Neural Networks Retrieving Boolean Patterns in a Sea of Gaussian Ones

Elena Agliari, Adriano Barra, Chiara Longo & Daniele Tantari

Journal of Statistical Physics
1

ISSN 0022-4715

J Stat Phys
DOI 10.1007/s10955-017-1840-9

Volume 122 • Number 1 • January

ONLINE FIRST

Journal of Statistical Physics

Available online
www.springerlink.com

10955 • ISSN 0022-4715
122(1) 1–196 (2006)

Springer

Springer

J Stat Phys
DOI 10.1007/s10955-017-1840-9

Neural Networks Retrieving Boolean Patterns in a Sea of Gaussian Ones

Elena Agliari[1,2] · **Adriano Barra**[2,3,4] · **Chiara Longo**[1] · **Daniele Tantari**[2,5]

Received: 15 March 2017 / Accepted: 6 July 2017

Abstract Restricted Boltzmann machines are key tools in machine learning and are described by the energy function of bipartite spin-glasses. From a statistical mechanical perspective, they share the same Gibbs measure of Hopfield networks for associative memory. In this equivalence, weights in the former play as patterns in the latter. As Boltzmann machines usually require real weights to be trained with gradient-descent-like methods, while Hopfield networks typically store binary patterns to be able to retrieve, the investigation of a *mixed* Hebbian network, equipped with both real (e.g., Gaussian) and discrete (e.g., Boolean) patterns naturally arises. We prove that, in the challenging regime of a high storage of real patterns, where retrieval is forbidden, an additional load of Boolean patterns can still be retrieved, as long as the ratio between the overall load and the network size does not exceed a critical threshold, that turns out to be the same of the standard Amit–Gutfreund–Sompolinsky theory. Assuming replica symmetry, we study the case of a low load of Boolean patterns combining the stochastic stability and Hamilton-Jacobi interpolating techniques. The result can be extended to the high load by a non rigorous but standard replica computation argument.

Keywords Neural networks · Hopfield model · Boltzmann machine

1 Introduction

In recent years we have witnessed a formidably fast development of research in Artificial Intelligence. Neural networks are playing an important role in this trend, mainly due

✉ Elena Agliari
agliari@mat.uniroma1.it

[1] Dipartimento di Matematica, Sapienza Università di Roma, Rome, Italy

[2] Istituto Nazionale di Alta Matematica (GNFM-INdAM), Rome, Italy

[3] Dipartimento di Matematica e Fisica “Ennio De Giorgi”, Università del Salento, Lecce, Italy

[4] Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Lecce, Lecce, Italy

[5] Scuola Normale Superiore, Centro “Ennio De Giorgi”, Pisa, Italy

Published online: 17 July 2017

to the ability of the so-called deep networks to solve difficult problems, after a proper training. Such problems are broadly ranged in sciences (from Particle Physics [9] to Computational Biology [46]), not to mention the applied world of technology, where their usage has become pervasive. Nevertheless, as admitted in [45], despite its remarkable successes, nobody yet understands exhaustively how the whole scaffold works, while there is wide agreement that achieving a full understanding of Deep Learning is an urgent priority.

The pivotal constituent of Deep Learning is the Restricted Boltzmann Machine (RBM) [37,39,44,49]. This is a network of units with a bipartite structure, the two parties being referred to as *visible layer* and *hidden layer*; units belonging to different layers are connected by links endowed with *weights* while nodes belonging to the same layer are not connected (see Fig. 1, left panel). In the jargon of statistical physicists, RBMs have the same energy of a bipartite spin-glass [8,15,21–23]. By marginalization over the hidden layer, RBMs have also been shown to share the same phase diagram of an Hopfield network [1,17,47,53,56], whose neurons, corresponding to the units of the visible layer (see Fig. 1 right panel), are connected each other via Hebbian couplings [38] and the number of stored patterns corresponds to the amount of hidden units. The Hopfield network is able to spontaneously retrieve such patterns, and therefore to work as an associative memory [5,44], as long as the ratio between the patterns to handle and the available neurons is not too large [6], or, in the dual perspective of the RBMs, as long as the size of the hidden layer is not too large compared to the size of the visible layer.

Crucially, the *weight vectors* learnt by the RBM after training play as *patterns* in Hopfield retrieval. Since RBMs typically work with real weight vectors, while standard Hopfield networks are built with Boolean patterns, studies on possible generalizations are in order and they are beginning to appear in the literature [11,56].

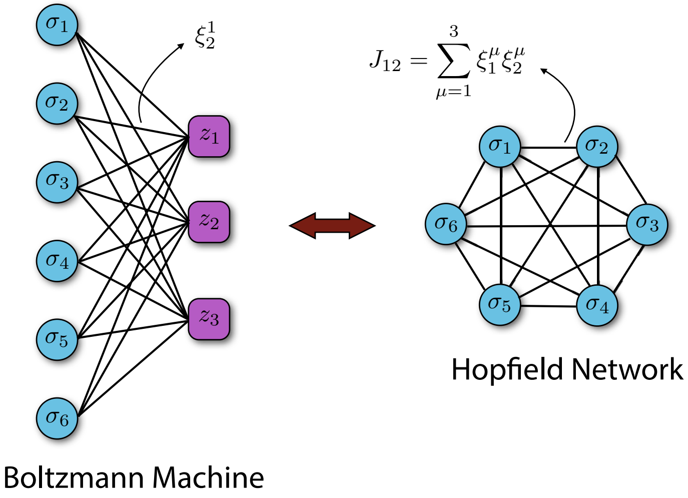

Fig. 1 Left: example of a RBM equipped with 6 visible neurons σ_i, $i \in (1, ..., 6)$ and 3 hidden units z_μ, $\mu \in (1, ..., 3)$. The weights connecting them form the $N \times P$ matrix ξ_i^μ. Right: example of the corresponding AHN, whose six visible neurons σ_i, $i \in (1, \ldots, 6)$ retrieve as patterns stored in the Hebb matrix $J_{ij} = \sum_\mu^P \xi_i^\mu \xi_j^\mu$ the three vectors ξ^μ, $\mu \in (1, ..., 3)$, each pertaining to a *feature*, i.e. one of the three z_μ hidden variables of the (corresponding) RBM

In fact, in the last years, an increasing number of semi-heuristic routes toward a rationale for Deep Learning have been introduced, while rigorous answers (e.g., avoiding the usage of the so called replica-trick [32,48,50]) to specific questions are hardly distilled (see e.g., [14,16,24,26–28,30,51,52,54,55]). However, beyond the replica-trick, other techniques (from cavity or message passing [41–43,47,53,58] to those based on interpolating structures [3,15,16,23]) to handle spin-glasses have recently appeared in the literature, hence an attempt should be made in using them to enlarge our knowledge on RBMs and generalized Hopfield networks also from a rigorous perspective. Here we aim to contribute to this goal by adopting interpolating techniques under the replica-symmetry assumption. The latter means that we tacitly impose that the order parameters do not fluctuate around their means in the thermodynamic limit [32,48]. This does not prove in any way that the replica symmetric scenario we are going to paint is rigorously correct, but solely that, if replica symmetry holds, the behavior and the properties of the system will be those inferred in this work. An alternative route (mathematically completely different and in spirit somehow complementary to the present one), aimed to prove the existence of regions where replica symmetry is preserved in neural networks, has been paved by Bovier and Gayrard [26,28] and by Talagrand [54,55].

In the following, we investigate a Hopfield networks endowed with patterns that are *mixed*, namely in part binary and in part real, by combining two mathematical approaches, i.e., stochastic stability [4,14,16,20,31] and Hamilton-Jacobi interpolation [2,13,19,35,36]. In this way we are able to describe the model free-energy and its phase diagram for pure state retrieval and we prove, at the replica symmetric level, that these mixed Hopfield networks are robustly capable of retrieving the digital information (i.e., the binary patterns) although "immersed" in the continuous (slow) noise generated by the real patterns (i.e., the *sea*).

More precisely, let us consider a system made of N Ising neurons dealing with a certain number of patterns, referred to as p or k according to whether the number scales linearly with N (i.e., $p = \alpha N$) or logarithmically with N ($k = \gamma \ln N$). These two cases correspond to the so-called *high storage* and *low storage* regimes, respectively [7]. As well known, in the low-storage regime the Hopfield model is able to retrieve patterns (i.e., to work as a distributed associative memory) for binary as well as real patterns [10,11], while, in the high-storage regime, only binary patterns can be retrieved because a linearly extensive (in N) amount of real patterns contains too much information for the $O(N^2)$ synaptic couplings to perform pattern recognition or similar tasks [10,25]. Indeed, in general, the high-storage case is much more tricky due to its intrinsic glassiness, whence tools from disordered statistical mechanics are in order to infer its properties [7,48]. On the contrary, standard statistical mechanical machineries are usually effective for the low-storage case [32]. For mixed Hebbian networks (where patterns are in part analog and in part digital) a first scenario we would figure out and clarify is their retrieval capabilities when they are constrained to keep an extensive amount of p real patterns (hence the worst case for retrieval) but they are also over-fed by a further low-load of k binary patterns. Exploiting Guerra's interpolating schemes we prove that there exists a region in the parameter space (corresponding to not-too-high values of both fast and slow noises), where mixed Hebbian network works as a distributed associative memory and the boundaries of such a region are evidenced by a first-order phase transition. Further, a fairly standard replica calculation, although not rigorous, suggests that this picture can be extended even to the case of an extensive load for both binary and real patterns, that is, there exists a retrieval region where pattern recognition for high-load digital information in a real sea seems possible. Remarkably, in all these cases, the boundary for the retrieval region turns out to be always the one identified by Amit-Gutfreund-Sompolinsky (AGS) in the 80's [5,6].

1.1 Associative Hopfield Networks and Restricted Boltzmann Machines

Let us deepen the ideas exposed so far, by introducing the standard definitions and concepts for associative Hopfield networks (AHN) and RBMs. Following classical notations [7,32], we shall consider N binary neurons (i.e., Ising spins [7]) and to each neuron i we assign a dichotomic variable σ_i that describes its activity: if $\sigma_i = +1$ the i-th neuron is spiking, while if $\sigma_i = -1$ it is quiescent.

