COMMENT

Comment on 'The capacity of the Hopfield model'

Anton Bovier†

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

Abstract.   In a recent paper 'The capacity of the Hopfield model', Feng and Tirozzi claim to prove rigorous results on the storage capacity that are in conflict with the predictions of the replica approach. We show that their results are in error and that their approach, even when the worst mistakes are corrected, does not give any mathematically rigorous results.

The paper [FT] by Feng and Tirozzi addresses the interesting question of the storage capacity of the Hopfield model. This value, namely the maximal ratio between the number of patterns to the number of neurons for which the Hopfield model works as a memory was first reported numerically by Hopfield to be about 0.14, and a value close to that was obtained analytically by Amit *et al* [AGS] with the use of the replica trick. Refined estimates using replica symmetry-breaking schemes were obtained later. However, the replica trick is mathematically not rigorous, and there have been many attempts to obtain such results in a mathematically rigorous way. The best results in this respect so far were rigorous lower bounds on α_c by Newman [N] which were recently improved by Loukianova [L1] and Talagrand [Ta]. These bounds are still by at least 50% off the expected value. Obtaining upper bounds on α_c has proven to be a much more difficult issue and the only result to our knowledge was obtained very recently by Loukianova [L2], who proved that for any $\alpha > 0$, the minimum of the Hamiltonian is at some finite distance away from the patterns, and that as α tends to infinity, this distance tends to at least 0.05. Although her proof is very nice and interesting, the numerical values are far from satisfactory.

The main result that Feng and Tirozzi claim to prove 'rigorously' is that, if a fraction δ of errors in the retrieval is allowed, then the critical $\alpha = \alpha(\delta)$ is given by $\alpha(\delta) = (1 - 2\delta)^2/(1 - \delta)^2$.

This result appears obviously false, since it gives $\alpha(\delta)$ close to one if δ is chosen close to 0, and $\alpha(\delta)$ close to zero, if δ is close to 1/2, contrary to what *has* to be the case. One might be first tempted to believe that this formula for $\alpha(\delta)$ is a misprint, but it is repeated consistently in the paper and moreover based on this formula, the authors argue that 'the replica trick approach to the capacity of the Hopfield model is only valid in the case $\alpha_N \to 0$ $(N \to \infty)$'.

Given the very strong and surprising claims made in this paper, it appears worthwhile to analyse their 'rigorous proof' in some detail in order to avoid misconceptions.

† E-mail address: bovier@wias-berlin.de

Feng and Tirozzi study the fixpoints of a deterministic gradient dynamic of the Hopfield model, i.e. solutions of the equation

$$\sigma_i = \text{sign}\left(\frac{1}{N}\sum_{\mu=1}^{p}\sum_{j\neq i}\xi_i^\mu\xi_j^\mu\sigma_j\right). \tag{1.1}$$

Since they are interested in solutions 'near' one, say the first, pattern, ξ^1, it is reasonable to index the configurations σ by the set $B\subset\{1,\ldots,N\}$ on which they differ from ξ^1, i.e. set†

$$\sigma_i^B \equiv \begin{cases} \xi_i^1 & \text{if } i\notin B \\ -\xi_i^1 & \text{if } i\in B. \end{cases} \tag{1.2}$$

The fixpoint equation (1.1) can then be written in the form

$$\sigma_i^B\left(\frac{1}{N}\sum_{\mu=1}^{p}\sum_{j\neq i}\xi_i^\mu\xi_j^\mu\sigma_j^B\right)\geqslant 0 \qquad \text{for all } 1\leqslant i\leqslant N. \tag{1.3}$$

After some elementary algebra, we can rewrite this as

$$\sigma_i^B\left(\frac{1}{N}\sum_{\mu=2}^{p}\sum_{j\neq i}\xi_i^\mu\xi_j^\mu\xi_j^1 - \frac{2}{N}\sum_{\mu=2}^{p}\sum_{j\in B, j\neq i}\xi_i^\mu\xi_j^\mu\xi_j^1 + \xi_i^1(1-|B|/N)\right) \tag{1.4}$$

where $|B|$ denotes the cardinality of the set B. Let us define the random variables‡

$$Z_i(N,B) \equiv \xi_i^1\left(\frac{1}{N}\sum_{\mu=2}^{p(N)}\sum_{j=1, j\neq i}\xi_i^\mu\xi_j^\mu\xi_j^1 - 2\frac{1}{N}\sum_{\mu=2}^{p(N)}\sum_{j\in B, j\neq i}\xi_i^\mu\xi_j^\mu\xi_j^1\right)\geqslant 0. \tag{1.5}$$

