Estimation of Local Microcanonical Averages in Two Lattice Mean-Field Models Using Coupling Techniques

Kalle Koskinen[1] · **Jani Lukkarinen**[1]

Abstract

We consider an application of probabilistic coupling techniques which provides explicit estimates for comparison of local expectation values between label permutation invariant states, for instance, between certain microcanonical, canonical, and grand canonical ensemble expectations. A particular goal is to obtain good bounds for how such errors will decay with increasing system size. As explicit examples, we focus on two well-studied mean-field models: the discrete model of a paramagnet and the mean-field spherical model of a continuum field, both of which are related to the Curie–Weiss model. The proof is based on a construction of suitable probabilistic couplings between the relevant states, using Wasserstein fluctuation distance to control the difference between the expectations in the thermodynamic limit.

Keywords Curie-Weiss · Mean field models · Wasserstein distance · Microcanonical ensemble · Coupling in probability

1 Introduction

We consider a novel method of analysis of convergence of local expectation values in probability distributions associated with microcanonical ensembles. Our approach aims at answering the following question which would be natural, for example, to control expectations in states arising in ergodic theory: Assume that the system is in a microcanonical state with one or

---

✉ Kalle Koskinen
   kalle.koskinen@helsinki.fi

   Jani Lukkarinen
   jani.lukkarinen@helsinki.fi

[1] Department of Mathematics and Statistics, University of Helsinki, P.O. Box 68, 00014 Helsingin yliopisto, Finland

two known fixed conserved quantities which are label permutation invariant. Consider an observable which depends only on a few degrees of freedom of some finite but large system, for example, consider local correlation functions. Assume that there is some other permutation invariant probability distribution, such as the corresponding grand canonical ensemble, in which the expectation of the observable can be computed, either via a simulation or analytically. Can we estimate the error which arises from replacing the typically not computable microcanonical expectation with the second result? Assuming that the two ensembles are thermodynamically equivalent, how fast does the error decrease with increasing system size?

From the perspective of uniform measures with constraints, we mainly focus on the related standard ensembles, i.e., microcanonical, canonical, and grand canonical ensembles, each with parameters associated with the thermodynamically relevant quantities. For the sake of completeness, we will give a heuristic overview of the standard ensemble theory in Sect. 1.1. There we also introduce notations and terminology which will be used later for defining the ensemble measures of the two models.

In the above standard ensemble set-up, the thermodynamic equivalence of ensembles can often be studied via relative entropy methods. In certain models, in particular, of discrete lattice fields, the relative entropy bounds can also provide an answer to the question stated in the beginning via the Pinsker inequality. However, relative entropy estimates are not always readily available, cannot be used between measures which are not absolutely continuous at least in one direction, and as we will show explicitly later, the estimates they provide might not be optimal.

The motivation to look for improvements of the well developed earlier methods comes from a recent result [13] for the supercritical Berlin–Kac spherical model [1]. This is a model with two thermodynamic quantities in a canonical ensemble where one of them becomes frustrated and forms a condensate. This results in an *nonequivalence* between the canonical and grand canonical ensembles. However, it was shown in [13] that, after separating the condensate modes, the state of the remaining modes is well described by a grand canonical state. Comparing canonical with this modified grand canonical ensemble yields local expectations which converge to each other in the thermodynamic limit. The result was proven using a suitably constructed coupling and relying on the translation invariance of the system, and the resulting estimates imply a power-law convergence in the system size of the errors between the two expectation values.

The coupling technique in [13] is, however, quite specific to the Berlin–Kac model, and relies partially on the existence of the condensate. Here, we explore the extension of these ideas to well-studied cases where equivalence and non-equivalence of various ensembles are known, and which are sufficiently simple to be fairly explicitly computable. For the models chosen here, translation invariance is being replaced by label permutation invariance, and it will serve as an important tool to lift the fairly crude coupling estimates into convergence of various local expectations. We also explore the idea of replacing the standard ensembles of some of these models with other, more accurate but still easy to evaluate, measures. The standard theory of ensemble equivalence will serve as a guide in this choice, and also in most of the cases studied here, it will suffice on its own.

The first of the models is the simple paramagnet which one can find in [11]. In working with this model, there will be a slight abuse of terminology. There is no associated Hamiltonian, but the magnetization is the "conserved quantity" of this model. The corresponding canonical ensemble then has a parameter associated with the control of expectation of magnetization. The paramagnet model is, however, closely related to the standard Curie–Weiss model since the ensemble expectations of the latter can be expressed as a convex combination of those of the former (this connection will be discussed also in Sect. 3.4).

The second model is a continuum modification of the Curie–Weiss model called the mean-field spherical model. The model has been studied in [9] and it is a simplification of the Berlin–Kac model introduced in [1]. In [9], the authors consider the thermodynamic properties of the microcanonical and canonical ensembles. In Sect. 4, we explore the mean-field spherical model in a slightly generalized set-up, namely, by also considering the density of the system to be a free parameter. This allows to study the properties of the grand canonical ensemble which is not explored in [9].

For both of these models, we will give detailed proofs of explicit rates of convergence of finite marginal distributions and/or finite moments of all order between the ensembles of the models. The main result here is the development of novel methods which employ rigorous and well-understood analysis of the thermodynamic properties of the ensembles in order to prove a form of weak convergence of the probability measures corresponding to the different ensembles.

For the simple paramagnet, we will supply two distinct proofs with explicit errors for the convergence of local observables. The first proof will utilize relative entropy and, as such, will mainly reprove and collect known results. The second proof will utilize a coupling argument. Due to the simple nature of this model, we can show explicitly that the error bounds given by the coupling method are strictly better than the bounds given by the relative entropy method that we used. For the mean-field spherical model, we will focus solely on application of the coupling methods to prove local convergence results.

The main mathematical tools for the rigorous control of expectations in the various ensemble measures are couplings of the ensemble measures and the related Wasserstein distance between them, with suitably chosen "cost functions." We give a brief review of couplings and Wasserstein metric in Sect. 2.1.

A crucial property of the ensemble measures and of the couplings constructed here is their invariance under permutations of the particle labels. The permutation invariance improves the control of differences of expectations under the ensemble measures, allowing to bound the error by the above Wasserstein distance. The method is similar to how translation invariance has been used in [13] for the supercritical Berlin–Kac spherical model, and it is described in detail in Sect. 2. Another tool for such an estimation is the Laplace method of asymptotic analysis for such integrals. The method and how it applies to the above error estimation is also discussed in Sect. 2 and A.2.

We postpone more detailed discussion about further related previous works, and how the present estimates connect to these, at the end of Introduction, to Sect. 1.2.

1.1 Equilibrium Ensembles with Two Thermodynamic Quantities

To fix our conventions, let us record here briefly our definitions and parametrization of the standard ensembles. For a review of further results and discussion about the thermodynamic equivalence of ensembles, we refer to [16] and for mathematical details also to [12].

In the following, $\mathcal{S}$ is some arbitrary state space with some fixed positive *reference measure* $d\phi$. The two thermodynamic observables, the "conserved quantities", will be called the *energy* $H : \mathcal{S} \to \mathbb{R}$ and the *particle number* $N : \mathcal{S} \to \mathbb{R}$. We use $V > 0$ to represent the *number of degrees of freedom* of the system and we focus on the properties of the system for large V. It is typically related to the "volume" of the state space $\mathcal{S}$ in some way.

We represent the constraints using, at the moment somewhat formal, δ-function notations; the rigorous meaning of the notations will be discussed later. Let us stress that we do not take the commonly used thin-shell smoothing of these measures since this would for our purposes

unnecessarily complicate the analysis. However, we then have to be careful in the choice of allowed parameter values in some of the ensembles below, to avoid instances where the normalization factor is zero or otherwise ill-defined.

The *microcanonical ensemble* with energy density $\varepsilon \in \mathbb{R}$ and particle density $\rho \in \mathbb{R}$ is then given by

$$\mu_{\mathrm{MC}}^{\varepsilon,\rho;V}(d\phi) := \frac{1}{Z_{\mathrm{MC}}(\varepsilon,\rho;V)} \delta(H[\phi] - \varepsilon V)\delta(N[\phi] - \rho V)\, d\phi.$$

The *canonical ensemble* with inverse temperature $\beta \in \mathbb{R}$ and particle density $\rho \in \mathbb{R}$ is given by

$$\mu_{\mathrm{C}}^{\beta,\rho;V}(d\phi) := \frac{1}{Z_{\mathrm{C}}(\beta,\rho;V)} e^{-\beta H[\phi]}\delta(N[\phi] - \rho V)\, d\phi.$$

Finally, the analogously defined *grand canonical ensemble* with inverse temperature $\beta \in \mathbb{R}$ and chemical potential $\mu \in \mathbb{R}$ is

$$\mu_{\mathrm{GC}}^{\beta,\mu;V}(d\phi) := \frac{1}{Z_{\mathrm{GC}}(\beta,\mu;V)} e^{-\beta H[\phi] - \mu N[\phi]}\, d\phi.$$

Let us remark that, for later convenience, we do not follow the standard physics conventions here using which our parameter "μ" should have been replaced by "$-\beta\mu$".

With the above definitions, there are a number of immediate, explicit relations between some of the above ensembles. In particular, we will need later the following two observations which allow representing an ensemble as a mixture of the more constrained ensemble:

$$\mu_{\mathrm{C}}^{\beta,\rho;V}(d\phi) = \frac{1}{\int d\varepsilon\, e^{-V\beta\varepsilon} Z_{\mathrm{MC}}(\varepsilon,\rho;V)} \int d\varepsilon\, e^{-V\beta\varepsilon} Z_{\mathrm{MC}}(\varepsilon,\rho;V)\mu_{\mathrm{MC}}^{\varepsilon,\rho;V}(d\phi) \quad (1.1)$$

and

$$\mu_{\mathrm{GC}}^{\beta,\mu;V}(d\phi) = \frac{1}{\int d\varepsilon d\rho\, e^{-V(\beta\varepsilon + \mu\rho)} Z_{\mathrm{MC}}(\varepsilon,\rho;V)} \int d\varepsilon d$$
$$\rho\, e^{-V(\beta\varepsilon + \mu\rho)} Z_{\mathrm{MC}}(\varepsilon,\rho;V)\mu_{\mathrm{MC}}^{\varepsilon,\rho;V}(d\phi).$$

Next, we define the *specific microcanonical entropy* or microcanonical entropy per degrees of freedom by

$$s(\varepsilon,\rho;V) := \frac{1}{V} \ln Z_{\mathrm{MC}}(\varepsilon,\rho;V).$$

We define the *specific canonical free energy* or canonical free energy per degrees of freedom by

$$f_{\mathrm{C}}(\beta,\rho;V) := -\frac{1}{V} \ln Z_{\mathrm{C}}(\beta,\rho;V).$$

Note that we do not divide here by β, as would be common for definition of a free energy: this would not be convenient for our models since also zero and negative values of β may occur here. Similarly, the *specific grand canonical free energy* or grand canonical free energy per degrees of freedom is defined here by

$$f_{\mathrm{GC}}(\beta,\mu;V) := -\frac{1}{V} \ln Z_{\mathrm{GC}}(\beta,\mu;V).$$

Now, in order to see the relationship to Laplace-type integrals, we note that

$$e^{-V\beta\varepsilon} Z_{MC}(\varepsilon, \rho; V) = e^{-V(\beta\varepsilon - s(\varepsilon,\rho;V))}, \quad e^{-V(\beta\varepsilon+\mu\rho)} Z_{MC}(\varepsilon, \rho; V)$$
$$= e^{-V(\beta\varepsilon+\mu\rho - s(\varepsilon,\rho;V))}.$$

Assuming that the limits exist, we define

$$s(\varepsilon, \rho) := \lim_{V\to\infty} s(\varepsilon, \rho; V), \quad f_C(\beta, \rho) = \lim_{V\to\infty} f_C(\beta, \rho; V), \quad f_{GC}(\beta, \mu)$$
$$:= \lim_{V\to\infty} f_{GC}(\beta, \mu; V).$$

Then either Laplace-type integral estimates or large deviation techniques [12] can often be used to show that the limit functions are related by a Legendre transform:

$$f_C(\beta, \rho) = \inf_{\varepsilon}\{\beta\varepsilon - s(\varepsilon, \rho)\}, \quad f_{GC}(\beta, \mu) = \inf_{\varepsilon,\rho}\{\beta\varepsilon + \mu\rho - s(\varepsilon, \rho)\}.$$

Typically, this results in a one-to-one correspondence between the parameters ε and ρ in the microcanonical ensemble with the associated free parameters β and μ. Assuming that the above thermodynamic limits exist and agree with each other using this correspondence, we say that the ensembles are *thermodynamically equivalent*.

The theory of Laplace-type integrals is well-developed and allows one to compute explicit asymptotics of such integrals. In particular, one is typically interested in second-order fluctuations. Indeed, from the specific free energies, we obtain

$$\frac{\langle H\rangle_C^{\beta,\rho;V}}{V} = \partial_\beta f_C(\beta, \rho; V), \quad \frac{\langle H^2\rangle_C^{\beta,\rho;V} - \left(\langle H\rangle_C^{\beta,\rho;V}\right)^2}{V} = -\partial_\beta^2 f_C(\beta, \rho; V).$$

Using the theory of Laplace-type integrals, we typically have

$$\lim_{V\to\infty} \frac{\langle H\rangle_C^{\beta,\rho;V}}{V} = \partial_\beta f_C(\beta, \rho), \quad \lim_{V\to\infty} \frac{\langle H^2\rangle_C^{\beta,\rho;V} - \left(\langle H\rangle_C^{\beta,\rho;V}\right)^2}{V} = -\partial_\beta^2 f_C(\beta, \rho).$$

The notion of "typical" here is rather vague and we refer the reader to A.2 for a more detailed account of the use of the Laplace method. The first limit implies that the energy density of the canonical system converges to a constant, which, in turn, implies that the energy density behaves like $O(1)$ for large V. The contents of the second limit imply that the standard deviation of the energy density of the canonical system behaves like $O(V^{-\frac{1}{2}})$. However, one should not rely on these formulae directly at phase transition points where the differentiability assumptions fail: in such cases, more refined tools, such as subdifferentials and convex analysis, will be needed to study the related behaviour.

However, in addition to analysing the thermodynamic properties of the system, the Laplace-type analysis offers us something more. Indeed, if we return to the alternative representation of the canonical ensemble and we denote the minimizing ε of $f_C(\beta, \rho)$ by ε^*, then, for some suitable class of observables $g(\phi)$, one might expect that

$$\lim_{V\to\infty} \left|\langle g\rangle_{MC}^{\varepsilon^*,\rho;V} - \langle g\rangle_C^{\beta,\rho;V}\right| = 0. \tag{1.2}$$

We then say that the two ensembles are equivalent in this observable class. For instance, if the above result would hold for every function $g : S \to \mathbb{C}$ which is Lipschitz continuous, we could say that the microcanonical and canonical ensembles are *Lipschitz observable equivalent*. Analogously, if the result holds for all polynomials g of the field whose degree is

not allowed to grow with V, we say that the ensembles are *equivalent in their local moments*. In this paper, we consider the suitable class of functions g, and the rate of convergence in (1.2) in more detail.

1.1.1 Clarification of Terminology

To avoid possible misunderstandings, let us explicitly record our usage of the terminology concerning ensembles and related objects such as the partition function and free energy. Most notably, we will need to make a distinction between thermodynamic and auxiliary statistical ensembles.

A statistical ensemble is a probability distribution describing the state of a system. A *thermodynamic ensemble* is a particular statistical ensemble which is determined by the physical properties of the system, in particular, by its dynamics. The most common examples start with a Hamiltonian defining the dynamics and then include any other relevant conserved quantities using one of the above discussed forms leading to microcanonical, canonical, and possibly one or more grand canonical ensembles. Partition functions and free energies can then be associated with these thermodynamic ensembles. We do make some choices of convenience to simplify the overall constant in the partition function: to avoid misunderstandings, we include also their explicit definitions in the following.

Here, we start from some some given thermodynamic ensemble in the microcanonical form. This yields the physical probability distribution whose local expectation values we aim to estimate. For this estimation, it turns out to be helpful to introduce new probability measures, i.e., statistical ensembles, on the system which we will call *auxiliary ensembles*. Since many of these auxiliary measures can be written in the same form as standard thermodynamic ensembles, it will be helpful to extend the standard terminology also there, leading, for example, to "auxiliary microcanonical ensemble with fixed magnetization density" for the Curie–Weiss model.

The auxiliary ensembles can be associated with "partition functions" and "free energies" in analogy with the standard ensembles, and this indeed will become a helpful shorthand notation in some of our computations. However, it should be stressed that the auxiliary ensembles usually do not have any thermodynamic meaning, for example, the magnetization defining the auxiliary ensemble above is not implied to be a conserved quantity in any dynamics leading to the Curie–Weiss model. In addition, when talking about phase transitions and their order, we will always refer to the parameters in the original thermodynamic ensembles, and not to those appearing in the auxiliary ensembles.

1.2 Related Works and Further Motivation

There has always been considerable interest in trying to classify the "correct" notions of convergence of the equilibrium ensembles. For a particularly illuminating and modern account on some of the various notions which have been considered, we refer to [16] and its references. Thermodynamic equivalence from the point of view of large deviations and convexity properties of entropy is considered in great generality in [16]. Here, we approach the problem more from the point of view of convergence of generic local expectation values, and the additional facilitating ingredient is label permutation invariance of the studied equilibrium ensembles. For rigorous applications of the ensembles in non-equilibrium phenomena, such as for estimating the accuracy of local thermal equilibrium while studying heat transport, it

would be important to be able to estimate the error in the approximation. This ultimate goal is the second motivation for starting with the simple example cases in the present contribution.

In fact, such rigorous proofs are already available in the literature, albeit for different systems from the ones studied here. A very detailed mathematical account of such a convergence has been given in [2] starting from uniform distributions on the intersection of a simplex and a sphere. By appropriately parametrizing the radius of the sphere, and considering the behaviour of finite dimensional marginals and moments of this uniform distribution as the dimension of the space is increased, the author is able to rigorously prove that a phase transition occurs for this specific system. In particular, the author is able to prove that in the high dimensional limit the finite marginal distributions of the given uniform distributions are of product form.

Another work in this direction, which cites the previous article, is given in [8]. In this work the authors consider the convergence of the microcanonical and grand canonical measures related to the Bose–Hubbard model. The commonality between both [2] and [8] is that the models they are considering are defined on state spaces with strictly positive unbounded elements. Such a feature seems to be a key property of these models since both of these works observe a phase transition into a state which can be characterized as containing a condensate.

In fact, a fairly satisfying account of ensembles with unbounded strictly positive phase spaces has been given in [14]. In this work the author proves a form of the equivalence of ensembles for systems with multiple constraints satisfying certain conditions, and the results are quite general as to their applicability. However, the main theorems presented there hold for phase spaces which are defined on $[0, \infty)^N$ rather than $\mathbb{R}^N$, and, furthermore, the assumptions of the main theorem do not hold for the ensembles we are considering here.

We also mention an extensive source for references to the relative entropy method and usage of the method in [7] and [3]. Some of these references will also be explicitly quoted later when discussing the usage of relative entropy.

Finally, let us mention the origin of the continuum model we are considering. First, we recall the (discrete) Curie–Weiss model. For a general overview of the discrete model, we refer to [10]. We also mention the classical work of Ellis in [6] which goes beyond the standard Curie–Weiss model. In [9], the authors consider a further simplification to the Berlin–Kac model introduced in [1]. In particular, the nearest neighbour Ising model is replaced by a mean-field Hamiltonian, and, as evidenced in the article, the thermodynamic properties of the microcanonical and canonical ensembles become exactly computable. However, the authors do not consider the properties of local observables in their analysis. The following references contain results about the phase structure of these models [9], as well as of their Potts model type generalizations to multicomponent cases [4].

Our approach differs significantly from those of the above previous works and their associated models. In particular, we will employ various coupling methods to prove convergence of finite dimensional marginal distributions and finite moments of all orders. In addition, our arguments do not hinge on definitions of the microcanonical ensembles with thin-set approximations. Instead, we define the microcanonical ensembles directly as constrained measures and explore their properties via analytic rather than probabilistic methods. For the first model, we refer to [11] for a considerably more detailed analysis of the various properties of the model. However, for the second model introduced in [9] there does not seem to be proofs pertaining to the convergence of finite dimensional marginals or finite moments. There is a considerable amount of fine structure which must be considered to give a full account of the local convergence at this level.

Finally, let us stress that the main purpose of this paper is to display the specific methods of coupling and their relationship with the local convergence properties of the equilibrium ensembles. The thermodynamic properties of these systems are already well-known and have been studied extensively, but we wish to give an alternate, simpler and more accurate, account of the two models present in this paper, with the hope that the ideas used here generalize to other, less explicitly tractable models.

2 Two Methods of Coupling and Main Lemmas

In this section, we will present definitions relevant to this article including the concept of coupling, the Wasserstein distance metric, and their two application methods which will be presented as theorems later on.