Neurons are embedded in a fully connected network, in such a way that *mean-field* approaches are suitable for the investigation. The synaptic potential h_i that the i-th neuron receives from the other $N-1$ is defined as

$$h_i = \sum_{j \neq i}^{N} J_{ij} \sigma_j,$$

where $J_{ij} = J_{ji}$ is the synaptic coupling between neuron j and neuron i, defined according to Hebb's learning rule [38] as

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu}. \tag{1}$$

Indeed, associative memory models are built to recognize a certain group of words, pixels, or, generically, patterns: a *pattern* ξ is defined as a sequence of random variables $\xi = (\xi_1, \ldots, \xi_N)$. If we want the network to memorize and retrieve a number p of patterns, we have to introduce another index to distinguish them: $\{\xi^1, \ldots, \xi^p\}$, and we shall assume that the set $\left\{\xi_i^{\mu}\right\}_{i,\mu}$ is made of $p \times N$ i.i.d. variables. Notice that, for a Shannon information compression argument, if the network is able to cope with this kind of patterns, then it certainly retains at least the same capacity in the case of correlated patterns [2,33]. Boolean binary patterns have entries such that $\mathbb{P}(\xi_i = +1) = \mathbb{P}(\xi_i = -1) = 1/2$, while Gaussian real patterns have entries drawn from $\mathbb{P}(\xi_i) \sim \mathcal{N}(0, 1)$.

The Hamiltonian $H_N^{AHN}(\sigma, \xi)$ of the AHN equipped with N Ising neurons σ and p patterns is defined as

$$H_N^{AHN}(\sigma, \xi) = -\frac{1}{2N} \sum_{i,j}^{N} \sum_{\mu=1}^{p} \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j. \tag{2}$$

Once introduced the (fast) noise $\beta = 1/T \in \mathbb{R}^+$, where T plays as a *temperature* in standard statistical mechanics, the partition function $Z_{N,p}^{AHN}(\beta)$ for the AHN is defined as

$$Z_{N,p}^{AHN}(\beta) = \sum_{\sigma} \exp\left\{ \frac{\beta}{2N} \sum_{\mu=1}^{p} \sum_{i,j}^{N} \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j \right\},$$

and the free energy as $N^{-1} \mathbb{E}_{\xi} \log Z_{N,p}^{AHN}(\beta)$. The analysis of the latter allows inferring the model phase-diagram in the thermodynamic limit ($N \to \infty$) [7,32]. Note that in the previous definitions we have introduced for simplicity also self-interactions, but we will see that their presence does not affect the thermodynamic state of the network because they contribute at most to a simple constant term in the free energy.

The Hamiltonian $H_N^{RBM}(\sigma, \xi)$ of the RBM, equipped with a visible layer of N binary (i.e. Boolean) units σ_i, $i \in (1, ..., N)$ and a hidden layer of p real (i.e. Gaussian) units z_{μ}, $\mu \in (1, ..., p)$, connected by the $N \times p$ weight matrix ξ_i^{μ}, is defined as

$$H_N^{RBM}(\sigma, \xi) = -\frac{1}{\sqrt{N}} \sum_{i,\mu}^{N,p} \xi_i^\mu \sigma_i z_\mu. \tag{3}$$

Again, considering β the fast noise of the network, the partition function $Z_{N,p}^{RBM}(\beta)$ for the RBM is introduced as

$$Z_{N,p}^{RBM}(\beta) = \sum_\sigma \int_{\mathbb{R}^p} d\mathcal{M}(z) \exp\left\{ \sqrt{\frac{\beta}{N}} \sum_{\mu=1}^{p} \sum_{i=1}^{N} \xi_i^\mu \sigma_i z_\mu \right\},$$

where $d\mathcal{M}(z) = \prod_{\mu=1}^{p} \frac{dz_\mu}{\sqrt{2\pi}} e^{z_\mu^2/2}$ is the p-dimensional centered Gaussian measure. The free-energy of the model is defined as before. It is just an exercise now to show (e.g., via standard Gaussian integration) that the partition functions of the AHN and of the RBM are the same, i.e.

$$Z_{N,p}^{AHN}(\beta) \equiv Z_{N,p}^{RBM}(\beta),$$

and thus the same equivalence holds for the two free energies as well.

Note that, while the identity $Z_N^{AHN}(\beta) \equiv Z_N^{RBM}(\beta)$ strictly holds only if we choose Gaussian hidden units z_μ, an analogous equivalence can be proved introducing a class of generalised AHN and RBM models with any unit priors [10,11].

Remark 1 Starting from a Master equation (see Eq. 4) for the evolution of the system, where $p_t(\sigma)$ denotes the probability of finding the network in the state σ at time t and $W(\sigma, \sigma')$ denotes the transition rate from the state σ' to the state σ, we notice that for symmetric couplings, i.e. $J_{ij} = J_{ji}$, the *detailed balance* (see Eq. 5) holds [7,32]. Consequently, any (non-pathological) network dynamics converges to the Gibbs measure of the Hamiltonian H (see Eq. 6), namely

$$p_{t+1}(\sigma) = \sum_{\sigma'} W(\sigma, \sigma') p_t(\sigma'), \tag{4}$$

$$W(\sigma, \sigma') p_\infty(\sigma') = W(\sigma', \sigma) p_\infty(\sigma), \tag{5}$$

$$p_\infty(\sigma) = Z^{-1} \exp(-\beta H). \tag{6}$$

Hence, if the Hamiltonian H displays the stored patterns as ground states (i.e., the ground states correspond to configurations $\boldsymbol{\sigma} = \boldsymbol{\xi}^\mu$, $\forall \mu = 1, ..., P$), and if the noise affecting the network is not too loud, any relaxation dynamics started within the basin of attraction of any of these minima should converge to it, tacitely coding for *retrieval capabilities*.[1] In particular, in [17] the joint dynamics of a Boltzmann machine coupled with its dual Hopfield representation is shown to respect the scheme above and such a dynamics holds for this mixed network too.

More in details, in order to investigate the capabilities of these networks to retrieve patterns, it is useful to introduce the concept of Mattis magnetization as follows: for any $\mu \in (1, ..., p)$, we define the Mattis magnetization, i.e. the overlap between the μ-th pattern and the neuron states, as

$$m_{\mu,N}(\sigma) = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \sigma_i, \tag{7}$$

[1] Actually this ideal scenario is an oversimplification due to the spontaneous formation of spurious and metastable states in the free energy landscape, but we remind to dedicated textbooks [7,32] or articles [24,28] as dynamical convergence to Gibbs equilibrium is not the focus of the present work.

and, in the following, if no confusion emerges, we drop the N or σ dependencies to lighten the notation. The magnitude of the Mattis magnetization m_μ encodes whether the pattern μ has been retrieved or not. Moreover we can rewrite the Hamiltonian (2) as a function of the order parameters m_μ's as

$$H_N^{AHN}(\sigma,\xi) = -\frac{N}{2}\sum_{\mu=1}^{p} m_\mu^2,$$

hence it becomes clear that its energy minima are located at large m_μ for some μ. This means that the energy function is minimized as the spins are aligned to some of the p patterns, thus indicating a retrieval state (i.e. the network overall works as a distributed associative memory).

Let us now turn our attention to the RBM case. Its energy function (3) can be rewritten as well in terms of Mattis magnetizations as

$$H_N^{RBM}(\sigma,\xi) = -\sqrt{N}\sum_{\mu=1}^{p} m_\mu z_\mu,$$

in such a way that, if the system is in the retrieval region, i.e., there is some pattern μ (say μ^*) that is retrieved by the dual Hopfield network, then its related Mattis magnetization m_{μ^*} raises from zero and acts as a *staggered magnetic field* over its related hidden variable z_{μ^*}. In the machine learning jargon, this condition corresponds to selecting a *feature*, among the p possible, and allows a statistically significant classification of the data [10,37,49].

2 Mixed Hebbian Networks

In our "hybrid" Hopfield model, we consider the case in which the network has stored a low load of Boolean patterns and a high load of Gaussian ones. We will assign the variables $\tilde{\xi}^\nu$, $\nu = 1,\dots,k = \gamma \ln N$ to the binary memories and ξ^μ, $\mu = 1,\dots,p = \alpha N$ to the real ones (with $\gamma,\alpha > 0$). We have

$$\begin{cases} \mathbb{P}\{\tilde{\xi}_i^\nu = +1\} = \mathbb{P}\{\tilde{\xi}_i^\nu = -1\} = \frac{1}{2} & \forall i = 1,\dots,N \text{ and } \nu = 1,\dots,k, \\ \mathbb{P}(\xi_i^\mu) \sim \mathcal{N}(0,1) & \forall i = 1,\dots,N \text{ and } \mu = 1,\dots,p. \end{cases}$$

Following the description of the standard Hopfield neural network given in Sect. 1.1, we give the following

Definition 1 The Hamiltonian $H_N^{MHN}(\sigma,\xi,\tilde{\xi})$ of the mixed Hebbian network (MHN), equipped with N Ising neurons, a low load of k binary patterns and a high load of p real patterns, reads as

$$H_N^{MHN}(\sigma,\xi,\tilde{\xi}) = -\frac{1}{N}\sum_{1\le i<j\le N}\left(\sum_{\nu=1}^{k}\tilde{\xi}_i^\nu\tilde{\xi}_j^\nu + \sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu\right)\sigma_i\sigma_j. \tag{8}$$

Notice that, splitting the above summations over (i, j), the Hamiltonian of the mixed Hebbian network can be written as

$$H_N(\sigma,\xi,\tilde{\xi}) = -\frac{1}{2N}\sum_{i,j=1}^{N}\left(\sum_{\nu=1}^{k}\tilde{\xi}_i^\nu\tilde{\xi}_j^\nu + \sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu\right)\sigma_i\sigma_j + \frac{1}{2N}\sum_{i=1}^{N}\sum_{\mu=1}^{p}\left(\xi_i^\mu\right)^2 + \frac{k}{2}, \tag{9}$$

hence the last term at the r.h.s. of the previous equation does not contribute at all in the thermodynamic limit, while the second-last term converges to

$$\lim_{N\to\infty}\left[\frac{1}{2N}\sum_{i=1}^{N}\sum_{\mu=1}^{p}\left(\xi_i^\mu\right)^2\right]=\frac{\alpha}{2}.$$