Then (1.4) can be written as

$$\begin{aligned} Z_i(N,B) &\geqslant -(1-2|B|/N) && \text{if } i\notin B \\ Z_i(N,B) &\leqslant -(1-2|B|/N) && \text{if } i\in B. \end{aligned} \tag{1.6}$$

The arguments of [FT] relies basically on their observation in equation (4) that for any fixed B, the random variables $Z_i(N,B)$ converge in distribution to $-\sqrt{\alpha}\zeta_i$§ (assuming that $p(N)/N$ tends to α), 'by the central limit theorem', where the ζ_i are i.i.d. (independent identically distributed) standard normal variables. The remainder of their analysis is then based on the study of the distribution of the kth maxima of i.i.d. Gaussian random variables.

This procedure involves an interchange of limits that is not justified. From the central limit theorem (CLT), one obtains the convergence of the variables in (1.1) in the sense that for a given B, *for any fixed, finite set I of indices*, the family $\{Z_i(N,B)\}_{i\in I}$ converges in distribution to a family of i.i.d. Gaussians. This does *not* imply that e.g. $\max_{i=1}^{N}Z_i(N,B)$ converges, for example, in distribution to the same limit as $\max_{i=1}^{N}\sqrt{\alpha}\zeta_i$! Maxima are not continuous functions with respect to the product topology, and therefore convergence in the distribution and the taking of maxima cannot be interchanged. (Take the following example: let $X_i(N)\equiv\zeta_i$, if $i<N$, and $X_N(N)=N$. This family converges to i.i.d. standard Gaussians, as N tends to infinity, but the maximum (which is always N) is totally different from the maximum of N i.i.d. Gaussians, which behaves like $\sqrt{\ln N}$!)

† I regret to have to introduce some notation that is different from that of Feng and Tirozzi.
‡ In [FT] these are defined in (4), and given the strange name $g(N, p(N))$, which makes reference neither to their dependence on the index i nor the set B. We need these attributes to make meaningful statements later.
§ The minus sign appearing here is rather unconventional, given that ζ_i and $-\zeta_i$ have the same distribution. However, it plays a role in the course of the mistakes they make later.

One should keep in mind that it is precisely this difficulty that has prevented reasonable upper bounds on α_c in the past. Loukianova, for instance uses a very clever idea of 'negative association' to compare the *dependent* variables $Z_i(N, B)$ to independent ones, but for this a price had to be paid that prevented sharp estimates.

While at this point it is clear that the arguments in [FT] are in no way 'rigorous', it may be still interesting to follow the sequel of their arguments in some more detail. Let us first consider the case of 'perfect retrieval', section 3.1. Here B is the empty set and their result relies on the assumption that the distribution of the maximum over i of the $Z_i(N, \emptyset)$ converges to that of i.i.d. Gaussians, which is not justified. There is no surprise in the fact that they obtain the same result as McEliece *et al* [MPRV], because the heuristic arguments of [MPRV] (which do not claim to give a rigorous proof!) are identical to those put forward here. I should stress that to my knowledge there is no rigorous proof of the 'if and only if' statement.

In section 3.2 [FT] study the case of non-perfect retrieval, i.e. they look for the critical $\alpha(\delta)$ such that a fixpoint σ^B will exist with $|B| = \delta N$. The way they seem to argue is as follows: $Z_i(N, B)/\sqrt{\alpha}$ converges to family of i.i.d. Gaussian random variables. Then, what is the fraction of N i.i.d. Gaussians that is larger than x? If this number is $g(x)$, then $g(-(1 - 2\delta)/\sqrt{\alpha})$ of the $Z_i(N, B)$ will be larger than $(1 - 2\delta)$, and so we will find a set B of size δN precisely when $(1 - \delta) = g(-(1 - 2\delta)/\sqrt{\alpha})$! At this point the authors claim (see the first phrase on page 3386, and figure 1) that the $[xN]$-largest of the N Gaussians ζ_i is of order x, so that $[xN]$ of them would be smaller than x, i.e. $[xN]$ of the $-\sqrt{\alpha}\zeta_i$ would be larger than $-\sqrt{\alpha}x$ which means that they take $g(x) = -x$. It does not become clear where they draw this claim from, but it is clear that it is totally wrong and leads to their absurd main result.

