AN ALMOST SURE CENTRAL LIMIT THEOREM FOR THE HOPFIELD MODEL[#]

Anton Bovier [1] **and Véronique Gayrard**[2,3,4]

Weierstraß–Institut
für Angewandte Analysis und Stochastik
Mohrenstrasse 39, D-10117 Berlin, Germany

Abstract: We prove a central limit theorem for the finite dimensional marginals of the Gibbs distribution of the macroscopic 'overlap'-parameters in the Hopfield model in the case where the number of random 'patterns', M, as a function of the system size N satisfies $\lim_{N\uparrow\infty} M(N)/N = 0$, without any assumptions on the speed of convergence. The covariance matrix of the limiting gaussian distributions is diagonal and independent of the disorder for almost all realizations of the patterns.

Keywords: Hopfield model, neural networks, central limit theorem, Brascamp-Lieb inequalities

AMS Subject Classification: 60F05, 60K35

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

1 e-mail: bovier@wias-berlin.de

2 e-mail: gayrard@wias-berlin.de,gayrard@cpt.univ-mrs

3 Permanent address: CPT-CNRS, Lunimy, Case 907, F-13288 Marseille Cedex, France

4 Supported by Graduiertenkolleg "Stochastische Prozesse und Probabilistische Analysis", Berlin

I. Introduction

Let us recall the definitions of the Hopfield model [Ho] and the main quantities of interest. For a more detailed exposition of the model we refer to [BG3]. Let N be an integer and $M:I\!N \to I\!N$ be a strictly increasing function. We set $\alpha(N) \equiv \frac{M(N)}{N}$. In the present work we will consider only the case where $\lim_{N\uparrow\infty} \alpha(N) = 0$. We denote by $\mathcal{S}_N \equiv \{-1,1\}^N$ and $\mathcal{S} \equiv \{-1,1\}^{I\!N}$ the set of spin configurations, σ, in finite, resp. infinite volume. We denote by σ_i the value of σ at i. Let $(\Omega, \mathcal{F}, I\!P)$ be an abstract probability space and let $\{\xi_i^\mu[\omega], i, \mu \in I\!N\}$, denote a family of independent identically distributed random variables on this space. For the purposes of this paper we will assume that $I\!P[\xi_i^\mu = \pm 1] = \frac{1}{2}$, but more general distributions can be considered.

We define random maps $m_N[\omega] : \mathcal{S}_N \to [-1,1]^{M(N)}$ whose components are given by

$$m_N^\mu[\omega](\sigma) \equiv \frac{1}{N} \sum_{i=1}^N \xi_i^\mu[\omega]\sigma_i \quad , \mu = 1, \dots, M(N). \tag{1.1}$$

The Hamiltonian of the Hopfield model is now defined as

$$\begin{aligned} H_N[\omega](\sigma) &\equiv -\frac{N}{2} \sum_{\mu=1}^{M(N)} \left(m_N^\mu[\omega](\sigma)\right)^2 \\ &= -\frac{N}{2} \|m_N[\omega](\sigma)\|_2^2 \end{aligned} \tag{1.2}$$

where $\|\cdot\|_2$ denotes the ℓ_2-norm in $I\!R^M$. With this Hamiltonian we define in a natural way finite volume Gibbs measures on $(\mathcal{S}_N, \mathcal{B}(\mathcal{S}_N))$ via

$$\mu_{N,\beta}[\omega](\sigma) \equiv \frac{2^{-N}}{Z_{N,\beta}[\omega]} e^{-\beta H_N[\omega](\sigma)} \tag{1.3}$$

where the parameter $\beta > 0$ denotes the inverse temperature and where the normalizing factor $Z_{N,\beta}[\omega]$ is given by

$$Z_{N,\beta}[\omega] \equiv 2^{-N} \sum_{\sigma \in \mathcal{S}_N} e^{-\beta H_N[\omega](\sigma)} \equiv I\!E_\sigma e^{-\beta H_N[\omega](\sigma)} \tag{1.4}$$

We furthermore introduce the measures on $(I\!R^{M(N)}, \mathcal{B}(I\!R^{M(N)}))$ induced by the Gibbs measures and the maps $m_N[\omega]$:

$$\mathcal{Q}_{N,\beta}[\omega] \equiv \mu_{N,\beta}[\omega] \circ m_N[\omega]^{-1} \tag{1.5}$$

Over the last few years a very satisfactory and complete description of the measures $\mathcal{Q}_{N,\beta}[\omega]$ has been obtained in the case $\lim_{N\uparrow 0} \frac{M(N)}{N} = 0$. In particular, in [BGP1], a law of large number type was proven for the random vectors $m_N[\omega]$, and in [BG1] the associated full large deviation principle was obtained, without any condition on the speed of convergence of $\frac{M(N)}{N}$ to zero. In such a situation it is natural to also expect a central limit theorem to hold. Such results were in fact proven in several papers by B. Gentz [Ge1], [Ge2] and [Ge3]. However, they required strong conditions on the speed at which $\frac{M(N)}{N}$ tends to zero, the weakest being $\lim_{N\uparrow 0}\frac{M^2(N)}{N}=0$ in [Ge3]. In this note we show that the central limit theorem holds under the sole hypothesis that $\lim_{N\uparrow 0}\frac{M(N)}{N}=0$. Thus, in this regime all the classical theorems of probability theory are now established.

We note that the proof of the CLT requires a far more detailed analysis of the local properties of the measures $\mathcal{Q}_{N,\beta}$ then all previous results in the same regime. The crucial ingredient is a local convexity estimate that was given in [BG2] and the crucial new analytic tool are Brascamp-Lieb inequalities [BL,HS,N,NS].

In order to state the results we need some more notation and definitions. Let $m^*(\beta)$ be the largest solution of the mean field equation $m=\tanh(\beta m)$. Note that $m^*(\beta)$ is strictly positive for all $\beta>1$, $\lim_{\beta\uparrow\infty}m^*(\beta)=1$, $\lim_{\beta\downarrow 1}\frac{(m^*(\beta))^2}{3(\beta-1)}=1$ and $m^*(\beta)=0$ if $\beta\leq 1$. Denoting by e^μ the μ-th unit vector of the canonical basis of $I\!R^M$ we set, for all $(\mu,s)\in\{-1,1\}\times\{1,\dots,M(N)\}$,

$$m^{(\mu,s)}\equiv sm^*(\beta)e^\mu, \tag{1.6}$$

and for any $\rho>0$ we define the balls

$$B_\rho^{(\mu,s)}\equiv\left\{x\in I\!R^M \,\middle|\, \|x-m^{(\mu,s)}\|_2\leq\rho\right\} \tag{1.7}$$

For any pair of indices (μ,s) and any $\rho>0$ we define the conditional measures[1]

$$\mathcal{Q}_{N,\beta,\rho}^{(\mu,s)}[\omega](\mathcal{A})\equiv\mathcal{Q}_{N,\beta}[\omega](\mathcal{A}\mid B_\rho^{(\mu,s)}),\quad \mathcal{A}\in\mathcal{B}(I\!R^{M(N)}) \tag{1.8}$$

Let X_N be a random vector distributed according to $\mathcal{Q}_{N,\beta,\rho}^{(\mu,s)}[\omega]$ and denote by $\overline{X}_{N,\beta,\rho}^{(\mu,s)}[\omega]$ it's expectation. We want to characterize the distribution of the normalized centered variable

$$\widetilde{X}_N\equiv\sqrt{N}(X_N-\overline{X}_{N,\beta,\rho}^{(\mu,s)}[\omega]) \tag{1.9}$$

To do so we consider it's Laplace transform (recall that $\widetilde{X}$ is $M(N)$-dimensional):

$$\mathcal{L}_{N,\beta,\rho}^{(\mu,s)}[\omega](t)\equiv\int e^{\sqrt{N}(t,x-\overline{X}_{N,\beta,\rho}^{(\mu,s)}[\omega])}d\mathcal{Q}_{N,\beta,\rho}^{(\mu,s)}[\omega](x)\,,\quad t\in I\!R^{M(N)} \tag{1.10}$$

where $(\cdot,\cdot)$ stands for the scalar product in $I\!R^{M(N)}$. We prove the following theorem:

---

[1] All the results of this paper could also be formulated in terms of "tilted Gibbs measure", i.e. with a symmetry breaking magnetic field added instead of the conditioning (see [BG3]) for precise definitions.

Theorem 1.1: *Assume* $\lim_{N\uparrow\infty}\alpha(N)=0$. *Assume that* $\beta\in I\!R^+\backslash\{1\}$ *and set*

$$\mathcal{C}(\beta)\equiv\begin{cases}\frac{1-(m^*(\beta))^2}{1-\beta(1-(m^*(\beta))^2)} & \text{if } \beta>1\\ \frac{1}{1-\beta} & \text{if } \beta<1\end{cases} \tag{1.11}$$

There exists a constant $c(\beta)>0$ *such that with probability one, for all but a finite number of indices* N, *if* ρ *satisfies*

$$\frac{1}{2}m^*>\rho>c(\beta)\left\{\tfrac{1}{N^{1/4}}\wedge\sqrt{\alpha(N)}\right\}+c_0\frac{\sqrt{\alpha(N)}}{m^*} \tag{1.12}$$

for some constant $c_0>0$, *then for all* t *with* $\|t\|_2<\infty$ *we have*

$$\lim_{N\uparrow\infty}\log\mathcal{L}^{(\mu,s)}_{N,\beta,\rho}[\omega](t)=\tfrac{1}{2}\mathcal{C}(\beta)\|t\|_2^2 \tag{1.13}$$

Corollary 1.2: *Under the assumptions of Theorem 1, for all* $k\in I\!N$, *the finite dimensional marginals of order* k *of the law of* $\widetilde{X}_N$ *under* $\mathcal{Q}^{(\mu,s)}_{N,\beta,\rho}[\omega]$ *converge weakly, as* N *diverges, to the gaussian measure on* $(I\!R^k,\mathcal{B}(I\!R^k))$ *with mean zero and covariance matrix* $\mathcal{C}(\beta)1\!\!1$ *where* $1\!\!1$ *is the identity matrix.*

Remark: The same result was obtained in [Ge3] under the stronger assumption $\lim_{N\uparrow 0}\frac{M^2(N)}{N}=0$.