2.1 Couplings and Wasserstein Distances

We collect some of the basic notions related to couplings here. More thorough introduction is available for instance in [17].

2.1.1 Couplings and Transport Maps

We will frequently make use of the notion of coupling between probability measures. Let X be a sample space and let Σ be a σ-algebra on X. Let μ_1 and μ_2 be two probability measures on X. Define the coordinate projections $P_1 : X \times X \to X$ and $P_2 : X \times X \to X$ by $P_1(x, y) := x$ and $P_2(x, y) := y$. A probability measure γ on a sample space $X \times X$ with a σ-algebra $\Sigma \otimes \Sigma$ is called a coupling if $\gamma \circ P_1^{-1} = \mu_1$ and $\gamma \circ P_2^{-1} = \mu_2$. Here, and in the following, P^{-1} will be used not only to denote the inverse of a mapping P, but also for the associated map which takes a set to its preimage under P.

In this paper, we will often give the definitions of probability measures with the explicit assumption that they can be constructed by simply giving suitable values of the expectations of measurable functions. For example, if X is a locally compact Hausdorff space and we are able to construct a bounded positive linear functional L on $C_c(X)$, the space of continuous functions with compact support equipped with the supremum norm, such that $\|L\| = 1$, then by the Riesz–Markov–Kakutani representation theorem, there exists a unique Radon probability measure[1] μ on X such that $L(f) = \langle f \rangle_\mu$ for all $f \in C_c(X)$.

For the contents of this paper, we will use the following equivalent notion of coupling. Let $f : X \to \mathbb{R}$ be a measurable function. A probability measure γ, as defined in the previous paragraph, is a coupling if

$$\langle f \circ P_1 \rangle_\gamma = \langle f \rangle_{\mu_1}, \ \langle f \circ P_2 \rangle_\gamma = \langle f \rangle_{\mu_2}$$

holds for all such functions f. One typically says that the marginal distributions of γ are given by μ_1 and μ_2.

In this paper, we will sometimes refer to specific types of couplings as transport maps. Let μ_1 be a probability measure as before, and let $T : X \to X$ be a measurable map. Define the probability measure μ_2 by setting $\mu_2(A) := \mu_1(T^{-1}(A))$ for all $A \in \Sigma$. Such

---

[1] Radon measures are a subclass of Borel measures with additional technical regularity properties, cf. Wikipedia or [15].

a probability measure μ_2 is called the *pushfoward measure* of μ_1 by the map T. We then denote $\mu_2 = T_*\mu_1$. This notion is also sometimes called the abstract change of variables due to the following equivalent definition of the pushforward measure: If $f : X \to \mathbb{R}_+$ is a characteristic function of a measurable set, we may set

$$\langle f \rangle_{\mu_3} = \int_X \mu_1(dx) \, f(T(x)), \tag{2.1}$$

and this defines a positive measure μ_3 on Σ. Then, it is straightforward to check that μ_3 indeed is a probability measure for which (2.1) holds for every non-negative measurable function f. In addition, $\mu_3 = \mu_2$, and thus (2.1) provides an alternative definition of $T_*\mu_1$.

When μ_2 and μ_1 are measures such that there is a measurable map T for which $\mu_2 = T_*\mu_1$, we call T a *transport map* from the measure μ_1 to μ_2. A transport map T can always be used to construct a coupling between μ_1 and μ_2 as follows: If $g : X \times X \to \mathbb{R}_+$ is a measurable function, we define a probability measure γ by setting

$$\langle g \rangle_\gamma = \int_X \mu_1(dx) \, g(x, T(x)).$$

One can go through analogous steps as above and show that γ is then indeed a coupling of μ_1 and $\mu_2 = T_*\mu_1$.

2.1.2 Wasserstein Distance and Coupling Optimization

For the moment, we will specialize to probability measures on $\mathbb{R}^n$. Let μ_1 and μ_2 be probability measures on $\mathbb{R}^n$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded 1-Lipschitz function with respect to the $\|\cdot\|_p$-norm for some $p \geq 1$. To be explicit, we require that f is a function for which its optimal Lipschitz constant K, defined by

$$K := \sup_{\phi \neq \psi} \frac{|f(\phi) - f(\psi)|}{\|\phi - \psi\|_p},$$

satisfies $K \leq 1$. This is a property which depends on the choice of norm, and restricts the class of allowed functions. Naturally, if f is a function with $K > 1$, then we can apply the results below to the 1-Lipschitz function $\frac{1}{K} f$, and the conclusions for the original function f will be the same, as long as the constant K remains bounded in n. We have chosen to use the "1-Lipschitz" assumption in order to remove one, otherwise quite relevant, constant from the estimates.

Suppose there exists a coupling γ of μ_1 and μ_2. Using the properties of probability measures, we have

$$\left| \langle f \rangle_{\mu_1} - \langle f \rangle_{\mu_2} \right| = \left| \langle f \circ P_1 - f \circ P_2 \rangle_\gamma \right| \leq \langle |f \circ P_1 - f \circ P_2| \rangle_\gamma \leq \langle \|x_1 - x_2\|_p \rangle_\gamma. \tag{2.2}$$

On the last line, we have used the short hand notation $x_i = P_i(x)$, $i = 1, 2$, for clarity. One should note that the coupling does not appear on the left hand side of this inequality, and, we are thus free to minimize this inequality with respect to all couplings γ. Since there always exists at least one coupling, given by the the product coupling $\gamma = \mu_1 \otimes \mu_2$, and since the functions f are bounded, then for any coupling the middle expression has a uniform upper bound. Therefore,

$$\left| \langle f \rangle_{\mu_1} - \langle f \rangle_{\mu_2} \right| \leq \inf_\gamma \langle \|x_1 - x_2\|_p \rangle_\gamma.$$

Naturally, we can swap the norm $|| \cdot ||_p$ for any cost function $c(x, y) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ with enough regularity as long as we can relate the difference of the expectations somehow to the given cost function.

For $p \geq 1$, define $\mathcal{P}_p(\mathbb{R}^n)$ to be the space of probability measures with finite p:th moments, i.e., assuming that $\langle \|x\|_p^p \rangle < \infty$. Consider $\mu_1, \mu_2 \in \mathcal{P}_p(\mathbb{R}^n)$. Given also some $q \geq 1$, we denote the p-Wasserstein distance between μ_1 and μ_2 with respect to the q-norm by $W_{p;q}(\mu_1, \mu_2)$. Explicitly,

$$W_{p;q}(\mu_1, \mu_2) = \left(\inf_{\gamma} \int_{\mathbb{R}^n \times \mathbb{R}^n} \gamma(dx, dy) \, ||x - y||_q^p \right)^{\frac{1}{p}},$$

and, since $\|x\|_q \leq n^{1/q} \max_j |x_j|$, is straightforward to check that then $W_{p;q}(\mu_1, \mu_2) < \infty$.

The p-Wasserstein distance has been studied comprehensively and applied in a great variety of circumstances; examples and discussion are provided in [17]. However, for the purposes of this paper, we will be more interested in slightly modified cost functions which are similar in nature to the p-Wasserstein distances. The main drawback of many of the methods and papers associated with the Wasserstein distances is that the focus has been on the case where the dimension of the space n is fixed. In the context of statistical mechanics, we are typically interested in asymptotic properties for arbitrarily large n.

2.2 Definitions

For the purposes of this section and for the definition of the lattice model later, let us fix some shorthand notations first. Given $N \in \mathbb{N}$, we denote the collection of first N integers as follows

$$[N] := \{1, 2, \ldots, N\}, \tag{2.3}$$

and we denote the group of permutations of its elements by S_N. Given a subset $I \subset [N]$, of a length $n := |I|$, there is a unique bijection $\pi_I : I \to [n]$ which retains the order of the elements in the subsequence. We let $\bar{\pi}_I \in S_N$ denote the extension of π_I which is obtained by permuting the elements in $[N] \setminus I$ in an order preserving manner into the set $[N] \setminus [n]$. In addition, every bijection π_I as above defines a projection $P_I : \mathbb{R}^N \to \mathbb{R}^n$ via the formula $(P_I x)_j := x_{\pi_I^{-1}(j)}$, $j \in [n]$. Analogously, given a permutation $\pi \in S_N$, the corresponding coordinate permutation will be denoted $Q_\pi : \mathbb{R}^N \to \mathbb{R}^N$; explicitly, we set $(Q_\pi x)_j := x_{\pi^{-1}(j)}$, $j \in [N]$ (note that using the inverse permutation in the formula will result in a map which will send coordinate i into coordinate $\pi(i)$).

Given $y \in \mathbb{R}$, there is a unique integer $k \in \mathbb{Z}$ for which $k \leq y < k + 1$, and we denote this by using the "floor" notation, $k := \lfloor y \rfloor$. In particular, given $n, N \in \mathbb{N}$ such that $n \leq N$ and setting $k = \lfloor N/n \rfloor$ we have $k \in \mathbb{N}$ and k satisfies $kn \leq N < (k + 1)n$.

Definition 2.1 (*Permutation invariance of measures on $\mathbb{R}^N$*) Given $N \in \mathbb{N}$, a probability measure μ on $\mathbb{R}^N$, we say that μ is permutation invariant, if for every integrable function $f : \mathbb{R}^N \to \mathbb{R}$ and a permutation $\pi \in S_N$, we have $f \circ Q_\pi \in L^1(\mu)$ and

$$\langle f \circ Q_\pi \rangle_\mu = \langle f \rangle_\mu.$$

Finally, instead of using a standard p-norm to measure distances in $\mathbb{R}^N$, we scale it suitably with N so that the Wasserstein cost function becomes an average over particle labels. The benefits of this definition will become apparent in Sect. 2.3.

Definition 2.2 (*Specific p-norm fluctuation distance*) Suppose $p \geq 1$ and $N \in \mathbb{N}$. Let μ_1 and μ_2 be two Radon probability measures on $\mathbb{R}^N$ such that the p:th moments under both measures are finite. Their specific p-norm fluctuation distance w_p is then defined as

$$w_p(\mu_1, \mu_2; N) := \left(\inf_\gamma \int_{\mathbb{R}^N \times \mathbb{R}^N} \gamma(dx, dy) \frac{1}{N} \sum_{i=1}^N |x_i - y_i|^p \right)^{\frac{1}{p}},$$

where the infimum is taken over all couplings of μ_1 and μ_2.

Clearly, this definition relates to the standard p-norm Wasserstein distance mentioned earlier via a scaling: $w_p = N^{-\frac{1}{p}} W_{p;p}$.

2.3 The Direct Coupling Method

To highlight the benefits of the above definitions, we offer the following fundamental Lemma which will be used to prove the main theorems of this paper. It should be stressed that the key assumption is to specialize to permutation invariant measures. We aim to consider local expectations, i.e., $\langle F \rangle$ for functions $F : \mathbb{R}^N \to \mathbb{R}$ which depend only on components x_i, $i \in I$, where $I \subset [N]$ can be otherwise arbitrary but it has a bounded size, i.e., $|I|$ remains bounded when $N \to \infty$. In particular, note that then there is some $f : \mathbb{R}^{|I|} \to \mathbb{R}$ such that $F = f \circ P_I$.

Lemma 2.3 *Suppose $p \geq 1$ and $N \in \mathbb{N}$. Let μ_1 and μ_2 be two permutation invariant Radon probability measures on $\mathbb{R}^N$ such that the p:th moments under both measures are finite. Consider a subset $I \subset [N]$. If $f : \mathbb{R}^{|I|} \to \mathbb{R}$ is a bounded 1-Lipschitz function with respect to the $|| \cdot ||_p$ norm, then we have*

$$\left| \langle f \circ P_I \rangle_{\mu_1} - \langle f \circ P_I \rangle_{\mu_2} \right| \leq \left(\frac{|I|}{1 - \frac{|I|}{N}} \right)^{\frac{1}{p}} w_p(\mu_1, \mu_2; N).$$

Proof For the proof, set $n := |I|$ and $k := \lfloor N/n \rfloor$ when $k \in \mathbb{N}$ and k satisfies $kn \leq N < (k+1)n$. We define the sets $I_i \subset [N]$, $i \in [k]$, by setting $I_1 := I$ and, for $i > 1$, we proceed inductively by selecting $|I|$ elements from the set $[N] \setminus \left(\bigcup_{j=1}^{i-1} I_j \right)$ to be the set I_i. The collection of sets I_i are disjoint and $\bigcup_{i=1}^k I_i \subset [N]$. For any i, there is a permutation in S_N which is bijection between I_i and I. Thus by the assumed permutation invariance of the measures, we have $\langle f \circ P_{I_i} \rangle = \langle f \circ P_I \rangle$ for either measure and all i. Therefore,

$$\langle f \circ P_I \rangle_{\mu_1} - \langle f \circ P_I \rangle_{\mu_2} = \frac{1}{k} \sum_{i=1}^k \left(\langle f \circ P_{I_i} \rangle_{\mu_1} - \langle f \circ P_{I_i} \rangle_{\mu_2} \right).$$

Suppose then that γ is a coupling between μ_1 and μ_2. Then $\langle f \circ P_{I_i} \rangle_{\mu_j} = \langle f \circ P_{I_i} \circ P_j \rangle_\gamma$ for both $j = 1, 2$. Again resorting to the shorthand notations $x_j := P_j x$, we can rewrite

$$\langle f \circ P_{I_i} \rangle_{\mu_1} - \langle f \circ P_{I_i} \rangle_{\mu_2} = \langle f(P_{I_i} x_1) - f(P_{I_i} x_2) \rangle_\gamma.$$

The absolute value of this expression can now be estimated using the assumed 1-Lipschitz property of f. Combining the results and using the triangle inequality we thus obtain

$$\left|\langle f \circ P_I \rangle_{\mu_1} - \langle f \circ P_I \rangle_{\mu_2}\right| \leq \frac{1}{k} \sum_{i=1}^{k} \langle \|P_{I_i} x_1 - P_{I_i} x_2\|_p \rangle_\gamma$$

$$\leq \left(\frac{1}{k} \sum_{i=1}^{k} \langle \|P_{I_i} x_1 - P_{I_i} x_2\|_p^p \rangle_\gamma \right)^{\frac{1}{p}},$$

where in the last step we have used Hölder's inequality. Since the sets I_i are disjoint, here $\sum_{i=1}^{k} \|P_{I_i} x_1 - P_{I_i} x_2\|_p^p \leq \sum_{j=1}^{N} |(x_1)_j - (x_2)_j|^p$. Therefore,

$$\left|\langle f \circ P_I \rangle_{\mu_1} - \langle f \circ P_I \rangle_{\mu_2}\right| \leq \left(\frac{1}{k} \langle \|x_1 - x_2\|_p^p \rangle_\gamma \right)^{\frac{1}{p}}.$$

Because the left hand side of the above estimate does not depend on the coupling γ, we can take the infimum over all possible couplings. Then using the relation between k and n stated in the beginning of the proof, we obtain

$$\left|\langle f \circ P_I \rangle_{\mu_1} - \langle f \circ P_I \rangle_{\mu_2}\right| \leq \left(\frac{n}{1 - \frac{n}{N}} \right)^{\frac{1}{p}} w_p(\mu_1, \mu_2; N),$$

as desired. □

The first Lemma concerned bounds on local observables which were bounded 1-Lipschitz functions. This next variant of the Lemma concerns estimation of arbitrary finite moments.

Theorem 2.4 *Suppose $p > 1$ and $N \in \mathbb{N}$. Let μ_1 and μ_2 be two permutation invariant Radon probability measures on $\mathbb{R}^N$ such that the p_0:th moments of both measures are finite for some $p_0 \geq p$.*

Let J be a finite sequence of elements in $[N]$ where elements may be repeated. Let $n_J := |J|$ denote the length of the sequence and $I \subset [N]$ the collection of elements occurring in the sequence, i.e., set $I := \{J_\ell \mid \ell \in [n_J]\}$. For any $x \in \mathbb{R}^N$, we then let x^J denote the power

$$x^J := \prod_{\ell=1}^{n_J} x_{J_\ell}.$$

Assuming also $n_J \leq p_0 + 1 - \frac{p_0}{p}$, it follows that

$$\left|\langle x^J \rangle_{\mu_1} - \langle x^J \rangle_{\mu_2}\right| \leq n_J M(J, p)^{n_J - 1} \left(\frac{|I|}{1 - \frac{|I|}{N}} \right)^{\frac{1}{p}} w_p(\mu_1, \mu_2; N),$$

where $M(J, p) = 1$ if $n_J = 1$, and otherwise

$$M(J, p) := \max_{i \in I} \left(\langle |x_i|^{q(n_J-1)} \rangle_{\mu_1}^{\frac{1}{q(n_J-1)}}, \langle |x_i|^{q(n_J-1)} \rangle_{\mu_2}^{\frac{1}{q(n_J-1)}} \right) < \infty,$$

with $q = \frac{p}{p-1}$.

Proof Generalized Hölder's inequality implies that

$$\langle |x^J| \rangle \leq \prod_{\ell=1}^{n_J} \langle |x_{J_\ell}|^{n_J} \rangle^{\frac{1}{n_J}}.$$

Since $1 \le n_J \le p_0$, the assumptions guarantee that x^J is integrable with respect to both μ_1 and μ_2. On the other hand, $n_J \le p_0 + 1 - \frac{p_0}{p}$ implies $q(n_J - 1) \le p_0$, so that also $M(J, p) < \infty$.

First, note that for $x, y \in \mathbb{R}^N$, we have

$$x^J - y^J = \sum_{i=1}^{n_J}(x_{J_i} - y_{J_i}) \prod_{j<i} x_{J_j} \prod_{k>i} y_{J_k}.$$

There are n_J factors in each of the products under the sum. Thus for any coupling γ between μ_1 and μ_2 and, for simplicity, replacing x_1, x_2 by x, y, we find using the generalized Hölder's inequality

$$\left|\left\langle x^J - y^J \right\rangle_\gamma\right| \le \sum_{i=1}^{n_J}\left\langle |x_{J_i} - y_{J_i}| \prod_{j<i} |x_{J_j}| \prod_{k>i} |y_{J_k}| \right\rangle_\gamma$$

$$\le \sum_{i=1}^{n_J} \left\langle |x_{J_i} - y_{J_i}|^p \right\rangle_\gamma^{\frac{1}{p}} \prod_{j<i} \left\langle |x_{J_j}|^{q'} \right\rangle_\gamma^{\frac{1}{q'}} \prod_{k>i} \left\langle |y_{J_k}|^{q'} \right\rangle_\gamma^{\frac{1}{q'}},$$

where $q' := q(n_J - 1)$, so that indeed $\frac{1}{p} + (n_J - 1)\frac{1}{q'} = 1$, as required by the Hölder's inequality. Apart from the first term, the remaining $n_J - 1$ terms are all bounded by $M(J, p)$. Therefore,

$$\left|\left\langle x^J - y^J \right\rangle_\gamma\right| \le M(J, p)^{n_J - 1} \sum_{i=1}^{n_J}\left\langle |x_{J_i} - y_{J_i}|^p \right\rangle_\gamma^{\frac{1}{p}}$$

$$\le M(J, p)^{n_J - 1} n_J \left(\frac{1}{n_J} \sum_{i=1}^{n_J}\left\langle |x_{J_i} - y_{J_i}|^p \right\rangle_\gamma \right)^{\frac{1}{p}},$$

where Hölder's inequality has been used in the second step. Here, even if there are repetitions in the sequence J, we have $\frac{1}{n_J} \sum_{i=1}^{n_J} |x_{J_i} - y_{J_i}|^p \le \sum_{i \in I} |x_i - y_i|^p =: \|x - y\|_{I,p}^p$. Therefore,

$$\left|\left\langle x^J - y^J \right\rangle_\gamma\right| \le n_J M(J, p)^{n_J - 1} \left\langle \|x - y\|_{I,p}^p \right\rangle_\gamma^{\frac{1}{p}}.$$

To finish the proof, one should notice that the label subset I which appears in this theorem can be regarded in the same way as in the proof of Lemma 2.3. Using the assumed permutation invariance to clone the labels yields collections of subsequences $J(\ell)$ and subsets I_ℓ for $\ell \in [k]$, where $k := \lfloor N/n \rfloor$, $n := |I|$. Since $\langle x^J \rangle_{\mu_i} = \langle x^{J(\ell)} \rangle_{\mu_i}$ by construction, we find using permutation invariance that

$$\left|\left\langle x^J \right\rangle_{\mu_1} - \left\langle x^J \right\rangle_{\mu_2}\right| \le \frac{1}{k} \sum_{\ell=1}^{k} \left|\left\langle x^{J(\ell)} - y^{J(\ell)} \right\rangle_\gamma\right| \le \frac{n_J M(J, p)^{n_J - 1}}{k} \sum_{\ell=1}^{k} \left\langle \|x - y\|_{I_\ell,p}^p \right\rangle_\gamma^{\frac{1}{p}}$$

$$\le n_J M(J, p)^{n_J - 1} \left(\frac{|I|}{1 - \frac{|I|}{N}} \right)^{\frac{1}{p}} w_p(\mu_1, \mu_2; N),$$

as desired. $\square$

2.4 Free Energy Method

By applying Lemma 2.3, we are now able to produce two distinct types of coupling proofs which concern the ensembles discussed in the introduction.

