

# Gibbs-non-Gibbs transition in the fuzzy Potts models with a Kac-type interaction: Closing the Ising gap

Florian Henning, \* †      Richard C. Kraaij, \* ‡  
Christof Külske \* §

February 20, 2018

## Abstract

We complete the investigation of the Gibbs properties of the fuzzy Potts model on the  $d$ -dimensional torus with Kac interaction which was started by Jahnelt and one of the authors in [JK17b]. As our main result of the present paper, we extend the previous sharpness result of mean-field bounds to cover all possible cases of fuzzy transformations, allowing also for the occurrence of Ising classes (**containing precisely two spin values**). The closing of this previously left open Ising-gap involves an analytical argument showing uniqueness of minimizing profiles for certain non-homogeneous conditional variational problems.

**AMS 2000 subject classification:** 82B20, 82B26.

**Keywords:** Potts model, Kac model, fuzzy Kac-Potts model, Gibbs versus non-Gibbs, large deviation principles, diluted large deviation principles.

---

\*Ruhr-Universität Bochum, Fakultät für Mathematik, D44801 Bochum, Germany

†Florian.Henning@rub.de

‡Richard.Kraaij@rub.de

§Christof.Kuelske@ruhr-uni-bochum.de,

<http://www.ruhr-uni-bochum.de/ffm/Lehrstuehle/Kuelske/kuelske.html> /~kuelske/

# 1 Introduction

The Gibbs property is an important regularity property of a large, or infinite system which comes in various versions, according to the setting considered. It can be formulated for lattice models in the infinite volume [EFS93; Geo11], for systems of point particles in euclidean space [Rue99] for mean-field systems [KLN07], or for Kac-systems [FHM14].

Often local transformations applied to realizations of a given large system which behaves according to a nice Gibbs distribution are of interest. Such transformations include discretization transformations, and various sampling procedures, which reduce information of the initial data, cf. [EFS93; HK04; JK17b]. Stochastic time-evolutions, motivated from physics, provide another very interesting type of such transformations, cf. [KLN07; EK10; FHM13; FHM14; HRZ15; JK17c; JK17a].

A relevant and natural question in this context is whether it is possible to describe the image system again as a system of a nice Gibbsian form, with effective, new interactions. This may be true (or not true) for reasons having to do with the absence (or the existence, and visibility) of internal phase transitions. These internal phase transitions may act as a switch and cause an infinite-range dependence on small variations of a conditioning. If we want to perform a concrete analysis to decide if such a mechanism shows up for a given system, given parameter values, and given transformation, both mean-field systems and Kac-systems are promising as they have a large parameter, which allows for an asymptotic description in terms of a variational principle.

Indeed, after some work, one sees that also Gibbs-non-Gibbs(GnG) transitions in the Kac-limit, are tied to *conditional* variational problems for profiles, via large deviations. There is also a natural link from Kac-models to mean-field models on the level of these variational principles, which one obtains by considering only profiles which do not have any spatial dependence.

In the present paper, we round up the previous investigation of the GnG transitions in the fuzzy Kac-Potts model of [JK17b], and fully decide on an open case which was left out, as we will explain now. The fuzzy Kac-Potts model is defined as the image of the Kac-version of Potts model, under the deterministic local transformation which distinguishes the possible spin values of the Potts local spin space  $\{1, 2, \dots, q\}$  only according to a fixed local partitioning into  $r$  subclasses. **Historically, the fuzzy Potts model was first introduced on the lattice in [MV95]. Bounds in parameter space ensuring Gibbsian, and bounds ensuring non-Gibbsian behavior on the lattice were derived in [Häg03], however, leaving gaps between both regions. The Kac-Potts model we treat here is formulated on a d-dimensional torus, a setup similar to that of [FHM14] in their study of the dynamical Kac-Ising model. For the fuzzy Kac-Potts model, we are now able to completely answer the question:**

*Are the non-Gibbs parameter regions equal in Kac-setup and mean-field setup?*

Or, putting this question in a slightly more refined way: Does the possibility for spatial inhomogeneity in the Kac-model create new and ‘worse’ bad configurations

which are seen in parameter regions beyond the bad mean-field regions? It turns out that worse bad configurations are indeed not possible, but for a non-trivial reason involving proper treatment of a non-homogeneous conditional variational problem, arising for the Ising classes (i.e. the classes which consist of precisely two values).

For class sizes strictly bigger than two, the corresponding conditional variational problem is related to a Potts model with a first-order transition whose jumps are producing the non-Gibbsianness. Spatial inhomogeneity in the conditioning can be treated for those classes with monotonicity arguments which show that ‘flat conditioning profiles are the worst’, but these arguments use the first-order nature of the phase transition and do not work for the Ising classes of size two.

In the present paper we provide different arguments for the classes of size two and arrive at the full treatment of all possible types of fuzzy maps, and this is what we mean by "closing the Ising gap". The progress made in the present paper relies essentially on the new analytical uniqueness of minimizers-result of Proposition 2.15, which leads to the proof of the main result Theorem 2.6.

The paper is organized as follows. We keep the presentation self-contained, at the same time streamlining and shortening some parts of the arguments. In Section 2 we introduce our model and state the main results. In particular, in Sections 2.1 to 2.3 we introduce the Potts-Kac, the fuzzy Potts-Kac model, the notion of sequential Gibbsianness and our main result, Theorem 2.6 on Gibbs-non-Gibbs transitions. In Section 2.4, we recall a representation of the conditional kernels that play an important role in the study of GnG transitions and recall the diluted large deviation principle from [JK17b]. In Section 2.5, we give the main technical result, Proposition 2.15, and conclude with a representation of the limiting conditional probability kernels in the Gibbs case, given in Theorem 2.16.

The result shows that Gibbsianness of a transformed system may very well hold in the presence of internal phase transitions, reflected here as the functional dependence of limiting kernels on non-trivial minimizing profiles which nevertheless themselves behave continuously as a function of conditioning profiles. While there is an internal phase transition in the Ising classes at sufficiently low temperatures, both phases always appear as symmetric mixtures, and there is no way to break their symmetry and select one of them in terms of the choice of a boundary condition. For a related phenomenon, compare the discussion of possible ‘invisibility of hidden phase transitions’ for the joint measures of a random bond Ising ferromagnet [Kül99, page 379].

The proofs follow in Section 3 and we conclude with two appendices on continuity of limiting functionals and convergence of random variables in the setting where an LDP holds with a finite number of minimizers.

**Acknowledgement** RK is supported by the Deutsche Forschungsgemeinschaft (DFG) via RTG 2131 High-dimensional Phenomena in Probability–Fluctuations and Discontinuity.

## 2 Model and main results

### 2.1 The Kac-Potts model

Let  $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$  be the  $d$ -dimensional unit torus. For  $n \in \mathbb{N}$ , let  $\mathbb{T}_n^d$  be the  $(1/n)$ -discretization of  $\mathbb{T}^d$  defined by  $\mathbb{T}_n^d := \Delta_n^d/n$ , with  $\Delta_n^d := \mathbb{Z}^d/n\mathbb{Z}^d$  the discrete torus of size  $n$ . For  $n \in \mathbb{N}$ , let  $\Omega_{q,n} := \{1, \dots, q\}^{\Delta_n^d}$  be the set of *Potts-spin configurations* on  $\Delta_n^d$ . We will call elements of  $\{1, \dots, q\}$  *colours*. The energy of the configuration  $\sigma := (\sigma(x))_{x \in \Delta_n^d} \in \Omega_{q,n}$  is given by the *Kac-type Hamiltonian*

$$H_n(\sigma) := -\frac{1}{n^d} \sum_{x,y \in \Delta_n^d} J\left(\frac{x-y}{n}\right) 1_{\sigma(x)=\sigma(y)}, \quad \sigma \in \Omega_{q,n} \quad (2.1)$$

where  $0 < J \in C(\mathbb{T}^d)$  is a continuous *interaction function* on  $\mathbb{T}^d$  which is symmetric and  $\int dv J(v) = 1$ . The Gibbs measure associated with  $H_n$  is given by

$$\mu_n(\sigma) := \frac{1}{Z_n} \exp(-\beta H_n(\sigma)), \quad \sigma \in \Omega_{q,n} \quad (2.2)$$

with  $\beta \in [0, \infty)$  the inverse temperature and  $Z_n$  the normalizing partition sum.

To study the limiting behaviour of the measures  $\mu_n$  for large  $n$ , we first embed our spin configurations into a common space of measures. For any Polish space  $E$ , let  $C_b(E)$  be the space of bounded continuous functions on  $E$  and let  $\mathcal{M}_+(E)$  and  $\mathcal{P}(E)$  be the space of non-negative, respectively probability, Borel measures on  $E$ . We equip both spaces with the weak topology, i.e. the metric topology induced by testing against functions in  $C_b(E)$ . Recall that if  $E$  is compact, then also  $\mathcal{P}(E)$  is weakly compact.

For  $\Lambda \subseteq \Delta_n^d$  let  $\pi_\Lambda : \Omega_{q,n} \mapsto \mathcal{P}(\mathbb{T}_n^d \times \{1, \dots, q\}) \subseteq \mathcal{P}(\mathbb{T}^d \times \{1, \dots, q\})$  be the *empirical colour measure vector* or *colour profiles* of  $\sigma$  inside the volume  $\Lambda$  defined by

$$\pi_\Lambda(\sigma) := \frac{1}{|\Lambda|} \left( \sum_{x \in \Lambda} 1_{\sigma(x)=1} \delta_{x/n}, \dots, \sum_{x \in \Lambda} 1_{\sigma(x)=q} \delta_{x/n} \right) \quad (2.3)$$

where  $\delta_u$  is the Dirac measure at  $u \in \mathbb{T}^d$ .

For any  $\nu \in \mathcal{P}(\mathbb{T}^d \times \{1, \dots, q\})$  we will write  $\nu[a]$  to indicate the evaluation of  $\nu$  at a colour  $a \in \{1, \dots, q\}$ . In particular, for  $\Lambda \subseteq \Delta_n^d$  and  $\sigma \in \Omega_{q,n}$ :

$$\pi_\Lambda[a](\sigma) := \frac{1}{|\Lambda|} \sum_{x \in \Lambda} 1_{\sigma(x)=a} \delta_{x/n} \in \mathcal{M}_+(\mathbb{T}^d).$$

For simplicity, we will write  $\pi_n$  and  $\pi_n[a]$  to denote  $\pi_{\Delta_n^d}$  and  $\pi_{\Delta_n^d}[a]$ . Using colour profiles, we can rewrite the Hamiltonian as

$$H_n(\sigma) = -n^d \sum_{a=1}^q F(\pi_n[a](\sigma)) \quad (2.4)$$

with  $F(\nu[a]) := \langle J * \nu[a], \nu[a] \rangle = \int \int \nu[a](du) \nu[a](dv) J(u-v)$  and where the convolution of a function and a measure is defined as  $(f * \mu)(x) = \int f(x-y) \mu(dy)$ .