The Gibbs measure for a generic function of the neurons $F(\sigma)$ at a given level of noise β is

$$\omega_N(F)=\frac{\sum_\sigma F(\sigma)e^{-\beta H_N(\sigma,\xi,\tilde{\xi})}}{Z_N(\beta)}, \tag{10}$$

and, once given s independent realizations (i.e., *replicas*) of the system, at the same level of noise β, and quenched patterns ξ and $\tilde{\xi}$, we define the s-replicated Gibbs measure as $\Omega=\omega^1\times\omega^2\times\ldots\times\omega^s$, i.e. for any function of the s neuron replicas $F(\sigma^{(1)},\ldots,\sigma^{(s)})$,

$$\begin{aligned}&\Omega\Big(\left(\sigma^{(1)},\ldots,\sigma^{(s)}\right)\Big)\\&\quad=\tfrac{1}{Z_N^s}\textstyle\sum_{\sigma^{(1)}}\cdots\sum_{\sigma^{(s)}}F\left(\sigma^{(1)},\ldots,\sigma^{(s)}\right)\exp\left\{-\beta\sum_{a=1}^{s}H_N\left(\sigma^{(a)},\xi,\tilde{\xi}\right)\right\}.\end{aligned} \tag{11}$$

Finally, the average over the quenched memories $\{\tilde{\xi}_i^\nu\}_{i,\nu}$ and $\{\xi_i^\mu\}_{i,\mu}$ for a generic function $F(\xi,\tilde{\xi})$ is introduced as

$$\mathbb{E}\left[F(\xi,\tilde{\xi})\right]=\int\prod_{\mu=1}^{p}\prod_{i=1}^{N}\frac{d\xi_i^\mu}{\sqrt{2\pi}}e^{-\frac{(\xi_i^\mu)^2}{2}}\times\prod_{\nu=1}^{k}\prod_{j=1}^{N}\sum_{\{\tilde{\xi}_j^\nu\}}\frac{1}{2}F(\xi,\tilde{\xi}),$$

and, overall, we define the average $\langle\cdot\rangle=\mathbb{E}\Omega(\cdot)$.

We continue by introducing the order parameters necessary to carry out the analysis of the mixed model. For any pattern, we define the Mattis magnetization as before, and we further introduce overlaps among replicas, as in [14,16], as follows: Given two replicas (a,b) of the network, the overlap q_{ab} between visible units is defined as

$$q_{ab}(\boldsymbol{\sigma})=\frac{1}{N}\sum_{i=1}^{N}\sigma_i^{(a)}\sigma_i^{(b)}\in[-1,1], \tag{12}$$

and the overlap p_{ab} between hidden units as

$$p_{ab}(\boldsymbol{z})=\frac{1}{p}\sum_{\mu=1}^{p}z_\mu^{(a)}z_\mu^{(b)}\in(-\infty,+\infty). \tag{13}$$

Finally, we introduce the free-energy density $A(\alpha,\beta)$ of the mixed Hebbian network as

$$A(\alpha,\beta)=\lim_{N\to\infty}A_{N,k,p}(\beta),\quad A_{N,k,p}(\beta)=\frac{1}{N}\mathbb{E}\ln Z_{N,k,p}(\beta), \tag{14}$$

where the partition function $Z_{N,k,p}(\beta)$ reads as

$$\begin{aligned}Z_{N,k,p}(\beta)&=\sum_\sigma\exp\left\{-\beta H_N^{MHN}(\sigma,\xi,\tilde{\xi})\right\}\\&=\sum_\sigma\exp\left\{\frac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\nu=1}^{k}\tilde{\xi}_i^\nu\tilde{\xi}_j^\nu\sigma_i\sigma_j+\frac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu\sigma_i\sigma_j\right\}.\end{aligned} \tag{15}$$

Therefore, the free-energy density at finite volume reads as

$$\begin{aligned}A_{N,k,p}(\beta) &= \tfrac{1}{N}\mathbb{E}\log Z_{N,k,p}(\beta) = \tfrac{1}{N}\mathbb{E}\Bigg[-\tfrac{\beta k}{2} - \tfrac{\beta}{2N}\textstyle\sum_{i=1}^{N}\sum_{\mu=1}^{p}(\xi_i^{\mu})^2\Bigg]\\ &\quad+\tfrac{1}{N}\mathbb{E}\log\Bigg(\textstyle\sum_{\sigma}\exp\Bigg\{\tfrac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\nu=1}^{k}\tilde{\xi}_i^{\nu}\tilde{\xi}_j^{\nu}\sigma_i\sigma_j + \tfrac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\mu=1}^{p-1}\xi_i^{\mu}\xi_j^{\mu}\sigma_i\sigma_j\Bigg\}\Bigg)\\ &= -O\left(\tfrac{\ln N}{N}\right) - \tfrac{\alpha_N\beta}{2} +\\ &\quad+\tfrac{1}{N}\mathbb{E}\ln\Bigg(\textstyle\sum_{\sigma}\exp\Bigg\{\tfrac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\nu=1}^{k}\tilde{\xi}_i^{\nu}\tilde{\xi}_j^{\nu}\sigma_i\sigma_j + \tfrac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\mu=1}^{p}\xi_i^{\mu}\xi_j^{\mu}\sigma_i\sigma_j\Bigg\}\Bigg),\end{aligned} \tag{16}$$

where the parameter α_N is such that $\alpha_N = \frac{p}{N} \to \alpha$ for $N \to \infty$.

We recall that, in the statistical mechanical treatment, finding an explicit expression for the free-energy density $A(\alpha, \beta)$ in terms of its order parameters m_μ, q_{ab}, p_{ab} is the first step for understanding the properties of the thermodynamic states of the system. This is because the solution of $A(\alpha, \beta)$ usually comes with a variational large deviation principle over the order parameters $\{m_\mu,\ q_{ab},\ p_{ab}\}$ [7,48,57] whose analysis allows inferring a phase diagram of the system behavior.

3 Sum Rules for the Mixed Hebbian Network's Free Energy

In this Section we explain and use the interpolating structure that we set up to obtain an expression for the free-energy density of the MHN, at the replica symmetric level,[2] as a variational principle over the order parameters. The solution of this optimization problem is encoded into a set of self-consistent equations that the order parameters have to satisfy, giving the phase diagram of the model by varying the tuneable parameters.

In particular, the question we are addressing in the present work is about the existence of a retrieval phase in such a phase diagram: we will prove that there is actually a region in the (α, β) plane where the NHM is able to retrieve, in particular where the signal conveyed by the binary patterns is detectable over the real noisy sea.

In a nutshell, we will adopt a combination of stochastic stability [4] and Hamilton-Jacobi [12,13] techniques: in this section we will show all the details regarding how to proceed by applying the former first and then the latter, while in the next section, we will briefly summarize the other route (starting with Hamilton-Jacobi and concluding with stochastic stability).

As a preliminary step, it is useful to apply the Gaussian integration to the partition function (15) to linearize the Gaussian section of the free energy density function $A_{N,k,p}(\beta)$ with respect to the bilinear quenched memories carried by $\xi_i^{\mu}\xi_j^{\mu}$, namely:

$$\begin{aligned}Z_{N,k,p}(\beta) &= \exp\Bigg\{-\tfrac{\beta k}{2} + \tfrac{\beta}{2N}\textstyle\sum_{i=1}^{N}\sum_{\mu=1}^{p}(\xi_i^{\mu})^2\Bigg\}\\ &\quad\times\textstyle\sum_{\sigma}\exp\left\{\tfrac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\nu=1}^{k}\tilde{\xi}_i^{\nu}\tilde{\xi}_j^{\nu}\sigma_i\sigma_j + \tfrac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\mu=1}^{p}\xi_i^{\mu}\xi_j^{\mu}\sigma_i\sigma_j\right\}\\ &= \exp\Bigg\{-\tfrac{\beta k}{2} + \tfrac{\beta}{2N}\textstyle\sum_{i=1}^{N}\sum_{\mu=1}^{p}(\xi_i^{\mu})^2\Bigg\}\textstyle\sum_{\sigma}\exp\left\{\tfrac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\nu=1}^{k}\tilde{\xi}_i^{\nu}\tilde{\xi}_j^{\nu}\sigma_i\sigma_j\right\}\\ &\quad\times\textstyle\int_{\mathbb{R}^p} d\mathcal{M}(z)\exp\left\{\sqrt{\tfrac{\beta}{N}}\sum_{\mu=1}^{p}\sum_{i=1}^{N}\xi_i^{\mu}\sigma_i z_\mu\right\},\end{aligned}$$

where $d\mathcal{M}(z) = \prod_{\mu=1}^{p}\frac{dz_\mu}{\sqrt{2\pi}}e^{z_\mu^2/2}$ is the p-dimensional Gaussian measure.

[2] We emphasize that the replica symmetric level of approximation is the standard one in the whole branch of Neural Networks [7,32] and, in a nutshell, it consists in preventing the order parameters to fluctuate around their means, i.e. they are self-averaging.

As anticipated earlier, to achieve our goal we shall now analyse a generalized problem, for which we give hereafter the definition in terms of its partition function: Once introduced $k+2$ scalar parameters $t \in \mathbb{R}^+$, $x \in \mathbb{R}^k$, $\psi \in [0,1]$, and three scalar fields $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, the generalized partition function $Z_N(t,x,\psi)$ for the MHN is defined as

$$\begin{aligned} Z_N\ (t,x,\psi) &= \exp\Bigg\{-\tfrac{\beta k}{2} - \tfrac{\beta}{2N}\textstyle\sum_{i=1}^N\sum_{\mu=1}^p(\xi_i^\mu)^2\Bigg\} \\ &\times \textstyle\sum_\sigma \displaystyle\int_{\mathbb{R}^p} d\mathcal{M}(z)\ \exp\Bigg\{\tfrac{t}{2N}\textstyle\sum_{i,j=1}^N\sum_{\nu=1}^k \tilde{\xi}_i^\nu\tilde{\xi}_j^\nu\sigma_i\sigma_j + \sum_{\nu=1}^k x_\nu \sum_{i=1}^N \tilde{\xi}_i^\nu\sigma_i\Bigg\} \\ &\times \exp\Bigg\{\sqrt{\psi}\sqrt{\tfrac{\beta}{N}}\textstyle\sum_{\mu=1}^p\sum_{i=1}^N \xi_i^\mu\sigma_i z_\mu\Bigg\} \times \exp\Bigg\{\mathcal{A}\sqrt{1-\psi}\textstyle\sum_{i=1}^N\eta_i\sigma_i\Bigg\} \\ &\times \exp\Bigg\{\mathcal{B}\sqrt{1-\psi}\textstyle\sum_{\mu=1}^p\theta_\mu z_\mu\Bigg\} \times \exp\Bigg\{\mathcal{C}\tfrac{1-\psi}{2}\textstyle\sum_{\mu=1}^p(z_\mu)^2\Bigg\}, \end{aligned} \tag{17}$$

with $\theta_\mu, \eta_i \sim \mathcal{N}(0,1)\ \forall \mu = 1,\ldots,p,\ i = 1,\ldots,N$.