Theorem 2.5 *Let $\mu_{MC}^{\varepsilon,\rho;N}$ be a permutation invariant probability measure corresponding to a microcanonical ensemble with energy density ε and particle density ρ. In addition, assume that if we fix a possible energy density ε', then for any other possible energy density ε there exists a constant $C(\varepsilon, \rho) > 0$ independent of ε' and N, but possibly dependent on ε and ρ, such that*

$$w_p(\mu_{MC}^{\varepsilon,\rho;N}, \mu_{MC}^{\varepsilon',\rho;N}; N) \leq C(\varepsilon, \rho)|\varepsilon - \varepsilon'|$$

for some $p \geq 1$. Suppose also that the microcanonical and canonical measures, for some parameter β, have finite p:th moments.

Fix $n < \infty$ and consider any $I \subset \mathbb{N}$ of length n. Let $f : \mathbb{R}^{|I|} \to \mathbb{R}$ be a bounded 1-Lipschitz function with respect to the $||\cdot||_p$ norm. Then

$$\left| \langle f \circ P_I \rangle_{MC}^{\varepsilon,\rho;N} - \langle f \circ P_I \rangle_C^{\beta,\rho;N} \right|$$

$$\leq C(\varepsilon, \rho) \left(\frac{|I|}{1 - \frac{|I|}{N}} \right)^{\frac{1}{p}} \left(\sigma_C^{\beta,\rho;N} \left(\frac{H}{N} \right) + \left| \varepsilon - \frac{\langle H \rangle_C^{\beta,\rho;N}}{N} \right| \right),$$

where the canonical standard deviation of energy density reads explicitly

$$\sigma_C^{\beta,\rho;N} \left(\frac{H}{N} \right) = \sqrt{\frac{\langle H^2 \rangle_C^{\beta,\rho;N} - \left(\langle H \rangle_C^{\beta,\rho;N} \right)^2}{N^2}}.$$

Using the notation of the specific free energies, the same result can be rewritten as

$$\left| \langle f \circ P_I \rangle_{MC}^{\varepsilon,\rho;N} - \langle f \circ P_I \rangle_C^{\beta,\rho;N} \right|$$

$$\leq C(\varepsilon, \rho) \left(\frac{|I|}{1 - \frac{|I|}{N}} \right)^{\frac{1}{p}} \left(\frac{1}{\sqrt{N}} \sqrt{-\partial_\beta^2 f_C(\beta, \rho; N)} + |\varepsilon - \partial_\beta f_C(\beta, \rho; N)| \right)$$

Proof By the relation (1.1), we have

$$\langle f \circ P_I \rangle_{MC}^{\varepsilon,\rho;N} - \langle f \circ P_I \rangle_C^{\beta,\rho;N} = \frac{1}{\int d\varepsilon' \, e^{-N\beta\varepsilon'} Z_{MC}(\varepsilon', \rho; N)} \int d\varepsilon' \, e^{-N\beta\varepsilon'} Z_{MC}$$

$$(\varepsilon', \rho; N) \left(\langle f \circ P_I \rangle_{MC}^{\varepsilon,\rho;N} - \langle f \circ P_I \rangle_{MC}^{\varepsilon',\rho;N} \right).$$

Applying Lemma 2.3 together with the assumptions of this theorem, we thus obtain

$$\left| \langle f \circ P_I \rangle_{MC}^{\varepsilon,\rho;N} - \langle f \circ P_I \rangle_C^{\beta,\rho;N} \right|$$

$$\leq \frac{1}{\int d\varepsilon' \, e^{-N\beta\varepsilon'} Z_{MC}(\varepsilon', \rho; N)} \int d\varepsilon' \, e^{-N\beta\varepsilon'} Z_{MC}(\varepsilon', \rho; N) \left(\frac{|I|}{1 - \frac{|I|}{N}} \right)^{\frac{1}{p}} C(\varepsilon, \rho)|\varepsilon - \varepsilon'|.$$

Since

$$|\varepsilon - \varepsilon'| \leq \left| \varepsilon - \frac{\langle H \rangle_C^{\beta,\rho;N}}{N} \right| + \left| \frac{\langle H \rangle_C^{\beta,\rho;N}}{N} - \varepsilon' \right|,$$

where the first term on the right hand side does not depend on ε', we obtain by Hölder's inequality an estimate

$$
\left| \langle f \circ P_I \rangle_{\mathrm{MC}}^{\varepsilon,\rho;N} - \langle f \circ P_I \rangle_{\mathrm{C}}^{\beta,\rho;N} \right|
$$

$$
\leq C(\varepsilon,\rho) \left(\frac{|I|}{1 - \frac{|I|}{N}} \right)^{\frac{1}{p}} \left[\left| \varepsilon - \frac{\langle H \rangle_{\mathrm{C}}^{\beta,\rho;N}}{N} \right| \right.
$$

$$
\left. + \left(\frac{1}{\int d\varepsilon'\, e^{-N\beta\varepsilon'}\, Z_{\mathrm{MC}}(\varepsilon',\rho;N)} \int d\varepsilon'\, e^{-N\beta\varepsilon'}\, Z_{\mathrm{MC}}(\varepsilon',\rho;N) \left| \frac{\langle H \rangle_{\mathrm{C}}^{\beta,\rho;N}}{N} - \varepsilon' \right|^2 \right)^{\frac{1}{2}} \right]
$$

$$
= C(\varepsilon,\rho) \left(\frac{|I|}{1 - \frac{|I|}{N}} \right)^{\frac{1}{p}} \left(\sigma_{\mathrm{C}}^{\beta,\rho;N} \left(\frac{H}{N} \right) + \left| \varepsilon - \frac{\langle H \rangle_{\mathrm{C}}^{\beta,\rho;N}}{N} \right| \right),
$$

as desired. Then we use the generic properties listed in Sect. 1.1 to express the result in terms of the canonical free energy. □

Following the theme of the direct coupling method, the approach can also then be applied to the case of finite moments.

Theorem 2.6 *Let $\mu_{\mathrm{MC}}^{\varepsilon,\rho;N}$ be a permutation invariant probability measure corresponding to a microcanonical ensemble with energy density ε and particle density ρ. In addition, assume that if we fix a possible energy density ε' then for any other possible energy density ε there exists a constant $C(\varepsilon,\rho) > 0$ independent of ε' and N, but possibly dependent on ε and ρ such that*

$$
w_p(\mu_{\mathrm{MC}}^{\varepsilon,\rho;N}, \mu_{\mathrm{MC}}^{\varepsilon',\rho;N}; N) \leq C(\varepsilon,\rho)|\varepsilon - \varepsilon'|
$$

for some $p > 1$. Suppose also that the microcanonical and canonical measures, for some parameter β, have finite p_0:th moments for some $p_0 \geq p$.

Let J be a finite sequence of elements in $[N]$ where elements may be repeated, let $n_J := |J|$, and suppose that $n_J \leq p_0 + 1 - \frac{p_0}{p}$. Collect into $I \subset [N]$ the elements occurring in the sequence. It follows that

$$
\left| \left\langle \phi^J \right\rangle_{\mathrm{MC}}^{\varepsilon,\rho;N} - \left\langle \phi^J \right\rangle_{\mathrm{C}}^{\beta,\rho;N} \right|
$$

$$
\leq C(\varepsilon,\rho) n_J M(J,p)^{n_J - 1} \left(\frac{|I|}{1 - \frac{|I|}{N}} \right)^{\frac{1}{p}} \left(\sigma_{\mathrm{C}}^{\beta,\rho;N} \left(\frac{H}{N} \right) + \left| \varepsilon - \frac{\langle H \rangle_{\mathrm{C}}^{\beta,\rho;N}}{N} \right| \right),
$$

where, using the dual exponent $q = \frac{p}{p-1}$,

$$
M(J,p) := \max_{i \in I} \left(\left\langle |\phi_i|^{q(n_J - 1)} \right\rangle_{\mathrm{MC}}^{\varepsilon,\rho;N}, \left\langle |\phi_i|^{q(n_J - 1)} \right\rangle_{\mathrm{MC}}^{\varepsilon',\rho;N} \right)^{\frac{1}{q(n_J - 1)}} < \infty.
$$

Proof The proof is almost identical to the proof of the previous theorem. In order to isolate the moments of the canonical ensemble, one needs an additional application of Hölder's inequality. □

For suitable ensembles, these theorems together imply that with bounded moments, one can achieve an explicit rate of convergence of the finite dimensional moments and marginals of the ensembles.

3 Simple Model of a Paramagnet

In this section, we will consider a simple model of a paramagnet discussed in [11]. We begin by defining the magnetization of the lattice system and the two probability measures on the lattice that we will consider.

Definition 3.1 Let Λ be a finite lattice with size $N := |\Lambda|$. We define the space of spin configurations $\mathcal{S} := \{-1, 1\}^{\Lambda}$. Define the magnetization $M : \mathcal{S} \to \mathbb{R}$ by

$$M[\phi] := \sum_{x \in \Lambda} \phi(x).$$

Furthermore, we define the magnetization density $m_N : \mathcal{S} \to \mathbb{R}$ by $m_N[\phi] = \frac{M[\phi]}{N}$.

In the following definitions, there will be a slight misuse of the ensemble terminology. We will refer to the fixed magnetization probability measures as an auxiliary microcanonical ensemble, and the probability measure with a parameter controlling the expectation of magnetization will be referred to as an auxiliary canonical ensemble.

Definition 3.2 (*Fixed magnetization ensemble/auxiliary microcanonical ensemble*) Let $m \in \mathrm{Ran}[m_N]$. For such an m, define the set $\mathcal{S}_m := \{\phi \in \mathcal{S} : m_N[\phi] = m\}$. The auxiliary microcanonical ensemble with magnetization density m is defined via its action on functions $f : \mathcal{S} \to \mathbb{R}$ by

$$\langle f \rangle_{\mathrm{MC}}^{m;N} := \frac{1}{|\mathcal{S}_m|} \sum_{\phi \in \mathcal{S}_m} f(\phi).$$

Definition 3.3 (*Fluctuating magnetization/auxiliary canonical ensemble*) Let $\mu \in \mathbb{R}$. The auxiliary canonical ensemble with magnetic potential μ is defined via its action functions $f : \mathcal{S} \to \mathbb{R}$ by

$$\langle f \rangle_{\mathrm{C}}^{\mu;N} := \frac{1}{\sum_{\phi \in \mathcal{S}} e^{-\mu M[\phi]}} \sum_{\phi \in \mathcal{S}} e^{-\mu M[\phi]} f(\phi)$$

$$= \frac{1}{\sum_{m \in \mathrm{Ran}[m_N]} e^{-\mu m N} |\mathcal{S}_m|} \sum_{m \in \mathrm{Ran}[m_N]} e^{-\mu m N} |\mathcal{S}_m| \langle f \rangle_{\mathrm{MC}}^{m;N}.$$

The second representation is called the magnetization representation of this ensemble.

We will also need the following standard "thermodynamic" properties of these ensembles. We have compiled them in the following lemma.

Lemma 3.4 *For $m \in \mathrm{Ran}[m_N] \setminus \{-1, 1\}$, the partition function of the fixed magnetization ensemble is given by*

$$Z_{\mathrm{MC}}(m; N) := |\mathcal{S}_m| = \binom{N}{\frac{1+m}{2} N} = \left((N+1) \int_0^1 dt \, e^{N\left(\frac{1+m}{2} \ln t + \frac{1-m}{2} \ln(1-t)\right)} \right)^{-1}.$$

The partition function of the fluctuating magnetization ensemble is given by

$$Z_C(\mu; N) := \sum_{\phi \in \mathcal{S}} e^{-\mu M[\phi]} = (2\cosh(\mu))^N,$$

and the specific free energy is given by

$$f_C(\mu; N) := -\frac{1}{N} \ln Z_C(\mu; N) = -\ln(2\cosh(\mu)).$$

The average and standard deviation of the magnetization density are given by

$$\frac{\langle M \rangle_C^{\mu;N}}{N} = -\tanh(\mu),$$

and

$$\sigma_C^{\mu;N}\left(\frac{M}{N}\right) = \frac{1}{\sqrt{N}}\sqrt{1 - \tanh^2(\mu)}.$$

Proof The calculation of the auxiliary microcanonical partition function is based on the fact that the number of positive spins in a field configuration fully defines the total magnetization of the configuration, and, as a result, one only needs to consider the number of configurations with a specific number of positive spins. The final equality follows from the representation of the beta function after opening up the combination and subsequent factorials.

The rest of the results concerning the auxiliary canonical ensemble follow by first computing the partition function, by noting that the structure of the measures is that of a product measure. Then we can differentiate the free energy with respect to μ and divide appropriately by the degrees of freedom N. □

Next, we are going to present two distinct methods with which to compute upper bounds for the rate of convergence of expectations of functions between these two probability measures.

3.1 Relative Entropy

We will, again, follow the presentation of this topic given in [11]. We begin with the definition of relative entropy.

Definition 3.5 (*Relative entropy*) Let λ_1 and λ_2 be two probability measures on a space X. If λ_1 is not absolutely continuous with respect to λ_2, the relative entropy $\mathcal{H}(\lambda_1||\lambda_2) = \infty$. If λ_1 is absolutely continuous with respect to λ_2, then we have

$$\mathcal{H}(\lambda_1 \| \lambda_2) := \int_X d\lambda_1 \ln \frac{d\lambda_1}{d\lambda_2}.$$

For the paramagnet, we have the following calculation of the relative entropy.

Proposition 3.6 (*Upper and lower bounds for specific relative entropy*) *For all* $m \in \mathrm{Ran}[m_N]\backslash\{-1, 1\}$, $\mu \in \mathbb{R}$, *and* $\eta > 0$, *there a cutoff* $N(m, \eta) \in \mathbb{N}$ *such that for* $N \geq N(m, \eta)$, *we have*

$$\left(\frac{1}{2} - \eta\right)\frac{\ln(N+1)}{N} \leq \frac{\mathcal{H}(\mu_{\mathrm{MC}}^{m;N} \| \mu_C^{\mu;N})}{N} - F(m, \mu) \leq \frac{\ln(N+1)}{N}, \qquad (3.1)$$

where

$$F(m, \mu) := \ln 2 + \ln \cosh(\mu) + \mu m + \frac{1+m}{2}\ln\left(\frac{1+m}{2}\right) + \frac{1-m}{2}\ln\left(\frac{1-m}{2}\right).$$
$$(3.2)$$

Every pair $(m, \mu) \in (\mathrm{Ran}[m_N] \setminus \{-1, 1\}) \times \mathbb{R}$ *which satisfies* $m = -\tanh \mu$ *solves* $F(m, \mu) = 0$ *and there are no other solutions to this equation for allowed values of m and* μ.

Proof We have

$$\frac{d\mu_{\mathrm{MC}}^{m;N}}{d\mu_{\mathrm{C}}^{\mu;N}}(\phi) = \frac{\mathbb{1}(\phi \in \mathcal{S}_m)}{Z_{\mathrm{MC}}(m; N)} \frac{Z_{\mathrm{C}}(\mu; N)}{e^{-\mu m N}}.$$

It follows that

$$\frac{\mathcal{H}(\mu_{\mathrm{MC}}^{m;N} \| \mu_{\mathrm{C}}^{\mu;N})}{N} = \ln 2 + \ln \cosh(\mu) + \mu m$$

$$+ \frac{\ln(N+1)}{N} + \frac{1}{N} \ln \int_0^1 dt \, e^{N\left(\frac{1+m}{2} \ln t + \frac{1-m}{2} \ln(1-t)\right)}.$$

Define the function $f(t) := \frac{1+m}{2} \ln t + \frac{1-m}{2} \ln(1-t)$. It can be shown by differentiation that the mapping f is strictly concave and thus attains its unique maximum on the interval $(0, 1)$. This maximum is attained at the point $t_0 := \frac{1+m}{2} \in (0, 1)$. Therefore,

$$\frac{1}{N} \ln \int_0^1 dt \, e^{N\left(\frac{1+m}{2} \ln t + \frac{1-m}{2} \ln(1-t)\right)}$$

$$\leq f\left(\frac{1+m}{2}\right) = \frac{1+m}{2} \ln\left(\frac{1+m}{2}\right) + \frac{1-m}{2} \ln\left(\frac{1-m}{2}\right).$$

Recalling the definition of $F(m, \mu)$ in (3.2), we conclude that

$$\frac{\ln(N+1)}{N} + \frac{1}{N} \ln \int_0^1 dt \, e^{Nf(t)} - f(t_0) \leq \frac{\mathcal{H}(\mu_{\mathrm{MC}}^{m;N} \| \mu_{\mathrm{C}}^{\mu;N})}{N} - F(m, \mu) \leq \frac{\ln(N+1)}{N}.$$

By computation, if $m \in (-1, 1)$ and $\mu \in \mathbb{R}$ are such that $m = -\tanh \mu$, then $\mu = \frac{1}{2} \ln\left(\frac{1-m}{1+m}\right)$ and thus $F(m, \mu) = 0$. For fixed m, by differentiation we find that $\mu \mapsto F(m, \mu)$ is a strictly convex function with a minimum at $\mu = \frac{1}{2} \ln\left(\frac{1-m}{1+m}\right)$. Hence, this value of μ is the only zero of the function.

To get a useful lower bound, we need to inspect the values of the function f more carefully. Suppose $0 < \delta < \frac{1}{2}$ and consider $t \in [t_0(1-\delta), t_0]$. Then

$$-f''(t) = \frac{t_0}{t^2} + \frac{1-t_0}{(1-t)^2} \leq \frac{1}{(1-\delta)^2 t_0} + \frac{1}{1-t_0} \leq \frac{4}{t_0} + \frac{1}{1-t_0}.$$

and since $f'(t_0) = 0$, we find from Taylor's theorem that there is $K > 0$ which depends only on m via t_0 such that for the above choice of t, which have $|t - t_0| \leq \delta t_0$,

$$f(t_0) - f(t) \leq K^2 \delta^2.$$

Therefore, by the positivity of the integrand, we obtain

$$\int_0^1 dt \, e^{Nf(t)} \geq \int_{(1-\delta)t_0}^{t_0} dt \, e^{Nf(t)} \geq e^{Nf(t_0) - NK\delta^2} t_0 \delta,$$

which implies

$$\frac{\ln(N+1)}{N} + \frac{1}{N}\ln\int_0^1 dt \, e^{Nf(t)} - f(t_0) \geq -K\delta^2 + \frac{1}{N}\ln(t_0(N+1)\delta) .$$

Consider then arbitrary $p \in (0, \frac{1}{2})$, and set $\delta := \frac{1}{2}(N+1)^{-\frac{1}{2}-p}$. The lower bound becomes $-\frac{K}{4}(N+1)^{-1-2p} + \frac{1}{N}\ln(t_0(N+1)^{\frac{1}{2}-p}/2) \geq \frac{1}{N}\left(-\frac{K}{4} + \ln(t_0/2) + \left(\frac{1}{2} - p\right)\ln(N+1)\right)$. Therefore, there exists N_0 which depends only on m and p such that for all $N \geq N_0$, the lower bound is greater than $\frac{1}{N}\left(\frac{1}{2} - 2p\right)\ln(N+1)$. If $\eta < 1$ we can set $p = \eta/2$, and the above estimates together thus prove (3.1) for all $N \geq N_0$. The lower bound is trivial if $\eta \geq 1$. This concludes the proof of the Proposition. □

3.2 Coupling

For the use of the coupling method, the previous thermodynamic calculations concerning the auxiliary canonical ensemble will be needed. The additional ingredient is the explicit coupling for certain observables presented in the following theorem.

Theorem 3.7 *Let $m \in Ran[m_N]$ and $m' \in Ran[m_N]$. We have*

$$w_1(\mu_{MC}^{m;N}, \mu_{MC}^{m';N}; N) = |m' - m|.$$

Proof Denote $M = mN$ and $M' = m'N$. By symmetry, it suffices to prove the result in the case $M' > M$; note that if $M' = M$, then $\mu_{MC}^{m;N} = \mu_{MC}^{m';N}$. First, consider any field configuration $\phi \in S$ such that $M[\phi] = M$. Note that $\Lambda_+[\phi]$ corresponds to the sites on the lattice for which ϕ has a positive spin. Now, consider another field configuration such that $\phi'|_{\Lambda_+[\phi]} = \phi$ and $M[\phi'] = M'$. Such a configuration ϕ' can be constructed by taking the configuration ϕ and flipping negative spins to positive spins or vice versa until we obtain the magnetization M'.

