HOPFIELD MODELS AS GENERALIZED RANDOM MEAN FIELD MODELS

Anton Bovier and Véronique Gayrard

L'intuition ne peut nous donner la rigeur,
ni même la certitude, on s'en est aperçu de plus en plus.
Henri Poincaré,
"La Valeur de La Science"

I. Introduction

Twenty years ago, Pastur and Figotin [FP1,FP2] first introduced and studied what has become known to be the Hopfield model and which turned out, over the years, as one of the more successful and important models of a disordered system. This is also reflected in the fact that several contributions in this book are devoted to it. The Hopfield model is quite versatile and models various situations: Pastur and Figotin introduced it as a simple model for a spin glass, while Hopfield, in 1982, independently considered it as a model for associative memory. The first viewpoint naturally put it in the context of equilibrium statistical mechanics, while Hopfield's main interest was its dynamics. But the great success of what became known as the Hopfield model came from the realization, mainly in the work of Amit, Gutfreund, and Sompolinsky [AGS] that a more complicated version of this model is reminiscent to a spin glass, and that the (then) recently developed methods of spin-glass theory, in particular the replica trick and Parisi's replica symmetry breaking scheme could be adapted to this model and allowed a "complete" analysis of the equilibrium statistical mechanics of the model and to recover some of the most prominent "experimentally" observed features of the model like the "storage capacity", and "loss of memory" in a precise analytical way. This observation sparked a surge of interest by theoretical physicists into neural network theory in general that has led to considerable progress in the field (the literature on the subject is extremely rich, and there are a great number of good review papers. See for example [A,HKP,GM,MR,DHS]). We will

Work partially supported by the Commission of the European Union under contract CHRX-CT93-0411

not review this development here. In spite of their success, the methods used in the analysis by theoretical physicist were of heuristic nature and involved mathematically unjustified procedures and it may not be too unfair to say that they do not really provide a deeper understanding for what is really going on in these systems. Mathematicians and mathematical physicists were only late entering this field; as a matter of fact, spin glass theory was (and is) considered a field difficult, if not impossible, to access by rigorous mathematical techniques.

As is demonstrated in this book, in the course of the last decade the attitude of at least some mathematicians and mathematical physicists towards this field has changed, and some now consider it as a major challenge to be faced rather than a nuisance to be avoided. And already, substantial progress in a rigorous mathematical sense has begun to be made. The Hopfield model has been for us the focal point of attention in this respect over the last five years and in this article we will review the results obtained by us in this spirit. Our approach to the model may be called "generalized random mean field models", and is in spirit close to large deviation theory. We will give a precise outlay of this general setting in the next section. Historically, our basic approach can be traced back even to the original papers by Pastur and Figotin. In this setting, the "number of patterns", M, or rather its relation to the system size N, is a crucial parameter and the larger it is, the more difficult things are getting. The case where M is is strictly bounded could be termed "standard disordered mean field", and it is this type of models that were studied by Pastur and Figotin in 1977, the case of two patterns having been introduced by Luttinger [Lut] shortly before that. Such "site-disorder" models were studied again intensely some years later by a number of people, emphasizing applications of large deviation methods [vHvEC,vH1,GK,vHGHK,vH2,AGS2,JK,vEvHP]. A general large deviation theory for such systems was obtained by Comets [Co] somewhat later. This was far from the "physically" interesting case where the ratio between M and N, traditionally called α, is a finite positive number [Ho, AGS]. The approach of Grensing and Kühn [GK], that could be described as the most straightforward generalization of the large deviation analysis of the Curie-Weiss model by combinatorial computation of the entropy (see Ellis' book [El] for a detailed exposition), was the first to be generalized to unbounded M by Koch and Piasko [KP] (but see also [vHvE]). Although their condition on M, namely $M < \frac{\ln N}{\ln 2}$, was quite strong, until 1992 this remained the only rigorous result on the thermodynamics of the model with an unbounded number of patterns and their analysis involved for the first time a non-trivial control on fluctuations of a free energy functional. Within their framework, however, the barrier $\ln N$ appeared unsurmountable, and some crucial new ideas were needed. They came in two almost simultaneous papers by Shcherbina and Tirozzi [ST] and Koch [K]. They proved that the free energy of the Hopfield model in the thermodynamic limit is equal to that of the Curie-Weiss model, provided only that $\lim_{N\uparrow\infty} \frac{M}{N} = 0$, without condition on the speed of convergence. In their proof this fact was linked to the convergence in norm of a certain random matrix constructed from the patterns to the identity matrix. Control on this matrix proved one key element in further progress. Building on this observation, in a paper with Picco [BGP1] we were able to give a construction of the extremal Gibbs states under the same hypothesis, and even get first results on the Gibbs states in the case $\frac{M}{N} = \alpha \ll 1$. Further progress in this latter case, however, required yet another key idea: the use of exponential concentration of measure estimates. Variance estimates based on the Yurinskii martingale construction had already appeared in [ST] where they were used to prove self-averaging of the free energy. With Picco [BGP3] we proved exponential estimates on "local" free energies and used this to show that disjoint Gibbs states corresponding to all patterns can be constructed for small enough α. A considerable refinement of this analysis that included a detailed analysis of the local minima near the Mattis states [Ma] was given in a later paper by the present authors [BG5]. The result is a fairly complete and rigorous picture of the Gibbs states and even metastable states in the small α regime, which is in good agreement with the heuristic results of [AGS]. During the preparation of this manuscript, a remarkable breakthrough was obtained by Michel Talagrand [T4]. He succeeded in proving that in a certain (nontrivial) range of the parameters β and α, the validity of the "replica symmetric solution" of [AGS] can be rigorously justified. It turns out that a result obtained in [BG5] can be used to give an alternative proof of that also yields some complementary information and in particular allows to analyse the convergence properties of the Gibbs measures in that regime. We find it particularly pleasant that, 10 years after the paper by Amit et al., we can present this development in this review.

In the present paper we will give a fairly complete and streamlined version of our approach, emphasizing generalizations beyond the standard Hopfield model, even though we will not work out all the details at every point. We have tried to give proofs that are either simpler or more systematic than the original ones and believe to have succeeded to some extent. At some places technical proofs that we were not able to improve substantially are omitted and reference is made to the original papers. In Section 2 we present a derivation of the Hopfield model as a mean field spin glass, introduce the concept of *generalized random mean field models* and discuss the thermodynamic formalism for such systems. We point out some popular variants of the Hopfield model and place them in this general framework. Section 3 discusses some necessary background on large deviations, emphasizing calculational aspects. This section is quite general and can be regarded as completely independent from particular models. Section 4 brings the last proof on exponential estimates on maximal and minimal eigenvalues of some matrices that are used throughout in the sequel. In Section 5 we show how large deviation estimates lead to estimates on Gibbs measures. Here the theme of *concentration of measure* appears in a crucial way. Section 6 as well as Section 7 are devoted to the study of the function Φ that emerged from Section 3 as a crucial instrument to control large deviations. Section 8, finally gives a rigorous derivation of the replica symmetric solution of [AGS] in an appropriate range of parameters, and the comstruction of the limiting distribution of the Gibbs measures (the "metastate" in the language of [NS]).

There are a number of other results on the Hopfield model that we do not discuss. We never talk here about the high temperature phase, and we also exclude the study of the zero temperature case. Also we do not speak about the case $\alpha = 0$ but will always assume $\alpha > 0$. However, all proofs work also when $\frac{M}{N} \downarrow 0$, with some trivial modifications necessary when $M(N)$ remains bounded or grows slowly. In this situation some more refined results, like large deviation principles [BG4] and central limit theorems [G1] can be obtained. Such results will be covered in other contributions to this volume.

Acknowledgements. We are grateful to Michel Talagrand for sending us copies of his work, in particular [T4] prior to publication. This inspired most of Section 8. We also are indebted to Dima Ioffe for suggesting at the right moment that the inequalities in [BL] could be the right tool to make use of Theorem 8.1. This proved a key idea. We thank Aernout van Enter for a careful reading of the manuscript and numerous helpful comments.

2. Generalized random mean field models

This section introduces the general setup of our approach, including a definition of the concept of "generalized random mean field model" and the corresponding thermodynamic formalism. But before giving formal definitions, we will show how such a class of models and the Hopfield model in particular arises naturally in the attempt to construct mean field models for spin glasses, or to construct models of autoassociative memory.

2.1. The Hopfield model as a mean field spin glass.

The derivation we are going to present does not follow the historical development. In fact, what is generally considered "the" mean field spin glass model, the *Sherrington-Kirkpatrick model* [SK], is different (although, as we will see, related) and not even, according to the definition we will use, a mean field model (a fact which may explain why it is so much harder to analyse than its inventors apparently expected, and which in many ways makes it much more interesting). What do we mean by "mean field model"? A spin system on a lattice is, roughly, given by a lattice, typically $\mathbb{Z}^d$, a local spin space $\mathcal{S}$, which could be some Polish space but which for the present we can think of as the discrete set $\mathcal{S} = \{-1, +1\}$, the *configuration space* $\mathcal{S}_\infty \equiv \mathcal{S}^{\mathbb{Z}^d}$ and its finite volume subspaces $\mathcal{S}_\Lambda \equiv \mathcal{S}^\Lambda$ for any finite $\Lambda \subset \mathbb{Z}^d$, and a Hamiltonian function H that for any finite Λ gives the energy of a configuration $\sigma \in \mathcal{S}_\infty$ in the volume Λ, as $H_\Lambda(\sigma)$. We will say that a spin system is a *mean field model* if its Hamiltonian depends on σ only through a set of so-called *macroscopic functions* or *order parameters*. By this we mean typically spatial averages of local functions of the configuration. If the mean field model is supposed to describe reasonably well a given spin system, a set of such functions should be used so that their equilibrium values suffice to characterize completely the *phase diagram* of the model. For instance, for a ferromagnetic spin system it suffices to consider the *total magnetization* in a volume Λ, $m_\Lambda(\sigma) \equiv \frac{1}{|\Lambda|}\sum_{i\in\Lambda}\sigma_i$ as order parameter. A mean field Hamiltonian for a ferromagnet is then $H_\Lambda^{fm}(\sigma) = -|\Lambda| E(m_\Lambda(\sigma))$; the physically most natural choice $E(m) = \frac{1}{2}m^2$ gives the *Curie-Weiss model.* Note that

$$H_\Lambda^{fm}(\sigma) = -\sum_{i\in\Lambda}\sigma_i\left[\frac{E(m_\Lambda(\sigma))}{m_\Lambda(\sigma)}\right] \tag{2.1}$$

which makes manifest the idea that in this model the spins σ_i at the site i interact only with the (non-local) *mean-field* $\frac{E(m_\Lambda(\sigma))}{m_\Lambda(\sigma)}$. In the Curie-Weiss case this mean field is of course the mean magnetization itself. Note that the order parameter $m_\Lambda(\sigma)$ measures how close the spin configuration in Λ is to the ferromagnetic ground states $\sigma_i \equiv +1$, resp. $\sigma_i \equiv -1$. If we wanted to model an antiferromagnet, the corresponding order parameter would be the *staggered magnetization* $m_\Lambda(\sigma) \equiv \frac{1}{|\Lambda|}\sum_{i\in\Lambda}(-1)^{\sum_{\gamma=1}^d i_\gamma}\sigma_i$.

In general, a natural choice for a set of order parameters will be given by the projections of the spin configurations to the *ground states* of the system. By ground states we mean configurations σ that for all Λ minimize the Hamiltonian H_Λ in the sense that $H_\Lambda(\sigma)$ cannot be made smaller by changing σ only within Λ[1]. So if $\xi^1,\dots,\xi^M$ are the ground states of our system, we should define the M order parameters $m^1_\Lambda(\sigma) = \frac{1}{|\Lambda|}\sum_{i\in\Lambda}\xi^1_i\sigma_i, \dots, m^M_\Lambda(\sigma) = \frac{1}{|\Lambda|}\sum_{i\in\Lambda}\xi^M_i\sigma_i$ and take as a Hamiltonian a function $H^{mf}_\Lambda(\sigma) = -|\Lambda| E\left(m^1_\Lambda(\sigma),\dots,m^M_\Lambda(\sigma)\right)$. For consistency, one should of course choose E in such a way that $\xi^1,\dots,\xi^M$ are ground states of the so defined $H^{mf}_\Lambda(\sigma)$. We see that in this spirit, the construction of a mean field model departs from assumptions on the ground states of the real model.

Next we should say what we mean by "spin glass". This is a more complicated issue. The generally accepted model for a lattice spin-glass is the Edwards-Anderson model [EA] in which Ising spins on a lattice $\mathbb{Z}^d$ interact via nearest-neighbour couplings J_{ij} that are independent random variables with zero mean. Little is known about the low-temperature properties of this model on a rigorous level, and even on the heuristic level there are conflicting opinions, and it will be difficult to find consensus within a reasonably large crowd of experts on what should be reasonable assumptions on the nature of ground states in a spin glass. But there will be some that would agree on the two following features which should hold in high enough dimension[2]

(1) The ground states are "disordered".
(2) The number of ground states is infinite.

Moreover, the most "relevant" ground states should be stationary random fields, although not much more can be said a priori on their distribution. Starting from these assumptions, we should choose some function $M(\Lambda)$ that tends to infinity as $\Lambda\uparrow\mathbb{Z}^d$ and $M(\Lambda)$ random vectors ξ^μ, defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ and taking values in $\mathcal{S}_\infty$ and define, for all $\omega\in\Omega$, a $M(\Lambda)$-dimensional vector of order parameters with components,

$$m^\mu_\Lambda[\omega](\sigma) \equiv \frac{1}{|\Lambda|}\sum_{i\in\Lambda}\xi^\mu_i[\omega]\sigma_i \tag{2.2}$$

[1] We are somewhat too simplistic here. The notion of ground states should in general not only be applied to individual configurations but rather to measures on configuration space (mainly to avoid the problem of local degeneracy); however, we will ignore such complications here.

[2] For arguments in favour of this, see e.g. [BF,vE], for a different view e.g. [FH].

and finally choosing the Hamiltonian as some function of this vector. The most natural choice in many ways is

$$\begin{aligned} H_\Lambda[\omega](\sigma) &= -\frac{|\Lambda|}{2} \|m_\Lambda[\omega](\sigma)\|_2^2 \\ &= -\frac{|\Lambda|}{2} \sum_{\mu=1}^{M(\Lambda)} [m_\Lambda[\omega](\sigma)]^2 \\ &= -\frac{1}{2|\Lambda|} \sum_{i,j\in\Lambda} \sum_{\mu=1}^{M(\Lambda)} \xi_i^\mu[\omega]\xi_j^\mu[\omega]\sigma_i\sigma_j \end{aligned} \tag{2.3}$$

If we make the additional assumption that the random variables ξ_i^μ are independent and identically distributed with $\mathbb{P}[\xi_i^\mu = \pm 1] = \frac{1}{2}$ we have obtained exactly the Hopfield model [Ho] in its most standard form[3]. Note that at this point we can replace without any loss Λ by the set $\{1, \dots, N\}$. Note also that many of the most common variants of the Hopfield model are simply obtained by a different choice of the function $E(m)$ or by different assumptions on the distribution of ξ.

In the light of what we said before we should check whether this choice was consistent, i.e. whether the ground states of the Hamiltonian (2.3) are indeed the vectors ξ^μ, at least with probability tending to one. This will depend on the behavior of the function $M(N)$. From what is known today, in a strict sense this is true only if $M(N) \leq c\frac{N}{\ln N}$ [McE,Mar] whereas under a mild relaxation (allowing deviations that are invisible on the level of the macroscopic variables m_N), this holds as long as $\lim_{N\uparrow\infty} \frac{M(N)}{N} = 0$ [BGP1]. It does not hold for faster growing $M(N)$ [Lu]. On the contrary, one might ask whether for given $M(\Lambda)$ consistency can be reached by the choice of a different distribution $\mathbb{P}$. This seems an interesting, and to our knowledge completely uninvestigated question.

2.2 The Hopfield model as an autoassociative memory.

Hopfield's purpose when deriving his model was not to model spin glasses, but to describe the capability of a neural network to act as a memory. In fact, the type of interaction for him was more or less dictated by assumptions on neural functioning. Let us, however, give another, fake, derivation of his model. By an *autoassociative memory* we will understand an algorithm that is capable of associating input

[3] Observe that the lattice structure of the set Λ plays no rôle anymore and we can consider it simply as a set of points

data to a preselected set of learned *patterns*. Such an algorithm may be deterministic or stochastic. We will generally only be interested in *complex* data, i.e. a pattern should contain a large amount of information. A pattern is thus naturally described as an element of a set $\mathcal{S}^N$, and a reasonable description of any possible datum $\sigma \in \mathcal{S}^N$ within that set in relation to the *stored* patterns $\xi^1, \dots \xi^M$ is in terms of its similarity to these patterns that is expressed in terms of the vector of *overlap parameters* $m(\sigma)$ whose components are $m^\mu(\sigma) = \frac{1}{N}\sum_{i=1}^N \xi_i^\mu \sigma_i$. If we agree that this should be all the information we care about, it is natural to construct an algorithm that can be expressed in terms of these variables only. A most natural candidate for such an algorithm is a Glauber dynamics with respect to a mean field Hamiltonian like (2.3). Functioning of the memory is then naturally interpreted by the existence of equilibrium measures corresponding to the stored patterns. Here the assumptions on the distribution of the patterns are dictated by a priori assumptions on the types of patterns one wants to store, and the maximal $M(N)$ for which the memory "functions" is called *storage capacity* and should be determined by the theory. In this paper we will not say much about this dynamical aspect, mainly because there are almost no mathematical results on this. It is clear from all we know about Glauber dynamics, that a detailed knowledge of the equilibrium distribution is necessary, but also "almost" sufficient to understand the main features of the long time properties of the dynamics. These things are within reach of the present theory, but only first steps have been carried out (See e.g. [MS]).

2.3 Definition of generalized random mean field models.

Having seen how the Hopfield model emerges naturally in the framework of mean field theory, we will now introduce a rather general framework that allows to encompass this model as well as numerous generalizations. We like to call this framework *generalized* random mean field models mainly due to the fact that we allow an unbounded number of order parameters, rather than a finite (independent of N) one which would fall in the classical setting of mean field theory and for which the standard framework of large deviation theory, as outlined in Ellis' book [El], applies immediately.

A generalized random mean field model needs the following ingredients.

(i) A single spin space $\mathcal{S}$ that we will always take to be a subset of some linear space, equipped with some a priori probability measure q.

(ii) A state space $\mathcal{S}^N$ whose elements we denote by σ and call *spin* *configurations*, equipped with the product measure $\prod_i q(d\sigma_i)$.

(iii) The dual space $(\mathcal{S}^N)^{*M}$ of linear maps $\xi^T_{N,M} : \mathcal{S}^N \to \mathbb{R}^M$.

(iv) A mean field potential which is some real valued function $E_M : \mathbb{R}^M \to \mathbb{R}$, that we will assume

(iv.1) Bounded below (w.l.g. $E_M(m) \geq 0$).

(iv.2) in most cases, convex and "essentially smooth", that is, it has a domain $\mathcal{D}$ with non-empty interior, is differentiable on its domain, and $\lim_{m\to\partial\mathcal{D}} |\nabla E_M(m)| = +\infty$ (see [Ro]).

(v) An abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and measurable maps $\xi^T : \Omega \to (\mathcal{S}^{\mathbb{N}})^{*\mathbb{N}}$. Note that if Π_N is the canonical projection $\mathbb{R}^{\mathbb{N}} \to \mathbb{R}^N$, then $\xi^T_{M,N}[\omega] \equiv \Pi_M \xi^T[\omega] \circ \Pi_N^{-1}$ are random elements of $(\mathcal{S}^N)^{*M}$.

(vi) The random order parameter

$$m_{N,M}[\omega](\sigma) \equiv \frac{1}{N}\xi^T_{M,N}[\omega]\sigma \in \mathbb{R}^M \tag{2.4}$$

(vii) A random Hamiltonian

$$H_{N,M}[\omega](\sigma) \equiv -NE_M\left(m_{N,M}[\omega](\sigma)\right) \tag{2.5}$$

Remark. The formulation above corresponds to what in large deviation theory is known as "level 1", i.e. we consider the Hamiltonian as a function of order parameters that are functions ("empirical averages") rather than as a function of empirical measures as in a "level 2" formulations. In some cases a level 2 formulation would be more natural, but since in our main examples everything can be done on level 1, we prefer to stick to this language.

With these objects we define the *finite volume Gibbs measures*, (which more precisely are probability measure valued random variables) $\mu_{\beta,N,M}$ on $(\mathcal{S}^N, \mathcal{B}(\mathcal{S}^N))$ through

$$\mu_{\beta,N,M}[\omega](d\sigma) = \frac{e^{-\beta H_{N,M}[\omega](\sigma)}}{Z_{\beta,N,M}[\omega]}\prod_{i=1}^N q(d\sigma_i) \tag{2.6}$$

where the normalizing factor, called *partition function*, is

$$Z_{\beta,N,M}[\omega] \equiv \mathbb{E}_\sigma e^{-\beta H_{N,M}[\omega](\sigma)} \tag{2.7}$$

where $\mathbb{E}_\sigma$ stands for the expectation with respect to the a priori product measure on $\mathcal{S}^N$. Due to the special feature of these models that $H_{N,M}[\omega]$ depends on σ only through $m_{N,M}[\omega](\sigma)$, the distribution of these quantities contains essentially all information on the Gibbs measures themselves (i.e. the measures $\mu_{\beta,N,M}[\omega]$ restricted to the level sets of the functions $m_{N,M}[\omega]$ are the uniform distribution on these sets) and thus play a particularly prominent rôle. They are measures on $(\mathbb{R}^M, \mathcal{B}(\mathbb{R}^M))$ and we will call them *induced measures* and denote them by

$$\mathcal{Q}_{\beta,N,M}[\omega] \equiv \mu_{\beta,N,M}[\omega] \circ \left(\frac{1}{N}\xi^T_{N,M}[\omega]\right)^{-1} \tag{2.8}$$

In the classical setting of mean field theory, N would now be considered as the large parameter tending to infinity while M would be some constant number, independent of N. The main new feature here is that both N and M are large parameters and that as N tends to infinity, we choose $M \equiv M(N)$ as some function of N that tends to infinity as well. However, we stress that the entire approach is geared to the case where at least $M(N) < N$, and even $M(N)/N \equiv \alpha$ is small. In fact, the passage to the induced measures $\mathcal{Q}$ appears reasonably motivated only in this case, since only then we work in a space of lower dimension. To study e.g. the Hopfield model for α large will require entirely different ideas which we do not have.

It may be worthwhile to make some remarks on randomness and self averaging at this point in a somewhat informal way. As was pointed out in [BGP1], the distribution $\mathcal{Q}$ of the order parameters can be expected to be much less "random" than the distribution of the spins. This is to be understood in a rather strong sense: Define

$$f_{\beta,N,M,\rho}[\omega](m) \equiv -\frac{1}{\beta N} \ln \mathcal{Q}_{\beta,N,M}[\omega]\left(B_\rho(m)\right) \tag{2.9}$$

where $B_\rho(m) \subset \mathbb{R}^M$ is the ball of radius ρ centered at m. Then by strong self-averaging we mean that (for suitably chosen ρ) f as a function of m is everywhere "close" to its expectation with probability close to one (for N large)). Such a fact holds in a sharp sense when M is bounded, but it remains "essentially" true as long as $M(N)/N \downarrow 0$ (This statement will be made precise in Section 6). This is the reason why under this hypothesis, these systems actually behave very much like ordinary mean field models. When $\alpha > 0$, what "close" can mean will depend on α, but for small α this will be controllable. This is the reason why it will turn out to be possible to study the situation with α small as a perturbation of the case $\alpha = 0$.

2.4 Thermodynamic limits

Although in some sense "only finite volume estimates really count", we are interested generally in asymptotic results as N (and M) tend to infinity, and it is suitable to discuss in a precise way the corresponding procedure of *thermodynamic limits.*

In standard spin systems with short range interactions there is a well established beautiful procedure of constructing infinite volume Gibbs measures from the set of all finite volume measures (with "boundary conditions") due to Dobrushin, Lanford and Ruelle (for a good exposition see e.g. [Geo]). This procedure cannot be applied in the context of mean field models, essentially because the finite volume Hamiltonians are not restrictions to finite volume of some formal infinite volume Hamiltonian, but contain parameters that depend in an explicit way on the volume N. It is however still possible to consider so called *limiting Gibbs measures* obtained as accumulation points of sequences of finite volume measures. This does, however require some discussion.

Observe first that it is of course trivial to extend the finite volume Gibbs measures $\mu_{\beta,N,M}$ to measures on the infinite product space $(\mathcal{S}^{\mathbb{N}}, \mathcal{B}(\mathcal{S}^{\mathbb{N}}))$, e.g. by tensoring it with the a priori measures q on the components $i > N$. Similarly, the induced measures can be extended to the space $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ by tensoring with the Dirac measure concentrated on 0. One might now be tempted to define the set of limiting Gibbs measures as the set of limit points, e.g.

$$\mathcal{C}_\beta[\omega] \equiv \text{clus}_{N\uparrow\infty} \left\{ \mu_{\beta,N,M(N)}[\omega] \right\} \tag{2.10}$$

where $\text{clus}_{N\uparrow\infty} a_N$ denotes the set of limit points ("cluster set") of the sequence a_N. However, it is easy to see that in general this set is not rich enough to describe the physical content of the model. E.g., if we consider the Curie-Weiss model (c.f. (2.1)) it is easy to see and well known that this cluster set would always consist of a single element, namely the measure $\frac{1}{2}\left(\prod_{i=1}^{\infty} q^{m^*(\beta)} + \prod_{i=1}^{\infty} q^{-m^*(\beta)}\right)$, where $q^a(\sigma_i) = \frac{e^{\beta a \sigma_i}}{2\cosh(\beta a)}$ and where $m^*(\beta)$ is the largest solution of the equation

$$x = \tanh \beta x \tag{2.11}$$

(and which we will have many occasions to meet in the sequel of this article). If $\beta > 1$, $m^*(\beta) > 0$, and the limiting measure is a mixture; we would certainly want to be allowed to call the two summands limiting Gibbs measures as well, and to consider them as *extremal*, with *all* limiting Gibbs measures convex combinations of them. The fact that more than one such extremal measure exists would be the sign of the occurrence of a *phase transition* if $\beta > 1$.

The standard way out of this problem is to consider a richer class of *tilted Gibbs measures*

$$\mu^h_{\beta,N,M}[\omega](d\sigma) \equiv \frac{e^{-\beta H_{N,M}[\omega](\sigma)+\beta N h(m_{N,M}[\omega](\sigma))}}{Z^h_{\beta,N,M}[\omega]} \prod_{i=1}^N q(d\sigma_i) \tag{2.12}$$

where $h : \mathbb{R}^M \to \mathbb{R}$ is a *small* perturbation that plays the rôle of a symmetry breaking term. In most cases it suffices to choose *linear* perturbations, $h\left(m_{N,M}[\omega](\sigma)\right) = (h, m_{N,M}[\omega](\sigma))$, in which case h can be interpreted as a *magnetic field.* Instead of (2.10) one defines then the set

$$\tilde{\mathcal{C}}_\beta[\omega] \equiv \text{clus}_{\|h\|_\infty \downarrow 0, N\uparrow\infty} \left\{ \mu^h_{\beta,N,M(M)}[\omega] \right\} \tag{2.13}$$

where we first consider the limit points that can be obtained for all $h \in \mathbb{R}^\infty$ and then collect all possible limit points that can be obtained as h is taken to zero (with respect to the sup-norm). Clearly $\mathcal{C}_\beta \subset \tilde{\mathcal{C}}_\beta$. If this inclusion is strict, this means that the infinite volume Gibbs measures depend in a discontinuous way on h at $h = 0$, which corresponds to the standard physical definition of a first order phase transition. We will call $\tilde{\mathcal{C}}_\beta[\omega]$ the set of *limiting Gibbs measures.*

The set $\tilde{\mathcal{C}}_\beta[\omega]$ will in general *not* be a convex set. E.g., in the Curie-Weiss case, it consists, for $\beta > 1$ of three elements, $\mu^+_{\beta,\infty}, \mu^-_{\beta,\infty}$, and $\frac{1}{2}(\mu^+_{\beta,\infty} + \mu^-_{\beta,\infty})$. (Exercise: Prove this statement!). However, we may still consider the convex closure of this set and call its extremal points *extremal Gibbs measures.* It is likely, but we are not aware of a proof, that all elements of the convex closure can be obtained as limit points if the limits $N \uparrow 0$, $\|h\|_\infty \downarrow 0$ are allowed to be taken jointly (Exercise: Prove that this is true in the Curie-Weiss model!).

Of course, in the same way we define the tilted induced measures, and the main aim is to construct, in a more or less explicit way, the set of limiting induced measures. We denote these sets by $\mathcal{C}^Q_\beta[\omega]$, and $\tilde{\mathcal{C}}^Q_\beta[\omega]$, respectively. The techniques used will basically of large deviation type, with some modifications necessary. We will discuss this formalism briefly in Section 3 and 5.

2.5 Convergence and propagation of chaos.

Here we would like to discuss a little bit the expected or possible behaviour of generalized random mean field models. Our first remark is that all the sets $\mathcal{C}_\beta[\omega]$ and $\tilde{\mathcal{C}}_\beta[\omega]$ will not be empty if $\mathcal{S}$ is compact. The same holds in most cases for $\mathcal{C}_\beta^{\mathcal{Q}}[\omega]$ and $\tilde{\mathcal{C}}_\beta^{\mathcal{Q}}[\omega]$, namely when the image of $\mathcal{S}^N$ under $\xi_{N,M}^T$ is compact. This may, however, be misleading. Convergence of a sequence of measures $\mathcal{Q}_{\beta,N,M(N)}$ on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ in the usual weak sense means simply convergence of all finite dimensional marginals. Now take the sequence $\delta_{e^{M(N)}}$, of Dirac-measures concentrated on the $M(N)$-th unit vector in $\mathbb{R}^\infty$. Clearly, this sequence converges to the Dirac measure concentrated on zero, and this observation obviously misses a crucial point about this sequence. Considered rather as a measure on the set of unit vectors, this sequence clearly does *not* converge. For most purposes it thus more appropriate to use a ℓ^2-topology rather than the more conventional product topology. In this sense, the above sequence of Dirac measures does, of course, not converge weakly, but converges vaguely to the zero measure.