We will be interested in weak limits of color profiles in  $\mathcal{P}(\mathbb{T}^d \times \{1, \dots, q\})$ , especially those having  $q$ -dimensional Lebesgue densities of the form  $\nu = \alpha\lambda = (\alpha[1]\lambda, \dots, \alpha[q]\lambda)$  with  $\alpha \in B$  where

$$B := \left\{ \alpha = (\alpha[1], \dots, \alpha[q])^T \mid 0 \leq \alpha[a] \in L^\infty(\mathbb{T}^d, \lambda) \right. \\ \left. \text{with } \sum_{a=1}^q \alpha[a](x) = 1 \text{ for } \lambda\text{-a.e. } x \in \mathbb{T}^d \right\}.$$

In the sequel, we will write

$$\mathcal{P}_q(B) := \left\{ \nu \in \mathcal{P}(\mathbb{T}^d \times \{1, \dots, q\}) \mid \exists \alpha \in B : \nu = \alpha\lambda \right\}.$$

We have the following straightforward result, proven for completeness in Appendix A, which indicates that  $\mathcal{P}_q(B)$  is the relevant space of limiting profiles.

**Lemma 2.1.** *If  $\sigma_n \in \Omega_{q,n}$  is a sequence such that  $\pi_n(\sigma_n)$  converges weakly to  $\nu$  in  $\mathcal{P}(\mathbb{T}^d \times \{1, \dots, q\})$ , then  $\nu \in \mathcal{P}_q(B)$ .*

## 2.2 Sequential Gibbsianness for the Kac models

For  $n \geq 1$  and  $u \in \mathbb{T}_n^d$ , we call  $\pi_n^{(u)} := \pi_{\Delta_n^d \setminus \{nu\}}$  the *colour profile perforated at  $u \in \mathbb{T}_n^d$* . We abbreviate  $\mathcal{M}_{q,n} := \pi_n(\Omega_{q,n}) \subseteq \mathcal{P}(\mathbb{T}^d \times \{1, \dots, q\})$  and  $\mathcal{M}_{q,n}^u := \pi_n^{(u)}(\Omega_{q,n}) \subseteq \mathcal{P}(\mathbb{T}^d \times \{1, \dots, q\})$  for the sets of possible profiles of mesh-size  $n^{-1}$  and possible profiles of mesh-size  $n^{-1}$  perforated at site  $u$ . Note that by Lemma 2.1 any limit  $\nu$  of profiles  $\nu_n \in \mathcal{M}_{q,n}$  or  $\mathcal{M}_{q,n}^{(u_n)}$  must lie in  $\mathcal{P}_q(B)$ .

We slightly change the definition of sequential Gibbsianness defined for Kac models as in [JK17b] and [FHM14]. This definition is essentially equivalent but notationally and mathematically more convenient.

**Definition 2.2.** Given any sequence  $(\mu_n)_{n \in \mathbb{N}}$  with  $\mu_n$  a probability measure on  $\Omega_{q,n}$  for every  $n \in \mathbb{N}$ , define the single-spin conditional probabilities at site  $u \in \mathbb{T}_n^d$  as

$$\gamma_n^u(\cdot \mid \nu_n^{(u)}) := \mu_n(\sigma(nu) = \cdot \mid \pi_n^{(u)}(\sigma) = \nu_n^{(u)}) \quad \nu_n^{(u)} \in \mathcal{M}_{q,n}^u. \quad (2.5)$$

- (a) We call a colour profile  $\nu \in \mathcal{P}_q(B)$  **good** for a sequence of probability measures  $(\mu_n)_{n \in \mathbb{N}}$ ,  $\mu_n \in \mathcal{M}_{q,n}$ , if there exists a neighbourhood  $\mathcal{N}_\nu \subseteq \mathcal{P}(\mathbb{T}^d \times \{1, \dots, q\})$  of  $\nu$  such that for all  $\tilde{\nu} \in \mathcal{N}_\nu \cap \mathcal{P}_q(B)$  and for all  $u \in \mathbb{T}^d$

$$\gamma^u(\cdot \mid \tilde{\nu}) := \lim_{n \uparrow \infty} \gamma_n^{u_n}(\cdot \mid \nu_n^{(u_n)}) \quad (2.6)$$

exists for all sequences  $u_n \in \mathbb{T}_n^d$  with  $u_n \rightarrow u$  and all sequences  $(\nu_n^{(u_n)})_{n \in \mathbb{N}}$  with  $\nu_n^{(u_n)} \in \mathcal{M}_{q,n}^{u_n}$  for every  $n \in \mathbb{N}$  such that  $\lim_{n \uparrow \infty} \nu_n^{(u_n)} = \tilde{\nu}$  in the weak sense. Moreover the limit must be independent of the choice of  $u_n$  and  $(\nu_n^{(u_n)})_{n \in \mathbb{N}}$ .

- (b) A colour profile  $\nu \in \mathcal{P}_q(B)$  is called **bad** for  $(\mu_n)_{n \in \mathbb{N}}$  if it is not good for  $(\mu_n)_{n \in \mathbb{N}}$ .
- (c)  $(\mu_n)_{n \in \mathbb{N}}$  is called sequentially Gibbs if it has no bad profiles in  $\mathcal{P}_q(B)$ .

**Remark 2.3.** By Lemma B.1 below, it follows that if  $(\mu_n)_{n \in \mathbb{N}}$  is sequentially Gibbs, then the map  $\gamma : \mathbb{T}^d \times \mathcal{P}_q(B) \rightarrow \mathcal{P}(\{1, \dots, q\})$ , defined by  $\gamma(u, \nu) = \gamma^u(\cdot | \nu)$  as in (2.6) is continuous.

## 2.3 Sequential Gibbsianness for the fuzzy Kac-Potts model

Next, we introduce the fuzzy Kac-Potts model. Consider a discretization map  $T : \{1, \dots, q\} \mapsto \{1, \dots, s\}$  where  $1 < s < q$ . More precisely, let  $R_1, \dots, R_s$  be a partition of  $\{1, \dots, q\}$  with  $r_i = |R_i|$  and  $\sum_{i=1}^s r_i = q$ , then  $T(a) = i$  if  $a \in R_i$ . The map  $T$  induces a local discretization map  $T : \Omega_{q,n} \rightarrow \Omega_{s,n}$  by applying  $T$  at every site. The *fuzzy Kac-Potts model* is obtained from the Kac-Potts model by applying the discretization map at every site:  $\mu_n^T := \mu_n \circ T^{-1} \in \mathcal{P}(\Omega_{s,n})$ , where  $\mu^n$  is the Kac-Potts measure.

We will refer to the sets  $R_1, \dots, R_s$  as *fuzzy classes* and in particular use the term *Ising class* for fuzzy classes of size two.

Since Potts distributions do not depend on any ordering on the local spin space (this will be further exploited in Section 3.5 below), all relevant information on the fuzzy transformation is given by the numbers  $(r_1, \dots, r_s)$ .

**Definition 2.4.** We call the generalized fuzzy Kac-Potts model sequentially Gibbs if all profiles  $\mathcal{P}_s(B)$  are good for the sequence  $\mu_n^T$ .

Jahnel and Külske studied the question of sequential Gibbsianness in [JK17b]. Their main result states that for the fuzzy Kac-Potts model the critical parameters for Gibbs-non-Gibbs transitions are the same as for the mean-field fuzzy Potts model if low-temperature Ising classes are avoided, see (a) below. We denote by  $\beta_c(r)$  the inverse critical temperature of the  $r$ -state mean-field Potts model.

**Theorem 2.5** (Theorem 2.7 in [JK17b]). *Consider the  $q$ -state Kac-Potts model at inverse temperature  $\beta$  and the corresponding fuzzy Kac-Potts model with spin partition  $(r_1, \dots, r_s)$ ,  $r_i \in \mathbb{N} \setminus \{0\}$ ,  $s < q$  and  $\sum_{i=1}^s r_i = q$ . Denote  $r_* := \min\{r \geq 3, r = r_i \text{ for some } i = 1, \dots, s\}$ .*

- (a) *Suppose that either  $\beta \leq \beta_c(2)$  or that  $r_i \neq 2$  for all  $i = 1, \dots, s$  and  $\beta < \beta_c(r_*)$ , then the fuzzy Kac-Potts model is sequentially Gibbs.*
- (b) *If  $r_i \geq 3$  for some  $i = 1, \dots, s$  and  $\beta \geq \beta_c(r_*)$ , then the fuzzy Kac-Potts model is non-Gibbs.*

This result should be compared to Theorem 1.2 in [HK04], which establishes a full characterization of the parameter regions for which the sequence of measures is sequentially Gibbs or not **in the mean-field fuzzy Potts model and the fuzzy**

**Potts model on trees.** Theorem 2.5 leaves open the question what happens if there are fuzzy classes of size 2 and  $\beta > \beta_c(2)$ . The main result of this paper establishes that the fuzzy Kac-Potts model is also sequentially Gibbs if all classes are of size 2 or  $\beta < \beta_c(r^*)$ , establishing a full characterization with the same parameter regime as for the mean-field fuzzy Potts model.

**Theorem 2.6.** *Consider the  $q$ -state Kac-Potts model at inverse temperature  $\beta$  and the corresponding fuzzy Kac-Potts model with spin partition  $(r_1, \dots, r_s)$ ,  $r_i \in \mathbb{N} \setminus \{0\}$ ,  $s < q$  and  $\sum_{i=1}^s r_i = q$ .*

(a) *Suppose that  $r_i \leq 2$  for all  $i = 1, \dots, s$ . Then the fuzzy Kac-Potts model is sequentially Gibbs for all  $\beta \geq 0$ .*

*Assume that  $r_i \geq 3$  for some  $i$  and denote  $r^* := \min\{r \geq 3, r = r_i \text{ for some } i = 1, \dots, s\}$ , then the following holds.*

(b) *If  $\beta < \beta_c(r^*)$  then the fuzzy Kac-Potts model is sequentially Gibbs.*

(c) *If  $\beta \geq \beta_c(r^*)$  then the fuzzy Kac-Potts model is not sequentially Gibbs.*

Like in the proof of Theorem 2.7 in [JK17b], i.e. Theorem 2.5, the analysis is based on a representation of the single-site kernels in terms of diluted Potts models, and an identification of the minimizers of an LDP for these diluted models. Unlike in the setting of classes of size larger or equal to three, for the ‘low temperature’ Ising classes we find exactly two minimizers with opposite local magnetization that are not spatially homogeneous. To give a representation for the limiting kernels in Theorem 2.16 below, we therefore need to introduce the corresponding representation and large deviation principle.

## 2.4 A representation for the fuzzy Potts-Kac model kernels and a diluted large deviation principle

In order to determine Gibbsianness of the fuzzy Kac-Potts model, similar to (2.5), as well as to give a representation for the limiting single-site kernels, we write for the single-site kernels

$$\gamma_{n,\beta,q,(r_1,\dots,r_s)}^u(k|\nu) := \mu_n^T(\sigma(nu) = k | \pi_n^{(u)}(\sigma) = \nu) \quad (2.7)$$

where  $\beta$  is the inverse temperature of the Kac-Potts model and  $\nu \in \mathcal{M}_{s,n}^u$ . It was shown in [JK17b] that the kernels can be re-expressed in terms of certain functionals integrated over Kac-Potts models on the fuzzy classes.

**Definition 2.7.** Define the following three objects:

(a) For a measure  $\nu \in \mathcal{M}_{s,n}$ ,  $u \in \mathbb{T}_n^d$  and  $i \in \{1, \dots, s\}$ , denote by

$$\Lambda_{i,u}(\nu) := \left\{ x \in \Delta_n^d \setminus \{un\} \mid \nu[i](x/n) = 1 \right\}$$

the set of sites with a spin-value in the  $i$ -th class.