Note that, by now, the scalar fields are given in full generality and they will be chosen later on, in order to ensure that the replica symmetric assumption is preserved at the end of the interpolation.

Note further that we can extend also the free energy density function to $A_{N,k,p}(t,x,\psi)$, the Gibbs measures to $\omega_{t,x,\psi}$ and $\Omega_{t,x,\psi}$ and the overall average to $\langle\cdot\rangle_{t,x,\psi}$. Of course, also these quantities recover the standard statistical mechanical scenario once evaluated at $t=\beta$, $x=0$ and $\psi=1$.

We begin the study of the free energy density function through the stochastic stability. First, exploiting the Fundamental Theorem of Calculus on $A_{N,k,p}(t,x,\psi)$ in the ψ variable, we write the following sum rule for the generalised free energy $A_{N,k,p}(t,x,\psi)$ of the MHN

$$\begin{aligned} A_{N,k,p}(t,x) &= A_{N,k,p}(t,x,\psi=1) \\ &= A_{N,k,p}(t,x,\psi=0) + \int_0^1 \left(d_{\psi'}A_{N,k,p}(t,x,\psi')\right)_{\psi'=\psi} d\psi. \end{aligned} \tag{18}$$

The original problem is therefore recast in the evaluation of the two terms at the r.h.s. of Eq. (18).

To compute the first term we start through a standard Gaussian integration, hence

$$\begin{aligned} A_{N,k,p}\ (t,x,\psi=0) &= -O\left(\tfrac{\ln N}{N}\right) - \tfrac{\alpha_N\beta}{2} \\ &+ \tfrac{1}{N}\mathbb{E}\Bigg[\log\textstyle\sum_\sigma \exp\Bigg\{\tfrac{t}{2N}\sum_{i,j=1}^N\sum_{\nu=1}^k\tilde{\xi}_i^\nu\tilde{\xi}_j^\nu\sigma_i\sigma_j \\ &+ \textstyle\sum_{\nu=1}^k x_\nu\sum_{i=1}^N\tilde{\xi}_i^\nu\sigma_i + \mathcal{A}\sum_{i=1}^N\eta_i\sigma_i\Bigg\} \\ &\times \int_{\mathbb{R}^p}\frac{dz_1\cdots dz_p}{(2\pi)^{p/2}}\exp\Bigg\{\sum_{\mu=1}^p\left(\mathcal{B}\theta_\mu z_\mu + \frac{\mathcal{C}-1}{2}z_\mu^2\right)\Bigg\}\Bigg] \\ &= -O\left(\tfrac{\ln N}{N}\right) - \tfrac{\alpha_N\beta}{2} + \tfrac{1}{N}\mathbb{E}\ln\left(\tfrac{1}{(1-\mathcal{C})^{p/2}}e^{\frac{\mathcal{B}^2\theta^2}{2(1-\mathcal{C})}p}\right) \\ &+ \tfrac{1}{N}\mathbb{E}\ln\textstyle\sum_\sigma\exp\Bigg\{\tfrac{t}{2N}\sum_{i,j=1}^N\sum_{\nu=1}^k\tilde{\xi}_i^\nu\tilde{\xi}_j^\nu\sigma_i\sigma_j \\ &+ \textstyle\sum_{\nu=1}^k x_\nu\sum_{i=1}^N\tilde{\xi}_i^\nu\sigma_i + \mathcal{A}\sum_{i=1}^N\eta_i\sigma_i\Bigg\}. \end{aligned} \tag{19}$$

It is now crucial to notice that the term in the last line of the previous Eq. (19) can be interpreted as the free energy density $\tilde{A}_{N,k}(t,x)$ of a Hopfield network with k binary patterns $\{\tilde{\xi}^\nu\}$ and

N external random fields $\mathcal{A}\eta_i$ which account for the slow noise supplied by the underlying *sea* of Gaussian patterns that can not be retrieved. The related generalized partition function $Z_{N,k}(t,x)$ is identified by the following expression

$$Z_{N,k}(t,x)=\sum_{\sigma}\exp\left\{\frac{tN}{2}\sum_{\nu=1}^{k}m_\nu^2+N\sum_{\nu=1}^{k}x_\nu m_\nu+\mathcal{A}\sum_{i=1}^{N}\eta_i\sigma_i\right\}.$$

and we can define the Guerra Action $\tilde{G}_{N,k}(t,x)$, for a unitary-mass point-particle moving in the $(1+k)$ dimensional (t,x) space, as the negative free energy density $\tilde{A}_N(t,x)$:

$$\tilde{G}_{N,k}(t,x)=-\tilde{A}_{N,k}(t,x)=-\frac{1}{N}\ln\tilde{Z}_N(t,x). \tag{20}$$

With this definition, the application of the Hamilton-Jacobi formalism for handling $\tilde{A}_{N,k}(t,x)$ is straightforward. In fact, one can check that $\tilde{A}_{N,k}(t,x)$ has the following properties

$$\partial_t\tilde{A}_{N,k}(t,x)=\frac{1}{2}\sum_{\nu=1}^{k}\langle m_\nu^2\rangle_{x,t},\qquad \partial_{x_\nu}\tilde{A}_{N,k}(t,x)=\langle m_\nu\rangle_{x,t}, \tag{21}$$

hence we can proceed according to the Hamilton-Jacobi prescription for $\tilde{G}_{N,k}(t,x)$. In fact, thanks to the Eq. (21), it is immediate to verify the next

Proposition 1 *The Guerra Action obeys the following Hamilton-Jacobi PDE*

$$\partial_t\big(\tilde{G}_{N,k}(t,x)\big)+\frac{1}{2}\big(\partial_x\tilde{G}_{N,k}(t,x)\big)^2+V_{N,k}(t,x)=0, \tag{22}$$

where the potential $V_{N,k}(t,x)$ is given by the sum over all the binary patterns of their related Mattis magnetization's variances, namely

$$V_{N,k}(t,x)=\frac{1}{2}\sum_{\nu}^{k}\big(\langle m_\nu^2\rangle_{t,x}-\langle m_\nu\rangle_{t,x}^2\big)=\frac{1}{2N}\partial_{xx}^2\tilde{G}_{N,k}(t,x).$$

Note that, as we are in the low-storage regime for binary patterns (i.e., $k\propto\ln N$), in the thermodynamic limit the Guerra Action paints a Galilean trajectory for the point-like particle: its evolution is simply a free motion as $\lim_{N\to\infty}V_{N,k}(t,x)=0$.[3] Hence, if we define a k-dimensional vector $\Gamma_N(t,x)$, whose components are $\Gamma_N^\nu(t,x)=\partial_{x_\nu}\tilde{G}_{N,k}(t,x)$, by deriving Eq. (22) with respect to x_ν we obtain the following set of k Burgers equations for the canonical momenta

$$\partial_t\Gamma_N^\nu(t,x)+\sum_{\tau=1}^{k}\Gamma_N^\tau(t,x)\times\partial_{x_\tau}\Gamma_N^\nu(t,x)=\frac{1}{2N}\sum_{\tau=1}^{k}\partial_{x_\tau x_\tau}^2\Gamma_N^\nu(t,x)\quad\forall\nu. \tag{23}$$

At present, the goal is thus to solve the Burgers equations and integrate back the solutions to get the original problem for $\tilde{G}_{N,k}(t,x)$ (and therefore for $\tilde{A}_N(t,x)$) solved too. As standard,

[3] Indeed, a standard signal-to-noise analysis applied to the slow noise built in by the not-retrieved patterns allows us to conclude that, as long as the amount of these patterns does not scale linearly with the number of neurons N, their interference with retrieval is negligible. In fact, suppose the network is in the basin of attraction of ξ^1: its equilibrium state, expected to be $\sigma=\xi^1$, is stable iff $\xi_i^1h_i>0$ for all $i\in(1,...,N)$, and it is immediate to check that $\xi_i^1h_i=N^{-1}\sum_{j\neq i}\sum_\mu\xi_i^\mu\xi_j^\mu\xi_i^1\xi_j^1=S+N$, where the signal $S=O(1)$, while the noise N is a zero average random walk whose variance grows as p/N, hence, as long as $\lim_{N\to\infty}p/N=0$, the signal will always be prevailing.

performing the Cole-Hopf transform $\Phi_{N,k}(t,x) := e^{N\tilde{A}_{N,k}(t,x)}$, we can assert that solving expression (23) is equal to solve the following Cauchy problem for the heat equation

$$\begin{cases} \partial_t \Phi_{N,k}(t,x) - \frac{1}{2N}\Delta\Phi_{N,k}(t,x) = 0 & t \in \mathbb{R}, x \in \mathbb{R}^k, \\ \Phi_{N,k}(0,x) = e^{N\tilde{A}_{N,k}(0,x)} & x \in \mathbb{R}^k. \end{cases} \tag{24}$$

We can now deal with the problem above through standard techniques. Namely, we write

$$\Phi_{N,k}(t,x) = \int_{\mathbb{R}^k} dx'_1 \cdots dx'_k G(t, x - x')\Phi_{N,k}(0,x'), \tag{25}$$

where G is the Green propagator $G(t,x) = \left(\frac{N}{2\pi t}\right)^{k/2} e^{-\frac{\sum_\nu x_\nu^2 N}{2t}}$.