It is an interesting question whether one can expect, in a random situation, that there exist *subsequences* of untilted measures converging weakly in the ℓ_2 topology in a phase transition region. Ch. Külske [Ku] recently constructed an example in which the answer to this question is negative. He also showed, that, as long as $M(N) < \ln N$, in the standard Hopfield model, the sets $\mathcal{C}_\beta^{\mathcal{Q}}[\omega]$ and $\tilde{\mathcal{C}}_\beta^{\mathcal{Q}}[\omega]$ coincide for almost all ω.

In conventional mean field models, the induced measures converge (if properly arranged) to Dirac measures, implying that in the thermodynamic limit, the macroscopic order parameters verify a *law of large numbers*. In the case of infinitely many order parameters, this is not obviously true, and it may not even seem reasonable to expect, if $M(N)$ is not considerably smaller than N. Indeed, it has been shown in [BGP1] that in the Hopfield model this holds if $\frac{M(N)}{N} \downarrow 0$. Another paradigm of mean field theory is *propagation of chaos* [Sn], i.e. the fact that the (extremal) limiting Gibbs measures are product measures, i.e. that any finite subset of spins forms a family of independent random variables in the thermodynamic limit. In fact, both historically and in most standard textbooks on statistical mechanics, *this* is the starting assumption for the derivation of mean field theory, while models such as the Curie-Weiss model are just convenient examples where these assumptions happen to be verified. In the situation of random models, this is a rather subtle issue, and we will come back to this in Section 8 where we will learn actually a lot about this.

2.6 Examples.

Before turning to the study of large deviation techniques, we conclude this section by presenting a list of commonly used variants of the Hopfield model and to show how they fit into the above framework.

2.6.1 The standard Hopfield model.

Here $\mathcal{S} = \{-1, 1\}$, q is the Bernoulli measure $q(1) = q(-1) = \frac{1}{2}$. $(\mathcal{S}^N)^*$ may be identified with $\mathbb{R}^N$ and $\xi^T_{N,M}$ are real $M \times N$-matrices. The mean field potential is $E_M(m) = \frac{1}{2}\|m\|_2^2$, where $\|\cdot\|_2$ denotes the 2-norm in $\mathbb{R}^M$. The measure $\mathbb{P}$ is such that ξ_i^μ are independent and identically distributed with $\mathbb{P}[\xi_i^\mu = \pm 1] = \frac{1}{2}$. The order parameter is the M-dimensional vector

$$m_{N,M}[\omega](\sigma) = \frac{1}{N}\sum_{i=1}^N \xi_i \sigma_i \tag{2.14}$$

and the Hamiltonian results as the one in (2.3).

2.6.2 Multi-neuron interactions.

This model was apparently introduced by Peretto and Niez [PN] and studied for instance by Newman [N]. Here all is the same as in the previous case, except that the mean field potential is $E_M(m) = \frac{1}{p}\|m\|_p^p$, $p > 2$. For (even) integer p, the Hamiltonian is then

$$H_{N,M}[\omega](\sigma) = -\frac{1}{N^p}\sum_{i_1,\dots,i_p} \sigma_{i_1}\dots\sigma_{i_p}\sum_{\mu=1}^M \xi_{i_1}^\mu \dots \xi_{i_p}^\mu \tag{2.15}$$

2.6.3 Biased Hopfield model.

Everything the same as in 2.6.1, but the distribution of ξ_i^μ is supposed to reflect an asymmetry (bias) between $+1$ and -1 (e.g. to store pictures that are typically more black than white). That is, we have (e.g.) $\mathbb{P}[\xi_i^\mu = 2x] = (1-x)$ and $\mathbb{P}[\xi_i^\mu = 2(1-x)] = x$. One may, of course, consider the model with yet different distributions of the ξ_i^μ.

2.6.4 Hopfield model with correlated patterns.

In the same context, also the assumption of independence of the ξ_i^μ is not always reasonable and may be dropped. One speaks of *semantic* correlation, if the components of each vector ξ^μ are independent, while the different vectors are correlated, and of *spatial* correlation, if the different vectors ξ^μ are independent, but have correlated components ξ_i^μ. Various reasons for considering such types of patterns can be found in the literature [FZ,Mi]. Other types of correlation considered include the case where $\mathbb{P}$ is the distribution of a family of Gibbs random fields [SW].

2.6.5 Potts-Hopfield model.

Here the space $\mathcal{S}$ is the set $\{1, 2, \ldots, p\}$, for some integer p, and q is the uniform measure on this set. We again have random patterns ξ_i^μ that are independent and the marginal distribution of $\mathbb{P}$ coincides with q. The order parameters are defined as

$$m_M^\mu[\omega](\sigma) = \frac{1}{N} \sum_{i=1}^{N} \left[\delta_{\sigma_i, \xi_i^\mu} - \tfrac{1}{p} \right] \tag{2.16}$$

for $\mu = 1, \ldots, M$. E_M is the same as in the standard Hopfield model. Note that the definition of m_M seems not to fit exactly our setting. The reader should figure out how this can be fixed. See also [G1]. A number of other interesting variants of the model really lie outside our setting. We mention two of them:

2.6.6 The dilute Hopfield model.

Here we are in the same setting as in the standard Hopfield model, except that the Hamiltonian is no longer a function of the order parameter. Instead, we need another family of, let us say independent, random variables, J_{ij}, with $(i, j) \in \mathbb{N} \times \mathbb{N}$ with distribution e.g. $\mathbb{P}[J_{ij} = 1] = x$, $\mathbb{P}[J_{ij} = 0] = 1 - x$, and the Hamiltonian is

$$H_{N,M}[\omega](\sigma) = -\frac{1}{2Nx} \sum_{i,j} \sigma_i \sigma_i J_{i,j}[\omega] \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu \tag{2.17}$$

This model describes a neural network in which each neuron interacts only with a fraction x of the other neurons, with the set of a priori connections between neuron described as a random graph [BG1,BG2]. This is certainly a more realistic assumption when one is modelling biological neural networks like the brain of a rat. The point here is that, while this model is not a generalized mean field model, if we replace the Hamiltonian (2.17) by its average with respect to the random variables J, we get back the original Hopfield Hamiltonian. On the other hand, it is true that

$$\sup_{\sigma \in \mathcal{S}^N} \left| H_{N,M}[\omega](\sigma) - \mathbb{E}\left[H_{N,M}[\omega](\sigma) \middle| \mathcal{F}_\xi \right] \right| \leq cN \sqrt{\frac{M}{xN}} \tag{2.18}$$

with overwhelming probability, which implies that in most respects the dilute model has the same behaviour as the normal one, provide $\frac{M}{xN}$ is small. The estimate (2.18) has been proven first in [BG2], but a much simpler proof can be found in [T4].

2.6.7 The Kac-Hopfield model.

This model looks similar to the previous one, but here some non-random geometry is introduced. The set $\{1,\dots,N\}$ is replaced by $\Lambda\subset\mathbb{Z}^d$, and the random J_{ij} by some deterministic function $J_\gamma(i-j)\equiv \gamma^d J(\gamma(i-j))$ with $J(x)$ some function with bounded support (or rapid decay) whose integral equals one. Here γ is a small parameter. This model had already been introduced by Figotin and Pastur [FP3] but has been investigated more thoroughly only recently [BGP2, BGP4]. It shows very interesting features and an entire article in this volume is devoted to it.

3. Large deviation estimates and transfer principle

The basic tools to study the models we are interested in are large deviation estimates for the induced measures $\mathcal{Q}_{\beta,N,M}$. Compared to the standard situations, there are two particularities in the setting of generalized random mean field models that require some special attention: (i) the dimension M of the space on which these measures are defined must be allowed to depend on the basic large parameter N and (ii) the measure $\mathcal{Q}_{\beta,N,M}$ is itself random. A further aspect is maybe even more important. We should be able to compute, in a more or less explicit form, the "rate function", or at least be able to identify its minima. In the setting we are in, this is a difficult task, and we will stress the calculational aspects here. We should mention that in the particular case of the Hopfield model with quadratic interaction, there is a convenient trick, called the *Hubbard-Stratonovich transformation* [HS] that allows one to circumvent the technicalities we discuss here. This trick has been used frequently in the past, and we shall come back to it in Section 8. The techniques we present here work in much more generality and give essentially equivalent results. The central result that will be used later is Theorem 3.5.

3.1. Large deviations estimates.

Let us start with the general large deviation framework adopted to our setting. Let M and N be two integers. Given a family $\{\nu_N,\ N\geq 1\}$ of probability measures on $(\mathbb{R}^M,\mathcal{B}(\mathbb{R}^M))$, and a function $E_M\colon \mathbb{R}^M\to\mathbb{R}$ (hypotheses on E_M will be specified later on), we define a new family $\{\mu_N,\ N\geq 1\}$ of probability measures on $(\mathbb{R}^M,\mathcal{B}(\mathbb{R}^M))$ via

$$\mu_N(\Gamma)\equiv\frac{\int_\Gamma e^{NE_M(x)}d\nu_N(x)}{\int_{\mathbb{R}^M} e^{NE_M(x)}d\nu_N(x)}\ ,\quad \Gamma\in\mathcal{B}(\mathbb{R}^M) \tag{3.1}$$

We are interested in the large deviation properties of this new family. In the case when M is a fixed integer, it follows from Varadhan's lemma on the asymptotics of integrals that, if $\{\nu_N, N \geq 1\}$ satisfies a large deviation principle with good rate function $I(\cdot)$, and if E_M is suitably chosen (we refer to [DS], Theorem 2.1.10 and exercise (2.1.24) for a detailed presentation of these results in a more general setting) then $\{\mu_N, N \geq 1\}$ satisfies a large deviation principle with good rate function $J(x)$ where

$$J(x) = -[E_M(x) - I(x)] + \sup_{y \in \mathbb{R}^M} [E_M(y) - I(y)] \tag{3.2}$$

Here we address the question of the large deviation behaviour of $\{\mu_N, N \geq 1\}$ in the case where $M \equiv M(N)$ is an unbounded function of N and where the measure ν_N is defined as follows:

Let ξ be a linear transformation from $\mathbb{R}^N$ to $\mathbb{R}^M$. To avoid complications, we assume that $M \leq N$ and ξ is non-degenerate, i.e. its image is all $\mathbb{R}^M$. We will use the same symbol to denote the corresponding $N \times M$ matrix $\xi \equiv \{\xi_{i,\mu}\}_{i=1,\dots N; \mu=1,\dots M}$ and we will denote by $\xi^\mu \equiv (\xi_1^\mu, \dots, \xi_N^\mu) \in \mathbb{R}^M$, respectively $\xi_i \equiv (\xi_i^1, \dots, \xi_i^M) \in \mathbb{R}^N$, the μ-th row vector and i-th column vector. The transposed matrix (and the corresponding adjoint linear transformation from $\mathbb{R}^M$ to $\mathbb{R}^N$) is denoted ξ^T. Consider a probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P})$ and its N-fold power $(\mathbb{R}^N, \mathcal{P}_N)$ where $\mathcal{P}_N = \mathcal{P}^{\otimes N}$. We set

$$\nu_N \equiv \mathcal{P}_N \circ \left(\tfrac{1}{N}\xi^T\right)^{-1} \tag{3.3}$$

In this subsection we will present upper and lower large deviation bounds for fixed N. More precisely we set, for any $\rho > 0$ and $x^* \in \mathbb{R}^M$,

$$Z_{N,\rho}(x^*) \equiv \int_{B_\rho(x^*)} e^{NE_M(x)} d\nu_N(x) \tag{3.4}$$

In the regime where $\lim_{N\to\infty} \frac{M}{N} = 0$, estimates on these quantities provide a starting point to prove a strong large deviation principle for $\{\mu_N, N \geq 1\}$ in a formulation that extends the "classical" Cramèr's formulation. This was done in [BG4] in the case of the standard Hopfield model. In the regime where $\lim_{N\to\infty} \frac{M}{N} = \alpha$ with $\alpha > 0$, we cannot anymore establish such a LDP. But estimates on $Z_{N,\rho}(x^*)$ will be used to establish concentration properties for $\mathcal{Q}_N$ asymptotically as N tends to infinity, as we will see later in the paper.

Following the classical procedure, we obtain an upper bound on $Z_{N,\rho}(x^*)$ by optimizing on a family of exponential Markov inequalities.

As is well known, this will require the computation of the conjugate of[4] the *logarithmic moment generating function*, defined as

$$\mathcal{L}_{N,M}(t) \equiv \frac{1}{N} \log \int_{\mathbb{R}^M} e^{N(t,x)} \nu_N(dx) \quad , t \in \mathbb{R}^M \tag{3.5}$$

In the setting we are in, the computation of this quantity is generally quite feasible. A recurrent theme in large deviation theory is that of the Legendre transform. To avoid complications that will not arise in our examples, we restrict the following discussion mainly to the case when the Legendre transform is well defined (and involutary) which is essentially the case where the convex function is *strictly convex* and *essentially smooth.* We recall from [Ro]:

Definition 3.1. *A real valued function g on a convex set C is said to be* strictly convex *on C if*

$$g((1-\lambda)x + \lambda y) < (1-\lambda)g(x) + \lambda g(y) \quad 0 < \lambda < 1 \tag{3.6}$$

for any two different points x and y in C. It it called proper *if it is not identically equal to $+\infty$.*

An extended-real-valued function h on R^M is essentially smooth *if it satisfies the following three conditions for $C = int(domh)$:*
(a) C is non empty;
(b) h is differentiable throughout C;
(c) $\lim_{i\to\infty} |\nabla h(x_i)| = +\infty$ whenever $x_1, x_2, \ldots,$ is a sequence in C converging to a boundary point x of C.

(Recall that $\text{dom} g \equiv \{x \in \mathbb{R}^M \mid g(x) < \infty\}$). Note that if a function E_M is essentially smooth, it follows (c.f. [RV], Theorem A and B and [Ro], pp. 263-272) that E_M attains a minimum value and the set on which this (global) minimum is attained consists of a single point belonging to the interior of it's domain. Without loss of generality we will assume in the sequel that $E_M(x) \geq 0$ and $E_M(0) = 0$.

All through this chapter we adopt the usual approach that consists in identifying a convex function g on $\text{dom} g$ with the convex function defined throughout the space $\mathbb{R}^M$ by setting $g(x) = +\infty$ for $x \notin \text{dom} g$.

[4] We have chosen to follow Rockafellar's terminology and speak about conjugacy correspondence and conjugate of a (convex) function instead of Legendre-Fenchel conjugate, as is often done. This will allow us to refer to [Ro] and the classical Legendre transform avoiding confusions that might otherwise arise.

Definition 3.2. *Let g be a proper convex function. The function g^* defined by*

$$g^*(x^*) = \sup_{x \in \mathbb{R}^M} \{(x, x^*) - g(x)\} \tag{3.7}$$

is called its (ordinary) conjugate.

For any set S in $\mathbb{R}^M$ we denote by $\mathrm{int} S$ its interior. For smooth g we denote by $\nabla g(x) \equiv \left(\frac{\partial g(x)}{\partial x^1}, \dots, \frac{\partial g(x)}{\partial x^\mu}, \dots, \frac{\partial g(x)}{\partial x^M}\right)$, $\nabla^2 g(x) \equiv \left(\frac{\partial^2 g(x)}{\partial x^\mu \partial x^\nu}\right)_{\mu,\nu=1,\dots,N}$ and $\Delta g(x) \equiv \sum_{\mu=1}^M \frac{\partial^2 g(x)}{\partial^2 x^\mu}$ respectively the gradient vector, the Hessian matrix, and the Laplacian of g at x.

The following lemma collects some well-known properties of $\mathcal{L}_{N,M}$ and its conjugate:

Lemma 3.3.
(a) $\mathcal{L}_{N,M}$ and $\mathcal{L}^*_{N,M}$ are proper convex functions from $\mathbb{R}^M$ to $\mathbb{R} \cup \infty$.
(b) $\mathcal{L}_{N,M}(t)$ is infinitely differentiable on

$\mathrm{int}(\mathrm{dom}\mathcal{L}_{N,M})$. *Defining the measure $\tilde{\nu}_{N,t}$ via $d\tilde{\nu}_{N,t}(X) \equiv \frac{\exp\{N(t,X)\}}{\int \exp\{N(t,X)\} d\nu_N(X)} d\nu_N(X)$, and denoting by $\tilde{\mathbb{E}}_t(\cdot)$, the expectation w.r.t. $\tilde{\nu}_{N,t}$ we have, for any t in* $\mathrm{dom}\mathcal{L}_{N,M}$,

$$\begin{aligned} \nabla \mathcal{L}_{N,M}(t) &= \tilde{\mathbb{E}}_t(X) = \left(\tilde{\mathbb{E}}_t(X_\mu)\right)_{\mu=1,\dots,M} \\ \tfrac{1}{N} \nabla^2 \mathcal{L}_{N,M}(t) &= \left(\tilde{\mathbb{E}}_t(X_\mu X_\nu) - \tilde{\mathbb{E}}_t(X_\mu)\tilde{\mathbb{E}}_t(X_\nu)\right)_{\mu,\nu=1,\dots,M} \end{aligned} \tag{3.8}$$

and, if $\mathcal{L}^*$ is smooth, the following three conditions on x are equivalent

$$\begin{aligned} &1) \quad \nabla \mathcal{L}_{N,M}(t) = x \\ &2) \quad \mathcal{L}^*_{N,M}(x) = (t, x) - \mathcal{L}_{N,M}(t) \\ &3) \quad (y, x) - \mathcal{L}_{N,M}(y) \quad \textit{achieves its supremum over } y \textit{ at } y = t \end{aligned} \tag{3.9}$$

(c) $\mathcal{L}^*_{N,M}(x) \geq 0$ and, if $\tilde{\mathbb{E}}_0(X) < \infty$, $\mathcal{L}^*_{N,M}(\tilde{\mathbb{E}}_0(X)) = 0$.

Proof. The proofs of statements (a) and (c) can be found in [DZ], as well as the proof of the differentiability property. The formulae (3.8) are simple algebra. Finally, the equivalence of the three conditions (3.9) is an application of Theorem 23.5 of [Ro] to the particular case of a differentiable proper convex function.

∎

Setting

$$\Psi_{N,M}(x) \equiv -E_M(x) + \mathcal{L}^*_{N,M}(x) \ \ , x \in \mathbb{R}^M \tag{3.10}$$

we have

Lemma 3.4. *For any x^* in $\mathbb{R}^M$, define $t^* \equiv t^*(x^*)$ through $\mathcal{L}^*_{N,M}(x^*) = (t^*, x^*) - \mathcal{L}_{N,M}(t^*)$ if such a t^* exists while otherwise $\|t^*\|_2 \equiv \infty$ (note that t^* need not be unique). We have, for any $\rho > 0$,*

$$\frac{1}{N} \log Z_{N,\rho}(x^*) \leq -\Psi_{N,M}(x^*) + \sup_{x \in B_\rho(x^*)} [E_M(x) - E_M(x^*)] + \rho\|t^*\|_2 \tag{3.11}$$

and

$$\begin{aligned} \frac{1}{N} \log Z_{N,\rho}(x^*) \geq -\Psi_{N,M}(x^*) + \inf_{x \in B_\rho(x^*)} [E_M(x) - E_M(x^*)] - \rho\|t^*\|_2 \\ + \tfrac{1}{N} \log(1 - \tfrac{1}{\rho^2 N} \Delta \mathcal{L}_{N,M}(t^*)) \end{aligned} \tag{3.12}$$

Proof. Analogous bounds were obtained in [BG4], Lemmata 2.1 and 2.2, in the special case of an application to the Hopfield model. The proofs of (3.11) and (3.12) follow the proofs of these lemmata with only minor modifications. We will only recall the main lines of the proof of the lower bound: the essential step is to perform an exponential change of measure i.e., with the definition of $\tilde{\nu}_{N,t}$ from Lemma 3.4, we have,

$$\frac{1}{N} \log Z_{N,\rho}(x^*) = \tilde{\mathbb{E}}_{t^*} \left(e^{N\{E_M(X) - (t^*,X)\}} \mathbb{1}_{\{B_\rho(x^*)\}} \right) \tilde{\mathbb{E}}_0 \left(e^{N(t^*,X)} \right) \tag{3.13}$$

from which, together with (3.5) and (3.9), we easily obtain,

$$\begin{aligned} \frac{1}{N} \log Z_{N,\rho}(x^*) \geq e^{N\left\{-\Psi_{N,M}(x^*) + \inf_{x \in B_\rho(x^*)}[E_M(x) - E_M(x^*)] - \rho\|t^*\|_2\right\}} \\ \times \tilde{\nu}_{N,t^*}(B_\rho(x^*)) \end{aligned} \tag{3.14}$$

When the law of large numbers is not available, as is the case here, the usual procedure to estimate the term $\tilde{\nu}_{N,t^*}(B_\rho(x^*))$ would be to use the upper bound. Here we simply use the Tchebychev inequality to write

$$1 - \tilde{\nu}_{N,t^*}(B_\rho(x^*)) = \tilde{\mathbb{E}}_{t^*} \left(\mathbb{1}_{\{\|X - x^*\|_2^2 > \rho^2\}} \right) \leq \tfrac{1}{\rho^2} \tilde{\mathbb{E}}_{t^*} \|X - x^*\|_2^2 \tag{3.15}$$

Now, by (3.9), t^* satisfies $\nabla\mathcal{L}_{N,M}(t^*) = x^*$, and it follows from (3.8) that

$$\tilde{\mathbb{E}}_{t^*}\|X - x^*\|_2^2 = \frac{1}{\rho^2}\sum_{\mu=1}^{M}\left[\tilde{\mathbb{E}}_{t^*}X_\mu^2 - \left(\tilde{\mathbb{E}}_{t^*}X_\mu\right)^2\right] = \tfrac{1}{\rho^2 N}\Delta\mathcal{L}_{N,M}(t^*) \tag{3.16}$$

Collecting (3.14), (3.15) and (3.16) proves (3.12).

■

Remark. The lower bound (3.12) is meaningful only if $\frac{1}{N\rho^2}\Delta\mathcal{L}_{M,N}(x) < 1$. But the Laplacian of a function on $\mathbb{R}^M$ has a tendency to be of order M. Thus, typically, the lower bound will be useful only if $\rho^2 \geq O(M/N)$. We see that if $\lim_{N\uparrow\infty}\frac{M}{N} = 0$, one may shrink ρ to 0 and get upper and lower bounds that are asymptotically the same (provided E_M is continuous), provided the norm of t^* remains bounded. Since t^* is random, this introduces some subtleties which, however, can be handled (see [BG4]). But if $\lim_{N\uparrow\infty}\frac{M}{N} = \alpha > 0$, we do not get a lower bound for balls of radius smaller than $O(\sqrt{\alpha})$ and there is no hope to get a large deviation principle in the usual sense from Lemma 3.4. What is more disturbing, is the fact that the quantities Ψ and t^* are more or less impossible to compute in an explicit form, and this makes Lemma 3.4 not a very good starting point for further investigations.

3.2. Transfer principle.

As we will show now, it is possible to get large deviation *estimates* that do not involve the computation of Legendre transforms. The price to pay will be that these will not be sharp everywhere. But as we will see, they are sharp at the locations of the extrema and thus are sufficient for many purposes. Let us define the function

$$\Phi_{N,M}(x) = -E_M(x) + (x, \nabla E_M(x)) - \mathcal{L}_{N,M}(\nabla E_M(x)) \tag{3.17}$$

Theorem 3.5.

(i) Let x^* be a point in $\mathbb{R}^M$ such that for some $\rho_0 > 0$, for all $x, x' \in B_{\rho_0}(x^*)$, $\|\nabla E_M(x) - \nabla E_M(x')\|_2 < c\|x - x'\|_2$. Then, for all $0 < \rho < \rho_0$

$$\frac{1}{N}\log Z_{N,\rho}(x^*) \leq -\Phi_{N,M}(x^*) + \frac{1}{2}c\rho^2 \tag{3.18}$$

(ii) Let x^* be a point such that $\nabla\mathcal{L}_{N,M}(\nabla E_M(x^*)) = x^*$. Then,

$$\frac{1}{N}\log Z_{N,\rho}(x^*) \geq -\Phi_{N,M}(x^*) + \tfrac{1}{N}\log(1 - \tfrac{1}{\rho^2 N}\Delta\mathcal{L}_{N,M}(\nabla E_M(x^*))) \tag{3.19}$$

Remark. The condition $\nabla\mathcal{L}_{N,M}(\nabla E_M(x^*)) = x^*$ is equivalent to the condition $\nabla\Psi_{N,M}(x^*) = 0$, if $\mathcal{L}^*$ is essentially smooth. This means that the lower bound holds at all critical points of the "true" rate function. It is easy to see that $\nabla\Psi_{N,M}(x) = 0$ implies $\nabla\Phi_{N,M}(x) = 0$, while the converse is not generally true. Fortunately, however, this *is* true for critical points of $\Phi_{N,M}$ that are minima. This fact will be established in the remainder of this section.

Remark. It is clear that we could get an upper bound with error term $C\rho$ without the hypothesis that ∇E_M is Lipshitz. However, when we apply Theorem 3.5, a good estimate on the error will be important[5], while local Lipshitz bounds on ∇E_M are readily available.

Proof. With the definition of $\tilde{\nu}_{N,t}$ from Lemma 3.4, we have,

$$\begin{aligned} Z_{N,\rho}(x^*) &= \tilde{\mathbb{E}}_t\left(e^{N\{E_M(X)-(t,X)\}}\mathbb{1}_{\{B_\rho(x^*)\}}\right)\tilde{\mathbb{E}}_0\left(e^{N(t,X)}\right) \\ &= e^{N\{\mathcal{L}_{N,M}(t)+E_M(x^*)-(t,x^*)\}} \\ &\times \tilde{\mathbb{E}}_t\left(e^{N\{E_M(X)-E_M(x^*)-(t,(X-x^*))\}}\mathbb{1}_{\{B_\rho(x^*)\}}\right) \end{aligned} \tag{3.20}$$

The strategy is now to chose t in such a way as to get optimal control over the last exponent in (3.20). By the fundamental theorem of calculus,

$$\begin{aligned} &|E_M(X) - E_M(x^*) - (t,(X-x^*))| \\ &= \left|\int_0^1 ds\,((\nabla E_M(sX+(1-s)x^*) - t),(X-x^*))\right| \\ &\leq \sup_{s\in[0,1]} \|(\nabla E_M(sX+(1-s)x^*) - t\|_2\|X-x^*\|_2 \end{aligned} \tag{3.21}$$

Of course we want a bound that is uniform in the set of X we consider, so that the best choice is of course $t \equiv \nabla E_M(x^*)$. Since $\nabla E_M(x)$ was

[5] The point is that the number of balls of radius ρ to cover, say, the unit ball is of the order $\rho^{-\alpha N}$, that is exponentially large. Therefore we want to use as large a ρ as possible with as small an error as possible. Such problems do not occur when the dimension of the space is independent of N.

assumed to be Lipshitz in $B_\rho(x^*)$ we get

$$\begin{aligned} Z_{N,\rho}(x^*) &\leq e^{N\{\mathcal{L}_{N,M}(\nabla E_M(x^*))+E_M(x^*)-(\nabla E_M(x^*),x^*)\}} e^{\frac{1}{2}Nc\rho^2} \\ &= e^{-N\Phi_{N,M}(x^*)} e^{\frac{1}{2}Nc\rho^2} \end{aligned} \tag{3.22}$$

where the last equality follows from the definition (3.17). This proves the upper bound (3.18). To prove the lower bound, note that since E_M is convex,

$$E_M(X) - E_M(x^*) - (\nabla E_M(x^*), (X - x^*)) \geq 0 \tag{3.23}$$

Using this in the last factor of (3.20), we get

$$Z_{N,\rho}(x^*) \geq e^{-N\{\Phi_{N,M}(x^*)\}} \tilde{\nu}_{N,t}(B_\rho(x^*)) \tag{3.24}$$

Now, just as in (3.15),

$$1 - \tilde{\nu}_{N,t}(B_\rho(x^*)) \leq \tfrac{1}{N}\tilde{\mathbb{E}}_t \|X - x^*\|_2^2 \tag{3.25}$$

and a simple calculation as in Section 3.1 shows that

$$\tilde{\mathbb{E}}_t \|X - x^*\|_2^2 = \tfrac{1}{\rho^2 N}\Delta\mathcal{L}_{N,M}(t) + \|\nabla\mathcal{L}_{N,M}(t) - x^*\|_2^2 \tag{3.26}$$

Here we see that the optimal choice for t would be the solution of $\nabla\mathcal{L}_{N,M}(t) = x^*$, an equation we did not like before. However, we now have by assumption, $\nabla\mathcal{L}_{N,M}(\nabla E_M(x^*)) = x^*$. This concludes the proof of Theorem 3.14.

■

Sometimes the estimates on the probabilities of ℓ_2-balls may not be the most suitable ones. A charming feature of the upper bound is that it can also be extended to sets that are adapted to the function E_M. Namely, if we define

$$\tilde{Z}_{N,\rho}(x^*) \equiv \int e^{NE_M(x)} \mathbb{I}_{\{\|\nabla E_M(x) - \nabla E_M(x^*)\|_2 \leq \rho\}} d\nu_N(x) \tag{3.27}$$

we get

Theorem 3.6. *Assume that for some $q \leq 1$ for all $y, y' \in B_{\rho_0}(\nabla E_M(x^*))$, $\|(\nabla E_M)^{-1}(y) - (\nabla E_M)^{-1}(y')\|_2 \leq c\|y - y'\|_2^q$, then* *for all* $0 < \rho < \rho_0$

$$\frac{1}{N} \log \tilde{Z}_{N,\rho}(x^*) \leq -\Phi_{N,M}(x^*) + \frac{1}{2} c\rho^{1+q} \tag{3.28}$$

The proof of this Theorem is a simple rerun of that of the upper bound in Theorem 3.5 and is left to the reader.