One may ask whether their arguments can be improved. First, how can we compute the size of the set of indices for which $Z_i(N, B) \geqslant x$? Obviously, one would want to study the quantity

$$G_N(x, B) \equiv \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{Z_i(N,B) \geqslant x} \tag{1.7}$$

which is nothing but (one minus) the distribution function of the empirical measure of the variables $Z_i(N, B)$. Now, if we replaced all the $Z_i(N, B)$ by $\sqrt{\alpha}\zeta_i$ (I take the freedom to drop their pointless minus sign), then by the strong law of large numbers

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\sqrt{\alpha}\zeta_i \geqslant x} = \frac{1}{\sqrt{2\pi}} \int_{x/\sqrt{\alpha}}^{\infty} e^{-\frac{1}{2}y^2 2} \, dy \equiv \Phi(-x/\sqrt{\alpha}) \qquad \text{a.s.} \tag{1.8}$$

This would then yield the somewhat more reasonable looking result $\alpha(\delta) = (1 - 2\delta)^2/(\Phi^{-1}(1 - \delta))^2$. But is that result to be trusted? First, the CLT again cannot justify the passage to the Gaussians, because just as with the maxima, the empirical measure is not a continuous function. If one is somewhat optimistic, one may hope to prove convergence of $G_N(x, B)$ to a Gaussian distribution function, *for fixed B*. (For example, Talagrand [Ta] proves this for the case $B = \emptyset$.) But even that would by no means imply that $Z_i(N, B)$ itself is becoming independent of B, as N increases, and that the set $B'(B)$ for which $Z_i(N, B) \geqslant -(1 - 2\delta)$ would coincide with B. One might want to argue that there should be a set B for which $B'(B) = B$, but then there is no reason why for this *random* set the convergence of the empirical measure should hold. To make such a statement, one would at least have to get control on the convergence of $G_N(x, B)$ *uniformly* in the different

possible sets B, i.e. we should need some estimate like

$$\mathbb{P}\left[\sup_{B \subset \{1,\ldots,N\}} \|G_N(x, B) - \Phi(x)\| > \epsilon\right] \downarrow 0 \qquad (1.9)$$

for all $\epsilon > 0$, for some norm $\|\cdot\|$. It does not seem likely that such a result is true, let alone that it can be proven. Note that the main difficulty here is that the number of sets B is exponentially large, and precisely this fact is responsible for the relatively poor lower bounds on α_c that exist in the literature.

   In conclusion, the paper by Feng and Tirozzi has unfortunately not contributed to progress in the mathematical understanding of this interesting and challenging problem. Even if the most obvious mistakes are corrected there remain fundamental problems in the basic approach, and even the improved prediction on $\alpha(\delta)$ is no more rigorous and rather less convincing than the predictions of the replica approach.

Acknowledgment

References

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

[FT]   Feng J and Tirozzi B 1997 Capacity of the Hopfield model *J. Phys. A: Math. Gen.* **30** 3383–91

[N]   Newman Ch M 1988 Memory capacity in neural network models: rigorous results *Neural Networks* **1** 223–38

[L1]   Loukianova D 1994 Capacité de mémoire dans le modèle de Hopfield *C. R. Acad. Sci., Paris* **318** Série 1 157–60

[L2]   Loukianova D 1997 Lower bounds on the restitution error in the Hopfield model *Prob. Theor. Rel. Fields* **107** 161–76

[MPRV]   McEliece R J, Posner E C, Rodemich E R and Venkatesh S S 1987 The capacity of the Hopfield associative memory *IEEE Trans. Inform. Theory* **33** 461–82

[Ta]   Talagrand M 1997 Rigorous results on the Hopfield model with many patterns *Prob. Theor. Rel. Fields.* to appear