The computations for the initial condition $\Phi_{N,k}(0,x)$ return

$$\Phi_N(0,x) = \exp\left\{N\ln 2 + \sum_{i=1}^N \mathbb{E}\ln\cosh\left(\sum_{\nu=1}^k \tilde{\xi}_i^\nu x_\nu + A\eta\right)\right\}. \tag{26}$$

Therefore, we can state the following

Proposition 2 *Assuming standard conditions on the existence of the solution for the problem in* (24)*, the latter is uniquely given by the following saddle point equation:*

$$\Phi_{N,k}(t,x) = \left(\frac{N}{2\pi t}\right)^{k/2} \int_{\mathbb{R}^k} dx'_1 \cdots dx'_k \, e^{-Ng(t,x,x')},$$

$$g(t,x,x') = \frac{1}{2t}\sum_{\nu=1}^k (x_\nu - x'_\nu)^2 - \ln 2 - \frac{1}{N}\sum_{i=1}^N \mathbb{E}\ln\cosh\left(\sum_{\nu=1}^k \tilde{\xi}_i^\nu x'_\nu + \mathcal{A}\eta_i\right). \tag{27}$$

Recalling that $\tilde{A}_{N,k}(t,x) = \frac{1}{N}\ln\Phi_N(t,x)$*, in the thermodynamic limit we have that*

$$\tilde{A}(t,x) = \lim_{N\to+\infty} \tilde{A}_{N,k}(t,x) = -\min_{x'\in\mathbb{R}^k} g(t,x,x'). \tag{28}$$

To get the full expression of the Guerra Action in the thermodynamic limit, we must finally set $t = \beta$, $x = 0$ and perform the minimization of the function g given in (27): with these values for t and x, we have to fix $x'_\nu = \beta\langle m_\nu\rangle$ $\forall\nu = 1,\ldots,k$.

At this point Eq. (18) is almost all explicit. We still need to calculate the integral term at the top right side of Eq. (18), for which it is sufficient to evaluate the ψ-derivative of the free-energy density $A_{N,k,p}(t,x,\psi)$ and write it in a way that allows extrapolating easily its replica symmetric approximation.

Here we just provide the final result, while the step-by-step calculations for the ψ-derivative are collected in Appendix A. So briefly,

$$\begin{aligned} \frac{dA_{N,k,p}(t,x,\psi)}{d\psi} = \frac{1}{N}\mathbb{E}\left[\frac{d_\psi Z_{N,k,p}(t,x,\psi)}{Z_{N,k,p}(t,x,\psi)}\right] = \frac{1}{2N}\left(\beta - \mathcal{B}^2 - \mathcal{C}\right)\textstyle\sum_{\mu=1}^p \mathbb{E}\Omega\left(z_\mu^2\right)_{t,x} \\ - \frac{\alpha_N\beta}{2}\langle q_{12}p_{12}\rangle_{t,x} - \frac{\mathcal{A}^2}{2}\left(1 - \langle q_{12}\rangle_{t,x}\right) + \frac{\alpha_N\beta^2}{2}\langle p_{12}\rangle_{t,x}. \end{aligned} \tag{29}$$

Fixing the free parameters $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ as

$$\mathcal{A} = \sqrt{\alpha\beta\bar{p}}, \qquad \mathcal{B} = \sqrt{\beta\bar{q}}, \qquad \mathcal{C} = \beta(1-\bar{q}), \tag{30}$$

and adding and subtracting the term $(\alpha_N\beta\cdot\bar{q}\bar{p})/2$ in Eq. (29) we have

$$\frac{dA_{N,k,p}(t,x,\psi)}{d\psi} = -\frac{\alpha_N\beta}{2}\bar{p}(1-\bar{q}) - \frac{\alpha_N\beta}{2}\langle(q_{12}-\bar{q})(p_{12}-\bar{p})\rangle, \tag{31}$$

In the replica symmetric regime, the order parameters m, q_{12}, p_{12} do not fluctuate with respect to their quenched averages in the thermodynamic limit, i.e. using a bar to denote their averages, $\langle m\rangle_{t,x} \to \bar{m}$, $\langle q_{12}\rangle_{t,x} \to \bar{q}_{12}$, $\langle p_{12}\rangle_{t,x} \to \bar{p}_{12}$ as $N \to \infty$. By choosing $\bar{p} = \bar{p}_{12}$ and $\bar{q} = \bar{q}_{12}$ the last term at the r.h.s. of the above expression goes to zero in the thermodynamic limit and the ψ-derivative can be integrated being constant over ψ. It holds [15,21–23] that the optimal values of $\bar{p}$ and $\bar{q}$ can simply be obtained by computing the two overlaps at $\psi = 0$ and this turns out to be equivalent to take the extremum of the trial free energy (18) w.r.t. $\bar{p}$ and $\bar{q}$ as stated in the following main theorem.

Theorem 1 *The replica-symmetric free-energy density of the mixed Hebbian network defined by the Hamiltonian* (9)*, in the thermodynamic limit, is determined by extremizing $A(\boldsymbol{m},\bar{q},\bar{p};\alpha,\beta)$ over $\boldsymbol{m},\bar{q},\bar{p}$, where*

$$\begin{aligned}A(\boldsymbol{m},\bar{q},\bar{p};\alpha,\beta) = &-\tfrac{\alpha\beta}{2} - \tfrac{\alpha}{2}\ln\big(1-\beta(1-\bar{q})\big) + \tfrac{\alpha\beta\bar{q}}{2\left(1-\beta(1-\bar{q})\right)} - \tfrac{\beta}{2}\textstyle\sum_\nu m_\nu^2\\ &+\ln 2 + \left\langle \ln\cosh\left(\beta\textstyle\sum_\nu \tilde{\xi}^\nu m_\nu + \sqrt{\alpha\beta\bar{p}}\eta\right)\right\rangle - \tfrac{\alpha\beta}{2}\bar{p}(1-\bar{q}),\end{aligned} \tag{32}$$

with $\eta \sim \mathcal{N}(0,1)$: the values of these order parameters are thus set via their following self-consistencies

$$\bar{p} = \frac{\beta\bar{q}}{\left(1-\beta(1-\bar{q})\right)^2}, \tag{33}$$

$$\bar{q} = \left\langle \tanh^2\left(\beta\sum_{\nu=1}^{k}\tilde{\xi}^\nu m_\nu + \sqrt{\alpha\beta\bar{p}}\eta\right)\right\rangle, \tag{34}$$

$$m_\nu = \left\langle \tilde{\xi}^\nu\tanh\left(\beta\sum_{\nu=1}^{k}\tilde{\xi}^\nu m_\nu + \sqrt{\alpha\beta\bar{p}}\eta\right)\right\rangle. \tag{35}$$

We highlight that for $\alpha = 0$ and $k = 1$ we recover the Curie-Weiss free energy density [12], while, if $\alpha > 0$ and $k = 0$ we recover the free energy density of the analog Hopfield model at high storage [14] and, finally, keeping $k = 0$, with $\alpha \to \infty$ (such that $\alpha\beta^2 = \beta'$, with β' finite), we recover the expression of the Sherrington-Kirkpatrick free energy density at noise level β' [15,16].

It is also instructive to comment on the physical meaning of the field amplitudes $\mathcal{A} \propto \sqrt{\bar{p}}$, $\mathcal{B} \propto \sqrt{\bar{q}}$: these correctly reproduce, at the mean-field level, the lowest order statistics of the internal field that each party (namely the digital one and the analog one, encoded by σ and by z, respectively) induces on the other one.

Remark 2 In order to get insights in the critical behavior exhibited by the system, in the expression (34), as standard when dealing with second-order phase transitions, we can expand for small q

$$q \simeq \frac{\beta^2\alpha}{(1-\beta)^2}q + o(q).$$

This procedure returns a (second order) transition line for ergodicity breaking at

$$\frac{\beta^2\alpha}{(1-\beta)^2} = 1 \quad\Leftrightarrow\quad \beta = \frac{1}{1+\sqrt{\alpha}},$$

that is the same as the one for the (standard, i.e. digital) Hopfield network [5,6] as well as for its analog counterpart [14,16]: this is not particularly surprising as here we are just checking the pure ergodic/spin-glass transition where universality is expected to hold [29,34].

A different intuition is needed when looking the boundary (i.e., the transition line) splitting the spin-glass phase from a (possible) region of retrieval. In order to find this first-order transition line we must compare the values of the two free-energies (the one under the *pure state* ansatz holding for retrieval and the other for no net magnetization accounting for the spin-glass phase), check that there is a region in the (α, β) plane where one prevails over the other and a complementary region where the opposite is true. The transition line is just given by the set of points in the parameter space where the two free energy balance. Our results return the same transition (hence the same retrieval region) of the standard (i.e. digital) Hopfield network. Its analog counterpart does not retrieve at all hence there is no line to compare for that case.

The whole can be restated in the following proposition

Proposition 3 *The mixed Hebbian network, equipped with an extensive load of real patterns and with a low load of binary patterns, is able to retrieve the binary patterns as long as the system stays confined within the standard AGS-retrieval region* [7,32].

In other words, bearing in mind the stochastic network dynamics as ruled by Eqs. $(4-6)$, starting at random within the boundaries of any basin of attraction of one of the possible minima, if the network's parameters do not exceed their thresholds (set up by the AGS phase transition), the relaxation of such stochastic process will result in a retrieved pattern.

Remark 3 Note that, in the $\alpha \to 0$ limit (hence neglecting the real sea), the critical point becomes $\beta_c = 1$. This is perfectly consistent with the emergence of a ferromagnetic phase (i.e., the point $(\beta = 1, \alpha = 0)$ is the Curie-Weiss or Mattis critical point).

