IOP Publishing Journal of Physics A: Mathematical and Theoretical
J. Phys. A: Math. Theor. **53** (2020) 074001 (19pp) https://doi.org/10.1088/1751-8121/ab6943

Interpolating between boolean and extremely high noisy patterns through minimal dense associative memories

Francesco Alemanno[1,2,3], **Martino Centonze**[1,3,4] **and Alberto Fachechi**[1,3]

[1] Dipartimento di Matematica e Fisica Ennio De Giorgi, Università del Salento, Lecce, Italy
[2] CNR-Nanotec, Sezione di Lecce, Italy
[3] Istituto Nazionale di Fisica Nucleare, Sezione di Lecce, Italy

E-mail: martinosalomone@gmail.com

Received 17 July 2019, revised 28 December 2019
Accepted for publication 8 January 2020
Published 27 January 2020

Abstract

Recently, Hopfield and Krotov introduced the concept of *dense associative memories* [DAM] (close to spin-glasses with P-wise interactions in a disordered statistical mechanical jargon): they proved a number of remarkable features these networks share and suggested their use to (partially) explain the success of the new generation of Artificial intelligence. Thanks to a remarkable ante-litteram analysis by Baldi & Venkatesh, among these properties, it is known these networks can handle a maximal amount of stored patterns K scaling as $K \sim N^{P-1}$.

In this paper, once introduced a *minimal dense associative network* as one of the most elementary cost-functions falling in this class of DAM, we sacrifice this high-load regime -namely we force the storage of *solely* a linear amount of patterns, i.e. $K = \alpha N$ (with $\alpha > 0$)- to prove that, in this regime, these networks can correctly perform pattern recognition even if pattern signal is $O(1)$ and is embedded in a sea of noise $O(\sqrt{N})$, also in the large N limit. To prove this statement, by extremizing the quenched free-energy of the model over its natural order-parameters (the various magnetizations and overlaps), we derived its phase diagram, at the replica symmetric level of description and in the thermodynamic limit: as a sideline, we stress that, to achieve this task, aiming at cross-fertilization among disciplines, we pave two hegemon routes in the statistical mechanics of spin glasses, namely the replica trick and the interpolation technique.

Both the approaches reach the same conclusion: there is a not-empty region, in the noise-T versus load-α phase diagram plane, where these

[4] Author to whom any correspondence should be addressed.

networks can actually work in this challenging regime; in particular we obtained a quite high critical (linear) load in the (fast) noiseless case resulting in $\lim_{\beta\to\infty}\alpha_c(\beta) = 0.65$.

Keywords: artificial neural networks, statistical mechanics of spin glasses, mean field methods

(Some figures may appear in colour only in the online journal)

1. Introduction

Due to an increase in the GPU processing power [34, 44], availability of large data-sets for training stages and the (deep) multi-layer architectures where neural networks can finally be embedded [45, 46], their impressive skills -overall termed *deep learning* [38]- keep achieving successes in the most disparate fields of science and technology [17, 33, 37, 39, 40, 48–50] (particularly outperforming at work in biomedical imaging, where they -noawadays- detect patterns possibly earlier than humans [29]).

Despite a number of remarkable progresses (e.g. [2, 5, 9, 13, 18, 20, 21, 23, 27, 28, 31, 32, 41–43]), these computational successes yet lack a full theoretical bulk behind (e.g. as made available in the pairwise limit of shallow networks as Hopfield and Boltzmann machines [19]), hence the quest for a *rationale* where different know-how(s) possibly merge is nowadays mandatory in the agenda of several research groups, ranging from computer science to applied mathematics (possibly crossing theoretical physics at its proliferative intersection offered by statistical mechanics of spin glasses).

In these regards, recently, Hopfield and Krotov proposed as an underlying bridge between deep neural networks and dense associative memories [24, 35, 36] (the latter being P-spin extensions [4, 25] of the celebrated Hopfield classical pairwise limit [30]) proving how these higher-order cost functions are more robust against adversarial and rubbish inputs.

Furthermore, this class of neural networks was deeply analyzed by Venkatesh & Baldi and Bovier & Niederhauser in the past [8, 16] and it is known that -calling K the number of patterns to handle, N the amount of neurons to accomplish the task and P the order of their interactions- their critical capacity scales as $K \propto N^{P-1}$ (and collapses to the standard one, i.e. $K = 0.14N$, in the known pairwise limit of $P = 2$ [7]).

Recently some of the authors addressed the statistical mechanical analysis of a generalized RBM introduced in the literature by Terrence Sejnowski in 1984 [47] and proved that it was able to perform pattern recognition of patterns whose intensity stays $O(1)$ even in a sea of noise $O(\sqrt{N})$ in the large N limit [6]. It was also shown a dual representation of this network in terms of a peculiar form of the class of models suggested by Hopfield and Krotov [6]: as in the pairwise counterpart [1, 11], this duality played as a crucial step to explain this skill of these machines as they can be obtained by keeping the network's load away from the maximal regime (the Baldi & Venkatesh limit [8]). We stress that the inspection of the low-storage regimes for these networks already started in [10].

Here we continue along this investigation [6, 10], focusing on pattern recognition at extremely low signal-to-noise ratios, by proving that such a skill is not peculiar to the Sejnowski machine (see appendix C): still focusing on four-wise interactions among discrete neurons, it holds for a broader class of Hopfield and Krotov models (w.r.t. one used in [6]) that we call *minimal dense associative memory* (MDAM). In particular we provide a phase diagram for the MDAM, at the replica symmetric level of description and in the linear-storage

regime, to show that there is a huge region in the plane of the two tunable parameters -load α and noise β- where this phenomenon happens (here $\lim_{\beta\to\infty}\alpha_c(\beta) = 0.65$).

For the sake of cross-fertilization, we present our results paving at first the standard route of the replica trick [19], then confirming the picture obtained by the RS-ansatz by suitably adapting to the case a Guerra's interpolation scheme [3, 12].

2. Minimal dense associative memory

Here we introduce a minimal cost function of the form suggested by Hopfield and Krotov, namely the *minimal dense associative memory* (MDAM).

Definition 1.

$$\mathcal{H}(\boldsymbol{\sigma}|\boldsymbol{\eta}) = -\frac{1}{2N^3}\sum_{\mu=1}^{K}\Big(\sum_{i,j=1}^{N}\eta_{ij}^{\mu}\sigma_i\sigma_j\Big)^2 \tag{2.1}$$

where $\sigma_i = \pm 1$, $i \in (1,...,N)$, are Ising spin and η_{ij}^{μ} is the symmetric synaptic tensor.

Our goal is to prove how this model can retrieve patterns of information also when they are immersed in a background of a $\mathcal{O}(\sqrt{N})$ Gaussian noise. This result can be achieved by requiring the network to store only $\mathcal{O}(N)$ patterns instead of the theoretical upper limit of $\mathcal{O}(N^3)$. In order to do that, we introduce the following decomposition of the synaptic tensor

Definition 2. The load of the network α, as anticipated, is defined as

$$\alpha = \lim_{N\to\infty}\frac{K}{N}, \tag{2.2}$$

while, the signal+noise decomposition reads

$$\eta_{ij}^{\mu} = \frac{1}{\sqrt{1+\alpha}}(\xi_{ij}^{\mu} + \sqrt{K}J_{ij}^{\mu}), \tag{2.3}$$

where ξ_{ij}^{μ} is the tensor, with entries ± 1, constituting the 'matrix' signal, while the noise is embedded in the symmetric tensor J_{ij}^{μ}, whose entries are *i.i.d.* $\mathcal{N}(0,1)$ variables.

Remark 1. The noise is given by the product $\sqrt{K}J_{ij}^{\mu}$, which in the thermodynamic limit globally scales as $\mathcal{O}(\sqrt{N})$, as $K \sim N$.

We are interested in the study of the retrieval phase of the network; for simplicity we restrict ourselves to the study of the retrieval of pure states. The retrieved pattern is arbitrary: we just denote it by ξ_{ij}^{1}; the remaining states ξ_{ij}^{μ}, with $\mu > 1$, will then constitute a quenched noise for the system. Hence, we perform a quenched average over the $P-1$ remaining states ξ_{ij}^{μ} together with the amplified J_{ij}^{μ} noise, by introducing the following expectation operator

$$\mathbb{E} \equiv \Big(\prod_{i,j,\mu>1}^{N,N,K}\frac{1}{2}\sum_{\xi_{ij}^{\mu}=\pm 1}\Big)\Big(\prod_{i,j,\mu\geqslant 1}^{N,N,K}\int DJ_{ij}^{\mu}\Big). \tag{2.4}$$

As usual, all the thermodynamic properties can be derived from the quenched pressure[5]

[5] Notice that the pressure is strictly related to the quenched intensive free energy f as $A = -\beta f$.