We now want to make the remark following Theorem 3.5 precise.

Proposition 3.7. *Assume that* E_M *is strictly convex, and essentially smooth. If* $\Phi_{N.M}$ *has a local extremum at a point* x^* *in the interior of its domain, then* $\nabla\mathcal{L}(\nabla E_M(x^*)) = x^*$*.*

Proof. To prove this proposition, we recall a fundamental Theorem on functions of Legendre type from [Ro].

Definition 3.8. *Let* h *be a differentiable real-valued function on a open subset* C *of* $\mathbb{R}^M$*. The Legendre conjugate of the pair* (C, h) *is defined to be the pair* (D, g) *where* $D = \nabla h(C)$ *and* g *is the function on* D *given by the formula*

$$g(x^*) = ((\nabla h)^{-1}(x^*), x^*) - h((\nabla h)^{-1}(x^*)) \tag{3.29}$$

Passing from (C, h) *to* (D, g)*, if the latter is well defined, is called the Legendre transformation.*

Definition 3.9. *Let* C *be an open convex set and* h *an essentially smooth and strictly convex function on* C*. The pair* (C, h) *will be called a convex function of Legendre type.*

The Legendre conjugate of a convex function of Legendre type is related to the ordinary conjugate as follows:

Theorem 3.10. ([Ro], Theorem 26.5) *Let* h *be a closed convex function. Let* $C = int(domh)$ *and* $C^* = int(domh^*)$*. Then* (C, h) *is a convex function of Legendre type if and only if* (C^*, h^*) *is a convex function of Legendre type. When these conditions hold,* (C^*, h^*) *is the Legendre conjugate of* (C, h)*, and* (C, h) *is in turn the Legendre conjugate of* (C^*, h^*)*. The gradient mapping is then one-to-one from the open convex set* C *onto the open convex set* C^**, continuous in both directions and*

$$\nabla h^* = (\nabla h)^{-1} \tag{3.30}$$

With this tool at our hands, let us define the function $\psi_{N,M}(x) \equiv E_M^*(x) - \mathcal{L}_{N,M}(x)$. The crucial point is that since E_M is of Legendre type, by Definition 3.8 and Theorem 3.10, we get

$$\Phi_{N,M}(x) = \psi_{N,M}(\nabla E_M(x)) \tag{3.31}$$

Moreover, since ∇E_M is one-to-one and continuous, $\Phi_{N,M}$ has a local extremum at x^* if and only if $\psi_{N,M}$ has a local extremum at the point $y^* = \nabla E_M(x^*)$. In particular, $\nabla\psi_{N,M}(y^*) = 0$. Thus, $\nabla E_M^*(y^*) = \nabla\mathcal{L}_{N,M}(y^*)$, and by (3.30), $(\nabla E_M)^{-1}(y^*) = \nabla\mathcal{L}_{N,M}(y^*)$, or $x^* = \nabla\mathcal{L}_{N,M}(\nabla E_M(x^*))$, which was to be proven.

■

The proposition asserts that at the minima of Φ, the condition of part (ii) of Theorem 3.5 is satisfied. Therefore, if we are interested in establishing localization properties of our measures, we only need to compute Φ and work with it *as if it was the true rate function.* This will greatly simplify the analysis in the models we are interested in.

Remark. *If* $\mathcal{L}$ is of Legendre type, it follows by the same type of argument that x^* is a critical point of Ψ if and only if $\nabla E_M(x^*)$ is a critical point of ψ. Moreover, at such critical points, $\Phi(x^*) = \psi(\nabla E_M(x^*))$. Thus in this situation, if x^* is a critical point of Ψ, than x^* is a critical point of Φ, and $\Psi(x^*) = \Phi(x^*)$. Conversely, by Proposition 3.7, if Φ has a local extremum at x^*, then x^* is a critical point of Ψ and $\Phi(x^*) = \Psi(x^*)$. Since generally $\Psi(x) \geq \Phi(x)$, this implies also that if Φ has a minimum at x^*, then Ψ has a minimum at x^*. One can build on the above observations and establish a more complete "duality principle" between the functions Φ and Ψ in great generality, but we will not make use of these observations. The interested reader will find details in [G2].

4. Bounds on the norm of random matrices

One of the crucial observations that triggered the recent progress in the Hopfield model was the observation that the properties of the random matrix $A(N) \equiv \frac{\xi^T\xi}{N}$ play a crucial rôle in this model, and that their main feature is that as long as M/N is small, $A(N)$ is close to the identity matrix. This observation in a sense provided the proper notion for the intuitive feeling that in this case, "all patterns are almost orthogonal to each other". Credit must go to both Koch [K] and Shcherbina and Tirozzi [TS] for making this observation, although the properties of the matrices $A(N)$ had been known a long time before. In fact it is known that under the hypothesis that ξ_i^μ are independent, identically distributed random variables with $\mathbb{E}\xi_i^\mu = 0$, $\mathbb{E}\left[\xi_i^\mu\right]^2 = 1$ and $\mathbb{E}\left[\xi_i^\mu\right]^4 < \infty$, the maximal and minimal eigenvalues of $A(N)$ satisfy

$$\lim_{N\uparrow\infty} \lambda_{max}(A(N)) = (1+\sqrt{\alpha})^2, \quad \text{a.s.} \tag{4.1}$$

This statement was proven in [YBK] under the above (optimal) hypotheses. For prior results under stronger assumptions, see [Ge,Si,Gi]. Such results are generally proven by tedious combinatorial methods, combined with truncation techniques. Estimates for deviations that were available from such methods give only subexponential estimates; the best bounds known until recently, to our knowledge, were due to Shcherbina and Tirozzi [ST] and gave, in the case where ξ_i^μ are symmetric Bernoulli random variables

$$\mathbb{P}\left[\|A(N)-\mathbb{1}\| > [(1+\sqrt{\alpha})^2-1](1+\epsilon)\right] \leq \exp\left(-\frac{\epsilon^{4/3}M^{2/3}}{K}\right) \tag{4.2}$$

with K a numerical constant and valid for small ϵ. More recently, a bound of the form $\exp\left(-\frac{\epsilon^2 N}{K}\right)$ was proven by the authors in [BG5], using a concentration estimate due to Talagrand. In [T4] a simplified version of that proof is given. We will now give the simplest proof of such a result we can think of.

Let us define for a $M\times M$-matrix A the norm

$$\|A\| \equiv \sup_{\substack{x\in\mathbb{R}^M\\ \|x\|_2=1}} (x, Ax) \tag{4.3}$$

For positive symmetric matrices it is clear that $\|A\|$ is the maximal eigenvalue of A. We shall also use the notation $\|A\|_2 \equiv \sqrt{\sum_{\mu,\nu} A_{\mu\nu}^2}$.

Theorem 4.1. *Assume that $\mathbb{E}\xi_i^\mu = 0$, $\mathbb{E}\left[\xi_i^\mu\right]^2 = 1$ and $|\xi_i^\mu| \leq 1$. Then there exists a numerical constant K such that for large enough N, the following holds for all $\epsilon \geq 0$ and all $\alpha \geq 0$*

$$\begin{aligned}&\mathbb{P}\left[|\|A(N)\| - (1+\sqrt{\alpha})^2| \geq \epsilon\right]\\ &\quad\leq K\exp\left(-N\frac{(1+\sqrt{\alpha})^2}{K}\left(\sqrt{\frac{\epsilon}{1+\sqrt{\alpha}}+1}-1\right)^2\right)\end{aligned} \tag{4.4}$$

Proof. Let us define for the rectangular matrix ξ

$$\|\xi\|_+ \equiv \sup_{\substack{x \in \mathbb{R}^M \\ \|x\|_2 = 1}} \|\xi x\|_2 \tag{4.5}$$

Clearly

$$\|A(N)\| = \|\xi/\sqrt{N}\|_+^2 \tag{4.6}$$

Motivated by this remark we show first that $\|\xi/\sqrt{N}\|_+$ has nice concentration properties. For this we will use the following theorem due to Talagrand:

Theorem 4.2. (Theorem 6.6 in [T2]) *Let f be a real valued function defined on $[-1,1]^N$. Assume that for each real number a, the set $\{f \leq a\}$ is convex. Suppose that on a convex set $B \subset [-1,1]^N$ the restriction of f to B satisfies for all $x, y \in B$*

$$|f(x) - f(y)| \leq l_B \|x - y\|_2 \tag{4.7}$$

for some constant $l_B > 0$. Let h denote the random variable $h = f(X_1, \dots, X_N)$.Then, if M_f is a median of h, for all $t > 0$,

$$\mathbb{P}\left[|h - M_f| \geq t\right] \leq 4b + \frac{4}{1-2b} \exp\left(-\frac{t^2}{16 l_B^2}\right) \tag{4.8}$$

where b denotes the probability of the complement of the set B.

To make use of this theorem, we show first that $\|\xi/\sqrt{N}\|_+$ is a Lipshitz function of the i.i.d. variables ξ_i^μ:

Lemma 4.3. *For any two matrices ξ, ξ', we have that*

$$\left|\|\xi\|_+ - \|\xi'\|_+\right| \leq \|\xi - \xi'\|_2 \tag{4.9}$$

Proof. We have

$$\begin{aligned}
|\|\xi\|_+ - \|\xi'\|_+| &\leq \sup_{\substack{x\in\mathbb{R}^M\\ \|x\|_2=1}} |\|\xi x\|_2 - \|\xi' x\|_2| \\
&\leq \sup_{\substack{x\in\mathbb{R}^M\\ \|x\|_2=1}} \|\xi x - \xi' x\|_2 \\
&\leq \sup_{\substack{x\in\mathbb{R}^M\\ \|x\|_2=1}} \sqrt{\sum_{i=1}^N x_i^2 \sum_{i=1}^N \sum_{\mu=1}^M (\xi_i^\mu - {\xi'}_i^\mu)^2} = \|\xi - \xi'\|_2
\end{aligned} \tag{4.10}$$

where in the first inequality we used that the modulus of the difference of suprema is bounded by the supremum of the modulus of the differences, the second follows from the triangle inequality and the third from the Schwarz inequality.

■

Next, note that as a function of the variables $\xi \in [-1,1]^{MN}$, $\|\xi\|_+$ is convex. Thus, by Theorem 4.2, it follows that for all $t > 0$,

$$\mathbb{P}\left[|\|\xi/\sqrt{N}\|_+ - \mathbb{M}_{\|\xi/\sqrt{N}\|_+}| \geq t\right] \leq 4e^{-N\frac{t^2}{16}} \tag{4.11}$$

where $\mathbb{M}_{\|\xi/\sqrt{N}\|_+}$ is a median of $\|\xi/\sqrt{N}\|_+$. Knowing that $\|A(N)\|$ converges almost surely to the values given in (4.1) we may without harm replace the median by this value. Thus

$$\begin{aligned}
&\mathbb{P}\left[\|A(N)\|_+ - (1+\sqrt{\alpha})^2 \geq \epsilon\right] \\
&= \mathbb{P}\left[\|\xi/\sqrt{N}\|_+ - (1+\sqrt{\alpha}) \geq (1+\sqrt{\alpha})\left(\sqrt{1+\frac{\epsilon}{(1+\sqrt{\alpha})^2}} - 1\right)\right] \\
&\leq 4\exp\left(-N(1+\sqrt{\alpha})^2\left(\sqrt{1+\frac{\epsilon}{(1+\sqrt{\alpha})^2}} - 1\right)^2 /16\right)
\end{aligned} \tag{4.12}$$

and similarly, for $0 \leq \epsilon \leq (1+\sqrt{\alpha})^2$

$$\begin{aligned}
&\mathbb{P}\left[\|A(N)\|_+ - (1+\sqrt{\alpha})^2 \leq -\epsilon\right] \\
&= \mathbb{P}\left[\|\xi/\sqrt{N}\|_+ - (1+\sqrt{\alpha}) \leq (1+\sqrt{\alpha})\left(\sqrt{1-\frac{\epsilon}{(1+\sqrt{\alpha})^2}}-1\right)\right] \\
&\leq 4\exp\left(-N(1+\sqrt{\alpha})^2\left(\sqrt{1-\frac{\epsilon}{(1+\sqrt{\alpha})^2}}-1\right)^2/16\right)
\end{aligned} \tag{4.13}$$

while trivially $\mathbb{P}\left[\|A(N)\|_+ - (1+\sqrt{\alpha})^2 \leq -\epsilon\right] = 0$ for $\epsilon > (1+\sqrt{\alpha})^2$. Using that for $0 \leq x \leq 1$, $(\sqrt{1-x}-1)^2 \geq (\sqrt{1+x}-1)^2$, we get Theorem 4.1.

■

Remark. Instead of using the almost sure results (4.1), it would also be enough to use estimates on the expectation of $\|A(N)\|$ to prove Theorem 4.1. We see that the proof required no computation whatsoever; it uses however that we know the medians or expectations. The boundedness condition on ξ_i^μ arises from the conditions in Talagrand's Theorem. It is likely that these could be relaxed.

Remark. In the sequel of the paper we will always assume that our general assumptions on ξ are such that Theorem 4.1 holds. Of course, since exponential bounds are mostly not really necessary, one may also get away in more general situations. On the other hand, we shall see in Section 6 that unbounded ξ_i^μ cause other problems as well.

5. Properties of the induced measures

In this section we collect the general results on the localization (or concentration) of the induced measures in dependence on properties of the function $\Phi_{\beta,N,M}$ introduced in the previous section. There are two parts to this. Our first theorem will be a rather simple generalization to what could be called the "Laplace method". It states, roughly, the (hardly surprising) fact that the Gibbs measures are concentrated "near" the absolute minima of Φ. A second, and less trivial remark states that quite generally, the Gibbs measures "respect the symmetry of the law of the disorder". We will make precise what that means.

5.1 Localization of the induced measures.

The following Theorem will tell us what we need to know about the function Φ in order to locate the support of the limiting measures

$\mathcal{Q}$.

Theorem 5.1. *Let $\mathcal{A} \subset \mathbb{R}^\infty$ be a set such that for all N sufficiently large the following holds:*

(i) There is $n \in \mathcal{A}$ such that for all $m \in \mathcal{A}^c$,

$$\Phi_{\beta,N,M(N)}[\omega](m) - \Phi_{\beta,N,M(N)}[\omega](n) \geq C\alpha \tag{5.1}$$

for $C > c$ sufficiently large, with c the constant from (i) of Theorem 3.5.

(ii) $\Delta\mathcal{L}_{N,M}(\nabla E_M(n)) \leq KM$ for some $K < \infty$, and $B_{K\sqrt{\alpha}}(n) \subset \mathcal{A}$. Assume further that Φ satisfies a tightness condition, i.e. there exists a constant, a, sufficiently small (depending on C), such that for all $r > C\alpha$

$$\ell\left(\{m \,|\Phi_{\beta,M,N}[\omega](m) - \Phi_{\beta,M,N}[\omega](n) \leq r\}\right) \leq r^{M/2} a^M M^{-M/2} \tag{5.2}$$

where $\ell(\cdot)$ denotes the Lebesgue measure. Then there is $L > 0$ such that

$$\mathcal{Q}_{\beta,N,M(N)}[\omega]\left(\mathcal{A}^c\right) \leq e^{-L\beta M} \tag{5.3}$$

and in particular

$$\lim_{N\uparrow\infty} \mathcal{Q}_{\beta,N,M(N)}[\omega]\left(\mathcal{A}\right) = 1 \tag{5.4}$$

Remark. Condition (5.2) is verified, e.g. if Φ is bounded from below by a quadratic function.

Proof. To simplify notation, we put w.r.g. $\Phi_{\beta,N,M}[\omega](n) = 0$. Note first that by (ii) and (3.19) we have that (for suitably chosen ρ)

$$\mathcal{Q}_{\beta,N,M(N)}[\omega]\left(\mathcal{A}\right) \geq \frac{1}{Z_{\beta,N,M(N)}[\omega]}\frac{1}{2}e^{-\beta N\Phi_{\beta,N,M}[\omega](n)} = \frac{1}{2Z_{\beta,N,M(N)}[\omega]} \tag{5.5}$$

It remains to show that the remainder has much smaller mass. Note that obviously, by (i),

$$\begin{aligned}\mathcal{Q}_{\beta,N,M(N)}[\omega]\left(\mathcal{A}^c\right) &\leq \int_{C\alpha}^\infty dr\, \mathcal{Q}_{\beta,N,M(N)}[\omega]\left(\mathcal{A}^c \cap \{m \,|\Phi_{\beta,N,M}(m) = r\}\right)\\ &\leq \int_{C\alpha}^\infty dr\, \mathcal{Q}_{\beta,N,M(N)}[\omega]\left(\{m \,|\Phi_{\beta,N,M}(m) = r\}\right)\end{aligned} \tag{5.6}$$

Now we introduce a lattice $\mathcal{W}_{M,\alpha}$ of spacing $1/\sqrt{N}$ in $\mathbb{R}^M$. The point here is that any domain $D \subset \mathbb{R}^M$ is covered by the union of balls of radius $\sqrt{\alpha}$ centered at the lattice points in D, while the number of lattice points in any reasonably regular set D is smaller than $\ell(D)N^{M/2}$ (see e.g. [BG5] for more details). Combining this observation with the upper bound (3.18), we get from (5.6) that

$$
\begin{aligned}
&Z_{\beta,N,M(N)}[\omega]\mathcal{Q}_{\beta,N,M(N)}[\omega]\left(\mathcal{A}^c\right) \\
&\leq \int_{C\alpha}^{\infty} dre^{-\beta Nr}\ell\left(\{m\,|\Phi_{\beta,N,M}(m)\leq r\}\right)N^{M/2}e^{\beta Mc/2} \\
&\leq \int_{C\alpha}^{\infty} dre^{-\beta Nr}r^{M/2}a^M\alpha^{-M/2}e^{\beta\alpha c/2} \\
&\leq a^M e^{\beta Mc/2}\alpha\int_{C}^{\infty} dre^{-\beta Mr}r^{M/2} \\
&\leq a^M e^{\beta\alpha c/2}e^{-\beta MC/2}\alpha\int_{C}^{\infty} dre^{-\beta Mr/2}r^{M/2} \\
&e^{-\beta M[C/2-c/2-\ln a/\beta]}N^{-1}\left[\frac{2}{e\beta}\right]^M
\end{aligned}
\tag{5.7}
$$

which clearly for $\beta \geq 1$ can be made exponentially small in M for C sufficiently large. Combined with (5.5) this proves (5.3). (5.4) follows by a standard Borel-Cantelli argument.

■

Remark. We see at this point why it was important to get the error terms of order ρ^2 in the upper bound of Theorem 3.5; this allows us to choose $\rho \sim \sqrt{\alpha}$. otherwise, e.g. when we are in a situation where we want use Theorem 5.6, we could of course choose ρ to be some higher power of α, e.g. $\rho = \alpha$. This then introduces an extra factor $e^{M|\ln\alpha|}$, which can be offset only by choosing $C \sim |\ln\alpha|$, which of course implies slightly worse estimates on the sets where $\mathcal{Q}$ is localized.

5.2 Symmetry and concentration of measure.

Theorem 5.1 allows us to localize the measure near the "reasonable candidates" for the absolute minima of Φ. As we will see, frequently, and in particular in the most interesting situation where we expect a *phase transition*, the smallest set $\mathcal{A}$ satisfying the hypothesis of Theorem 5.1 we can find will still be a union of disjoint sets. The components of this set are typically linked by "symmetry". In such a situation we would like to be able to compare the exact mass of the individual components, a task that goes beyond the possibilities of the explicit large deviation estimates. It is the idea of concentration of measure that allows us to make use of the symmetry of the distribution $\mathbb{P}$ here. This fact was first noted in [BGP3], and a more elegant proof in the Hopfield model that made use of the Hubbard-Stratonovich transformation was given first in [BG5] and independently in [T4].

Here we give a very simple proof that works in more general situations. The basic problem we are facing is the following. Suppose we are in a situation where the set $\mathcal{A}$ from Theorem 5.1 can be decomposed as $\mathcal{A} = \cup_k \mathcal{A}_k$ for some collection of disjoint sets $\mathcal{A}_k$. Define

$$f_N[\omega](k) \equiv -\frac{1}{\beta N} \ln \mathbb{E}_\sigma \, e^{-\beta H_{N,M}[\omega](\sigma)} \mathbb{1}_{\{m_{N,M}[\omega](\sigma)\in\mathcal{A}_k\}} \tag{5.8}$$

Assume that by for all k

$$\mathbb{E} f_N[\omega](k) = \mathbb{E} f_N[\omega](1) \tag{5.9}$$

(Think of $\mathcal{A}_k = B_\rho(m^* e^k)$ in the standard Hopfield model). We want so show that this implies that for all k, $|f_N[\omega](k) - f_N[\omega](1)|$ is "small" with large probability. Of course we should show this by proving that each $f_N[\omega](k)$ is close to its mean, and such a result is typically given by concentration estimates. To prove this would be easy, if it were not for the indicator function in (5.8), whose argument depends on the random parameter ω as well as the Hamiltonian. Our strategy will be to introduce quantities $f_N^\epsilon(k)$ that are close to $f_N(k)$, and for which it is easy to prove the concentration estimates. We will then control the difference between $f_N^\epsilon(k)$ and $f_N(k)$. We set

$$f_N^\epsilon(k) \equiv -\frac{1}{\beta N} \ln \mathbb{E}_\sigma \left(\frac{\beta N}{2\pi\epsilon}\right)^{M/2} \int_{\mathcal{A}_k} dm \, e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}(\sigma)-m\|_2^2} e^{\beta N E_M(m)} \tag{5.10}$$

Note that the idea is that $\left(\frac{\beta N}{2\pi\epsilon}\right)^{M/2} e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}(\sigma)-m\|_2^2}$ converges to the Dirac distribution concentrated on $m_{N,M}(\sigma)$, so that $f_N^\epsilon(k)$ converges to $f_N(k)$ as $\epsilon \downarrow 0$. Of course we will have to be a bit more careful than just that. However, Talagrand's Theorem 6.6 of [T2] gives readily

Proposition 5.2. *Assume that ξ verifies the assumptions of Theorem 4.1 and $\mathcal{S}$ is compact. Then there is a finite universal constant C such that for all $\epsilon > 0$,*

$$\mathbb{P}\left[|f_N^\epsilon(k) - \mathbb{E} f_N^\epsilon(k)| > x\right] \leq Ce^{-M} + Ce^{-\frac{x^2\epsilon^2 N}{C}} \tag{5.11}$$

Proof. We must establish a Lipshitz bound for $f_N^\epsilon[\omega](k)$. For notational simplicity we drop the superfluous indices N and k and set $f^\epsilon[\omega] \equiv f_N^\epsilon[\omega](k)$. Now

$$
\begin{aligned}
&|f^\epsilon[\omega] - f^\epsilon[\omega']| \\
&= \frac{1}{\beta N}\left|\ln \frac{\mathbb{E}_\sigma \int_{\mathcal{A}_k} dm\, e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}[\omega](\sigma)-m\|_2^2} e^{\beta N E_M(m)}}{\mathbb{E}_\sigma \int_{\mathcal{A}_k} dm\, e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}[\omega'](\sigma)-m\|_2^2} e^{\beta N E_M(m)}}\right| \\
&= \frac{1}{\beta N}\left|\ln \frac{\mathbb{E}_\sigma \int_{\mathcal{A}_k} dm\, e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}[\omega'](\sigma)-m\|_2^2} e^{\beta N E_M(m)}}{\mathbb{E}_\sigma \int_{\mathcal{A}_k} dm\, e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}[\omega'](\sigma)-m\|_2^2} e^{\beta N E_M(m)}}\right. \\
&\qquad \left.\frac{e^{-\frac{\beta N}{2\epsilon}\left(\|m_{N,M}[\omega](\sigma)-m\|_2^2 - \|m_{N,M}[\omega'](\sigma)-m\|_2^2\right)}}{}\right| \\
&\leq \frac{1}{\epsilon} \sup_{\sigma\in\mathcal{S}^N, m\in\mathcal{A}_k} \left|\|m_{N,M}[\omega](\sigma) - m\|_2^2 - \|m_{N,M}[\omega'](\sigma) - m\|_2^2\right|
\end{aligned}
\tag{5.12}
$$

But

$$
\begin{aligned}
&\left|\|m_{N,M}[\omega](\sigma) - m\|_2^2 - \|m_{N,M}[\omega'](\sigma) - m\|_2^2\right| \\
&\leq \|m_{N,M}[\omega'](\sigma) - m_{N,M}[\omega](\sigma)\|_2 \|2m - m_{N,M}[\omega](\sigma) - m_{N,M}[\omega'](\sigma)\|_2 \\
&\leq \frac{1}{\sqrt{N}} \|\xi[\omega'] - \xi[\omega]\|_2 \left[R + c(\sqrt{\|A[\omega]\|} + \sqrt{\|A[\omega']\|})\right]
\end{aligned}
\tag{5.13}
$$

where R is a bound for m on $\mathcal{A}_k$. We wrote $A[\omega] \equiv \frac{\xi^T[\omega]\xi[\omega]}{N}$ to make the dependence of the random matrices on the random parameter explicit. Note that this estimate is uniform in σ and m. It is easy to see that $f^\epsilon[\omega]$ has convex level sets so that the assumptions of Theorem 6.6 of [T1] are verified. Proposition 5.2 follows from here and the bounds on $\|A[\omega]\|$ given by Theorem 4.1.

■

We see from Proposition 5.2 that we can choose an $\epsilon = N^{-\delta_1}$, and an $x = N^{-\delta_2}$ with $\delta_1, \delta_2 > 0$ and still get a probability that decays faster than any power with N.

Let us now see more precisely how $f^\epsilon[\omega]$ and $f^0[\omega]$ are related. Let us introduce as an intermediate step the ϵ-smoothed measures

$$
\tilde{\mathcal{Q}}^\epsilon_{\beta,N,M}[\omega] \equiv \mathcal{Q}_{\beta,N,M}[\omega] \star \mathcal{N}\left(0, \frac{\epsilon}{\beta N}\right) \tag{5.14}
$$

where $\mathcal{N}\left(0, \frac{\epsilon}{\beta N}\right)$ is a M-dimensional normal distribution with mean 0 and variance $\frac{\epsilon}{\beta N}\mathbb{1}$. We mention that in the case $E_M(m) = \frac{1}{2}\|m\|_2^2$, the choice $\epsilon = 1$ is particularly convenient. This convolution is then known as the "Hubbard-Stratonovich transformation" [HS]. Its use simplifies to some extent that particular case and has been used frequently, by us as well as other authors. It allows to avoid the complications of Section 3 altogether.

We set $\tilde{f}^\epsilon[\omega] \equiv -\frac{1}{\beta N}\ln\left(Z_{\beta,N,M}\tilde{\mathcal{Q}}^\epsilon_{\beta,N,M}(\mathcal{A}_k)\right)$. But

$$\begin{aligned}
&Z_{\beta,N,M}\tilde{\mathcal{Q}}^\epsilon_{\beta,N,M}(\mathcal{A}_k)\\
&= \mathbb{E}_\sigma\left(\frac{\beta N}{2\pi\epsilon}\right)^{M/2}\int_{\mathcal{A}_k} dm\, e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}(\sigma)-m\|_2^2} e^{\beta N E_M(m_{N,M}(\sigma))}\\
&= \mathbb{E}_\sigma\left(\frac{\beta N}{2\pi\epsilon}\right)^{M/2}\int_{\mathcal{A}_k} dm \mathbb{1}_{\{\|m_{N,M}(\sigma)-m\|_2\leq\delta\}}\, e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}(\sigma)-m\|_2^2}\\
&\quad\times e^{\beta N E_M(m_{N,M}(\sigma))}\\
&+ \mathbb{E}_\sigma\left(\frac{\beta N}{2\pi\epsilon}\right)^{M/2}\int_{\mathcal{A}_k} dm \mathbb{1}_{\{\|m_{N,M}(\sigma)-m\|_2>\delta\}}\, e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}(\sigma)-m\|_2^2}\\
&\quad\times e^{\beta N E_M(m_{N,M}(\sigma))}\\
&\equiv (I) + (II)
\end{aligned} \tag{5.15}$$

for $\delta > 0$ to be chosen. We will assume that on $\mathcal{A}_k$, E_M is uniformly Lipshitz for some constant C_L. Then

$$\begin{aligned}
(I) &\leq e^{+\beta N C_L\delta}\mathbb{E}_\sigma\left(\frac{\beta N}{2\pi\epsilon}\right)^{M/2}\int_{\mathcal{A}_k} dm\, e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}(\sigma)-m\|_2^2} e^{\beta N E_M(m)}\\
&= e^{\beta N C_L\delta} e^{-\beta N f^\epsilon[\omega]}
\end{aligned} \tag{5.16}$$

and

$$\begin{aligned}
(II) &\leq \mathbb{E}_\sigma 2^{M/2}\left(\frac{\beta N}{4\pi\epsilon}\right)^{M/2} e^{-\frac{\beta N}{4\epsilon}\delta^2}\\
&\quad\times\int_{\mathcal{A}_k} dm\, e^{-\frac{\beta N}{4\epsilon}\|m_{N,M}(\sigma)-m\|_2^2} e^{\beta N E_M(m_{N,M}(\sigma))}\\
&\leq 2^{M/2} e^{-\frac{\beta N}{4\epsilon}\delta^2}\mathbb{E}_\sigma e^{\beta N E_M(m_{N,M}(\sigma))}\left(\frac{\beta N}{4\pi\epsilon}\right)^{M/2}\\
&\quad\times\int dm e^{-\frac{\beta N}{4\epsilon}\|m_{N,M}(\sigma)-m\|_2^2} = 2^{M/2} e^{-\frac{\beta N}{4\epsilon}\delta^2} Z_{\beta,N,m}
\end{aligned} \tag{5.17}$$

In quite a similar way we can also get a lower bound on (I), namely

$$
\begin{aligned}
(I) \geq\, & e^{-\beta N C_L \delta}\mathbb{E}_\sigma \left(\frac{\beta N}{2\pi\epsilon}\right)^{M/2} \int_{\mathcal{A}_k} dm\, e^{-\frac{\beta N}{2\epsilon}\|m_{N,M}(\sigma)-m\|_2^2} e^{\beta N E_M(m)} \\
& - e^{-\beta N C_L \delta} 2^{M/2} e^{-\frac{\beta N}{4\epsilon}\delta^2} e^{\beta N \sup_{m\in\mathcal{A}_k} E_M(m)} \\
=\, & e^{-\beta N C_L \delta} e^{-\beta N f^\epsilon[\omega]} - e^{-\beta N C_L \delta} 2^{M/2} e^{-\frac{\beta N}{4\epsilon}\delta^2} e^{\beta N \sup_{m\in\mathcal{A}_k} E_M(m)}
\end{aligned}
\tag{5.18}
$$

Since we anticipate that $\epsilon = N^{-\delta_1}$, the second term in (5.18) is negligible compared to the first, and (II) is negligible compared to (I), with room even to choose δ tending to zero with N; e.g., if we choose $\delta = \epsilon^{1/4}$, we get that

$$
|\tilde{f}^\epsilon[\omega] - f^\epsilon[\omega]| \leq const.\epsilon^{1/4} \tag{5.19}
$$

for sufficiently small ϵ. (We assume that $|f^\epsilon[\omega]| \leq C$).