Remark 4 Once we have fixed the parameters $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ (and, in particular, noting that $\mathcal{A} = \sqrt{\alpha\beta\bar{p}}$) and we have obtained an explicit expression for the MHN free-energy density (see Eq. 32), via its self-consistency for $\langle m_\nu \rangle$, we can appreciate how the high load of real patterns acts as a disturbing noise against the signal carried by the Booleans. Indeed, while Eq. 32 looks almost identical to its standard AGS-counterpart (see, e.g. [7, Eq. 6.73] or [32, Eq. 21.78]), its physical interpretation is quite different. In the standard AGS scenario, there are solely binary patterns and they are in a high storage. The Mattis contribution is given by the condensed pattern (i.e. the retrieved one(s)), and all the other (whose amount is linearly extensive in the volume of the neurons) overall introduce a quenched (slow) noise acting against retrieval. In the present picture, instead, the binary patterns again contribute to the Mattis retrieval, but -as they are in the low storage- do not generate any slow noise, rather, the term against retrieval is due to the high load of analog patterns.

This remark suggests that a fairly standard usage of the replica-trick allows to extend the previous result to the case of a high load of Boolean patterns too. Since it is not a rigorous argument we state the following as a

Conjecture 1 *Assuming a high storage of both real patterns (hence $p = \alpha N$) as well as binary patterns (hence $k = \gamma N$), Theorem 1 still holds as long as we replace $\alpha \to \alpha + \gamma$.*

Indeed, to check this within the replica trick framework, once introuced n replicas of the system, we can consider the following partition function

$$Z_N(\beta, \alpha, \gamma) = \sum_{\sigma} \exp\left(\frac{\beta}{2N} \sum_{i,j=1}^{N} \sum_{\mu=1}^{p+k} \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j \right), \tag{36}$$

where the first $p = \alpha N$ patterns have real entries sampled from i.i.d. Gaussians $\mathcal{N}(0,1)$ and the last $k = \gamma N$ have Boolean entries ± 1. We want to compute the averaged replicated partition function $\mathbb{E}_\xi Z^n$, taking then the $\lim_{n\to 0}(\mathbb{E}_\xi Z^n - 1)/n = \mathbb{E}_\xi \log Z$ en route for the free energy [18,48]. We get

$$\mathbb{E}_\xi Z^n = \sum_\sigma \int d\mu(z) \mathbb{E}_\xi \exp\left(\frac{\beta N}{2} \sum_{a=1}^n m_1^2(\sigma^a) + \sqrt{\frac{\beta}{N}} \sum_{a=1}^n \sum_{i=1}^N \sum_{\mu=1}^{p+k-1} \xi_i^\mu \sigma_i^a z_\mu^a \right), \quad (37)$$

where, via the first term inside the exponential, we highlighted the pattern to retrieve while for the other we have introduced $n \times (p+k-1)$ Gaussian random variables to linearize the interactions. Indeed, the resulting second term represents the noise stemming from non-condensed patterns [10,11] and we are going to see that it is universal w.r.t. the pattern distribution. In fact,

$$\begin{aligned}\mathbb{E}_\xi \exp\left(\sqrt{\frac{\beta}{N}} \sum_{a=1}^n \sum_{i=1}^N \sum_{\mu=1}^{p+k-1} \xi_i^\mu \sigma_i^\alpha z_\mu^\alpha \right) &= \exp\left[\sum_{i=1}^N \sum_{\mu=1}^{p+k-1} u_{\xi_i^\mu} \left(\sqrt{\frac{\beta}{N}} \sum_{a=1}^n \sigma_i^a z_\mu^a \right) \right] \\ &\sim \exp\left(\sum_{i=1}^N \sum_{\mu=1}^{p+k-1} \frac{\beta}{2N} \sum_{a,b=1}^n \sigma_i^a \sigma_i^b z_\mu^a z_\mu^b \right).\end{aligned}$$

where we used that the pattern distributions are both symmetric and with unitary variance, hence $u_\xi(x/\sqrt{N}) = \log \mathbb{E}_\xi(e^{\xi x/\sqrt{N}}) = x^2/N + o(1/N)$. Then, the effective load of the network is given by the term

$$\begin{aligned}&\int d\mu(z) \exp\left(\sum_{i=1}^N \sum_{\mu=1}^{p+k-1} \frac{\beta}{2N} \sum_{a,b=1}^n \sigma_i^a \sigma_i^b z_\mu^a z_\mu^b \right) \\ &= \exp\left[(\alpha+\gamma) N \int d\mu(z) \exp\left(\sum_{i=1}^N \frac{\beta}{2N} \sum_{a,b=1}^n \sigma_i^a \sigma_i^b z^a z^b \right) \right], \end{aligned} \quad (38)$$

that is the usual [7,32] slow noise but proportional to the total load $\alpha + \gamma$ this time.

4 The Inverse Process

In this final section we briefly illustrate that proceeding first with the Hamilton-Jacobi formalism and then with the stochastic stability is equivalent to the process we described in Sect. 3.

In this route, instead of the generalized partition function defined in (17), we have the following:

$$\begin{aligned}Z_{N,k,p}(t,x) = &\sum_\sigma \exp\left\{ \frac{t}{2N} \sum_{i,j=1}^N \sum_{\nu=1}^k \tilde{\xi}_i^\nu \tilde{\xi}_j^\nu \sigma_i \sigma_j + \sum_{\nu=1}^k x_\nu \sum_{i=1}^N \tilde{\xi}_i^\nu \sigma_i \right\} \\ &\times \exp\left\{ -\frac{k\beta}{2} - \frac{\beta}{2N} \sum_{i=1}^N \sum_{\mu=1}^p (\xi_i^\mu)^2 + \frac{\beta}{2N} \sum_{i,j=1}^N \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j \right\},\end{aligned}$$

where we can notice the Hamilton–Jacobi scaffold in the interpolation of the Boolean section of the system. We recover the proper partition function if we put $t = \beta$ and $x = 0$, while if $t = 0$ and $x = 1$ we obtain a one-body problem for the Boolean memories.

Even though the generalized free energy is now defined through this new partition function, the equations for its derivatives expressed in (21) still hold and therefore we can proceed with the Hamilton-Jacobi formalism adopting the same argument we used in Sec. 3. So Eqs. (22), (23) and (24) still hold, but now the initial state function $A_{N,k,p}(0,x)$ is

$$\begin{aligned}A_{N,k,p}(0,x) =& \frac{1}{N}\mathbb{E}\ln\Bigg\{\exp\Bigg[-\frac{k\beta}{2} - \frac{\beta}{2N}\sum_{i=1}^{N}\sum_{\mu=1}^{p}(\xi_i^\mu)^2\Bigg]\sum_\sigma \exp\Bigg[\sum_{\nu=1}^{k} x_\nu \sum_{i=1}^{N}\tilde{\xi}_i^\nu \sigma_i\Bigg]\\ &\times \exp\Bigg[\frac{\beta}{2N}\sum_{i,j=1}^{N}\sum_{\mu=1}^{p}\xi_i^\mu \xi_j^\mu \sigma_i\sigma_j\Bigg]\Bigg\}.\end{aligned}$$

This function is now interpretable as the free energy density at a finite volume N of a Hopfield network with p real patterns and an external field (that this time contains patterns of information), so we can now use the stochastic stability technique to write an explicit form of the expression above. To do so, we introduce the variable $\psi \in [0,1]$ and the interpolated free energy density:

$$\begin{aligned}A_{N,k,p}(0,x,\psi) =& -O\left(\frac{\ln N}{N}\right) - \frac{\alpha_N\beta}{2} + \frac{1}{N}\mathbb{E}\ln\Bigg(\sum_\sigma \exp\Bigg\{\sum_\nu\sum_i x_\nu\tilde{\xi}_i^\nu\sigma_i\Bigg\}\\ &\times \int_{\mathbb{R}^p} Dz\,\exp\Bigg\{\sqrt{\psi}\sqrt{\frac{\beta}{N}}\sum_{\mu=1}^{p}\sum_{i=1}^{N}\xi_i^\mu\sigma_i z_\mu\Bigg\} \times \exp\Bigg\{\mathcal{A}\sqrt{1-\psi}\sum_{i=1}^{N}\eta_i\sigma_i\Bigg\}\\ &\times \exp\Bigg\{\mathcal{B}\sqrt{1-\psi}\sum_{\mu=1}^{p}\theta_\mu z_\mu\Bigg\} \times \exp\Bigg\{\mathcal{C}\frac{1-\psi}{2}\sum_{\mu=1}^{p}z_\mu^2\Bigg\}\Bigg).\end{aligned}$$

Mirroring the previous modus operandi, we can now apply the Fundamental Theorem of Calculus in ψ, perform analogous calculations and substitute the values of the free parameters according to (30). What we obtain is:

$$\begin{aligned}A_{N,k,p}(0,x) &= A_{N,k,p}(0,x,\psi=1) = A_{N,k,p}(0,x,\psi=0)\\ &\quad + \int_0^1 d\psi\,(d'_\psi A_{N,k,p}(0,x,\psi'))_{\psi'=\psi}\\ &= -O\left(\frac{\ln N}{N}\right) - \frac{\alpha_N\beta}{2} + \ln 2 + \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\ln\cosh\Bigg(\sum_{\nu=1}^{k}\tilde{\xi}_i^\nu + \sqrt{\alpha_N\beta\bar{p}}\,\eta_i\Bigg)\\ &\quad - \frac{\alpha_N}{2}\ln\big(1-\beta(1-\bar{q})\big) + \frac{\alpha_N\beta\bar{q}}{2\big(1-\beta(1-\bar{q})\big)} - \frac{\alpha_N\beta}{2}\bar{p}(1-\bar{q}).\end{aligned}$$

Now, recalling that the solution to (24) is defined by (25), and that $\Phi_{N,k,p} = e^{NA_{N,k,p}}$, we can write the free-energy density at a finite volume N as

$$A_{N,k,p}(t,x) = -O\left(\frac{\ln N}{N}\right) - \frac{\alpha_N\beta}{2} + \frac{1}{N}\ln\left(\frac{N}{2\pi t}\right)^{k/2} + \frac{1}{N}\ln\int_{\mathbb{R}^k} e^{-Ng(t,x,x')},$$

where

$$g(t,x,x') = \sum_{i=1}^{N} \frac{\left(x_i - x_i'\right)^2}{2t} - \ln 2 - \frac{1}{N}\sum_{i=1}^{N} \mathbb{E} \ln\cosh\left(\sum_{\nu=1}^{k} \xi_i^{\nu} x_{\nu} + \sqrt{\alpha_N \beta \bar{p}}\,\eta_i\right)$$
$$+ \frac{\alpha_N}{2} \ln\left(1 - \beta(1-\bar{q})\right) - \frac{\alpha_N \beta \bar{q}}{2\big(1-\beta(1-\bar{q})\big)} + \frac{\alpha_N \beta}{2} \bar{p}(1-\bar{q}).$$

In the thermodynamic limit the free-energy density is consequently obtained by (28) with the help of a saddle point argument. So, fixing the parameters t and x to be $t = \beta$, $x = 0$ and finding that the minimum of the function g is determined by $x_\nu' = \beta\langle m_\nu\rangle$, we obtain again Eq. (32).