Definition 3. In the thermodynamic limit, the quenched pressure reads

$$A_N = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\ln Z_N, \tag{2.5}$$

where Z_N is the partition function, defined as

$$Z_N = \sum_{\boldsymbol{\sigma}} \exp\left(-\beta\mathcal{H}\right) = \sum_{\boldsymbol{\sigma}} \exp\Big(\frac{\beta}{2N^3}\sum_{\mu=1}^{K}\big(\sum_{i,j=1}^{N}\eta_{ij}^{\mu}\sigma_i\sigma_j\big)^2\Big). \tag{2.6}$$

Remark 2. The partition function Z_N can be written by introducing auxiliary Gaussian variables as follows

$$Z_N = \int Dz \sum_{\boldsymbol{\sigma}} \exp\Big(\sqrt{\frac{\beta}{N^3}}\sum_{\mu=1}^{K}\sum_{i,j=1}^{N}\eta_{ij}^{\mu}\sigma_i\sigma_j z_\mu\Big), \tag{2.7}$$

where $\int Dz \equiv \int\prod_{\mu=1}^{K} Dz_\mu$ and Dz_μ the $\mathcal{N}(0,1)$ measure relative to the μ component of the vector z_μ. We stress that, written in this form, the partition function is equivalent to that of a bi-partite system, with the hidden layer $\boldsymbol{z}$ added to the visible one, $\boldsymbol{\sigma}$. The hidden layer is therefore filled with real valued gaussian $\mathcal{N}(0,1)$ neurons.

In the following two sub-sections, we will tackle the problem of finding an explicit expression for the above pressure in the thermodynamic limit in terms of the natural order parameters of the model, defined only after having introduced n replicas of the system (as usual in the context of replica trick and interpolation technique calculations).

Definition 4. The overlap q_{ab} among two replicas $(a, b = 1, .., n)$ of the system is defined as

$$q_{ab} = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^a\sigma_i^b. \tag{2.8}$$

Equivalently, the overlap relative to the hidden layer is defined as:

$$p_{ab} = \frac{1}{K-1}\sum_{\mu=2}^{K} z_\mu^a z_\mu^b. \tag{2.9}$$

The Mattis magnetization, for a generic pattern ξ_i^μ, and for a generic replica a of the system, reads:

$$m_\mu^a = \frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu\sigma_i^a. \tag{2.10}$$

We here introduce the matrix magnetization (also relative to the ath replica):

$$M_\mu^a = \frac{1}{N^2}\sum_{i,j=1}^{N}\xi_{ij}^\mu\sigma_i^a\sigma_j^a. \tag{2.11}$$

2.1. Route one: replica trick

The replica trick is based on the following identity:

$$A = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\ln Z_N = \lim_{N\to\infty}\lim_{n\to 0} \frac{\ln \mathbb{E}Z_N^n}{nN}. \tag{2.12}$$

Introducing the decomposition (2.3), the $\mathbb{E}Z_N^n$ partition function becomes

$$\mathbb{E}Z_N^n = \Big(\prod_{a=1}^{n}\sum_{\boldsymbol{\sigma}^a}\Big)\int\Big(\prod_{a=1}^{n}Dz^a\Big)\mathbb{E}\exp\Big(\sqrt{\frac{\beta}{(1+\alpha)N^3}}\sum_{a=1}^{n}\sum_{\mu=1}^{K}\sum_{i,j=1}^{N}\xi_{ij}^{\mu}\sigma_i\sigma_j z_\mu + \sqrt{\frac{\beta\alpha}{(1+\alpha)N^2}}\sum_{a=1}^{n}\sum_{\mu=1}^{K}\sum_{i,j=1}^{N}J_{ij}^{\mu}\sigma_i\sigma_j z_\mu\Big). \tag{2.13}$$

We now assume that only a single pattern (say ξ^1) is candidate for retrieval. Therefore, all patterns with $\mu \geqslant 2$ will contribute to the noise. We can therefore factorize the signal ($\mu = 1$) from the global noise ($\mu > 1$) in the partition function. Thus, the quenched average of the nth power of the partition function reads

$$\begin{aligned}\mathbb{E}Z^n =& \Big(\prod_{a=1}^{n}\sum_{\boldsymbol{\sigma}^a}\Big)\mathbb{E}\exp\Big[\frac{\beta}{2(1+\alpha)}\sum_{a=1}^{n}\Big(\sqrt{N}M_1^a + \frac{\sqrt{\alpha}}{N}\sum_{i,j=1}^{N}J_{ij}^1\sigma_i^a\sigma_j^a\Big)^2\Big]\times \\ &\times\int\Big(\prod_{a=1}^{n}D\{z^a\}_{\mu>1}\Big)\exp\Big(\sqrt{\frac{\beta}{(1+\alpha)N^3}}\sum_{a=1}^{n}\sum_{\mu>1}^{K}\sum_{i,j=1}^{N}\Big(\xi_{ij}^{\mu}+\sqrt{\alpha N}J_{ij}^{\mu}\Big)\sigma_i^a\sigma_j^a z_\mu^a\Big).\end{aligned} \tag{2.14}$$

In the first line, we can simply drop out the Gaussian contribution from the signal term, since

$$\frac{1}{N}\sum_{i,j}J_{ij}^1\sigma_i^a\sigma_j^a \sim \mathcal{O}(1), \tag{2.15}$$

for each $a = 1,\dots,n$[6]. Then, the we can split the nth power of the partition function as

$$\mathbb{E}Z^n = \Big(\prod_{a=1}^{n}\sum_{\boldsymbol{\sigma}^a}\Big)Z_{\text{signal}}Z_{\text{noise}}, \tag{2.17}$$

where

$$\begin{aligned}Z_{\text{signal}} &= \exp\Big[\frac{\beta N}{2(1+\alpha)}\sum_{a=1}^{n}(M_1^a)^2\Big], \\ Z_{\text{noise}} &= \int\Big(\prod_{a=1}^{n}D\{z^a\}_{\mu>1}\Big)\mathbb{E}\exp\Big(\sqrt{\frac{\beta}{(1+\alpha)N^3}}\sum_{a=1}^{n}\sum_{\mu>1}^{K}\sum_{i,j=1}^{N}\Big(\xi_{ij}^{\mu}+\sqrt{\alpha N}J_{ij}^{\mu}\Big)\sigma_i^a\sigma_j^a z_\mu^a\Big).\end{aligned} \tag{2.18}$$

First, we focus on the signal term. The matrix magnetization (2.11) measures the overlap of the product $\sigma_i\sigma_j$ in the direction specified by the matrix ξ_{ij}[7]. However, the network

[6] Recall that only terms that are linear extensive in N do contribute in the exponent, as lower order terms disappear in the thermodynamic limit. The $\sqrt{N}M_1^a$ term has the correct scaling

$$\frac{1}{N^{3/2}}\sum_{i,j}\xi_{ij}^1\sigma_i^a\sigma_j^a \sim \mathcal{O}(N^{1/2}), \tag{2.16}$$

which becomes $\mathcal{O}(N)$, given the presence of the square in equation (2.14) and for this reason it cannot be neglected: it represents the signal in the network.

[7] We omit the 'upper' index in ξ_{ij}^1, since it plays no role in what follows.

configuration is fixed by specifying the value of the N variables σ_i, while the ξ_{ij} has $\sim N^2$ degrees of freedom. This means that the network configuration could not retrieve a general tensor[8]. This issue is removed by working directly with factorized information patterns, i.e. in the form $\xi_{ij} \equiv \xi_i \xi_j$. This has an interesting consequence: the matrix magnetization factorizes in the square of the Mattis magnetization:

$$M^a = \Big(\frac{1}{N}\sum_{i=1}^{N} \xi_i \sigma_i^a\Big)^2 = (m_a)^2. \tag{2.19}$$

Hence, the signal term is simply[9]

$$Z_{\text{signal}} = \int \Big(\prod_a \mathrm{d}m_a \mathrm{d}\hat{m}_a\Big) \exp\Big(-\mathrm{i}N\sum_a m_a\hat{m}_a - \mathrm{i}\sum_a \hat{m}_a \sum_i \xi_i\sigma_i^a + N\frac{\beta}{2(1+\alpha)}\sum_a m_a^4\Big), \tag{2.20}$$

where $\hat{m}_a$ is the conjugated momentum of m_a, and naturally arises from the Fourier representation of the Dirac delta

$$1 = \int \prod_a \mathrm{d}m_a \delta(m_a - \frac{1}{N}\sum_i \xi_i\sigma_i^a). \tag{2.21}$$

