, \infty)\) and that \(s(\beta_c, 0) = (q - 2)/(q - 1)\) is a global minimum. The Lemma follows once we prove that \(\beta q x_{s,1} x_{s,q} - 1 < 0\) and \(\beta x_{s,q} - 1 < 0\). The second inequality follows directly from Proposition 4.1. To prove \(\beta q x_{s,1} x_{s,q} - 1 < 0\), we first consider the case \(h = h_0\). First of all we note that the minima are the solutions of
\[f(s(\beta, h_0)) = \log(1 + (q - 1)s(\beta, h_0)) - \log(1 - s(\beta, h_0)) - \beta s(\beta, h_0) - h_0 = 0.\]
If \(\beta < \beta_0\) we have that \(\partial f(s(\beta, h_0))/\partial s > 0\). Rearranging this equality and using the parametrization of \(x_s\) yields the result. If \(\beta > \beta_0\) we use \(\nabla G_{\beta, h_0}(x_s) = 0\) to obtain
\[\log(x_{s,1}) - \beta x_{s,1} - h_0 = \log(x_{s,q}) - \beta x_{s,q}.\]

This equation yield
\[\beta q x_{s,1} x_{s,q} - 1 = \left(\frac{\log(x_{s,1}) - \log(x_{s,q}) - h_0}{x_{s,1} - x_{s,q}}\right) q x_{s,1} x_{s,q} - 1 = \left(\log\left(\frac{x_{s,1}}{x_{s,q}}\right) - h_0\right) q x_{s,1} x_{s,q} - 1.\]

Using the fact that \(x_{s,1} + (q - 1)x_{s,q} = 1\) we obtain
\[\beta q x_{s,1} x_{s,q} - 1 = \left(\log\left(\frac{(q - 1)x_{s,1}}{1 - x_{s,1}}\right) - h_0\right) q x_{s,1} x_{s,q} - 1 = \left(\log\left(\frac{x_{s,1}(1 - x_{s,1})}{(q - 1)x_{s,1} - (1 - x_{s,1})}\right) - h_0\right) q x_{s,1} x_{s,q} - 1.\]
\[ \log \left( \frac{(q - 1)x_{s,1}}{1 - x_{s,1}} \right) - h_0 \right) \frac{x_{s,1}(1 - x_{s,1})}{q x_{s,1} - 1} - 1 = \frac{x_{s,1}(1 - x_{s,1})}{q x_{s,1} - 1} f_{h_0}(x_{s,1}), \]

where

\[ f_h(x_{s,1}) := \log \left( \frac{(q - 1)x_{s,1}}{1 - x_{s,1}} \right) - h - \frac{q x_{s,1} - 1}{q x_{s,1}(1 - x_{s,1})}. \]

For \( \beta_0 \) the global minimizer is given by \( y = (1/2, 1/2(q - 1), \ldots, 1/2(q - 1)) \) and it follows easily that \( f_{h_0} \left( \frac{1}{2} \right) = 0 \). Additionally \( \frac{\partial f_{h_0}}{\partial x}(x) < 0 \) for \( x \in [2^{-1}, 1] \). Thus we obtain that \( f_{h_0}(x_{s,1}) < 0 \) for \( x_{s,1} \in [2^{-1}, 1] \), and this is equivalent to \( \beta q x_{s,1} x_{s,q} - 1 < 0 \).

Now we consider the case: \( h \neq h_0 \). If \( \beta < \beta_0 \), the proof is identical to the case of \( \beta < \beta_0 \) for \( h = h_0 \). If \( \beta \geq \beta_0 \) we have

\[ \beta q x_{s,1} x_{s,q} - 1 = \frac{x_{s,1}(1 - x_{s,1})}{q x_{s,1} - 1} f_h(x_{s,1}). \]

For \( \beta \geq \beta_0 \) we know that \( 1 > s(\beta, h) \geq s(\beta_0, h) \geq s(\beta_0, 0) = (q - 2)/(q - 1) \). The rest of the proof is identical to the case (ii) of the proof of Proposition 2.2 in [13].

\[ \square \]

References

[1] A. D. Barbour, *Stein’s method for diffusion approximations*, Probab. Theory Related Fields 84 (1990), no. 3, 297–322. MR 91d:60081

[2] M. Biskup, L. Chayes, and N. Crawford, *Mean-field driven first-order phase transitions in systems with long-range interactions*, J. Stat. Phys. 122 (2006), no. 6, 1139–1193. MR 2219531 (2007g:82014)

[3] P. Blanchard, D. Gandolfo, J. Ruiz, and M. Wouts, *Thermodynamic vs topological phase transition: Cusp in the Kertész line*, Europhys. Lett. 82 (2008), 1–5.