Finally we must argue that $\tilde{f}^\epsilon[\omega]$ and $f^0[\omega]$ differ only by little. This follows since $\mathcal{N}(0, \frac{\beta N}{\epsilon})$ is sharply concentrated on a sphere of radius $\frac{\epsilon\alpha}{\beta}$ (although this remark alone would be misleading). In fact, arguments quite similar to those that yield (5.19)(and that we will not reproduce here) give also

$$
|\tilde{f}^\epsilon[\omega] - f^0[\omega]| \leq const.\epsilon^{1/4} \tag{5.20}
$$

Combining these observations with Proposition 5.2 gives

Theorem 5.3. *Assume that ξ verifies the assumptions of Theorem 4.1 and $\mathcal{S}$ is compact. Assume that $\mathcal{A}_k \subset \mathbb{R}^M$ verifies*

$$
\mathcal{Q}_{\beta,N,M}[\omega](\mathcal{A}_k) \geq e^{-\beta N c} \tag{5.21}
$$

for some finite constant c, with probability greater than $1 - e^{-M}$. Then there is a finite constant C such that for $\epsilon > 0$ small enough for any k, l,

$$
\mathbb{P}\left[|f_N(k) - f_N(l)| \geq C\epsilon^{1/4} + x\right] \leq Ce^{-M} + Ce^{-\frac{x^2\epsilon^2 N}{C}} \tag{5.22}
$$

6. Global estimates on the free energy function

After the rather general discussion in the last three sections, we see that all results on a specific model depend on the analysis of the (effective) rate function $\Phi_{\beta,N}[\omega](x)$. The main idea we want to follow here is to divide this analysis in two steps:

(i) Study the average $\mathbb{E}\Phi_{\beta,N}[\omega](x)$ and obtain explicit bounds from which the locations of the global minima can be read off. This part is typically identical to what we would have to do in the case of finitely many patterns.

(ii) Prove that with large probability, $|\Phi_{\beta,N}[\omega](x) - \mathbb{E}\Phi_{\beta,N}[\omega](x)|$ is so small that the deterministic result from (i) holds essentially outside small balls around the locations of the minima for $\Phi_{\beta,N}[\omega](x)$ itself.

These results then suffice to use Theorems 5.1 and 5.3 in order to construct the limiting induced measures. The more precise analysis of Φ close to the minima is of interest in its own right and will be discussed in the next section.

We mention that this strict separation into two steps was not followed in [BG5]. However, it appears to be the most natural and reasonable procedure. Gentz [G1] used this strategy in her proof of the central limit theorem, but only in the regime $M^2/N \downarrow 0$. To get sufficiently good estimates when $\alpha > 0$, a sharper analysis is required in part (ii).

To get explicit results, we will from now on work in a more restricted class of examples that includes the Hopfield model. We will take $\mathcal{S} = \{-1, 1\}$, with $q(\pm 1) = 1/2$ and $E_M(m)$ of the form

$$E_M(m) \equiv \frac{1}{p}\|m\|_p^p \tag{6.1}$$

with $p \geq 2$ and we will only require of the variables ξ_i^μ to have mean zero, variance one and to be bounded. To simplify notation, we assume $|\xi_i^\mu| \leq 1$. We do not strive to get optimal estimates on constants in this generality, but provide all the tools necessary do so in any specific situation, if desired[5].

A simple calculation shows that the function of Theorem 3.5 defined in (3.17) in this case is given by (we make explicit reference to p and

[5] A word of warning is due at this point. We will treat these generalized models assuming always $M=\alpha N$. But from the memory point of view, these models should and do work with $M=\alpha N^{p-1}$ (see e.g. [Ne] for a proof in the context of storage capacity). For $p>2$ our approach appears perfectly inadequate to deal with so many patterns, as the description of system in terms of so many variables (far more than the original spins!) seems quite absurd. Anyhow, there is some fun in these models even in this more restricted setting, and since this requires only a little more work, we decided to present those results.

β, but drop the M)

$$\Phi_{p,\beta,N}[\omega](m) = \frac{1}{q}\|m\|_p^p - \frac{1}{\beta N}\sum_{i=1}^{N}\ln\cosh\left(\beta\sum_{\mu=1}^{M}\xi_i^\mu[\omega]|m_\mu|^{p-1}\text{sign}(m_\mu)\right) \tag{6.2}$$

where $\frac{1}{p}+\frac{1}{q}=1$.[6] Moreover

$$\Phi_{p,\beta,N}[\omega](m) = \left(\psi_{q,\beta,N}[\omega]\circ\nabla E_M\right)(m) \tag{6.3}$$

where $\psi_{q,\beta,N}[\omega] : \mathbb{R}^M \to \mathbb{R}$ is given by

$$\psi_{q,\beta,N}[\omega](x) = \frac{1}{q}\|x\|_q^q - \frac{1}{\beta N}\sum_{i=1}^{N}\ln\cosh\left(\beta(\xi_i[\omega],x)\right) \tag{6.4}$$

and $\nabla E_M : \mathbb{R}^M \to \mathbb{R}^M$, by

$$\nabla E_M(m) = \left(\nabla_1 E_M(m_1),\dots,\nabla_\mu E_M(m_\mu),\dots,\nabla_M E_M(m_M)\right) \tag{6.5}$$

where

$$\nabla_\mu E_M(m_\mu) = \text{ sign }(m_\mu)|m_\mu|^{p-1} \tag{6.6}$$

Since $\nabla_\mu E_M$ is a continuous and strictly increasing function going to $+\infty$, resp. $-\infty$, as m_μ goes to $+\infty$, resp. $-\infty$, (and being zero at $m_\mu = 0$) its inverse $\nabla_\mu E_M^{-1}$ exists and has the same properties as $\nabla_\mu E_M$. It is thus enough, in order to study the structure of the minima of $\Phi_{p,\beta,N}[\omega]$, to study that of $\psi_{q,\beta,N}[\omega]$.

Before stating our main theorem we need to make some comments on the generalized Curie-Weiss functions

$$\phi_{q,\beta}(z) = \frac{1}{q}|z|^q - \frac{1}{\beta}\ln\cosh(\beta z) \tag{6.7}$$

The standard Curie-Weiss case $q = 2$ is well documented (see e.g. [El]), but the general situation can be analyzed in the same way. In a quite general setting, this can be found in [EE]. A new feature for $q < 2$ is that now zero is always a *local* minimum and that there is a range of temperatures where three local minima exist while the absolute minimum is the one at zero. For sufficiently low temperatures, however,

[6] Throughout this section, q will stand for the conjugate of p.

the two minima away from zero are always the lowest ones. The critical temperature β_c is defined as the one where $\phi_{q,\beta}$ takes the same value at all three local minima. Thus a particular feature for all $q < 2$ is that for $\beta \geq \beta_c$, the position of the deepest minimum, $x^*(\beta)$, satisfies $x^*(\beta) \geq x_q^*(\beta_c) > 0$. Of course $x_q^*(\beta_c)$ tends to 0 as q tends to 2. For integer $p \geq 3$ we have thus the situation that $x^*(\beta) = O(1)$, and only in the case $p = 2$ do we have to take the possible smallness of $x^*(\beta)$ near the critical point into account.

Proposition 6.1. *Assume that ξ_i^μ are i.i.d., symmetric bounded random variables with variance 1. Let either $p = 2$ or $p \geq 3$. Then for all $\beta > \beta_c(p)$ there exists a strictly positive constant $C_p(\beta)$ and a subset $\Omega_1 \subset \Omega$ with $\mathbb{P}[\Omega_1] \geq 1 - O(e^{-\alpha N})$ such that for all $\omega \in \Omega_1$ the following holds for all x for which $x_\mu = sign(m_\mu)|m_\mu|^{p-1}$ with $\|m\|_2 \leq 2$: There is $\gamma_a > 0$ and a finite numerical constant c_1 such that for all $\gamma \leq \gamma_a$ if $\inf_{s,\mu} \|x - se^\mu x_q^*\|_2 \geq \gamma c_1 x_q^*$,*

$$\psi_{p,\beta,N}[\omega](x) - \frac{1}{q}(x_q^*)^q + \frac{1}{\beta}\ln\cosh(\beta x_q^*) \geq C_p(\beta) \inf_{s,\mu} \|x - se^\mu x_q^*\|_2^2 \quad (6.8)$$

where $C_2(\beta) \sim (m^*(\beta))^2$ as $\beta \downarrow 1$, and $C_p(\beta) \geq C_p > 0$ for $p \geq 3$. The infima are over $s \in \{-, 1, +1\}$ and $\mu = 1, \dots, M$.

Remark. Estimates on the various constants can be collected from the proofs. In case (i), $C_2(\beta)$ goes like 10^{-5}, and $\gamma_a \sim 10^{-8}$ and $c_1 \sim 10^{-7}$. These numbers are of course embarrassing.

From Proposition 6.1 one can immediately deduce localization properties of the Gibbs measure with the help of the theorems in Section 5. In fact one obtains

Theorem 6.2. *Assume that ξ_i^μ are i.i.d. Bernoulli random variables taking the values ± 1 with equal probability. Let either $p = 2$ or $p \geq 3$. Then there exists a finite constant c_p such that for all $\beta > \beta_c(p)$ there is subset $\Omega_1 \subset \Omega$ with $\mathbb{P}[\Omega_1] \geq 1 - O(e^{-\alpha N})$ such that for all $\omega \in \Omega_1$ the following holds:*
(i) In the case $p = 2$,

$$\mathcal{Q}_{\beta,N,M(N)}[\omega]\left(\cup_{s,\mu} B_{c_2\gamma m^*}(se^\mu m^*)\right) \geq 1 - \exp\left(-KM(N)\right) \quad (6.9)$$

(ii) In the case $p \geq 3$,

$$\mathcal{Q}_{\beta,N,M(N)}[\omega]\left(\cup_{s,\mu}\left\{m \in \mathbb{R}^M \,|x(m) \in B_{c_p\alpha|\ln\alpha|}(se^\mu x_q^*)\right\}\right) \\ \geq 1 - \exp\left(-KM(N)\right) \quad (6.10)$$

Moreover, for $h = \epsilon s e^\mu$, and any $\epsilon > 0$, for $p = 2$

$$\mathcal{Q}^h_{\beta,N,M(N)}[\omega]\left(B_{c_2\gamma m^*}(se^\mu m^*)\right) \geq 1 - \exp\left(-K(\epsilon)M(N)\right) \quad (6.11)$$

and for $p \geq 3$,

$$\begin{aligned}&\mathcal{Q}^h_{\beta,N,M(N)}[\omega]\left(\left\{m \in \mathbb{R}^M \,|x(m) \in B_{c_p\alpha|\ln\alpha|}(se^\mu x^*_q)\right\}\right)\\ &\quad \geq 1 - \exp\left(-K(\epsilon)M(N)\right)\end{aligned} \quad (6.12)$$

with $K(\epsilon) \geq const.\epsilon > 0$.

Remark. Theorem 6.2 was first proven, for the case $p = 2$, with imprecise estimates on the radii of the balls in [BGP1,BGP3]. The correct asymptotic behaviour (up to constants) given here was proven first in [BG5]. A somewhat different proof was given recently in [T4], after being announced in [T3] (with additional restrictions on β). The case $p \geq 3$ is new. It may be that the $|\ln\alpha|$ in the estimates there can be avoided. We leave it to the reader to deduce Theorem 6.2 from Proposition 6.1 and Theorems 5.1 and 5.3. In the case $p \geq 3$, Theorem 3.6 and the remark following the proof of Theorem 5.1 should be kept in mind.

Proof of Proposition 6.1. We follow our basic strategy to show first that the mean of $\psi_{q,\beta,N}[\omega]$ has the desired properties and to control the fluctuations via concentration estimates. We rewrite $\psi_{q,\beta,N}[\omega](x)$ as

$$\begin{aligned}\psi_{q,\beta,N}[\omega](x) = &+ \mathbb{E}\left\{\frac{1}{q}|(\xi_1,x)|^q - \frac{1}{\beta}\ln\cosh(\beta(\xi_1,x))\right\}\\ &+ \frac{1}{q}\|x\|_q^q - \frac{1}{q}\mathbb{E}|(\xi_1,x)|^q\\ &+ \frac{1}{\beta N}\sum_{i=1}^N\left\{\mathbb{E}\ln\cosh(\beta(\xi_i,x)) - \ln\cosh(\beta(\xi_i,x))\right\}\end{aligned} \quad (6.13)$$

We will study the first, and main, term in a moment. The middle term "happens" to be positive:

Lemma 6.3. *Let $\{X_j, j = 1,\dots,n\}$ be i.i.d. random variables with $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$, and let $x = (x_1,\dots,x_n)$ be a vector in $\mathbb{R}^n$. Then, for $1 < q \leq 2$,*

$$\|x\|_q^q - \mathbb{E}\left|\sum_{j=1}^n x_j X_j\right|^q \geq 0 \quad (6.14)$$

Equality holds if all but one component of the x_j are zero.

Proof. A straightforward application of the Hölder inequality yields

$$\mathbb{E}|\sum_{j=1}^n x_j X_j|^q \leq (\mathbb{E}|\sum_{j=1}^n x_j X_j|^2)^{\frac{q}{2}} = \|x\|_q^q \tag{6.15}$$

■

Let us now consider the first term in (6.13). For $q = 2$ we have from [BG5] the following bound: Let

$$\hat{c}(\beta) \equiv \frac{\ln\cosh(\beta x^*)}{\beta (x^*)^2} - \frac{1}{2} \tag{6.16}$$

Then for all $\beta > 1$ and for all z

$$\phi_{2,\beta}(z) - \phi_{2,\beta}(x^*) \geq \hat{c}(\beta)(|z| - x^*)^2 \tag{6.17}$$

Moreover $\hat{c}(\beta)$ tends to $\frac{1}{2}$ as $\beta \uparrow \infty$, and behaves like $\frac{1}{12}(x^*(\beta))^2$, as $\beta \downarrow 1$.

Proposition 6.4. *Assume that ξ_1^μ are i.i.d., symmetric and $\mathbb{E}(\xi_1^\mu)^2 = 1$ and $|\xi_1^\mu| \leq 1$. Let either $p = 2$ or $p \geq 3$. Then for all $\beta > \beta_c$ (of p) there exists a positive constant $C_q(\beta)$ such that for all x such that $x_\mu = sign(m_\mu)|m_\mu|^{p-1}$ with $\|m\|_2 \leq 2$,*

$$\begin{aligned} \mathbb{E}\left\{\frac{1}{q}|(\xi_1, x)|^q - \frac{1}{\beta}\ln\cosh(\beta(\xi_1, x))\right\} - \frac{1}{q}(x^*)^q + \frac{1}{\beta}\ln\cosh(\beta x^*) \\ \geq C_q(\beta)\inf_{\mu,s}\|x - s e^\mu x^*\|_2^2 \end{aligned} \tag{6.18}$$

where x^* is the largest solution of the equation $x^{q-1} = \tanh\beta x$. In the case $q = 2$ $C_2(\beta) = \frac{1}{5000}\left(\frac{\ln\cosh(\beta x^*)}{\beta(x^*)^2} - \frac{1}{2}\right) \approx \frac{1}{600000}(x^*)^2$ for $\beta \downarrow 1$.

Remark. Note that nothing depends on α in this proposition. The constants appearing here are quite poor, but the proof is fairly nice and universal. In a very recent paper [T4] has a similar result where the constant seems to be $1/256L$, but so far we have not been able to figure out what his estimate for L would be. Anyway, there are other options if the proof below is not to your taste!

Proof. It is not difficult to convince oneself of the fact that there exist positive constants $\tilde{C}_q(\beta)$ such that for all $Z = (\xi_1, x)$ satisfying the assumption of the proposition

$$\frac{1}{q}|Z|^q - \frac{1}{\beta}\ln\cosh(\beta Z) - \frac{1}{q}(x^*)^q + \frac{1}{\beta}\ln\cosh(\beta x^*) \geq \tilde{C}_q(\beta)\left(|Z| - x^*\right)^2 \tag{6.19}$$

For $q = 2$ this follows from Lemma (6.17). For $q \geq 3$, note first that the allowed $|Z|$ are bounded. Namely,

$$|(\xi_1, x)| \leq \left|\sum_{\mu=1}^{M} |\xi_1^\mu||m_\mu^{p-1}|\right| \leq \sqrt{\sum_\mu m_\mu^2}\sqrt{\sum_\mu |m_\mu|^{2(p-2)}} \leq \|m\|_2^2 \tag{6.20}$$

using that $\|m\|_\infty \leq 1$ and the Hölder inequality in the case $p > 3$. Moreover since by definition $\pm x^*$ are the only points where the function $\phi_{q,\beta}(z)$ takes its absolute minimum, and x^* is uniformly bounded away from 0, it is clear that a lower bound of the form (6.19) can be constructed on the bounded interval $[-2, 2]$.
We have to bound the expectation of the right hand side of (6.19).

Lemma 6.5. *Let $Z = X + Y$ where X, Y are independent real valued random variables. Then for any $\epsilon > 0$*

$$\begin{aligned}\mathbb{E}(|Z| - x^*)^2 \geq \frac{1}{2}\left(\sqrt{\mathbb{E}Z^2} - x^*\right)^2 + \frac{1}{2}\epsilon^2\mathbb{P}[|X| > \epsilon] \\ \times \min\left(\mathbb{P}[Y > \epsilon], \mathbb{P}[Y < -\epsilon]\right)\end{aligned} \tag{6.21}$$

Proof. First observe that, since $\mathbb{E}|Z| \leq \sqrt{\mathbb{E}Z^2}$,

$$\begin{aligned}\mathbb{E}(|Z| - x^*)^2 &= \left(\sqrt{\mathbb{E}Z^2} - x^*\right)^2 + 2x^*\mathbb{E}\left(\sqrt{\mathbb{E}Z^2} - |Z|\right) \\ &\geq \left(\sqrt{\mathbb{E}Z^2} - x^*\right)^2\end{aligned} \tag{6.22}$$

On the other hand, Tchebychev's inequality gives that for any positive ϵ,

$$\mathbb{E}\left(|Z| - x^*\right)^2 \geq \epsilon^2\mathbb{P}\left[||Z| - x^*| > \epsilon\right] \tag{6.23}$$

Now it is clear that if $|X| > \epsilon$, then $||X + Y| - x^*| > \epsilon$ either if $Y > \epsilon$ or if $Y < -\epsilon$ (or in both cases). This gives the desired estimate. Thus

(6.23) implies that

$$\mathbb{E}\left(|Z|-x^*\right)^2 \geq \epsilon^2 \mathbb{P}[|X|>\epsilon] \min\left(\mathbb{P}[Y>\epsilon], \mathbb{P}[Y<-\epsilon]\right) \tag{6.24}$$

(6.22) and (6.24) together imply (6.21).

■

In the case of symmetric random variables, the estimate simplifies to

$$\mathbb{E}(|Z|-x^*)^2 \geq \frac{1}{2}\left(\sqrt{\mathbb{E}Z^2}-x^*\right)^2+\frac{1}{4}\epsilon^2 \mathbb{P}[|X|>\epsilon]\mathbb{P}[|Y|>\epsilon] \tag{6.25}$$

which as we will see is more easy to apply in our situations. In particular, we have the following estimates.

Lemma 6.6. *Assume that* $X=(x,\xi)$ *where* $|\xi^\mu| \leq 1$, $\mathbb{E}\xi^\mu=0$ *and* $\mathbb{E}(\xi^\mu)^2=1$. *Then for any* $1>g>0$,

$$\mathbb{P}\left[|X|>g\|x\|_2\right] \geq \frac{1}{4}\left(1-g^2\right)^2 \tag{6.26}$$

Proof. A trivial generalization of the Paley-Zygmund inequality [Ta1] implies that for any $1>g>0$

$$\mathbb{P}\left[|X|^2 \geq g^2 \mathbb{E}|X|^2\right] \geq (1-g^2)^2 \frac{(\mathbb{E}|X|^2)^2}{\mathbb{E}X^4} \tag{6.27}$$

On the other hand, the Marcinkiewicz-Zygmund inequality (see [CT], page 367) yields that

$$\mathbb{E}|(x,\xi)|^4 \leq 4\mathbb{E}\left(\sum_\mu x_\mu^2 (\xi^\mu)^2\right)^2 \leq 4\|x\|_2^4 \tag{6.28}$$

while $\mathbb{E}X^2=\|x\|_2^2$. This gives (6.26).

■

Combining these two results we arrive at

Lemma 6.7. *Assume that* $Z = (x, \xi)$ *with* ξ^μ *as in Lemma 6.6 and symmetric. Let* $I \subset \{1, \dots M\}$ *and set* $\tilde{x}_\mu = x_\mu$, *if* $\mu \in I$, $\tilde{x}_\mu = 0$ *if* $\mu \notin I$. *Put* $\hat{x} = m - \tilde{x}$. *Assume* $\|\tilde{x}\|_2 \geq \|\hat{x}\|_2$. *Then*

$$\mathbb{E}(|Z| - x^*)^2 \geq \frac{1}{2}\left(\|x\|_2 - x^*\right)^2 + \frac{1}{500}\|\hat{x}\|_2^2 \tag{6.29}$$

Proof. We put $\epsilon = g\|\hat{x}\|_2$ in (6.25) and set $g^2 = \frac{1}{5}$. Then Lemma 6.6 gives the desired bound.

■

Lemma 6.8. *Let* Z *be as in Lemma 6.7. Then there is a finite positive constant* c *such that*

$$\mathbb{E}(|Z| - x^*)^2 \geq c \inf_{\mu, s} \|x - se^\mu x^*\|_2^2 \tag{6.30}$$

where $c \geq \frac{1}{4000}$.

Proof. We assume w.r.g. that $x_{\geq}|x_2| \geq |x_3| \geq \dots \geq |x_M|$ and distinguish three cases. **Case 1:** $x_1^2 \geq \frac{1}{2}\|x\|_2^2$. Here we set $\hat{x} \equiv (0, x_2, \dots, x_M)$. We have that

$$\begin{aligned}\|x - e^1 x^*\|_2^2 &= \|\hat{x}\|_2^2 + (x_1 - x^*)^2 \\ &\leq \|\hat{x}\|_2^2 + 2(x_1 - \|x\|_2)^2 + 2(\|x\|_2 - x^*)^2 \\ &\leq 3\|\hat{x}\|_2^2 + 2(\|x\|_2 - x^*)^2\end{aligned} \tag{6.31}$$

Therefore (6.29) yields

$$\begin{aligned}&\frac{1}{2}\left(\|x\|_2 - x^*\right)^2 + \frac{1}{500}\|\hat{x}\|_2^2 \geq \frac{1}{3 \cdot 500}\left(3\|\hat{x}\|_2^2 + 1500/2(\|x\|_2 - x^*)^2\right) \\ &\geq \frac{1}{3 \cdot 500}\|x - e^1 x^*\|_2^2\end{aligned} \tag{6.32}$$

which is the desired estimate in this case.

Case 2: $x_1^2 < \frac{1}{2}\|x\|_2^2$, $x_2^2 \geq \frac{1}{4}\|x\|_2^2$. Here we may choose $\hat{x} = (0, x_2, 0, \dots, 0)$. We set $\tilde{x} = (0, 0, x_3, \dots, x_M)$. Then

$$\|x - e^1 x^*\|_2^2 \leq (x_1 - x^*)^2 + \|\hat{x}\|_2^2 + \|\tilde{x}\|_2^2 \tag{6.33}$$

But $\|\tilde{x}\|_2^2 \leq \|x\|_2^2 - \frac{1}{2}\|x\|_2^2 \leq 2\|\hat{x}\|_2$ and

$$(x_1 - x^*)^2 \leq (\tfrac{1}{2}\|x\|_2 - x^*)^2 \leq 2(\|x\|_2 - x^*)^2 + \frac{1}{2}\|x\|_2^2 \frac{1}{2(1-\epsilon)} x^* \|\hat{x}\|_2$$
$$\leq 2(\|x\|_2 - x^*)^2 + 2\|\hat{x}\|_2^2 \tag{6.34}$$

Thus $\|x - e^1 x^*\|_2^2 \leq 4\|\hat{x}\|_2^2 + 2(\|x\|_2 - x^*)^2$, from which follows as above that

$$\frac{1}{2}\left(\|x\|_2 - x^*\right)^2 + \frac{1}{500}\|\hat{x}\|_2^2 \geq \frac{1}{4 \cdot 500}\|x - e^1 x^*\|_2^2 \tag{6.35}$$

Case 3: $x_1^2 < \frac{1}{2}\|x\|_2^2$, $x_2 < \frac{1}{4}\|x\|_2^2$. In this case it is possible to find $1 \leq t < M$ such that $\tilde{x} = (x_1, x_2, \dots, x_t, 0, \dots, 0)$ and $\hat{x} = (0, \dots, 0, x_{t+1}, \dots, x_M)$ satisfy $|\|\tilde{x}\|_2^2 - \|\hat{x}\|_2^2| \leq \frac{1}{4}\|x\|_2^2$. In particular, $\|\tilde{x}\|_2^2 \leq \frac{5}{3}\|\hat{x}\|_2^2$, and $(x^*)^2 \leq 2(\|x\|_2 - x^*)^2 + 2\|x\|_2^2 \leq 2(\|x\|_2 - x^*)^2 + \frac{16}{3}\|\hat{x}\|_2^2$. Thus

$$\|x - e^1 x^*\|_2^2 \leq (x^*)^2 + \|\tilde{x}\|_2^2 + \|\hat{x}\|_2^2 \leq 2(\|x\|_2 - x^*)^2 + 8\|\hat{x}\|_2^2 \tag{6.36}$$

and thus

$$\frac{1}{2}\left(\|x\|_2 - x^*\right)^2 + \frac{1}{500}\|\hat{x}\|_2^2 \geq \frac{1}{8 \cdot 500}\|x - e^1 x^*\|_2^2 \tag{6.37}$$

Choosing the worst estimate for the constants of all three cases proves the lemma. Proposition 6.4 follows by putting al together.

■

We thus want an estimate on the fluctuations of the last term in the r.h.s. of (6.13). We will do this uniformly inside balls $B_R(x) \equiv \left\{x' \in \mathbb{R}^M \mid \|x - x'\|_2 \leq R\right\}$ of radius R centered at the point $x \in \mathbb{R}^M$.

Proposition 6.9. *Assume* $\alpha \leq 1$. *Let* $\{\xi_i^\mu\}_{i=1,\dots,N;\mu=1,\dots,M}$ *be i.i.d. random variables taking values in* $[-1,1]$ *and satisfying* $\mathbb{E}\xi_i^\mu = 0$, $\mathbb{E}(\xi_i^\mu)^2 = 1$. *For any* $R < \infty$ *and* $x_0 \in \{sm^* e^\mu, s = \pm 1, \mu = 1, \dots, M\}$ *we have:*

i) For $p = 2$ *and* $\beta < 11/10$, *there exist finite numerical constants* C, *K such that* [7]

$$\mathbb{P}\left[\sup_{x\in B_R(x_0)}\left|\frac{1}{\beta N}\sum_{i=1}^{N}\{\mathbb{E}\ln\cosh(\beta(\xi_i,x))-\ln\cosh(\beta(\xi_i,x))\}\right| \geq C\sqrt{\alpha}R(m^*+R)+C\alpha m^*+4\alpha^3(m^*+R)\right] \leq \ln\left(\tfrac{R}{\alpha^3}\right)e^{-\alpha N}+e^{-\alpha^2 N} \tag{6.38}$$

ii) For $p\geq 3$ and $\beta>\beta_c$, and for $p=2$ and $\beta\geq 11/10$,

$$\mathbb{P}\left[\sup_{x\in B_R(x_0)}\left|\frac{1}{\beta N}\sum_{i=1}^{N}\{\mathbb{E}\ln\cosh(\beta(\xi_i,x))-\ln\cosh(\beta(\xi_i,x))\}\right| > C\sqrt{\alpha}R(R+\|x_0\|_2)+C\alpha+4\alpha^3\right] \leq \ln\left(\tfrac{R}{\alpha^3}\right)e^{-\alpha N}+e^{-\alpha^2 N} \tag{6.39}$$

Proof. We will treat the case (i) first, as it is the more difficult one. To prove Proposition 6.9 we will have to employ some quite heavy machinery, known as "chaining" in the probabilistic literature[8](see [LT]; we follow closely the strategy outlined in Section 11.1 of that book). Our problem is to estimate the probability of a supremum over an M-dimensional set, and the purpose of chaining is to reduce this to an estimate of suprema over countable (in fact finite) sets. Let us use in the following the abbreviations $f(z)\equiv\beta^{-1}\ln\cosh(\beta z)$ and $F(\xi,x)\equiv\frac{1}{N}\sum_{i=1}^{N}f((\xi_i,x))$. We us denote by $\mathcal{W}_{M,r}$ the lattice in $\mathbb{R}^M$ with spacing $r/\sqrt{M}$. Then, for any $x\in\mathbb{R}^M$ there exists a lattice point $y\in\mathcal{W}_{M,r}$ such that $\|x-y\|_2\leq r$. Moreover, the cardinality of the set of lattice points inside the ball $B_R(x_0)$ is bounded by[9]

$$\left|\mathcal{W}_{M,r}\bigcap B_R(x_0)\right|\leq e^{\alpha N[\ln(R/r)+2]} \tag{6.40}$$

We introduce a set of exponentially decreasing numbers $r_n=e^{-n}R$ (this choice is somewhat arbitrary and maybe not optimal) and set

---

[7] The absurd number 11/10 is of course an arbitrary choice. It so happens that, numerically, $m^*(1.1)\approx 0.5$ which seemed like a good place to separate cases.