- (b)  $\mu_{\Lambda,\beta,r}$  denotes the Kac-Potts model in the subvolume  $\Lambda \subseteq \Delta_n^d$ ,  $\Lambda \neq \emptyset$ , with Hamiltonian

$$H_{\Lambda}(\sigma) := -\frac{1}{|\Lambda|} \sum_{x,y \in \Lambda} J \left( \frac{x-y}{n} \right) 1_{\sigma(x)=\sigma(y)},$$

inverse temperature  $\beta$  and  $r$  local states. Denote by  $P_{\Lambda,\beta,r} \in \mathcal{P}(\mathcal{P}(\mathbb{T}^d \times \{1, \dots, r\}))$  the push-forward of  $\mu_{\Lambda,\beta,r}$  under  $\pi_{\Lambda}$ :

$$P_{\Lambda,\beta,r} := \mu_{\Lambda,\beta,r} \circ \pi_{\Lambda}^{-1}.$$

- (c) For each class size  $r$ , define  $\mathcal{A}_r : \mathcal{P}(\mathbb{T}^d \times \{1, \dots, r\}) \times \mathbb{T}^d \times \mathbb{R}^+ \rightarrow [r, \infty)$  as

$$\mathcal{A}_r(\pi, u, \beta) := \sum_{i=1}^r \exp \{2\beta (J * \pi[i])(u)\}.$$

Note that the functional  $\mathcal{A}$  that we introduce is not equal to  $A$  in [JK17b]. The functional  $A$  includes an integral, which  $\mathcal{A}$  does not. This is done to allow for a better control on and representation of the limiting problem, and the limiting kernels to be considered below.

**Proposition 2.8** (Proposition 2.4 in [JK17b]). *Fix  $n \in \mathbb{N} \setminus \{0\}$ ,  $u \in \mathbb{T}_n^d$  and  $\nu_n \in \mathcal{M}_{s,n}$ . Write  $\beta_{l,u}(\nu_n) = \beta n^{-d} |\Lambda_{l,u}(\nu)| = \beta \nu_n[l](\mathbb{T}_n^d \setminus \{u\})$  for the renormalized inverse temperature in class  $l$ . Suppose that  $\Lambda_{l,u}(\nu_n) \neq \emptyset$  for all  $l$ . Then we have the representation*

$$\gamma_{n,\beta,q,(r_1,\dots,r_s)}^u(k | \nu_n) = \frac{\int \mathcal{A}_{r_k}(\pi, u, \beta_{k,u}(\nu_n)) P_{\Lambda_{k,u}(\nu_n), \beta_{k,u}(\nu_n), r_k}(\mathrm{d}\pi)}{\sum_{l=1}^s \int \mathcal{A}_{r_l}(\pi, u, \beta_{l,u}(\nu_n)) P_{\Lambda_{l,u}(\nu), \beta_{l,u}(\nu_n), r_l}(\mathrm{d}\pi)}. \quad (2.8)$$

For a classes  $l$  of size 0, i.e.  $\Lambda_{l,u}(\nu_n) = \emptyset$ , the weight

$$\int \mathcal{A}_{r_l}(\pi, u, \beta_{l,u}(\nu_n)) P_{\Lambda_{l,u}(\nu_n), \beta_{l,u}(\nu_n), r_l}(\mathrm{d}\pi),$$

should be replaced with  $\mathcal{A}_{r_l}(\pi, u, 0) = r_k$ , consistent with the fact that  $\Lambda_{l,u}(\nu_n) = \emptyset$  implies that  $\beta_{l,u}(\nu_n) = 0$ .

**Remark 2.9.** In [JK17b], classes of size 0 were not treated. The proof of the representation, however, shows that indeed the weight equals the size of the class.

We find that, to study the limiting behaviour of the kernels, we need to study the limiting behaviour of the measures  $P_{\Lambda_{l,u}(\nu_n), \beta_{l,u}(\nu_n), r_l}$  when  $\nu_n \rightarrow \nu$  and, additionally, continuity properties of  $\mathcal{A}_r$  when  $(\beta_n, u_n) \rightarrow (\beta, u)$ . We start by considering  $\mathcal{A}_r$ .

**Lemma 2.10.** *Let  $r \geq 1$  and let  $\mathcal{A}_r$  be the map defined in Definition 2.7. Then*

- (a) *For each pair  $(u, \beta) \in \mathbb{T}^d \times \mathbb{R}^+$  the map  $\pi \mapsto \mathcal{A}_r(\pi, u, \beta)$  is an element of  $C_b(\mathcal{P}(\mathbb{T}^d \times \{1, \dots, r\}))$ ,*

(b) Let  $\{(u_n, \beta_n)\}_{n \geq 1}$  be a sequence of pairs in  $\mathbb{T}^d \times \mathbb{R}^+$  converging to  $(u, \beta) \in \mathbb{T}^d \times \mathbb{R}^+$ , then  $\mathcal{A}_r(\cdot, u_n, \beta_n)$  converges uniformly as a function on  $\mathcal{P}(\mathbb{T}^d \times \{1, \dots, r\})$  to the function  $\mathcal{A}_r(\cdot, u, \beta)$ .

We immediately obtain the following result, corresponding to the fact that  $\int f_n d\mu_n \rightarrow \int f d\mu$  if  $f_n$  converges uniformly to  $f$  and  $\mu_n$  weakly to  $\mu$ .

**Lemma 2.11.** *Let  $\nu_n \rightarrow \nu$  and  $u_n \rightarrow u$ . Suppose that for each fuzzy class  $l \in \{1, \dots, s\}$  with  $\nu[l](\mathbb{T}^d) > 0$  there exists some measure  $P^l$  such that*

$$P_{\Lambda_l, u_n(\nu_n), \beta_l, u_n(\nu_n), r_l} \rightarrow P^l$$

weakly, *see Definition 2.7 (a),(b). Then*

$$\gamma_{n, \beta, q, (r_1, \dots, r_s)}^{u_n}(k | \nu_n) \rightarrow \frac{\int \mathcal{A}_{r_k}(\pi, u, \beta_k(\nu)) P^k(d\pi)}{\sum_{l=1}^s \int \mathcal{A}_{r_l}(\pi, u, \beta_l(\nu)) P^l(d\pi)}.$$

As above, if a fuzzy class  $l$  has no limiting mass, i.e.  $\nu[l](\mathbb{T}^d) = 0$ , then the weight should be replaced by  $\mathcal{A}_{r_l}(\cdot, u, 0) = r_l$ .

In the situation of the lemma above, we will denote by

$$P_n^l := P_{\Lambda_l, u_n(\nu_n), \beta_l, u_n(\nu_n), r_l},$$

with the right-hand side as defined in Definition 2.7 (a),(b), the distribution of the colour-profile associated with fuzzy class  $l$  on the discrete torus of size  $n$ . Consistent with the notation in the lemma above we write  $P^l := \lim_{n \rightarrow \infty} P_n^l$ , if the limit exists in the weak topology.

To study the limiting properties of the sequences  $(P_n^l)_{n \in \mathbb{N}}$ , we proceed with the large deviation principle proven in [JK17b] for the measures  $P_{\Lambda_l, u_n(\nu_n), \beta_l, u_n(\nu_n), r_l}$ .

**Definition 2.12.** We say that a sequence of volumes  $\Lambda_n \subseteq \Delta_n^d = n\mathbb{T}_n^d$  converges to some measure  $\rho \in \mathcal{M}_+(\mathbb{T}^d)$ , denoted by  $\Lambda_n \Rightarrow \rho$  if the empirical measure

$$\frac{1}{n^d} \sum_{i \in \Lambda_n} \delta_{x/n} \in \mathcal{M}_+(\mathbb{T}^d)$$

converges weakly to  $\rho$ . Abusing notation, if  $\rho$  has a density with respect to the Lebesgue measure, then we will denote this density by  $\rho$ .

**Proposition 2.13** (Proposition 2.5 in [JK17b]). *(Diluted version of LDP for empirical color profiles). Consider a sequence of inverse temperatures  $\tilde{\beta}_n \rightarrow \tilde{\beta}$  and a sequence of diluted sets  $\Lambda_n \subseteq \Delta_n^d$  with  $\Lambda_n \Rightarrow \rho$  for some Lebesgue density  $\rho$  with  $N_\rho := \rho\lambda(\mathbb{T}^d) > 0$ . Denote  $\tilde{\rho}(u) := N_\rho^{-1}\rho(u)$ , then the measures  $\mu_{\Lambda_n, \tilde{\beta}_n, q} \circ (\pi_{\Lambda_n})^{-1}$  satisfy an LDP on the space  $\mathcal{P}(\mathbb{T}^d \times \{1, \dots, q\})$  with rate  $|\Lambda_n|$  and rate function  $I_{\tilde{\rho}, \tilde{\beta}} - \inf_{\varphi \in \mathcal{P}(\mathbb{T}^d \times \{1, \dots, q\})} I_{\tilde{\rho}, \tilde{\beta}}(\varphi)$  where*

$$I_{\tilde{\rho}, \tilde{\beta}}(\varphi) = \begin{cases} -\tilde{\beta} \sum_{a=1}^q \langle J * \tilde{\rho}\alpha[a], \tilde{\rho}\alpha[a] \rangle + \langle S(\alpha|eq), \tilde{\rho}\lambda \rangle & \text{if } \varphi[a] = \tilde{\rho}\alpha[a]\lambda, \alpha \in B \\ \infty & \text{otherwise.} \end{cases} \quad (2.9)$$

**Remark 2.14.** Proposition 2.5 in [JK17b] was originally proven for fixed  $\tilde{\beta}$ . Changing the result to include  $\tilde{\beta}_n \rightarrow \tilde{\beta}$  follows immediately from Varadhan's lemma for a uniformly converging sequence of bounded continuous functions.

