NON-CONVEX MULTI-SPECIES HOPFIELD MODELS

ELENA AGLIARI, DANILA MIGLIOZZI, AND DANIELE TANTARI

ABSTRACT. In this work we introduce a multi-species generalization of the Hopfield model for associative memory, where neurons are divided into groups and both inter-groups and intra-groups pair-wise interactions are considered, with different intensities. Thus, this system contains two of the main ingredients of modern Deep neural network architectures: Hebbian interactions to store patterns of information and multiple layers coding different levels of correlations. The model is completely solvable in the low-load regime with a suitable generalization of the Hamilton-Jacobi technique, despite the Hamiltonian can be a non-definite quadratic form of the magnetizations. The family of multi-species Hopfield model includes, as special cases, the 3-layers Restricted Boltzmann Machine (RBM) with Gaussian hidden layer and the Bidirectional Associative Memory (BAM) model.

1. INTRODUCTION

Artificial Intelligence (AI) is nowadays playing a major role in our everyday lives and has grown extremely fast in the past few years, both in popularity and in scientific advances. The rise over the past decade has been strongly correlated with the increased processing power of modern computers: the progressions have made complex computations more accessible and allowed for AI-models to train on data-sets so large and multifaceted it would not have been possible a few years ago.

Neural networks are playing an important and renewed role in this trend mainly because of the success of Deep Learning [1, 2] in several applications, ranging from engineering and computer science to neuro and computational biology. From a mathematical perspective a neural network consists of many simple, connected processors (called neurons), each associated to an activity state depending on the state of the other neurons and on a possible external stimulation. In the family of neural network models the Hopfield model plays a major role. It was introduced to mimic the ability of the brain to retrieve information previously stored (see e.g., [3, 4, 5, 6]) but recently it has been investigated also because of its equivalence with the fundamental constituent of a Deep architecture, namely a two-layer neural network called Restricted Boltzmann Machine (RBM) [7, 8, 9, 10, 11].

This bridge is particularly interesting since, while a solid theoretical description does exist for the mechanisms behind the retrieval capabilities of the Hopfield model, a clear and exhaustive theoretical scaffold for the performances of Deep Networks is still needed and most of the advances in technology are often achieved by specific home-made recipes, without a global and systematic understanding of the methodology. For example in [12, 13, 14, 15, 16] it has been shown how the statistical mechanics analysis of the Hopfield model phase diagram allows answering some issues concerning both the way RBMs extract features from data and their efficiency in terms of the training-set size necessary for a good generalization.

The heart of the Hopfield model is the particular shape of the interaction between neurons that follows the well-known Hebbian rule [17]. The structural property of a deep architecture is the presence of several layers or groups of units (neurons) where only neurons belonging to two consecutive layers are connected: the aim of each layer is to codify more and more abstract levels of correlations. In this work we consider an extension of the standard Hopfield model, with multiple groups of neurons and different Hebbian-like interactions for each couple of groups. The multi-species model considered in this paper contains also, as special cases corresponding to particular choices of the parameters, the 3-layers RBM and the Bidirectional Associative Memory (BAM) model. The latter was introduced [18, 19, 20] to mimic the ability of the human brain to retrieve informations through association of ideas.

Remarkably, in our model interactions between groups can be suitably tuned in such a way that inter-groups interaction may possibly result stronger than the intra-group one. As a consequence, the model turns out to be in the family of non-convex multi-species systems. It is known that non-convexity can yield to strong difficulties in the statistical mechanics analysis, especially when dealing with disordered systems. For example, the multi-species Sherrington Kirkpatrick (SK) model for spin glasses [21, 22,

23] still presents several mathematical challenges in the non-convex region, while it was completely explored in the convex regime [21, 24]. Some properties of the multi-species SK model, in particular the overlaps synchronization, have recently driven towards the investigation of a multipartite version of the Generalized Random Energy model [25]. In the case of multipartite ferromagnets the analysis is relatively viable [26] and a suitable strategy based on convexification of the problem has been recently introduced [27] and used to compute the model free energy, the equilibrium states and the thermodynamic fluctuations. Hebbian models are disordered systems but they are purely ferromagnetic in the low load regime, see Section 2. For this reason in this paper we can generalize the strategy in [27] to find an exact solution of the multi-species Hopfield model in that regime.

The remaining of the paper is organized as follows. In Sec. 2 we review the basic Hopfield model, that is the starting point for our work. In Sec. 3 we introduce the multi-species Hopfield model and the main observables used for its investigation. We also give the main theorem concerning the variational principle for the free energy and the equations for the optimal parameters. In Sec. 4 we discuss two particular cases of the model, corresponding to the BAM and the 3-layer RBM. In Sec. 5 we present the proof of the main theorem. Finally, Sec. 6 is left for conclusions and outlooks. Some technical details about the solution are collected in the Appendix.