[8] Physicists would more likely call this "coarse graining" of even "renormalization".

[9] For the (simple) proof see [BG5].

$\mathcal{W}(n) \equiv \mathcal{W}_{M,r_n} \cap B_{r_{n-1}}(0)$. The point is that if $r_0 = R$, any point $x \in B_R(x_0)$ can be subsequently approximated arbitrarily well by a sequence of points $k_n(x)$ with the property that

$$k_n(x) - k_{n-1}(x) \in \mathcal{W}(n) \tag{6.41}$$

As a consequence, we may write, for any n^* conveniently chosen,

$$\begin{aligned} |F(\xi,x) - \mathbb{E}F(\xi,x)| &\le |F(\xi,k_0(x)) - \mathbb{E}F(\xi,k_0(x))| \\ &+ \sum_{n=1}^{n^*} |F(\xi,k_n(x)) - F(\xi,k_{n-1}(x)) - \mathbb{E}(F(\xi,k_n(x)) - F(\xi,k_{n-1}(x)))| \\ &+ |F(\xi,x) - F(\xi,k_{n^*}(x)) - \mathbb{E}(F(\xi,x) - F(\xi,k_{n^*}(x)))| \end{aligned} \tag{6.42}$$

At this point it is useful to observe that the functions $F(\xi,x)$ have some good regularity properties as functions of x.

Lemma 6.10. *For any $x \in \mathbb{R}^M$ and $y \in \mathbb{R}^M$,*

$$\begin{aligned} &\frac{1}{\beta N}\left|\sum_{i=1}^{N} \{\ln\cosh(\beta(\xi_i,x)) - \ln\cosh(\beta(\xi_i,y))\}\right| \\ &\le \begin{cases} \|x-y\|_2 \max(\|x\|_2,\|y\|_2)\|A\| & \text{if } \beta < 11/10 \\ \|x-y\|_2\|A\|^{1/2} & \text{if } \beta \ge 11/10 \end{cases} \end{aligned} \tag{6.43}$$

Proof of Lemma 6.10. Defining F as before, we use the mean value theorem to write that, for some $0 < \theta < 1$,

$$\begin{aligned} |F(\xi,x) - F(\xi,y)| &= \frac{1}{N}\sum_{i=1}^{N}(x-y,\xi_i)f'((\xi_i, x+\theta(y-x))) \\ &\le \left[\frac{1}{N}\sum_{i=1}^{N}(x-y,\xi_i)^2\right]^{\frac12}\left[\frac{1}{N}\sum_{i=1}^{N}\Big(f'\big((\xi_i,x+\theta(y-x))\big)\Big)^2\right]^{\frac12} \end{aligned} \tag{6.44}$$

By the Schwarz inequality we have.

$$\frac{1}{N}\sum_{i=1}^{N}(x-y,\xi_i)^2 \le \|x-y\|_2^2\|A\| \tag{6.45}$$

To treat the last term in the r.h.s. of (6.44) we will distinguish the two cases $\beta \le \frac{11}{10}$ and $\beta \ge \frac{11}{10}$.

1) If $\beta \leq \frac{11}{10}$ we use that $|f'(x)| = |\tanh(\beta x)| \leq \beta|x|$ to write

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\Big(f'\big((\xi_i, x+\theta(y-x))\big)\Big)^2 &\leq \left(\tfrac{11}{10}\right)^2 \frac{1}{N}\sum_{i=1}^{N}(\xi_i, x+\theta(y-x))^2 \\ &\leq \left(\tfrac{11}{10}\right)^2 \frac{1}{N}\sum_{i=1}^{N}(\theta(x,\xi_i)^2 + (1-\theta)(y,\xi_i)^2) \\ &= \left(\tfrac{11}{10}\right)^2 (\theta\|x\|_2^2 + (1-\theta)\|y\|_2^2)\|A\| \\ &\leq \left(\tfrac{11}{10}\right)^2 \max(\|x\|_2^2, \|y\|_2^2)\|A\|\end{aligned} \tag{6.46}$$

which, together with (6.44) and (6.45), yields

$$|F(\xi,x) - F(\xi,y)| \leq \|x-y\|_2 \max(\|x\|_2, \|y\|_2)\|A\| \tag{6.47}$$

2) If A $\beta \geq \frac{11}{10}$ we use that $|f'(x)| \leq 1$ to get

$$|F(\xi,x) - F(\xi,y)| \leq \|x-y\|_2\|A\|^{1/2} \tag{6.48}$$

This concludes the proof of Lemma 6.10.

■

Lemma 6.10 implies that the last term in (6.42) satisfies

$$|F(\xi,x) - F(\xi,k_{n^*}(x)) - \mathbb{E}(F(\xi,x) - F(\xi,k_{n^*}(x)))| \leq const. r_{n^*} \tag{6.49}$$

which can be made irrelevantly small by choosing, e.g., $r_{n^*} = \alpha^3$. From this it follows that for any sequence of positive real numbers t_k

such that $\sum_{n=0}^{\infty} t_n \leq t$, we have the estimate

$$
\begin{aligned}
&\mathbb{P}\left[\sup_{x\in B_R(x_0)} |F(\xi,x)-\mathbb{E}F(\xi,x)| \geq t+\bar t + r_{n^*}\|x\|_2(\|A\|+\mathbb{E}\|A\|)\right]\\
&\leq \mathbb{P}\left[|F(\xi,x_0)-\mathbb{E}F(\xi,x_0)|\geq \bar t\right]\\
&+\mathbb{P}\Bigg[\sup_{x\in B_R(x_0)}\Big|F(\xi,k_0(x))-F(\xi,x_0)\\
&\qquad\qquad -\mathbb{E}\left(F(\xi,k_0(x))-F(\xi,x_0)\right)\Big|\geq t_0\Bigg]\\
&+\sum_{n=1}^{n^*}\mathbb{P}\Bigg[\sup_{x\in B_R(x_0)}\Big|F(\xi,k_n(x))-F(\xi,k_{n-1}(x))\\
&\qquad\qquad -\mathbb{E}(F(\xi,k_n(x))-F(\xi,k_{n-1}(x)))\Big|\geq t_n\Bigg]\\
&\leq \mathbb{P}\left[|F(\xi,x_0)-\mathbb{E}F(\xi,x_0)|\geq \bar t\right]\\
&+e^{M[\ln\frac{R}{r_0}+2]}\mathbb{P}\left[||F(\xi,x)-\mathbb{E}F(\xi,x)|\geq t_0\right]\\
&+\sum_{n=1}^{n^*}e^{M[\ln\frac{R}{r_n}+2]}\mathbb{P}\Big[|F(\xi,k_n(x))-F(\xi,k_{n-1}(x))\\
&\qquad\qquad -\mathbb{E}(F(\xi,k_n(x))-F(\xi,k_{n-1}(x)))|\geq t_n\Big]
\end{aligned}
\tag{6.50}
$$

where we used that the cardinality of the set

$$
\begin{aligned}
&\mathrm{Card}\Big\{|F(\xi,k_n(x))-F(\xi,k_{n-1}(x))\\
&\qquad -\mathbb{E}(F(\xi,k_n(x))-F(\xi,k_{n-1}(x)))|\,;x\in B_R(x_0)\Big\}\\
&\leq \mathrm{Card}\{\mathcal{W}_{M,r_{n-1}}\cap B_R(x_0)\}\leq \exp\left(M[\ln\tfrac{R}{r_n}+2]\right)
\end{aligned}
\tag{6.51}
$$

We must now estimate the probabilities occurring in (6.50); the first one is simple and could be bounded by using Talagrand' s Theorem 6.6 cited in Section 4. Unfortunately, for the other terms this does not seem possible since the functions involved there do not satisfy the hypothesis of convex level sets. We thus proceed by elementary methods, exploiting the particularly simple structure of the functions F as sums over independent terms. Thus we get from the exponential Tchebychev inequality that

$$\begin{aligned}
&\mathbb{P}\left[F(\xi,x)-F(\xi,y)-\mathbb{E}[F(\xi,x)-F(\xi,y)]\geq\delta\right]\\
&\leq\inf_{s\geq 0}e^{-\delta s}\prod_{i=1}^{N}\mathbb{E}e^{+\frac{s}{N}(f((\xi_i,x))-f((\xi_i,y))-\mathbb{E}[f((\xi_i,x))-f((\xi_i,y))])}\\
&\leq\inf_{s\geq 0}e^{-\delta s}\prod_{i=1}^{N}\left[1+\frac{s^2}{2N^2}\mathbb{E}\Big(f((\xi_i,x))-f((\xi_i,y))\right.\\
&\left.-\mathbb{E}[f((\xi_i,x))-f((\xi_i,y))]\Big)^2 e^{\frac{s}{N}|f((\xi_i,x))-f((\xi_i,y))-\mathbb{E}[f((\xi_i,x))-f((\xi_i,y))]|}\right]
\end{aligned}\tag{6.52}$$

We now use that both $|\tanh(\beta x)|\leq 1$ and $|\tanh(\beta x)|\leq\beta|x|$ to get that

$$|f((\xi_i,x))-f((\xi_i,y))|\leq|(\xi_i,(x-y))|\max_z|f'((\xi_i,z))|\leq|(\xi_i,(x-y))|\tag{6.53}$$

and

$$\begin{aligned}
|f((\xi_i,x))-f((\xi_i,y))&\leq|(\xi_i,(x-y))|\\
&\leq\beta|(\xi_i,(x-y))|\max\left(|(\xi_i,x)|,|(\xi_i,y)|\right)
\end{aligned}\tag{6.54}$$

The second inequality will only be used in the case $p\geq 3$ and if $\beta\leq 1.1$ Using the Schwarz inequality together with (6.53) we get

$$\begin{aligned}
&\mathbb{E}\left(f((\xi_i,x))-f((\xi_i,y))-\mathbb{E}[f((\xi_i,x))-f((\xi_i,y))]\right)^2\\
&\quad\times e^{\frac{s}{N}|f((\xi_i,x))-f((\xi_i,y))-\mathbb{E}[f((\xi_i,x))-f((\xi_i,y))]|}\\
&\leq\left[8\mathbb{E}\left(f((\xi_i,x))-f((\xi_i,y))\right)^4\right]^{1/2}\\
&\quad\times\left[\mathbb{E}e^{\frac{2s}{N}|f((\xi_i,x))-f((\xi_i,y))-\mathbb{E}(f((\xi_i,x))-f((\xi_i,y)))|}\right]^{1/2}\\
&\leq\sqrt{8}\left[\mathbb{E}(\xi_i,x-y)^4\right]^{1/2}\left[\mathbb{E}e^{\frac{2s}{N}|(\xi_i,(x-y))|}\right]^{1/2}e^{\frac{s}{N}\mathbb{E}|(\xi_i,(x-y))|}
\end{aligned}\tag{6.55}$$

Using (6.54) and once more the Schwarz inequality we get an alternative bound for this quantity by

$$\begin{aligned}
&\sqrt{8}\beta^2\left[\mathbb{E}(\xi_i,x-y)^8\right]^{1/4}\max\left(\left[\mathbb{E}(\xi_i,x)^8\right]^{1/4}\left[\mathbb{E}(\xi_i,y)^8\right]^{1/4}\right)\\
&\times\left[\mathbb{E}e^{\frac{2s}{N}|(\xi_i,(x-y))|}\right]^{1/2}e^{\frac{s}{N}\mathbb{E}|(\xi_i,(x-y))|}
\end{aligned}\tag{6.56}$$

The last line is easily bounded using essentially Khintchine resp. Marcinkiewicz-Zygmund inequalities (see [CT], pp. 366 ff.), in particular

$$\begin{aligned} \mathbb{E}|(\xi_i,x)| &\leq \sqrt{2}\|x\|_2 \qquad \text{reps.} \quad \|x\|_2/\sqrt{2} \quad \text{if } \xi_i^\mu \text{ are Bernoulli} \\ \mathbb{E}e^{s|(\xi_i,x)|} &\leq 2e^{\frac{s^2}{2}c} \qquad \text{with } c=1 \text{ if } \xi_i^\mu \text{ are Bernoulli} \\ \mathbb{E}(\xi_i,(x-y))^{2k} &\leq 2^{2k}k^k\|(x-y)\|_2^{2k} \qquad \text{no } 2^{2k} \text{ if } \xi_i^\mu \text{ are Bernoulli} \end{aligned} \tag{6.57}$$

Thus

$$\begin{aligned} &\mathbb{E}\left(f((\xi_i,x)) - f((\xi_i,y)) - \mathbb{E}[f((\xi_i,x)) - f((\xi_i,y))]\right)^2 \\ &\quad \times e^{\frac{s}{N}|f((\xi_i,x))-f((\xi_i,y))-\mathbb{E}[f((\xi_i,x))-f((\xi_i,y))]|} \\ &\leq \sqrt{8}\sqrt{32}\|(x-y)\|_2^2 e^{\frac{s}{N}\sqrt{2}\|x-y\|_2+c\frac{2s^2}{N^2}\|x-y\|_2^2} \\ &\quad \text{respectively} \\ &\leq \beta^2\sqrt{8}2^4 4^2\|x-y\|_2^2\left(\|x\|_2+\|y\|_2\right)^2 e^{\frac{s}{N}\sqrt{2}\|x-y\|_2+c\frac{2s^2}{N^2}\|x-y\|_2^2} \end{aligned} \tag{6.58}$$

In the Bernoulli case the constants can be improved to $2\sqrt{8}$ and $\sqrt{8}4^2$, resp., and $c=1$.
Inserting (6.58) into (6.52), using that $1+x\leq e^x$ and choosing s gives the desired bound on the probabilities. The trick here is not to be tempted to choose s depending on δ. Rather, depending on which bound we use, we choose $s=\frac{N\sqrt{\alpha}}{\|x-y\|_2}$ or $s=\frac{N\sqrt{\alpha}}{\|x-y\|_2\|(\|x\|_2+\|y\|_2)}$. This gives

$$\begin{aligned} &\mathbb{P}\left[F(\xi,x)-F(\xi,y)-\mathbb{E}[F(\xi,x)-F(\xi,y)]\geq\delta\right] \\ &\leq \exp\left(-N\frac{\sqrt{\alpha}\delta}{\|x-y\|_2}+8\alpha Ne^{\sqrt{2\alpha}+2c\alpha}\right) \end{aligned} \tag{6.59}$$

respectively

$$\begin{aligned} &\mathbb{P}\left[F(\xi,x)-F(\xi,y)-\mathbb{E}[F(\xi,x)-F(\xi,y)]\geq\delta\right] \\ &\leq \exp\left(-N\frac{\sqrt{\alpha}\delta}{\|x-y\|_2(\|x\|_2+\|y\|_2)}+\alpha N\beta^2\sqrt{8}2^4 4^2 e^{\frac{2\sqrt{\alpha}}{\|x\|_2+\|y\|_2}+\frac{c\alpha}{(\|x\|_2+\|y\|_2)^2}}\right) \end{aligned} \tag{6.60}$$

In particular

$$\begin{aligned}\mathbb{P}\Big[\big|F(\xi,k_n(x))-F(\xi,k_{n-1}(x))\\ -\mathbb{E}[|F(\xi,k_n(x))-F(\xi,k_{n-1}(x))]\big|\geq t_n\Big]\\ \leq 2\exp\left(-N\frac{t_n\sqrt{\alpha}}{r_{n-1}}+N8\alpha Ne^{\sqrt{2\alpha}+2c\alpha}\right)\end{aligned} \tag{6.61}$$

and

$$\begin{aligned}\mathbb{P}\Big[\big|F(\xi,k_n(x))-F(\xi,k_{n-1}(x))\\ -\mathbb{E}[|F(\xi,k_n(x))-F(\xi,k_{n-1}(x))]\big|\geq t_n\Big]\\ \leq 2\exp\left(-N\frac{t_n\sqrt{\alpha}}{r_{n-1}(R+\|x_0\|_2)}+\alpha N\beta^2\sqrt{8}2^4 4^2 e^{\frac{2\sqrt{\alpha}}{R+\|x_0\|_2}+\frac{c\alpha}{(\|x_0\|_2+R)^2}}\right)\end{aligned} \tag{6.62}$$

We also have

$$\begin{aligned}\mathbb{P}\left[|F(\xi,k_0(x))-F(\xi,x_0)-\mathbb{E}\left[F(\xi,k_0(x))-F(\xi,x_0)\right]|\geq t_0\right]\\ \leq 2\exp\left(-N\frac{t_0\sqrt{\alpha}}{R}+8\alpha Ne^{\sqrt{2\alpha}+2c\alpha}\right)\end{aligned} \tag{6.63}$$

and

$$\begin{aligned}\mathbb{P}\left[|F(\xi,k_0(x))-F(\xi,x_0)-\mathbb{E}\left[F(\xi,k_0(x))-F(\xi,x_0)\right]|\geq t_0\right]\\ \leq 2\exp\left(-N\frac{t_0\sqrt{\alpha}}{R(\|x_0\|_2+R)}+\alpha N\beta^2\sqrt{8}2^4 4^2 e^{\frac{2\sqrt{\alpha}}{R+\|x_0\|_2}+\frac{c\alpha}{(\|x_0\|_2+R)^2}}\right)\end{aligned} \tag{6.64}$$

Since $\|x_0\|_2+R\geq x^*$ so that $\frac{2\sqrt{\alpha}}{R+\|x_0\|_2}+\frac{c\alpha}{(\|x_0\|_2+R)^2}\leq\sqrt{2}\gamma x^*+c\gamma^2(x^*)2\leq c'\gamma$. Thus in the case $\beta\leq 1.1$ we may choose t_0 and t_n as

$$t_0=\sqrt{\alpha}R(\|x_0\|_2+R)\left[1+2+\beta^2 2^4 4^2\sqrt{8}e^{c'\gamma}+1\right] \tag{6.65}$$

and

$$t_n=\sqrt{\alpha}e^{-(n-1)}R(\|x_0\|_2+R)\left[n+2+\beta^2 2^4 4^2 e^{c'\gamma}+1\right] \tag{6.66}$$

Finally a simple estimate gives that (for $x_0 = \pm m^* e^\mu$)

$$\mathbb{P}\left[|F(\xi, x_0) - \mathbb{E}F(\xi, x_0)| \geq \bar{t}\right] \leq 2e^{-\frac{\bar{t}^2}{20(m^*)^2}N} \tag{6.67}$$

Choosing $\bar{t} = m^*\alpha$, setting $n^* = \ln\left(\frac{\alpha^3}{R}\right)$, and putting all this into (6.50) we get that

$$\begin{aligned}\mathbb{P}\Bigg[&\sup_{x\in B_R(x_0)} |F(\xi, x) - \mathbb{E}F(\xi, x)| \geq \\ &\sqrt{\alpha}R(\|x_0\|_2 + R)\Big(4 + \beta^2 2^4 4^2 \sqrt{8} e^{c'\gamma} \\ &+ 3 + \tfrac{e}{e-1} + \beta^2 2^4 4^2 e^{c'\gamma}\Big) + m^*\alpha + \alpha^3(\|x_0\|_2 + R)\Bigg] \\ &\leq \ln\left(\tfrac{\alpha^3}{R}\right) e^{-\alpha N} + 2e^{-\alpha^2 N/20}\end{aligned} \tag{6.68}$$

This proves part (i) of Proposition 6.9 and allows us to estimate the constant C in (6.38). In the same way, but using (6.61) and (6.63), we get the analogous bound in case (ii), namely

$$\begin{aligned}\mathbb{P}\Bigg[&\sup_{x\in B_R(x_0)} |F(\xi, x) - \mathbb{E}F(\xi, x)| \geq \\ &\sqrt{\alpha}R(\|x_0\|_2 + R)\left(4 + 8e^{c'\gamma} + 3 + \tfrac{e}{e-1} + 8e^{c'\gamma}\right) + 4\alpha^3\Bigg] \\ &\leq \ln\left(\tfrac{\alpha^3}{R}\right) e^{-\alpha N}\end{aligned} \tag{6.69}$$

This concludes the proof of Proposition 6.9.

■

Remark. The reader might wonder whether this heavy looking chaining machinery used in the proof of Proposition 6.9 is really necessary. Alternatively, one might use just a single lattice approximation and use Lemma 6.10 to estimate how far the function can be from the lattice values. But for this we need at least a lattice with $r = \sqrt{\alpha}$, and this would force us to replace the $\sqrt{\alpha}$ terms in (6.38) and (6.39) by $\sqrt{\alpha|\ln\alpha|}$. While this may not look too serious, it would certainly spoil the correct scaling between the critical α and $\beta - 1$ in the case $p = 2$.

We are now ready to conclude the proof of Proposition 6.1. To do this, we consider the $2M$ sectors in which sx^*e^μ is the closest of all the points sx^*e^ν and use Proposition 6.9 with $x_0 = se^\mu x^*$ and R the distance from that point. One sees easily that if that distance is sufficiently large (as stated in the theorem), then with probability exponentially close to one, the modulus of the last term in (6.13) is bounded by one half of the lower bound on the first term given by Proposition 6.4. Since it is certainly enough to consider a discrete set of radii (e.g. take $R \in \mathbb{Z}/N$), and the individual estimates fail only with a probability of order $\exp(-\alpha N)$, it is clear that the estimates on ψ hold indeed uniformly in x with probability exponentially close to one. This concludes the proof of Proposition 6.1.

■

7. Local analysis of Φ

To obtain more detailed information on the Gibbs measures requires to look more precisely at the behaviour of the functions $\Phi_{p,\beta,N}(m)$ in the vicinities of points $\pm m^*(\beta)e^\mu$. Such an analysis has first been performed in the case of the standard Hopfield model in [BG5]. The basic idea was simply to use second order Taylor expansions combined with careful probabilistic error estimates. One can certainly do the same in the general case with sufficiently smooth energy function $E_M(m)$, but since results (and to some extent techniques) depend on specific properties of these functions, we restrict our attention again to the cases where $E_M(m) = \frac{1}{p}\|m\|_p^p$, with $p \geq 2$ integer, as in the previous section. For reasons that will become clear in a moment, the (most interesting) case $p = 2$ is special, and we consider first the case $p \geq 3$. Also throughout this section, the ξ_i^μ take the values ± 1.

7.1. The case $p \geq 3$.

As a matter of fact, this case is "misleadingly simple"[7]. Recall that we deal with the function $\Phi_{p,\beta,N}(m)$ given by (6.2). Let us consider without restriction of generality the vicinity of m^*e^1. Write $m = m^*e^1 + v$ where v is assumed "small", e.g. $\|v\|_2 \leq \epsilon < m^*$. We have to consider mainly the regions over which Proposition 6.1 does not give control, i.e. where $\|\operatorname{sign}(m)|m|^{p-1} - e^1(m^*)^{p-1}\|_2 \leq c_1\sqrt{\alpha}$ (recall (6.6)). In terms of the variable v this condition implies that both $|v_1|^2 \leq C\sqrt{\alpha}$ and $\|\hat{v}\|_{2p-2}^{2p-2} \leq C\sqrt{\alpha}$ for some constant C (depending on

[7] But note that we consider only the case $M \sim \alpha N$ rather than $M \sim \alpha N^{p-1}$

p), where we have set $\hat{v} = (0, v_2, v_3, \dots, v_M)$. Under these conditions we want to study

$$\begin{aligned}\Phi_{p,\beta,N}(m^*e^1+v)-\Phi_{p,\beta,N}(m^*e^1) &= \frac{1}{q}\left((m^*+v_1)^p-(m^*)^p+\|\hat{v}\|_p^p\right)\\ &-\frac{1}{\beta N}\sum_{i=1}^N\left[\ln\cosh\left(\beta((m^*+v^1)^{p-1}+\sum_{\mu\geq 2}\hat{\xi}_i^\mu v_\mu^{p-1})\right)\right.\\ &\left.-\ln\cosh(\beta(m^*)^{p-1})\right]\end{aligned} \tag{7.1}$$

where we have set $\hat{\xi}_i^\mu \equiv \xi_i^1\xi_i^\mu$. The crucial point is now that we can expand each of the terms in the sum over i without any difficulty: for $|(m^*+v_1)^{p-1}-(m^*)^{p-1}| \leq |v_1|(p-1)(m^*+|v_1|)^{p-2} \leq C|v_1|$, and, more importantly, the Hölder inequality gives

$$\left|\sum_{\mu\geq 2}\hat{\xi}_i^\mu \text{sign}(v_\mu)|v_\mu|^{p-1}\right| \leq \|\hat{v}\|_2^2\|\hat{v}\|_\infty^{p-3} \tag{7.2}$$

As explained earlier, we need to consider only v for which $\|v\|_2 \leq 2$, and $\|\hat{v}\|_\infty \leq \|\hat{v}\|_{2p-2} \leq (C\sqrt{\alpha})^{1/(2p-2)}$ is small on the set we consider. Such a result does not hold if $p=2$, and this makes the whole analysis much more cumbersome in that case — as we shall see.

What we can already read off from (7.1) otherwise is that v_1 and $\hat{v}$ enter in a rather asymmetric way. We are thus well-advised to treat $|v_1|$ and $\|\hat{v}\|_2$ as independent small parameters. Expanding, and using that $m^* = \tanh(\beta(m^*)^{p-1})$ gives therefore

$$\begin{aligned}&\Phi_{p,\beta,N}(m^*e^1+v)-\Phi_{p,\beta,N}(m^*e^1)\\ &= v_1^2\frac{p-1}{2}(m^*)^{p-2}\left[1-\beta(1-(m^*)^2)(m^*)^{p-2}(p-1)\right]\\ &+\frac{1}{q}\|\hat{v}\|_p^p-\frac{\beta}{2}(1-(m^*)^2)\left(\text{sign}\,(\hat{v})|\hat{v}|^{p-1},\frac{\xi^T\xi}{N}\text{sign}\,(\hat{v})|\hat{v}|^{p-1}\right)\\ &-\frac{1}{N}\sum_{i=1}^N\left(\hat{\xi}_i,\,\text{sign}\,(\hat{v})|\hat{v}|^{p-1}\right)\left[m^*+\frac{\beta}{2}(1-(m^*)^2)(m^*)^{p-2}v_1\right]\\ &+R(v)\end{aligned} \tag{7.3}$$

where

$$
\begin{aligned}
|R(v)| &\leq |v_1|^3 \frac{(p-1)(p-2)(p-3)}{6}(m^*+|v_1|)^{p-3} \\
&+ \frac{2^{9/4}}{6}\left[|v_1|^3(p-1)^3(m^*+|\epsilon|)^{3(p-2)} + \frac{1}{N}\sum_{i=1}^N |(\hat\xi_i, \operatorname{sign}(\hat v)|\hat v|^{p-1})|^3\right] \\
&\times \frac{2\beta^2 \tanh\beta\left((m^*+|v_1|)^{p-1} + \|\hat v\|_2^2\|\hat v\|_\infty^{p-3}\right)}{\cosh^2\beta\left((m^*-|v_1|)^{p-1} - \|\hat v\|_2^2\|\hat v\|_\infty^{p-3}\right)}
\end{aligned}
\tag{7.4}
$$

where the last factor is easily seen to be bounded uniformly by some constant, provided $|v_1|$ and $\|\hat v\|_2$ are small compared to $m^*(\beta)$. Recall that the latter is, for $\beta \geq \beta_c$, bounded away from zero if $p \geq 3$. (Note that we have used that for positive a and b, $(a+b)^3 \leq 2^{9/4}(a^3+b^3)$). Note further that

$$
\begin{aligned}
&\left(\operatorname{sign}(\hat v)|\hat v|^{p-1}, \frac{\xi^T\xi}{N}\operatorname{sign}(\hat v)|\hat v|^{p-1}\right) \leq \|A(N)\|\sum_{\mu\geq 2} v_\mu^{2p-2} \\
&\leq \|A(N)\|\|\hat v\|_p^p\|\hat v\|_\infty^{p-2}
\end{aligned}
\tag{7.5}
$$

and

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^N |(\hat\xi_i, \operatorname{sign}(\hat v)|\hat v|^{p-1})|^3 &\leq \frac{1}{N}\sum_{i=1}^N |(\hat\xi_i, \operatorname{sign}(\hat v)|\hat v|^{p-1})|^2\|\hat v\|_2^2\|\hat v\|_\infty^{p-3} \\
&\leq \|A(N)\|\|\hat v\|_p^p\|\hat v\|_\infty^{2p-5}\|\hat v\|_2^2
\end{aligned}
\tag{7.6}
$$

so that in fact

$$
\begin{aligned}
&\Phi_{p,\beta,N}(m^*e^1+v) - \Phi_{p,\beta,N}(m^*e^1) \\
&= v_1^2\frac{p-1}{2}(m^*)^{p-2}\left[1-\beta(1-(m^*)^2)(m^*)^{p-2}(p-1)\right] + \frac{1}{q}\|\hat v\|_p^p \\
&- \frac{1}{N}\sum_{i=1}^N\left(\hat\xi_i, \operatorname{sign}(\hat v)|\hat v|^{p-1}\right)\left[m^* + \frac{\beta}{2}(1-(m^*)^2)(m^*)^{p-2}v_1\right] + R(v)
\end{aligned}
\tag{7.7}
$$

where $|\tilde R(v)| \leq c\left(|v_1|^3 + \|\hat v\|_p^p\|\hat v\|_\infty^{p-2}\right)$.