## 2.5 Minimizers of the rate function for Ising classes and a limiting form of the single-site kernels

A more careful analysis of the rate function for the Ising classes gives an extension of the representation of the limiting kernels of Theorem 2.7 in [JK17b]. For any  $\varphi = \alpha\tilde{\rho}\lambda$ , we first re-express  $\alpha$  in terms of the local magnetization

$$m^\alpha(u) := \alpha[1](u) - \alpha[2](u).$$

**Proposition 2.15.** *Consider the rate function in Proposition 2.13 for an Ising class with  $\Lambda_n \Rightarrow \rho \neq 0$ . Denote by  $P_n := \mu_{\Lambda_n, \tilde{\rho}_n, 2} \circ (\pi_{\Lambda_n})^{-1}$  the distribution of the colour profile associated to the Ising class on the discrete torus of size  $n$ . Then the following two situations can occur.*

- (a) *There is exactly one global minimizer  $\varphi^* = \varphi^*(\tilde{\rho}, \tilde{\beta})$  for  $I_{\tilde{\rho}, \tilde{\beta}}$ , which is the spatially homogeneous equi-distribution, i.e. corresponding to local magnetization profile  $m_{\tilde{\rho}, \tilde{\beta}} = 0$ . We have  $P_n \rightarrow \delta_{\varphi^*}$ .*
- (b) *There are exactly two global minimizers  $\varphi^{*,+} = \varphi^{*,+}(\tilde{\rho}, \tilde{\beta})$  and  $\varphi^{*,-} = \varphi^{*,-}(\tilde{\rho}, \tilde{\beta})$  for  $I_{\tilde{\rho}, \tilde{\beta}}$ . Let  $m_{\tilde{\rho}, \tilde{\beta}}$  denote the local magnetization of  $\varphi^{*,+}$ . Then  $m_{\tilde{\rho}, \tilde{\beta}}$  is everywhere positive, and the local magnetization of  $\varphi^{*,-}$  equals  $-m_{\tilde{\rho}, \tilde{\beta}}$ . We have  $P_n \rightarrow \frac{1}{2}(\delta_{\varphi^{*,+}} + \delta_{\varphi^{*,-}})$ .*

Thus, for classes of size 2, and  $\rho \neq 0$ , we have a well defined local magnetization  $m_{\tilde{\rho}, \tilde{\beta}}$ . For a class of size 2 and  $\rho = 0$ , set  $m_{\tilde{\rho}, \tilde{\beta}} = 0$ . Define for each  $r \geq 2$  the function

$$D_r : \mathbb{T}^d \times P_r(B) \times \mathbb{R}^+ \rightarrow [1, \infty), \quad D_r(u, \tilde{\rho}, \tilde{\beta}) := \mathbb{1}_{\{r \neq 2\}} + \mathbb{1}_{\{r=2\}} \frac{1}{\sqrt{1 - m_{\tilde{\rho}, \tilde{\beta}}(u)^2}}.$$

**Theorem 2.16.** *Consider the setting of Theorem 2.6. Let  $\nu_n \rightarrow \nu$  and  $u_n \rightarrow u$ . Suppose the parameters are in setting (a) or (b). Then the limiting conditioning kernel is given by*

$$\lim_{n \rightarrow \infty} \gamma_{n, \beta, q, (r_1, \dots, r_s)}^u(k | \nu_n) = \frac{r_k \exp \left\{ 2\beta r_k^{-1} \int dv \nu[k](v) J(u-v) \right\} D_{r_k}(u, \tilde{\nu}_k, \tilde{\beta}_k)}{\sum_{i=1}^s r_i \exp \left\{ 2\beta r_i^{-1} \int dv \nu[i](v) J(u-v) \right\} D_{r_i}(u, \tilde{\nu}_i, \tilde{\beta}_i)}$$

where  $\tilde{\nu}_i := \nu[i](\mathbb{T}^d)^{-1} \nu[i]$  and  $\tilde{\beta}_i = \nu[i](\mathbb{T}^d) \beta$ .

## 3 Proofs of main results

We start by proving our main Theorem 2.6 using Lemmas 2.10, 2.11 and Proposition 2.15. We prove the two lemmas immediately afterwards.

Then, in Sections 3.3 to 3.6, we analyse the set of minimizers of the rate function for Ising classes which leads to a proof of Proposition 2.15. We conclude in Section 3.5 with a proof of Theorem 2.16..

### 3.1 Proof of Theorem 2.6

We proceed with the proof of our main result, which is based on Lemmas 2.10 and 2.11 and Proposition 2.15.

*Proof of Theorem 2.6.* First we prove (a), i.e. all classes have size 1 or 2. Let  $\nu_n \rightarrow \nu$  and  $u_n \rightarrow u$ . For classes  $l \in \{1, \dots, s\}$  and  $n \geq 1$  such that  $\Lambda_{l, u_n}(\nu_n) \neq \emptyset$ , recall the notation  $P_n^l := P_{\Lambda_{l, u_n}(\nu_n), \beta_{l, u_n}(\nu_n), r_l}$ .

To prove our result, it suffices to verify the conditions for Lemma 2.11. Note that we do not need to consider classes with no limiting mass. Thus, without loss of generality, we assume all fuzzy classes have non-zero limiting mass.

First suppose that class  $l$  has size 1. Then it follows that the measures  $P_n^l$  converge to  $\tilde{\nu}_l$ . For classes of size 2 with non-zero mass in the limit, we find by Proposition 2.13 with  $\Lambda_n = \Lambda_{l, u_n}(\nu_n)$ ,  $\tilde{\beta}_n = \beta_{l, u_n}(\nu_n)$  and  $\rho = \nu[l]$  that  $P_n^l$  satisfies a large deviation principle with good rate function. By Proposition 2.15 there is a measure  $P^l$ , independent of the sequences  $\nu_n$  and  $u_n$ , such that  $P_n^l \rightarrow P^l$ . Thus Lemma 2.11 implies the limit of the kernels exists and is independent of the sequences  $\nu_n$  and  $u_n$ .

Cases (b) and (c) follow from the arguments of [JK17b], combined with the result of the present paper that even when Ising-classes are present, they do not provide a source of discontinuity. □

### 3.2 Proof of Lemmas 2.10 and 2.11

*Proof of Lemma 2.10.* Fix  $r \geq 1$  Recall that  $\mathcal{A}_r : \mathcal{P}(\mathbb{T}^d \times \{1, \dots, r\}) \times \mathbb{T}^d \times \mathbb{R}^+ \rightarrow [1, \infty)$  was defined as

$$\mathcal{A}_r(\pi, u, \beta) := \sum_{i=1}^r \exp \{2\beta (J * \pi[i])(u)\}. \quad (3.1)$$

We start with proving (a), the continuity of  $\pi \mapsto \mathcal{A}_r(\pi, u, \beta)$  for all  $(u, \beta) \in \mathbb{T}^d \times \mathbb{R}^+$ . Clearly, this result follows if  $\pi \mapsto (J * \pi[i])(u)$  is continuous for all  $i$ . But this is immediate as  $J$  is a continuous function and  $\mathcal{M}_+(\mathbb{T}^d)$  is equipped with the weak topology.

We proceed with the proof of (b). Let  $(u_n, \beta_n) \rightarrow (u, \beta)$ . Note that the uniform convergence, again, follows by proving that the function  $f_{n,i}(\pi) := (J * \pi[i])(u_n)$  converges uniformly to  $f_i(\pi) := (J * \pi[i])(u)$ . This, however, follows immediately from the uniform continuity of the function  $J(\mathbb{T}^d)$  is compact). □

*Proof of Lemma 2.11.* Let  $\nu_n \rightarrow \nu$  and  $u_n \rightarrow u$ . By the representation for the kernels given in Proposition 2.8, using that the weight of class  $l$  is bounded below by  $r_l$ , we find that the limiting statement holds if for all classes  $l \in \{1, \dots, s\}$ , we have that

$$\int \mathcal{A}_{r_l}(\pi, u_n, \beta_{l, u_n}(\nu_n)) P_n^l(d\pi) \rightarrow \int \mathcal{A}_{r_l}(\pi, u, \beta_l) P^l(d\pi), \quad (3.2)$$

where the weights need to be replaced by  $r_l$  if the set  $\Lambda_{l,u_n}(\nu_n) = \emptyset$  on the left-hand side, or if  $\nu[l](\mathbb{T}^d)$  on the right-hand side.

Fix  $l$ . First, suppose that  $\nu[l](\mathbb{T}^d) = 0$  and hence  $\Lambda_{l,u_n}(\nu_n) \Rightarrow 0$ . If the set  $\Lambda_{l,u_n}(\nu_n) = \emptyset$ , then the weight in (2.8) equals  $r_l$ . If the set is not empty, then, we need to consider the integral

$$\int \mathcal{A}_{r_l}(\pi, u, \beta_{l,u_n}(\nu_n)) P_{\Lambda_{l,u_n}(\nu_n), \beta_{l,u_n}(\nu_n), r_l}(\mathrm{d}\pi).$$

Regardless of which case we have, we have that  $\beta_{l,u_n}(\nu_n) \rightarrow 0$ , and **by definition of  $\mathcal{A}_{r_l}$ , see (3.1)**,

$$\mathcal{A}_{r_l}(\cdot, \beta_{l,u_n}(\nu_n), u_n) \rightarrow \mathcal{A}_{r_l}(\cdot, 0, u) = r_l$$

uniformly by Lemma 2.10. This implies that the weight of class  $l$  has a unique limit, which equals  $r_l$ .

Next, we assume that  $\nu[l](\mathbb{T}^d) > 0$ . This implies that for sufficiently large  $n$ , the set  $\Lambda_{l,u_n}(\nu_n) \neq \emptyset$ . As  $\beta_{l,u_n}(\nu_n) = \beta n^{-d} |\Lambda_{l,u_n}(\nu_n)| = \beta \nu_n[l](\mathbb{T}^d)$ , i.e. we find by the weak convergence of  $\nu_n \rightarrow \nu$  that  $\beta_{l,u_n}(\nu_n) \rightarrow \beta_l := \beta \nu[l](\mathbb{T}^d)$ . Because also  $u_n \rightarrow u$ , we find that the continuous functions  $\mathcal{A}_{r_l}(\cdot, u_n, \beta(\nu_n))$  converge uniformly to the continuous function  $\mathcal{A}_{r_l}(\cdot, u, \beta)$  by Lemma 2.10. Thus, the weak convergence  $P_n^l \rightarrow P^l$  establishes (3.2).  $\square$

### 3.3 Preliminaries for the Ising-class analysis

Before identifying minimizers of the rate function of Proposition 2.13 for Ising classes, we first reparametrize our Ising profiles in terms of the magnetization of the profile. Afterwards, we rewrite our rate function in terms of a local term, analogous to that of the Curie-Weiss model, and a global term that expresses the non-local interactions.

Consider  $\tilde{\beta} \geq 0$  and a profile  $\rho\lambda \in \mathcal{M}_+(\mathbb{T}^d)$  with Lebesgue density  $\rho$  and assume that  $N_\rho := \rho\lambda(\mathbb{T}^d) > 0$ . Set  $\tilde{\rho} = N_\rho^{-1}\rho$  as in Proposition 2.13. We study the minimizers of  $I_{\tilde{\rho},\tilde{\beta}}$ , which is equivalent to studying the minimizers of the rate function  $I_{\tilde{\rho},\tilde{\beta}} - \inf_\varphi I_{\tilde{\rho},\tilde{\beta}}(\varphi)$ , where

$$I_{\tilde{\rho},\tilde{\beta}}(\varphi) = \begin{cases} \int \mathrm{d}u \tilde{\rho}(u) \left\{ -\tilde{\beta} \int \mathrm{d}v \tilde{\rho}(v) (\alpha[1](u)\alpha[1](v) + \alpha[2](u)\alpha[2](v)) J(u-v) \right. \\ \quad \left. + S(\alpha[\cdot](u) | eq) \right\} & \text{if } \varphi = \alpha\tilde{\rho}\lambda, \alpha \in B_2 \\ \infty & \text{else} \end{cases}$$

Thus,  $I_{\tilde{\rho}}(\varphi)$  is expressed in terms of  $\alpha \in B_2$ , where  $\varphi = \alpha\tilde{\rho}\lambda$ . First, we will rewrite the rate function in terms of the local magnetization  $m$  in the closed ball

$$\text{Ball}(L^\infty) := \left\{ m \in L^\infty(\mathbb{T}^d, \lambda^d) \mid \text{ess sup } |m| \leq 1 \right\},$$

defined by  $m(u) := \alpha[1](u) - \alpha[2](u)$ . To re-express the diluted rate function in terms of magnetization functions, we express the quadratic term in terms of  $m$ :

$$\alpha[1](v)\alpha[1](u) + \alpha[2](u)\alpha[2](v) = \frac{1 + m(u)m(v)}{2}.$$