As before, define $M_+ := |\Lambda_+[\phi]|$ and $M_- := N - M_+$, and set also $M'_+ := |\Lambda_+[\phi']|$. Then $M_+ = \frac{N+M}{2}$, $M_- = \frac{N-M}{2}$, and $M_+ < M'_+$. Denote $\Delta := M'_+ - M_+ = \frac{M'-M}{2} > 0$. The number of field configurations with magnetization M is given by $\binom{N}{M_+}$ or, equivalently, by $\binom{N}{M_-}$. In order to go from magnetization M to M', we must flip Δ negative sites to positive sites. The number of ways to do this is $\binom{M_-}{\Delta}$. Define $\gamma : S \times S \to \mathbb{R}$ by

$$\gamma(\phi, \phi') := \mathbb{1}(\phi \in S_m, \, \phi' \in S_{m'}, \, \phi'|_{\Lambda_+[\phi]} = \phi|_{\Lambda_+[\phi]}) \frac{1}{\binom{N}{M_-}} \frac{1}{\binom{M_-}{\Delta}}.$$

By construction, we have

$$\sum_{\phi' \in S} \gamma(\phi, \phi') = \frac{\mathbb{1}(\phi \in S_m)}{\binom{N}{M_-}} \frac{1}{\binom{M_-}{\Delta}} \sum_{\phi' \in S_{m'}} \mathbb{1}(\phi'|_{\Lambda_+(\phi)} = \phi|_{\Lambda_+[\phi]}) = \frac{\mathbb{1}(\phi \in S_m)}{\binom{N}{M_-}} \frac{1}{\binom{M_-}{\Delta}} \binom{M_-}{\Delta}$$

$$= \frac{\mathbb{1}(\phi \in S_m)}{\binom{N}{M_-}} = \frac{\mathbb{1}(\phi \in S_m)}{|S_m|}.$$

In the other direction, if we fix ϕ' with magnetization M' and consider the number of ways to go to a field configuration ϕ which agrees on the positive lattice sites of ϕ, then clearly we must take $\Delta = \frac{M'-M}{2}$ positive sites of ϕ' and flip them negative. The number of ways to do this is given by $\binom{M'_+}{\Delta}$ and we thus have

$$\sum_{\phi \in \mathcal{S}} \gamma(\phi, \phi') = \frac{\mathbb{1}(\phi' \in \mathcal{S}_{m'})}{\binom{N}{M_-}} \frac{1}{\binom{M_-}{\Delta}} \sum_{\phi \in \mathcal{S}_m} \mathbb{1}(\phi'|_{\Lambda_+[\phi]} = \phi|_{\Lambda_+[\phi]}) = \frac{\mathbb{1}(\phi' \in \mathcal{S}_{m'})}{\binom{N}{M_-}} \frac{1}{\binom{M_-}{\Delta}} \binom{M'_+}{\Delta}.$$

Now, we have the following simple binomial coefficient manipulations

$$\frac{\binom{M'_+}{\Delta}\binom{N}{M'_+}}{\binom{N}{M_-}\binom{M_-}{\Delta}} = \frac{\frac{(M'_+)!N!}{(M'_+-\Delta)!\Delta!(N-M'_+)!(M'_+)!}}{\frac{N!(M_-)!}{(N-M_-)!(M_-)!(M_--\Delta)!\Delta!}} = \frac{(N-M_-)!}{(M'_+-\Delta)!} \frac{(M_--\Delta)!}{(N-M'_+)!}$$

$$= \frac{(M_+)!}{(M_+)!} \frac{(M_- - M'_+ + M_+)!}{(M_- + M_+ - M'_+)!} = 1.$$

It follows that

$$\sum_{\phi \in \mathcal{S}} \gamma(\phi, \phi') = \frac{\mathbb{1}(\phi' \in \mathcal{S}_{m'})}{\binom{N}{M'_+}} = \frac{\mathbb{1}(\phi' \in \mathcal{S}_{m'})}{|\mathcal{S}_{m'}|}.$$

This verifies that γ is indeed a coupling between the fixed magnetization density ensembles with different magnetizations m and m'. For such a coupling, by construction, we have

$$\gamma(\phi, \phi') ||\phi - \phi'||_1 = 2\Delta\gamma(\phi, \phi') = (M' - M)\gamma(\phi, \phi'),$$

from which it follows that

$$w_1(\mu_{\text{MC}}^{m;N}, \mu_{\text{MC}}^{m';N}; N) \leq \frac{M' - M}{N} = m' - m.$$

On the other hand, if η is any other coupling of $\mu_{\text{MC}}^{m;N}$ and $\mu_{\text{MC}}^{m';N}$, we also have

$$\frac{1}{N} \int_{\mathcal{S} \times \mathcal{S}} \eta(d\phi, d\phi') \sum_{x \in \Lambda} |\phi(x) - \phi'(x)| \geq \frac{1}{N} \left| \left\langle \sum_{x \in \Lambda} \phi(x) - \sum_{x \in \Lambda} \phi'(x) \right\rangle_\eta \right| = m' - m.$$

This implies that the coupling γ is an optimal coupling, and, we have

$$w_1(\mu_{\text{MC}}^{m;N}, \mu_{\text{MC}}^{m';N}; N) = m' - m.$$

This completes the proof assuming $M' > M$, and hence by symmetry, also the proof of the Theorem. □

3.3 Convergence of Local Observables

In this subsection, we will present two distinct proofs of the convergence of local observables based on the two previously introduced objects. The first proof will be based on utilizing the Pinsker inequality and its relationship with relative entropy. This type of argument can be found in [7]. The argument uses the information divergence related methods from [5] in which a statement concerning the relationship between weak convergence and relative entropy is also given.

We must also emphasize that whatever we refer to here as the relative entropy method is precisely the collection of arguments and theorems that will be presented shortly. We are not stating that the classical inequalities could not be, for instance, strengthened or leveraged with other theorems in order to produce better results. For an example of this sort of work, we refer to [3].

3.3.1 Relative Entropy Method

Again, we will follow the example of [11]. We have the following theorem.

Theorem 3.8 *Let* $m \in Ran[m_N] \setminus \{-1, 1\}$ *and* $\mu \in \mathbb{R}$. *Let* $I \subset \Lambda$ *and suppose* $f : \{-1, 1\}^{|I|} \to \mathbb{R}$. *It follows that*

$$\frac{\left| \langle f \circ P_I \rangle_{MC}^{m;N} - \langle f \circ P_I \rangle_C^{\mu;N} \right|^2}{2} \leq |I| \left(\max_{\phi \in \{-1,1\}^{|I|}} |f(\phi)| \right)^2 \frac{\mathcal{H}(\mu_{MC}^{m;N} \| \mu_C^{\mu;N})}{N}$$

Proof The following proof can be considered a sketch of the standard argument presented for the same model in [11].

Since f is a bounded function, and the measures are absolutely continuous with respect to each other, we have

$$\left| \langle f \circ P_I \rangle_{MC}^{m;N} - \langle f \circ P_I \rangle_C^{\mu;N} \right| \leq \left(\max_{\phi \in \{-1,1\}^{|I|}} |f(\phi)| \right) ||\mu_{MC}^{m;N}|_I - \mu_C^{\mu;N}|_I||_{TV}.$$

By Pinsker's inequality, we have

$$\frac{||\mu_{MC}^{m;N}|_I - \mu_C^{\mu;N}|_I||_{TV}^2}{2} \leq \mathcal{H}(\mu_{MC}^{m;N}|_I \| \mu_C^{\mu;N}|_I).$$

Now, let $N \geq |I|$ and let $I_k \subset \Lambda$ be disjoint copies of size $|I|$ on the lattice such that $\Lambda \subset \bigcup_{k=1}^{K} I_k$ and $\bigcup_{k=1}^{K-1} I_k \subset \Lambda$. It follows that $(K-1)|I| \leq |\Lambda| \leq K|I|$ which implies that $\frac{1}{K} \leq \frac{|I|}{N}$. Now, by utilizing the fact that $\mu_C^{\mu;N}$ is a product measure, and that both measures are permutation invariant, it follows that

$$\mathcal{H}(\mu_{MC}^{m;N}|_I \| \mu_C^{\mu;N}|_I) \leq |I| \frac{\mathcal{H}(\mu_{MC}^{m;N} \| \mu_C^{\mu;N})}{N}.$$

The statement follows by combining these calculations. □

We give the full convergence rate in the following corollary. Here, and in the following, we employ the standard rigorous definition of the "O"-notation: given $g(N) \geq 0$ for $N \in \mathbb{N}$, "$X(N) = O(g(N))$" refers to the limit $N \to \infty$, i.e., it means there is an N-independent constant C and some $N_0 \in \mathbb{N}$ such that $|X(N)| \leq Cg(N)$ for all $N \geq N_0$. However, in these results, the constant C is allowed to depend on possible other parameters of the setup: for example, no uniformity of C in the parameters m, $|I|$ or $\|f\|_\infty$ is claimed below.

Corollary 3.9 *Let* $m \in Ran[m_N] \setminus \{-1, 1\}$ *and* $\mu = \tanh^{-1}(-m)$. *Let* $I \subset \Lambda$ *be a fixed size index set, and let* $f : \{-1, 1\}^{|I|} \to \mathbb{R}$. *It follows that*

$$\left| \langle f \circ P_I \rangle_{MC}^{m;N} - \langle f \circ P_I \rangle_C^{\mu;N} \right| \leq \sqrt{2|I|} \left(\max_{\phi \in \{-1,1\}^{|I|}} |f(\phi)| \right) \sqrt{\frac{\mathcal{H}(\mu_{MC}^{m;N} \| \mu_C^{\mu;N})}{N}}$$

$$\leq \sqrt{2|I|} \left(\max_{\phi \in \{-1,1\}^{|I|}} |f(\phi)| \right) \sqrt{\frac{\log(N+1)}{N}}.$$

By applying the relative entropy method, we have

$$\langle f \circ P_I \rangle_{MC}^{m;N} = \langle f \circ P_I \rangle_C^{\mu;N} + O\left(\sqrt{\log(N)}N^{-\frac{1}{2}}\right).$$

The factor of $\sqrt{\log(N)}$ cannot be removed by using this specific inequality since if we choose $\eta = \frac{1}{4}$, there exists a cutoff $N(m) \in \mathbb{N}$ such that for $N \geq N(m)$, we have

$$\frac{\ln(N+1)}{4N} \leq \frac{\mathcal{H}(\mu_{\text{MC}}^{m;N} \| \mu_{\text{C}}^{\mu;N})}{N}.$$

Proof The corollary follows directly by combining the contents and bounds from proposition 3.6 and theorem 3.8. □

3.3.2 Coupling Method

The main theorems formulated in the coupling section concern 1-Lipschitz functions with respect to some norm $|| \cdot ||_p$. Since the domain set is finite, all functions $f : \{-1, 1\}^{|I|} \to \mathbb{R}$ are automatically Lipschitz functions with respect to all of these norms. The choice of using $p = 1$ norm below is partially a matter of convenience, due to equivalence of the finite set p-norms, but one should be careful in the application of the result if the size of the set I is allowed to become unbounded as $N \to \infty$. We recall from Sect. 2.1.2 that the optimal Lipschitz constant does depend on the choice of norm, and it will affect the overall constant in the bounds, unless scaled to one, as we require here.

We can now state the full convergence theorem. We continue to use the notations introduced before Corollary 3.9, i.e., "$A(N) = B(N) + O(N^{-\frac{1}{2}})$" here means that there exists $N_0 \in \mathbb{N}$ and $C \geq 0$ such that $|A(N) - B(N)| \leq CN^{-\frac{1}{2}}$ for all $N \geq N_0$.

Theorem 3.10 *Let $m \in Ran[m_N] \setminus \{-1, 1\}$ and $\mu = \tanh^{-1}(-m)$. Let $I \subset \Lambda$ be a fixed size index set, and let $f : \{-1, 1\}^{|I|} \to \mathbb{R}$ be a bounded 1-Lipschitz function with respect to the $|| \cdot ||_1$ norm. It follows that*

$$\langle f \circ P_I \rangle_{MC}^{m;N} = \langle f \circ P_I \rangle_{C}^{\mu;N} + O(N^{-\frac{1}{2}}).$$

Proof The result follows by applying the free energy method presented in Theorem 2.5, along with the w_1 fluctuation distance bound presented in Theorem 3.7, and the equations in Lemma 3.4. □

For the auxiliary microcanonical ensemble with fixed magnetization density, the w_1 choice of cost function is natural since the $|| \cdot ||_1$-norm satisfies

$$|M[\phi] - M[\phi']| \leq ||\phi - \phi'||_1.$$

3.4 Relationship to the Curie–Weiss Model

Let us first recall the Curie–Weiss Hamiltonian.

Definition 3.11 (*Curie–Weiss Hamiltonian*) Let Λ be a finite lattice with size $N := |\Lambda|$. We define the space of spin configurations $\mathcal{S} := \{-1, 1\}^{\Lambda}$. Let $J > 0$ and $h \in \mathbb{R}$. Define the Hamiltonian $H : \mathcal{S} \to \mathbb{R}$ by

$$H[\phi] := -\frac{J}{2N} \sum_{x,y \in \Lambda} \phi(x)\phi(y) - h \sum_{x \in \Lambda} \phi(x).$$

The energy density $\varepsilon_N : \mathcal{S} \to \mathbb{R}$ is defined by $\varepsilon_N[\phi] := \frac{H[\phi]}{N}$.

Note that the Hamiltonian can be written in terms of the magnetization by

$$H[\phi] = -\frac{J}{2N}M[\phi]^2 - hM[\phi] = -\frac{J}{2N}\left(M[\phi] + \frac{hN}{J}\right)^2 + \frac{h^2N}{2J}.$$

This relation leads to a simplification when studying the microcanonical ensemble of the Curie–Weiss model, defined as follows.

Definition 3.12 (*Fixed energy density/Microcanonical ensemble*) Let $\varepsilon \in \mathbb{R}$ be such that $\varepsilon \in \mathrm{Ran}[\varepsilon_N]$. Define the set $\mathcal{S}_\varepsilon := \{\phi \in \mathcal{S} : \varepsilon_N[\phi] = \varepsilon\}$. The microcanonical ensemble with energy density ε is defined via its action on functions $f : \mathcal{S} \to \mathbb{R}$ by

$$\langle f \rangle_{\mathrm{MC}}^{\varepsilon;N} := \frac{1}{|\mathcal{S}_\varepsilon|} \sum_{\phi \in \mathcal{S}_\varepsilon} f(\phi).$$

We will always use the lower case letter m to specify fixed magnetization densities introduced in Definition 3.2, and ε for fixed energy densities so that there is no ambiguity.

In some sense, the fixed energy ensemble for some values of J and h is not necessarily fundamental as it can be represented as a convex combination of fixed magnetizations. The energy density can be written in terms of the magnetization density as

$$\varepsilon_N[\phi] = -\frac{J}{2}\left(m_N[\phi] + \frac{h}{J}\right)^2 + \frac{h^2}{2J} \iff m_{N,\pm}[\phi] = -\frac{h}{J} \pm \sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon_N[\phi]}{J}},$$

from which it is clear that for some values of h and J there are multiple magnetization densities which give the same energy density. The following lemma makes the previous statements more quantitative.

Lemma 3.13 *Let $\varepsilon \in \mathbb{R}$ be such that $\varepsilon \in \mathrm{Ran}[\varepsilon_N]$ and define $m_\pm = -\frac{h}{J} \pm \sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}$. We have*

$$\langle f \rangle_{\mathrm{MC}}^{\varepsilon;N} = \frac{|\mathcal{S}_{m_+}|}{|\mathcal{S}_{m_+}| + |\mathcal{S}_{m_-}|} \langle f \rangle_{\mathrm{MC}}^{m_+;N} + \frac{|\mathcal{S}_{m_-}|}{|\mathcal{S}_{m_+}| + |\mathcal{S}_{m_-}|} \langle f \rangle_{\mathrm{MC}}^{m_-;N}, \tag{3.3}$$

with the convention that $\langle f \rangle_{\mathrm{MC}}^{m;N} = 0$ and $|\mathcal{S}_m| = 0$ if $m \notin \mathrm{Ran}[m_M]$.

Proof Since $\varepsilon \in \mathrm{Ran}[\varepsilon_N]$, it follows that $\varepsilon \leq \frac{h^2}{2J}$. If $\varepsilon = \frac{h^2}{2J}$, we have $m_+ = m_-$, $\mathcal{S}_m = \mathcal{S}_\varepsilon$ for $m = m_\pm$, and thus $\langle f \rangle_{\mathrm{MC}}^{\varepsilon;N} = \frac{1}{2}\langle f \rangle_{\mathrm{MC}}^{m_+;N} + \frac{1}{2}\langle f \rangle_{\mathrm{MC}}^{m_-;N}$. Hence, (3.3) holds in this case.

We may thus assume that $\varepsilon < \frac{h^2}{2J}$, when $m_- < m_+$. We have $\mathcal{S}_\varepsilon = \mathcal{S}_{m_+} \cup \mathcal{S}_{m_-}$ and $\mathcal{S}_{m_+} \cap \mathcal{S}_{m_-} = \emptyset$. Thus, $|\mathcal{S}_\varepsilon| = |\mathcal{S}_{m_-}| + |\mathcal{S}_{m_+}|$ and

$$\frac{1}{|\mathcal{S}_\varepsilon|} \sum_{\phi \in \mathcal{S}_\varepsilon} f(\phi) = \frac{1}{|\mathcal{S}_{m_-}| + |\mathcal{S}_{m_+}|}\left(|\mathcal{S}_{m_+}|\frac{1}{|\mathcal{S}_{m_+}|} \sum_{\phi \in \mathcal{S}_{m_+}} f(\phi) + |\mathcal{S}_{m_-}|\frac{1}{|\mathcal{S}_{m_-}|} \sum_{\phi \in \mathcal{S}_{m_-}} f(\phi)\right),$$

from which the statement follows. □

From the previous lemma, it is apparent that if one has knowledge of the auxiliary microcanonical partition function and local observables of the fixed magnetization ensemble, then, in principle, one has a full description of the weak convergence properties of corresponding fixed energy density ensemble. In the later sections, we will give an example of this exact kind of analysis with the mean-field spherical model.

3.5 Final Remarks and Comparison of Methods

From the results in the previous subsections, we can see that the coupling method generates a strictly better convergence rate by removing the factor of $\sqrt{\log(N)}$ which is an irremovable part of the magnitude of specific relative entropy in this case. For the relative entropy method, the primary object of interest is the calculation of the specific relative entropy. As can be seen, the selection of the magnetization density m and the parameter μ comes down to solving an equation which relates these parameters. Once the pair has been realized from this equation, we obtain the desired upper bound. In our calculation, the partition functions of both the auxiliary microcanonical ensemble and auxiliary canonical ensemble needed to be estimated sufficiently accurately (other more refined approaches for using relative entropy may be found in [12]).

For the coupling method, one needed to come up with the coupling of the auxiliary microcanonical ensembles resulting in Theorem 3.7. In addition, one uses the standard "thermodynamic" relations for the auxiliary canonical ensemble, given in Lemma 3.4. Because the auxiliary canonical ensemble was in product form, the calculations were particularly simple.

From these observations, the main difference between the methods concerns the treatment of the auxiliary microcanonical ensemble. In the above direct relative entropy computation, we need to calculate the partition function of the auxiliary microcanonical ensemble and the auxiliary canonical ensemble, but the "thermodynamic" relations of the auxiliary canonical ensemble do not seem important. For the coupling method, the auxiliary microcanonical partition function does not play the same role, and the coupling is the most important object along with the "thermodynamic" relations from the auxiliary canonical ensemble.

4 Mean-Field Spherical Model Aka Continuum Curie–Weiss Model

For this model, we will need to clarify the goals and priority of the limiting measures for the local convergence result. Our main goal is to analyse explicitly the probability measure associated with the microcanonical ensemble. To that end, there will be some local convergence results in which thermodynamic equivalence might hold between two ensembles, but we will opt for a simpler auxiliary ensemble measure as the approximating measure.

In particular, for this model, we will see that the simplest limit measure to consider will be either a Gaussian measure or a convex combination of Gaussian measures. We will also prove some local convergence results where the limiting measure is not a product measure.

In the second model considered here, the "spin-field" ϕ is allowed to take all real values otherwise being similar to the discrete Curie–Weiss model.

Definition 4.1 (*Continuum Curie–Weiss Hamiltonian*) Let Λ be a finite lattice with size $N := |\Lambda|$. We define the space of field configurations $\mathcal{S} := \mathbb{R}^\Lambda$. Let $J > 0$ and $h \in \mathbb{R}$. Define the Hamiltonian $H : \mathcal{S} \to \mathbb{R}$ and the particle number $N : \mathcal{S} \to \mathbb{R}$ by

$$H[\phi] := -\frac{J}{2N} \sum_{x,y \in \Lambda} \phi(x)\phi(y) - h \sum_{x \in \Lambda} \phi(x), \quad N[\phi] := \sum_{x \in \Lambda} \phi(x)^2.$$

We also define the magnetization $M : \mathcal{S} \to \mathbb{R}$ by

$$M[\phi] := \sum_{x \in \Lambda} \phi(x).$$

Furthermore, we define the energy density $\varepsilon : \mathcal{S} \to \mathbb{R}$ and magnetiztion density $m : \mathcal{S} \to \mathbb{R}$ by $\varepsilon[\phi] := \frac{H[\phi]}{N}$ and $m[\phi] := \frac{M[\phi]}{N}$.