We will see in the sequel that, due to the sharp concentration properties of the measure $\mathcal{Q}^{(\mu,s)}_{N,\beta,\rho}[\omega]$, the centering $\overline{X}^{(\mu,s)}_{N,\beta,\rho}[\omega]$ obeys the following bound:

Lemma 1.3: *Under the assumption of Theorem 1.1, with probability one, for all but a finite number of indices* N,

$$\left\|\overline{X}^{(\mu,s)}_{N,\beta,\rho}[\omega]-m^{(\mu,s)}\right\|_2\leq\tilde{\rho} \tag{1.14}$$

where

$$\tilde{\rho}\equiv\tilde{c}_0\frac{\sqrt{\alpha(N)}}{m^*} \tag{1.15}$$

for some constant $\tilde{c}_0>0$.

The remainder of this paper is organized as follows. We only present the proof of Theorem 1 in the case where $\beta>1$, the case $\beta<1$ being trivial [2]. Moreover, in order to avoid having to distinguish several cases and since we are mainly interested in the regime of parameters not covered in [Ge3] , we will assume that $M(N)>(\log N)^2$. It is however not difficult at all to treat the case $M(N)\leq(\log N)^2$. In fact, wherever estimates of the form e^{-cM} appear, they can be replaced by $e^{-c\sqrt{N}}$ if so desired by trivial modifications. The basic structure of the proof is as follows:

[2] The situation at $\beta=1$ as well as the limits $\beta\to 1$ taken in various ways are up tp now completely uninvestigated and promise a rather rich and complex structure.

(i) Using the Hubbard-Stratonovich transformation, show that for ρ chosen as in (1.12), the Laplace transform (1.10), $\mathcal{L}^{(\mu,s)}_{N,\beta,\rho}$, can be expressed in terms of the Laplace transform $\widetilde{\mathcal{L}}^{(\mu,s)}_{N,\beta,\rho}$ of a smoothed version $\widetilde{\mathcal{Q}}^{(\mu,s)}_{N,\beta,\rho}$ of the measure $\mathcal{Q}^{(\mu,s)}_{N,\beta,\rho}$.

(ii) Show that the measures $\widetilde{\mathcal{Q}}^{(\mu,s)}_{N,\beta,\rho}$ for all ρ satisfying (1.12) are equivalent.

(iii) Choose ρ as the lower bound in (1.12) and, using the results of [BG2], show that the corresponding measures have densities of the forms $e^{-NV(x)}$ with V strictly convex; moreover, the Hessian of V is uniformly close to a multiple of the identity.

(iv) The Brascamp-Lieb inequalities, together with a simple reverse [DGI], now yield asymptotically coinciding upper and lower bounds on the Laplace transform which imply Theorem 1.1.

Assuming (ii), we present (i), (iii) and (iv) in Section 2. This represents the essential and original part of the proof. While the results of (ii) use by now quite standard techniques and are not very original, they require rather lengthy computations. We give them in Section 3; readers not interested in these technicalities are advised not to read that section.

Notation and conventions: Before giving the proofs, let us fix some general conventions on notation. From now on the parameter ρ in $\overline{X}^{(\mu,s)}_{N,\beta,\rho}[\omega]$ is fixed and chosen as in (1.12). We will then simply write

$$\overline{X}^{(\mu,s)}[\omega] \equiv \overline{X}^{(\mu,s)}_{N,\beta,\rho}[\omega] \tag{1.16}$$

and no confusion should arise from this. In general, in order not to overburden the notation, we will suppress part of or all of the subscripts β, N, ρ when we feel that this cannot be confusing. We will also often suppress the explicit dependence of several quantities on β and N: mostly we will write $m^* \equiv m^*(\beta)$, $M \equiv M(N)$, $\alpha \equiv \alpha(N)$. Finally, let us insist that to simplify the notation, the dependance of various random quantities on ω will be made explicit only when we want to stress the random nature of these quantities.

Acknowledgements: We thank J.-D. Deuschel, G. Giacomin and D. Ioffe for informing us of their results in [DGI] concerning the reverse Brascamp-Lieb inequalities prior to publication.

2. Proof of Theorem 1.1

In this section we give the main part of the proof of Theorem 1.1. We recall first the Hubbard-Stratonovich transformation [H,S].

Let $\mathcal{N}^M_{\beta N}$ be the gaussian measure on $(I\!R^M, \mathcal{B}(I\!R^M))$ with density $\left(\frac{\beta N}{2\pi}\right)^{M/2} \exp\left\{-\frac{1}{2}\beta N\|z\|_2^2\right\}$ with respect to Lebesgue measure in $I\!R^M$. The Hubbard-Stratonovich approach consists in considering the convolution

$$\widetilde{\mathcal{Q}}_{N,\beta} \equiv \mathcal{Q}_{N,\beta} \star \mathcal{N}_{\beta N}^{M} \tag{2.1}$$

instead of the measure $\mathcal{Q}_{N,\beta}$ itself. The resulting measure $\widetilde{\mathcal{Q}}_{N,\beta}$ is absolutely continuous and has density

$$\frac{1}{Z_{N,\beta}} \exp\left\{-\Phi_{N,\beta}(z)\right\} \tag{2.2}$$

with respect to Lebesgue's measure in $I\!R^M$. The function $\Phi_{N,\beta}(z)$ can be computed explicitly and is given by

$$\Phi_{N,\beta}(z) = \frac{1}{2}\|z\|_2^2 - \frac{1}{\beta N}\sum_{i=1}^{N} \ln\cosh\beta(\xi_i, z)\,, \quad z \in I\!R^M \tag{2.3}$$

Note that under our assumptions on α, the measures $\widetilde{\mathcal{Q}}_{N,\beta}$ and $\mathcal{Q}_{N,\beta}$ have the same convergence properties as for large enough N, the gaussian $\mathcal{N}_{\beta N}^{M}$ gets concentrated sharply on a sphere of radius $\sqrt{\alpha/\beta}$.

In complete analogy with (1.8) to (1.10), we introduce the conditional measures

$$\widetilde{\mathcal{Q}}^{(\mu,s)}_{N,\beta,\rho}(\mathcal{A}) \equiv \mathcal{Q}_{N,\beta}(\mathcal{A} \mid B_\rho^{(\mu,s)}), \quad \mathcal{A} \in \mathcal{B}(I\!R^M) \tag{2.4}$$

and, for Z_N distributed according to $\widetilde{\mathcal{Q}}^{(\mu,s)}_{N,\beta,\rho}$ we consider the Laplace transform

$$\widetilde{\mathcal{L}}^{(\mu,s)}_{N,\beta,\rho}(t) \equiv \int e^{\sqrt{N}(t, z - \overline{Z}^{(\mu,s)})} d\widetilde{\mathcal{Q}}^{(\mu,s)}_{N,\beta,\rho}(z)\,, \quad t \in I\!R^M \tag{2.5}$$

of the normalized centered variable $\widetilde{Z}_N \equiv \sqrt{N}(Z_N - \overline{Z}^{(\mu,s)})$, where $\overline{Z}^{(\mu,s)}$ is the expectation of Z_N. For later convenience, we also introduce the quantities

$$\widetilde{\mathcal{L}}^{(\mu,s)}_{N,\beta,\rho}(t) \equiv e^{\sqrt{N}(t, \overline{Z}^{(\mu,s)} - \overline{X}^{(\mu,s)})} \widetilde{\mathcal{L}}^{(\mu,s)}_{N,\beta,\rho}(t) \tag{2.6}$$

We will proof in the remainder of this section the analog of Theorem 1.1 for the function $\widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\bar{\rho}}(t)$ with $\bar{\rho} = \bar{\rho}(N)$ that tends to zero as N tends to infinity.

The following proposition, whose proof will be given in Section 3, assures that this implies that $\mathcal{L}^{(\mu,s)}_{\beta,N,\bar{\rho}}(t)$ converges to the same limit.

