A Free Energy Bound for the Hopfield Model

Hans Koch *
Department of Mathematics, University of Texas at Austin
Austin, TX 78712

Abstract. We give a simple upper and lower bound on the free energy density of the Hopfield model of size N with p stored patterns, in the limit where N and p tend to infinity with $p/N \to \alpha < 1$. The two bounds coincide for $\alpha = 0$.

The Hamiltonian of the Hopfield model with p stored N–bit patterns is a random function on the set $\mathcal{S} = \{+1, -1\}^N$, with values given by the equation

$$H_{p,N,\xi}(S) = -\frac{1}{2N} \sum_{i,j=1}^{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu S_i S_j \,, \qquad S \in \mathcal{S} \,, \tag{1}$$

where ξ is a $p \times N$ matrix whose elements ξ_i^μ are independent random variables with values ± 1 and mean zero. The quantity considered here is the expected value of the free energy density

$$f_{p,N,\beta}(\xi) = -\frac{1}{\beta N} \ln\Big[\sum_{S \in \mathcal{S}} e^{-\beta H_{p,N,\xi}(S)}\Big] \tag{2}$$

at positive inverse temperatures β. In order to state our result, we define

$$\phi_\delta(\beta, h) = -\frac{1}{\beta} \ln\big[2\cosh(\beta h)\big] + \frac{1-\delta}{2} h^2 \,. \tag{3}$$

Theorem. *Let $0 \le \alpha < 1$ and $\delta < 1$ such that $\delta - 4\sqrt{\alpha}(1-\delta)$ is positive. Then*

$$\min_{h \in \mathbb{R}} \phi_\delta(\beta, h) + \frac{\alpha}{2\beta} \ln[\delta - 4\sqrt{\alpha}(1-\delta)] \le \lim_{\substack{N,p \to \infty \\ p/N \to \alpha}} \mathbb{E}\, f_{p,N,\beta}(\xi) \le \min_{h \in \mathbb{R}} \phi_0(\beta, h) \,. \tag{4}$$

* Supported in Part by the National Science Foundation under Grant No. DMS–9103590.

Remarks. The upper bound in (4) is well known: It is the free energy density of the Curie–Weiss model, which is obtained by dropping all terms with $\mu > 1$ from the Hamiltonian (1). For $\alpha = 0$ the lower bound in (4) coincides with the upper bound as $\delta \downarrow 0$. In this case we recover a recent result of Shcherbina and Tirozzi [2]. For $\beta < 1$, the lower bound with $\delta = 1-\beta$ agrees with the correct free energy density up to $\mathcal{O}(\alpha^{3/2})$; see e.g. the references [1] and [3].

Proof. Consider the overlap matrix A,

$$A_{\mu\nu} = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu - \delta_{\mu\nu}\,, \qquad \mu,\nu = 1,\ldots,p. \tag{5}$$

An estimate in [2] implies that there exists a real number λ and an even positive integer n, both depending on N, such that

$$\lambda^{-n}\mathbb{E}\,\mathrm{tr}A^n \leq o(N^{-1})\,, \qquad \lambda = 4\sqrt{\alpha} + o(1)\,, \tag{6}$$

as $N \to \infty$. A stronger version of this estimate (inequality (11)), which could be of independent interest, will be proved below.

Assume now that α and δ satisfy the hypothesis of the theorem, and that N is sufficiently large such that $\delta - \lambda(1-\delta) > 0$. Denote by $\langle .,. \rangle$ the standard inner product on $\mathbb{R}^p$, and denote by ξ_i the i–th column of the matrix ξ. If ξ is such that the operator norm of A is less than or equal to λ, we get an upper bound on the sum in (2) from the identity

$$\begin{aligned}
\sum_{S\in\mathcal{S}} e^{-\beta H_{p,N,\xi}(S)} &= \left(\frac{N\beta}{2\pi}\right)^{p/2} \int d^p m\, e^{-\frac{1}{2}N\beta\langle m,m\rangle} \sum_{S\in\mathcal{S}} e^{\beta\sum_{i=1}^{N}\langle m,\xi_i\rangle S_i} \\
&= \left(\frac{N\beta}{2\pi}\right)^{p/2} \int d^p m\, e^{-\frac{1}{2}N\beta\langle m,m\rangle} \prod_{i=1}^{N} 2\cosh\bigl(\beta\langle m,\xi_i\rangle\bigr) \\
&= \left(\frac{N\beta}{2\pi}\right)^{p/2} \int d^p m\, e^{-\frac{1}{2}N\beta\langle m,[\delta+(1-\delta)A]m\rangle - \beta\sum_{i=1}^{N}\phi_\delta(\beta,\langle m,\xi_i\rangle)}\,,
\end{aligned} \tag{7}$$

if we replace the matrix A by $-\lambda$ and the function $\phi_\delta(\beta,.)$ by its minimum value. The resulting lower bound on the free energy density can be extended to all patterns ξ, by adding to it the trivial bound $-[p/2 + const]$, multiplied by the positive factor $\lambda^{-n}\mathrm{tr}A^n$ which is larger than 1 whenever $||A|| > \lambda$:

$$f_{p,N}(\beta,\xi) \geq \min_{h\in\mathbb{R}} \phi_\delta(\beta,h) + \frac{p}{2\beta N}\ln[\delta - \lambda(1-\delta)] - [p/2 + const]\lambda^{-n}\mathrm{tr}A^n\,. \tag{8}$$

The assertion now follows from (8) and (6).