5 Conclusions

The Hopfield neural network and the restricted Boltzmann machine are amongst the best known and intensively studied models in Artificial Intelligence. The former is meant to mimic retrieval, namely the capacity of (the *neurons* of) a machine to recall a pattern of information previously stored. The latter is meant to mimic learning, namely the capacity of (the *synapses* of) a machine to be trained to encode selected patterns of information. Remarkably, Hopfield networks and Boltzmann machines share the same thermodynamics. This equivalence has several implications and, in particular, it implies that the conditions under which the former is able to retrieve are the same conditions under which the latter is able to identify features in the input data. In fact, in this equivalence, the patterns of information retrieved by the Hopfield model corresponds to the optimized weights of the trained Boltzmann machine.

However, in the wide Literature concerning these models, the patterns handled by the Hopfield model are typically binary, while the weights the Boltzmann Machine usually ends up with are real: this gap looks structural since the retrieval of real patterns (at least in the high-load regime) is beyond the Hopfield model capabilities [10,11]. While numerical understanding in the field increases at an impressive rate, analytical improvements proceed more slowly. In order to get further insights into this point through the analytic perspective, in this work we considered a mixed Hopfield network, where patterns are partly real and partly binary and we studied its statistical mechanical properties (i.e., we focused on the behavior of *averaged systems* in *the thermodynamic limit*, which is not the typical benchmark for reseachers in Computer Science).

In particular, we rigorously answered (positively) to the question of whether such a hybrid network with a high-load of analog patterns and a low-load of binary patterns is able to retrieve the latter, under replica symmetry assuption. On the other hand, the retrieval of a high-load of analog patterns is already known to be unfeasible [10,25]. We proved that the hybrid model shares the same phase diagram of the classic Hopfield network with a high storage of Boolean patters only: in the parameter space, where parameters are given by the fast noise (i.e., the temperature) and by the slow-noise (i.e., the "sea" of analog patterns), there exists a retrieval region bounded by a first-order transition line.

This result has been achieved by developing a novel interpolating technique stemming from Guerra's interpolation schemes (see [13–16]). In a nutshell, exploiting the above mentioned equivalence, we recast the hybrid Hopfield model in terms of its related Boltzmann machine and then we ask for stochastic stability of the bulk of patterns (hence the real ones). We interpolate between the free energy of the mixed Hopfield model and two one-body random systems (whose factorized treatment becomes straightforward). This approach allows us to recognize, within the free energy contribution due to real patterns, another nestling free-energy density due to the Boolean contribution of the binary patterns. The latter can then be extracted via the Hamilton-Jacobi route in terms of its natural order parameters. This approach allows detecting when the signal carried by a logarithmic load of Booleans is strong enough to shine over the noisy sea generated by the extensive storage of Gaussian patterns.

The case of high load of real as well as binary patterns can be addressed via a fairly standard replica-trick calculation obtaining evidence that the outlined scenario is preserved as long as the sum of the two slow noises (stemming from the two contributions of real and binary patterns) does not exceed the standard threshold found for the original Hopfield network [7].

Finally, we stress that the mathematical machinery we exploited here does not allow us to check when the replica symmetric solution is the correct solution (i.e. it does not prove or find out boundaries of validity for the overlap distributions to be delta-peaked over their averages in the thermodynamic limit), rather it assumes this and shows its implications. A mathematically completely different route, developed by Bovier and Gayard [26,28] and Talagrand [54,55] for the standard Hopfield model, just investigates the existence of regions in the tuneable parameter space where this assumption is feasible: a rigorous inspection through these techniques on the existence of such allowed retrieval-regions is mandatory also for this mixed Hebbian network and its analysis would be the next natural step of the present investigation.

Acknowledgements E.A. acknowledges financial support from GNFM-INdAM (Progetto Giovani Agliari-2016). A.B. acknowledges financial support from Salento University and by INFN Sezione di Lecce. D.T. acknowledges financial support from GNFM-INdAM (Progetto Giovani Tantari-2016).

Appendix A: Calculating the ψ-Streaming of the Interpolating Free Energy

As anticipated in Sect. 3, in this appendix we will illustrate the calculations of the ψ-derivative of the generalized free energy density $A_{N,k,p}(t,x,\psi)$ written in Eq. (29).

When evaluating the streaming $d_\psi A_{N,k,p}(t,x,\psi)$ we get the sum of four terms: *I*, *II*, *III* and *IV*, that we analyse shortly. Each one comes as a consequence of the derivation of a corresponding exponential term appearing into the expression of the generalized free energy density, whose generalized partition function $Z_{N,k,p}(t,x,\psi)$ is defined in (17).

We remind that we introduced in Sect. 3 the generalized average $\langle\cdot\rangle_{t,x,\psi}$, that naturally extends the Gibbs measure encoded in the interpolating scheme (and is reduced to the proper one whenever setting $t=\beta$, $x=0$ and $\psi=1$). To lighten the expressions, we introduce the function $B_{N,k,p}(t,x,\psi)$ that stands for the generalized Boltzmann factor.
We can now show the calculations of terms I, *II*, *III* and *IV*:

$$I = \frac{1}{N}\mathbb{E}\Big[\sum_\sigma \int d\mathcal{M}(z)\sqrt{\frac{\beta}{N}}\sum_{i=1}^{N}\sum_{\mu=1}^{p}\xi_i^\mu \sigma_i z_\mu \times \frac{1}{2\sqrt{\psi}} B_{N,k,p}(t,x,\psi)\Big] \tag{A.1}$$

$$= \frac{\sqrt{\beta}}{2N\sqrt{N\psi}}\sum_{i=1}^{N}\sum_{\mu=1}^{p}\mathbb{E}\Big[\xi_i^\mu \omega_{t,x,\psi}(\sigma_i z_\mu)\Big] \tag{A.2}$$

$$= \frac{\sqrt{\beta}}{2N\sqrt{N\psi}}\sum_{i=1}^{N}\sum_{\mu=1}^{p}\mathbb{E}\Big[\partial_{\xi_i^\mu}\omega_{t,x,\psi}(\sigma_i z_\mu)\Big] \tag{A.3}$$

$$= \frac{\beta}{2N}\sum_{\mu=1}^{p}\mathbb{E}\omega_{t,x,\psi}(z_\mu^2) - \frac{\alpha_N\beta}{2}\langle q_{12}p_{12}\rangle_{t,x,\psi}, \tag{A.4}$$

$$II = \frac{1}{N}\mathbb{E}\Bigg[\frac{1}{Z_N(t,x,\psi)}\sum_\sigma\int d\mathcal{M}(z)\,\frac{-\mathcal{A}}{2\sqrt{1-\psi}}\sum_{i=1}^{N}\eta_i\sigma_i B_{N,k,p}(t,x,\psi)\Bigg] \tag{A.5}$$

$$= -\frac{\mathcal{A}}{2N\sqrt{1-\psi}}\sum_{i=1}^{N}\mathbb{E}\Big[\eta_i\omega_{t,x,\psi}(\sigma_i)\Big] \tag{A.6}$$

$$= -\frac{\mathcal{A}}{2N\sqrt{1-\psi}}\sum_{i=1}^{N}\mathbb{E}\Big[\partial_{\eta_i}\omega_{t,x,\psi}(\sigma_i)\Big] \tag{A.7}$$

$$= -\frac{\mathcal{A}^2}{2}\big(1-\langle q_{12}\rangle_{t,x,\psi}\big), \tag{A.8}$$

$$III = \frac{1}{N}\mathbb{E}\Bigg[\frac{1}{Z_N(t,x,\psi)}\sum_\sigma\int d\mathcal{M}(z)\,\frac{-\mathcal{B}}{2\sqrt{1-\psi}}\sum_{\mu=1}^{p}\theta_\mu z_\mu B_{N,k,p}(t,x,\psi)\Bigg] \tag{A.9}$$

$$= -\frac{\mathcal{B}}{2N\sqrt{1-\psi}}\sum_{\mu=1}^{p}\mathbb{E}\Big[\theta_\mu\omega_{t,x,\psi}(z_\mu)\Big] \tag{A.10}$$

$$= -\frac{\mathcal{B}}{2N\sqrt{1-\psi}}\mathbb{E}\Big[\partial_{\theta_\mu}\omega_{t,x,\psi}(z_\mu)\Big] \tag{A.11}$$

$$= -\frac{\mathcal{B}^2}{2N}\sum_{\mu=1}^{p}\mathbb{E}\omega_{t,x,\psi}(z_\mu^2) + \frac{\alpha_N\mathcal{B}^2}{2}\langle p_{12}\rangle_{t,x,\psi}. \tag{A.12}$$

In these three equations we used integration by parts (Wick's Theorem), and we manipulated the expressions in order to let the order parameters q_{12} and p_{12} appear (for their general definitions see Eqs. (12) and (13)). Term IV is easily computed through the standard Gaussian integration:

$$\begin{aligned} IV &= \tfrac{1}{N}\mathbb{E}\Big[\tfrac{1}{Z_{N,k,p}(t,x,\psi)}\textstyle\sum_\sigma\int d\mathcal{M}(z)\,\tfrac{-\mathcal{C}}{2}\sum_{\mu=1}^{p} z_\mu^2 B_{N,k,p}(t,x,\psi)\Big] \\ &= \tfrac{-\mathcal{C}}{2N}\textstyle\sum_{\mu=1}^{p}\mathbb{E}\omega_{t,x,\psi}(z_\mu^2). \end{aligned} \tag{A.13}$$

Summing the final expressions of Eqs. (A.4), (A.8), (A.12) and (A.13) we have:

$$\begin{aligned} \tfrac{dA_{N,k,p}}{d\psi}(t,x,\psi) = {} & \tfrac{1}{2N}\big(\beta-\mathcal{B}^2-\mathcal{C}\big)\textstyle\sum_{\mu=1}^{p}\mathbb{E}\omega_{t,x,\psi}(z_\mu^2) \\ & -\tfrac{\alpha_N\beta}{\langle q_{12}p_{12}\rangle_{t,x,\psi}} - \tfrac{\mathcal{A}^2}{2}\big(1-\langle q_{12}\rangle_{t,x,\psi}\big) + \tfrac{\alpha_N\mathcal{B}^2}{2}\langle p_{12}\rangle_{t,x,\psi}, \end{aligned}$$

which is what we reported in Eq. (29).