Proposition 2.1: *Assume that $\beta > 1$. There exist finite*

positive constants $c \equiv c(\beta), \tilde{c} \equiv \tilde{c}(\beta), \bar{c} \equiv \bar{c}(\beta)$ such that, with a probability one, for all but a finite number of indices N, if ρ satisfies

$$\frac{1}{2}m^* > \rho > c(\beta)\left\{\tfrac{1}{N^{1/4}} \wedge \sqrt{\alpha}\right\} \tag{2.7}$$

then, for all t with $\|t\|_2 < \infty$,

i)
$$\mathcal{L}^{(\mu,s)}_{\beta,N,\rho}(t)\left(1-e^{-\tilde{c}M}\right) \le e^{-\frac{1}{2\beta}\|t\|_2^2}\widehat{\mathcal{L}}^{(\mu,s)}_{\beta,N,\rho}(t) \le e^{-\tilde{c}M} + \mathcal{L}^{(\mu,s)}_{\beta,N,\rho}(t)\left(1+e^{-\tilde{c}M}\right) \tag{2.8}$$

ii) for any $\bar{\rho}$ satisfying (2.7)
$$\widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\bar{\rho}}(t)\left(1-e^{-\bar{c}M}\right) \le \widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\rho}(t) \le e^{-\bar{c}M} + \widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\bar{\rho}}(t)\left(1+e^{-\bar{c}M}\right) \tag{2.9}$$

iii) for any $\bar{\rho}$ satisfying (2.7)
$$\left|\left(\overline{X}^{(\mu,s)} - \overline{Z}^{(\mu,s)}_{\rho}, t\right)\right| \le \|t\|_2 e^{-\bar{c}M} \tag{2.10}$$

Remark: Note that (iii) implies that
$$\left|\ln \widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\rho}(t) - \ln \widehat{\mathcal{L}}^{(\mu,s)}_{\beta,N,\rho}(t)\right| \le \sqrt{N}\|t\|_2 e^{-\bar{c}M} \tag{2.11}$$

which under our assumption $M(N) \ge (\ln N)^2$ tends to zero.

We now want to compute the Laplace transform $\widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\bar{\rho}}(t)$ for $\bar{\rho} \equiv \bar{\rho}(N)$ that tends to zero as N tends to infinity.

Proposition 2.2: *Assume that $\beta \in I\!R^+ \backslash \{1\}$ and set:*
$$\lambda(\beta) \equiv 1 - \beta(1-(\beta m^*(\beta))^2)\ . \tag{2.12}$$

Let $\alpha(N)$ and $\bar{\rho}(N)$ be decreasing functions of N that go to zero as N goes to infinity and satisfy
$$\bar{\rho}(N) \ge 2\sqrt{\tfrac{\alpha(N)}{\lambda(\beta)}}\ . \tag{2.13}$$

Then with probability one, for all but a finite number of indices N,
$$\frac{\|t\|_2(1-3\bar{\rho}e^{-M})}{2\beta(\lambda(\beta)+\gamma(N))} \le \ln \widetilde{\mathcal{L}}^{(\mu,s)}_{\beta,N,\bar{\rho}}(t) \le \tfrac{\|t\|_2(1+2\bar{\rho}e^{-M})}{2\beta(\lambda(\beta)-\gamma(N))} \tag{2.14}$$

where
$$\gamma(N) = \beta\left[3(\sqrt{\bar{\rho}}+\sqrt{\alpha}) + c\sqrt{\tfrac{\ln N}{N}} + c'\tfrac{8}{m^*}\bar{\rho}\right] \tag{2.15}$$

for some strictly positive constants c and c'.

We recall a few notation and definitions. Let S and T be two $M \times M$ real symmetric matrices. The matrix norm is defined by
$$\|T\| \equiv \sup_{x:\|x\|_2=1} |(x,Tx)| \tag{2.16}$$

We say that T is non negative, and we write $T \geq 0$, if $(x, Tx) \geq 0$ for any $x \in I\!R^M$. More generally, we say that $T \geq S$ or $S \leq T$ if $T - S \geq 0$. For any function V: $I\!R^M \to I\!R$, we will denote by $\nabla^2 V(x)$ it's Hessian matrix at x.

Lemma 2.3: *Let $\alpha(N)$ and $\rho(N)$ be decreasing functions of N that go to zero as N goes to infinity. Assume that $\beta \neq 1$. Then with a probability one, for all but a finite number of indices N, for all v in the set $\{v \in I\!R^M : \|v\|_2 \leq \rho(N)\}$, we have:*

$$0 < (\lambda(\beta) - \gamma(N))\mathbb{1} \leq \nabla^2 \Phi_{\beta,N}(m^{(1,1)} + v) \leq (\lambda(\beta) + \gamma(N))\mathbb{1} \tag{2.17}$$

where $\gamma(N)$ is defined in (2.15).

Proof: We will only give the proof of the upper bound. The proof of the lower bound is very similar and can already be found in [BG2], [BG3]. A straightforward computation gives

$$\nabla^2 \Phi_{\beta,N}(m^{(1,1)} + v) = \mathbb{1} - \beta A + \frac{\beta}{N} \sum_{i=1}^{N} \xi_i^t \xi_i \tanh^2(\beta(m^* \xi_i^1 + (\xi_i, v))) \tag{2.18}$$

Our strategy will be to show that $\nabla^2 \Phi$ can be rewritten as it's dominant contribution, $\lambda(\beta)\mathbb{1}$, plus terms that will either have small norm or be non negative. We will then make use the two following facts: for any real symmetric matrices T and S,

i) if $S \geq 0$ then $T + S \geq T$.

ii) if $T = t\mathbb{1} + S$ with $\|S\| \leq s$ for some constants t and s, then $T \geq (t - s)\mathbb{1}$.

Introducing a parameter $0 < \tau \ll 1$ that will be appropriately chosen later, we decompose $\nabla^2 \Phi$ as

$$\nabla^2 \Phi_{\beta,N}(m^{(1,1)} + v) = \lambda(\beta)\mathbb{1} + T_1 + T_2 + T_3 + T_4 + T_5 \tag{2.19}$$

where

$$\begin{aligned}
T_1 \equiv& \beta[\tanh^2(\beta m^*(1-\tau)) - \tanh^2(\beta m^*)]\mathbb{1} \\
T_2 \equiv& \beta(1 - \tanh^2(\beta m^*(1-\tau)))(\mathbb{1} - A) \\
T_3 \equiv& -\tanh^2(\beta m^*(1-\tau))\frac{\beta}{N}\sum_{i=1}^{N} \xi_i^t \xi_i \mathbb{1}_{\{|(\xi_i, v)| \geq \tau m^*\}} \\
T_4 \equiv& \tfrac{\beta}{N}\sum_{i=1}^{N} \xi_i^t \xi_i \mathbb{1}_{\{|(\xi_i, v)| < \tau m^*\}}[\tanh^2(\beta(m^* \xi_i^1 + (\xi_i, v))) - \tanh^2(\beta m^*(1-\tau))] \\
T_5 \equiv& \tfrac{\beta}{N}\sum_{i=1}^{N} \xi_i^t \xi_i \mathbb{1}_{\{|(\xi_i, v)| \geq \tau m^*\}} \tanh^2(\beta(m^* \xi_i^1 + (\xi_i, v)))
\end{aligned} \tag{2.20}$$

It is easy to verify that $T_4 \geq 0$ and $T_5 \geq 0$. Thus, by i),

$$\nabla^2 \Phi_{\beta,N}(m^{(1,1)} + v) \geq \lambda(\beta) \mathbb{1} + T_1 + T_2 + T_3 \tag{2.21}$$

and we are left to show that T_1, T_2 and T_3 have small norms. Let us treat T_1 first. Trivially,

$$\begin{aligned} \|T_1\| \equiv & \beta |\tanh^2(\beta m^*(1-\tau)) - \tanh^2(\beta m^*)| \\ \leq & 2\beta |\tanh(\beta m^*(1-\tau)) - \tanh(\beta m^*)| \leq \frac{\beta\tau}{1-\tau} \end{aligned} \tag{2.22}$$

Let $A(N) \equiv A(N)$ denote the $M \times M$ random matrix with elements $\frac{1}{N}\sum_{i=1}^N \xi_i^\mu \xi_i^\nu$ The smallness of $\|T_2\|$ comes from the well know fact (see e.g. [G]) that, for small α, the matrix $A(N)$ is very close to the identity. In particular, it follows from Theorem 4.1 of [BG3] that, for large enough N, there exists a numerical constant K such that, for all $\epsilon \geq 0$,

$$I\!P\left[\|A(N) - \mathbb{1}\| \geq 2\sqrt{\alpha} + \alpha + \epsilon\right] \leq K \exp\left(-N\frac{(1+\sqrt{\alpha})^2}{K}\left(\sqrt{\frac{\epsilon}{1+\sqrt{\alpha}} + 1} - 1\right)^2\right) \tag{2.23}$$

In particular, choosing $\epsilon \equiv c\sqrt{\ln N/N}$ for some constant $c > 0$ sufficiently large, (2.23) reduces to

$$I\!P\left[\|A(N) - \mathbb{1}\| \geq 2\sqrt{\alpha} + \alpha + c\sqrt{\ln N/N}\right] \leq \frac{1}{N^2}$$