Substituting the representation into the rate function gives

$$I_{\tilde{\rho},\tilde{\beta}}(\varphi) = \int du \tilde{\rho}(u) \left\{ -\frac{\tilde{\beta}}{2} \int dv \tilde{\rho}(v) (1 + m(u)m(v)) J(u-v) + I(m(u)) \right\}$$

where  $I$  denotes the entropy-term in the Curie-Weiss rate function

$$I(x) = \begin{cases} \frac{1+x}{2} \log(1+x) + \frac{1-x}{2} \log(1-x) & \text{if } x \in (-1, 1), \\ \log(2) & \text{if } x \in \{-1, +1\}. \end{cases}$$

Note, that, since the final rate function equals  $I_{\tilde{\rho},\tilde{\beta}}$  plus some constant we can omit any terms in the integrand which do not depend on  $m$ . We will further write  $I_{\tilde{\rho},\tilde{\beta}}(m)$  instead of  $I_{\tilde{\rho},\tilde{\beta}}(\varphi)$  if  $\varphi$  has a magnetization represented by  $m$ . We conclude that it suffices to study minimizers of the functional

$$I_{\tilde{\rho},\tilde{\beta}}(m) = \int du \tilde{\rho}(u) \left\{ -\frac{\tilde{\beta}}{2} \int dv \tilde{\rho}(v) m(u)m(v) J(u-v) + I(m(u)) \right\} \quad (3.3)$$

on the set  $\text{Ball}(L^\infty)$ . Analogous to the discussion in the case of all classes of size at least three in [JK17b], we may rewrite the rate function (3.3) by

$$I_{\tilde{\rho},\tilde{\beta}}(m) = \frac{\beta}{4} \int du \tilde{\rho}(u) \int dv \tilde{\rho}(v) [m(u) - m(v)]^2 J(u-v) + \int du \tilde{\rho}(u) \left[ -\frac{1}{2} b_{\tilde{\beta},\tilde{\rho},J}(u) m^2(u) + I(m(u)) \right] \quad (3.4)$$

with the local inverse temperature at site  $u$  given by the convolution

$$b_{\tilde{\beta},\tilde{\rho},J}(u) := \tilde{\beta} \int dv \tilde{\rho}(v) J(u-v). \quad (3.5)$$

To elucidate the equivalence between the bracketed expression in the second term of  $I_{\tilde{\rho}}$  and the Curie-Weiss rate function at (site-dependent) inverse temperature  $b_{\tilde{\beta},\tilde{\rho},J}(u)$  and magnetization  $m(u)$ , we further use the notation

$$\Phi_u(m) := -\frac{1}{2} b_{\tilde{\beta},\tilde{\rho},J}(u) m^2 + I(m), \quad u \in \mathbb{T}^d, m \in [-1, 1].$$

We see that the rate functional expresses the competition between the local Curie-Weiss term and a global term that penalizes spatial inhomogeneity.

To visualize this competition, Figure 1 shows three different magnetization profiles on the one-dimensional torus at fixed interaction-function  $J$ , (conditioning) density  $\tilde{\rho}$  and inverse temperature  $\tilde{\beta}$ . The profile  $m_{\text{loc}}$  (see Lemma 3.1 below) is given as

$$m_{\text{loc}}(u) := \arg \min_{m \in [0,1]} (\Phi_u(m)),$$

whereas  $m_{\text{flat}}$  is the minimizer to the local Curie-Weiss term in the class of non-negative spatial homogeneous profiles. The profile  $m_{\text{stat}}$  arises from the necessary condition of vanishing gradient at minimizing profiles to  $I_{\tilde{\beta},\tilde{\rho}}$ .

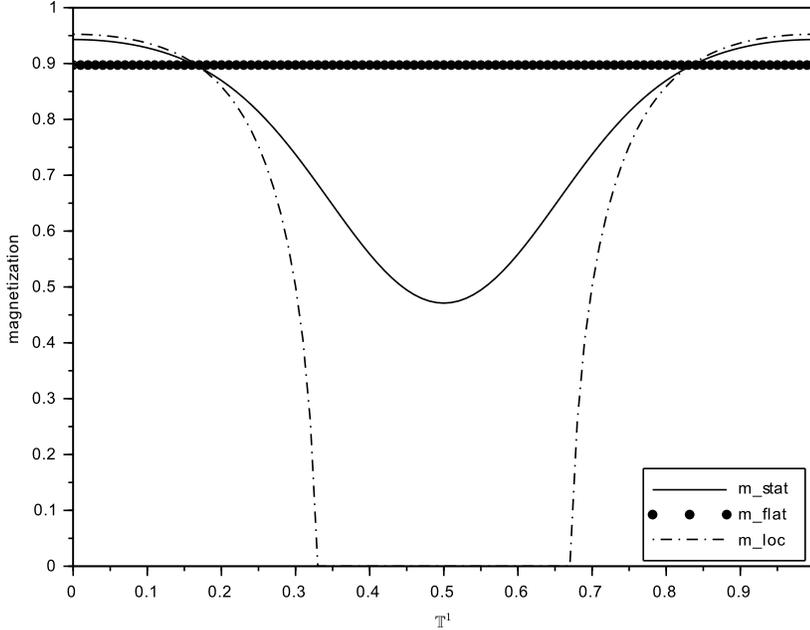


Figure 1:  $\tilde{\beta} = 1.3$ ,  $\tilde{\rho}(\cdot) = 1 + \cos(2\pi\cdot)$ ,  $J(\cdot) = 1 + \cos(2\pi\cdot)$

### 3.4 Identifying the minimizers of the rate function for Ising classes

We exploit the representation of our rate function in terms of the local magnetization  $m$  to study minimizers of  $I$ . We first argue based on the assumption that a local minimizer exists. Based on this assumption and natural symmetry properties in the rate function, we show that we can find a local minimizer with lower cost that is non-negative. In addition, we show, using the decomposition of  $I$  into a local and global term that local minimizers are bounded away from  $-1$  and  $1$ .

**Lemma 3.1.** *A local (global) minimizer  $m$  of the rate function  $I_{\tilde{\rho}, \tilde{\beta}}$  has the following properties:*

- (a) *By symmetry of  $I_{\tilde{\rho}, \tilde{\beta}}$ , the profile  $-m$  is also a local (global) minimizer.*
- (b) *We can always find a non-negative profile  $\tilde{m}$ , such that  $I_{\tilde{\rho}, \tilde{\beta}}(\tilde{m}) \leq I_{\tilde{\rho}}(m)$ . if both  $\tilde{\rho}\lambda(\{m > 0\}) > 0$  and  $\tilde{\rho}\lambda(\{m < 0\}) > 0$  then the inequality is strict.*
- (c)  *$m$  is an interior point of  $\text{Ball}(L^\infty)$ , i.e.  $\text{ess sup}(|m|) < 1$ .*

*Proof.* (a) is immediate. For (b) consider the profile  $\tilde{m} := |m|$  given as the pointwise absolute value of  $m$ . The entropy functional  $I(\cdot)$  is symmetric, implying

$$\int du \tilde{\rho}(u) I(\tilde{m}(u)) = \int du \tilde{\rho}(u) I(m(u)).$$

At any  $(u, v) \in \mathbb{T}^d \times \mathbb{T}^d$  we further have

$$-\tilde{m}(u)\tilde{m}(v) \begin{cases} = -m(u)m(v) & \text{if } (u, v) \in \{m \geq 0\}^2 \cup \{m < 0\}^2, \\ < -m(u)m(v) & \text{else,} \end{cases}$$

which leads to

$$\begin{aligned} & \int du \tilde{\rho}(u) \left\{ -\frac{\tilde{\beta}}{2} \int dv \tilde{\rho}(v) \tilde{m}(u)\tilde{m}(v) J(u-v) \right\} \\ & \leq \int du \tilde{\rho}(u) \left\{ -\frac{\tilde{\beta}}{2} \int dv \tilde{\rho}(v) m(u)m(v) J(u-v) \right\} \end{aligned}$$

with the inequality being strict if  $\tilde{\rho}\lambda(\{m > 0\}) > 0$  and  $\tilde{\rho}\lambda(\{m < 0\}) > 0$ . We conclude

$$I_{\tilde{\rho}, \tilde{\beta}}(\tilde{m}) \leq I_{\tilde{\rho}, \tilde{\beta}}(m)$$

with strict inequality in the case discussed above.

For the proof of (c) let  $m \geq 0$  be any magnetization-profile with  $\text{ess sup}(m) = 1$ . Consider the representation (3.4) of  $I_{\tilde{\rho}, \tilde{\beta}}$  and define

$$m_{\text{loc}}(u) := \arg \min_{m \in [0, 1]} (\Phi_u(m)).$$

By continuity, the local inverse temperature  $b_{\tilde{\rho}, J, \tilde{\rho}}$  is bounded on  $\mathbb{T}^d$ , so  $m_{\text{loc}}$ , given as non-negative solution to the mean-field equation

$$m_{\text{loc}}(u) = \tanh(b_{\tilde{\rho}, J, \tilde{\rho}}(u)m_{\text{loc}}(u)), \quad (3.6)$$

is bounded from above by a constant  $c < 1$ . Since  $\text{ess sup}(m) = 1$ , we clearly have  $\tilde{\rho}\lambda(m > c) > 0$ . Thus, the profile  $\tilde{m}(\cdot) = m(\cdot) \wedge c$  is an interior point of  $\text{Ball}(L^\infty)$ . By construction, the contribution of the local part of (3.4) for  $\tilde{m}$  is lower than for  $m$ . By a straightforward verification, the same follows for the global part, as  $|\tilde{m}(u) - \tilde{m}(v)| \leq |m(u) - m(v)|$  for all  $u, v$ . We find  $I_{\tilde{\rho}, \tilde{\beta}}(\tilde{m}) < I_{\tilde{\rho}, \tilde{\beta}}(m)$ .  $\square$

By (c) of Lemma 3.1, we know that minimizers lie in the interior of  $\text{Ball}(L^\infty)$ . This implies we can differentiate the function  $I_{\tilde{\rho}, \tilde{\beta}}$  to get further conditions on minimizers. We start with a technical lemma that identifies the gradient of  $I_{\tilde{\rho}}$ .