References

1. Agliari, E., et al.: Multitasking associative networks. Phys. Rev. Lett. **109**, 268101 (2012)
2. Agliari, E., Barra, A., De Antoni, A., Galluzzi, A.: Parallel retrieval of correlated patterns: from hopfield networks to Boltzmann machines. Neural Netw. **38**, 52–63 (2013)

3. Agliari, E., et al.: Retrieval capabilities of hierarchical networks: from dyson to hopfield. Phys. Rev. Lett. **114**, 028103 (2015)
4. Aizenman, M., Contucci, P.: On the stability of the quenched state in mean-field spin-glass models. J. Stat. Phys. **92**(5), 765–783 (1998)
5. Amit, D.J., Gutfreund, H., Sompolinsky, H.: Spin glass model of neural networks. Phys. Rev. A **32**, 1007–1018 (1985)
6. Amit, D.J., Gutfreund, H., Sompolinsky, H.: Storing infinite numbers of patterns in a spin glass model of neural networks. Phys. Rev. Lett. **55**, 1530–1533 (1985)
7. Amit, D.J.: Modeling Brain Function: The World of Attractor Neural Networks. Cambridge University Press, Cambridge (1992)
8. Auffinger, A., Chen, W.K.: Free energy and complexity of spherical bipartite models. J. Stat. Phys. **157**, 40 (2014)
9. Baldi, P., Sadowski, P., Whiteson, D.: Searching for exotic particles in high-energy physics with deep learning. Nat. Commun. **5**, 12 (2014)
10. Barra, A., Genovese, G., Sollich, P., Tantari, D.: Phase diagram of Restricted Boltzmann Machines and Generalised Hopfield Networks with arbitrary priors. preprint arXiv:1702.05882 (2017)
11. Barra, A., Genovese, G., Sollich, P., Tantari, D.: Phase transitions in Restricted Boltzmann Machines with generic priors. preprint arXiv:1612.03132 (2016)
12. Barra, A.: The mean field Ising model trough interpolating techniques. J. Stat. Phys. **132**(5), 787–809 (2008)
13. Barra, A., Di Biasio, A., Guerra, F.: Replica symmetry breaking in mean-field spin glasses through the Hamilton Jacobi technique. J. Stat. Mech. **2010**(09), P09006 (2010)
14. Barra, A., Genovese, G., Guerra, F.: The replica symmetric approximation of the analogical neural network. J. Stat. Phys. **140**(4), 784–796 (2010)
15. Barra, A., Genovese, G., Guerra, F.: Equilibrium statistical mechanics of bipartite spin systems. J. Phys. A **44**, 245002 (2011)
16. Barra, A., Genovese, G., Guerra, F., Tantari, D.: How glassy are neural networks? J. Stat. Mech. **07**, P07009 (2012)
17. Barra, A., Bernacchia, A., Santucci, E., Contucci, P.: On the equivalence of Hopfield networks and Boltzmann machines. Neural Nets **34**, 1–9 (2012)
18. Barra, A., Guerra, F., Mingione, E.: Interpolating the SherringtonKirkpatrick replica trick. Philos. Mag. **92**, 78–97 (2012)
19. Barra, A., Del Ferraro, G., Tantari, D.: Mean field spin glasses treated with pde techniques. Eur. Phys. J. B **86**(7), 1–10 (2013)
20. Barra, A., Genovese, G., Guerra, F., Tantari, D.: About a solvable mean field model of a Gaussian spin glass. J. Phys. A **47**(15), 155002 (2014)
21. Barra, A., Galluzzi, A., Guerra, F., Pizzoferrato, A., Tantari, D.: Mean field bipartite spin models treated with mechanical techniques. Eur. Phys. J. B **87**(3), 74 (2014)
22. Barra, A., Contucci, P., Mingione, E., Tantari, D.: Multi-species mean field spin glasses. Rigorous results. Ann. Henri Poincaré **16**(3), 691–708 (2015)
23. Barra, A., Guerra, F.: About the ergodic regime in the analogical Hopfield neural networks: moments of the partition function. J. Math. Phys. **49**, 125217 (2008)
24. Bovier, A., Gayrard, V., Picco, P.: Gibbs states of the Hopfield model with extensively many patterns. J. Stat. Phys. **79**, 395–414 (1995)
25. Bovier, A., van Enter, A.C.D., Niederhauser, B.: Stochastic symmetry-breaking in a Gaussian Hopfield model. J. Stat. Phys. **95**(1–2), 181–213 (1999)
26. Bovier, A., Gayrard, V.: The retrieval phase of the Hopfield model, A rigorous analysis of the overlap distribution. Probab. Theor. Relat. Fields **107**, 61–98 (1995)
27. Bovier, A., Gayrard, V.: Hopfield models as generalized random mean field models. In: Bovier, A., Picco, P. (eds.) Progress in Probability. Birkauser, Boston (1997)
28. Bovier, A., Gayrard, V.: Metastates in the Hopfield model in the replica symmetric regime. Math. Phys. (Anal. Geom.) **1**(2), 107–144 (1998)
29. Carmona, P., Hu, Y.: Universality in Sherrington—Kirkpatrick's spin glass model. Ann. Henri Poincaré (B) **42**(2), 215–222 (2006)
30. Choromanska, A., Henaff, M., Mathieu, M., Arous, G.B., LeCun, Y.: The Loss Surfaces of Multilayer Networks. In AISTATS (2015)
31. Contucci, P., Giardiná, C.: Perspectives on Spin Glasses. Cambridge University Press, Cambridge (2013)
32. Coolen, A.C.C., Kühn, R., Sollich, P.: Theory of Neural Information Processing Systems. Oxford Press, Oxford (2005)

33. Gardner, E.J., Wallace, D.J., Stroud, N.: Training with noise and the storage of correlated patterns in a neural network model. J. Phys. A **22**(12), 2019 (1989)
34. Genovese, G.: Universality in bipartite mean field spin glasses. J. Math. Phys. **53**(12), 123304 (2012)
35. Genovese, G., Tantari, D.: Non-convex multipartite ferromagnets. J. Stat. Phys. **163**(3), 492–513 (2016)
36. Guerra, F.: Sum rules for the free energy in the mean field spin glass model. Fields Inst. Commun. **30**, 161 (2001)
37. Hackley, D.H., Hinton, G.E., Sejnowski, T.J.: A learning alghoritm for Boltzmann machines. Cogn. Sci. **9**(1), 147 (1985)
38. Hebb, O.D.: The Organization of Behaviour: A Neuropsychological Theory. Psychology Press, New York (1949)
39. Hinton, G.E., Osindero, S., Teh, Y.W.: A fast algorithm for deep belief nets. Neural Comput. **18**, 1527–1554 (2006)
40. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA **79**, 2554–2558 (1982)
41. Huang, H.: Statistical mechanics of unsupervised feature learning in a restricted Boltzmann machine with binary synapses, arXiv preprint arXiv:1612.01717 (2016)
42. Huang, H., Toyoizumi, T.: Advanced mean-field theory of the restricted Boltzmann machine. Phys. Rev. E **91**, 050101 (2015)
43. Huang, H., Toyoizumi, T.: Unsupervised feature learning from finite data by message passing: discontinuous versus continuous phase transition. Phys. Rev. E **94**, 062310 (2016)
44. Larocelle, H., Mandel, M., Pascanu, R., Bengio, Y.: Learning algorithms for the classification restricted Boltzmann machine. J. Mach. Learn. **13**, 643–669 (2012)
45. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
46. Lobo, D., Levin, M.: Inferring regulatory networks from experimental morphological phenotypes: a computational method reverse-engineers planarian regeneration. PLoS Comput. Biol. **11**(6), e1004295 (2015)
47. Mezard, M.: Mean-field message-passing equations in the Hopfield model and its generalizations, arXiv:1608.01558v1 (2016)
48. Mézard, M., Parisi, G., Virasoro, M.A.: Spin Glass Theory and Beyond. World Scientific, Singapore (1987)
49. Salakhutdinov, R., Hinton, G.E.: Deep Boltzmann machines. AISTATS **1**, 3 (2009)
50. Seung, H.S., Sompolinsky, H., Tishby, N.: Statistical mechanics of learning from examples. Phys. Rev. A **45**(8), 6056 (1992)
51. Shcherbina, M.: Some mathematical problems of neural networks theory. In: Proceedings of the 4th European Congress in Mathematics. EMS Publishing house (2005)
52. Shcherbina, M., Tirozzi, B.: Rigorous solution of the gardner problem. Commun. Math. Phys. **234**, 383–422 (2003)
53. Sollich, P., Tantari, D., Annibale, A., Barra, A.: Extensive parallel processing on scale free networks. Phys. Rev. Lett. **113**, 238106 (2014)
54. Talagrand, M.: Rigorous results for the Hopfield model with many patterns. Probab. Theory Relat. Fields **110**(2), 177–275 (1998)
55. Talagrand, M.: Exponential inequalities and convergence of moments in the replica-symmetric regime of the Hopfield model. Ann. Probab. **12**, 1393–1469 (2000)
56. Tubiana, J., Monasson, R.: Emergence of compositional representations in restricted Boltzmann machines. Phys. Rev. Lett. **118**, 138301 (2017)
57. Varadhan, S.R.: Large Deviations and Applications. Society for Industrial and Applied Mathematics, Philadelphia (1984)
58. Zhen, H., Wang, S.N., Zhou, H.J.: Unsupervised prototype learning in an associative-memory network, arXiv:1704.02848 (2017)