Finally we are left to estimate $\|T_3\|$. But this was already done in [BG2] (see equations (4.77)-(4.79) together with Proposition 4.8). We rephrase this result hereafter in the particular (simpler but weaker) form we need: for all $\rho \geq 0$,

$$I\!P\left[\sup_{v \in B_\rho} \|\sum_{i=1}^N \xi_i^t \xi_i \mathbb{1}_{\{|(\xi_i, v)| \geq \tau m^*\}}\| \geq 2\Gamma(\alpha, \tau m^*/\rho)\right] \leq \frac{4}{N^2} \tag{2.24}$$

where

$$\Gamma(\alpha, a/\rho) \leq C\left[e^{-(1-2\sqrt{\alpha})^2 \frac{a^2}{4\rho^2}} + \alpha(|\ln \alpha| + 2) + 2\frac{\ln N}{N}\right] \tag{2.25}$$

for some constant $C < \infty$ ($C \approx 25$). Therefore, collecting (2.22), (2.23), (2.24) and the definitions of T_2 and T_3 we have, with a probability larger than $1 - \frac{5}{N^2}$,

$$\|T_1\| + \|T_2\| + \|T_3\| \leq \frac{\beta\tau}{1-\tau} + \beta(2\sqrt{\alpha} + \alpha + c\sqrt{\ln N/N}) + 2\beta\Gamma(\alpha, \tau m^*/\rho) \tag{2.26}$$

where we made used of the trivial bounds $0 \leq \tanh^2 x \leq 1$. It only remains to choose the parameter τ. Setting $\tau \equiv \sqrt{\rho}$, we get that, for large enough N,

$$\|T_1\| + \|T_2\| + \|T_3\| \leq \beta\left[2\sqrt{\rho} + \left(2\sqrt{\alpha} + \alpha + c\sqrt{\frac{\ln N}{N}}\right) + 25\left(\frac{8}{m^*}\rho + \alpha(|\ln \alpha| + 2) + 2\frac{\ln N}{N}\right)\right] \tag{2.27}$$

where the r.h.s. is easily seen to be bounded by $\gamma(N)$ if α and ρ are small. The lower bound in (2.17) then follows from (2.27) and (2.21) by ii) and an application of the Borel-Cantelli Lemma. This concludes the proof of Lemma 2.3 $\diamondsuit$

The following slight generalization of the Brascamp-Lieb inequalities will be our crucial tool to exploit Lemma 2.3.

Lemma 2.4: *Assume that the positive numbers ℓ, δ and ρ satisfy the relations*

$$\rho > K\sqrt{\tfrac{\alpha(N)}{\ell-\delta}} \tag{2.28}$$

where

$$K\frac{\ell-\delta}{\ell+\delta} \geq 2 + \ln K \tag{2.29}$$

Let $V\colon \mathbb{R}^M \to \mathbb{R}$ be a non-negative function such that for all $x \in B_\rho$

$$0 < (\ell-\delta)\mathbb{1} \leq \nabla^2 V(x) \leq (\ell+\delta)\mathbb{1} \tag{2.30}$$

Denote by $\mathbb{E}_V$ the expectation with respect to the probability measure on $(B_\rho, \mathcal{B}(B_\rho))$

$$\frac{e^{-NV(x)}\mathbb{1}_{\{x\in B_\rho\}}}{\int_{B_\rho} e^{-NV(x)}d^M x} d^M x \tag{2.31}$$

Then, for any $t \in \mathbb{R}^M$,

$$\frac{\|t\|_2^2}{\ell+\delta} - er \leq \mathbb{E}_V\left(\sqrt{N}(t, x - \mathbb{E}_V(x))\right)^2 \leq \frac{\|t\|_2^2}{\ell-\delta}1 + er \tag{2.32}$$

and

$$\frac{\|t\|_2^2}{2(\ell+\delta)}1 - er \leq \ln \mathbb{E}_V e^{\sqrt{N}(t, x - \mathbb{E}_V(x))} \leq \frac{\|t\|_2^2}{2(\ell-\delta)} + er \tag{2.33}$$

where

$$er \equiv \frac{2\rho e^{-M}}{\ell-\delta}\|t\|_2^2 \tag{2.34}$$

Proof: Let us first consider (2.32). Note that the upper bound would simply obtain from an application of the Brascamp-Lieb inequalities [BL] if it were not that the measure (2.31) has support in a ball of finite radius. Because of this we will have to be a little more careful and take into account "boundary effects" which, as we shall see, do nothing but create asymptotically negligible small terms. Similarly, the lower bound will essentially result from a "reverse" Brascamp-Lieb inequality recently obtained by [DGI]. Our proof is based on a representation which was originally introduced by Helffer and Sjöstrand [HS]. It was recently used by Naddaf [N] and Naddaf and Spencer [NS] who noticed in particular that this representation provides a very simple way of proving the Brascamp-Lieb inequalities.

We proceed exactly as in [HS]: given a temperate function $f{:}I\!R^M \to I\!R$ and a constant b, we consider the differential equation

$$f = Lu + b \tag{2.35}$$

where the operator L is defined as

$$L \equiv -(\nabla V)\cdot\nabla + \Delta = e^{NV}\nabla e^{-NV}\nabla$$

Then observe that integrating by parts,

$$b = I\!E_V f - \int_{B_\rho} \nabla\cdot\left((\nabla u(x))\frac{e^{-NV(x)}}{\int_{B_\rho} e^{-NV(x)}d^Mx}\right) d^Mx \tag{2.36}$$

Assume that u is a solution of (2.35). Integrating by parts again, the correlation $I\!E_V[fg]$ of two temperate functions f and g with (to simplify) $I\!E_V f = I\!E_V g = 0$ can be expressed as

$$I\!E_V[fg] = I\!E_V(\nabla g, \nabla u) + \int_{B_\rho} \nabla\cdot\left(g(\nabla u(x))\frac{e^{-NV(x)}}{\int_{B_\rho} e^{-NV(x)}d^Mx}\right) d^Mx \tag{2.37}$$

while

$$\nabla f = (L + \nabla^2 V)\nabla u \tag{2.38}$$

But L is positive and, by assumption, so is $\nabla^2 V(x)$ for all x in B_ρ. Therefore $L + \nabla^2 V$ is invertible and (2.37) together with (3.53) entails

$$\begin{aligned} I\!E_V[fg] &= \frac{1}{N} I\!E_V(\nabla g, [L+\nabla^2 V(x)]^{-1}\nabla f) + \int_{B_\rho} \nabla\cdot\left(g(\nabla u(x))\frac{e^{-NV(x)}}{\int_{B_\rho} e^{-NV(x)}d^Mx}\right) d^Mx \\ &\equiv [1] + [2] \end{aligned} \tag{2.39}$$

We are now ready to prove (2.32). Set $f(x) = g(x) = \sqrt{N}(t, x - I\!E_V(x))$. Then

$$[1] \equiv \frac{1}{N} I\!E_V(\nabla f(x), [L+\nabla^2 V(x)]^{-1}\nabla f(x)) = I\!E_V(t, [L+\nabla^2 V(x)]^{-1} t) \tag{2.40}$$

Upper and lower bounds on the latter quantity are easily established. From the positivity of L and the assumption that $\nabla^2 V(x) \geq (\ell - \delta)1\!\!1$ uniformly in the ball B_ρ, it immediately follows that

$$I\!E_V(t, [L+\nabla^2 V(x)]^{-1} t) \leq \frac{\|t\|_2^2}{(\ell-\delta)} \tag{2.41}$$

On the other hand, as noted in [DGI], the Legendre-Fenchel transform of the quadratic form $\frac{1}{2}(t, [L+\nabla^2 V(x)]^{-1} t)$ being well defined we can write:

$$\begin{aligned} I\!E_V(t, [L+\nabla^2 V(x)]^{-1} t) =& 2 I\!E_V\left\{ \sup_{t^*\in I\!R^M} (t^*, t) - \tfrac{1}{2}(t^*, [L+\nabla^2 V(x)]t^*) \right\} \\ \geq& 2 \sup_{t^*\in I\!R^M} I\!E_V\left[(t^*, t) - \tfrac{1}{2}(t^*, [L+\nabla^2 V(x)]t^*)\right] \\ =& 2 \sup_{t^*\in I\!R^M} I\!E_V\left[(t^*, t) - \tfrac{1}{2}(t^*, \nabla^2 V(x) t^*)\right] \\ \geq& 2 \sup_{t^*\in I\!R^M} \left\{(t^*, t) - \tfrac{1}{2}(\ell+\delta)\|t^*\|_2^2\right\} \\ =& \frac{\|t\|_2^2}{(\ell+\delta)} \end{aligned} \tag{2.42}$$

where we used in the fourth line that $Lt^* = 0$ and in the fifth line that, by assumption, $\nabla^2 V(x) \leq (\ell + \delta)\mathbb{1}$ uniformly in the ball B_ρ.