**Lemma 3.2.** *For any two magnetization-profiles  $m_1$  and  $m_2$  and any  $t \in [0, 1]$  such that there is an open neighborhood  $t \in U \subseteq \mathbb{R}$  with*

$$m_1 + \hat{t}m_2 \in \text{Ball}(L^\infty) \quad \text{for all } \hat{t} \in U$$

the derivative  $\frac{d}{d\hat{t}} I_{\tilde{\rho}, \tilde{\beta}}(m_1 + \hat{t}m_2)|_{\hat{t}=t}$  exists and is given by

$$\begin{aligned} & \frac{d}{d\hat{t}} I_{\tilde{\rho}, \tilde{\beta}}(m_1 + \hat{t}m_2)|_{\hat{t}=t} \\ & = \int du \tilde{\rho}(u) m_2(u) \left\{ -\tilde{\beta} \int dv \tilde{\rho}(v) (m_1 + tm_2)(v) J(u-v) + I'((m_1 + tm_2)(u)) \right\}. \end{aligned}$$

*Proof.* We have

$$\begin{aligned} & \frac{d}{d\tilde{t}} I_{\tilde{\rho}}(m_1 + \tilde{t}m_2)|_{\tilde{t}=t} \\ &= -\frac{\tilde{\beta}}{2} \frac{d}{d\tilde{t}} \int du \tilde{\rho}(u) \int dv \tilde{\rho}(v) (m_1 + \tilde{t}m_2)(u) (m_1 + \tilde{t}m_2)(v) J(u-v)|_{\tilde{t}=t} \\ & \quad + \frac{d}{d\tilde{t}} \int du \tilde{\rho}(u) I((m_1 + \tilde{t}m_2)(u))|_{\tilde{t}=t}. \end{aligned}$$

We first consider the first term on the right-hand side. As the integrand in the first term is continuously differentiable in  $\tilde{t}$  and using the fact that both  $m_1$  and  $m_2$  are essentially bounded, we can interchange integration and differentiation, cf. [Kle14, Lemma 6.28] or [Bog07, Exercise 5.8.135]:

$$\begin{aligned} & -\frac{\tilde{\beta}}{2} \int du \tilde{\rho}(u) \int dv \tilde{\rho}(v) \frac{d}{d\tilde{t}} \left\{ (m_1 + \tilde{t}m_2)(u) (m_1 + \tilde{t}m_2)(v) J(u-v) \right\} |_{\tilde{t}=t} \\ &= -\frac{\tilde{\beta}}{2} \int du \tilde{\rho}(u) m_2(u) \int dv \tilde{\rho}(v) (m_1 + tm_2)(v) J(u-v) \\ & \quad - \frac{\tilde{\beta}}{2} \int du \tilde{\rho}(u) (m_1 + tm_2)(u) \int dv m_2(v) J(u-v). \end{aligned}$$

Using the Fubini-Tonelli-theorem on the second term on the right-hand side and then switching roles of  $u$  and  $v$  we find

$$\begin{aligned} & -\frac{\tilde{\beta}}{2} \int du \tilde{\rho}(u) m_2(u) \left\{ \int dv \tilde{\rho}(v) (m_1 + tm_2)(v) (J(u-v) + J(v-u)) \right\} \\ &= -\tilde{\beta} \int du \tilde{\rho}(u) m_2(u) \left\{ \int dv \tilde{\rho}(v) (m_1 + tm_2)(v) J(u-v) \right\}. \end{aligned}$$

In the same way we have:

$$\begin{aligned} \frac{d}{d\tilde{t}} \int du \tilde{\rho}(u) I((m_1 + \tilde{t}m_2)(u))|_{\tilde{t}=t} &= \int du \tilde{\rho}(u) \frac{d}{d\tilde{t}} \{ I((m_1 + \tilde{t}m_2)(u)) \} |_{\tilde{t}=t} \\ &= \int du \tilde{\rho}(u) m_2(u) I'(m_1(u) + tm_2(u)), \end{aligned}$$

concluding the proof of the lemma.  $\square$

By differentiation of  $I_{\tilde{\rho}, \tilde{\beta}}$  at a local minimizer, we immediately find the following necessary condition for minimizers.

**Lemma 3.3.** *A minimizer  $m$  to  $I_{\tilde{\rho}, \tilde{\beta}}$  satisfies the stationarity equation*

$$-\tilde{\beta} \int dv \tilde{\rho}(v) m(v) J(u-v) + I'(m(u)) = 0 \quad \text{for } \tilde{\rho}\lambda\text{-a.e. } u \in \mathbb{T}^d, \quad (3.7)$$

or equivalently,

$$m(u) = \tanh \left( \tilde{\beta} \int dv \tilde{\rho}(v) m(v) J(u-v) \right) \quad \text{for } \tilde{\rho}\lambda\text{-a.e. } u \in \mathbb{T}^d. \quad (3.8)$$

*Proof.* Let  $m$  be a local minimum and  $\tilde{m}$  an arbitrary profile. By Lemma 3.1 we may assume  $m$  to be an interior point of (3.2), so, by Lemma 3.2 the derivative

$$\frac{d}{dt} I_{\tilde{\rho}, \tilde{\beta}}(m + t\tilde{m})|_{t=0}$$

exists and must vanish for  $m$  being a local minimum of  $I_{\tilde{\rho}, \tilde{\beta}}$ . Rewriting  $I' = \operatorname{arctanh}$  leads to the second expression.  $\square$

An immediate consequence of the stationarity equation is a further restriction of the set of possible minimizers.

**Lemma 3.4.** *A nontrivial non-negative minimizer to  $I_{\tilde{\rho}, \tilde{\beta}}$  is strictly positive.*

*Proof.* By assumption,  $J > 0$ , which implies by the stationarity equation (3.8) that a minimizers is strictly positive.  $\square$

For the arguments that follow, we only need one side of the stationarity equation.

**Definition 3.5.** We say that  $m$  satisfies the *stationarity inequality* if

$$-\tilde{\beta} \int dv \tilde{\rho}(v) m(v) J(u-v) + I'(m(u)) \leq 0 \quad \text{for } \tilde{\rho}\lambda\text{-a.e. } u \in \mathbb{T}^d, \quad (3.9)$$

or equivalently,

$$m(u) \leq \tanh \left( \tilde{\beta} \int dv \tilde{\rho}(v) m(v) J(u-v) \right) \quad \text{for } \tilde{\rho}\lambda\text{-a.e. } u \in \mathbb{T}^d.$$

We start with a technical result based on the stationarity inequality and convexity that, afterwards, will allow us to establish the uniqueness of non-negative minimizers.

**Lemma 3.6.** *For any two different (i.e.  $\tilde{\rho}\lambda(\{m_1 \neq m_2\}) > 0$ ) non-negative solutions  $m_1$  and  $m_2$  to the stationarity inequality (3.9) we have*

$$\tilde{\rho}\lambda(\{m_2 < m_1\}) = 0 \quad \Rightarrow \quad I_{\tilde{\rho}}(m_2) < I_{\tilde{\rho}}(m_1).$$

*Proof.* Assume, that  $m_1$  and  $m_2$  are two different non-negative solutions to (3.9) where  $\tilde{\rho}\lambda(\{m_2 < m_1\}) = 0$  (so,  $\tilde{\rho}\lambda(\{m_2 > m_1\}) > 0$ ). Consider the linear interpolation

$$F : [0, 1] \times \mathbb{T}^d \rightarrow [0, 1] \quad ; \quad F(t, u) := (1-t)m_1(u) + tm_2(u).$$

Then  $I_{\tilde{\rho}, \tilde{\beta}}F$  is a continuous, differentiable function in  $t$ , so at any  $0 \leq s \leq 1$  the Fundamental theorem of calculus gives

$$I_{\tilde{\rho}, \tilde{\beta}}((1-s)m_1 + sm_2) - I_{\tilde{\rho}}(m_1) = \int_0^s (I_{\tilde{\rho}, \tilde{\beta}}F)'(t) dt$$

where by Lemma 3.2

$$\begin{aligned} & (I_{\tilde{\rho}, \tilde{\beta}} F)'(t) \\ &= \int du \tilde{\rho}(u) (m_2(u) - m_1(u)) \left\{ -\tilde{\beta} \int dv \tilde{\rho}(v) F(t, v) J(u - v) + I'(F(t, u)) \right\}. \end{aligned} \quad (3.10)$$

Inserting  $\pm((1-t)I'(m_1(u)) + tI'(m_2(u)))$ , we may restate the bracketed expression in (3.10) in terms of the stationarity inequality:

$$\begin{aligned} & -\tilde{\beta} \int dv \tilde{\rho}(v) F(t, v) J(u - v) + I'(F(t, u)) \\ &= \left( -\tilde{\beta} \int dv \tilde{\rho}(v) F(t, v) J(u - v) + (1-t)I'(m_1(u)) + tI'(m_2(u)) \right) \\ &\quad + I'(F(t, u)) - ((1-t)I'(m_1(u)) + tI'(m_2(u))) \\ &= (1-t) \left( -\tilde{\beta} \int dv \tilde{\rho}(v) m_1(v) J(u - v) + I'(m_1(u)) \right) \\ &\quad + t \left( -\tilde{\beta} \int dv \tilde{\rho}(v) m_2(v) J(u - v) + I'(m_2(u)) \right) \\ &\quad + I'(F(t, u)) - ((1-t)I'(m_1(u)) + tI'(m_2(u))). \end{aligned}$$

Thus, using that  $m_1, m_2$  satisfy the stationarity inequality (3.9), we find

$$\begin{aligned} & -\tilde{\beta} \int dv \tilde{\rho}(v) F(t, v) J(u - v) + I'(F(t, u)) \\ &\leq I'((1-t)m_1(u) + tm_2(u)) - ((1-t)I'(m_1(u)) + tI'(m_2(u))). \end{aligned}$$

The function  $I'(\cdot) = \operatorname{arctanh}(\cdot)$  is strictly convex on  $[0, \infty)$ . Using that  $\tilde{\rho}\lambda(\{m_2 < m_1\}) = 0$  and  $\tilde{\rho}\lambda(\{m_2 > m_1\}) > 0$ , we find by (3.10) that for all  $t \in (0, 1)$ :

$$(I_{\tilde{\rho}, \tilde{\beta}} F)'(t) < 0$$

which establishes the claim.  $\square$

**Lemma 3.7.** *A non-negative global minimizer to the rate function  $I_{\tilde{\rho}, \tilde{\beta}}$  is unique.*

*Proof.* Assume that there are two different non-negative global minimizers  $m_1$  and  $m_2$  for  $I_{\tilde{\rho}, \tilde{\beta}}$ . Then both of them satisfy the stationarity equation (3.7) and, hence, the stationarity inequality (3.9). Since  $m_1$  and  $m_2$  are different, at least one of them is different to the point-wise maximum  $m_1 \vee m_2$ , and both  $m_1$  and  $m_2$  are bounded from above by  $m_1 \vee m_2$ . Therefore, showing, that  $m_1 \vee m_2$  satisfies the stationarity inequality (3.9), one may apply the Lemma 3.6 to get the contradiction

$$I_{\tilde{\rho}, \tilde{\beta}}(m_1 \vee m_2) < I_{\tilde{\rho}, \tilde{\beta}}(m_1) \vee I_{\tilde{\rho}, \tilde{\beta}}(m_2) = I_{\tilde{\rho}, \tilde{\beta}}(m_1) \wedge I_{\tilde{\rho}, \tilde{\beta}}(m_2).$$

By non-negativity of  $J$  and  $\tilde{\rho}$  and the application of the stationarity inequality (3.9) to  $m_1$  and  $m_2$ , we have at  $\tilde{\rho}\lambda$ -a.e.  $u \in \mathbb{T}^d$ :

$$\begin{aligned} \tanh\left(\tilde{\beta}((\tilde{\rho}(m_1 \vee m_2)) * J)(u)\right) &\geq \tanh\left(\tilde{\beta}((\tilde{\rho}m_1) * J)(u)\right) \geq m_1(u), \\ \tanh\left(\tilde{\beta}((\tilde{\rho}(m_1 \vee m_2)) * J)(u)\right) &\geq \tanh\left(\tilde{\beta}((\tilde{\rho}m_2) * J)(u)\right) \geq m_2(u), \end{aligned}$$

so,

$$\tanh\left(\tilde{\beta}((\tilde{\rho}(m_1 \vee m_2)) * J)(u)\right) \geq (m_1 \vee m_2)(u) \quad \tilde{\rho}\lambda\text{-a.s.}$$

which concludes the proof.  $\square$

A combination of the results above allows us to identify the minimizers of  $I_{\tilde{\rho}, \tilde{\beta}}$ .