These bounds give control over the local minima near the Mattis states. In fact, we can compute easily the first corrections to their precise (random) positions. The approximate equations for them have the form

$$
\begin{aligned}
v_1 &= c_1(\beta)\frac{1}{\sqrt{N}}(z, \operatorname{sign}(\hat{v})|\hat{v}|^{p-1}) \\
v_\mu &= \frac{1}{\sqrt{N}}z_\mu(m^* + c_2(\beta)v_1), \quad \text{for } \mu \neq 1
\end{aligned}
\tag{7.8}
$$

where $z_\mu = \frac{1}{\sqrt{N}}\sum_i \hat{\xi}_i^\mu$ and $c_1(\beta), c_2(\beta)$ are constants that can be read off (7.7). These equations are readily solved and give

$$
\begin{aligned}
v_1 &= \frac{1}{\sqrt{N}}c_1 X \\
v_\mu &= \frac{1}{\sqrt{N}}z_\mu\left(m^* + \frac{c_1c_2}{\sqrt{N}}X\right)
\end{aligned}
\tag{7.9}
$$

where X is the solution of the equation

$$
X = \frac{1}{N^{(p-1)/2}}\|z\|_p^p\left(m^* + \frac{c_1c_2}{\sqrt{N}}X\right)^p \tag{7.10}
$$

Note that for N large,

$$
\mathbb{E}\|z\|_p^p \approx \frac{M}{N^{(p-1)/2}}\frac{(1-(-1)^p)2^{p/2}\Gamma\left(\frac{1+p}{2}\right)}{2\sqrt{\pi}} \tag{7.11}
$$

Moreover, an estimate of Newman ([N], Proposition 3.2) shows that

$$
\mathbb{P}\left[\left|\|z\|_p^p - \mathbb{E}\|z\|_p^p\right| > \gamma M^{\frac{p-2}{2p-2}}\right] \leq 2e^{-c_p(\gamma)M^{1/(p-1)}} \tag{7.12}
$$

for some function $c_p(\gamma) > 0$ for $\gamma > 0$. This implies in particular that $\frac{X}{\sqrt{N}}$ is sharply concentrated around the value $\frac{M}{N^{p/2}}$ (which tends to zero rapidly for our choices of M). Thus under our assumptions on M, the location of the minimum in the limit as N tends to infinity is $v_1 = 0$ and $v_\mu = \frac{m^*}{\sqrt{N}}z_\mu$, and at this point $\Phi_{p,\beta,N}(m^*e^1 + v) - \Phi_{p,\beta,N}(m^*e^1) = O\left(M/N^{p/2}\right)$.

On the other hand, for $\|\hat{v}\|_p \geq 2\sqrt{\alpha}(m^* + c)$,

$$
\Phi_{p,\beta,N}(m^*e^1 + v) - \Phi_{p,\beta,N}(m^*e^1) \geq c_1v_1^2 + c_3\|\hat{v}\|_p^p > 0 \tag{7.13}
$$

which completes the problem of localizing the minima of Φ in the case $p \geq 3$. Note the very asymmetric shape of the function in their vicinity.

7.2 The case $p = 2$.

The case of the standard Hopfield model turns out to be the more difficult, but also the most interesting one. The major source of this is the fact that an inequality like (7.2) does *not* hold here. Indeed, it is easy to see that there exist v such that $\left|\sum_\mu \hat{\xi}_i^\mu v_\mu\right| = \sqrt{M}\|\hat{v}\|_2$. The idea, however, is that this requires that v be adapted to the particular $\hat{\xi}_i$, and that it will be impossible, typically, to find a v such that for *many* indices, i, $\left|\sum_\mu \hat{\xi}_i^\mu v_\mu\right|$ would be much bigger than $\|v\|_2$ and to take advantage of that fact. The corresponding analysis has been carried out in [BG5] and we will not repeat all the intermediate technical steps here. We will however present the main arguments in a streamlined form. The key idea is to perform a Taylor expansion like in the previous case only for those indices i for which (ξ_i, v) is small, and to use a uniform bound for the others. The upper and lower bounds must be treated slightly differently, so let us look first at the lower bound.

The uniform bound we have here at our disposal is that

$$-\frac{1}{\beta}\ln\cosh\beta x \geq \frac{(m^*)^2}{2} - \frac{1}{\beta}\ln\cosh\beta m^* - \frac{x^2}{2} \tag{7.14}$$

Using this we get, for suitably chosen parameter $\tau > 0$, by a simple computation that for some $0 \leq \theta \leq 1$,

$$\begin{aligned}
&\Phi_{2,\beta,N}(m^*e^1 + v) - \Phi_{2,\beta,N}(m^*e^1) \\
&\geq \frac{1}{2}\|v\|_2^2 - \frac{1}{2}\beta(1-(m^*)^2)\frac{1}{N}\sum_{i=1}^N(\xi_i, v)^2 - \frac{m^*}{N}\sum_{i=1}^N(\hat{\xi}_i, \hat{v}) \\
&- \frac{1}{6}\frac{1}{N}\sum_{i=1}^N \mathbb{I}_{\{|(\xi_i,v)|\leq\tau m^*\}}|(\xi_i,v)|^3 2\beta^2\frac{\tanh\beta(m^*+\theta(\xi_i,v))}{\cosh^2\beta(m^*+\theta(\xi_i,v))} \\
&- \frac{1}{2}(1-\beta(1-(m^*)^2))\frac{1}{N}\sum_{i=1}^N \mathbb{I}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i,v)^2
\end{aligned} \tag{7.15}$$

The first two lines are the main second order contributions. The third line is the standard third order remainder, but improved by the characteristic function that forces (ξ_i, v) to be small. The last line is the price we have to pay for that, and we will have to show that with large probability this is also very small. This is the main “difficulty”; for the third order remainder one may use simply that

$$
\begin{aligned}
&\frac{1}{6}\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\{|(\xi_i,v)|\leq\tau m^*\}}|(\xi_i,v)|^3 2\beta^2\frac{\tanh\beta(m^*+\theta(\xi_i,v))}{\cosh^2\beta(m^*+\theta(\xi_i,v))}\\
&\leq\frac{1}{2N}\sum_{i=1}^{N}(\xi_i,v)^2\tau m^*\frac{1}{3}\beta^2\frac{\tanh\beta(m^*(1+\tau))}{\cosh^2\beta(m^*(1-\tau))}\\
&\leq\frac{1}{2N}\sum_{i=1}^{N}(\xi_i,v)^2\tau(1+\tau)(m^*)^2\frac{\beta^3}{3}\cosh^{-2}\beta(m^*(1-\tau))
\end{aligned}
\tag{7.16}
$$

For τ somewhat small, say $\tau\leq 0.1$, it is not difficult to see that $\frac{\beta^3}{3}\cosh^{-2}\beta(m^*(1-\tau))$ is bounded uniformly in β by a constant of order 1. Thus we can for our purposes use

$$
\begin{aligned}
&\frac{1}{6N}\sum_{i=1}^{N}\mathbb{1}_{\{|(\xi_i,v)|\leq\tau m^*\}}|(\xi_i,v)|^3 2\beta^2\frac{\tanh\beta(m^*+\theta(\xi_i,v))}{\cosh^2\beta(m^*+\theta(\xi_i,v))}\\
&\quad\leq\tau(1+\tau)(m^*)^2\frac{1}{2N}\sum_{i=1}^{N}(\xi_i,v)^2
\end{aligned}
\tag{7.17}
$$

which produces just a small perturbation of the quadratic term. Setting

$$
X_a(v)\equiv\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\{|(\xi_i,v)|>a\}}(\xi_i,v)^2 \tag{7.18}
$$

we summarize our finding so far as

Lemma 7.1. *There exists $\tau_c>0$ (≈ 0.1) such that for all β, for $\tau\leq\tau_c$,*

$$
\begin{aligned}
&\Phi_{2,\beta,N}(m^*e^1+v)-\Phi_{2,\beta,N}(m^*e^1)\\
&\geq\frac{1}{2}\left(v,\left[\mathbb{1}-(\beta(1-(m^*)^2)+\tau(1+\tau)(m^*)^2)\frac{\xi^T\xi}{N}\right]v\right)-\frac{m^*}{N}\sum_{i=1}^{N}(\hat{\xi}_i,\hat{v})\\
&-\frac{1}{2}(1-\beta(1-(m^*)^2))X_{\tau m^*}(v)
\end{aligned}
\tag{7.19}
$$

Before turning to the study of $X_a(v)$, we derive corresponding lower bounds. For this we need a complement to (7.14). Using the Taylor formula with second order remainder we have that for some $\tilde{x}$

$$
\begin{aligned}
-\frac{1}{\beta}\ln\cosh\beta x &\leq \frac{(m^*)^2}{2} - \frac{1}{\beta}\ln\cosh\beta m^* - \frac{x^2}{2} \\
&\quad + \frac{(x-m^*)^2}{2}\left[1-\beta\left(1-\tanh^2\beta(\tilde{x})\right)\right] \\
&\leq \frac{(m^*)^2}{2} - \frac{1}{\beta}\ln\cosh\beta m^* - \frac{x^2}{2} + \frac{(x-m^*)^2}{2}
\end{aligned}
\tag{7.20}
$$

By a similar computation as before this gives

Lemma 7.2. *There exists $\tau_c > 0$ (≈ 0.1) such that for all β, for $\tau \leq \tau_c$,*

$$
\begin{aligned}
&\Phi_{2,\beta,N}(m^*e^1 + v) - \Phi_{2,\beta,N}(m^*e^1) \\
&\leq \frac{1}{2}\left(v, \left[\mathbb{1} - (\beta(1-(m^*)^2) - \tau(1+\tau)(m^*)^2)\frac{\xi^T\xi}{N}\right] v\right) \\
&- \frac{m^*}{N}\sum_{i=1}^{N}(\hat{\xi}_i, \hat{v}) + \frac{1}{2}\beta(1-(m^*)^2))X_{\tau m^*}(v)
\end{aligned}
\tag{7.21}
$$

To make use of these bounds, we need to have uniform control over the $X_a(v)$. In [BG5] we have proven for this the following

Proposition 7.3. *Define*

$$
\begin{aligned}
\Gamma(\alpha, a/\rho) = &\left(2\sqrt{2\sqrt{2}e^{-\frac{(1-3\sqrt{\alpha})^2}{(1-\sqrt{\alpha})^2}\frac{a^2}{4\rho^2}} + \alpha(|\ln\alpha|+2)} + \alpha\sqrt{1+r(\alpha)}\right)^2 \\
&+ 2\alpha^2(1+r(\alpha)) + \tfrac{1}{2}\alpha\left(2e^{-\frac{a^2}{\alpha\rho^2}} + 2\sqrt{3\alpha(|\ln\alpha|+2)}\right)
\end{aligned}
\tag{7.22}
$$

Then

$$
\mathbb{P}\left[\sup_{v\in B_\rho} X_a(v) \geq \rho^2\Gamma(\alpha, a/\rho)\right] \leq e^{-\alpha N} + \mathbb{P}[\|A - \mathbb{1}\| \geq r(\alpha)] \tag{7.23}
$$

We see that $\Gamma(\alpha, a, \rho)$ is small if α is small and ρ^2 is small compared to a which for us is fine: we need the proposition with $a = \tau m^*$ and with $\rho \leq \gamma m^* c_1$, where γ is our small parameter. The proof of this proposition can be found in [BG5]. It is quite technical and uses a chaining procedure quite similar to the one used in Section 6 in the proof of Proposition 6.9. Since we have not found a way to simplify or improve it, we will not reproduce it here. Although in [BG5] only the Bernoulli case was considered, but the extension to centered bounded ξ_i^μ poses no particular problems and can be left to the reader; of course constants will change, in particular if the variables are asymmetric.

The expression for $\Gamma(\alpha, a, \rho)$ looks quite awful. However, for α small (which is all we care for here), it is in fact bounded by

$$\Gamma(\alpha, a/\rho) \leq C\left[e^{-(1-2\sqrt{\alpha})^2 \frac{a^2}{4\rho^2}} + \alpha(|\ln\alpha| + 2)\right] \tag{7.24}$$

with $C \approx 25$. We should now choose τ in an optimal way. It is easy to see that in (7.19), for $\rho \leq c\gamma m^*$, this leads to $\tau \sim \gamma\sqrt{|\ln\gamma|}$, uniformly in $\beta > 1$. This uses that the coefficient of $X_{\tau m^*}(v)$ is proportional to $(m^*)^2$. Unfortunately, that is not the case in the upper bound of Lemma 7.2, so that it turns out that while this estimate is fine for β away from 1 (e.g. $\beta > 1.1$, which means $m^* > 0.5$), for β near one we have been too careless! This is only just: replacing $\beta(1-\tanh^2\beta\tilde{x})$ by zero and hoping to get away with it was overly optimistic. This is, however, easily remedied by dealing more carefully with that term. We will not give the (again somewhat tedious) details here; they can be found in [BG5]. We just quote from [BG5] (Theorem 4.9)

Lemma 7.4. *Assume that $\beta \leq 1.1$. Then there exists $\tau_c > 0$ (≈ 0.1) such that for $\tau \leq \tau_c$,*

$$\begin{aligned}
&\Phi_{2,\beta,N}(m^*e^1 + v) - \Phi_{2,\beta,N}(m^*e^1) \\
&\leq \frac{1}{2}\left(v, \left[\mathbb{1} - (\beta(1-(m^*)^2) - \tau(1+\tau)(m^*)^2)\frac{\xi^T\xi}{N}\right]v\right) - \frac{m^*}{N}\sum_{i=1}^N(\hat{\xi}_i, \hat{v}) \\
&+ \frac{1}{2}(m^*)^2\|v\|_2^2\left(\gamma + 240e^{-(1-2\sqrt{\alpha})^2\frac{(m^*)^2}{4\|v\|_2^2}}\right)
\end{aligned} \tag{7.25}$$

For the range of v we are interested in, all these bounds combine to

Theorem 7.5. *For all $\beta > 1$ and for all $\|v\|_2 \leq c\gamma m^*$, there exists a* *finite numerical constant $0 < C < \infty$ such that*

$$\left| \Phi_{2,\beta,N}(m^*e^1 + v) - \Phi_{2,\beta,N}(m^*e^1) \right.$$
$$\left. - \frac{1}{2}\left[1 - \beta(1-(m^*)^2)\right] \|v\|_2^2 - \frac{m^*}{N}\sum_{i=1}^N (\hat{\xi}_i, \hat{v}) \right| \leq \gamma\sqrt{|\ln\gamma|}C(m^*)^2\|v\|_2^2 \tag{7.26}$$

with probability greater that $1 - e^{-\alpha N}$.

As an immediate consequence of this bound we can localize the position of the minima of Φ near m^*e^μ rather precisely.

Corollary 7.6. *Let v^* denote the position of the lowest minimum of the function $\Phi_{2,\beta,N}(m^*e^1 + v)$ in the ball $\|v\|_2 \leq c\gamma m^*$. Define the vector $z^{(\nu)} \in \mathbb{R}^M$ with components*

$$z_\mu^{(\nu)} \equiv \begin{cases} \frac{1}{N}\sum_i \xi_i^\nu \xi_i^\mu, & \text{for } \mu \neq \mu \\ 0, & \text{for } \mu = \nu \end{cases} \tag{7.27}$$

There exists a finite constant C such that

$$\left\| v^* - \frac{m^*}{1-\beta(1-(m^*)^2)} z^{(1)} \right\|_2 \leq C\ \gamma\sqrt{|\ln\gamma|}\frac{\|z^{(1)}\|_2 (m^*)^3}{(1-\beta(1-(m^*)^2))^2} \tag{7.28}$$

with the same probability as in Theorem 7.5. Moreover, with probability greater than $1 - e^{-4M/5}$,

$$\|z^{(1)}\|_2 \leq 2\sqrt{\alpha} \tag{7.29}$$

so that in fact

$$\left\| v^* - \frac{m^*}{1-\beta(1-(m^*)^2)} z^{(1)} \right\|_2 \leq C\gamma^2\sqrt{|\ln\gamma|}m^* \tag{7.30}$$

Proof. (7.28) is straightforward from Theorem 7.5. The bound on $\|z^{(1)}\|_2$ was given in [BG5], Lemma 4.11 and follows from quite straightforward exponential estimates.

■

Remark. We will see in the next section that for β not too large (depending on α), there is actually a unique minimum for $\|v\|_2 \leq c\gamma m^*$.

8. Convexity, the replica symmetric solution, convergence

In this final section we restrict our attention to the standard Hopfield model. Most of the results presented here were inspired by a recent paper of Talagrand [T4].

In the last section we have seen that the function Φ is locally bounded from above and below by quadratic functions. A natural question is to ask whether this function may even be locally covex. The following theorem (first proven in [BG5]) shows that this is true under some further restrictions on the range of the parameters.

Theorem 8.1. *Assume that $1<\beta<\infty$. If the parameters α,β,ρ are such that for $\epsilon>0$,*

$$\begin{aligned}\inf_{\tau}\Big(&\beta(1-\tanh^2(\beta m^*(1-\tau)))(1+3\sqrt{\alpha})\\ &+2\beta\tanh^2(\beta m^*(1-\tau))\Gamma(\alpha,\tau m^*/\rho)\Big)\leq 1-\epsilon\end{aligned}\tag{8.1}$$

Then with probability one for all but a finite number of indices N, $\Phi_{N,\beta}[\omega](m^*e^1+v)$ is a twice differentiable and strictly convex function of v on the set $\{v:\|v\|_2\leq\rho\}$, and $\lambda_{min}\left(\nabla^2\Phi_{N,\beta}[\omega](m^*e^1+v)\right)>\epsilon$ on this set.

Remark. The theorem should of course be used for $\rho=c\gamma m^*$. One checks easily that with such ρ, the conditions mean: (i) For β close to 1: γ small and, (ii) For β large: $\alpha\leq c\beta^{-1}$.

Remark. In deviation from our general policy not to speak about the high-temperature regime, we note that it is of course trivial to show that $\lambda_{min}\left(\nabla^2\Phi_{N,\beta}[\omega](m)\right)\geq\epsilon$ for all m if $\beta\leq\frac{1-2\epsilon}{(1+\sqrt{\alpha})^2}$. Therefore all the results below can be easily extended into that part of the high-temperature regime. Note that this does *not* cover *all* of the high temperature phase, which starts already at $\beta^{-1}=1+\sqrt{\alpha}$.

Proof. The differentiability for fixed N is no problem. The non-trivial assertion of the theorem is the local convexity. Since $\frac{d^2}{dx^2}\ln\cosh(\beta x)=$

$\beta\left(1-\tanh^2(\beta x)\right)$ we get

$$\begin{aligned}\nabla^2\Phi(m^*e^1+v) &= \mathbb{1}-\frac{1}{N}\sum_{i=1}^N f''_\beta(m^*\xi_i^1+(\xi_i,v))\xi_i^T\xi_i\\ &= \mathbb{1}-\frac{\beta}{N}\sum_{i=1}^N\xi_i^T\xi_i+\frac{\beta}{N}\sum_i\xi_i^T\xi_i\tanh^2(\beta(m^*\xi_i^1+(\xi_i,v)))\\ &\geq \mathbb{1}-\beta\frac{\xi^T\xi}{N}+\frac{\beta}{N}\sum_i\xi_i^T\xi_i\mathbb{1}_{\{|(\xi_i,v)|\leq\tau m^*\}}\tanh^2(\beta m^*(1-\tau))\\ &= \mathbb{1}-\beta\left[1-\tanh^2(\beta m^*(1-\tau))\right]\frac{\xi^T\xi}{N}\\ &\quad -\beta\tanh^2(\beta m^*(1-\tau))\frac{1}{N}\sum_i\xi_i^T\xi_i\mathbb{1}_{\{|(\xi_i,v)|>\tau m^*\}}\end{aligned} \tag{8.2}$$

Thus

$$\begin{aligned}\lambda_{min}\left(\nabla^2\Phi(m^*e^1+v)\right) &\geq 1-\beta\left[1-\tanh^2(\beta m^*(1-\tau))\right]\|A(N)\|\\ &\quad -\beta\tanh^2(\beta m^*(1-\tau))\left\|\tfrac{1}{N}\sum_{i=1}^N\mathbb{1}_{\{|(\xi_i,v)|>\tau m^*\}}\xi_i^T\xi_i\right\|\end{aligned} \tag{8.3}$$

What we need to do is to estimate the norm of the last term in (8.3). Now,

$$\begin{aligned}&\sup_{v\in B_\rho}\left\|\tfrac{1}{N}\sum_{i=1}^N\mathbb{1}_{\{|(\xi_i,v)|>\tau m^*\}}\xi_i^T\xi_i\right\|\\ &=\sup_{v\in B_\rho}\sup_{w:\|w\|_2=\rho}\tfrac{1}{\rho^2}\tfrac{1}{N}\sum_{i=1}^N\mathbb{1}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i,w)^2\\ &\leq\tfrac{1}{\rho^2}\sup_{v\in B_\rho}\sup_{w\in B_\rho}\tfrac{1}{N}\sum_{i=1}^N\mathbb{1}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i,w)^2\end{aligned} \tag{8.4}$$

To deal with this last expression, notice that

$$\begin{aligned}&(\xi_i,w)^2\\ &\quad=\mathbb{1}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i,w)^2\left(\mathbb{1}_{\{|(\xi_i,w)|<|(\xi_i,v)|\}}+\mathbb{1}_{\{|(\xi_i,w)|\geq|(\xi_i,v)|\}}\right)\\ &\quad\leq\mathbb{1}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i,v)^2+\mathbb{1}_{\{|(\xi_i,w)|>\tau m^*\}}(\xi_i,w)^2\end{aligned} \tag{8.5}$$

Thus

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_{\{|(\xi_i,v)|>\tau m^*\}}(\xi_i,w)^2 = X_{\tau m^*}(v) + X_{\tau m^*}(w) \tag{8.6}$$

and so we are reduced to estimating the same quantities as in Section 7. Thus using Proposition 7.3 and the estimate (4.12) with $\epsilon = \sqrt{\alpha}$, we obtain therefore that with probability greater than $1 - e^{-const.\alpha N}$ for all v with norm less than ρ,

$$\begin{aligned}\lambda_{min}\left(\nabla^2\Phi(m^*e^1+v)\right) \geq 1 - \beta\left[1-\tanh^2(\beta m^*(1-\tau))\right](1+3\sqrt{\alpha}) \\ - 2\beta\tanh^2(\beta m^*(1-\tau))\Gamma(\alpha,\tau m^*/\rho)\end{aligned} \tag{8.7}$$

Optimizing over τ gives the claim of the theorem.

■

Remark. Note that the estimates derived from (8.7) become quite bad if β is large. Thus local convexity appears to break down for some critical $\beta_{conv}(\alpha)$ that tends to infinity, as $\alpha \downarrow 0$. In the heuristic picture [AGS] such a critical line appears as the boundary of the region where the so-called replica symmetry is supposed to hold. It is very instructive to read what Amit et al. write on replica symmetry breaking in the retrieval phases: *"....the very occurrence of RSB*[8] *implies that the energy landscape of the basin of each of the retrieval phases has features that are similar to the SG*[9]*phase. In particular, each of the retrieval phases represents many degenerate retrieval states. All of them have the same macroscopic overlap m, but they differ in the location of the errors. These states are organized in an ultrametric structure"* ([AGS], page 59). Translated to our language, this means that replica symmetry breaking is seen as a failure of local convexity and the appearance of many local minima. On this basis we conjectured in [BG5] that replica symmetry is closely related to the local convexity of the free energy functional [10]

[8] = replica symmetry breaking

[9] = spin glass

[10] We should note, however, that our condition for local convexity (roughly $\beta^{-1}>\alpha$) does *not* have the same behaviour as is found for the stability of the replica symmetric solution in [AGS] ($\beta^{-1}>\exp(-1/2\alpha)$). It is rather clear that our condition for convexity cannot be substantially improved. On the other hand, Talagrand has informed us that his method of deriving the replica symmetric solution which

We can now make these observations more precise. While we have so far avoided this, now is the time to make use of the *Hubbard-Stratonovich transformation* [HS] for the case of quadratic E_M. That is, we consider the new measures $\tilde{\mathcal{Q}}_{\beta,N,M} \equiv \tilde{\mathcal{Q}}^1_{\beta,N,M}$ defined in (5.14). They have the remarkable property that they are absolutely continuous w.r.t. Lebesgue measure with density

$$\frac{1}{Z_{\beta,N,M}} \exp\left(-\beta N \Phi_{\beta,N,M}(z)\right) \tag{8.8}$$

(do the computation or look it up in [BGP1]). Moreover, in many computations it can conveniently replace the original measure $\mathcal{Q}$. In particular, the following identity holds for all $t \in \mathbb{R}^M$.

$$\int d\mathcal{Q}_{\beta,N,M}(m) e^{(t,m)} = e^{\frac{\|t\|_2^2}{\beta N}} \int d\tilde{\mathcal{Q}}_{\beta,N,M}(z) e^{(t,z)} \tag{8.9}$$

Since for t with bounded norm the first factor tends to one rapidly, this shows that the exponential moments of $\mathcal{Q}$ and $\tilde{\mathcal{Q}}$ are asymptotically equal. We will henceforth assume that we are in a range of β and α such that the union of the balls $B_{\rho(\epsilon)}(sm^* e^\mu)$ has essentially full mass under $\tilde{\mathcal{Q}}$.

To study one of the balls, we define for simplicity the conditional measures

$$\tilde{\mathcal{Q}}^{(1,1)}_{\beta,N,M}(\cdot) \equiv \tilde{\mathcal{Q}}_{\beta,N,M}\left(\cdot \,\middle|\, z \in B_{\rho(\epsilon)}(m^* e^1)\right) \tag{8.10}$$

with $\rho(\epsilon)$ such that Theorem 8.1 holds. (Alternatively we could consider tilted measures with h proportional to e^1 and arbitrarily small). For notational convenience we will introduce the abbreviation $\mathbb{E}_{\tilde{\mathcal{Q}}}$ for the expectation w.r.t. the measure $\tilde{\mathcal{Q}}^{(1,1)}_{\beta,N,M}$.

Now intuitively one would think that since $\tilde{\mathcal{Q}}^{(1,1)}_{\beta,N,M}$ has a density of the form $e^{-NV(z)}$ with a convex V with strictly positive second derivative, this measure should have similar properties as for quadratic V. It turns out that this is to some extent true. For instance, we have:

Theorem 8.2. *Under the hypothesis of Theorem 8.1, and with the same probability as in the conclusion of that theorem, for any* $t \in \mathbb{R}^M$

---

does not require convexity, can be extended to work under essentially the conditions of [AGS].

with $\|t\|_2 \leq C < \infty$,

$$e^{(t,\mathbb{E}_{\tilde{\mathcal{Q}}} z)} - O(e^{-M}) \leq \mathbb{E}_{\tilde{\mathcal{Q}}} e^{(t,z)} \leq e^{(t,\mathbb{E}_{\tilde{\mathcal{Q}}} z)} e^{\|t\|_2^2/\epsilon N} + O(e^{-M}) \quad (8.11)$$

In particular, the marginal distributions of $\mathcal{Q}$ *converge to Dirac distributions concentrated on the corresponding projections of* $\mathbb{E}_{\tilde{\mathcal{Q}}} z$.

Proof. The main tool in proving this Theorem are the so-called *Brascamp-Lieb inequalities*[8][BL]. We paraphrase them as follows.

Lemma 8.3. *[Brascamp-Lieb[BL]]Let* $V : \mathbb{R}^M \to \mathbb{R}$ *be non-negative and strictly convex with* $\lambda_{min}(\nabla^2 V) \geq \epsilon$. *Denote by* $\mathbb{E}_V$ *expectation with respect to the probability measure*

$$\frac{e^{-NV(x)} d^M x}{\int e^{-NV(x)} d^M x} \quad (8.12)$$

Let $f : \mathbb{R}^M \to \mathbb{R}$ *be any continuously differentiable function. Then*

$$\mathbb{E}_V (f - \mathbb{E}_V f)^2 \leq \frac{1}{\epsilon N} \mathbb{E}_V (\|\nabla f\|_2^2) \quad (8.13)$$

We see that we are essentially in a situation where we can apply Lemma 8.3. The only difference is that our measures are supported only on a subset of $\mathbb{R}^M$. This is however no problem: we may either continue the function $\Phi(m)$ as a strictly convex function to all $\mathbb{R}^M$ and study the corresponding measures noting that all reasonable expectations differ only by exponentially small terms, or one may run through the proof of Lemma 8.3 to see that the boundary terms we introduce only lead to exponentially small error terms in (8.13). We will disregard this issue in order not to complicate things unnecessarily. To see how Lemma 8.3 works, we deduce the following

Corollary 8.4. *Let* $\mathbb{E}_V$ *be as in Lemma 8.3. Then*

(i) $\mathbb{E}_V \|x - \mathbb{E}_V x\|_2^2 \leq \frac{M}{\epsilon N}$

(ii) $\mathbb{E}_V \|x - \mathbb{E}_V x\|_4^4 \leq 4 \frac{M}{\epsilon^2 N^2}$

(iii) For any function f *such that* $V_t(x) \equiv V(x) - tf(x)/N$ *for* $t \in [0,1]$ *is still strictly convex and* $\lambda_{min}(\nabla^2 V_t) \geq \epsilon' > 0$, *then*

$$0 \leq \ln \mathbb{E}_V e^f - \mathbb{E}_V f \leq \frac{1}{2\epsilon' N} \sup_{t\in[0,1]} \mathbb{E}_{V_t} \|\nabla f\|_2^2 \quad (8.14)$$

[8] We thank Dima Ioffe for having brought these to our attention

In particular,

(iv) $\ln \mathbb{E}_V e^{(t,(x-\mathbb{E}_V x))} \leq \frac{\|t\|_2^2}{2\epsilon N}$

(v) $\ln \mathbb{E}_V e^{\|x-\mathbb{E}_V x\|_2^2} - \mathbb{E}_V \|x-\mathbb{E}_V x\|_2^2 \leq \frac{M}{\epsilon^2 N^2}$

Proof. (i) Choose $f(x) = x_\mu$ in (8.13). Insert and sum. (ii) Choose $f(x) = x_\mu^2$ and use (i). (iii) Note that

$$\begin{aligned} \ln \mathbb{E}_V e^f &= \mathbb{E}_V f + \int_0^1 ds \int_0^s ds' \frac{\mathbb{E}_V \left[e^{s'f} \left(f - \frac{\mathbb{E}_V e^{s'f} f}{\mathbb{E}_V e^{s'f}} \right)^2 \right]}{\mathbb{E}_V e^{s'f}} \\ &= \mathbb{E}_V f + \int_0^1 ds \int_0^s ds' \mathbb{E}_{V_{s'}} \left(f - \mathbb{E}_{V_{s'}} f \right)^2 \end{aligned} \tag{8.15}$$

where by assumption $V_s(x)$ has the same properties as V itself. Thus using (8.13) gives (8.15) (iv) and (v) follow with the corresponding choices for f easily.