Concerning the noise term, it can be evaluated as (see appendix A.1)

$$\begin{aligned} Z_{\text{noise}} = &\int \Big(\prod_{a,b} \mathrm{d}q_{ab}\mathrm{d}p_{ab}\mathrm{d}\hat{q}_{ab}\mathrm{d}\hat{p}_{ab}\Big) \exp\Big(-\frac{\alpha N}{2}\ln\det(\mathbb{I} + 2\mathrm{i}\hat{\mathbb{P}})\Big) \\ &\times \exp\Big(\mathrm{i}N\sum_{a,b} q_{ab}\hat{q}_{ab} - \mathrm{i}\sum_i\sum_{a,b}\hat{q}_{ab}\sigma_i^a\sigma_i^b + \mathrm{i}\alpha N\sum_{a,b} p_{ab}\hat{p}_{ab} + \frac{\beta\alpha^2}{2(1+\alpha)}\sum_{a,b=1}^{n} q_{ab}^2 p_{ab}\Big). \end{aligned} \tag{2.22}$$

Again, the parameters $\hat{q}_{ab}$ and $\hat{p}_{ab}$ are the conjugate momenta of q_{ab} and p_{ab}. Putting together our results, we end up with the following expression:

$$\begin{aligned} \mathbb{E}Z^n = \int &\Big(\prod_{a,b} \mathrm{d}q_{ab}\mathrm{d}p_{ab}\mathrm{d}\hat{q}_{ab}\mathrm{d}\hat{p}_{ab}\Big)\Big(\prod_a \mathrm{d}m_a\mathrm{d}\hat{m}_a\Big) \exp\Big(-\frac{\alpha N}{2}\ln\det(\mathbb{I} + 2\mathrm{i}\hat{\mathbb{P}}) \\ &+ \mathrm{i}N\sum_{a,b} q_{ab}\hat{q}_{ab} + \mathrm{i}\alpha N\sum_{a,b} p_{ab}\hat{p}_{ab} + \frac{\beta\alpha^2}{2(1+\alpha)}\sum_{a,b=1}^{n} q_{ab}^2 p_{ab} \\ &+ \mathrm{i}N\sum_a m_a\hat{m}_a + N\frac{\beta}{2(1+\alpha)}\sum_a m_a^4\Big) \\ &\times \Big(\prod_{a=1}^{n}\sum_{\boldsymbol{\sigma}^a}\Big) \exp\Big(-\mathrm{i}\sum_i\sum_{a,b}\hat{q}_{ab}\sigma_i^a\sigma_i^b - \mathrm{i}\sum_a \hat{m}_a\sum_i \xi_i\sigma_i^a\Big). \end{aligned} \tag{2.23}$$

The last line in the latter equation can be written as

[8] It can be shown that, when the pattern ξ_{ij} is not fully factorized in the product of two copies of the same vector ξ_i, there are no possible spin configurations $\boldsymbol{\sigma}$ giving $|M| = 1$. Roughly speaking, this is due to the fact that the matrix ξ_{ij} has $\mathcal{O}(N^2)$ degrees of freedom, in contrast to the solely $\mathcal{O}(N)$ of a N-spin network.

[9] We neglect the irrelevant term $\left(\frac{N}{2\pi}\right)^{2n}$, as it gives no contribution in equation (2.12).

$$\begin{aligned}&\Big(\prod_{a=1}^{n}\sum_{\boldsymbol{\sigma}^a}\Big)\exp\Big(-\mathrm{i}\sum_i\sum_{a,b}\hat{q}_{ab}\sigma_i^a\sigma_i^b-\mathrm{i}\sum_a\hat{m}_a\sum_i\xi_i\sigma_i^a\Big)\\&=\exp\Big[N\Big\langle\ln\Big(\prod_{a=1}^{n}\sum_{\sigma^a=\pm1}\Big)\exp\Big(-\mathrm{i}\sum_{a,b}\hat{q}_{ab}\sigma^a\sigma^b-\mathrm{i}\sum_a\hat{m}_a\xi\sigma^a\Big)\Big\rangle_\xi\Big]\end{aligned}\tag{2.24}$$

where the average over $\boldsymbol{\xi}$ has been defined as

$$\langle g(\boldsymbol{\xi})\rangle_\xi\equiv\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}g(\boldsymbol{\xi}_i).\tag{2.25}$$

Assuming the commutativity of the two limits $N\to\infty$ and $n\to0$ (following the replica trick paradigm [15]), we can compute the statistical pressure in the thermodynamic limit through the saddle point method, which gives

$$A=\lim_{n\to0}\frac{1}{n}\mathrm{Extr}\,\phi,\tag{2.26}$$

where ϕ is the argument of the exponential in the partition function (see equations (2.23) and (2.24)), i.e.:

$$\begin{aligned}\phi=&\,\mathrm{i}\sum_{a,b}q_{ab}\hat{q}_{ab}+\mathrm{i}\alpha\sum_{a,b}p_{ab}\hat{p}_{ab}+\frac{\beta\alpha^2}{2(1+\alpha)}\sum_{a,b=1}^{n}q_{ab}^2p_{ab}-\frac{\alpha}{2}\ln\det(\mathbb{I}-2\mathrm{i}\hat{\mathbb{P}})\\&+\mathrm{i}\sum_a m_a\hat{m}_a+\frac{\beta}{2(1+\alpha)}\sum_a m_a^4+\Big\langle\ln\sum_{\boldsymbol{\sigma}}\exp\Big(-\mathrm{i}\sum_{a,b}\hat{q}_{ab}\sigma^a\sigma^b-\mathrm{i}\sum_a\hat{m}_a\xi\sigma^a\Big)\Big\rangle_\xi.\end{aligned}\tag{2.27}$$

We can drop out the conjugates momenta by imposing the saddle point conditions on p,q and m respectively, which correspondingly give

$$\hat{p}_{ab}=\frac{\mathrm{i}}{2}\frac{\beta\alpha}{1+\alpha}q_{ab}^2,\qquad\hat{q}_{ab}=\mathrm{i}\frac{\beta\alpha^2}{1+\alpha}q_{ab}p_{ab},\qquad\hat{m}_a=2\mathrm{i}\frac{\beta}{1+\alpha}m_a^3.\tag{2.28}$$

With these conditions, we obtain a simpler form for ϕ, namely:

$$\begin{aligned}\phi=&-\frac{\beta\alpha^2}{1+\alpha}\sum_{a,b}q_{ab}^2p_{ab}^2-\frac{\alpha}{2}\ln\det(\mathbb{I}+\frac{\beta\alpha}{1+\alpha}\mathbb{Q})-\frac{3}{2}\frac{\beta}{1+\alpha}\sum_a m_a^4\\&+\Big\langle\ln\sum_{\boldsymbol{\sigma}}\exp\Big(\frac{\beta\alpha^2}{1+\alpha}\sum_{a,b}q_{ab}p_{ab}\sigma^a\sigma^b-\frac{2\beta}{1+\alpha}\xi\sum_a m_a^3\sigma^a\Big)\Big\rangle_\xi,\end{aligned}\tag{2.29}$$

where the matrix $\mathbb{Q}$ has been defined as $\mathbb{Q}_{ab}\equiv q_{ab}^2$.

Definition 5. The replica symmetric ansatz (*RS*) for this network model reads

$$q_{ab}=\delta_{ab}+q(1-\delta_{ab}),\qquad p_{ab}=p_D\delta_{ab}+p(1-\delta_{ab}),\qquad m_a=m.\tag{2.30}$$

We are now able to enunciate the following proposition regarding the quenched pressure (2.5) in the RS ansatz:

Proposition 1. *The replica symmetric expression of the quenched pressure related to the model (1) reads*

$$A^{RS} = \ln 2 - \frac{\beta\alpha^2}{1+\alpha}(qp - q^2 p) - \frac{\alpha}{2}\ln\left(1 - \frac{\beta\alpha}{1+\alpha}(1-q^2)\right) + \frac{\alpha}{2}\frac{\beta\alpha}{1+\alpha}\frac{q^2}{1-\frac{\beta\alpha}{1+\alpha}(1-q^2)} - \frac{3}{2}\frac{\beta}{1+\alpha}m^4 + \int Dx\ \ln\cosh\left(\sqrt{2\frac{\beta\alpha^2}{1+\alpha}pq}\,x + \frac{2\beta}{1+\alpha}m^3\right). \tag{2.31}$$

Proof. The details of the RS ansatz computations are reported in appendix A.2. □

2.2. Route two: interpolation method

We now proceed to check the validity of the replica trick computation with an alternative route, i.e. the Guerra's interpolation method. Given the expression of Z_N in equation (2.7), and substituting the explicit form of η (according to definition 2) in terms of signal and noise, the statistical pressure in the thermodynamic limit reads

$$A = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\ln\sum_{\sigma}\int Dz\ \exp\Big(\sqrt{\frac{\beta}{(1+\alpha)N^3}}\sum_{\mu=1}^{K}\sum_{i,j=1}^{N}\xi_{ij}^{\mu}\sigma_i\sigma_j z_\mu + \sqrt{\frac{\beta\alpha}{(1+\alpha)N^2}}\sum_{\mu=1}^{K}\sum_{i,j=1}^{N}J_{ij}^{\mu}\sigma_i\sigma_j z_\mu\Big). \tag{2.32}$$

Again, we isolate the signal ($\mu = 1$) from the noise ($\mu > 1$), always neglecting the irrelevant term because of equation (2.15). Thus

$$A = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\ln\sum_{\sigma}\int Dz\exp\Big(\sqrt{\frac{\beta}{(1+\alpha)N^3}}\sum_{\mu>1}^{K}\sum_{i,j=1}^{N}\xi_i^{\mu}\xi_j^{\mu}\sigma_i\sigma_j z_\mu + \sqrt{\frac{\beta\alpha}{(1+\alpha)N^2}}\sum_{\mu>1}^{K}\sum_{i,j=1}^{N}J_{ij}^{\mu}\sigma_i\sigma_j z_\mu + \frac{\beta N}{2(1+\alpha)}\Big(\frac{1}{N}\sum_{i,j=1}^{N}\xi_i^1\sigma_i\Big)^4\Big). \tag{2.33}$$

Notice that we already adopted the signal factorization $\xi_{ij} = \xi_i\xi_j$, which allows us to directly express everything in terms of the Mattis magnetization m_1 associated to the retrieved pattern $\mu = 1$.