As in the discrete case, the Hamiltonian can be written in terms of the magnetization,

$$H[\phi] = -\frac{J}{2N} M[\phi]^2 - h M[\phi] = -\frac{J}{2N} \left(M[\phi] + \frac{hN}{J} \right)^2 + \frac{h^2 N}{2J}.$$

In this model, the particle number function $N[\cdot]$ is much more relevant than in the discrete case. For this Hamiltonian, we will need to consider probability measures described by products of delta functions. To properly resolve them, we begin with an observation concerning a matrix relevant to the definitions of the ensembles. In the following, we employ the notation $M_N(\mathbb{R})$ for the collection of real $N \times N$ matrices.

Lemma 4.2 *Define $M \in M_N(\mathbb{R})$ by $M_{ij} := 1$ for all $i \in [N]$ and $j \in [N]$. There exists an orthogonal matrix $U \in M_N(\mathbb{R})$ which diagonalizes M such that for $x \in \mathbb{R}^N$, we have*

$$(Ux)_1 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i.$$

Proof A simple analysis shows that M has an eigenvalue N with no degeneracy, and an eigenvalue 0 with $N-1$ fold degeneracy. Collecting the eigenvalues into a diagonal matrix results in $D \in M_N(\mathbb{R})$ for which $D_{11} := N$ and $D_{ij} := 0$ for all other $i \in [N]$ and $j \in [N]$. Since M is a real symmetric matrix, then there exists an orthogonal matrix $Q \in M_N(\mathbb{R})$ such that

$$Q^T D Q = M.$$

Writing out the above matrix multiplication componentwise explicitly, we find for all i, j

$$N Q_{1i} Q_{1j} = 1.$$

In particular, then $|Q_{1i}| = \frac{1}{\sqrt{N}}$ for all $i \in [N]$, and thus for each i there is $\sigma_i \in \{\pm 1\}$ such that $Q_{1i} = \sigma_i \frac{1}{\sqrt{N}}$. Using a proof by contradiction, one can see that, in fact, the elements Q_{1i} must either all be negative or all be positive. Now, define $U \in M_N(\mathbb{R})$ by $U := -Q$ if the elements Q_{1i} are all negative, and $U := Q$ if the elements Q_{1i} are all positive. It follows that U is an orthogonal matrix and, by definition, we have

$$(Ux)_1 := \sum_{j=1}^{N} U_{1j} x_j = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} x_i.$$

This completes the proof of the Lemma.   □

Next, we will give two examples of how to apply δ-function calculation rules to resolve the ones relevant to the Curie–Weiss system, for both microcanonical and canonical ensembles. It is possible to prove the validity of these manipulations under the assumptions made in the Examples, for instance, following the discussion in Appendix A of [13].

Example 4.3 Let $U : \mathcal{S} \to \mathbb{R} \times \mathbb{R}^{N-1}$ be an orthogonal matrix satisfying Lemma 4.2. Suppose $\rho > 0$ and $m \in (-\sqrt{\rho}, \sqrt{\rho})$. For bounded 1-Lipschitz functions $f : \mathcal{S} \to \mathbb{R}$, we have

$$\int_{\mathcal{S}} d\phi \, \delta \left(\sum_{x \in \Lambda} \phi(x) - mN \right) \delta \left(\sum_{x \in \Lambda} \phi(x)^2 - \rho N \right) f(\phi)$$

$$= \int_{\mathbb{R}} dz \int_{\mathbb{R}^{N-1}} d\psi \; \delta(\sqrt{N}z - mN)\delta(z^2 + ||\psi||^2 - \rho N) \left(f \circ U^{-1}\right)(z, \psi)$$

$$= \frac{(N(\rho - m^2))^{\frac{N-3}{2}}}{2\sqrt{N}} \int_{\mathbb{S}^{N-2}} d\Omega \; \left(f \circ U^{-1}\right)(m\sqrt{N}, \sqrt{N(\rho - m^2)}\Omega).$$

Here we have first made a change of variables to $(z, \psi) = U\phi$ and then used spherical coordinates system to integrate out the resulting δ-functions. Since the left hand side does not depend on the choice of the matrix U, all choices must result in the same value for the integral on the right hand side.

Example 4.4 Let $U : \mathcal{S} \to \mathbb{R} \times \mathbb{R}^{N-1}$ be an orthogonal matrix satisfying Lemma 4.2. Fix $h \in \mathbb{R}$ and $\rho > 0$, and suppose $\varepsilon \in \mathbb{R}$ satisfies $\varepsilon < \frac{h^2}{2J}$. Define then

$$m_+ := -\frac{h}{J} + \sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}, \quad m_- := -\frac{h}{J} - \sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}.$$

Then, m_-, m_+ are distinct real numbers, and we assume furthermore that $m_+^2 < \rho$ and $m_-^2 < \rho$.

Then for bounded 1-Lipschitz functions $f : \mathcal{S} \to \mathbb{R}$, we may proceed as in the previous Example to conclude that

$$\int_{\mathcal{S}} d\phi \; \delta \left(-\frac{J}{2N} \sum_{x,y \in \Lambda} \phi(x)\phi(y) - h \sum_{x \in \Lambda} \phi(x) - \varepsilon N\right) \delta \left(\sum_{x \in \Lambda} \phi(x)^2 - \rho N\right) f(\phi)$$

$$= \int_{\mathbb{R}} dz \int_{\mathbb{R}^{N-1}} d\psi \; \delta \left(-\frac{J}{2}z^2 - h\sqrt{N}z - \varepsilon N\right) \delta(z^2 + ||\psi||^2 - \rho N) \left(f \circ U^{-1}\right)(z, \psi)$$

$$= \frac{1}{J\sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}\sqrt{N}} \int_{\mathbb{R}} dz \int_{\mathbb{R}^{N-1}} d\psi \; \delta \left(z - m_+\sqrt{N}\right) \delta(z^2 + ||\psi||^2 - \rho N) \left(f \circ U^{-1}\right)(z, \psi)$$

$$+ \frac{1}{J\sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}\sqrt{N}} \int_{\mathbb{R}} dz \int_{\mathbb{R}^{N-1}} d\psi \; \delta \left(z - m_-\sqrt{N}\right) \delta(z^2 + ||\psi||^2 - \rho N) \left(f \circ U^{-1}\right)(z, \psi)$$

$$= \frac{1}{J\sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}\sqrt{N}} \frac{(N(\rho - m_+^2))^{\frac{N-3}{2}}}{2\sqrt{N}} \int_{\mathbb{S}^{N-2}} d\Omega \; \left(f \circ U^{-1}\right)(m_+\sqrt{N}, \sqrt{N(\rho - m_+^2)}\Omega)$$

$$+ \frac{1}{J\sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}\sqrt{N}} \frac{(N(\rho - m_-^2))^{\frac{N-3}{2}}}{2\sqrt{N}} \int_{\mathbb{S}^{N-2}} d\Omega \; \left(f \circ U^{-1}\right)(m_-\sqrt{N}, \sqrt{N(\rho - m_-^2)}\Omega).$$

We will utilize these forms for more explicit definitions of the ensembles and in the proof concerning the boundedness of moments of the microcanonical ensembles.

4.1 Microcanonical Analysis

From here on, whenever the mapping U is present, we are always referring to the mapping U defined by a matrix satisfying Lemma 4.2. We fix the choice of this matrix in the following. We begin with definitions of the two ensembles, related to fixed magnetization and to fixed energy.

Definition 4.5 (*Fixed magnetization density and particle density/auxiliary microcanonical ensemble*) Let $\rho > 0$ and $m \in (-\sqrt{\rho}, \sqrt{\rho})$. The auxiliary microcanonical ensemble with

particle density ρ and magnetization density m is defined via its action on bounded 1-Lipschitz functions $f : \mathcal{S} \to \mathbb{R}$ by

$$\langle f \rangle_{\mathrm{MC}}^{m,\rho;N} := \frac{1}{|\mathbb{S}^{N-2}|} \int_{\mathbb{S}^{N-2}} d\Omega \ \left(f \circ U^{-1} \right) (m\sqrt{N}, \sqrt{N(\rho - m^2)}\Omega).$$

Furthermore, we define the specific auxiliary microcanonical partition function by

$$Z_{\mathrm{MC}}(m, \rho; N) := (\rho - m^2)^{\frac{N-3}{2}},$$

and the specific auxiliary microcanonical entropy by

$$s_{\mathrm{MC}}(m, \rho) := \frac{1}{2} \ln(\rho - m^2).$$

Definition 4.6 (*Fixed energy density and particle density/microcanonical ensemble*) Let $\rho > 0$, and ε, m_+, and m_- be as in Example 4.4, in particular, assume $\varepsilon < \frac{h^2}{2J}$ and $m_-^2, m_+^2 < \rho$. The microcanonical ensemble with energy density ε and particle density $\rho > 0$ is then defined via its action on bounded 1-Lipschitz functions $f : \mathcal{S} \to \mathbb{R}$ by

$$\langle f \rangle_{\mathrm{MC}}^{\varepsilon,\rho;N} := \frac{Z_{\mathrm{MC}}(m_+, \rho; N)}{Z_{\mathrm{MC}}(m_+, \rho; N) + Z_{\mathrm{MC}}(m_-, \rho; N)} \langle f \rangle_{\mathrm{MC}}^{m_+,\rho;N}$$
$$+ \frac{Z_{\mathrm{MC}}(m_-, \rho; N)}{Z_{\mathrm{MC}}(m_+, \rho; N) + Z_{\mathrm{MC}}(m_-, \rho; N)} \langle f \rangle_{\mathrm{MC}}^{m_-,\rho;N}.$$

If $\varepsilon < \frac{h^2}{2J}$ but $\min(m_-^2, m_+^2) < \rho \leq \max(m_-^2, m_+^2)$, we set $\langle f \rangle_{\mathrm{MC}}^{\varepsilon,\rho;N} := \langle f \rangle_{\mathrm{MC}}^{m,\rho;N}$ with $m := m_+$ if $|m_+| < |m_-|$, and $m := m_-$, otherwise.

If $\varepsilon = \frac{h^2}{2J}$ and $\rho > \frac{h^2}{J^2}$, we set $\langle f \rangle_{\mathrm{MC}}^{\varepsilon,\rho;N} := \langle f \rangle_{\mathrm{MC}}^{m,\rho;N}$ with $m := -\frac{h}{J}$.

One can indeed verify by using the calculations in Examples 4.3 and 4.4 that these measures correspond to δ-function definitions, resolved in the manner used in the Examples. The second definition serves as an explanation of the choice of multiplicative constants in the definition of the microcanonical partition function and specific entropy. In the degenerate case $\varepsilon = \frac{h^2}{2J}$, we would have above $m_+ = m_- = -\frac{h}{J} = m$, and in this case the δ-function definition in Example 4.4 does not really make sense since it would contain a singular term $\delta\left((z - m\sqrt{N})^2\right)$. Following these observations, we define the fixed energy microcanonical ensemble via the corresponding fixed magnetization ensemble.

Note that in addition to values of (ε, ρ) for which there are no solutions to the constraints, we have also left undefined the degenerate energy ensembles for which $\varepsilon \leq \frac{h^2}{2J}$ but $\min(m_-^2, m_+^2) = \rho$, as well as the degenerate magnetization ensembles with $m^2 = \rho$. In theses cases, the dimensionality of the solution manifold does not increase with N since all solutions have $\psi = 0$. As such, the resulting degenerate ensemble does not have standard thermodynamic behaviour.

We begin by estimating the fluctuation distance of two fixed magnetization ensembles by constructing a suitable transport map between them.

Theorem 4.7 *Let $\rho > 0$ and $m, m' \in (-\sqrt{\rho}, \sqrt{\rho})$. Consider some $N \in \mathbb{N}$. We have*

$$w_2(\mu_{MC}^{m,\rho;N}, \mu_{MC}^{m',\rho;N}; N) \leq \left(1 + \frac{2}{\sqrt{1 - m^2/\rho}}\right) |m - m'|.$$

Proof Define $T : \mathbb{R} \times \mathbb{R}^{N-1} \to \mathbb{R} \times \mathbb{R}^{N-1}$ by

$$T(z, \psi) := \left(m' \sqrt{N}, \sqrt{N(\rho - m'^2)} \frac{\psi}{||\psi||_2} \right).$$

Note that then for any $\Omega \in \mathbb{S}^{N-2}$ we have $T(m\sqrt{N}, \sqrt{N(\rho - m^2)}\Omega) = (m'\sqrt{N}, \sqrt{N(\rho - (m')^2)}\Omega)$. In order for this mapping to act on the correct coordinate space, we define $T' : \mathbb{R}^\Lambda \to \mathbb{R}^\Lambda$ by $T' := U^{-1} \circ T \circ U$. Then for any observable f we obtain directly from the definitions a relation

$$\langle f \circ T' \rangle_{\mathrm{MC}}^{m,\rho;N} = \langle f \rangle_{\mathrm{MC}}^{m',\rho;N}.$$

Therefore, T' is a transport map from the measure $\mu_{\mathrm{MC}}^{m,\rho;N}$ to $\mu_{\mathrm{MC}}^{m',\rho;N}$. Let γ denote the associated coupling as defined in Sect. 2.1.1. This yields an estimate

$$w_2(\mu_{\mathrm{MC}}^{m,\rho;N}, \mu_{\mathrm{MC}}^{m',\rho;N}; N)^2 \le \int \gamma(\mathrm{d}\phi, \mathrm{d}\psi) \frac{1}{N} \|\phi - \psi\|_2^2 = \int \mu_{\mathrm{MC}}^{m,\rho;N}(\mathrm{d}\phi) \frac{1}{N} \|\phi - T'\phi\|_2^2$$

$$= \frac{1}{N} \frac{1}{|\mathbb{S}^{N-2}|} \int_{\mathbb{S}^{N-2}} d\Omega \left\| U^{-1}(m\sqrt{N}, \sqrt{N(\rho - m^2)}\Omega) - U^{-1}(m'\sqrt{N}, \sqrt{N(\rho - m'^2)}\Omega) \right\|^2$$

$$= \frac{1}{N} \frac{1}{|\mathbb{S}^{N-2}|} \int_{\mathbb{S}^{N-2}} d\Omega \left\| (m\sqrt{N}, \sqrt{N(\rho - m^2)}\Omega) - (m'\sqrt{N}, \sqrt{N(\rho - m'^2)}\Omega) \right\|^2$$

$$= (m - m')^2 + \left(\sqrt{\rho - m^2} - \sqrt{\rho - m'^2} \right)^2 \le \left(1 + \frac{4\rho}{\rho - m^2} \right)(m - m')^2.$$

Since $1 + x^2 \le (1 + x)^2$ for $x \ge 0$, we obtain the stated bound after taking a square root. $\square$

In the previous theorem, the fluctuation distance is seemingly bounded asymmetrically with respect to the magnetization densities m and m'. By symmetry, the bound holds for either choice, and thus a symmetric bound can also be straightforwardly derived. The reason for the asymmetric choice is that while using the fluctuation distance, we will always consider one of the magnetization densities to be fixed.

To study the fixed energy ensembles, we begin with a Lemma which implies that, for $h \ne 0$, one of the fixed magnetization measures dominates in the fixed energy ensemble.

Lemma 4.8 *Consider some $N \in \mathbb{N}$. If $h \ne 0$, then*

$$\max\{Z_{MC}(m_+, \rho; N), Z_{MC}(m_-, \rho; N)\} = \left(\rho - \left(\frac{|h|}{J} - \sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}} \right)^2 \right)^{\frac{N-3}{2}}.$$

Proof If we consider the mapping $m \mapsto Z_{\mathrm{MC}}(m, \rho; N)$, then it is clear that $Z_{\mathrm{MC}}(m, \rho; N) \ge Z_{\mathrm{MC}}(m', \rho; N)$ for all $|m| \le |m'| < \sqrt{\rho}$. Now, note that

$$m_{\pm}^2 = \frac{h^2}{J^2} + \left(\frac{h^2}{J^2} - \frac{2\varepsilon}{J} \right) \mp 2\frac{h}{J}\sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}.$$

If $h > 0$, then $m_-^2 > m_+^2 \implies |m_-| > |m_+|$, and, if $h < 0$, then $m_+^2 < m_-^2 \implies |m_+| > |m_-|$. The result follows by plugging in the values in the partition function. $\square$

With the above computations, we can also now give a simpler definition of the set of allowed energies, i.e., of those values of ε for which the fixed energy ensemble is defined using Definition 4.6.

Definition 4.9 For $h \in \mathbb{R}$ and $\rho > 0$, we define the set of possible energy densities $\mathcal{E}_{h,\rho}$ by

$$\mathcal{E}_{h,\rho} := \left\{ \varepsilon \in \mathbb{R} : \varepsilon \le \frac{h^2}{2J}, \left| \frac{|h|}{J} - \sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}} \right| < \sqrt{\rho} \right\}.$$

We remark that for all of the above h, ρ the set $\mathcal{E}_{h,\rho}$ contains $\varepsilon = 0$ and an interval of negative values of ε. In particular, $\mathcal{E}_{h,\rho}$ is non-empty. Also, in case $h = 0$, we have $\mathcal{E}_{0,\rho} = \left(-\frac{\rho J}{2}, 0 \right]$.

Theorem 4.10 *Let $h \ne 0$, $\rho > 0$, and suppose $\varepsilon \in \mathcal{E}_{h,\rho}$. For integrable functions $f : \mathcal{S} \to \mathbb{R}$ satisfying $\langle |f| \rangle_{MC}^{m_\pm,\rho;N} \le K$ for some $K \ge 0$ and whenever the ensemble is defined, we have*

$$\left| \langle f \rangle_{MC}^{\varepsilon,\rho;N} - \langle f \rangle_{MC}^{m,\rho;N} \right| \le 2K \left| \frac{\rho - \left(\frac{|h|}{J} + \sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}} \right)^2}{\rho - \left(\frac{|h|}{J} - \sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}} \right)^2} \right|^{\frac{N-3}{2}}, \tag{4.1}$$

where

$$m = -\frac{h}{J} + \operatorname{sgn}(h)\sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}.$$

In addition,

$$\langle f \rangle_{MC}^{\varepsilon,\rho;N} = \langle f \rangle_{MC}^{m,\rho;N} + O(e^{-cN})$$

for some positive constant $c > 0$.

Proof If $\varepsilon = \frac{h^2}{2J}$, we have here $m = -\frac{h}{J}$, and the Theorem is trivially true since then $\langle f \rangle_{MC}^{\varepsilon,\rho;N} = \langle f \rangle_{MC}^{m,\rho;N}$. The same holds for those values where $\varepsilon < \frac{h^2}{2J}$ and $\rho \le \max(m_-^2, m_+^2)$.

In the remaining cases, we necessarily have $\varepsilon < \frac{h^2}{2J}$ and $m_-^2, m_+^2 < \rho$. Applying lemma 4.8, we have the following estimate

$$\left| \langle f \rangle_{MC}^{\varepsilon,\rho;N} - \langle f \rangle_{MC}^{m,\rho;N} \right| \le \left(\langle |f| \rangle_{MC}^{m_+,\rho;N} + \langle |f| \rangle_{MC}^{m_-,\rho;N} \right) \frac{Z_{MC}(m_{-\operatorname{sgn}(h)}, \rho; N)}{Z_{MC}(m_{\operatorname{sgn}(h)}, \rho; N)}.$$

The results follow since the term inside the absolute values on the right hand side in (4.1) is strictly less than one. □

If $h = 0$, there exists a suitable coupling which can be constructed from the couplings used for the fixed magnetization ensembles.

Theorem 4.11 *Let $h = 0$ and $\varepsilon', \varepsilon \in \left(-\frac{\rho J}{2}, 0 \right]$. For $\varepsilon \ne 0$, we have*

$$w_2(\mu_{MC}^{\varepsilon,\rho;N}, \mu_{MC}^{\varepsilon',\rho;N}; N) \le \frac{2}{J} \frac{1}{\sqrt{-\frac{2\varepsilon}{J}}\sqrt{1 - \left(-\frac{2\varepsilon}{J\rho} \right)}} |\varepsilon - \varepsilon'|,$$

and, for $\varepsilon = 0$, we have

$$w_2(\mu_{MC}^{0,\rho;N}, \mu_{MC}^{\varepsilon',\rho;N}; N) \le \frac{2}{\sqrt{J}} |\varepsilon'|^{\frac{1}{2}}.$$

Proof Now $m_\pm = \pm\sqrt{\frac{-2\varepsilon}{J}}$, and thus if $\varepsilon \neq 0$, we have $0 < |m_\pm| < \rho$. Therefore, these ensembles are defined as in the first case in Definition 4.6.