**Proposition 3.8** (Characterization of the set of minimizers to  $I_{\tilde{\rho}}$ ). *The rate function  $I_{\tilde{\rho}, \tilde{\beta}}$  is either uniquely minimized by the trivial profile  $m \equiv 0$ , or there are exactly two minimizers  $m_1$  and  $m_2$  where*

a)  $m_1$  is strictly positive and

b)  $m_2 = -m_1$ .

*Proof.* Since  $I_{\tilde{\rho}, \tilde{\beta}}$  has compact level sets, there is at least one global minimizer  $m$ . Suppose this particular minimizer  $m$  is not equal to 0. Then by Lemma 3.1 (b), there are two global minimizers  $m_1, m_2$  with  $m_1 \geq 0$  and  $m_1 = -m_2$  and in addition  $m \in \{m_1, m_2\}$ . By (c),  $m_1$  is an interior point of  $\text{Ball}(L^\infty)$ , which implies by Lemma 3.4 that  $m_1$  is strictly positive.

The uniqueness of a non-negative profile was established in Lemma 3.7.  $\square$

### 3.5 Convergence of profiles, proof of Proposition 2.15

For the proof of Proposition 2.15, we will critically exploit symmetry properties of the Potts model. We start with a short discussion to establish the permutation invariance of the profiles that are obtained from the  $r$ -state Potts model under the maps  $\pi_\Lambda$ .

Suppose our fuzzy class has  $r$  elements. Let  $s$  be an element from the permutation group of  $\{1, \dots, r\}$ . Also denote by  $s$  the map induced by  $s$  on  $\{1, \dots, r\}^\Lambda$  by acting coordinate-wise. In turn, we can define a map  $\hat{s} : \mathcal{P}(\mathbb{T}^d \times \{1, \dots, r\}) \rightarrow \mathcal{P}(\mathbb{T}^d \times \{1, \dots, r\})$ , by  $(\hat{s}\pi)(u, i) = \pi(u, s(i))$ . Recall the definition of  $\pi_\Lambda$  in (2.3). By construction, we have the following permutation property:  $\pi_\Lambda \circ s = \hat{s} \circ \pi_\Lambda$ . Also, by definition of the Potts model, we have  $\mu_{\Lambda, \beta, r} \circ s^{-1} = \mu_{\Lambda, \beta, r}$ .

Combining these statements, we obtain

$$P_{\Lambda, \beta, r} \circ \hat{s}^{-1} = \mu_{\Lambda, \beta, r} \circ \pi_\Lambda^{-1} \circ \hat{s}^{-1} = \mu_{\Lambda, \beta, r} \circ s^{-1} \circ \pi_\Lambda^{-1} = \mu_{\Lambda, \beta, r} \circ \pi_\Lambda^{-1} = P_{\Lambda, \beta, r},$$

i.e. the permutation symmetry of the Potts model carries over to that of the density profiles.

*Proof of Proposition 2.15.* The set

$$U := \left\{ Q \in \mathcal{P}(\mathcal{P}(\mathbb{T}^d \times \{1, 2\})) \mid \text{for all permutations } s : Q \circ \hat{s}^{-1} = Q \right\}$$

is closed for the weak topology, due to the continuity of the maps  $s$ . It follows that any limit point of the measures  $P_n$  must be in  $U$  as well.

By Lemma C.1, it follows that limiting points must be concentrated on the minimizers of the large deviation rate function, which have been identified in Proposition 3.8.

In both setting (a) and (b), the intersection of these two sets contains only one element, proving the result.  $\square$

### 3.6 The limiting kernel, proof of Theorem 2.16

To establish the result of Theorem 2.16, it suffices by Lemma 2.11 to find a unique limiting measure for each fuzzy class. In the proof of Theorem 2.7 in [JK17b], it has been shown for fuzzy classes  $l$  of size at least 3 that if the profile  $\nu$  is good then the limiting measure is the spatially homogeneous equi-distribution. Integrating this measure against the function  $\mathcal{A}_{r_l}$  yields one of the factors in the limiting kernel. Thus, it suffices to give an explicit formula for the integral of  $\mathcal{A}_2$  for the Ising classes.

**Lemma 3.9.** *Let  $\beta \geq 0$  and let  $\rho$  be a density profile on  $\mathbb{T}^d$  with  $\rho\lambda(\mathbb{T}^d) > 0$ . Denote  $\tilde{\beta} = \beta\rho\lambda(\mathbb{T}^d)$  and  $\tilde{\rho} = (\rho\lambda(\mathbb{T}^d))^{-1}\rho$ . Suppose that  $\varphi^{*,+}$  is the unique minimizer of  $I_{\tilde{\rho},\tilde{\beta}}$  with non-negative local magnetization  $m_{\tilde{\rho}}$ . Set  $P \in \mathcal{P}(\mathbb{T}^d \times \{1,2\})$  by*

$$P = \frac{1}{2}(\delta_{\varphi^{*,+}} + \delta_{\varphi^{*, -}}).$$

*Note that if the  $I_{\tilde{\rho},\tilde{\beta}}$  has a unique global minimizer  $\varphi^* = \varphi^{*,+} = \varphi^{*, -}$  with 0 local magnetization, then the formula for  $P$  remains valid also.*

*Then we have that*

$$\int \mathcal{A}_2(\pi, u, \tilde{\beta})P(d\pi) = 2 \exp \left\{ \tilde{\beta} \int dv \tilde{\rho}(v)J(u-v) \right\} \frac{1}{\sqrt{1 - m_{\tilde{\rho}}(u)^2}}.$$

*Proof.* Recall, that the magnetization profiles  $\pm m_{\tilde{\rho}}$  correspond to color profiles having densities

$$\frac{d\pi[1]}{d\tilde{\rho}\lambda}(\cdot) = \frac{1 \pm m_{\tilde{\rho}}(\cdot)}{2} \quad \text{and} \quad \frac{d\pi[2]}{d\tilde{\rho}\lambda}(\cdot) = \frac{1 \mp m_{\tilde{\rho}}(\cdot)}{2}.$$

We thus obtain

$$\begin{aligned} & \int \mathcal{A}_2(\pi, u, \tilde{\beta})P(d\pi) \\ &= \int \sum_{i=1}^2 \exp \left\{ 2\tilde{\beta}(J * \pi[i])(u) \right\} \left( \frac{1}{2}\delta_{\varphi^{*,+}} + \frac{1}{2}\delta_{\varphi^{*, -}} \right) (d\pi) \\ &= 2 \left( \frac{1}{2} \exp \left\{ 2\tilde{\beta} \int \frac{1 + m_{\tilde{\rho}}(v)}{2} J(u-v)\tilde{\rho}(v)dv \right\} \right. \\ & \quad \left. + \frac{1}{2} \exp \left\{ 2\tilde{\beta} \int \frac{1 - m_{\tilde{\rho}}(v)}{2} J(u-v)\tilde{\rho}(v)dv \right\} \right) \\ &= \exp \left\{ \tilde{\beta} \int \tilde{\rho}(v)J(u-v)dv \right\} \left( \exp \left\{ \tilde{\beta} \int \tilde{\rho}(v)m_{\tilde{\rho}}(v)J(u-v)dv \right\} \right. \\ & \quad \left. + \exp \left\{ -\tilde{\beta} \int \tilde{\rho}(v)m_{\tilde{\rho}}(v)J(u-v)dv \right\} \right). \end{aligned} \quad (3.11)$$

We first simplify the term involving two exponentials with  $m_{\tilde{\rho}}$  factors. By application of the stationarity equation (3.8)

$$\tilde{\beta} \int \tilde{\rho}(v)m_{\tilde{\rho}}(v)J(u-v)dv = \operatorname{arctanh}(m_{\tilde{\rho}}(u)) = \frac{1}{2} \log \left( \frac{1 + m_{\tilde{\rho}}(u)}{1 - m_{\tilde{\rho}}(u)} \right),$$

the braced expression further becomes

$$\begin{aligned} & \exp \left\{ \tilde{\beta} \int \tilde{\rho}(v) m_{\tilde{\rho}}(v) J(u-v) dv \right\} + \exp \left\{ -\tilde{\beta} \int \tilde{\rho}(v) m_{\tilde{\rho}}(v) J(u-v) dv \right\} \\ &= \sqrt{\frac{1+m_{\tilde{\rho}}(u)}{1-m_{\tilde{\rho}}(u)}} + \sqrt{\frac{1-m_{\tilde{\rho}}(u)}{1+m_{\tilde{\rho}}(u)}} = \frac{2}{\sqrt{1-m_{\tilde{\rho}}(u)^2}}. \end{aligned}$$

Inserting  $\tilde{\beta} = \beta \rho \lambda(\mathbb{T}^d)$  and  $\tilde{\rho} = \frac{\rho}{\rho \lambda(\mathbb{T}^d)}$  into the first factor of (3.11), we get the result

$$\int \mathcal{A}_2(\pi, u, \tilde{\beta}) P(d\pi) = 2 \exp \left\{ \beta \int dv \rho_k(v) J(u-v) \right\} \frac{1}{\sqrt{1-m_{\tilde{\rho}}(u)^2}}.$$

□

## A Absolute continuity of limiting measures

We start by establishing the result of Lemma 2.1.

**Lemma A.1.** *For each  $n$  let  $\Lambda_n \subseteq \mathbb{T}_n^d$ . Set*

$$\nu_n := \frac{1}{|\mathbb{T}_n^d|} \sum_{i \in \Lambda_n} \delta_i.$$

*Then any weak limit point  $\nu$  of the sequence  $\nu_n$  can be considered as an element in the unit ball of  $L^\infty(\mathbb{T}^d, dx)$ .*

*Proof.* Let  $\nu$  be a weak limit point of the sequence  $\nu_n$ . Without loss of generality, we can assume  $\nu_n \rightarrow \nu$  weakly.

First, we prove that  $\nu \in L^1(\mathbb{T}^d, dx)$ . As the set  $L^1(\mathbb{T}^d, dx)$  can be interpreted as a subset of  $\mathcal{M}(\mathbb{T}^d)$ , it suffices to show that for any element  $f \in L^\infty(\mathbb{T}^d, dx)$  we have that  $\langle f, \nu \rangle \leq \|f\|_{L^\infty(\mathbb{T}^d, dx)}$  as this implies  $\|\nu\|_{L^1(\mathbb{T}^d, dx)} \leq 1$ .

Fix  $\varepsilon > 0$ . By Lusin's theorem, cf. [Bog07, Theorem 7.1.13], there is a continuous function  $f_\varepsilon$  with  $\nu(f \neq f_\varepsilon) \leq \varepsilon$  and  $\|f_\varepsilon\| \leq \|f\|$ . Thus

$$\left| \int f d\nu \right| \leq 2 \|f\| \varepsilon + \left| \int f_\varepsilon d\nu \right|.$$

The right hand side is approximated by  $\langle f_\varepsilon, \nu_n \rangle$ , which by assumption is bounded by  $\|f\|$  uniformly in  $n$ . Sending  $\varepsilon \downarrow 0$ , we find

$$\left| \int f d\nu \right| \leq \|f\|$$

establishing that  $\nu \in L^1(\mathbb{T}^d, dx)$ .