We are now ready to set up the interpolation strategy. We introduce an interpolating parameter $t \in (0,1)$ such that (in its extrema) it compares the original model (recovered for $t = 1$) and a *simpler* model at $t = 0$. Hence we introduce the next

Definition 6. The Guerra's interpolating pressure for the MDAM coded by the cost function (1) reads as

$$\mathcal{A}(t) = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\ln\sum_{\sigma}\int Dz\exp\Big(\sqrt{t}\sqrt{\frac{\beta}{(1+\alpha)N^3}}\sum_{\mu>1}^{K}\sum_{i,j=1}^{N}\xi_i^{\mu}\xi_j^{\mu}\sigma_i\sigma_j z_\mu + \sqrt{t}\sqrt{\frac{\beta\alpha}{(1+\alpha)N^2}}\sum_{\mu>1}^{K}\sum_{i,j=1}^{N}J_{ij}^{\mu}\sigma_i\sigma_j z_\mu + t\frac{\beta N}{2(1+\alpha)}m_1^4 + \sqrt{1-t}\mathcal{W} + (1-t)\mathcal{D}\Big), \tag{2.34}$$

where $\mathcal{W}$ and $\mathcal{D}$ are defined as

$$\mathcal{W} = \sqrt{\frac{\beta}{1+\alpha}} C_1 \sum_i J_i \sigma_i + \sqrt{\frac{\beta}{1+\alpha}} C_2 \sum_\mu J_\mu z_\mu, \tag{2.35}$$

$$\mathcal{D} = C_3 \frac{\beta}{1+\alpha} \sum_\mu \frac{z_\mu^2}{2} + C_4 \frac{\beta N}{(1+\alpha)} m_1, \tag{2.36}$$

and C_1, ..., C_4 are constants whose explicit values will be set later, see equation (2.45).

The *interpolating variables* J_i and J_μ are, respectively, N-component and K-component vectors of i.i.d. $\mathcal{N}(0,1)$ variables. Therefore, the expectation $\mathbb{E}$ is now extended to include these new degrees of freedom.

Proposition 2. *The quenched pressure related to the model (1) can thus be recovered using the fundamental theorem of calculus:*

$$A = \mathcal{A}(t=1) = \mathcal{A}(t=0) + \int_0^1 \mathrm{d}t\, \partial_t \mathcal{A}(t). \tag{2.37}$$

Following the scheme used in [12, 14, 22, 26], we can evaluate separately $\partial_t \mathcal{A}$ and $\mathcal{A}(0)$. Tackling the t-streaming first and keeping as order parameters those defined in equations (2.8), (2.9) and (2.11), we obtain

$$\begin{aligned} \partial_t \mathcal{A}(t) = \frac{\beta}{2(1+\alpha)} \mathbb{E}\big[& \alpha^2 \langle p_{11} \rangle - \alpha^2 \langle p_{12} q_{12}^2 \rangle - C_1^2 + C_1^2 \langle q_{12} \rangle - \alpha C_2^2 \langle p_{11} \rangle + \alpha C_2^2 \langle p_{12} \rangle \\ & - \alpha C_3 \langle p_{11} \rangle + \langle m_1^4 \rangle - 2 C_4 \langle m_1 \rangle \big], \end{aligned} \tag{2.38}$$

where we use the standard notation $\langle . \rangle$ for the Boltzmann average[10]. The terms involving the Mattis magnetizations for $\mu > 1$ do not appear in the streaming equation since their contribution is subleading in the thermodynamic limit.

The expected Boltzmann averages $\langle q_{12} \rangle$, $\langle p_{12} \rangle$, $\langle p_{11} \rangle$ and $\langle m_1 \rangle$ are difficult to compute, but recall that we are interested in the replica symmetric evaluation of the quenched free energy (and, thus, also of the replica symmetric expression of all the order parameters). Introducing the fluctuations of the order parameter (centered around their quenched mean values q, p and m, see equation (2.30))

$$\Delta_q = q_{12} - q, \tag{2.39}$$

$$\Delta_p = p_{12} - p, \tag{2.40}$$

$$\Delta_m = m_1 - m, \tag{2.41}$$

and recalling that, in the *RS* approximation, they vanish in the thermodynamic, we can rewrite the interaction terms in equation (2.38) as

$$\begin{aligned} \langle p_{12} q_{12}^2 \rangle &= -2pq^2 + q^2 \langle p_{12} \rangle + 2pq \langle q_{12} \rangle, \\ \langle m_1^4 \rangle &= -3m^4 + 4m^3 \langle m_1 \rangle. \end{aligned} \tag{2.42}$$

[10] Notice that the Boltzmann average has a functional dependence from the interpolating parameter t, as every thermodynamic observable is computed from the general interpolating pressure (2.34).

By substitution inside the streaming equation we obtain

$$\partial_t \mathcal{A} = \frac{\beta}{2(1+\alpha)} \mathbb{E}\big[\alpha^2\langle p_{11}\rangle + 2\alpha^2 pq^2 - \alpha^2 q^2\langle p_{12}\rangle - 2\alpha^2 pq\langle q_{12}\rangle - C_1^2 + C_1^2\langle q_{12}\rangle - \alpha C_2^2\langle p_{11}\rangle + \alpha C_2^2\langle p_{12}\rangle - \alpha C_3\langle p_{11}\rangle - 3m^4 + 4m^3\langle m_1\rangle - 2C_4\langle m_1\rangle\big]. \tag{2.43}$$

Recall that we have four free parameters: $C_1,\ C_2,\ C_3$ and C_4. They can be chosen *a fortiori* in order to eliminate the expected Boltzmann averages $\langle q_{12}\rangle$, $\langle p_{12}\rangle$, $\langle p_{11}\rangle$ and $\langle m_1\rangle$ in favour of their replica-symmetric expectations in the thermodynamic limit. With this idea in mind, we rewrite the latter equation as

$$\partial_t \mathcal{A} = \frac{\beta}{2(1+\alpha)} \mathbb{E}\big[(\alpha^2 - \alpha C_2^2 - \alpha C_3)\langle p_{11}\rangle + (\alpha C_2^2 - \alpha^2 q^2)\langle p_{12}\rangle + (C_1^2 - 2\alpha^2 pq)\langle q_{12}\rangle + (4m^3 - 2C_4)\langle m_1\rangle - C_1^2 + 2\alpha^2 pq^2 - 3m^4\big]. \tag{2.44}$$

It is now clear that, with the following choice:

$$C_1 = \sqrt{2\alpha^2 pq}, \qquad C_2 = \sqrt{\alpha} q, \qquad C_3 = \alpha(1-q^2), \qquad C_4 = 2m^3, \tag{2.45}$$

we can achieve our goal and simplify the streaming term further, obtaining

$$\partial_t \mathcal{A} = -\frac{\beta}{2(1+\alpha)}\big[2\alpha^2 pq(1-q) + 3m^4\big]. \tag{2.46}$$

Now we are left with the one body term:

$$\mathcal{A}(0) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\ln \sum_{\sigma} \int Dz \exp\Big(\sqrt{\frac{\beta}{1+\alpha}} C_1 \sum_i J_i\sigma_i + \sqrt{\frac{\beta}{1+\alpha}} C_2 \sum_\mu J_\mu z_\mu + C_3 \frac{\beta}{1+\alpha}\sum_\mu \frac{z_\mu^2}{2} + C_4 \frac{\beta}{1+\alpha}\sum_{i=1}^{N} \xi_i^1\sigma_i\Big). \tag{2.47}$$

It can be easily computed, returning

$$\mathcal{A}(0) = -\frac{\alpha}{2}\ln\Big(1 - \frac{\beta\alpha}{1+\alpha}(1-q^2)\Big) + \frac{\alpha}{2}\frac{\beta\alpha}{1+\alpha}\frac{q^2}{1-\frac{\beta\alpha}{1+\alpha}(1-q^2)} + \int Dx\ \ln\cosh\Big(\sqrt{2\frac{\beta\alpha^2}{1+\alpha}pq}\,x + \frac{2\beta}{1+\alpha}m^3\Big). \tag{2.48}$$

Combining equations (2.46) and (2.48), after some rearrangements we get the same result already derived through the replica trick for the RS pressure (2.31).