Suppose first that $\varepsilon, \varepsilon' < 0$ and let $m_\pm$ and $m'_\pm$ be corresponding positive and negative magnetization densities to ε and ε', respectively. To obtain a transport map, we define $T :$ $\mathbb{R} \times \mathbb{R}^{N-1} \to \mathbb{R} \times \mathbb{R}^{N-1}$ by

$$T(z, \psi) := \mathbb{1}(z \geq 0) \left(m'_+ \sqrt{N}, \sqrt{(\rho - m'^2_+)N} \frac{\psi}{||\psi||_2} \right)$$
$$+ \mathbb{1}(z < 0) \left(m'_- \sqrt{N}, \sqrt{(\rho - m'^2_-)N} \frac{\psi}{||\psi||_2} \right).$$

Now, let $U \in M_N(\mathbb{R})$ be the same unitary mapping as before. By setting $T' := U^{-1} \circ T \circ U$ and going through the same calculations as earlier, one can confirm that $\langle f \circ T' \rangle^{\varepsilon, \rho; N}_{MC} = \langle f \rangle^{\varepsilon', \rho; N}_{MC}$ for all observables f. Thus T' is a transport map and the associated coupling yields a bound

$$w_2(\mu^{\varepsilon, \rho; N}_{MC}, \mu^{\varepsilon', \rho; N}_{MC}; N)^2 \leq \frac{1}{2} \sum_{\sigma = \pm 1} \left[(m_\sigma - m'_\sigma)^2 + \left(\sqrt{\rho - m^2_\sigma} - \sqrt{\rho - (m'_\sigma)^2} \right)^2 \right]$$

$$= \frac{2}{J}(\sqrt{-\varepsilon} - \sqrt{-\varepsilon'})^2 + \left(\sqrt{\rho - \frac{-2\varepsilon}{J}} - \sqrt{\rho - \frac{-2\varepsilon'}{J}} \right)^2$$

$$\leq \frac{4}{J^2} \frac{\rho}{\left(-\frac{2\varepsilon}{J}\right)\left(\rho - \left(-\frac{2\varepsilon}{J}\right)\right)} (\varepsilon - \varepsilon')^2.$$

If $\varepsilon = 0 > \varepsilon'$, we have $m = 0$ but it still holds that $\langle f \rangle^{\varepsilon, \rho; N}_{MC} = \frac{1}{2}\langle f \rangle^{m, \rho; N}_{MC} + \frac{1}{2}\langle f \rangle^{m, \rho; N}_{MC}$. Proceeding as above then yields an estimate

$$w_2(\mu^{0, \rho; N}_{MC}, \mu^{\varepsilon', \rho; N}_{MC}; N)^2 \leq \frac{2}{J}(-\varepsilon') + \left(\sqrt{\rho} - \sqrt{\rho - \frac{-2\varepsilon'}{J}} \right)^2 \leq \frac{2}{J}|\varepsilon'| + \frac{4}{\rho J^2}\varepsilon'^2 \leq \frac{4}{J}|\varepsilon'|.$$

The bound is also trivially true if $\varepsilon = \varepsilon' = 0$. Combining the above estimates proves the statement in the Theorem. $\square$

4.2 Canonical Analysis

Compared to the regular Curie-Weiss model from [10], the canonical ensemble is somewhat simpler to analyse. In particular, one should note that when applying the identity

$$e^{\frac{x^2}{2}} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} dz \, e^{-\frac{z^2}{2} - xz},$$

which is sometimes referred to as Gaussian linearization, to solve the partition function of the regular Curie-Weiss model, we are effectively adding an extra variable over which to integrate when using Laplace's method. To this end, in order to avoid multi-dimensional Laplace analysis, we can forego the use of Gaussian linearization, and use a direct 1-dimensional Laplace method.

In the following definitions and calculations, there will be a significant difference in the treatment of the models depending on whether we are dealing with $h = 0$ or $h \neq 0$. In the previous section, for $h = 0$, we constructed an explicit coupling between the fixed energy density ensembles. In the case of $h \neq 0$, it will turn out that the coupling between the fixed

magnetization density ensembles is the more important object of study. This phenomenon seems to be closely related to the phase transitions in the mean-field spherical model which are thoroughly presented and analysed in [9].

We will define the canonical ensembles with the help of the microcanonical ensembles.

Definition 4.12 (*Fluctuating magnetization and fixed particle density/auxiliary canonical ensemble*) Let $\rho > 0$ and $\mu \in \mathbb{R}$. The auxiliary canonical ensemble with magnetic potential μ and particle density ρ is defined via its action on bounded 1-Lipschitz functions $f : \mathcal{S} \to \mathbb{R}$ by

$$\langle f \rangle_{\mathrm{C}}^{\mu,\rho;N} := \frac{1}{\int_{-\sqrt{\rho}}^{\sqrt{\rho}} dm \, e^{-\mu N m} Z_{\mathrm{MC}}(m, \rho; N)} \int_{-\sqrt{\rho}}^{\sqrt{\rho}} dm \, e^{-\mu N m} Z_{\mathrm{MC}}(m, \rho; N) \langle f \rangle_{\mathrm{MC}}^{m,\rho;N}$$

$$= \frac{1}{Z_{\mathrm{C}}(\mu, \rho; N)} \int_{-\sqrt{\rho}}^{\sqrt{\rho}} dm \, e^{-N(\mu m - s_{\mathrm{MC}}(m,\rho))} (\rho - m^2)^{-\frac{3}{2}} \langle f \rangle_{\mathrm{MC}}^{m,\rho;N},$$

where we define the auxiliary canonical partition function by

$$Z_{\mathrm{C}}(\mu, \rho; N) := \int_{-\sqrt{\rho}}^{\sqrt{\rho}} dm \, e^{-N(\mu m - s_{\mathrm{MC}}(m,\rho))} (\rho - m^2)^{-\frac{3}{2}},$$

and the specific auxiliary canonical free energy by

$$f_{\mathrm{C}}(\mu, \rho; N) := -\frac{1}{N} \ln Z_{\mathrm{C}}(\mu, \rho; N).$$

The case $h \neq 0$ is taken care of by the previous definition. We will refer to the special case of $h = 0$ as the fluctuating energy density ensemble.

Definition 4.13 (*Fluctuating energy density and fixed particle density/canonical ensemble*) Let $h = 0$. Suppose $\rho > 0$ and $\beta \in \mathbb{R}$. The canonical ensemble with inverse temperature β and particle density ρ is defined via its action on bounded 1-Lipschitz functions $f : \mathcal{S} \to \mathbb{R}$ by

$$\langle f \rangle_{\mathrm{C}}^{\beta,\rho;N} := \frac{1}{\int_{-\frac{\rho J}{2}}^{0} d\varepsilon \, e^{-\beta \varepsilon N} Z_{\mathrm{MC}}(\varepsilon, \rho; N)} \int_{-\frac{\rho J}{2}}^{0} d\varepsilon \, e^{-\beta \varepsilon N} Z_{\mathrm{MC}}(\varepsilon, \rho; N) \langle f \rangle_{\mathrm{MC}}^{\varepsilon,\rho;N}$$

$$= \frac{1}{Z_{\mathrm{C}}(\beta, \rho; N)} \int_{-\frac{\rho J}{2}}^{0} d\varepsilon \, e^{-N(\beta \varepsilon - s_{\mathrm{MC}}(\varepsilon,\rho))} \left(-\frac{2\varepsilon}{J} \right)^{-\frac{1}{2}} \left(\rho + \frac{2\varepsilon}{J} \right)^{-\frac{3}{2}} \langle f \rangle_{\mathrm{MC}}^{\varepsilon,\rho;N},$$

where the microcanonical partition function $Z_{\mathrm{MC}}(\varepsilon, \rho; N)$ is defined by

$$Z_{\mathrm{MC}}(\varepsilon, \rho; N) := \frac{\left(\rho + \frac{2\varepsilon}{J} \right)^{\frac{N-3}{2}}}{\sqrt{-\frac{2\varepsilon}{J}}},$$

the specific microcanonical entropy $s_{\mathrm{MC}}(\varepsilon, \rho)$ is defined by

$$s_{\mathrm{MC}}(\varepsilon, \rho) := \frac{1}{2} \ln \left(\rho + \frac{2\varepsilon}{J} \right),$$

and the canonical partition function $Z_{\mathrm{C}}(\beta, \rho; N)$ by

$$Z_C(\beta, \rho; N) := \int_{-\frac{\rho J}{2}}^{0} d\varepsilon \, e^{-N(\beta\varepsilon - s_{MC}(\varepsilon, \rho))} \left(-\frac{2\varepsilon}{J}\right)^{-\frac{1}{2}} \left(\rho + \frac{2\varepsilon}{J}\right)^{-\frac{3}{2}},$$

and the specific canonical free energy $f_C(\beta, \rho; N)$ by

$$f_C(\beta, \rho; N) := -\frac{1}{N} \ln Z_C(\beta, \rho; N).$$

One can verify by formal calculations that this corresponds to the typical definition of an ensemble with a fixed average constraint. We have overloaded the notation here similarly as was done in the previous section: for example, the functions $Z_C(\mu, \rho; N)$ and $Z_C(\beta, \rho; N)$ are different, but the name of the first parameter will uniquely determine to which we refer in the following. Proceeding as before, we first present the asymptotics of the derivatives of the partition function.

Theorem 4.14 *Let $\rho > 0$ and $\mu \in \mathbb{R}$. Define $\psi_{\mu,\rho} : (-\sqrt{\rho}, \sqrt{\rho}) \to \mathbb{R}$ by*

$$\psi_{\mu,\rho}(m) := \mu m - \frac{1}{2} \ln(\rho - m^2).$$

Employing the shorthand notation

$$\langle F(m) \rangle' := \frac{1}{Z_C(\mu, \rho; N)} \int_{-\sqrt{\rho}}^{\sqrt{\rho}} dm \, e^{-N\psi_{\mu,\rho}(m)} (\rho - m^2)^{-\frac{3}{2}} F(m),$$

we have

$$\left\langle \frac{M}{N} \right\rangle_C^{\mu,\rho;N} = \langle m \rangle' \quad and \quad \sigma_C^{\mu,\rho;N} \left(\frac{M}{N}\right) = \sqrt{\langle m^2 \rangle' - (\langle m \rangle')^2}.$$

Furthermore, if we fix $\rho > 0$, then for every $\mu \in \mathbb{R}$ there exists $m \in (-\sqrt{\rho}, \sqrt{\rho})$ such that $\psi_{\mu,\rho}$ is minimized at m, and, for every $m \in (-\sqrt{\rho}, \sqrt{\rho})$, there exists $\mu \in \mathbb{R}$ such that $\psi_{\mu,\rho}$ is minimized at m. The following asymptotics hold

$$\left\langle \frac{M}{N} \right\rangle_C^{\mu,\rho;N} = \mathbb{1}(\mu \neq 0) \left(\frac{1}{2\mu} - \mathrm{sgn}(\mu)\sqrt{\left(\frac{1}{2\mu}\right)^2 + \rho}\right)$$

$$+ O(N^{-\frac{1}{2}}), \quad \sigma_C^{\mu,\rho;N} \left(\frac{M}{N}\right) = O(N^{-\frac{1}{2}}).$$

Proof The first part of the theorem follows directly by differentiating the specific free energies with respect to μ and dividing by the degrees of freedom N appropriately.

Next, for fixed $\rho > 0$, we compute

$$\psi'_{\mu,\rho}(m) = \mu + \frac{m}{\rho - m^2}, \quad \psi''_{\mu,\rho}(m) = \frac{1}{\rho - m^2} + \frac{2m^2}{(\rho - m^2)^2} = \frac{\rho + m^2}{(\rho - m^2)^2} > 0.$$

It follows that the map $\psi_{\mu,\rho}$ is strictly concave for all $\mu \in \mathbb{R}$, and we can check that there is a unique global minimum at $m \in (-\sqrt{\rho}, \sqrt{\rho})$ which satisfies $\psi'_{\mu,\rho}(m) = 0$. First, if $\mu = 0$, then clearly the minimizing $m = 0$. If $\mu \neq 0$, we have

$$\psi'_{\mu,\rho}(m) = 0 \iff m = \frac{1}{2\mu} \pm \sqrt{\left(\frac{1}{2\mu}\right)^2 + \rho}.$$

Next, if $\mu > 0$, then $\frac{1}{2\mu} + \sqrt{\left(\frac{1}{2\mu}\right)^2 + \rho} > \sqrt{\rho}$, and thus the minimizing m must be

$$m = \frac{1}{2\mu} - \sqrt{\left(\frac{1}{2\mu}\right)^2 + \rho} \in (-\sqrt{\rho}, 0).$$

If $\mu < 0$, then $\frac{1}{2\mu} - \sqrt{\left(\frac{1}{2\mu}\right)^2 + \rho} = -\left(\sqrt{\left(\frac{1}{2\mu}\right)^2 + \rho} - \frac{1}{2\mu}\right) < -\sqrt{\rho}$, and thus the minimizing m must satisfy

$$m = \frac{1}{2\mu} + \sqrt{\left(\frac{1}{2\mu}\right)^2 + \rho} \in (0, \sqrt{\rho}).$$

The conclusion is that, if $|m| < \sqrt{\rho}$, then

$$\psi_{\mu,\rho}'(m) = 0 \iff m = \mathbb{1}(\mu \neq 0)\left(\frac{1}{2\mu} - \operatorname{sgn}(\mu)\sqrt{\left(\frac{1}{2\mu}\right)^2 + \rho}\right).$$

Furthermore, the above relation goes both ways. For every $\mu \in \mathbb{R}$ there exists a unique minimizing m for the above equation, and, for every $m \in (-\sqrt{\rho}, \sqrt{\rho})$, there exists $\mu \in \mathbb{R}$ such that the given m is the minimizing term. This can be seen by simply studying the given equation above and considering the limits $|\mu| \to 0$ and $|\mu| \to \infty$ and using the continuity on the open intervals $(-\infty, 0)$ and $(0, \infty)$.

The asymptotics of the average and standard deviation of magnetization density are given by the asymptotics of Laplace type integrals. We have

$$\left\langle \frac{M}{N} \right\rangle_C^{\mu,\rho;N} = \mathbb{1}(\mu \neq 0)\left(\frac{1}{2\mu} - \operatorname{sgn}(\mu)\sqrt{\left(\frac{1}{2\mu}\right)^2 + \rho}\right)$$

$$+ O(N^{-\frac{1}{2}}), \quad \sigma_C^{\mu,\rho;N}\left(\frac{M}{N}\right) = O(N^{-\frac{1}{2}}),$$

as desired. $\square$

Next, we present the asymptotics of the $h = 0$ case.

Theorem 4.15 *Let $h = 0$. Let $\rho > 0$ and $\beta \in \mathbb{R}$. Define $\psi_{\beta,\rho} : \left(-\frac{\rho J}{2}, 0\right] \to \mathbb{R}$ by*

$$\psi_{\beta,\rho}(\varepsilon) := \beta\varepsilon - \frac{1}{2}\ln\left(\rho + \frac{2\varepsilon}{J}\right).$$

Employing the shorthand notation

$$\langle F(\varepsilon)\rangle' := \frac{1}{Z_C(\beta, \rho; N)} \int_{-\frac{\rho J}{2}}^{0} d\varepsilon \, e^{-N\psi_{\beta,\rho}(\varepsilon)}\left(-\frac{2\varepsilon}{J}\right)^{-\frac{1}{2}}\left(\rho + \frac{2\varepsilon}{J}\right)^{-\frac{3}{2}} F(\varepsilon),$$

we have

$$\left\langle \frac{H}{N} \right\rangle_C^{\beta,\rho;N} = \langle\varepsilon\rangle' \quad and \quad \sigma_C^{\beta,\rho;N}\left(\frac{H}{N}\right) = \sqrt{\langle\varepsilon^2\rangle' - (\langle\varepsilon\rangle')^2}.$$

Furthermore, if we fix $\rho > 0$, then for every $\beta \geq \frac{1}{J\rho}$, there exists $\varepsilon \in \left(-\frac{\rho J}{2}, 0\right]$ such that $\psi_{\beta,\rho}$ is minimized at ε, and, for every $\varepsilon \in \left(-\frac{\rho J}{2}, 0\right]$, there exists $\beta \geq \frac{1}{J\rho}$ such that ψ_β is minimized at ε. For such β and ε, the following asymptotics holds

$$\left\langle \frac{H}{N} \right\rangle_C^{\beta,\rho;N} = -\frac{J\rho}{2}\left(1 - \frac{1}{\beta J\rho}\right) + O(N^{-\frac{1}{2}}), \quad \sigma_C^{\beta,\rho;N}\left(\frac{H}{N}\right) = O(N^{-\frac{1}{2}}). \qquad (4.2)$$

If $\beta < \frac{1}{J\rho}$, the mapping ψ_β is always minimized at 0, and the following asymptotics hold

$$\left\langle \frac{H}{N} \right\rangle_C^{\beta,\rho;N} = O(N^{-1}), \quad \sigma_C^{\beta,\rho;N}\left(\frac{H}{N}\right) = O(N^{-1}).$$

Proof The first part of this theorem follows directly by differentiating the specific free energies with respect to β and dividing by the degrees of freedom N appropriately.

Next, fix $\rho > 0$. We have

$$\psi'_{\beta,\rho}(\varepsilon) = \beta - \frac{\frac{1}{J}}{\rho + \frac{2\varepsilon}{J}}, \quad \psi''_{\beta,\rho}(\varepsilon) = \frac{2}{J^2}\frac{1}{\left(\rho + \frac{2\varepsilon}{J}\right)^2}.$$

It follows that $\psi_{\beta,\rho}$ is strictly convex and obtains a unique global minimum when $\psi'_{\beta,\rho}(\varepsilon) = 0$. Computing it from the above, we see that $\psi'_{\beta,\rho}(\varepsilon) = 0 \iff \varepsilon = -\frac{J\rho}{2}\left(1 - \frac{1}{\beta J\rho}\right)$. In particular, we see that for every $\varepsilon \in \left(-\frac{J\rho}{2}, 0\right]$ there exists $\beta \geq \frac{1}{J\rho}$ such that the given ε minimizes $\psi_{\beta,\rho}$, and, conversely, for every $\beta \geq \frac{1}{\rho J}$ there exists a minimizing value $\varepsilon \in \left(-\frac{\rho J}{2}, 0\right]$. Furthermore, if $\beta < \frac{1}{\rho J}$, then $\psi'_{\beta,\rho}$ is strictly negative on the entire interval, and, as a result $\psi_{\beta,\rho}$ is minimized for $\varepsilon = 0$.

For the asymptotics, if $\beta \geq \frac{1}{J\rho}$, then the asymptotics are standard and we have

$$\left\langle \frac{H}{N} \right\rangle_C^{\beta,\rho;N} = -\frac{J\rho}{2}\left(1 - \frac{1}{\beta J\rho}\right) + O(N^{-\frac{1}{2}}), \quad \sigma_C^{\beta,\rho;N}\left(\frac{H}{N}\right) = O(N^{-\frac{1}{2}}).$$

If $\beta = \frac{1}{J\rho}$, then we need to choose half-integer values of "α" in the Laplace method, but this will not alter the scaling of the asymptotics for the above ratios. However, if $\beta < \frac{1}{J\rho}$, then $\psi'_{\beta,\rho}(\varepsilon) < 0$ for all ε, and since then "$\mu = 1$" in the Laplace method, it follows that

$$\left\langle \frac{H}{N} \right\rangle_C^{\beta,\rho;N} = O(N^{-1}), \quad \sigma_C^{\beta,\rho;N}\left(\frac{H}{N}\right) = O(N^{-1}).$$

This completes the proof of the Theorem. $\qquad\qquad\square$

4.3 Grand Canonical Analysis

Finally, we will present the grand canonical ensemble and auxiliary grand canonical ensemble and the direct coupling method. If one considers microcanonical to be the most fundamental ensemble, this will result in substantial simplification of computation of its expectation values in the thermodynamic limit since these can now be computed using the grand canonical ensemble which is a Gaussian measure.

Definition 4.16 (*Fluctuating magnetization and particle density/auxiliary grand canonical ensemble*) Let $\mu \in \mathbb{R}$ and $\eta > 0$. The auxiliary grand canonical ensemble with magnetic potential μ and chemical potential η is defined via its action on bounded 1-Lipschitz functions $f : \mathcal{S} \to \mathbb{R}$ by

$$\langle f \rangle_{GC}^{\mu,\eta;N} := \frac{1}{\int_{\mathbb{R}^N} d\phi\, e^{-\mu M[\phi] - \eta N[\phi]}} \int_{\mathbb{R}^N} d\phi\, e^{-\mu M[\phi] - \eta N[\phi]} f(\phi).$$

The definition may be rewritten using the same parametrization of the integrals as for the auxiliary microcanonical ensemble. The result is summarized in the following Lemma.