To complete the proof of our theorem we will now derive a bound that implies (6). Let $n \geq 4$. Then $N^n\mathbb{E}\,\mathrm{tr}A^n$ is equal to the number of ordered products

$$\Pi_0 \equiv \xi_{i_1}^{\mu_1}\xi_{i_1}^{\mu_2}\xi_{i_2}^{\mu_2}\xi_{i_2}^{\mu_3}\cdots\xi_{i_{k-1}}^{\mu_k}\bigl\{\xi_{i_k}^{\mu_k}\cdots\xi_{i_l}^{\mu_l}\bigr\}\xi_{i_l}^{\mu_{l+1}}\cdots\xi_{i_n}^{\mu_n}\xi_{i_n}^{\mu_1}\,, \tag{9}$$

(ignoring the braces) with $1 \leq \mu_k \leq p$ and $1 \leq i_k \leq N$ for all k, subject to the constraint that $\mu_1 \neq \mu_2 \neq \ldots \neq \mu_n \neq \mu_1$, and that each random variable ξ_i^μ appears an even number of times.

Such products will be referred to as *contributing* products. Note that the constraint implies that the cardinalities $u = |\{\mu_1, \ldots, \mu_n\}|$ and $v = |\{i_1, \ldots, i_n\}|$ satisfy $u+v \leq n+1$ and $2v \leq n$.

A contributing product will be called *simple* if its $2n$ factors can be grouped into n pairs of identical random variables in such a way that if each pair is marked by braces as shown in equation (9) for the pair $(\xi^{\mu_k}_{i_k}, \xi^{\mu_l}_{i_l})$, then the "sets" corresponding to different pairs are either nested or disjoint. Note that such pairings can be specified by giving only the n left braces. Thus, since each pairing determines whether or not any two indices can have different values, the number of simple products can be bounded by

$$c_n(p) \equiv \binom{2n}{n} \max_{\substack{u+v \leq n+1 \\ v \leq n/2}} p^u N^v \,. \tag{10}$$

Assume now that Π_0 is a non–simple contributing product. Below we will give a procedure for cutting Π_0 between some top–linked pairs (adjacent factors that have a common upper index, or the first and last factor) and regrouping the pieces into a simple product Π_m that has the following property P: *The first factor of* Π_m *is the same as that of* Π_0*; and if* ω *is any closed walk on the factors of* Π_m *such that every odd–numbered (even–numbered) step is between two factors that are top–linked in* Π_m (Π_0)*, then, with one possible exception, all pairs connected by an odd–numbered step of* ω *are pairs of identical factors.* In particular, if we only consider walks whose first step goes to the right, and for which all but the last (or all) odd–numbered steps are between identical pairs, then we can find a set W of walks of this type, such that every factor of Π_m is visited by exactly one walk in W, and only once. Thus, it is possible to encode the original product Π_0 in a modified version Π'_m of Π_m, where, along each walk $(\omega_0, \omega_1, \ldots, \omega_\ell = \omega_0) \in W$ of length $\ell \geq 4$, the upper index in the identical factors ω_{2k-2} and ω_{2k-1} has been replaced by a pointer to the factor ω_{2k}, for $1 \leq k < \ell/2$. Since the upper indices of Π'_m can take at most $p + 2n$ different "values", it follows that the number of contributing products is bounded by $c_n(p+2n)$, and hence

$$\mathbb{E}\, \mathrm{tr} A^n \leq N^{-n} c_n(p+2n) \leq 4^n (p+2n) \Big(\frac{p+2n}{N}\Big)^{n/2} \,, \qquad\qquad p+2n \leq N. \tag{11}$$

We get (6) by taking e.g. $\lambda = 4(1+n^{-1/2})\big(\frac{p+2n}{N}\big)^{1/2}$ with $n =$ twice the integer part of $\sqrt{N}$.

Finally, to construct Π_m from Π_0 we iterate the following step. Assume that Π_{m-1} is non–simple. Then, in this product, it is possible to find an odd number ≥ 3 of consecutive factors $\xi^\mu_{i'} \xi^\mu_i \cdots \xi^\mu_{j'}$ such that $\xi^\mu_{j'}$ is top–linked to a factor ξ^μ_j that is identical to ξ^μ_i, with $i \neq i'$ and $j \neq j'$. After choosing such a sub–product, we now reverse the order of the factors $\xi^\mu_i \cdots \xi^\mu_{j'}$ in Π_{m-1} and define Π_m to be the result of this operation. Note that this creates a new top–linked pair (ξ^μ_i, ξ^μ_j) of identical factors. Thus, the iteration ends (with a simple product) after less than n steps. It is easy to check that each step preserves the abovementioned property P, and that contributing products are mapped to contributing products.

References

[1] D.J. Amit, H. Gutfreund, and H. Sompolinsky, *Spin–Glass Models of Neural Networks*, Phys. Rev. A **32**:1007 (1985).

[2] M. Shcherbina and B. Tirozzi, *Exact Mean Field Theory for Some Hopfield Model*, Preprint (1991).

[3] X. Scacciatelli and B. Tirozzi, *Fluctuation of the Free Energy in the Hopfield Model*, Preprint (1991).