2.3. Phase diagram

Before moving on, we rewrite here the *RS* pressure for the reader's convenience:

$$A^{RS} = \ln 2 - \frac{\beta\alpha^2}{1+\alpha}(qp - q^2 p) - \frac{\alpha}{2}\ln\Big(1 - \frac{\beta\alpha}{1+\alpha}(1-q^2)\Big) + \frac{\alpha}{2}\frac{\beta\alpha}{1+\alpha}\frac{q^2}{1-\frac{\beta\alpha}{1+\alpha}(1-q^2)} - \frac{3}{2}\frac{\beta}{1+\alpha}m^4 + \int Dx\ \ln\cosh\Big(\sqrt{2\frac{\beta\alpha^2}{1+\alpha}pq}\,x + \frac{2\beta}{1+\alpha}m^3\Big). \tag{2.49}$$

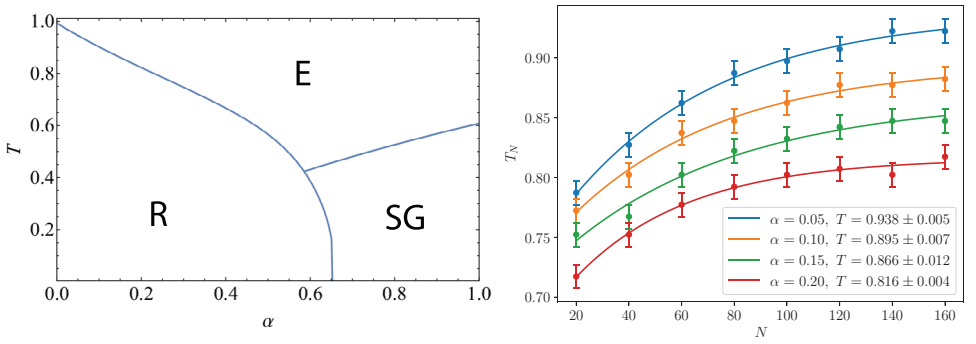

Figure 1. (left) The phase diagram of the MDAM obtained by solving the self-consistencies, see equations (2.50). We highlight three regions: a pure ergodic one (E), a spin glass phase (SG) and a retrieval one (R). (right) Different critical lines T_N depicting the E-R transition relative to different loads of the network, i.e. for $\alpha = 0.05, 0.10, 0.15$ and 0.20, as function of N. For each load α we performed Monte Carlo simulations for different sizes N, ranging from $N = 20$ to $N = 160$, with leaps of $\Delta N = 20$, thus obtaining the different points interpolated in the right panel of the figure. The critical temperatures T (for different values of α) in the bottom right legend of the figure are so obtained; they are consistent with our theoretical results.

Extremizing the statistical pressure with respect to the parameters q, p and m, we end up with the self-consistency equations

$$\begin{aligned} m &= \int Dx \, \tanh\left(\sqrt{2\frac{\beta\alpha^2}{1+\alpha}pq}\, x + \frac{2\beta}{1+\alpha}\, m^3\right), \\ q &= \int Dx \, \tanh^2\left(\sqrt{2\frac{\beta\alpha^2}{1+\alpha}pq}\, x + \frac{2\beta}{1+\alpha}\, m^3\right), \\ p &= \frac{\frac{\beta\alpha}{1+\alpha}q^2}{\left(1 - \frac{\beta\alpha}{1+\alpha}\,(1-q^2)\right)^2}. \end{aligned} \tag{2.50}$$

We numerically solve these equations and paint the phase diagram reported in figure 1(left panel), made by three different phases: the retrieval (R), characterized by non-zero values of the two order parameters m and q, i.e. $m \neq 0$ and $q \neq 0$; the spin glass phase (SG), where $m = 0$, $q \neq 0$, and the ergodic phase (E), with $m = q = 0$. In the retrieval region pure states are always global minima for the free energy. Since mixture states (which are present as well as pure ones) are not global minima of the free energy, we refer to the R phase as a 'pure retrieval' phase. We argue that this is due to the decomposition (2.3).

Furthermore we performed Monte Carlo simulations in order to check out our assumptions (e.g. the RS ansatz). In particular, we focused on the E-R critical line (see the phase diagram, figure 1(left)). We performed a finite size scaling analysis (see figure 1, (right)), which led to the critical temperatures for different values of α depicted in the bottom right panel, figure 1, (right): numerical outcomes are in excellent agreement with the theoretical predictions.

3. Conclusions

Along the lines of our recent research [6, 10], in this paper we extensively relied upon tools typical of the statistical mechanics of spin-glasses to quantify the high pattern recognition capacity of, possibly, the simplest neural network falling in the class of *dense associative memories*. The latter were recently proposed by Hopfield and Krotov [35, 36] as a candidate benchmark to inspect for possibly explaining (part of) the impressive skills that artificial neural architectures experience nowadays.

In particular we have shown that such a network, equipped with *solely* a linear storage of patterns K—in the volume N—but where patterns are split in a $\mathcal{O}(1)$ signal term and an $\mathcal{O}(\sqrt{N})$ noisy term, is able to extensively de-noise the perceived inputs such as to accomplish pattern recognition despite the prohibitive level of noise: this is ultimately due to the dense connections where redundant representations of patterns are possible [6]. The critical capacity in this regime of operation -at least at the replica symmetric level of description- is quite huge, resulting in $\alpha_c(\beta \to \infty) \sim 0.65$ (and we checked numerically that the replica symmetric assumption is tolerated as shown by extensive Monte Carlo runs). In particular, at present -to our knowledge- this is the largest critical capacity for networks presenting this high pattern recognition skill, (the network of [6] has to respect $\alpha_c \leqslant 0.5$).

Furthermore the retrieval region is characterized by the fact that pure states are always global minima for the free energy. We conjecture this is a consequence of the choice of the decomposition (2.3). We also notice that the phase diagram resembles that of the Sherrington–Kirkpatrick model (with the role of J_0, the mean value for the Gaussian fields J_{ij} in the *S-K* model, here played by some decreasing function of the load α); this is not casual, as, in this model, the quenched noise given by the boolean fields $\boldsymbol{\xi}^\mu$, for $\mu > 1$, is negligible in the thermodynamic limit (see appendix A.1). Therefore, the only 'seen' by the network is the Gaussian noise (given by the fields J_{ij}^μ, still for $\mu > 1$). Consequently, the model can be seen as a spin glass (a $P = 4$ generalization of the Sherrington–Kirkpatrick model) able to retrieve a pattern $\boldsymbol{\xi}$ from the Gaussian noise.

Finally, with the aim of promoting cross-fertilization among the two disciplines of Machine Learning and Disordered Statistical Mechanics, we collected the outlined results by using two among the most used methods to deal with spin-glasses, namely the replica trick [19] and the interpolation method [12], discussing both of them in great detail. The authors are grateful to Unisalento, Istituto Nazionale di Fisica Nucleare, Sezione di Lecce, and CNR-Nanotec, Sezione di Lecce, for partial fundings.

Acknowledgment

The authors are grateful to Unisalento, Istituto Nazionale di Fisica Nucleare, Sezione di Lecce, and CNR-Nanotec, Sezione di Lecce, for partial fundings?

Appendix A. Replica trick computations: details

In this appendix, we report some details on the replica trick computation.