Lemma 4.17 *Let $\mu \in \mathbb{R}$ and $\eta > 0$. We have for all bounded 1-Lipschitz functions $f : \mathcal{S} \to \mathbb{R}$*

$$\langle f \rangle_{GC}^{\mu,\eta;N} = \frac{1}{Z_{GC}(\mu, \eta; N)} \int_{-\infty}^{\infty} dz\, e^{-\eta\left(z + \frac{\mu\sqrt{N}}{2\eta}\right)^2} \int_0^{\infty} dr\, r^{N-2} e^{-\eta r^2}$$

$$\times \frac{1}{|\mathbb{S}^{N-2}|} \int_{\mathbb{S}^{N-2}} d\Omega\, \left(f \circ U^{-1}\right)(z, r\Omega),$$

where

$$Z_{GC}(\mu, \eta; N) := \int_{-\infty}^{\infty} dz\, e^{-\eta\left(z + \frac{\mu\sqrt{N}}{2\eta}\right)^2} \int_0^{\infty} dr\, r^{N-2} e^{-\eta r^2}.$$

We can now construct a direct coupling between the auxiliary microcanonical ensemble and the auxiliary grand canonical ensemble.

Theorem 4.18 *Suppose $\rho > 0$, $m \in (-\sqrt{\rho}, \sqrt{\rho})$, $\mu \in \mathbb{R}$ and $\eta > 0$ satisfy the relations*

$$m = -\frac{\mu}{2\eta}, \quad \rho = \frac{1}{2\eta} + \frac{\mu^2}{4\eta^2}.$$

Then,

$$w_2(\mu_{GC}^{\mu,\eta;N}, \mu_{MC}^{m,\rho;N}; N) \leq \frac{1}{\sqrt{\rho - m^2}} \frac{1}{\sqrt{N}},$$

which implies

$$w_2(\mu_{GC}^{\mu,\eta;N}, \mu_{MC}^{m,\rho;N}; N) = O(N^{-\frac{1}{2}}).$$

Proof Define $T : \mathbb{R} \times \mathbb{R}^{N-1} \to \mathbb{R} \times \mathbb{R}^{N-1}$ by

$$T(z, \psi) := \left(m\sqrt{N}, \sqrt{N(\rho - m^2)} \frac{\psi}{||\psi||_2}\right),$$

and set $T' := U^{-1} \circ T \circ U$. It follows that $\langle f \circ T \rangle_{GC}^{\mu,\eta;N} = \langle f \rangle_{MC}^{m,\rho;N}$, and thus T' is a transport map. Therefore, using the related coupling we find an estimate

$$w_2(\mu_{GC}^{\mu,\eta;N}, \mu_{MC}^{m,\rho;N}; N)^2$$

$$\leq \frac{1}{N} \frac{1}{\int_{-\infty}^{\infty} dz\, e^{-\eta\left(z + \frac{\mu\sqrt{N}}{2\eta}\right)^2}} \int_{-\infty}^{\infty} dz\, e^{-\eta\left(z + \frac{\mu\sqrt{N}}{2\eta}\right)^2} \left(z - m\sqrt{N}\right)^2$$

$$+ \frac{1}{N} \frac{1}{\int_{\mathbb{R}^{N-1}} d\psi\, e^{-\eta||\psi||^2}} \int_{\mathbb{R}^{N-1}} d\psi\, e^{-\eta||\psi||^2} \left|||\psi|| - \sqrt{N(\rho - m^2)}\right|^2$$

$$= \frac{1}{N} \frac{1}{\int_{-\infty}^{\infty} dz\, e^{-\frac{z^2}{2}}} \int_{-\infty}^{\infty} dz\, e^{-\frac{z^2}{2}} \left(\frac{z}{\sqrt{2\eta}} - \left(m\sqrt{N} + \frac{\mu\sqrt{N}}{2\eta} \right) \right)^2$$

$$+ \frac{1}{N} \frac{1}{\int_{\mathbb{R}^{N-1}} d\psi\, e^{-\frac{||\psi||^2}{2}}} \int_{\mathbb{R}^{N-1}} d\psi\, e^{-\frac{||\psi||^2}{2}} \left| \frac{||\psi||}{\sqrt{2\eta}} - \sqrt{N(\rho - m^2)} \right|^2.$$

We compute

$$\frac{\mu\sqrt{N}}{2\eta} = -m\sqrt{N}, \quad \frac{1}{\sqrt{2\eta}} = \sqrt{\rho - m^2} \iff \mu = -\frac{m}{\rho - m^2}, \quad \eta = \frac{1}{2(\rho - m^2)}.$$

The converse result states that

$$m = -\frac{\mu}{2\eta}, \quad \rho = \frac{1}{2\eta} + \frac{\mu^2}{4\eta^2}.$$

It follows that for every pair (m, ρ) for which the auxiliary microcanonical ensemble exists, there exists a pair (μ, η) such that the auxiliary grand canonical ensemble exists, and, the converse result holds as well. For such a pair satisfying the equations given above, we have

$$\frac{1}{N} \frac{1}{\int_{-\infty}^{\infty} dz\, e^{-\frac{z^2}{2}}} \int_{-\infty}^{\infty} dz\, e^{-\frac{z^2}{2}} \left(\frac{z}{\sqrt{2\eta}} - \left(m\sqrt{N} + \frac{\mu\sqrt{N}}{2\eta} \right) \right)^2 = \frac{1}{N} \frac{1}{2\eta} = \frac{1}{N} \frac{1}{4(\rho - m^2)},$$

and

$$\frac{1}{N} \frac{1}{\int_{\mathbb{R}^{N-1}} d\psi\, e^{-\frac{||\psi||^2}{2}}} \int_{\mathbb{R}^{N-1}} d\psi\, e^{-\frac{||\psi||^2}{2}} \left| \frac{||\psi||}{\sqrt{2\eta}} - \sqrt{N(\rho - m^2)} \right|^2$$

$$= \frac{1}{N} \frac{1}{2\eta} \frac{1}{\int_{\mathbb{R}^{N-1}} d\psi\, e^{-\frac{||\psi||^2}{2}}} \int_{\mathbb{R}^{N-1}} d\psi\, e^{-\frac{||\psi||^2}{2}} \left| ||\psi|| - \sqrt{N} \right|^2.$$

We have $||\psi||^2 - N = 1 + \sum_{i=1}^{N-1}(\psi_i^2 - 1)$, and thus

$$\left(||\psi||^2 - N \right)^2 = \sum_{i=1}^{N-1}(\psi_i^2 - 1)^2 + \sum_{i \neq j}^{N-1}(\psi_i^2 - 1)(\psi_j^2 - 1) - 2\sum_{i=1}^{N-1}(\psi_i^2 - 1) + 1.$$

Therefore,

$$\left| ||\psi|| - \sqrt{N} \right|^2 = \frac{\left(||\psi||^2 - N \right)^2}{\left(||\psi|| + \sqrt{N} \right)^2}$$

$$\leq \frac{1}{N} \left(\sum_{i=1}^{N-1}(\psi_i^2 - 1)^2 + \sum_{i \neq j}^{N-1}(\psi_i^2 - 1)(\psi_j^2 - 1) - 2\sum_{i=1}^{N-1}(\psi_i^2 - 1) + 1 \right).$$

It follows that

$$\frac{1}{\int_{\mathbb{R}^{N-1}} d\psi\, e^{-\frac{||\psi||^2}{2}}} \int_{\mathbb{R}^{N-1}} d\psi\, e^{-\frac{||\psi||^2}{2}} \left| ||\psi|| - \sqrt{N} \right|^2 \leq \frac{2N - 1}{N}.$$

Combining all the terms, we find

$$w_2(\mu_{GC}^{\mu,\eta;N}, \mu_{MC}^{m,\rho;N}; N)^2 \leq \frac{1}{N}\frac{1}{2\eta}\frac{3N-1}{N} = \frac{1}{4(\rho - m^2)}\frac{3N-1}{N^2} \leq \frac{1}{(\rho - m^2)N},$$

which implies the bound stated in the Theorem. □

If $h \neq 0$, the microcanonical energy ensemble is well-approximated by an auxiliary microcanonical magnetization ensemble whose auxiliary grand canonical theory we already covered above. For the case of $h = 0$, we consider the following grand canonical energy ensembles.

Definition 4.19 (*Fluctuating energy and particle density/grand canonical ensemble*) Suppose $h = 0$, $\mu > 0$ and $\beta < \frac{2\mu}{J}$. The grandcanonical ensemble with inverse temperature β and chemical potential μ is defined via its action on bounded 1-Lipschitz functions $f : S \to \mathbb{R}$ by

$$\langle f \rangle_{GC}^{\beta,\mu;N} := \frac{1}{\int_{\mathbb{R}^N} d\phi\, e^{-\beta H[\phi] - \mu N[\phi]}} \int_{\mathbb{R}^N} d\phi\, e^{-\beta H[\phi] - \mu N[\phi]} f(\phi).$$

The definition may be rewritten using the same parametrization of the integrals as for the microcanonical ensemble. The result is summarized in the following Lemma.

Lemma 4.20 *Let $h = 0$. Let $\mu > 0$ and $\beta < \frac{2\mu}{J}$. Then*

$$\langle f \rangle_{GC}^{\beta,\mu;N} = \frac{1}{Z_{GC}(\beta, \mu; N)} \int_{-\infty}^{\infty} dz\, e^{-\left(\mu - \frac{\beta J}{2}\right)z^2}$$
$$\int_0^{\infty} dr\, r^{N-2} e^{-\mu r^2} \frac{1}{|\mathbb{S}^{N-2}|} \int_{\mathbb{S}^{N-2}} d\Omega\, \left(f \circ U^{-1}\right)(z, r\Omega),$$

where

$$Z_{GC}(\beta, \mu; N) := \int_{-\infty}^{\infty} dz\, e^{-\left(\mu - \frac{\beta J}{2}\right)z^2} \int_0^{\infty} dr\, r^{N-2} e^{-\mu r^2}.$$

For the fixed energy density ensemble, there is only a single value of energy density for which a direct coupling can be constructed.

Theorem 4.21 *Suppose $h = 0$ and $\mu > 0$ are given. Define $\rho = \frac{1}{2\mu}$. Then, for all $\beta < \frac{2\mu}{J}$, we have*

$$w_2(\mu_{GC}^{\beta,\mu;N}, \mu_{MC}^{\varepsilon,\rho;N}|_{\varepsilon=0}; N)^2 \leq \frac{1}{N}\frac{1}{2\left(\mu - \frac{\beta J}{2}\right)} + \frac{1}{N}\frac{1}{\mu} = \frac{1}{N}\left(\frac{\rho}{1 - \rho\beta J} + 2\rho\right),$$

implying

$$w_2(\mu_{GC}^{\beta,\mu;N}, \mu_{MC}^{\varepsilon,\rho;N}|_{\varepsilon=0}; N) = O(N^{-\frac{1}{2}}).$$

Proof Let us begin by considering the more general case with $h \in \mathbb{R}$ and $\mu, \rho > 0$ arbitrary. Let m_+ and m_- be the corresponding negative and positive magnetization densities to the given ε. Define $T : \mathbb{R} \times \mathbb{R}^{N-1} \to \mathbb{R} \times \mathbb{R}^{N-1}$ by

$$T(z, \psi) := \mathbb{1}(z \geq 0)\left(m_+\sqrt{N}, \sqrt{N(\rho - m_+^2)}\frac{\psi}{||\psi||_2}\right)$$

$$+ \mathbb{1}(z < 0) \left(m_- \sqrt{N}, \sqrt{N(\rho - m_-^2)} \frac{\psi}{||\psi||_2} \right).$$

Define $T' := U^{-1} \circ T \circ U$. It follows that $\langle f \circ T' \rangle_{\text{GC}}^{\beta,\mu;N} = \langle f \rangle_{\text{MC}}^{\varepsilon,\rho;N}$, and T' is a transport map. Using the associated coupling, we find

$$w_2(\mu_{\text{GC}}^{\beta,\mu;N}, \mu_{\text{MC}}^{\varepsilon\rho;N}; N)^2 \leq \frac{1}{N} \frac{1}{\int_0^\infty dz \, e^{-\left(\mu - \frac{\beta J}{2}\right)z^2}} \int_0^\infty dz \, e^{-\left(\mu - \frac{\beta J}{2}\right)z^2} \left(z - \sqrt{-\frac{2\varepsilon}{J}} \sqrt{N} \right)^2$$

$$+ \frac{1}{N} \frac{1}{\int_{\mathbb{R}^{N-1}} d\psi \, e^{-\mu||\psi||^2}} \int_{\mathbb{R}^{N-1}} d\psi \, e^{-\mu||\psi||^2} \left| ||\psi|| - \sqrt{N \left(\rho - \left(-\frac{2\varepsilon}{J} \right) \right)} \right|^2$$

$$= \frac{1}{N} \frac{1}{\int_0^\infty dz \, e^{-\frac{z^2}{2}}} \int_0^\infty dz \, e^{-\frac{z^2}{2}} \left(\frac{z}{\sqrt{2 \left(\mu - \frac{\beta J}{2} \right)}} - \sqrt{-\frac{2\varepsilon}{J}} \sqrt{N} \right)^2$$

$$+ \frac{1}{N} \frac{1}{\int_{\mathbb{R}^{N-1}} d\psi \, e^{-\frac{||\psi||^2}{2}}} \int_{\mathbb{R}^{N-1}} d\psi \, e^{-\frac{||\psi||^2}{2}} \left| \frac{||\psi||}{\sqrt{2\mu}} - \sqrt{N \left(\rho - \left(-\frac{2\varepsilon}{J} \right) \right)} \right|^2.$$

Note that for $\varepsilon \neq 0$, we have

$$\frac{1}{N} \frac{1}{\int_0^\infty dz \, e^{-\frac{z^2}{2}}} \int_0^\infty dz \, e^{-\frac{z^2}{2}} \left(\frac{z}{\sqrt{2 \left(\mu - \frac{\beta J}{2} \right)}} - \sqrt{-\frac{2\varepsilon}{J}} \sqrt{N} \right)^2 \sim -\frac{2\varepsilon}{J},$$

which does not provide any additional convergence for the local expectation error estimates. However, under the assumptions listed in the Theorem, i.e., if $h = 0 = \varepsilon$, $\mu > 0$, $\rho = \frac{1}{2\mu}$, we find via the same computation as above that

$$w_2(\mu_{\text{GC}}^{\beta,\mu;N}, \mu_{\text{MC}}^{\varepsilon,\rho;N}; N)^2 \leq \frac{1}{N} \frac{1}{2 \left(\mu - \frac{\beta J}{2} \right)} + \frac{1}{N} \frac{1}{2\mu} 2 = \frac{1}{N} \left(\frac{\rho}{\rho - \beta J} + \rho \right).$$

Note that the above holds for all $\beta < \frac{2\mu}{J} \iff \beta < \frac{1}{\rho J}$. $\qquad\square$

For the cases $\beta \geq \frac{1}{\rho J}$, we must introduce another class of auxiliary measures.

Definition 4.22 Let $\overline{\mu} \geq 0$ and $\eta > 0$. We define an alternate auxiliary grand canonical ensemble with parameters $\overline{\mu}$ and η via its action on bounded 1-Lipschitz functions $f : \mathcal{S} \to \mathbb{R}$ by

$$\langle f \rangle_{\text{AGC}}^{\overline{\mu},\eta;N} := \frac{1}{Z_{\text{AGC}}(\overline{\mu}, \eta; N)} \int_{\mathcal{S}} d\phi \, e^{-\eta||\phi||^2} \cosh(\overline{\mu} M[\phi]) f(\phi),$$

where

$$Z_{\text{AGC}}(\overline{\mu}, \eta; N) = \int_{\mathcal{S}} d\phi \, e^{-\eta||\phi||^2} \cosh(\overline{\mu} M[\phi]).$$

By direct computation, we also have then

$$\langle f \rangle_{\mathrm{AGC}}^{\overline{\mu},\eta;N} = \frac{1}{2} \langle f \rangle_{\mathrm{GC}}^{\overline{\mu},\eta;N} + \frac{1}{2} \langle f \rangle_{\mathrm{GC}}^{-\overline{\mu},\eta;N}.$$

Comparing this to the definition of the microcanonical ensemble, we find using Lemma 4.18 that if $h = 0$ and $-\frac{\rho J}{2} < \varepsilon < 0$, then, with $\overline{\mu} = \sqrt{\frac{-2\varepsilon}{J}} \frac{1}{\rho + \frac{2\varepsilon}{J}}$ and $\eta = \frac{J}{2(J\rho+2\varepsilon)}$,

$$w_2(\mu_{\mathrm{AGC}}^{\overline{\mu},\eta;N}, \mu_{\mathrm{MC}}^{\varepsilon,\rho;N}; N) = O(N^{-\frac{1}{2}}).$$

One should note that there is no direct coupling of this new alternate auxiliary grand canonical ensemble to the microcanonical ensemble because the probability measures are not disjoint. However, the individual grand canonical ensembles do converge suitably to the fixed magnetization density case, and thus we still have the desired local convergence properties. We also remark that the case $\overline{\mu} = 0$ corresponds to the regular grand canonical ensemble given by a Gaussian measure with $\beta = 0$.

4.4 Convergence of Finite Marginal Distributions and Finite Moments

In this subsection, we will collect and apply the upper bounds and error estimates presented for the continuum model to formulate the main local convergence theorems. Since our main goal is to prove convergence theorems and error estimates for local observables of the mean-field spherical microcanonical ensemble, we will present a variety of target measures which the local observables can converge to. Some of these target measures will come from thermodynamic ensembles and others from auxiliary ensembles. In particular, the value of the constant h will have a significant impact on the choice of target measure.

The main theorems concerning the convergence of moments required the boundedness of single moments of all degrees. To this end, we will employ the following lemma.

Lemma 4.23 *Let $f : \mathbb{R} \to \mathbb{R}$ be integrable with respect to any Gaussian measure. Let $x \in \Lambda$ and define $P_x : \mathcal{S} \to \mathbb{R}$ by $P_x(\phi) = \phi_x$. It follows that for all $\rho > 0$ and $m \in (-\sqrt{\rho}, \sqrt{\rho})$, we have*

$$\langle f \circ P_x \rangle_{MC}^{m,\rho;N} = O(1).$$

Proof A direct calculation using the delta function definition of the measures shows that

$$\langle f \circ P_x \rangle_{\mathrm{MC}}^{m,\rho;N} = \frac{1}{C(m,\rho;N)} \int_{-\infty}^{\infty} d\phi_x \, f(\phi_x) \mathbb{1}$$

$$\times \left((\phi_x - m)^2 \le (\rho - m^2)(N-1) \right) \left(1 - \frac{(\phi_x - m)^2}{(\rho - m^2)(N-1)} \right)^{\frac{N-4}{2}},$$

where

$$C(m,\rho;N) := \int_{-\infty}^{\infty} d\phi_x \, \mathbb{1} \left((\phi_x - m)^2 \le (\rho - m^2)(N-1) \right) \left(1 - \frac{(\phi_x - m)^2}{(\rho - m^2)(N-1)} \right)^{\frac{N-4}{2}}.$$

For $N \ge 5$, we have

$$\mathbb{1} \left((\phi_x - m)^2 \le (\rho - m^2)(N-1) \right) \left(1 - \frac{(\phi_x - m)^2}{(\rho - m^2)(N-1)} \right)^{\frac{N-4}{2}}$$

$$\le \mathbb{1} \left((\phi_x - m)^2 \le (\rho - m^2)(N-1) \right) e^{-\frac{(\phi_x - m)^2}{2(\rho - m^2)} \frac{N-4}{N-1}}$$

$$\leq e^{-\frac{(\phi_x - m)^2}{8(\rho - m^2)}}.$$

By the dominated convergence theorem, using the assumed integrability of f, we have

$$\lim_{N \to \infty} \langle f \circ P_x \rangle_{\text{MC}}^{m,\rho;N} = \frac{1}{\int_{-\infty}^{\infty} d\phi_x \, e^{-\frac{(\phi_x - m)^2}{2(\rho - m^2)}}} \int_{-\infty}^{\infty} d\phi_x \, f(\phi_x) e^{-\frac{(\phi_x - m)^2}{2(\rho - m^2)}} < \infty,$$

which implies that

$$\langle f \circ P_x \rangle_{\text{MC}}^{m,\rho;N} = O(1).$$

$\square$

Lemma 4.24 *Let $f : \mathbb{R} \to \mathbb{R}$ be integrable with respect to any Gaussian measure. Let $x \in \Lambda$ and define $P_x : \mathcal{S} \to \mathbb{R}$ by $P_x(\phi) = \phi_x$. Then for all $\rho > 0$, $h \in \mathbb{R}$, and $\varepsilon \in \mathcal{E}_{h,\rho}$, we have*

$$\langle f \circ P_x \rangle_{MC}^{\varepsilon,\rho;N} = O(1).$$

Proof By definition, $\langle f \circ P_x \rangle_{\text{MC}}^{\varepsilon,\rho;N}$ is either equal to one of the expectations studied in the previous Lemma, or it is a convex combination of $\langle f \circ P_x \rangle_{\text{MC}}^{m_+,\rho;N}$ and $\langle f \circ P_x \rangle_{\text{MC}}^{m_-,\rho;N}$. In both of these cases, the result remains bounded as $N \to \infty$. $\square$

As we remarked earlier, the phase transitions in the mean-field spherical model result in the need for different limiting measures outside of the standard ensembles when using the coupling method. Of particular importance is the parameter $h \in \mathbb{R}$. In the following convergence results, we will always explicitly state for which different parameters and limiting measures the convergence results hold.