To conclude the proof of (2.32) we are left to estimate the the second term, [2], in the right hand side of (2.37). Notice that the difference between Helffer and Sjöstrand formulation of the covariance and (2.37) lies in the presence of this term only. Inserting our choice of f in [2] we get, by and application of the Schwartz inequality together with the Gauss-Green-Ostrogradskii-Stokes formula on exterior derivatives [A],

$$[2] \leq \frac{2\|t\|_2^2 \rho}{\ell - \delta} \frac{\int_{S_\rho} e^{-NV(x)} d^{M-1}(x)}{\int_{B_\rho} e^{-NV(x)} d^M x} \tag{2.43}$$

where S_ρ denotes the sphere in $I\!R^M$ of radius ρ and centered at zero. Remembering the assumption (2.28) on ρ and making use once again of the upper and lower bound (2.30) on $\nabla^2 V(x)$, classical gaussian type estimates yield:

$$\frac{\int_{S_\rho} e^{-NV(x)} d^{M-1}(x)}{\int_{B_\rho} e^{-NV(x)} d^M x} \leq \exp\left\{-M\left[\tfrac{\ell-\delta}{\ell+\delta}K - 1 - \ln K\right]\right\} \leq e^{-M} \tag{2.44}$$

To prove (2.33) let us set $f(x) \equiv \sqrt{N}(t, x)$ and $V_s(x) \equiv V(x) + s\frac{f(x)}{N}$ for $s \in [0, 1]$. Then note that on one hand,

$$\ln I\!E_V e^{(f - I\!E_V f)} = \int_0^1 ds \int_0^s ds' I\!E_{V_{s'}}(f - I\!E_{V_{s'}} f)^2 \tag{2.45}$$

while on the other, $\nabla^2 V_s(x) = \nabla^2 V(x)$. Therefore applying (2.32) to the summand in the r.h.s. of (2.45) immediately yields (2.33). Thus Lemma 2.4 is proven.$\diamondsuit$

Proof of Proposition 2.2: Proposition 2.2 this is an immediate consequence of Lemma 2.3 and Lemma 2.4 since the condition (1.12) on ρ always allows us to chose $\rho \equiv \bar\rho(N)$ with $\bar\rho(N)$ a decreasing function of N that goes to zero as N diverges. This concludes the proof of Proposition 2.2. $\diamondsuit$

Assuming Proposition 2.1, this concludes the proof of Theorem 1.1., since the difference between the logarithms of all Laplace transforms for different ρ goes to zero by (ii), the difference arising from the different centering between $\widetilde{\mathcal{L}}$ and $\widehat{\mathcal{L}}$ goes to zero by (iii), and the original Laplace transforms $\mathcal{L}$ are related to $\widehat{\mathcal{L}}$ by (i). $\diamondsuit\diamondsuit$

3. Proof of Proposition 2.1

We conclude the proof of Theorem 1.1 by proving Proposition 2.1. It is largely based on results from [BG2] which we collect in Theorem 3.1 and Lemma 3.2 below. For any $\rho > 0$ and $x \in I\!R^M$

we define the ball $B_\rho(x) \equiv \{y \in I\!R^M \,|\, \|x-y\|_2 \leq \rho\}$, and we denote by $B^c_\rho(x)$ it's complement in $I\!R^M$. We recall from (1.7) that $B_\rho^{(\mu,s)} \equiv B_\rho(m^{(\mu,s)})$ where the points $m^{(\mu,s)}$ are defined in (1.6) and we denote by $\mathcal{R}_\rho$ the complement of the union of these balls:

$$\mathcal{R}_\rho \equiv \left(\bigcup_{\mu\in\{1,\dots,M(N)\}, s\in\{-1,1\}} B_\rho^{(\mu,s)} \right)^c \tag{3.1}$$

Note that the balls in the previous union are disjoint provided that $\rho < m^*/\sqrt{2}$.

Theorem 3.1: **([BG2], Theorem 1).** *There exists $\gamma_a > 0$ and finite positive constants $c_0 \geq 1/2$, $c_1 > 0$, such that for all $\beta > 1$, for $\sqrt{\alpha} < \gamma_a (m^*)^2$, if $\rho \geq c_0 \frac{\sqrt{\alpha}}{m^*}$ then, with probability one, for all but a finite number of indices N, for all $z \in \mathcal{R}_\rho$,*

$$\Phi_{\beta,N}(z) - \Phi_{\beta,N}(m^{(1,1)}) \geq c_1 (m^*)^2 \inf_{\mu\in\{1,\dots,M(N)\}, s\in\{-1,1\}} \|z - s m^*(\beta) e^\mu\|_2^2 \tag{3.2}$$

With the notation of Theorem 3.1 we have:

Lemma 3.2: *For all $\beta > 1$ and $\sqrt{\alpha} < \gamma_a (m^*)^2$, if $c_0 \frac{\sqrt{\alpha}}{m^*} < \rho < m^*/\sqrt{2}$ then, with probability one, for all but a finite number of indices N, for all $\mu \in \{1,\dots,M(N)\}$, $s \in \{-1,1\}$,*

i)

$$\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right) \geq e^{-\frac{1}{2}\beta M} \tag{3.3}$$

ii) there exists a constant $c_2 > 0$ such that

$$\frac{\widetilde{\mathcal{Q}}_{\beta,N}\left(\mathcal{R}_\rho\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)} \leq e^{-c_2 \beta M} \tag{3.4}$$

iii) for all $b > 0$ such that $\rho + b < \sqrt{2} m^*$,

$$1 \leq \frac{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho+b}^{(\mu,s)}\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)} \leq 1 + e^{-c_2 \beta M} \tag{3.5}$$

where c_2 is the constant appearing in (3.4).

Proof: (3.3) and (3.4) were proved in [BG2]. As it will be useful to us later on, let us mention that in the course of the proof of (3.3), we established in particular that:

$$\frac{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right) Z_{\beta,N}}{e^{-\beta N \Phi_{\beta,N}(m^{(1,1)})}} = \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M z e^{-\beta N \{\Phi_{\beta,N}(z) - \Phi_{\beta,N}(m^{(1,1)})\}} 1\!\!1_{\left\{z \in B_\rho^{(\mu,s)}\right\}} \geq e^{-c_3 \beta M} \tag{3.6}$$

for some constant $c_3 > 0$. The lower bound of (3.5) is immediate while it's upper bound is a direct consequence of (3.4). To obtain it we write $B_{\rho+b}^{(\mu,s)} = B_\rho^{(\mu,s)} \cup B_{\rho+b}^{(\mu,s)} \setminus B_\rho^{(\mu,s)}$. But if $\rho + b < m^*/\sqrt{2}$ then $B_{\rho+b}^{(\mu,s)} \setminus B_\rho^{(\mu,s)} \subseteq \mathcal{R}_\rho$ so that

$$\frac{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho+b}^{(\mu,s)}\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)} \leq 1 + \frac{\widetilde{\mathcal{Q}}_{\beta,N}\left(\mathcal{R}_\rho\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)} \leq 1 + e^{-c_2 \beta M} \tag{3.7}$$

$\diamondsuit$

A main tool to compare the measures $\widetilde{\mathcal{Q}}_{\beta,N}$ and $\mathcal{Q}_{\beta,N}$ will be to use the strong concentration properties of $\mathcal{N}_{\beta N}^M$. This is the content of the next

Lemma 3.3: *For all $\delta > 0$ set $a \equiv \delta + \sqrt{\frac{M}{\beta N}}$. Then, for all $\rho > a$,*

$$\mathcal{N}_{\beta N}^M\left(B_{\rho-a}(y-x)\right) - e^{-\frac{1}{2}\beta N \delta^2} \leq \mathbb{1}_{\{x \in B_\rho(y)\}} \leq \mathcal{N}_{\beta N}^M\left(B_{\rho+a}(y-x)\right) + 2e^{-\frac{1}{2}\beta N \delta^2} \tag{3.8}$$

Before proving Lemma 3.3 let us show how it enables to relate the measures of balls:

Lemma 3.4: *For all $\delta > 0$ set $a \equiv \delta + \sqrt{\frac{M}{\beta N}}$. Then, for all $\rho > a$,*

$$\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho-a}(m^{(\mu,s)})\right) - e^{-\frac{1}{2}\beta N \delta^2} \leq \mathcal{Q}_{\beta,N}\left(B_\rho(m^{(\mu,s)})\right) \leq \widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho+a}(m^{(\mu,s)})\right) + 2e^{-\frac{1}{2}\beta N \delta^2} \tag{3.9}$$

Proof: By definition of $\widetilde{\mathcal{Q}}$ we have:

$$\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right) = \int \mathcal{N}_{\beta N}^M\left(B_\rho(m^{(\mu,s)} - x)\right) d\mathcal{Q}_{\beta,N}(x) \tag{3.10}$$

Lemma 3.4 is an immediate consequence of the above identity and the estimates (3.8) of Lemma 3.3.$\diamondsuit$

Proof of Lemma 3.3: The basic ingredient of the proof is the following gaussian isoperimetric type inequality:

Lemma 3.5: *for all $\delta > 0$ set $a \equiv \delta + \sqrt{\frac{M}{\beta N}}$. Then,*

$$\mathcal{N}_{\beta N}^M\left(B_a^c(0)\right) \leq e^{-\frac{1}{2}\beta N \delta^2} \tag{3.11}$$

Proof: It is a simple consequence of the following well-known "gaussian" concentration inequality (see e.g. [LT]): denoting by $\mathbb{E}_\mathcal{N}$ the expectation with respect to $\mathcal{N}_{\beta N}^M$ we have, for all $\delta > 0$:

$$\mathcal{N}_{\beta N}^M\left(\{\|z\|_2 > \mathbb{E}_\mathcal{N}\|z\|_2 + \delta\}\right) \leq e^{-\beta N \frac{\delta^2}{2}} \tag{3.12}$$

As direct computation yields $\mathbb{E}_{\mathcal{N}}\|z\|_2^2 = \frac{M}{\beta N}$ we have, by the Schwartz inequality, $\mathbb{E}_{\mathcal{N}}\|z\|_2 \leq \sqrt{\frac{M}{\beta N}}$, which together with (3.12) entails (3.11).◇