A.1. Evaluation of the noise term

In this section, we evaluate the noise term in the splitted partition function (2.17), which we report here for convenience:

$$Z_{\text{noise}} = \int \Big(\prod_{a=1}^{n} D\{z^a\}_{\mu>1}\Big)\mathbb{E}\exp\Big(\sqrt{\frac{\beta}{(1+\alpha)N^3}}\sum_{a=1}^{n}\sum_{\mu>1}^{K}\sum_{i,j=1}^{N}\Big(\xi_{ij}^{\mu}+\sqrt{\alpha N}J_{ij}^{\mu}\Big)\,\sigma_i^a\sigma_j^a z_\mu^a\Big). \tag{A.1}$$

Because of the independence of the signal and noise term in the pattern decomposition (2.3), we can perform the averages separately. We start with performing the average over the ξ's, which leads to

$$\exp\Big(\sum_{i,j,\mu>1}\ln\cosh\Big(\sqrt{\frac{\beta}{(1+\alpha)N^3}}\sum_{a}\sigma_i^a\sigma_j^a z_\mu^a\Big)\Big). \tag{A.2}$$

In the large N limit, we can expand in powers of $1/N$ the $\ln\cosh$ function, keeping only the leading contribution (as all higher order corrections vanish in the thermodynamic limit). Then, we are left with

$$\exp\Big(\frac{\beta}{2(1+\alpha)N^3}\sum_{i,j,\mu>1}\Big(\sum_{a}\sigma_i^a\sigma_j^a z_\mu^a\Big)^2\Big). \tag{A.3}$$

However, the exponent in the latter equation is of order $\mathcal{O}(1)$, thus it is a subleading contribution w.r.t. to the Gaussian part of the noise term. Therefore, it can been neglected in the large N limit; this is an important result, as it means that the quenched noise given by the boolean fields $\boldsymbol{\xi}^{\mu}$, for $\mu > 1$, is totally hidden by the Gaussian noise, introduced by the decomposition (2.3). The result is

$$Z_{\text{noise}} = \int DJ\int\Big(\prod_{a=1}^{n}D\{z^a\}_{\mu>1}\Big)\exp\Big(\sqrt{\frac{\beta\alpha}{(1+\alpha)N^2}}\sum_{a=1}^{n}\sum_{\mu>1}^{K}\sum_{i,j=1}^{N}J_{ij}^{\mu}\sigma_i^a\sigma_j^a z_\mu^a\Big). \tag{A.4}$$

Now, we can directly average over the J variables, obtaining

$$\begin{aligned} Z_{noise} &= \int\Big(\prod_{a=1}^{n}D\{z^a\}_{\mu>1}\Big)\exp\Big(\frac{\beta\alpha}{2(1+\alpha)N^2}\sum_{\mu>1}^{K}\sum_{i,j=1}^{N}\Big(\sum_{a=1}^{n}\sigma_i^a\sigma_j^a z_\mu^a\Big)^2\Big)\\ &= \int\Big(\prod_{a=1}^{n}D\{z^a\}_{\mu>1}\Big)\exp\Big(\frac{\beta\alpha}{2(1+\alpha)N^2}\sum_{\mu>1}^{K}\sum_{i,j=1}^{N}\sum_{a,b=1}^{n}\sigma_i^a\sigma_i^b\sigma_j^a\sigma_j^b z_\mu^a z_\mu^b\Big). \end{aligned} \tag{A.5}$$

In the last line, the dependence on the order parameters q_{ab} and p_{ab} is clear. Hence, we can now introduce a product of delta functions by using

$$1 = \int\Big(\prod_{a,b}\mathrm{d}q_{ab}\mathrm{d}p_{ab}\,\delta\big(q_{ab}-\frac{1}{N}\sum_{i=1}^{N}\sigma_i^a\sigma_i^b\big)\,\delta\big(p_{ab}-\frac{1}{K-1}\sum_{\mu=2}^{K}z_\mu^a z_\mu^b\big)\Big). \tag{A.6}$$

After this manipulation, we get

$$\begin{aligned} Z_{\text{noise}} = &\int\Big(\prod_{a=1}^{n}D\{z^a\}_{\mu>1}\Big)\Big(\prod_{a,b}\mathrm{d}q_{ab}\mathrm{d}p_{ab}\,\delta\big(q_{ab}-\frac{1}{N}\sum_{i=1}^{N}\sigma_i^a\sigma_i^b\big)\,\delta\big(p_{ab}-\frac{1}{K-1}\sum_{\mu=2}^{K}z_\mu^a z_\mu^b\big)\Big)\\ &\times\exp\Big(N\frac{\beta\alpha^2}{2(1+\alpha)}\sum_{a,b=1}^{n}q_{ab}^2 p_{ab}\Big). \end{aligned} \tag{A.7}$$

At this point, we use the Fourier representation of the Dirac delta:

$$\delta\big(q_{ab}-\frac{1}{N}\sum_{i}\sigma_i^a\sigma_i^b\big) = \frac{N}{2\pi}\int\mathrm{d}\hat{q}_{ab}\exp\Big(\mathrm{i}N\hat{q}_{ab}\big(q_{ab}-\frac{1}{N}\sum_{i}\sigma_i^a\sigma_i^b\big)\Big), \tag{A.8}$$

and similarly for p_{ab}. Then, the noise term now reads[11]:

$$\int \Big(\prod_{a,b} \mathrm{d}q_{ab}\mathrm{d}p_{ab}\mathrm{d}\hat{q}_{ab}\mathrm{d}\hat{p}_{ab}\Big)\Big(\prod_{a} D\{z^a\}_{\mu>1}\Big)\exp\Big(-\mathrm{i}\sum_{\mu>1}\sum_{a,b}\hat{p}_{ab}z^a_\mu z^b_\mu + \mathrm{i}N\sum_{a,b} q_{ab}\hat{q}_{ab} - \mathrm{i}\sum_i\sum_{a,b}\hat{q}_{ab}\sigma^a_i\sigma^b_i + \mathrm{i}\alpha N\sum_{a,b}p_{ab}\hat{p}_{ab} + \frac{\beta\alpha^2}{2(1+\alpha)}\sum_{a,b=1}^{n} q^2_{ab}p_{ab}\Big). \tag{A.9}$$

The integral over the z variables can be easily performed, leading to

$$\int \Big(\prod_{a} D\{z^a\}_{\mu>1}\Big)\exp\Big(-\mathrm{i}\sum_{\mu>1}\sum_{a,b}\hat{p}_{ab}z^a_\mu z^b_\mu\Big) = \prod_{\mu>1}^{K}\det(\mathbb{I}+2\mathrm{i}\hat{\mathbb{P}})^{-1/2}, \tag{A.10}$$

where $\mathbb{I}_{ab}=\delta_{ab}$ is the $n\times n$ identity matrix and $\hat{\mathbb{P}}_{ab}\equiv\hat{p}_{ab}$. We therefore end with the final expression for the noise term

$$Z_{\text{noise}} = \int \Big(\prod_{a,b}\mathrm{d}q_{ab}\mathrm{d}p_{ab}\mathrm{d}\hat{q}_{ab}\mathrm{d}\hat{p}_{ab}\Big)\exp\Big(-\frac{\alpha N}{2}\ln\det(\mathbb{I}+2\mathrm{i}\hat{\mathbb{P}})\Big) \times \exp\Big(\mathrm{i}N\sum_{a,b}q_{ab}\hat{q}_{ab} - \mathrm{i}\sum_i\sum_{a,b}\hat{q}_{ab}\sigma^a_i\sigma^b_i + \mathrm{i}\alpha N\sum_{a,b}p_{ab}\hat{p}_{ab} + \frac{\beta\alpha^2}{2(1+\alpha)}\sum_{a,b=1}^{n}q^2_{ab}p_{ab}\Big). \tag{A.11}$$

A.2. The replica symmetric ansatz

In this section, we compute term by term the contributions appearing in (2.29) after adopting the RS ansatz. Since the statistical pressure presents an overall factor $1/n$ in (2.26), only the $\mathcal{O}(n)$ terms are relevant for our purposes (since we have to evaluate the $n\to 0$ limit). For the first term, the leading contribution is

$$\sum_{a,b} q^2_{ab}p^2_{ab} \sim n(p_D - pq^2). \tag{A.12}$$

For the second one, we have

$$\ln\det\Big(\mathbb{I}-\frac{\beta\alpha}{1+\alpha}\mathbb{Q}\Big) \sim n\ln\Big(1-\frac{\beta\alpha}{1+\alpha}(1-q^2)\Big) - n\frac{\beta\alpha}{1+\alpha}\frac{q^2}{1-\frac{\beta\alpha}{1+\alpha}(1-q^2)}. \tag{A.13}$$

The m-dependent contribution is trivial, and reads

$$\sum_a m^4_a = n\,m^4. \tag{A.14}$$

Finally, the last term can be straightforwardly evaluated as follows

$$\Big\langle \ln\sum_{\boldsymbol{\sigma}}\exp\Big(\frac{\beta\alpha^2}{1+\alpha}\sum_{a,b}q_{ab}p_{ab}\sigma^a\sigma^b + \frac{2\beta}{1+\alpha}\xi\sum_a m^3_a\sigma^a\Big)\Big\rangle_\xi = n\frac{\beta\alpha^2}{1+\alpha}(p_D-pq) + n\ln 2 + n\int Dx\,\ln\cosh\Big(\sqrt{2\frac{\beta\alpha^2}{1+\alpha}pq}\,x + \frac{2\beta}{1+\alpha}m^3\Big). \tag{A.15}$$

Putting all these results in the expression for the statistical pressure, we get the result (2.31).