First, we will state the convergence result for the auxiliary microcanonical, auxiliary canonical, microcanonical, and canonical ensembles.

Theorem 4.25 *Let $\rho > 0$ and $m \in (-\sqrt{\rho}, \sqrt{\rho})$ and define $\mu := -\frac{m}{\rho - m^2}$. Let $I \subset \Lambda$ be a fixed size index set, and let $f : \mathbb{R}^{|I|} \to \mathbb{R}$ be a bounded 1-Lipschitz function with respect to the $||\cdot||_2$-norm. It follows that*

$$\langle f \circ P_I \rangle_{MC}^{m,\rho;N} = \langle f \circ P_I \rangle_{C}^{\mu,\rho;N} + O(N^{-\frac{1}{2}}).$$

$\square$

Proof The result follows by applying the free energy coupling presented in Theorem 2.5, along with the w_2 bound presented in Theorem 4.7, and with the asymptotics presented in Theorem 4.14. $\square$

Theorem 4.26 *Suppose $h \neq 0$ and $\rho > 0$. Assume $\varepsilon \in \mathcal{E}_{h,\rho}$. Define $m := -\frac{h}{J} + \text{sgn}(h)\sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}$ and $\mu := -\frac{m}{\rho - m^2}$. Let $I \subset \Lambda$ be a fixed size index set, and let $f : \mathbb{R}^{|I|} \to \mathbb{R}$ be a bounded 1-Lipschitz function with respect to the $||\cdot||_2$-norm. It follows that*

$$\langle f \circ P_I \rangle_{MC}^{\varepsilon,\rho;N} = \langle f \circ P_I \rangle_{C}^{\mu,\rho;N} + O(N^{-\frac{1}{2}}).$$

Proof By Theorem 4.10, we have

$$\langle f \circ P_I \rangle_{MC}^{\varepsilon,\rho;N} = \langle f \circ P_I \rangle_{MC}^{m,\rho;N} + O(e^{-cN}),$$

and, by Theorem 4.25, we have

$$\langle f \circ P_I \rangle_{MC}^{m,\rho;N} = \langle f \circ P_I \rangle_{C}^{\mu,\rho;N} + O(N^{-\frac{1}{2}}),$$

from which the result follows. $\qquad\square$

Theorem 4.27 *Let $h = 0$. Suppose $\rho > 0$ and $\varepsilon \in \left(-\frac{J\rho}{2}, 0\right]$. Let $I \subset \Lambda$ be a fixed size index set, and let $f : \mathbb{R}^{|I|} \to \mathbb{R}$ be a bounded 1-Lipschitz function with respect to the $|| \cdot ||_2$-norm. For $\varepsilon \in \left(-\frac{J\rho}{2}, 0\right)$, define $\beta := \frac{\frac{1}{J}}{\rho + \frac{2\varepsilon}{J}}$. It follows that*

$$\langle f \circ P_I \rangle_{MC}^{\varepsilon,\rho;N} = \langle f \circ P_I \rangle_{C}^{\beta,\rho;N} + O(N^{-\frac{1}{2}}).$$

For $\varepsilon = 0$, let $\beta < \frac{1}{\rho J}$ be arbitrary. It follows that

$$\langle f \circ P_I \rangle_{MC}^{0,\rho;N} = \langle f \circ P_I \rangle_{C}^{\beta,\rho;N} + O(N^{-\frac{1}{2}}).$$

Proof If $\varepsilon \in \left(-\frac{J\rho}{2}, 0\right)$, the result follows by applying the free energy coupling presented in Theorem 2.5, along with the w_2 bound presented in Theorem 4.11, and the asymptotics presented in Theorem 4.15.

If $\varepsilon = 0$, then observe that the w_2 bound in Theorem 4.11 is not Lipschitz in the appropriate sense to directly apply Theorem 2.5. However, following the proof of Theorem 2.5, we can apply the following inequality

$$\left\langle w_2 \left(\mu_{MC}^{0,\rho;N}, \mu_{MC}^{\frac{H}{N},\rho;N}; N \right) \right\rangle_{C}^{\beta,\rho;N} \leq \frac{2}{\sqrt{J}} \left(\langle -H/N \rangle_{C}^{\beta,\rho;N} \right)^{\frac{1}{2}},$$

where the upper-index $\frac{H}{N}$ is a non-positive random variable of the canonical ensemble. It follows that

$$\left| \langle f \circ P_I \rangle_{MC}^{0,\rho;N} - \langle f \circ P_I \rangle_{C}^{\beta,\rho;N} \right| \leq C \left(\langle -H/N \rangle_{C}^{\beta,\rho;N} \right)^{\frac{1}{2}},$$

for a global constant $C > 0$. By considering the asymptotics presented in Theorem 4.15, we have

$$\langle f \circ P_I \rangle_{MC}^{0,\rho;N} = \langle f \circ P_I \rangle_{C}^{\beta,\rho;N} + O(N^{-\frac{1}{2}}).$$

$\qquad\square$

Finally, we will state the convergence result for the auxiliary microcanonical, auxiliary grand canonical, alternate auxiliary grand canonical, microcanonical, and grand canonical ensembles.

Theorem 4.28 *Let $\rho > 0$ and $m \in (-\sqrt{\rho}, \sqrt{\rho})$ and define $\mu := -\frac{m}{\rho - m^2}$ and $\eta := \frac{1}{2(\rho - m^2)}$. Let $I \subset \Lambda$ be a fixed size index set, and let $f : \mathbb{R}^{|I|} \to \mathbb{R}$ be a bounded 1-Lipschitz function with respect to the $|| \cdot ||_2$-norm or let f be a finite product of finite order moments. It follows that*

$$\langle f \circ P_I \rangle_{MC}^{m,\rho;N} = \langle f \circ P_I \rangle_{GC}^{\mu,\eta;N} + O(N^{-\frac{1}{2}}).$$

Proof The result follows by applying the direct coupling method in Lemma 2.3 along with the w_2 bound given in Theorem 4.18. The convergence of the finite dimensional moments follows from Theorem 2.4 and the fact that the moments of both relevant ensembles are bounded by Lemma 4.23. $\qquad\square$

Theorem 4.29 *Suppose $h \neq 0$ and $\rho > 0$. Assume $\varepsilon \in \mathcal{E}_{h,\rho}$. Define $m := -\frac{h}{J} +$ $\mathrm{sgn}(h)\sqrt{\frac{h^2}{J^2} - \frac{2\varepsilon}{J}}$ and set then $\mu := -\frac{m}{\rho - m^2}$ and $\eta := \frac{1}{2(\rho - m^2)}$. Let $I \subset \Lambda$ be a fixed size index set, and let $f : \mathbb{R}^{|I|} \to \mathbb{R}$ be a bounded 1-Lipschitz function with respect to the $|| \cdot ||_2$-norm or let f be a finite product of finite order moments. It follows that*

$$\langle f \circ P_I \rangle_{MC}^{\varepsilon,\rho;N} = \langle f \circ P_I \rangle_{GC}^{\mu,\eta;N} + O(N^{-\frac{1}{2}}).$$

Proof The proof follows from Theorem 4.28 via the same steps as in the proof of Theorem 4.26. □

Theorem 4.30 *Suppose $h = 0$ and $\rho > 0$. Assume $\varepsilon \in \left(-\frac{J\rho}{2}, 0\right]$ and define $\overline{\mu} := \frac{\sqrt{-\frac{2\varepsilon}{J}}}{\rho - \left(-\frac{2\varepsilon}{J}\right)}$ and $\eta := \frac{1}{2\left(\rho - \left(-\frac{2\varepsilon}{J}\right)\right)}$. Let $I \subset \Lambda$ be a fixed size index set, and let $f : \mathbb{R}^{|I|} \to \mathbb{R}$ be a bounded 1-Lipschitz function with respect to the $|| \cdot ||_2$-norm or a finite product of finite order moments. It follows that*

$$\langle f \circ P_I \rangle_{MC}^{\varepsilon,\rho;N} = \langle f \circ P_I \rangle_{AGC}^{\overline{\mu},\eta;N} + O(N^{-\frac{1}{2}}).$$

Proof The result follows by splitting the fixed energy density ensemble into its fixed magnetization ensity ensembles and applying Theorem 4.28. □

4.5 Remark on Choice of Cost Function

For this model, it should be observed that the w_2 convergence is a natural choice of convergence from the perspective that it implies both w_1 and w_2 convergence simultaneously, which in turn implies that the magnetization density converges along with the energy density. Without this property, a coupling of suitable strength between the microcanonical and grand canonical measures seems unlikely. Observe that

$$w_1\left(\mu_{MC}^{m,\rho;N}, \mu_{MC}^{m',\rho;N}; N\right) \leq w_2\left(\mu_{MC}^{m,\rho;N}, \mu_{MC}^{m',\rho;N}; N\right).$$

Employing the lower bound in the triangle inequality, we have

$$w_1\left(\mu_{MC}^{m,\rho;N}, \mu_{MC}^{m',\rho;N}\right) \geq |m - m'|.$$

Now, if we consider Theorem 4.7, then we have

$$|m - m'| \leq w_2\left(\mu_{MC}^{m,\rho;N}, \mu_{MC}^{m',\rho;N}; N\right) \leq \left(1 + \frac{2}{\sqrt{1 - m^2/\rho}}\right)|m - m'|.$$

Thus, even though optimality of the transport was not necessarily achieved as in the discrete case, the best possible scaling in the dependence on changes in the parameter m was still obtained here.

A Rigorous Asymptotic Analysis of Laplace-Type Integrals

First, we will fix some notation and definitions.

Definition A.1 (*Asymptotic equivalence*) Let $f, g : \mathbb{R} \to \mathbb{R}$ be suitable functions so that the following limits and quotients exist. We say that f and g are asymptotically equivalent at $a \in \overline{\mathbb{R}} := [-\infty, \infty]$ if

$$\lim_{x \to a} \frac{f(x)}{g(x)} = 1.$$

Asymptotic equivalence will be denoted $f \sim g$ without reference to the limiting point a if it is clear from context.

Furthermore, we say that a function f admits an asymptotic power series representation at a point $a \in \mathbb{R}$ if there is a sequence of constants $(a_k)_{k \in \mathbb{N}_0}$ and some $\mu \geq 0$ such that

$$\lim_{x \to a} (x - a)^{-N-\mu} \left(f(x) - \sum_{k=0}^{N} a_k (x - a)^{k+\mu} \right) = 0$$

for all $N \in \mathbb{N}_0$. Since this implies that $f(x) - \sum_{k=0}^{N-1} a_k (x - a)^{k+\mu} \sim a_N (x - a)^{N+\mu}$ whenever $a_N \neq 0$, we will use the notation

$$f(x) \sim \sum_{k=0}^{\infty} a_k (x - a)^{k+\mu}$$

to denote the above without reference to N, even if the power series on the right does not converge.

Analogously, we define asymptotic power series representation as $x \to \infty$ by requiring that $x \mapsto f(1/x)$, $x > 0$, has an asymptotic power series at 0. Explicitly, we then require

$$\lim_{x \to \infty} x^{N+\mu} \left(f(x) - \sum_{k=0}^{N} a_k x^{-k-\mu} \right) = 0 \,,$$

for all $N \in \mathbb{N}_0$, and denote this by

$$f(x) \sim \sum_{k=0}^{\infty} a_k x^{-k-\mu}$$

The asymptotic analysis of Laplace-type integrals has been studied extensively. For completeness, we will present below a general form of the asymptotics of Laplace-type integrals.

Theorem A.2 *Let $h : [a, b] \to \mathbb{R}$ be a function satisfying the following conditions:*

- *h attains a global minimum at the left end-point a, $h(x) > h(a)$ for all $x \in (a, b)$, and for every $\delta > 0$ we have $\inf_{x \in [a+\delta, b)} \{h(x) - h(a)\} > 0$.*
- *h admits a power series representation at the left end-point a of the form*

$$h(x) \sim h(a) + \sum_{s=0}^{\infty} a_s (x - a)^{s+\mu}$$

for some $\mu > 0$ and with $a_0 \neq 0$.

- *h is differentiable in a neighbourhood of a and the previous power series representation can be term-wise differentiated to give*

$$h'(x) \sim \sum_{s=0}^{\infty} a_s (s + \mu)(x - a)^{s+\mu-1}.$$

- *h' is continuous in a neighbourhood of a except possibly at a.*

Suppose also that $\varphi : [a, b] \to \mathbb{R}$ is a function satisfying all of the following:

- *φ is continuous in a neighbourhood of a except possibly at a.*
- *φ admits a power series representation at the left end-point a of the form*

$$\varphi(x) \sim \sum_{s=0}^{\infty} b_s (x - a)^{s+\alpha-1}$$

for some $\alpha \in \mathbb{C}$ such that $\operatorname{Re} \alpha > 0$, and with $b_0 \neq 0$.

Furthermore, suppose that there exists $M > 0$ such that for all $\lambda \geq M$, the integral $I(\lambda)$ defined by

$$I(\lambda) := \int_a^b dx \, \varphi(x) e^{-\lambda h(x)},$$

converges absolutely.

Then, as $\lambda \to \infty$,

$$I(\lambda) \sim e^{-\lambda h(a)} \sum_{s=0}^{\infty} \Gamma\left(\frac{s + \alpha}{\mu}\right) \frac{c_s}{\lambda^{\frac{s+\alpha}{\mu}}},$$

where the coefficients c_s are expressible in terms of a_s and b_s, and, in particular, we have

$$c_0 = \frac{b_0}{\mu a_0^{\frac{\alpha}{\mu}}}.$$

Proof The proof is given [18], chapter 2 "Classical Procedures", section 1 "Laplace's method". □

The previous theorem can be applied to all the Laplace-type integrals that will be used in this paper. To be explicit, the most typical usage of this theorem will be for the case where $h : [a, c] \to \mathbb{R}$ is a twice continuously differentiable strictly convex function, which implies that $h''(x) > 0$ for all $x \in [a, c]$. If there exists $b \in (a, c)$ such that $h'(b) = 0$, then this point b is the global minimum of h and one can consider the function h on the intervals $[a, b]$ and $[b, c]$. Note that the previous theorem holds precisely for h on the interval $[b, c]$ since h attains its global minimum at the left-end point b. For the interval $[a, b]$, one instead considers the mapping $\tilde{h}(x) := h(-x)$ defined on the interval $[-b, -a]$. One finds that the

mapping $\tilde{h}$ obtains a global minimum at $-b$ and again the contents of the previous theorem hold.

The role of the mapping $\varphi : [a, c] \to \mathbb{R}$ does not change. In particular, if φ admits a power series representation at any point on this interval, then it necessarily also admits power series representations when using one sided limits.

In the case of a strictly convex h, at the global minimum b, we have $\mu = 2$ and $a_0 = \frac{1}{2} h''(b) \neq 0$. The mapping φ is of more importance. In particular, suppose that φ is a smooth function such that for some finite $i \in \mathbb{N}$ and for all $k < i$, we have $\varphi^{(k)}(b) = 0$ and $\varphi^{(i)}(b) \neq 0$. In the notation of the previous theorem, this would correspond to the situation where $\alpha = i + 1$ and $b_0 = \frac{1}{i!} \varphi^{(i)}(b)$. Applying the previous theorem, we then would have

$$\int_a^c dx \, \varphi(x) e^{-\lambda h(x)} \sim e^{-\lambda h(b)} \Gamma\left(\frac{i+1}{2}\right) \frac{1}{i!} \varphi^{(i)}(b) \frac{1}{2 \left(\frac{1}{2} h''(b)\right)^{\frac{i+1}{2}}} \frac{1}{\lambda^{\frac{i+1}{2}}},$$

and

$$\frac{\int_a^c dx \, \varphi(x) e^{-\lambda h(x)}}{\int_a^c dx \, e^{-\lambda h(x)}} \sim \frac{\Gamma\left(\frac{i+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \frac{1}{i!} \varphi^{(i)}(b) \frac{1}{\left(\frac{1}{2} h''(b)\right)^{\frac{i}{2}}} \frac{1}{\lambda^{\frac{i}{2}}}.$$

The primary message from this is that the order of the first non-zero derivative of φ determines the rate of vanishing of such Laplace-type integrals, in particular, in this case we would have

$$\frac{\int_a^c dx \, \varphi(x) e^{-\lambda h(x)}}{\int_a^c dx \, e^{-\lambda h(x)}} = O(\lambda^{-\frac{i}{2}}), \quad \lambda \to \infty.$$

Furthermore, there will be some cases in which $h'(x) \neq 0$ for any $x \in (a, c)$. In such a situation, it will also typically be so that $h'(x) < 0$ for all $x \in (a, c)$, this implies that h is minimized at the right end point c, and, by considering the mapping $\tilde{h}(x) = h(-x)$ again, we see that the mapping $\tilde{h}$ is now minimized at its left end point $-c$ and thus the theorems hold again. In such a situation, we have $\mu = 1$ instead of $\mu = 2$.

References

1. Berlin, T.H., Kac, M.: The spherical model of a ferromagnet. Phys. Rev. **86**, 821–835 (1952). https://doi.org/10.1103/PhysRev.86.821
2. Chatterjee, S.: A note about the uniform distribution on the intersection of a simplex and a sphere. J. Topol. Anal. **09**(04), 717–738 (2017). https://doi.org/10.1142/s1793525317500224
3. Chleboun, P., Großkinsky, S.: Condensation in stochastic particle systems with stationary product measures. J. Stat. Phys. **154**(1–2), 432–465 (2013)
4. Costeniuc, M., Ellis, R.S., Touchette, H.: Complete analysis of phase transitions and ensemble equivalence for the Curie–Weiss–Potts model. J. Math. Phys. **46**(6), 063301 (2005)
5. Csiszar, I.: I-divergence geometry of probability distributions and minimization problems. Ann. Probab. **3**(1), 146–158, 02 (1975)
6. Ellis, R.S., Newman, C.M.: The statistics of Curie–Weiss models. J. Stat. Phys. **19**(2), 149–161 (1978). https://doi.org/10.1007/BF01012508
7. Großkinsky, S.: Equivalence of ensembles for two-species zero-range invariant measures. Stoch. Process. Appl. **118**(8), 1322–1350 (2008)
8. Huveneers, F., Theil, E.: Equivalence of ensembles, condensation and glassy dynamics in the Bose–Hubbard hamiltonian. J. Stat. Phys. **177**(5), 917–935 (2019). https://doi.org/10.1007/s10955-019-02396-z
9. Kastner, M., Schnetz, O.: On the mean-field spherical model. J. Stat. Phys. **122**(6), 1195–1214 (2006). https://doi.org/10.1007/s10955-005-8031-9

10. Kochmański, M., Paszkiewicz, T., Wolski, S.: Curie–Weiss magnet: a simple model of phase transition. Eur. J. Phys. **34**(6), 1555–1573 (2013). https://doi.org/10.1088/0143-0807/34/6/1555
11. Lewis, J., Pfister, C.-E., Sullivan, W.: The equivalence of ensembles for lattice systems: some examples and a counterexample. J. Stat. Phys. **77**, 10 (1994)
12. Lewis, J., Pfister, C.-E., Sullivan, W.: Entropy, concentration of probability and conditional limit theorems. Markov Process. Relat. Fields **1**(3), 319–386 (1995)
13. Lukkarinen, J.: Multi-state condensation in Berlin–Kac spherical models. Commun. Math. Phys. **373**(1), 389–433 (2019)
14. Nam, K.: Large deviations and localization of the microcanonical ensembles given by multiple constraints. ArXiv preprint, 41 pp. (2018). arXiv:1809.04138
15. Rudin, W.: Real and Complex Analysis. Mathematics Series. McGraw-Hill, London (1987)
16. Touchette, H.: Equivalence and nonequivalence of ensembles: thermodynamic, macrostate, and measure levels. J. Stat. Phys. **159**(5), 987–1016 (2015). https://doi.org/10.1007/s10955-015-1212-2
17. Villani, C.: Optimal Transport. Springer Berlin Heidelberg (2009). https://doi.org/10.1007/978-3-540-71050-9
18. Wong, R.: Asymptotic Approximations of Integrals. Elsevier, Amsterdam (1989). https://doi.org/10.1016/c2013-0-07651-7