[4] S. Chatterjee, P. Diaconis, and E. Meckes, *Exchangeable pairs and Poisson approximation*, Probab. Surv. 2 (2005), 64–106 (electronic). MR 2121796 (2007b:60087)

[5] S. Chatterjee and E. Meckes, *Multivariate normal approximation using exchangeable pairs*, ALEA Lat. Am. J. Probab. Math. Stat. 4 (2008), 257–283. MR MR2453473

[6] S. Chatterjee and Q.-M. Shao, *Stein’s method of exchangeable pairs with application to the Curie-Weiss model*, to appear in Ann. Appl. Prob., 2010.

[7] M. Costeniuc, R. S. Ellis, and P. T.-H. Otto, *Multiple critical behavior of probabilistic limit theorems in the neighborhood of a tricritical point*, J. Stat. Phys. 127 (2007), no. 3, 495–552. MR 2316196 (2008i:60044)

[8] M. Costeniuc, R. S. Ellis, and H. Touchette, *Complete analysis of phase transitions and ensemble equivalence for the Curie-Weiss-Potts model*, J. Math. Phys. 46 (2005), no. 6, 063301, 25. MR 2149837 (2006d:82025)

[9] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, New York, 1998.

[10] P. Eichelsbacher and M. Löwe, *Stein’s method for dependent variables occurring in statistical mechanics*, Electron. J. Probab. 15 (2010), no. 30, 962–988. MR 2570014

[11] R. S. Ellis, K. Haven, and B. Turkington, *Large deviation principles and complete equivalence and nonequivalence results for pure and mixed ensembles*, J. Statist. Phys. 101 (2000), no. 5-6, 999–1064. MR 1806714 (2001m:82003)

[12] R. S. Ellis and C. M. Newman, *Limit theorems for sums of dependent random variables occurring in statistical mechanics*, Z. Wahrsch. Verw. Gebiete 44 (1978), no. 2, 117–139.
[13] R. S. Ellis and K. Wang, *Limit theorems for the empirical vector of the Curie-Weiss-Potts model*, Stochastic Process. Appl. 35 (1990), no. 1, 59–79. MR 91m:82023

[14] ———, *Limit theorems for maximum likelihood estimators in the Curie-Weiss-Potts model*, Stochastic Process. Appl. 40 (1992), no. 2, 251–288. MR 93c:62046

[15] D. Gandolfo, J. Ruiz, and M. Wouts, *Limit theorems and coexistence probabilities for the Curie-Weiss Potts model with an external field*, Stochastic Process. Appl. 120 (2010), no. 1, 84–104. MR 2565853 (2010j:82021)

[16] F. Götze, *On the rate of convergence in the multivariate CLT*, Ann. Probab. 19 (1991), no. 2, 724–739. MR MR1106283 (92g:60028)

[17] R. B. Griffiths and P. A. Pearce, *Potts model in the many–component limit*, J. Phys. A 13 (1980), 2143–2148. MR 42 #1412

[18] H. Kesten and R. H. Schonmann, *Behavior in large dimensions of the Potts and Heisenberg models*, Rev. Math. Phys. 1 (1989), no. 2-3, 147–182. MR 91g:82015

[19] G. Reinert and A. Röllin, *Multivariate normal approximation with Stein’s method of exchangeable pairs under a general linearity condition*, Ann. Probab. 37 (2009), no. 6, 2150–2173. MR 2573554

[20] Y. Rinott and V. Rotar, *A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics*, J. Multivariate Anal. 56 (1996), no. 2, 333–350. MR 1379533 (97a:60035)

[21] ———, *On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted $U$-statistics*, Ann. Appl. Probab. 7 (1997), no. 4, 1080–1105. MR 1484798 (99g:60050)

[22] B. Simon and R. B. Griffiths, *The $(\phi^4)_2$ field theory as a classical Ising model*, Comm. Math. Phys. 33 (1973), 145–164. MR MR0428998 (55 #2018)

[23] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory (Berkeley, Calif.), Univ. California Press, 1972, pp. 583–602. MR 53 #6687

[24] C. Stein, *Approximate computation of expectations*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7, Institute of Mathematical Statistics, Hayward, CA, 1986. MR MR882007 (88j:60055)

[25] C. Stein, P. Diaconis, S. Holmes, and G. Reinert, *Use of exchangeable pairs in the analysis of simulations*, Stein’s method: expository lectures and applications, IMS Lecture Notes Monogr. Ser., vol. 46, Inst. Math. Statist., Beachwood, OH, 2004, pp. 1–26. MR MR2118600 (2005j:65005)

[26] K. Wang, *Solutions of the variational problem in the Curie-Weiss-Potts model*, Stochastic Process. Appl. 50 (1994), no. 2, 245–252. MR 1273773 (95a:60029)

[27] F. Y. Wu, *The Potts model*, Rev. Modern Phys. 54 (1982), no. 1, 235–268. MR 84d:82033