■

Theorem 8.2 is thus an immediate consequence of (iv).

■

We now come to the main result of this section. We will show that Theorem 8.1 in fact implies that the replica symmetric solution of [AGS] is correct in the range of parameters where Theorem 8.1 holds. Such a result was recently proven by Talagrand [T4], but we shall see that using Theorem 8.1 and the Brascamp-Lieb inequalities, we can give a greatly simplified proof.

Theorem 8.5. *Assume that the parameters β, α are such that the conditions both of Theorem 6.2 and of Theorem 8.1 are satisfied, with $\epsilon > 0$ and $\rho \geq c\gamma m^*$, where c is such that the mass of the complement of the set $\cup_{s,\mu} B_{c\gamma m^*}(sm^* e^\mu)$ is negligible. Then, the replica symmetric solution of [AGS] holds in the sense that, asymptotically, as $N \uparrow \infty$, $\mathbb{E}_{\tilde{Q}} z_1$, and $\mathbb{E}\|\mathbb{E}_{\tilde{Q}} \hat{z}\|_2^2$ (recall that $\hat{z} \equiv (0, z_2, \ldots)$ converge almost surely to the positive solution $\hat{\mu}$ and r of the system of equations*

$$\hat{\mu} = \int d\mathcal{N}(g) \tanh(\beta(\hat{\mu} + \sqrt{\alpha r} g)) \tag{8.16}$$

$$q = \int d\mathcal{N}(g) \tanh^2(\beta(\hat{\mu} + \sqrt{\alpha r} g)) \tag{8.17}$$

$$r = \frac{q}{(1-\beta+\beta q)^2} \tag{8.18}$$

(note that q is an auxiliary variable that could be eliminated).

Remark. As far as Theorem 8.5 is considered as a result on *conditional measures* only, it is possible to extend its validity beyond the regime of Theorem 6.2. In that case, what is needed is only Theorem 8.1 *and* the control of the location of the local minima given by Theorem 7.5. One may also, in this spirit, consider the extension of this result to other local minima (corresponding to the so-called "mixed patterns"), which would, of course, require to prove the analogues of Theorem 7.5, 8.1 in this case, as well as carrying out the stability analysis of a certain dynamical system (see below). We do not doubt that this can be done.

Remark. We will not enter into the discussion on how these equations were originally derived with the help of the replica trick. This is well explained in [AGS]. In [T4] it is also shown how one can derive on this basis the formula for the free energy as a function of $\hat{\mu}, r$, and q that is given in [AGS] and for which the above equations are the saddle point equations. We will not repeat these arguments here.

Remark. In [PST] it was shown that the replica symmetric solution holds if the so-called Edwards-Anderson parameter, $\frac{1}{N}\sum_i [\mu_{\beta,N,M}(\sigma_i)]^2$ is self-averaging. Some of the basic ideas in that paper are used both in Talagrand's and in our proof below. In fact we follow the strategy of [PST] more closely than Talagrand, and we will see that this leads immediately to the possibility of studying the limiting Gibbs measures.

Proof. It may be well worthwhile to outline the strategy of the proof in a slightly informal way before we go into the details. This may also give a new explanation to the mysterious looking equations above. It turns out that in a very specific sense, the idea of these equations and their derivation is closely related to the original idea of "mean field theory". Let us briefly recall what this means. The standard derivation of "mean field" equations for homogeneous magnets in most textbooks on statistical mechanics does not start from the Curie-Weiss model but from (i) the hypothesis that in the infinite volume limit, the spins are independent and identically distributed under the limiting (extremal) Gibbs measure and that (ii) their distribution is of the form $e^{\beta\sigma_i m}$ where m is the mean value of the spin under this same measure, and that is assumed to be an almost sure constant with respect to the Gibbs measure. The resulting consistency equation is then $m = \tanh \beta m$. This derivation breaks down in random systems, since it would be unreasonable to think that the spins are identically distributed. Of course one may keep the assumption of independence, and write down a set of consistency equations (in the spin-glass case, these are know as TAP-equations [TAP]). Let us try the idea in Hopfield model. The spin σ_i here couples to a "mean field" $h_i(\sigma) = (\xi_i, m(\sigma))$, which is a function of the entire vector of magnetizations. To obtain a self-consistent set of equations we would have to compute all of these, leading to the system

$$m_\mu = \frac{1}{N} \sum_i \xi_i^\mu \tanh(\beta(\xi_i, m)) \tag{8.19}$$

Solving this is a hopelessly difficult task when M is growing somewhat fast with N, and it is not clear why one should expect these quantities to be constants when $M = \alpha N$.

But now suppose it were true that we could somehow compute the *distribution* of $h_i(\sigma)$ *a priori* as a function of a small number of parameters, not depending on i. Assume further that these parameters are again functions of the distribution of the mean field. Then we could write down consistency conditions for them and (hopefully) solve them. In this way the expectation of σ_i could be computed. The tricky part is thus to find the *distribution* of the mean field [8]. Miraculously, this can be done, and the relevant parameters turn out to be the quantities $\hat{\mu}$ and r, with (8.16)-(8.18) the corresponding consistency equations[9]

We will now follow these ideas and give the individual steps a precise meaning. In fact, the first step in our proof corresponds to proving a version of Lemma 2.2 of [PST], or if one prefers, a sharpened version of Lemma 4.1 of [T4]. Note that we will never introduce any auxiliary Gaussian fields in the Hamiltonian, as is done systematically in [PST] and sometimes in [T4]; all comparison to quantities in these

[8] This idea seems related to statements of physicists one finds sometimes in the literature that in spin glasses, that the relevant "order parameter" is a actually a probability distribution.

[9] In fact, we will see that the situation is just a bit more complicated. For finite N, the distribution of the mean field will be seen to depend essentially on three N-dependent, non-random quantities whose limits, *should they exist*, are related to $\hat{\mu}$, r and q. Unfortunately, one of the notorious problems in disordered mean field type models is that one cannot prove a priori such intuitively obvious facts like that the mean values of thermodynamic quantities (such as the free energy, etc.) converge, even when it is possible to show that their fluctuations converge to zero (this sad fact is sometimes overlooked). We shall see that convergence of the quantities involved here can be proven in the process, using properties of the recurrence equations for which the equations above are the fixed point equations, and a priori control on the overlap distribution as results from Theorem 6.2 (or 7.5).

papers is thus understood modulo removal of such terms. Let us begin by mentioning that the crucial quantity $u(\tau)$ defined in Definition 5 of [PST] has the following nice representation[10]

$$u(\tau) = \ln \int d\tilde{\mathcal{Q}}^{(1,1)}_{\beta,N,M}(z) e^{\tau\beta(\eta,z)} \tag{8.20}$$

where, like Talagrand in [T4], we singled out the site $N+1$ (instead of 1 as in [PST]) and set $\xi_{N+1} = \eta$. For notational simplicity we will denote the expectation w.r.t. the measure $\tilde{\mathcal{Q}}^{(1,1)}_{\beta,N,M}$ by $\mathbb{E}_{\tilde{\mathcal{Q}}}$ and we will set $\bar{z} = z - \mathbb{E}_{\tilde{\mathcal{Q}}} z$.

Lemma 8.6. *Under the hypotheses of Theorem 8.5 we have that*
(i) With probability exp. close to 1,

$$\mathbb{E}_\eta \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})} = e^{\frac{\tau^2\beta^2}{2}\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}\|_2^2 + R} \tag{8.21}$$

where $|R| \leq \frac{C}{N}$.
(ii) Moreover,

$$\mathbb{E}_\eta \left(\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})} - \mathbb{E}_\eta \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})} \right)^2 \leq \frac{C}{N} \tag{8.22}$$

Proof. Note first that

$$\mathbb{E}_\eta \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})} \leq \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\frac{\tau^2\beta^2}{2}\|\bar{z}\|_2^2} \tag{8.23}$$

and also

$$\mathbb{E}_\eta \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})} \geq \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\frac{\tau^2\beta^2}{2}\|\bar{z}\|_2^2 - \frac{\tau^4\beta^4}{4}\|\bar{z}\|_4^4} \tag{8.24}$$

(8.23) looks most encouraging and (ii) of Corollary 8.4 leaves hope for the $\|\bar{z}\|_4^4$ to be irrelevant. Of course for this we want the expectation to move up into the exponent. To do this, we use (iii) of Corollary 8.4 with f chosen as $\frac{\tau^2\beta^2}{2}\|\bar{z}\|_2^2$ and $\frac{\tau^2\beta^2}{2}\|\bar{z}\|_2^2 - \frac{\tau^4\beta^4}{12}\|\bar{z}\|_4^4$, respectively. For this we have to check the strict convexity of $\Phi + \frac{s}{N} f$ in these cases. But a simple computation shows that in both cases $\lambda_{\min}\left(\nabla^2(\Phi + \frac{s}{N} f)\right) \geq \epsilon - \frac{\tau\beta}{N}$, so that for any τ, β there is no problem if N is large enough (Note that the quartic term has the good sign!). A straightforward calculation shows

[10] Actually, our definition differs by an irrelevant constant from that of [PST].

that this gives (8.21).
To prove (ii), it is enough to compute

$$\mathbb{E}_\eta \left(\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})}\right)^2 = \mathbb{E}_\eta \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z}+\bar{z}')} \tag{8.25}$$

where we (at last!) introduced the "replica" z' that is an independent copy of the random variable z. By some abuse of notation $\mathbb{E}_{\tilde{\mathcal{Q}}}$ also denotes the product measure for these two copies. By the same token as in the proof of (i), we see that,

$$\mathbb{E}_\eta \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z}+\bar{z}')} = e^{\frac{\tau^2\beta^2}{2}\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}+\bar{z}'\|_2^2 + O(1/N)} \tag{8.26}$$

Finally,

$$\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}+\bar{z}'\|_2^2 = 2\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}\|_2^2 + 2\mathbb{E}_{\tilde{\mathcal{Q}}}(\bar{z},\bar{z}') = 2\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}\|_2^2 \tag{8.27}$$

Inserting this and (8.21) into the left hand side of (8.22) establishes that bound. This concludes the proof of Lemma 8.6.

∎

An easy corollary gives what Talagrand's Lemma 4.1 should be:

Corollary 8.7. *Under the hypotheses of Lemma 8.6, there exists a finite numerical constant c such that*

$$u(\tau) = \beta\tau(\eta, \mathbb{E}_{\tilde{\mathcal{Q}}} z) + \frac{\tau^2\beta^2}{2}\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}\|_2^2 + R_N \tag{8.28}$$

where

$$\mathbb{E}|R_N|^2 \leq \frac{c}{N} \tag{8.29}$$

Proof. Obviously

$$\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,z)} = e^{\tau\beta(\eta,\mathbb{E}_{\tilde{\mathcal{Q}}} z)}\mathbb{E}_\eta \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})} \frac{\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})}}{\mathbb{E}_\eta \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})}} \tag{8.30}$$

Taking logarithms, the first two factors in (8.30) together with (8.21) give the two first terms in (8.28) plus a remainder of order $\frac{1}{N}$. For the last factor, we notice first that by Corollary 8.4, (iii),

$$e^{-\tau^2\beta\frac{M}{\epsilon N}} \leq \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})} \leq e^{\tau^2\beta\frac{M}{\epsilon N}} \tag{8.31}$$

so that for α small, τ and $\beta\alpha$ bounded, $\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})}$ is bounded away from 0 and infinity; we might for instance think that $\frac{1}{2} \leq \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})} \leq 2$. But for A, B in a compact interval of the positive half line not containing zero, there is a finite constant C such that $|\ln\frac{A}{B}| = |\ln A - \ln B| \leq C|A-B|$. Using this gives

$$\mathbb{E}_\eta\left[\ln\frac{\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})}}{\mathbb{E}_\eta\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})}}\right]^2 \leq C^2\mathbb{E}_\eta\left(\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})} - \mathbb{E}_\eta\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\tau\beta(\eta,\bar{z})}\right)^2 \tag{8.32}$$

From this and (8.22) follows the estimate (8.29).

■

We have almost proven the equivalent of Lemma 2.2 in [PST]. What remains to be shown is

Lemma 8.8: *Under the assumptions of Theorem 8.1* $(\eta, \mathbb{E}_{\tilde{\mathcal{Q}}} z)$ *converges in law to* $\eta_1\hat{\mu} + \sqrt{\alpha r} g$ *where* $\hat{\mu} = \lim_{N\uparrow\infty}\mathbb{E}_{\tilde{\mathcal{Q}}} z_1$ *and* $r \equiv \alpha^{-1}\lim_{N\uparrow\infty}\left\|\mathbb{E}_{\tilde{\mathcal{Q}}}\hat{z}\right\|_2^2$, *where* $\hat{z} \equiv (0, z_2, z_3, \dots,)$ *and* g *is a standard normal random variable.*

Quasiproof:[PST] The basic idea behind this lemma is that for all $\mu > 1$, $\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu$ tends to zero, the η_μ are independent amongst each other and of the $\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu$ and that therefore $\sum_{\mu>1}\eta_\mu\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu$ converge to a Gaussians with variance $\lim_{N\uparrow\infty}\left\|\mathbb{E}_{\tilde{\mathcal{Q}}}\hat{z}\right\|_2^2$.

■

To make this idea precise is somewhat subtle. First, to prove a central limit theorem, one has to show that some version of the Lindeberg condition [CT] is satisfied in an appropriate sense. To do this we need some more facts about self-averaging. Moreover, one has to make precise to what extent the quantities $\mathbb{E}_{\tilde{\mathcal{Q}}} z_1$ and $\left\|\mathbb{E}_{\tilde{\mathcal{Q}}}\hat{z}\right\|_2^2$ converge, as N tends to infinity. There is no way to prove this a priori, and only at the end of the proof of Theorem 8.5 will it be clear that this is the case. Thus we cannot and will not use Lemma 8.8 in the proof of the Theorem, but a weaker statement formulated as Lemma 8.13 below.

The following lemma follows easily from the proof of Talagrand's Proposition 4.3 in [T5].

Lemma 8.9. *Assume that $f(x)$ is a convex random function defined on some open neighborhood $U \subset \mathbb{R}$. Assume that f verifies for all $x \in U$ that $|(\mathbb{E}f)''(x)| \leq C < \infty$ and $\mathbb{E}(f(x) - \mathbb{E}f(x))^2 \leq S^2$. Then, if $x \pm S/C \in U$*

$$\mathbb{E}\left(f'(x) - \mathbb{E}f'(x)\right)^2 \leq 12CS \tag{8.33}$$

But as so often in this problem, variance estimates are not quite sufficient. We will need the following, sharper estimate (which may be well known):

Lemma 8.10. *Assume that $f(x)$ is a random function defined on some open neighborhood $U \subset \mathbb{R}$. Assume that f verifies for all $x \in U$ that for all $0 \leq r \leq 1$,*

$$\mathbb{P}\left[|f(x) - \mathbb{E}f(x)| > r\right] \leq c \exp\left(-\frac{Nr^2}{c}\right) \tag{8.34}$$

and that, at least with probability $1-p$, $|f'(x)| \leq C$, $|f''(x)| \leq C < \infty$ both hold uniformly in U. Then, for any $0 < \zeta \leq 1/2$, and for any $0 < \delta < N^{\zeta/2}$,

$$\mathbb{P}\left[|f'(x) - \mathbb{E}f'(x)| > \delta N^{-\zeta/2}\right] \leq \frac{32C^2}{\delta^2} N^{\zeta} \exp\left(-\frac{\delta^4 N^{1-2\zeta}}{256c}\right) + p \tag{8.35}$$

Proof. Let us assume that $|U| \leq 1$. We may first assume that the boundedness conditions for the derivatives of f hold uniformly; by standard arguments one shows that if they only hold with probability $1-p$, the effect is nothing more than the final summand p in (8.35). The first step in the proof consists in showing that (8.34) together with the boundedness of the derivative of f implies that $f(x) - \mathbb{E}f(x)$ is uniformly small. To see this introduce a grid of spacing ϵ, i.e. let

$U_\epsilon = U \cap \epsilon\mathbb{Z}$. Clearly

$$
\begin{aligned}
&\mathbb{P}\left[\sup_{x\in U}|f(x)-\mathbb{E}f(x)|>r\right]\\
&\leq \mathbb{P}\left[\sup_{x\in U_\epsilon}|f(x)-\mathbb{E}f(x)|\right.\\
&\qquad\left.+\sup_{x,y:|x-y|\leq\epsilon}|f(x)-f(y)|+|\mathbb{E}f(x)-\mathbb{E}f(y)|>r\right]\\
&\leq \mathbb{P}\left[\sup_{x\in U_\epsilon}|f(x)-\mathbb{E}f(x)|>r-2C\epsilon\right]\\
&\leq \epsilon^{-1}\mathbb{P}\left[|f(x)-\mathbb{E}f(x)|>r-2C\epsilon\right]
\end{aligned}
\tag{8.36}
$$

If we choose $\epsilon = \frac{r}{4C}$, this yields

$$
\mathbb{P}\left[\sup_{x\in U}|f(x)-\mathbb{E}f(x)|>r\right]\leq\frac{4C}{r}\exp\left(-\frac{Nr^2}{4c}\right)
\tag{8.37}
$$

Next we show that *if* $\sup_{x\in U}|f(x)-g(x)|\leq r$ for two functions f, g with bounded second derivative, then

$$
|f'(x)-g'(x)|\leq\sqrt{8Cr}
\tag{8.38}
$$

For notice that

$$
\left|\frac{1}{\epsilon}[f(x+\epsilon)-f(x)]-f'(x)\right|\leq\frac{\epsilon}{2}\sup_{x\leq y\leq x+\epsilon}f''(y)\leq C\frac{\epsilon}{2}
\tag{8.39}
$$

so that

$$
\begin{aligned}
|f'(x)-g'(x)|&\leq\frac{1}{\epsilon}|f(x+\epsilon)-g(x+\epsilon)-f(x)+g(x)|+C\epsilon\\
&\leq\frac{2r}{\epsilon}+C\epsilon
\end{aligned}
\tag{8.40}
$$

Choosing the optimal $\epsilon=\sqrt{2r/C}$ gives (8.38). It suffices to combine (8.38) with (8.37) to get

$$
\mathbb{P}\left[|f'(x)-\mathbb{E}f'(x)|>\sqrt{8rC}\right]\leq\frac{4C}{r}\exp\left(-\frac{Nr^2}{4c}\right)
\tag{8.41}
$$

Setting $r=\frac{\delta^2}{CN^\zeta}$, we arrive at (8.35).

■

We will now use Lemma 8.10 to control $\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu$. We define

$$f(x) = \frac{1}{\beta N} \ln \int_{B_\rho(m^* e^1)} d^M z e^{\beta N x z_\mu} e^{-\beta N \Phi_{\beta,N,M}(z)} \tag{8.42}$$

and denote by $\mathbb{E}_{\tilde{\mathcal{Q}}^x}$ the corresponding modified expectation. As has by now been shown many times [T2,BG5,T4], $f(x)$ verifies (8.34). Moreover, $f'(x) = \mathbb{E}_{\tilde{\mathcal{Q}}^x} z_\mu$ and

$$f''(x) = \beta N \mathbb{E}_{\tilde{\mathcal{Q}}^x} \left(z_\mu - \mathbb{E}_{\tilde{\mathcal{Q}}^x} z_\mu\right)^2 \tag{8.43}$$

Of course the addition of the linear term to Φ does not change its second derivative, so that we can apply the Brascamp-Lieb inequalities also to the measure $\mathbb{E}_{\tilde{\mathcal{Q}}^x}$. This shows that

$$\mathbb{E}_{\tilde{\mathcal{Q}}^x} \left(z_\mu - \mathbb{E}_{\tilde{\mathcal{Q}}^x} z_\mu\right)^2 \leq \frac{1}{\epsilon N \beta} \tag{8.44}$$

which means that $f(x)$ has a second derivative bounded by $c = \frac{1}{\epsilon}$.
Remark. In the sequel we will use Lemma 8.10 only in situations where p is irrelevantly small compared to the main term in (8.35). We will thus ignore its existence for simplicity.

This gives the

Corollary 8.11. *Under the assumptions of Theorem 8.1, there are finite positive constants c, C such that, for any $\zeta \leq \frac{1}{2}$ and $\delta \leq N^{\zeta/2}$, for any μ,*

$$\mathbb{P}\left[\left|\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu - \mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu\right| \geq \delta N^{-\zeta/2}\right] \leq \frac{C}{\delta^2} N^\zeta \exp\left(-\frac{\delta^4 N^{1-2\zeta}}{c}\right) \tag{8.45}$$

This leaves us only with the control of $\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu$. But by symmetry, for all $\mu > 1$, $\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu = \mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}} z_2$ while on the other hand

$$\sum_{\mu=2}^{M} (\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu)^2 \leq c^2 \gamma^2 (m^*)^2 \tag{8.46}$$

so that $|\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu| \leq \frac{c}{m^*} N^{-1/2}$. Therefore, with probability of order, say $1 - \exp(-N^{1-2\zeta})$ it is true that for all $\mu > 2$, $|\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu| \leq \delta N^{-\zeta/2}$.

Finally we must control the behaviour of the prospective variance of our gaussian. We set $T_N \equiv \sum_{\mu=2}^{M(N)} (\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu)^2$. Let us introduce

$$g(x) \equiv \frac{1}{\beta N} \ln \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\beta N x (\hat{z}, \hat{z}')} \tag{8.47}$$

where $\mathbb{E}_{\tilde{\mathcal{Q}}}$ is understood as the product measure for the two independent copies z and z'. The point is that $T_N = g'(0)$. On the other hand, g satisfies the same self-averaging conditions as the function f before, and its second derivative is bounded (for $x \leq \epsilon/2$), since

$$\begin{aligned} g''(x) &= \beta N \mathbb{E}_{\tilde{\mathcal{Q}}^x} \left((\hat{z}, \hat{z}') - \mathbb{E}_{\tilde{\mathcal{Q}}^x} (\hat{z}, \hat{z}') \right)^2 \\ &\leq \frac{2\beta}{\epsilon} 2 \mathbb{E}_{\tilde{\mathcal{Q}}^x} \|\hat{z}\|_2^2 \leq 2\rho \frac{\beta}{\epsilon} \end{aligned} \tag{8.48}$$

where here $\mathbb{E}^x_{\tilde{\mathcal{Q}}}$ stands for the coupled measure corresponding to (8.47) (and is not the same as the the measure with the same name in (8.43)). Thus we get our second corollary:

Corollary 8.12. *Under the assumptions of Theorem 8.1, there are finite positive constants c, C such that, for any $\zeta \leq \frac{1}{2}$ and $\delta \leq N^{\zeta/2}$,*

$$\mathbb{P}\left[|T_N - \mathbb{E}T_N| \geq \delta N^{-\zeta/2} \right] \leq \frac{C}{\delta^2} N^\zeta \exp\left(-\frac{\delta^4 N^{1-2\zeta}}{c} \right) \tag{8.49}$$

Thus T_N converges almost surely to a constant if $\mathbb{E}T_N$ converges. We are now in a position to prove

Lemma 8.13. *Consider the random variables $X_N \equiv \frac{1}{\sqrt{\mathbb{E}T_N}} \sum_{\mu=2}^{M(N)} \eta_\mu \mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu$. Then, if the hypotheses of Theorem 8.5 are satisfied, X_N converges weakly to a gaussian random variable of mean zero and variance one.*

Proof. Let us show that $\mathbb{E}e^{itX_N}$ converges to $e^{-t^2/2}$. To see this, let Ω_N denote the subset of Ω on which the various nice things we want to impose on $\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu$ hold; we know that the complement of that set has measure smaller than $O(e^{-N^{1-2\zeta}})$. We write

$$\begin{aligned} \mathbb{E}e^{itX_N} &= \mathbb{E}_\xi \left[\mathbb{1}_{\Omega_N} \mathbb{E}_\eta e^{itX_N} + \mathbb{1}_{\Omega_N^c} \mathbb{E}_\eta e^{itX_N} \right] \\ &= \mathbb{E}_\xi \left[\mathbb{1}_{\Omega_N} \prod_\mu \cos\left(\frac{t}{\sqrt{\mathbb{E}T_N}} \mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu \right) \right] + O\left(e^{-N^{1-2\zeta}} \right) \end{aligned} \tag{8.50}$$

Thus the second term tends to zero rapidly and can be forgotten. On the other hand, on Ω_N,

$$\sum_{\mu=2}^{M}(\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu)^4 \leq \delta^2 N^{-\zeta} \sum_{\mu=2}^{M}(\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu)^2 \leq \delta^2 N^{-\zeta} \frac{c\alpha}{(m^*)^2} \tag{8.51}$$

tends to zero, so that using for instance $|\ln\cos x - x^2/2| \leq cx^4$ for $|x| \leq 1$,

$$\begin{aligned} &\mathbb{E}_\xi \mathbb{I}_{\Omega_N} \mathbb{E}_\eta e^{itX_N} \\ &\leq e^{-t^2/2} \sup_{\Omega_N} \left[\exp\left(-\frac{T_N - \mathbb{E}T_N}{2\mathbb{E}T_N} + c\frac{t^4\delta^2 N^{-\zeta}}{(\mathbb{E}T_N)^2}\right)\right] \mathbb{P}_\xi(\Omega_N) \end{aligned} \tag{8.52}$$

Clearly, since also $|T_N - \mathbb{E}T_N| \leq \delta N^{-\zeta/2}$, the right hand side converges to $e^{-t^2/2}$ and this proves the lemma.

■

Corollary 8.7 together with Lemma 8.13 represent the complete analogue of Lemma 2.2 of [PST]. To derive from here the equations (8.16)-(8.18) requires actually a little more, namely a corresponding statement on the convergence of the derivative of $u(\tau)$. Fortunately, this is not very hard to show.

Lemma 8.14. *Set $u(\tau) = u_1(\tau) + u_2(\tau)$, where $u_1(\tau) = \tau\beta(\eta, \mathbb{E}_{\tilde{\mathcal{Q}}} z)$ and $u_2(\tau) = \ln \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\beta\tau(\eta,\bar{z})}$. Then under the assumption of Corollary 8.13,*

(i) $\frac{1}{\beta\sqrt{\mathbb{E}T_N}}\frac{d}{d\tau}u_1(\tau)$ converges weakly to a standard gaussian random variable.

(ii) $\left|\frac{d}{d\tau}u_2(\tau) - \tau\beta^2\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}\|_2^2\right|$ converges to zero in probability.

Proof. (i) is obvious from Corollary 8.13. To prove (ii), note that $u_2(\tau)$ is convex and $\frac{d^2}{d\tau^2}u_2(\tau) \leq \frac{\beta\alpha}{\epsilon}$. Thus, *if* $\mathrm{var}\,(u_2(\tau)) \leq \frac{C}{\sqrt{N}}$, then $\mathrm{var}\left(\frac{d}{d\tau}u_2(\tau)\right) \leq \frac{C'}{N^{1/4}}$ by Lemma 8.9. On the other hand, $|\mathbb{E}u_2(\tau) - \frac{\tau^2\beta^2}{2}\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}\|_2^2| \leq \frac{K}{\sqrt{N}}$, by Corollary 8.7, which, together with the boundedness of the second derivative of $u_2(\tau)$ implies that $|\frac{d}{d\tau}\mathbb{E}u_2(\tau) - \tau\beta^2\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}\|_2^2| \downarrow 0$. This means that $\mathrm{var}\,(u_2(\tau)) \leq \frac{C}{\sqrt{N}}$ implies the Lemma. Since we already know that $\mathbb{E}R_N^2 \leq \frac{K}{N}$, it is enough to prove $\mathrm{var}\left(\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}\|_2^2\right) \leq \frac{C}{\sqrt{N}}$. But this is a, by now, familiar exercise.