Denote by  $g$  the density of  $\nu$  with respect to the Lebesgue measure. We prove it is bounded by using the same trick as above, but by pairing with functions

$f \in L^1(\mathbb{T}^d, dx)$ . First, recall that the space of bounded continuous functions is dense in  $L^1(\mathbb{T}^d, \nu)$ , thus, we can approximate  $f$  in  $L^1$  by continuous bounded functions  $f_k$ . We find

$$\left| \int f g dx \right| = \lim_k \left| \int f_k g dx \right| = \lim_k \lim_n \langle f_k, \nu_n \rangle.$$

For each fixed  $k$ , the function  $f_k$  is uniformly continuous. As the measures  $\nu_n$  are constructed by putting Dirac-masses at a grid with distance  $n^{-1}$  between neighboring points, we can bound the integrals  $|\langle f_k, \nu_n \rangle| \leq \left| \int f_k dx \right| + \delta_k(n)$ , where  $\lim_n \delta_k(n) = 0$ . It follows that  $\left| \int f g dx \right| \leq \|f\|_{L^1(\mathbb{T}^d, dx)}$  implying that  $g \in L^\infty(\mathbb{T}^d, dx)$  and  $\|g\|_{L^\infty(\mathbb{T}^d, dx)} \leq 1$ . □

## B Continuity of limiting functionals

The following lemma is an adaptation from Lemma 2.1 in [HRZ15], taking into account that for each  $n$ , we consider a different space.

**Lemma B.1.** *Let  $\{\mathcal{X}_n\}_{n \geq 1}$  be a sequence of spaces and let  $(\mathcal{X}, d_{\mathcal{X}})$  and  $(\mathcal{Y}, d_{\mathcal{Y}})$  be two metric spaces. Suppose there are maps  $\eta_n : \mathcal{X}_n \rightarrow \mathcal{X}$  in such a way that for every  $x \in \mathcal{X}$  there are  $x_n \in \mathcal{X}_n$  with  $\eta_n(x_n) \rightarrow x$ .*

*Let  $f_n : \mathcal{X}_n \rightarrow \mathcal{Y}$ ,  $f : \mathcal{X} \rightarrow \mathcal{Y}$  and suppose that for all sequences  $x_n \in \mathcal{X}_n$  with  $\eta_n(x_n) \rightarrow x$ , we have  $f_n(x_n) \rightarrow f(x)$ . Then  $f$  is continuous.*

*Proof.* As  $\mathcal{X}$  is metric, it suffices to prove for any sequence  $x_n \in \mathcal{X}$  converging to  $x \in \mathcal{X}$  that  $f(x_n) \rightarrow f(x)$ .

For every  $n$  let  $x_{n,k} \in \mathcal{X}_k$  be such that  $\eta_k(x_{n,k}) \rightarrow x_n$ . By assumption, we find  $f_k(x_{n,k}) \rightarrow f(x_n)$ . Using that  $\mathcal{X}$  and  $\mathcal{Y}$  are metric, we can choose  $k(n)$  large enough such that

$$d_{\mathcal{X}}(\eta_{k(n)}(x_{n,k(n)}), x_n) \leq \frac{1}{n}, \quad d_{\mathcal{Y}}(f_{k(n)}(x_{n,k(n)}), f(x_n)) \leq \frac{1}{n}. \quad (\text{B.1})$$

The first inequality, combined with  $\eta_n(x_n) \rightarrow x$ , implies that  $\lim_n \eta_{k(n)}(x_{n,k(n)}) = x$ . By assumption, we find that  $\lim_n f_{k(n)}(x_{n,k(n)}) = f(x)$ . Therefore, using the second inequality in (B.1)

$$\lim_n d_{\mathcal{Y}}(f(x_n), f(x)) \leq \lim_n d_{\mathcal{Y}}(f(x_n), f_{k(n)}(x_{n,k(n)})) + d_{\mathcal{Y}}(f_{k(n)}(x_{n,k(n)}), f(x)) = 0.$$

We conclude that  $f$  is continuous. □

## C Limits when the LDP has multiple minimizers

The following result is folklore, for which the authors could not find a good reference.

**Lemma C.1.** *Let  $(\mu_n)_{n \geq 1}$  be a sequence of probability-measures on a Polish space  $X$  satisfying an large deviation upper bound at speed  $\{r(n)\}_{n \geq 1}$  with good rate function  $I$ . Further assume, that the rate function only possesses a finite number of minimizers  $m_1, \dots, m_s$ .*

*Then, for any weakly convergent subsequence*

$$\mu_{n_k} \xrightarrow{k \rightarrow \infty} \mu \in \mathcal{M}_1(X, \mathcal{B}(X))$$

*there are  $\lambda_1, \dots, \lambda_s \geq 0$  with  $\sum_{i=1}^s \lambda_i = 1$  such that  $\mu = \sum_{i=1}^s \lambda_i \delta_{m_i}$*

*Proof.* Assume, that  $\mu \notin \{\sum_{i=1}^s \lambda_i \delta_{m_i} : \lambda_1, \dots, \lambda_s \geq 0 \text{ and } \sum_{i=1}^s \lambda_i = 1\}$ . Then we have  $\mu(\{m_1, \dots, m_s\}^c) > 0$ . By inner regularity of the probability measure  $\mu$  (e.g. [Bog07, Theorem 7.1.7]) this still holds true for some compact subset  $K \subseteq \{m_1, \dots, m_s\}^c$ . Moreover, employing the fact, that  $X$  is metrizable, we can find some open set  $O \subseteq X$  where

$$K \subseteq O \subseteq \overline{O} \subseteq \{m_1, \dots, m_s\}^c.$$

The large deviation upper bound, using that  $\overline{O}$  is a closed set not containing the minimizers  $\{m_1, \dots, m_s\}$  of  $I$ , yields

$$\limsup_n \frac{1}{r(n)} \log \mu_n(\overline{O}) \leq - \inf_{x \in \overline{O}} I(x) < 0$$

and, hence,

$$\lim_n \mu_n(\overline{O}) = 0. \tag{C.1}$$

On the other hand, using the lower bound in the Portmanteau theorem, we find

$$\liminf_{k \rightarrow \infty} \mu_{n_k}(\overline{O}) \geq \liminf_{k \rightarrow \infty} \mu_{n_k}(O) \geq \mu(O) \geq \mu(K) > 0,$$

which contradicts (C.1). □

## References

- [Bog07] V. I. Bogachev. *Measure Theory*. Springer-Verlag, 2007.
- [EFS93] A. C. Enter, R. Fernández, and A. D. Sokal. “Regularity properties and pathologies of position-space renormalization-group transformations: Scope and limitations of Gibbsian theory”. In: *Journal of Statistical Physics* 72.5-6 (1993), pp. 879–1167.
- [EK10] V. Ermolaev and C. Külske. “Low-Temperature Dynamics of the Curie-Weiss Model: Periodic Orbits, Multiple Histories, and Loss of Gibbsianness”. In: *Journal of Statistical Physics* 141.5 (2010), pp. 727–756. ISSN: 0022-4715. DOI: 10.1007/s10955-010-0074-x.

- [FHM13] R. Fernández, F. den Hollander, and J. Martínez. “Variational Description of Gibbs-non-Gibbs Dynamical Transitions for the Curie-Weiss Model”. In: *Communications in Mathematical Physics* 319.3 (2013), pp. 703–730. ISSN: 0010-3616. DOI: 10.1007/s00220-012-1646-1.
- [FHM14] R. Fernández, F. den Hollander, and J. Martínez. “Variational Description of Gibbs-Non-Gibbs Dynamical Transitions for Spin-Flip Systems with a Kac-Type Interaction”. In: *Journal of Statistical Physics* 156.2 (2014), pp. 203–220. ISSN: 0022-4715. DOI: 10.1007/s10955-014-1004-0.
- [Geo11] H.-O. Georgii. *Gibbs measures and phase transitions*. Second. Vol. 9. de Gruyter Studies in Mathematics. Walter de Gruyter & Co., Berlin, 2011, pp. xiv+545. ISBN: 978-3-11-025029-9. DOI: 10.1515/9783110250329.
- [HK04] O. Häggström and C. Külske. “Gibbs properties of the fuzzy Potts model on trees and in mean field”. In: *Markov Process. Related Fields* 10.3 (2004), pp. 477–506. ISSN: 1024-2953.
- [HRZ15] F. den Hollander, F. Redig, and W. van Zuijlen. “Gibbs-non-Gibbs dynamical transitions for mean-field interacting Brownian motions”. In: *Stochastic Processes and their Applications* 125.1 (2015), pp. 371–400. ISSN: 0304-4149. DOI: <http://dx.doi.org/10.1016/j.spa.2014.09.011>.
- [Häg03] O. Häggström. “Is the fuzzy Potts model Gibbsian?” In: *Annales de l’Institut Henri Poincaré (B) Probability and Statistics* 39.5 (2003), pp. 891–917. ISSN: 0246-0203. DOI: [https://doi.org/10.1016/S0246-0203\(03\)00026-8](https://doi.org/10.1016/S0246-0203(03)00026-8). URL: <http://www.sciencedirect.com/science/article/pii/S0246020303000268>.
- [JK17a] B. Jahnel and C. Külske. “Gibbsian representation for point processes via hyperedge potentials”. In: *arXiv preprint arXiv:1707.05991* (2017).
- [JK17b] B. Jahnel and C. Külske. “Sharp thresholds for Gibbs-non-Gibbs transitions in the fuzzy Potts model with a Kac-type interaction”. In: *Bernoulli* 23.4A (Nov. 2017), pp. 2808–2827. DOI: 10.3150/16-BEJ828.
- [JK17c] B. Jahnel and C. Külske. “The Widom-Rowlinson model under spin flip: Immediate loss and sharp recovery of quasilocality”. In: *Ann. Appl. Probab.* 27.6 (Dec. 2017), pp. 3845–3892. DOI: 10.1214/17-AAP1298. URL: <https://doi.org/10.1214/17-AAP1298>.
- [Kle14] A. Klenke. *Probability theory*. Second. Universitext. A comprehensive course. Springer, London, 2014, pp. xii+638. ISBN: 978-1-4471-5360-3; 978-1-4471-5361-0. DOI: 10.1007/978-1-4471-5361-0.

- [KLN07] C. Külske and A. Le Ny. “Spin-Flip Dynamics of the Curie-Weiss Model: Loss of Gibbsianness with Possibly Broken Symmetry”. In: *Communications in Mathematical Physics* 271.2 (2007), pp. 431–454. ISSN: 0010-3616. DOI: 10.1007/s00220-007-0201-y.
- [Kül99] C. Külske. “(Non-) Gibbsianness and phase transitions in random lattice spin models”. In: *Markov Process. Related Fields* 5.4 (1999), pp. 357–383. ISSN: 1024-2953.
- [MV95] C. Maes and K. Velde. “The fuzzy Potts model”. In: *Journal of Physics A: Mathematical and General* 28.15 (1995), p. 4261.
- [Rue99] D. Ruelle. *Statistical mechanics: Rigorous results*. World Scientific, 1999.