Now to prove the lower bound in (3.8) let us consider the quantity $\mathcal{N}_{\beta N}^{M}\left(B_{\rho-a}(y-x)\right)$ and rewrite it as

$$\mathcal{N}_{\beta N}^{M}\left(B_{\rho-a}(y-x)\right) = I_{\rho-a} + J_{\rho-a} \tag{3.13}$$

where

$$\begin{aligned} I_{\rho-a} &\equiv \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M z e^{-\frac{1}{2}\beta N\|z\|_2^2} \mathbb{1}_{\{z+x\in B_{\rho-a}(y)\}} \mathbb{1}_{\{z\in B_a(0)\}} \\ J_{\rho-a} &\equiv \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M z e^{-\frac{1}{2}\beta N\|z\|_2^2} \mathbb{1}_{\{z+x\in B_{\rho-a}(y)\}} \mathbb{1}_{\{z\in B_a^c(0)\}} \end{aligned} \tag{3.14}$$

Then note that on the one hand,

$$J_{\rho-a} \leq \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M z e^{-\frac{1}{2}\beta N\|z\|_2^2} \mathbb{1}_{\{z\in B_a^c(0)\}} \leq e^{-\frac{1}{2}\beta N\delta^2} \tag{3.15}$$

where the last inequality is nothing but Lemma 2.3. On the other hand, using that $\mathbb{1}_{\{z+x\in B_{\rho-a}(y)\}}\mathbb{1}_{\{z\in B_a(0)\}} \leq \mathbb{1}_{\{x\in B_\rho(y)\}}\mathbb{1}_{\{z\in B_a(0)\}}$,

$$I_{\rho-a} \leq \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M z e^{-\frac{1}{2}\beta N\|z\|_2^2} \mathbb{1}_{\{x\in B_\rho(y)\}} \mathbb{1}_{\{z\in B_a(0)\}} \leq \mathbb{1}_{\{x\in B_\rho(y)\}} \tag{3.16}$$

and inserting (3.15) and (3.16) in (3.13) gives the lower bound of (3.8). To prove the upper bound, we consider the quantity $\mathcal{N}_{\beta N}^{M}\left(B_{\rho+a}(y-x)\right)$ and rewrite it as

$$\mathcal{N}_{\beta N}^{M}\left(B_{\rho+a}(y-x)\right) = I_{\rho+a} + J_{\rho+a} \tag{3.17}$$

Trivially, $J_{\rho+a} \geq 0$ while, using that $\mathbb{1}_{\{z+x\in B_{\rho+a}(y)\}}\mathbb{1}_{\{z\in B_a(0)\}} \leq \mathbb{1}_{\{x\in B_\rho(y)\}}\mathbb{1}_{\{z\in B_a(0)\}}$,

$$\begin{aligned} I_{\rho+a} &\geq \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M z e^{-\frac{1}{2}\beta N\|z\|_2^2} \mathbb{1}_{\{x\in B_\rho(y)\}} \mathbb{1}_{\{x\in B_\rho(y)\}} \mathbb{1}_{\{z\in B_a(0)\}} \\ &\geq \mathbb{1}_{\{x\in B_\rho(y)\}} \left[1 - 2\left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M z e^{-\frac{1}{2}\beta N\|z\|_2^2} \mathbb{1}_{\{z\in B_a^c(0)\}}\right] \\ &\geq \mathbb{1}_{\{x\in B_\rho(y)\}} \left[1 - 2e^{-\frac{1}{2}\beta N\delta^2}\right] \end{aligned} \tag{3.18}$$

where the last inequality again follows from Lemma 2.3. (3.18) together with (3.17) gives the bound of (3.8). This concludes the proof of Lemma 3.5.◇

We are now ready to prove Proposition 3.1.

Proof of proposition 3.1, part i): Assume that ρ and a are chosen in such a way that both satisfy the assumptions of Lemma 3.3 and Lemma 3.4. We will first prove an upper bound on $\widehat{\mathcal{L}}_\rho^{(\mu,s)}$

in terms of $\mathcal{L}_\rho^{(\mu,s)}$. Remembering the definition (2.5) of $\widehat{\mathcal{L}}_\rho^{(\mu,s)}$ we have:

$$
\begin{aligned}
&\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)\widehat{\mathcal{L}}_{\beta,N,\rho}^{(\mu,s)}(t)\\
=&\frac{1}{Z_{\beta,N}}\left(\frac{\beta N}{2\pi}\right)^{M/2}\int d^M z e^{-\beta N\Phi_{\beta,N}(z)+\sqrt{N}(t,z-\overline{X}^{(\mu,s)})}\mathbb{1}_{\left\{z\in B_\rho^{(\mu,s)}\right\}}\\
=&\frac{1}{Z_{\beta,N}}\left(\frac{\beta N}{2\pi}\right)^{M/2}\int d^M z e^{-\frac{1}{2}\beta N\|z\|_2^2+\sqrt{N}(t,z-\overline{X}^{(\mu,s)})}\mathbb{E}_\sigma e^{\beta N(m_N(\sigma),z)}\mathbb{1}_{\left\{z\in B_\rho^{(\mu,s)}\right\}}\\
=&\frac{e^{-\sqrt{N}(t,\overline{X}^{(\mu,s)})}}{Z_{\beta,N}}\mathbb{E}_\sigma e^{\frac{1}{2}\beta N\|m_N(\sigma)+\frac{t}{\beta\sqrt{N}}\|_2^2}\left(\frac{\beta N}{2\pi}\right)^{M/2}\int d^M z e^{-\frac{1}{2}\beta N\|z-(m_N(\sigma)+\frac{t}{\beta\sqrt{N}})\|_2^2}\mathbb{1}_{\left\{z\in B_\rho^{(\mu,s)}\right\}}\\
=&\frac{e^{-\sqrt{N}(t,\overline{X}^{(\mu,s)})}}{Z_{\beta,N}}\mathbb{E}_\sigma e^{\frac{1}{2}\beta N\|m_N(\sigma)+\frac{t}{\beta\sqrt{N}}\|_2^2}\mathcal{N}_{\beta N}^M\left(B_\rho(m^{(\mu,s)}-(m_N(\sigma)+\tfrac{t}{\beta\sqrt{N}}))\right)
\end{aligned}
\tag{3.19}
$$

Therefore using the lower bound of (3.8),

$$
\widehat{\mathcal{L}}_{\beta,N,\rho}^{(\mu,s)}(t)\le T_1+T_2 \tag{3.20}
$$

where

$$
\begin{aligned}
T_1&\equiv\frac{e^{-\frac{1}{2}\beta N\delta^2+\frac{1}{2\beta}\|t\|_2^2}}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)Z_{\beta,N}}\mathbb{E}_\sigma e^{\frac{1}{2}\beta N\|m_N(\sigma)\|_2^2+\sqrt{N}(t,m_N(\sigma)-\overline{X}^{(\mu,s)})}\\
T_2&\equiv\frac{e^{\frac{1}{2\beta}\|t\|_2^2}}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)Z_{\beta,N}}\mathbb{E}_\sigma e^{\frac{1}{2}\beta N\|m_N(\sigma)\|_2^2+\sqrt{N}(t,m_N(\sigma)-\overline{X}^{(\mu,s)})}\mathbb{1}_{\left\{m_N(\sigma)+t/\sqrt{N}\in B_{\rho+a}^{(\mu,s)}\right\}}
\end{aligned}
\tag{3.21}
$$

To bound T_1 we will first make use of (1.14) from Lemma 1.3 to write that $\|\overline{X}^{(\mu,s)}\|_2\le\tilde{\rho}+m^*$ and thus

$$
T_1\le\frac{e^{-\frac{1}{2}\beta N\delta^2+\frac{1}{2\beta}\|t\|_2^2}}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)Z_{\beta,N}}e^{\sqrt{N}\|t\|_2(\tilde{\rho}+m^*)}\mathbb{E}_\sigma e^{\frac{1}{2}\beta N\|m_N(\sigma)\|_2^2+\sqrt{N}(t,m_N(\sigma))} \tag{3.22}
$$

On the other hand, it is immediate to see that for N large enough, on a subset of Ω of probability going to one exponentially fast, $\|m_N(\sigma)\|_2\le 2$ (c.f. e.g. [BG2]). Therefore, on that subset,

$$
\begin{aligned}
T_1\le&\frac{e^{-\frac{1}{2}\beta N\delta^2+\frac{1}{2\beta}\|t\|_2^2}}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)}e^{\sqrt{N}\|t\|_2(\tilde{\rho}+m^*+2)}\\
\le&e^{-\frac{1}{2}\beta N\delta^2+\frac{1}{2\beta}\|t\|_2^2+(\tilde{\rho}+m^*+2)\sqrt{N}\|t\|_2+\frac{1}{2}\beta M}\\
\le&e^{-\frac{1}{2}\beta N\delta^2+4\sqrt{N}\|t\|_2+\frac{1}{2}\beta M}
\end{aligned}
\tag{3.23}
$$

where the second inequality follows from (3.3) and where we used in the third one that $\tilde{\rho}\downarrow 0$ as $N\uparrow\infty$ while $m^*\le 1$ and $\|t\|_2$ is finite. Let us now turn to the term T_2. Just note that $\left\{\sigma\mid m_N(\sigma)+t/\sqrt{N}\in B_{\rho+a}^{(\mu,s)}\right\}\subset\left\{\sigma\mid m_N(\sigma)\in B_{\rho+a+\|t\|_2/\sqrt{N}}^{(\mu,s)}\right\}$ so that we immediately have