The point is to use that $\mathbb{E}_{\tilde{\mathcal{Q}}}\|\bar{z}\|_2^2 = \frac{d}{dx}\tilde{g}(x)$, where

$$\tilde{g}(x) \equiv \frac{1}{\beta N} \ln \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\beta N x \|\bar{z}\|_2^2} \tag{8.53}$$

and to prove that $\text{var}\,(\tilde{g}(x)) \leq \frac{K}{N}$. using what we know about $\|\mathbb{E}_{\tilde{\mathcal{Q}}} z\|_2$ this follows as in the case of the function $g(x)$. The proof is finished. ■

From here we can follow [PST]. Let us denote by $\mathbb{E}_{\mathcal{Q}}$ the expectation with respect to the (conditional) induced measures $\mathcal{Q}^{(1,1)}_{\beta,N,M}$. Note first that (8.9) implies that[11] $\mathbb{E}_{\mathcal{Q}} m_\mu = \mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu$. On the other hand,

$$\mathbb{E}_{\mathcal{Q}} m_\mu = \frac{1}{N}\sum_{i=1}^N \xi_i^\mu \mu^{(1,1)}_{\beta,N,M}(\sigma_i) \tag{8.54}$$

and so, by symmetry

$$\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}_{N+1}}(z_\mu) = \eta^\mu \mathbb{E}\mu_{\beta,N+1,M}(\sigma_{N+1}) \tag{8.55}$$

Note that from here on we will make the N-dependence of our mesures explicit, as we are going to derive recursion relations. Now, $u(\tau)$ was defined such that

$$\begin{aligned}\mathbb{E}\mu_{\beta,N+1,M}(\sigma_{N+1}) &= \mathbb{E}\frac{e^{u(1)} - e^{u(-1)}}{e^{u(1)} + e^{u(-1)}}\\ &= \mathbb{E}\tanh(\beta(\eta_1 \mathbb{E}_{\tilde{\mathcal{Q}}_N} z_1 + \sqrt{\mathbb{E}T_N}X_N)) + o(1)\end{aligned} \tag{8.56}$$

Thus, if $\mathbb{E}_{\tilde{\mathcal{Q}}_N} z_1$ and $\mathbb{E}T_N$ converge, by Lemma 8.13, the limit must satisfy (8.16). Of course we still need an equation for $\mathbb{E}T_N$ which is somewhat tricky. Let us first *define* a quantity $\mathbb{E}Q_N$ by

$$\mathbb{E}Q_N \equiv \mathbb{E}\tanh^2(\beta(\eta_1 \mathbb{E}_{\tilde{\mathcal{Q}}_N} z_1 + \sqrt{\mathbb{E}T_N}X_N)) \tag{8.57}$$

This corresponds of course to (8.17). Now note that $T_N = \|\mathbb{E}_{\tilde{\mathcal{Q}}_N} z\|_2^2 -$

[11] This relation is exact, if the tilted measures are considered, and it is true up to irrelevant error terms if one considers the conditioned measures.

$(\mathbb{E}_{\tilde{\mathcal{Q}}_N} z_1)^2$ and

$$\begin{aligned}\mathbb{E}\|\mathbb{E}_{\tilde{\mathcal{Q}}_{N+1}} z\|_2^2 &= \sum_{\mu=1}^{M} \mathbb{E}\left(\frac{1}{N+1}\sum_{i=1}^{N+1} \xi_i^\mu \mu_{\beta,N+1,M}(\sigma_i)\right)^2 \\ &= \frac{M-1}{N+1}\mathbb{E}\left(\mu^{(1,1)}_{\beta,N+1,M}(\sigma_{N+1})\right)^2 \\ &+ \sum_{\mu=1}^{M} \mathbb{E}\xi^\mu_{N+1}\mu^{(1,1)}_{\beta,N+1,M}(\sigma_{N+1})\left(\frac{1}{N+1}\sum_{i=1}^{N}\xi_i^\mu \mu_{\beta,N+1,M}(\sigma_i)\right)\end{aligned} \tag{8.58}$$

We see that the first term gives, by definition and (8.56), $\alpha\mathbb{E}Q_N$. For the second term, we use the identity form [PST]

$$\sum_{\mu=1}^{M}\xi^\mu_{N+1}\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu\mu_{\beta,N+1,M}(\sigma_i)\right) = \beta^{-1}\frac{\sum_{\tau=\pm1}u'(\tau)e^{u(\tau)}}{\sum_{\tau=\pm1}e^{u(\tau)}} \tag{8.59}$$

which it is not too hard to verify. Together with Lemma 8.14 one concludes that in law up to small errors

$$\begin{aligned}\sum_{\mu=1}^{M}\xi^\mu_{N+1}\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu\mu_{\beta,N+1,M}(\sigma_i)\right) &= \xi^1_{N+1}\mathbb{E}_{\tilde{\mathcal{Q}}_N} z_1 + \sqrt{\mathbb{E}T_N}X_N \\ &+ \beta\mathbb{E}_{\tilde{\mathcal{Q}}_N}\|\bar z\|_2^2 \tanh\beta\left(\xi^1_{N+1}\mathbb{E}_{\tilde{\mathcal{Q}}_N} z_1 + \sqrt{\mathbb{E}T_N}X_N\right)\end{aligned} \tag{8.60}$$

and so

$$\begin{aligned}\mathbb{E}\|\mathbb{E}_{\tilde{\mathcal{Q}}_{N+1}} z\|_2^2 = \alpha\mathbb{E}Q_N + \mathbb{E}\Bigg[&\tanh\beta\left(\xi^1_{N+1}\mathbb{E}_{\tilde{\mathcal{Q}}_N} z_1 + \sqrt{\mathbb{E}T_N}X_N\right) \\ &\times\left[\xi^1_{N+1}\mathbb{E}_{\tilde{\mathcal{Q}}_N} z_1 + \sqrt{\mathbb{E}T_N}X_N\right]\Bigg] \\ &+ \beta\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}_N}\|\bar z\|_2^2\tanh^2\beta\left(\xi^1_{N+1}\mathbb{E}_{\tilde{\mathcal{Q}}_N} z_1 + \sqrt{\mathbb{E}T_N}X_N\right)\end{aligned} \tag{8.61}$$

Using the self-averaging properties of $\mathbb{E}_{\tilde{\mathcal{Q}}_N}\|\bar z\|_2^2$, the last term is of course essentially equal to

$$\beta\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}_N}\|\bar z\|_2^2\mathbb{E}Q_N \tag{8.62}$$

The appearance of $\mathbb{E}_{\tilde{\mathcal{Q}}_N}\|\bar z\|_2^2$ is disturbing, as it introduces a new quantity into the system. Fortunately, it is the last one. The point is that proceeding as above, we can show that

$$\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}_{N+1}}\|z\|_2^2 = \alpha + \mathbb{E}\Bigg[\tanh\beta\left(\xi^1_{N+1}\mathbb{E}_{\tilde{\mathcal{Q}}_N}z_1 + \sqrt{\mathbb{E}T_N}X_N\right) \times \left[\xi^1_{N+1}\mathbb{E}_{\tilde{\mathcal{Q}}_N}z_1 + \sqrt{\mathbb{E}T_N}X_N\right]\Bigg] + \beta\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}_N}\|\bar{z}\|_2^2\mathbb{E}Q_N \tag{8.63}$$

so that setting $U_N \equiv \mathbb{E}_{\tilde{\mathcal{Q}}_N}\|\bar{z}\|_2^2$, we get, subtracting (8.61) from (8.63), the simple recursion

$$\mathbb{E}U_{N+1} = \alpha(1-\mathbb{E}Q_N) + \beta(1-\mathbb{E}Q_N)\mathbb{E}U_N \tag{8.64}$$

From this we get (since all quantities considered are self-averaging, we drop the $\mathbb{E}$ to simplify the notation), setting $M_N \equiv \mathbb{E}_{\tilde{\mathcal{Q}}_N}z_1$,

$$\begin{aligned} T_{N+1} &= -(M_{N+1})^2 + \alpha Q_N + \beta U_N Q_N \\ &\quad + \int d\mathcal{N}(g)[M_N + \sqrt{T_N}g]\tanh\beta(M_N + \sqrt{T_N}g) \\ &= M_{N+1}(M_N - M_{N+1}) + \beta U_N Q_N + \beta T_N(1-Q_N) + \alpha Q_N \end{aligned} \tag{8.65}$$

where we used integration by parts. The complete system of recursion relations can thus be written as

$$\begin{aligned} M_{N+1} &= \int d\mathcal{N}(g)\tanh\beta\left(M_N + \sqrt{T_N}g\right) \\ T_{N+1} &= M_{N-1}(M_N - M_{N+1}) + \beta U_N Q_N + \beta T_N(1-Q_N) + \alpha Q_N \\ U_{N+1} &= \alpha(1-Q_N) + \beta(1-Q_N)U_N \\ Q_{N+1} &= \int d\mathcal{N}(g)\tanh^2\beta\left(M_N + \sqrt{T_N}g\right) \end{aligned} \tag{8.66}$$

We leave it to the reader to check that the fixed points of this system lead to the equations (8.16)-(8.18) with $r = \lim_{N\uparrow\infty} T_N/\alpha$, $q = \lim_{N\uparrow\infty} Q_N$ and $m_1 = \lim_{N\uparrow\infty} M_N$ (where the variable $u = \lim_{N\uparrow\infty} U_N$ is eliminated).

We have dropped both the $o(1)$ errors and the fact that the parameters β and α are slightly changed on the left by terms of order $1/N$. The point is that, as explained in [T4], these things are irrelevant. The point is that from the localization results of the induced measures we know a priori that for all N, if α and β are in the appropriate domain, the four quantities are in a well defined domain. Thus, if this domain is attracted by the "pure" recursion (8.66), then we may choose some function $f(N)$ tending (slowly) to infinity (e.g. $f(N) = \ln N$) would be a good choice) and iterate $f(N)$ times; letting N tend to infinity then gives the desired convergence to the fixed point.

The necessary stability analysis, which is finally an elementary analytical problem can be found in [T4], Lemma 7.9 where it was apparently carried out for the first time in rigorous form (a numerical investigation can of course be found in [AGS]). It shows that all is well if $\alpha\beta$ and γ are small enough.

■

It is a particularly satisfying feature of the proof of Theorem 8.5 that in the process we have obtained via Corollary 8.7 and Lemma 8.13 control over the limiting probability distribution of the "mean field", (ξ_i, m), felt by an individual spin σ_i. In particular, the facts we have gathered also prove Lemma 8.8. Indeed, since $u(\tau)$ is the logarithm of the Laplace transform of that field we can identify it with a gaussian of variance $\mathbb{E}\mathbb{E}_{\tilde{\mathcal{Q}}_N} \|\bar{z}\|_2^2$ and mean $\mathbb{E}_{\tilde{\mathcal{Q}}_N} z_1 + \sqrt{\alpha r} g_i$, where g_i is itself a standard gaussian. Moreover, esssentially the same analysis allows to control not only the distribution of a single field (ξ, m), but of any *finite* collection, $(\xi_i, m)_{i\in V}$, of them. Form this we are able to reconstruct *the probability distribution of the Gibbs measures*:

Theorem 8.15. *Under the conditions of Theorem 8.5, for any finite set $V \subset \mathbb{N}$, the corresponding marginal distributions of the Gibbs measures $\mu^{(1,1)}_{\beta,N,M(N)}(\sigma_i = s_i, \forall i \in V)$ converge in law to*

$$\prod_{i\in V} \frac{e^{\beta s_i(\hat{\mu}\xi_i^1 + \sqrt{\alpha r} g_i)}}{2\cosh(\beta(\hat{\mu}\xi_i^1 + \sqrt{\alpha r} g_i)}$$

where g_i, $i \in V$ are independent standard gaussian random variables.

Remark. In the language of Newman [NS] the above theorem identifies the limiting Aizenman-Wehr metastate[12]for our system. Note that there seems to be no (reasonable) way to enforce almost sure convergence of Gibbs states for $\alpha > 0$. In fact, the g_i are continuous unbounded random variables, and by chosing suitable random subsequences N_i, we can construct *any* desired product measure as limiting measure!! Thus in the sense of the definition of limiting Gibbs states in Section II, we must conclude that for positive α, all product measures

[12] It would be interesting to study also the "empirical metastate'.'

are extremal measures for our system, a statement that may seem surprising and that misses most of the interesting information contained in Theorem 8.12. Thus we stress that this provides an example where the only way to express the full available information on the asymptotics of the Gibbs measures is in terms of their probability distribution, i.e. through metastates. Note that in our case, the metatstate is concentrated on product mesures which can be seen as a statement on "propagation of chaos" [Sn]. Beyond the "replica symmetric regime" this should no longer be true, and the metastate should then live on mixtures of product measures.

Proof. We will give a brief sketch of the proof of Theorem 8.15. More details are given in [BG6]. It is a simple matter to show that

$$\begin{aligned}\mu^{(1,1)}_{\beta,N,M}&(\sigma_i = s_i, \forall i \in V)\\ &= \frac{\int_{B_\rho(m^* e^1)} d^M z e^{-\beta N\left[\frac{\|z\|_2^2}{2} - \frac{1}{\beta N}\sum_{i\notin V} \ln\cosh(\beta(\xi_i,z))\right]} e^{\beta\sum_{i\in V} s_i(\xi_i,z)}}{\int_{B_\rho(m^* e^1)} d^M z e^{-\beta N\left[\frac{\|z\|_2^2}{2} - \frac{1}{\beta N}\sum_{i\notin V} \ln\cosh(\beta(\xi_i,z))\right]} \prod_{i\in V} 2\cosh(\beta(\xi_i,z))}\end{aligned} \tag{8.67}$$

Note that there is, for V fixed and N tending to infinity, virtually no difference between the function $\Phi_{\beta,N,M}$ and $\frac{\|z\|_2^2}{2} - \frac{1}{\beta N}\sum_{i\notin V} \ln\cosh(\beta(\xi_i,z))$ so we will simply pretend they are the same. So we may write in fact

$$\mu^{(1,1)}_{\beta,N,M}(\sigma_i = s_i, \forall i \in V) = \frac{\mathbb{E}_{\tilde{\mathcal{Q}}_{N-|V|}} e^{\beta\sum_{i\in V} s_i(\xi_i,z)}}{\sum_{\sigma_V} \mathbb{E}_{\tilde{\mathcal{Q}}_{N-|V|}} e^{\beta\sum_{i\in V} \sigma_i(\xi_i,z)}} \tag{8.68}$$

Now we proceed as in Lemma 8.6.

$$\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\beta\sum_{i\in V} s_i(\xi_i,z)} = e^{\beta\sum_{i\in V} s_i(\xi_i,\mathbb{E}_{\tilde{\mathcal{Q}}} z)} \mathbb{E}_{\tilde{\mathcal{Q}}} e^{\beta\sum_{i\in V} s_i(\xi_i,\bar{z})} \tag{8.69}$$

The second factor is controlled just as in Lemma 8.6, and up to terms that converge to zero in probability is independent of s_V. It will thus drop out in the ratio in (8.68). The exponent in the first term is treated as in Lemma 8.8; since all the ξ_i, $i \in V$ are independent, we obtain that the $(\xi_i, \mathbb{E}_{\tilde{\mathcal{Q}}}\hat{z})$ converge indeed to independent gaussian random variables. We omit the details of the proof of the analogue of Lemma 8.9; but note that $(\xi_i, \mathbb{E}_{\tilde{\mathcal{Q}}}\hat{z})$ are uncorrelated, and this is enough to get independence in the limit (since uncorrelated gaussians are independent). From here the proof of Theorem 8.15 is obvious.

■

We stress that we have proven that the Gibbs measures converge weakly in law (w.r.t. to $\mathbb{P}$) to some random product measure on the spins. Moreover it should be noted that the probabilities of local events (i.e. the expressions considered in Theorem 8.15) in the limit are not measurable with respect to a local sigma-algebra, since they involve the gaussians g_i. These are, as we have seen, obtained in a most complicated way from the entire set of the $\mathbb{E}_{\tilde{\mathcal{Q}}} z_\mu$, which depend of course on all the ξ_i. It is just fortunate that the covariance structure of the family of gaussians g_i, $i \in V$, is actually deterministic. This means in particular that if we take a fixed configuration of the ξ and pass to the limit, we cannot expect to converge.

Fianlly let us point out that to get propagation of chaos not all what was needed to prove Theorem 8.8 is really necessary. The main fact we used in the proof is the self-averaging of the quantity $\mathbb{E}_{\tilde{\mathcal{Q}}} e^{\beta \sum_{i \in V} s_i(\xi_i, \tilde{z})}$, i.e. essentially (ii) of Lemma 8.6, while (i) is not needed. The second property is that $(\xi_i, \mathbb{E}_{\tilde{\mathcal{Q}}} z)$ converges in law, while it is irrelevant what the limit would be (these random variables might well be dependent). Unfortunately(?), to prove (ii) of Lemma 8.6 requires more or less the same hypotheses as everything else (i.e. we need Theorem 8.1!), so this observation makes little difference. Thus ist may be that propagation of chaos and the exactness of the replica symmetric solution always go together (as the results in [PST] imply).

While in our view the results presented here shed some light on the "mystery of the replica trick", we are still far from understanding the really interesting phenomenon of "replica symmetry breaking". This remains a challenge for the decade to come.

References

[A] D.J. Amit, "Modelling brain function", Cambridge University Press, Cambridge (1989).

[AGS] D.J. Amit, H. Gutfreund and H. Sompolinsky, "Statistical mechanics of neural networks near saturation", Ann. Phys. **173**, 30-67 (1987).

[AGS2] D.J. Amit, H. Gutfreund and H. Sompolinsky, "Spin-glass model of neural network", Phys. Rev. **A 32**, 1007-1018 (1985).

[B] A. Bovier, "Self-averaging in a class of generalized Hopfield models, J. Phys. **A 27**, 7069-7077 (1994).

[BG1] A. Bovier and V. Gayrard, "Rigorous bounds on the storage capacity for the dilute Hopfield model", J. Stat.Phys. **69**, 597-627 (1992)

[BG2] A. Bovier and V. Gayrard, "Rigorous results on the thermodynamics of the dilute Hopfield model", J. Stat. Phys. **72**, 643-664 (1993).

[BG3] A. Bovier and V. Gayrard, "Rigorous results on the Hopfield model of neural networks", Resenhas do IME-USP **2**, 161-172 (1994).

[BG4] A. Bovier and V. Gayrard, "An almost sure large deviation principle for the Hopfield model", Ann. Probab **24**, 1444-1475 (1996).

[BG5] A. Bovier and V. Gayrard, "The retrieval phase of the Hopfield model, A rigorous analysis of the overlap distribution", Prob. Theor. Rel. Fields **107**, 61-98 (1995).

[BG6] A. Bovier and V. Gayrard, in preparation.

[BGP1] A. Bovier, V. Gayrard, and P. Picco, "Gibbs states of the Hopfield model in the regime of perfect memory", Prob. Theor. Rel. Fields **100**, 329-363 (1994).

[BGP2] A. Bovier, V. Gayrard, and P. Picco, "Large deviation principles for the Hopfield model and the Kac-Hopfield model", Prob. Theor. Rel. Fields **101**, 511-546 (1995).

[BGP3] A. Bovier, V. Gayrard, and P. Picco, "Gibbs states of the Hopfield model with extensively many patterns", J. Stat. Phys. **79**, 395-414 (1995).

[BGP4] A. Bovier, V. Gayrard, and P. Picco, "Distribution of overlap profiles in the one-dimensional Kac-Hopfield model", WIAS-preprint 221, to appear in Commun. Math. Phys. (1997).

[BF] A. Bovier and J. Fröhlich, "A heuristic theory of the spin glass phase", J. Stat. Phys. **44**, 347-391 (1986).

[BL] H.J. Brascamp and E.H. Lieb, "On extensions of the Brunn-Minkowski and Pékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation", J. Funct. Anal. **22**, 366-389 (1976).

[Co] F. Comets, "Large deviation estimates for a conditional probability distribution. Applications to random Gibbs measures", Probab. Theor. Rel. Fields **80**, 407-432 (1989).

[CT] Y.S. Chow and H. Teicher, "Probability theory", 2nd edition, Springer, Berlin-Heidelberg-New York (1988)

[DHS] E. Domany, J.L. van Hemmen, and K. Schulten (eds.), "Models of neural networks", Springer Verlag, Berlin (1991).

[DS] J.-D. Deuschel and D. Stroock, "Large deviations", Academic Press, Boston (1989).

[DZ] A. Dembo and O. Zeitouni, Large deviation techniques and applications, Jones and Bartlett, Boston (1992).

[El] R.S. Ellis, "Entropy, large deviations, and statistical mechanics", Springer-Verlag, Berlin (1985).

[EA] S.F. Edwards and P.W. Anderson, "Theory of spin glasses", J. Phys. **F 5**, 965-974 (1975).

[EE] T. Eisele and R.S. Ellis, "Multiple phase transitions in the generalized Curie-Weiss model", J. Stat. Phys. **52**, 161-202 (1988).

[vE] A.C.D. van Enter, "Stiffness exponent, number of pure states, and Almeida-Thouless line in spin glasses", J. Stat. Phys. **60**, 275-279 (1990).

[vEvHP] A.C.D. van Enter, J.L. van Hemmen and C. Pospiech, "Mean-field theory of random- site q-state Potts models", J. Phys. **A 21**, 791-801 (1988).

[FH] D.S. Fisher and D.A. Huse, "Pure phases in spin glasses", J. Phys. **A 20**, L997-L1003 (1987); "Absence of many states in magnetic spin glasses", J. Phys. **A 20**, L1005-L1010 (1987).

[FP1] L.A. Pastur and A.L. Figotin, "Exactly soluble model of a spin glass", Sov. J. Low Temp. Phys. **3(6)**, 378-383 (1977).

[FP2] L.A. Pastur and A.L. Figotin, "On the theory of disordered spin systems", Theor. Math. Phys. **35**, 403-414 (1978).

[FP2] L.A. Pastur and A.L. Figotin, "Infinite range limit for a class of disordered spin systems", Theor. Math. Phys. **51**, 564-569 (1982).

[FZ] C. Fassnacht and A. Zippelius, "Recognition and Categorization in a structured neural network with attractor dynamics", Network **2**, 63-84 (199?).

[G1] V. Gayrard, "The thermodynamic limit of the q-state Potts-Hopfield model with infinitely many patterns", J. Stat. Phys. **68**, 977-1011 (1992).

[G2] V. Gayrard, "Duality formulas and the transfer principle", in preparation.

[Ge] S. Geman, "A limit theorem for the norms of random matrices", Ann. Probab. **8**, 252-261 (1980).

[Geo] H.-O. Georgii, "Gibbs measures and phase transitions", Walter de Gruyter (de Gruyter Studies in Mathematics, Vol. 19), Berlin-New York (1988).

[Gi] V.L. Girko, "Limit theorems for maximal and minimal eigenvalues of random matrices", Theor. Prob. Appl. **35**, 680-695 (1989).

[GK] D. Grensing and K. Kühn, "On classical spin-glass models", J. Physique **48**, 713-721 (1987).

[GM] E. Golez and S. Martínez, "Neural and automata networks", K-luwer Academic Publ., Dodrecht (1990)

[HKP] J. Hertz, A. Krogh, and R. Palmer, "Introduction to the theory of neural computation", Addison-Wesley, Redwood City (1991).

[Ho] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", Proc. Natl. Acad. Sci. USA **79**, 2554-2558 (1982).

[HS] R.L. Stratonovich, "On a method of calculating quantum distribution functions", Doklady Akad. Nauk S.S.S.R. **115**, 1097 (1957)[translation: Soviet Phys. Doklady **2**, 416-419 (1958)],
J. Hubbard, "Calculation of partition functions", Phys. Rev. Lett. **3**, 77-78 (1959).

[JK] J. Jedrzejewski and A. Komoda, "On equivalent-neighbour, random-site models of disordered systems", Z. Phys. **B 63**, 247-257 (1986).

[vH1] J.L. van Hemmen, "Equilibrium theory of spin-glasses: mean-field theory and beyond", in "Heidelberg colloquium on spin glasses", Eds. J.L. van Hemmen and I.Morgenstern, 203-233 (1983), LNP 192 Springer, Berlin-Heidelberg-New York (1983)

[vH2] J.L. van Hemmen, "Spin glass models of a neural network", Phys. Rev. A **34**, 3435-3445 (1986).

[vHGHK] J.L. van Hemmen, D. Grensing, A. Huber and R. Kühn, "Elementary solution of classical spin-glass models, Z. Phys. **B 65**, 53-63 (1986).

[vHvE] J.L. van Hemmen and A.C.D.van Enter, "Chopper model for pattern recognition", Phys.Rev. **A 34**, 2509-2512,(1986).

[vHvEC] J.L. van Hemmen, A.C.D. van Enter, and J. Canisius. "On a classical spin-glass model", Z. Phys. **B 50**, 311-336 (1983).

[K] H. Koch, "A free energy bound for the Hopfield model", J. Phys. **A 26**, L353-L355 (1993).

[KP] H. Koch and J. Piasko, "Some rigorous results on the Hopfield neural network model", J. Stat. Phys. **55**, 903-928 (1989).

[Ku] Ch. Külske, private communiction.

[LT] M. Ledoux and M. Talagrand, "Probability in Banach spaces", Springer, Berlin-Heidelberg-New York (1991).

[Lu] D. Loukianova, "Two rigorous bounds in the Hopfield model of associative memory", to appear in Probab. Theor. Rel. Fields (1996).

[Lut] J.M. Luttinger, "Exactly Soluble Spin-Glass Model", Phys.Rev. Lett. **37**, 778-782 (1976).

[Mar] S. Martinez, "Introduction to neural networks", preprint, Temuco, (1992).

[Ma] D.C. Mattis, "Solvable spin system with random interactions", Phys. Lett. **56A**, 421-422 (1976).

[McE] R.J. McEliece, E.C. Posner, E.R. Rodemich and S.S. Venkatesh, "The capacity of the Hopfield associative memory", IEEE Trans. Inform. Theory **33**, 461-482 (1987).

[Mi] Y. Miyashita, "Neuronal correlate of visual associative long term memory in the primate temporal cortex", Nature **335**, 817-819 (1988).

[MPR] E. Marinari, G. Parisi, and F. Ritort, "On the 3D Ising spin glass", J. Phys. **A 27**, 2687-2708 (1994).

[MPV] M. Mézard, G. Parisi, and M.A. Virasoro, "Spin-glass theory and beyond", World Scientific, Singapore (1988).

[MR] B. Müller and J. Reinhardt, "Neural networks: an introduction", Springer Verlag, Berlin (1990).

[MS] V.A. Malyshev and F.M. Spieksma, "Dynamics of binary neural networks with a finite number of patterns". Part 1: General picture of the asynchronous zero temperature dynamics", MPEJ **3**, 1-36 (1997).

[N] Ch.M. Newman, "Memory capacity in neural network models: Rigorous results", Neural Networks **1**, 223-238 (1988).

[NS] Ch.M. Newman and D.L. Stein, "Non-mean-field behaviour in realistic spin glasses", Phys. Rev. Lett. **76**, 515-518 (1996); "Spatial inhomogeneity and thermodynamic chaos", Phys. Rev. Lett. **76**, 4821-4824 (1996); "Topics in disordered systems", to appear in Birkhäuser, Boston (1997); "Thermodynamic chaos and the structure of short range spion glasses", this volume.

[P] D. Petritis, "Thermodynamic formalism of neural computing", preprint Université de Rennes (1995)

[PS] L. Pastur and M. Shcherbina, "Absence of self-averaging of the order parameter in the Sherrington-Kirkpatrick model", J. Stat. Phys. **62**, 1-19 (1991).

[PST] L. Pastur, M. Shcherbina, and B. Tirozzi, "The replica symmetric solution without the replica trick for the Hopfield model", J. Stat. Phys. **74**, 1161-1183 (1994).

[PN] P. Peretto and J.J. Niez, "Long term memory storage capacity of multiconnected neural networks", Biological Cybernetics **39**, 53-63 (1986).

[Ro] R.T. Rockafellar, "Convex analysis", Princeton University Press, Princeton (1970).

[RV] A.W. Roberts and D.E. Varberg, "Convex functions", Academic Press, New York and London (1973).

[SK] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin glass", Phys. Rev. Lett. **35**, 1792-1796 (1972).

[Si] J. Silverstein, "Eigenvalues and eigenvectors of large sample covariance matrices", Contemporary Mathematics **50**, 153-159 (1986).

[Sn] A.-S. Snitzman, "Equations de type Boltzmann spatialement homogènes", Z. Wahrscheinlichkeitstheorie verw. Gebiete **66**, 559-592

(1986).

[ST] M. Shcherbina and B. Tirozzi, "The free energy for a class of Hopfield models", J. Stat. Phys. **72**, 113-125 (1992).

[SW] M. Schlüter and E. Wagner, Phys. Rev. **E49**, 1690 (1994).

[Yu] V.V. Yurinskii, "Exponential inequalities for sums of random vectors", J. Multivariate Anal. **6**, 473-499 (1976).

[T1] M. Talagrand, "Concentration of measure and isoperimetric inequalities in product space", Publ. Math. I.H.E.S., **81**, 73-205 (1995).

[T2] M. Talagrand, "A new look at independence", Ann. Probab. **24**, 1-34 (1996).

[T3] M. Talagrand, "Résultats rigoureux pour le modèle de Hopfield", C. R. Acad. Sci. Paris, **t. 321, Série I**, 109-112 (1995).

[T4] M. Talagrand, "Rigorous results for the Hopfield model with many patterns", preprint 1996, to appear in Probab. Theor. Rel. Fields.

[T5] M. Talagrand, "The Sherrington-Kirkpatrick model: A challenge for mathematicians", preprint 1996, to appear in Prob. Theor. Rel. Fields.

[TAP] D.J. Thouless, P.W. Anderson, and R.G. Palmer, Phil. Mag. **35**,593 (1977).

[YBK] Y.Q. Yin, Z.D. Bai, and P.R. Krishnaiah, "On the limitof the largest eigenvalue of the large dimensional sample covariance matrix", Probab. Theor. Rel. Fields **78**, 509-521 (1988).

[Yu] V.V. Yurinskii, "Exponential inequalities for sums of random vectors", J. Multivariate Anal. **6**, 473-499 (1976).

Weierstrass-Insitut für Angewandte Analysis und Stochastik, Mohrenstrasse 39, 10117 Berlin, Germany

Centre de Physique Théorique-CNRS, Lunimy, Case 907, F-13288 Marseille Cedex 9, France