[11] Again, we neglect the irrelevant factors $\left(\frac{N}{2\pi}\right)^{n^2}\left(\frac{\alpha N}{2\pi}\right)^{n^2}$, as they give no contribution in equation (2.12).

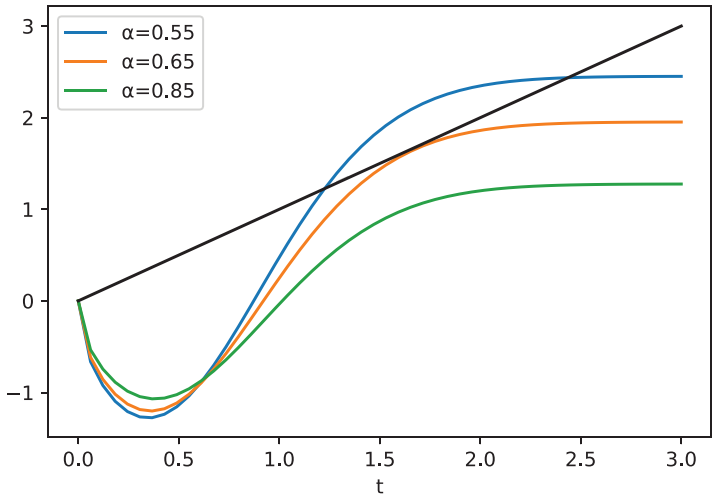

Figure B1. Comparison between the l.h.s. (black dashed line) and r.h.s. (blue solid curves) of equation (B.5) for $\alpha = 0.55, 0.65, 0.85$.

Appendix B. Zero-temperature critical capacity analysis

In order to estimate the zero-temperature critical capacity $\alpha_c(T = 0)$, we start from the self-consistency equations (2.50). Upon eliminating the conjugate parameter p, we get

$$\begin{aligned} m &= \int Dx \, \tanh\left(\frac{\beta}{1+\alpha}\left(\frac{\sqrt{2\alpha^3 q^3}}{1-\frac{\beta\alpha}{1+\alpha}(1-q^2)}\, x + 2\, m^3\right)\right), \\ q &= \int Dx \, \tanh^2\left(\frac{\beta}{1+\alpha}\left(\frac{\sqrt{2\alpha^3 q^3}}{1-\frac{\beta\alpha}{1+\alpha}(1-q^2)}\, x + 2\, m^3\right)\right). \end{aligned} \tag{B.1}$$

In the limit $\beta \to \infty$, it is easy to check that $q \to 1$, then $1 - q^2 \to 0$ in the zero temperature limit. The quantity $C = \beta(1 - q^2)$, which satisfies the self-consistency equation

$$C = \beta - \beta\left(1 - \frac{1+\alpha}{2\beta}\frac{\partial}{\partial(m^3)}\int Dx \tanh(g(m,q))\right)^2, \tag{B.2}$$

where $g(m, q)$ is the argument of the hyperbolic tangent in (B.1), is finite in the $\beta \to \infty$ limit. Using $\tanh(\beta x) \to \mathrm{sgn}(x)$ in the large β limit, then the self-consistency equations can be evaluated as

$$\begin{aligned} m &= \mathrm{erf}\left(\frac{m^3}{\alpha^{3/2}}\left(1 - \frac{\alpha}{1+\alpha}(1-C)\right)\right), \\ C &= (\alpha+1)\frac{1-\frac{\alpha}{1+\alpha}C}{\alpha^{3/2}}\frac{2}{\sqrt{\pi}}\exp\left(-\frac{m^6}{\alpha^3}(1-\frac{\alpha}{1+\alpha}C)^2\right). \end{aligned} \tag{B.3}$$

By introducing the quantity

$$t = \frac{m^3}{\alpha^{3/2}}\left(1 - \frac{\alpha}{1+\alpha}(1-C)\right), \tag{B.4}$$

after some rearrangements, we end up with a single equation

$$t = \frac{1}{\alpha^{3/2}}\mathrm{erf}^3(t) - \frac{2t}{\sqrt{\alpha\pi}}\exp(-t^2). \tag{B.5}$$

Then, the critical storage capacity is the value of α leading to non-trivial solutions for equation (B.5). A comparison between the two sides of the equation for various α values is reported in figure B1. By numerically solving the equation (B.5), we found a critical storage capacity $\alpha_c(T=0) \simeq 0.651$, which is in perfect agreement with the phase diagram.

Appendix C. Signal-to-noise analysis

We here perform a signal-to-noise analysis, motivating the decomposition equation (2.3). Introducing the 'internal' field h_i seen by the ith spin σ_i, defined as

$$h_i = \frac{1}{2N^3}\sum_{\mu=1}^{K}\sum_{j,k,l=1}^{N}\eta_{ij}^{\mu}\eta_{kl}^{\mu}\sigma_j\sigma_k\sigma_l, \tag{C.1}$$

we can write the hamiltonian of the model as:

$$H = -\sum_{i=1}^{N} h_i\sigma_i. \tag{C.2}$$

By virtue of the pattern decomposition, this field can be rewritten as

$$h_i = \frac{1}{2N^3}\sum_{\mu=1}^{K}\sum_{j,k,l=1}^{N}\left(\xi_i^{\mu}\xi_j^{\mu}\xi_k^{\mu}\xi_l^{\mu} + \sqrt{K}\xi_i^{\mu}\xi_j^{\mu}J_{kl}^{\mu} + \sqrt{K}\xi_k^{\mu}\xi_l^{\mu}J_{ij}^{\mu} + KJ_{ij}^{\mu}J_{kl}^{\mu}\right)\sigma_j\sigma_k\sigma_l. \tag{C.3}$$

Probing the alignment to the pattern $\boldsymbol{\xi}^1 = (\xi_1^1, .., \xi_N^1)$, we set $\boldsymbol{\sigma} = \boldsymbol{\xi}^1$, by which the following standard decomposition holds (we simply separate the 'signal' characterized by $\mu = 1$ from the 'noise' $\mu > 1$):

$$h_i\sigma_i = \mathcal{S} + \mathcal{N} \tag{C.4}$$

where

$$\mathcal{S} = \frac{1}{2}\left[1 + \frac{\sqrt{K}}{N}\sum_j J_{ij}^1\xi_i^1\xi_j^1 + \frac{\sqrt{K}}{N^2}\sum_{k,l}J_{kl}^1\xi_k^1\xi_l^1 + \frac{K}{N^3}\sum_{j,k,l}\xi_i^1\xi_j^1\xi_k^1\xi_l^1 J_{il}^1 J_{kl}^1\right] \tag{C.5}$$

is the 'signal' and

$$\mathcal{N} = \frac{1}{2N^3}\sum_{\mu>1}^{K}\sum_{j,k,l=1}^{N}\left(\xi_i^{\mu}\xi_i^1\xi_j^{\mu}\xi_j^1\xi_k^{\mu}\xi_k^1\xi_l^{\mu}\xi_l^1 + \sqrt{K}\xi_i^{\mu}\xi_i^1\xi_j^{\mu}\xi_j^1\xi_k^1\xi_l^1 J_{kl}^{\mu} + \sqrt{K}\xi_k^{\mu}\xi_k^1\xi_l^{\mu}\xi_l^1\xi_i^1\xi_j^1 J_{ij}^{\mu} + K\xi_i^1\xi_j^1\xi_k^1\xi_l^1 J_{ij}^{\mu}J_{kl}^{\mu}\right) \tag{C.6}$$

the 'noise'. Recall that the J_{ij}^{μ} tensors are all *i.i.d.* variables distributed as $\mathcal{N}(0,1)$. In order to compute the signal-to-noise ratio $\mathcal{S}/\mathcal{N}$, we firstly perform the standard Gaussian expectation $\mathbb{E}$ over the signal $\mathcal{S}$. This results in

$$\mathbb{E}\left[\mathcal{S}\right] = \frac{1}{2}\left(1 + \frac{K}{N^2}\right) \to \frac{1}{2} \tag{C.7}$$

as $N \to \infty$ in the thermodynamic limit. In order to get this result we consider that

$$\mathbb{E}\left[J_{ij}^1\right] = \mathbb{E}\left[J_{kl}^1\right] = 0 \tag{C.8}$$

and

$$\mathbb{E}\left[J_{ij}^1 J_{kl}^1\right] = \delta_{ik}\delta_{jl} \tag{C.9}$$

and the fact that the products similar to $J_{ij}^1 \xi_i^1 \xi_j^1$ give new *i.i.d* $\mathcal{N}(0,1)$ variables as the J's are.