$$
T_2\le e^{\frac{1}{2\beta}\|t\|_2^2}\mathcal{L}_{\beta,N,\rho}^{(\mu,s)}(t)\frac{\mathcal{Q}_{\beta,N}\left(B_{\rho+a+\|t\|_2/\sqrt{N}}^{(\mu,s)}\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)} \tag{3.24}
$$

and all we need to do is to show that the last ratio is close to one. Treating it's numerator with the help of the upper bound of lemma 3.5 we have, if $\rho + 2a + \|t\|_2/\sqrt{N} < m^*/\sqrt{2}$,

$$\frac{\mathcal{Q}_{\beta,N}\left(B^{(\mu,s)}_{\rho+a+\|t\|_2/\sqrt{N}}\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B^{(\mu,s)}_{\rho}\right)} \leq \frac{\widetilde{\mathcal{Q}}_{\beta,N}\left(B^{(\mu,s)}_{\rho+2a+\|t\|_2/\sqrt{N}}\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B^{(\mu,s)}_{\rho}\right)} + \frac{2e^{-\frac{1}{2}\beta N\delta^2}}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B^{(\mu,s)}_{\rho}\right)} \leq 1 + e^{-c_2\beta M} + 2e^{-\frac{1}{2}\beta N\delta^2 + \frac{1}{2}\beta M} \tag{3.25}$$

where we have used the upper bound from Lemma 3.3, iii), to bound the first term in the right hand side of the first line and the estimate (3.3) from Lemma 3.3, i), to bound the second term. Finally inserting (3.24) in (3.25) yields

$$T_2 \leq e^{\frac{1}{2\beta}\|t\|_2^2}\mathcal{L}^{(\mu,s)}_{\beta,N,\rho}(t)\left(1 + e^{-c_2\beta M} + 2e^{-\frac{1}{2}\beta N\delta^2+\frac{1}{2}\beta M}\right) \tag{3.26}$$

and (3.23), (3.26) together with (3.20) give

$$e^{-\frac{1}{2\beta}\|t\|_2^2}\widehat{\mathcal{L}}^{(\mu,s)}_{\beta,N,\rho}(t) \leq e^{-\frac{1}{2}\beta N\delta^2+4\sqrt{N}\|t\|_2+\frac{1}{2}\beta M} + \mathcal{L}^{(\mu,s)}_{\beta,N,\rho}(t)\left(1 + e^{-c_2\beta M} + 2e^{-\frac{1}{2}\beta N\delta^2+\frac{1}{2}\beta M}\right) \tag{3.27}$$

From this we see that δ must be chosen in such a way that the first term in (3.27) vanishes while at the same time the last factor goes to one as N goes to infinity. Clearly the only constraint lies in making the first term small. Distinguishing the two cases $M < \sqrt{N}$ and $M > N$ we set

$$\delta^2 = \begin{cases} (\beta + 8\|t\|_2 + 2c_4)\frac{1}{\beta\sqrt{N}} & \text{if } M < \sqrt{N} \\ (\beta + 8\|t\|_2 + 2c_4)\frac{M}{\beta N} & \text{if } M \geq \sqrt{N} \end{cases} \tag{3.28}$$

for some constant $c_4 > 0$. With this choice we have,

$$e^{-\frac{1}{2}\beta N\delta^2+4\sqrt{N}\|t\|_2+\frac{1}{2}\beta M} \leq e^{-c_4\{\sqrt{N}\wedge M\}} \leq e^{-c_4 M} \tag{3.29}$$

and,

$$e^{-c_2\beta M} + 2e^{-\frac{1}{2}\beta N\delta^2+\frac{1}{2}\beta M} \leq 3e^{-c_5 M} \tag{3.30}$$

for some new constant $c_5 > 0$. Collecting (3.27), (3.28) and (3.30) finally yields,

$$e^{-\frac{1}{2\beta}\|t\|_2^2}\widehat{\mathcal{L}}^{(\mu,s)}_{\beta,N,\rho}(t) \leq e^{-c_4 M} + \mathcal{L}^{(\mu,s)}_{\beta,N,\rho}(t)\left(1 + e^{-c_5 M}\right) \tag{3.31}$$

The proof of the corresponding upper bound on $\widehat{\mathcal{L}}$ in terms of $\mathcal{L}$ follows the same pattern. Starting from (3.19) and using this time the upper bound of (3.8) we get

$$\widehat{\mathcal{L}}^{(\mu,s)}_{\beta,N,\rho}(t) \geq T_1' + T_2' \tag{3.32}$$

where $T_1' = 2T_1 \geq 0$ and T_2' is defined exactly as T_2 but with the characteristic function of the event $\left\{m_N(\sigma) + t/\sqrt{N} \in B_{\rho+a}^{(\mu,s)}\right\}$ replaced by that of the event $\left\{m_N(\sigma) + t/\sqrt{N} \in B_{\rho-a}^{(\mu,s)}\right\}$. Thus, assuming that $\rho - (2a + \|t\|_2/\sqrt{N}) > 0$

$$T_2' \geq e^{\frac{1}{2\beta}\|t\|_2^2} \mathcal{L}_{\beta,N,\rho}^{(\mu,s)}(t) \frac{\mathcal{Q}_{\beta,N}\left(B_{\rho-a-\|t\|_2/\sqrt{N}}^{(\mu,s)}\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho}^{(\mu,s)}\right)} \tag{3.33}$$

and proceeding as in (3.25), substituting the upper bounds by the appropriate lower ones,

$$\begin{aligned} \frac{\mathcal{Q}_{\beta,N}\left(B_{\rho-a-\|t\|_2/\sqrt{N}}^{(\mu,s)}\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho}^{(\mu,s)}\right)} &\geq \frac{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho-2a-\|t\|_2/\sqrt{N}}^{(\mu,s)}\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho}^{(\mu,s)}\right)} - \frac{e^{-\frac{1}{2}\beta N\delta^2}}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho}^{(\mu,s)}\right)} \\ &\geq 1 - e^{-\frac{1}{2}\beta N\delta^2 + \frac{1}{2}\beta M} \\ &\geq 1 - e^{-c_4 M} \end{aligned} \tag{3.34}$$

where the last line follows from the choice of δ made in (3.28). Therefore, inserting (3.34) in (3.33),

$$T_2' \geq e^{\frac{1}{2\beta}\|t\|_2^2} \mathcal{L}_{\beta,N,\rho}^{(\mu,s)}(t)\left(1 - e^{-c_4 M}\right) \tag{3.35}$$

From this and the fact that $T_1' \geq 0$, (3.32) yields

$$e^{-\frac{1}{2\beta}\|t\|_2^2} \widehat{\mathcal{L}}_{\beta,N,\rho}^{(\mu,s)}(t) \geq \mathcal{L}_{\beta,N,\rho}^{(\mu,s)}(t)\left(1 - e^{-c_4 M}\right) \tag{3.36}$$

If thus ρ satisfies the various constraints appearing in the course of the proof, then (3.31) and (3.36) are the desired upper and lower bounds of (2.7). We are left to show that our choice of δ allows us to choose such a ρ. More precisely, this will be the case if we can choose ρ such that $\rho + 2a + \|t\|_2/\sqrt{N} < m^*/\sqrt{2}$, and $\rho - (2a + \|t\|_2/\sqrt{N}) > c_0 \frac{\sqrt{\alpha}}{m^*}$. Inserting our choice of δ in the definition of a we have

$$\begin{cases} a < [(\beta + 8\|t\|_2 + 2c_4) + 1]^{\frac{1}{2}} \frac{1}{\sqrt{\beta}} \frac{1}{N^{1/4}} & \text{if } M < \sqrt{N} \\ a = [(\beta + 8\|t\|_2 + 2c_4) + 1]^{\frac{1}{2}} \frac{1}{\sqrt{\beta}} \sqrt{\frac{M}{N}} & \text{if } M \geq \sqrt{N} \end{cases} \tag{3.37}$$

Since $\|t\|_2$ is finite, $\|t\|_2/\sqrt{N} \ll a$ for large enough N in both cases so that our conditions are fulfilled for ρ taken as in (2.7). This concludes the proof of part i) of Proposition 3.1. $\diamondsuit$

Proof of proposition 3.1, part ii): Let ρ and $\bar{\rho}$ satisfy the assumptions of the proposition and take δ as in (3.39). Remembering the definition (2.5) of $\widetilde{\mathcal{L}}_{\rho}^{(\mu,s)}$ (see also the first equality in (3.19)) we have

$$\widetilde{\mathcal{L}}_{\beta,N,\rho}^{(\mu,s)}(t) = \widetilde{T}_1 + \widetilde{T}_2 \tag{3.38}$$

where

$$
\begin{aligned}
\widetilde{T}_1 \equiv & \frac{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\bar{\rho}}^{(\mu,s)}\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho}^{(\mu,s)}\right)} \widetilde{\mathcal{L}}_{\beta,N,\bar{\rho}}^{(\mu,s)}(t) \\
\widetilde{T}_2 \equiv & \frac{1}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho}^{(\mu,s)}\right) Z_{\beta,N}} \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M z e^{-\beta N \Phi_{\beta,N}(z)+\sqrt{N}(t,z-\overline{X}^{(\mu,s)})} \mathbb{1}_{\left\{z \in B_{\rho}^{(\mu,s)} \setminus B_{\bar{\rho}}^{(\mu,s)}\right\}}
\end{aligned}
\tag{3.39}
$$