Consider now the noise $\mathcal{N}$. The first term in the parenthesis (C.6) can be decomposed in a sum of various contributions, given the four summations in μ, j, k, l. We get a contribution by setting $j = k = l = i$, which is of order $\mathcal{O}(N^{-2})$ given the overall factor $1/N^3$ in front of each term in the noise; we have several contributions from $l \neq k$ with $k = j = i$, and cyclic permutations (i.e. $l \neq j$ with $k = j = i$ and so on), which overall give a contribution of order $\mathcal{O}(N^{-2})$; then we have to consider the terms coming from $l \neq i, k \neq i, j = i$ and similar, giving $\mathcal{O}(N^{-3/2})$ contributions and, the remaining ones coming from $l \neq i, k \neq i, j \neq i$ and similar, which are $\mathcal{O}(N^{-1})$. We see therefore that, in the thermodynamic limit, the first term in the noise is zero.

The remaining terms have to be evaluated via the Gaussian expectation operator, therefore we can easily apply similar considerations to those used in the evaluation of $\mathbb{E}\left[\mathcal{S}\right]$. This results in vanishing contributions from the second and the third term in the noise decomposition in the thermodynamic limit. The only non-zero contribution comes from the last term, the fourth, which however is non-zero only for $k = i, j = l$, giving $\alpha^2/2$.

The ratio $\mathcal{S}/\mathcal{N}$ can now easily computed, giving $1/\alpha^2$, which is of order $\mathcal{O}(1)$. It can be shown that this is the minimal $\mathcal{S}/\mathcal{N}$ value given the decomposition in equation (2.3): attempting to overtake the linear load $K = \alpha N$ (e.g. by considering super-linear regimes such as $K \sim N^2$) leads to a vanishing signal to noise ratio in the thermodynamic limit. Hence, in super-linear regimes the network cannot retrieve any pattern of information: our decomposition equation (2.3) is therefore the worst scenario from the network's point of view, i.e. it gives the maximal noise to the network maintaining its retrieval capability.

ORCID iDs

Francesco Alemanno https://orcid.org/0000-0003-1065-2590
Martino Centonze https://orcid.org/0000-0002-7945-4392

References

[1] Agliari E *et al* 2013 Parallel retrieval of correlated patterns: from hopfield networks to boltzmann machines *Neural Netw.* **38** 52
[2] Agliari E *et al* 2012 Multitasking associative networks *Phys. Rev. Lett.* **109** 268101
[3] Agliari E *et al* 2017 Neural networks retrieving binary patterns in a sea of real ones *J. Stat. Phys.* **168** 1085
[4] Agliari E *et al* 2012 Notes on the P-spin glass studied via Hamilton–Jacobi and smooth cavity techniques *J. Math. Phys.* **53** 063304
[5] Alemanno F *et al* 2019 Dreaming neural networks: rigorous results *J.Stat. Mech.* 083503

[6] Agliari E, Alemanno F, Barra A, Centonze M and Fachechi A 2019 Neural networks with redundant representations: detecting the undetectable *Phys. Rev. Lett.* **124** 028301
[7] Amit D J 1989 *Modeling Brain Functions* (Cambridge: Cambridge University Press)
[8] Baldi P and Venkatesh S S 1987 Number of stable points for spin-glasses and neural networks of higher orders *Phys. Rev. Lett.* **58** 913
[9] Barlow H 2001 Redundancy reduction revisited *Network: Comput. Neur. Syst.* **12** 241
[10] Barra A, Beccaria M and Fachechi A 2018 A new mechanical approach to handle generalized Hopfield neural networks *Neural Netw.* **106** 205–22
[11] Barra A, Bernacchia A, Santucci E and Contucci P 2012 On the equivalence of Hopfield networks and Boltzman machines *Neural Netw.* **34** 1–9
[12] Barra A, Genovese G and Guerra F 2010 The replica symmetric approximation of the analogical neural network *J. Stat. Phys.* **140** 784
[13] Barra A, Genovese G, Sollich P and Tantari D 2018 Phase diagram of restricted Boltzmann machines and generalized Hopfield networks with arbitrary priors *Phys. Rev.* E **97** 022310
[14] Haiping H and Kabashima Y 2013 Adaptive Thouless–Anderson–Palmer approach to inverse Ising problems with quenched random fields *Phys. Rev.* E **87** 062129
[15] Barra A, Guerra F and Mingione E 2012 Interpolating the Sherrington–Kirkpatrick replica trick *Phil. Mag.* **91** 78
[16] Bovier A and Niederhauser B 2001 The spin-glass phase-transition in the Hopfield model with p-spin interactions (arXiv cond-mat/0108235)
[17] Ching T *et al* 2018 Opportunities and obstacles for deep learning in biolny and medicine *J. R. Soc. Interface* **15** 20170387
[18] Cocco S, Leibler S and Monasson R 2009 Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods *Proc. Natl Acad. Sci.* **106** 14058
[19] Coolen A C C, Kühn R and Sollich P 2005 *Theory of Neural Information Processing Systems* (Oxford: Oxford University Press)
[20] Decelle A, Hwang S, Rocchi J and Tantari D 2019 Inverse problems for structured datasets using parallel TAP equations and RBM (arXiv:1906.11988)
[21] Elad M 2012 Sparse and redundant representation modeling: what next? *IEEE Signal Process. Lett.* **19** 12
[22] Fachechi A, Agliari E and Barra A 2018 Dreaming neural networks: forgetting spurious memories and reinforcing pure ones *Neural Netw.* **112** 24
[23] Gabriè M *et al* 2018 Entropy and mutual information in models of deep neural networks *Neural Information Processing Systems (Conf. Montreal)*
[24] Gardner E 1988 The space of interactions in neural network models *J. Phys. A: Math. Gen.* **21** 257
[25] Gardner E 1985 Spin glasses with P-spin interactions *Nucl. Phys.* B **257** 747
[26] Genovese G 2012 Universality in bipartite mean field spin glasses *J. Math. Phys.* **53** 123304
[27] Goodfellow I, Bengio Y and Courville A 2017 *Deep Learning* (Cambridge, MA: MIT Press)
[28] Hinton G E and Salakhutdinov R R 2006 Reducing the dimensionality of data with neural networks *Science* **313** 504
[29] Hosny A *et al* 2018 Artificial intelligence in radiolny *Nat. Rev. Cancer* **18** 500
[30] Hopfield J J 1982 Neural networks and physical systems with emergent collective computational abilities *Proc. Natl Acad. Sci.* **79** 2554–8
[31] Huang H and Kabashima Y 2014 Origin of the computational hardness for learning with binary synapses *Phys. Rev.* E **90** 052813
[32] Krzakala F *et al* 2012 Statistical-physics-based reconstruction in compressed sensing *Phys. Rev.* X **2** 021005
[33] Kim K H and Kim S J 2000 Neural spike sorting under nearly 0 dB signal-to-noise ration using non-linear energy operator and artificial neural network classifier *IEEE Trans. Biomed. Eng.* **47** 10
[34] Kirk D 2007 NVIDIA CUDA software and GPU parallel computing architecture *ISMM.7* **7** 103–4
[35] Krotov D and Hopfield J J 2016 Dense associative memory for pattern recognition *Adv. Neur. Inf. Process. Syst.* 1172–80
[36] Krotov D and Hopfield J J 2018 Dense associative memory is robust to adversarial inputs *Neur. Comput.* **30** 3151–67
[37] Lehnert J and Pursley M B 1987 Error probabilities for binary direct-sequence spread-spectrum communications with random signature sequences *IEEE Trans. Commun.* **35** 87

[38] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436
[39] Lewicki M S and Sejnowski T J 2000 Learning overcomplete representations *Neural Comput.* **12** 337
[40] Li R B, Fazio D and Zeira A 2002 A low bias algorithm to estimate negative SNRs in an AWGN channel *IEEE Commun. Lett.* **6** 469
[41] Martin O C, Monasson R and Zecchina R 2001 Statistical mechanics methods and phase transitions in optimization problems *Theor. Comput. Sci.* **265** 3
[42] Mehta P and Schwab D J 2014 An exact mapping between the variational renormalization group and deep learning (arXiv:1410.3831)
[43] Mezard M 2017 Mean-field message-passing equations in the Hopfield model and its generalizations *Phys. Rev.* E **95** 022117
[44] Nickolls J *et al* 2008 Scalable parallel programming with CUDA *Queue* **6** 40
[45] Salakhutdinov R and Hinton G 2009 Deep Boltzmann machines *Artif. Intell. Stat.* 448–55
[46] Schmidhuber J 2015 Deep learning in neural networks: an overview *Neural Netw.* **61** 85
[47] Sejnowski T J 1986 Higher-order Boltzmann machines *AIP Conf. Proc.* **151** 398–403
[48] Tandra A and Sahai R 2008 SNR walls for signal detection *IEEE Select. Topics Sign. Proc.* **2** 4–17
[49] Xie J, Xu L and Chen E 2012 Image denoising and inpainting with deep neural networks *Adv. Neur. Inf. Process. Syst.* 341–9
[50] Zbilut J P, Giuliani A and Webber C L Jr 1998 Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification *Phys. Lett.* A **246** 122