With our choice of ρ and δ and since $\bar{\rho} > c_0 \frac{\sqrt{\alpha}}{m^*}$, Lemma 3.3, iii), applies and yields

$$
1 \geq \frac{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\bar{\rho}}^{(\mu,s)}\right)}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho}^{(\mu,s)}\right)} \geq \frac{1}{1+e^{-c_2 \beta M}} \geq 1 - e^{-c_2 \beta M}
\tag{3.40}
$$

Thus

$$
\widetilde{\mathcal{L}}_{\beta,N,\bar{\rho}}^{(\mu,s)}(t) \geq \widetilde{T}_1 \geq \widetilde{\mathcal{L}}_{\beta,N,\bar{\rho}}^{(\mu,s)}(t)\left(1 - e^{-c_2 \beta M}\right)
\tag{3.41}
$$

It now remains to prove an upper bound for $\widetilde{T}_2$. Making use of Theorem 3.2 we have:

$$
\begin{aligned}
\widetilde{T}_2 \leq & \frac{e^{-\beta N \Phi_{\beta,N}(m^{(1,1)})}}{\widetilde{\mathcal{Q}}_{\beta,N}\left(B_{\rho}^{(\mu,s)}\right) Z_{\beta,N}} \\
& \times \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M z e^{-\beta N c_1 (m^*)^2 \|z-m^{(\mu,s)}\|_2^2 + \sqrt{N}(t,z-\overline{X}^{(\mu,s)})} \mathbb{1}_{\left\{z \in B_{\rho}^{(\mu,s)} \setminus B_{\bar{\rho}}^{(\mu,s)}\right\}} \\
\leq & e^{-c\beta M} e^{\sqrt{N}(t,m^{(\mu,s)}-\overline{X}^{(\mu,s)})} \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M u e^{-\beta N c_1 (m^*)^2 \|u\|_2^2 + \sqrt{N}(t,u)} \mathbb{1}_{\{\|u\|_2 \geq \bar{\rho}\}}
\end{aligned}
\tag{3.42}
$$

where we used (3.6) to bound the ratio appearing in the first line, and performed the change of variable $u = z - m^{(\mu,s)}$. Now by (1.14) $(t, m^{(\mu,s)} - \overline{X}^{(\mu,s)}) \leq \|t\|_2 \bar{\rho}$. Therefore, classical gaussian tails estimates yield

$$
\begin{aligned}
\widetilde{T}_2 \leq & e^{-c_3 \beta M} e^{\sqrt{N}\|t\|_2 \bar{\rho}} \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M u e^{-\beta N c_1 (m^*)^2 \|u\|_2^2 + \sqrt{N}(t,u)} \mathbb{1}_{\{\|u\|_2 \geq \bar{\rho}\}} \\
\leq & e^{-c_3 \beta M} e^{\sqrt{N}\|t\|_2 \bar{\rho}} e^{\frac{\|t\|_2^2}{\beta c_1 (m^*)^2}} \left(\frac{\beta N}{2\pi}\right)^{M/2} \int d^M u e^{-\beta N c_1 (m^*)^2 \|u\|_2^2} \mathbb{1}_{\left\{\|u\|_2 \geq \bar{\rho} - \frac{\|t\|_2}{\beta \sqrt{N} c_1 (m^*)^2}\right\}} \\
\leq & e^{-c_3 \beta M} e^{\sqrt{N}\|t\|_2 \bar{\rho}} e^{\frac{\|t\|_2^2}{\beta c_1 (m^*)^2}} e^{-\gamma \beta N c_1 (m^*)^2 \left(\bar{\rho} - \frac{\|t\|_2}{\beta \sqrt{N} c_1 (m^*)^2}\right)^2} \left(\frac{1}{c_1 (m^*)^2 (1-\gamma)}\right)^{M/2}
\end{aligned}
\tag{3.43}
$$

for any $0 < \gamma < 1$. As $N\tilde{\rho}^2 = \tilde{c}_0^2 M$ and since $\bar{\rho} > \tilde{\rho}$, we easily see that, for N and $c(\beta)$ large enough,

$$
0 \leq \widetilde{T}_2 \leq e^{-c_6 M}
\tag{3.44}
$$

for some constant $c_6 > 0$. Combining (3.44), (3.41) and (3.38) proves part ii) of Proposition 3.1.

Proof of Proposition 2.1, part (iii): : We proceed as in the proof of part (i). Note that

$$
\begin{aligned}
&\widetilde{\mathcal{Q}}_{\beta,N}\left(B_\rho^{(\mu,s)}\right)\left(\overline{Z}^{(\mu,s)},t\right)\\
=&\frac{1}{Z_{\beta,N}}\left(\frac{\beta N}{2\pi}\right)^{M/2}\int d^M z e^{-\beta N\Phi_{\beta,N}(z)}(z,t)\mathbb{1}_{\left\{z\in B_\rho^{(\mu,s)}\right\}}\\
=&\frac{1}{Z_{\beta,N}}\left(\frac{\beta N}{2\pi}\right)^{M/2}\int d^M z e^{-\beta N\Phi_{\beta,N}(z)}(m_N(\sigma),t)\mathbb{1}_{\left\{z\in B_\rho^{(\mu,s)}\right\}}\\
=&\frac{1}{Z_{\beta,N}}\left(\frac{\beta N}{2\pi}\right)^{M/2}\int d^M z e^{-\frac{1}{2}\beta N\|z\|_2^2}\mathbb{E}_\sigma e^{\beta N(m_N(\sigma),z)}(z-m_N(\sigma)+m_N(\sigma),t)\mathbb{1}_{\left\{z\in B_\rho^{(\mu,s)}\right\}}\\
=&\frac{1}{Z_{\beta,N}}\left(\frac{\beta N}{2\pi}\right)^{M/2}\int d^M z e^{-\frac{1}{2}\beta N\|z\|_2^2}\mathbb{E}_\sigma e^{\beta N(m_N(\sigma),z)}(m_N(\sigma),t)\mathbb{1}_{\left\{z\in B_\rho^{(\mu,s)}\right\}}\\
+&\frac{1}{Z_{\beta,N}}\left(\frac{\beta N}{2\pi}\right)^{M/2}\int d^M z e^{-\beta N\Phi_{\beta,N}(z)}\left(z-\frac{1}{N}\sum_{i=1}^N\xi_i\tanh(\beta(\xi_i,z)),t\right)\mathbb{1}_{\left\{z\in B_\rho^{(\mu,s)}\right\}}\\
&\equiv(I)+(II)
\end{aligned}
\tag{3.45}
$$

Term (I) is dealt with exactly as was done in the estimations for the Laplace transforms. We do not repeat the details. (II) is a boundary term: Namely,

$$
(II)=\frac{1}{\beta N Z_{\beta,N}}\left(\frac{\beta N}{2\pi}\right)^{M/2}\int_{B_\rho^{(\mu,s)}} d^M z\nabla\cdot\left(te^{-\beta N\Phi_{\beta,N}(z)}\right) \tag{3.46}
$$

and so just as in the estimate of the term [2] from (2.39), we get that

$$
|(II)|\leq\frac{\|t\|_2}{\beta N}\exp\left(-M\left[2\frac{\lambda(\beta)-\gamma(N)}{\lambda(\beta)+\gamma(N)}-1-\ln 2\right]\right)\leq\frac{\|t\|_2}{\beta N}e^{-M} \tag{3.47}
$$

Putting this together concludes the proof of the Proposition and hence Theorem 1.1.◊

References

[A] V.I. Arnold, "Mathematical methods of classical mechanics", Spinger, Berlin-Heidelberg-New York (1978).

[BG1] A. Bovier and V. Gayrard, "An almost sure large deviation principle for the Hopfield model", to appear in Ann. Probab **24** (1996).

[BG2] A. Bovier and V. Gayrard, "The retrieval phase of the Hopfield model, A rigorous analysis of the overlap distribution", to appear in Prob. Theor. Rel. Fields **107** (1996).

[BG3] A. Bovier and V. Gayrard, "Hopfield models as generalized random mean field models", WIAS-preprint 253, to appear in *Mathematics of spin glasses and neural networks*, A. Bovier and P. Picco, eds., Birkhäuser, Basel, Boston, Stuttgart (1997). Preprint WIAS-253 (1996).

[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).

[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).

[DGI] J.-D. Deuschel, G. Giacomin, and D. Ioffe, in preparation.

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

[Ge1] B. Gentz, "An almost sure central limit theorem for the overlap parameters in the Hopfield model", Stochastic Process. Appl. **62**, 243-262 (1996).

[Ge2] B. Gentz, "A central limit theorem for the overlap in the Hopfield model", to appear in Ann. Probab. (1996).

[Ge3] B. Gentz, "A central limit theorem for the overlap in the Hopfield model", Ph. D. Thesis, University of Zürich (1996).

[H] J. Hubbard, "Calculation of partition functions", Phys. Rev. Lett. **3**, 77-78 (1959).

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

[HS] B. Helffer and J. Sjöstrand, "On the correlation for Kac-like models in the convex case", J. Stat. Phys. **74**, 349-409 (1994).

[S] 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)].

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

[N] A. Naddaf, "On homogenization and scaling limit of some gradient perturbations of a massless free field", Ph.D.-Thesis, New York University, 1995.

[NS] A. Naddaf and T. Spencer, "On homogenization and scaling limit of some gradient perturbations of a massless free field", to appear in Commun. Math. Phys. (1996).