2. A brief review on the Hopfield model

Before introducing the extended model considered in this work it is worth reviewing the standard Hopfield model (see also e.g., [3, 4, 28] for a more extensive explanation).

We consider N binary neurons (i.e., Ising spins for statistical physicists [3] or McCulloch-Pitts neurons for computer scientists [29]) and to each neuron i we assign a variable σ_i that describes its activity: if $\sigma_i = +1$ the i-th neuron is spiking, while if $\sigma_i = -1$ the i-th neuron is quiescent. We denote with $\boldsymbol{\sigma} \in \{-1, +1\}^N$ the overall configuration of the system.
Neurons are embedded on a fully-connected network in such a way that the internal field h_i acting on the i-th neuron is given by

$$h_i(\boldsymbol{\sigma}) = \sum_{\substack{j=1 \\ j \neq i}}^{N} J_{ij}\sigma_j,$$

where $\mathbf{J} = \{J_{ij}\}_{i,j=1,...,N}$ is the *synaptic* coupling between neuron j and neuron i. Inspired by neurophysiological mechanisms, one can introduce a dynamic rule for the spin configuration as follows: extract randomly and uniformly a spin i and update its state according to

$$\sigma_i(t) \to \sigma_i(t+1) = \mathrm{sgn}\left[h_i(\boldsymbol{\sigma}(t)) + Tz_i(t)\right], \tag{2.1}$$

where t is the discrete time unit, z is a random variable[1] and $T := 1/\beta$ is a measure of the degree of noise within the system, usually referred to as temperature. As long as $\boldsymbol{J}$ is symmetric (i.e., $J_{ij} = J_{ji}, \forall i, j$) and devoid of self-interactions (i.e., $J_{ii} = 0, \forall i$), this dynamic is ergodic and there exists an invariant measure given by

$$p_{\mathrm{eq}}(\beta) = \frac{1}{Z_N(\beta|\mathbf{J})}e^{-\beta H_N(\boldsymbol{\sigma}|\mathbf{J})}, \tag{2.2}$$

where $Z_N(\beta|\mathbf{J})$ is a normalization factor, also called partition function, and

$$H_N(\boldsymbol{\sigma}|\mathbf{J}) = -\sum_{1 \leqslant i < j \leqslant N} \sigma_i J_j \sigma_j. \tag{2.3}$$

In a statistical-mechanics context the distribution (2.2) corresponds to the Boltzmann-Gibbs measure used to describe the canonical equilibrium of a system described by the Hamiltonian (2.3).

Now, the goal of the system is to be able to recognize and retrieve a certain group of words, pixels, or, generically, patterns. A *pattern* is defined as a sequence of random variables $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \ldots, \xi_N^\mu)$ with $\xi_i^\mu \in \{-1, +1\}, \forall i = 1, ..., N$, and $\forall \mu = 1, ..., P$, where the label μ distinguishes different patterns. In the following we shall assume the set $\{\xi_i^\mu\}_{i,\mu}$ made of i.i.d. random variables such that

$$\mathbb{P}(\xi_i^\mu = +1) = \mathbb{P}(\xi_i^\mu = -1) = 1/2, \ \forall i, \mu. \tag{2.4}$$

More general patterns distribution can be investigated: correlated patterns [30, 31, 32], diluted patterns, i.e. which may have zero entries [33, 34, 35, 36, 37, 38], Gaussian patterns [39, 40, 41, 42, 43], weighted patterns [44]. With such premises we can give a definition of *retrieval*: we say that the system is able

---

[1]The random variable z is chosen symmetrically distributed and, typically, its probability density is taken as $p(z) = [1 - \tanh^2(z)]/2$.

to retrieve the μ-th pattern if, given a suitable starting point (i.e., a configuration belonging to the attraction basin of $\boldsymbol{\xi}^{\mu}$), the spin configuration $\boldsymbol{\sigma}$ converges to $\boldsymbol{\xi}^{\mu}$ under the dynamics (2.1). One can prove that a coupling matrix defined according to Hebb's learning rule [17] as

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^{\mu} \xi_j^{\mu} \tag{2.5}$$

ensures the attractiveness of the P patterns, as long as the noise T and the number of patterns P are not too large. For instance, by taking P finite (or, still, sublinear with respect to N) in the thermodynamic limit $N \to \infty$, one has retrieval capabilities as long as $T < 1$. A modification of the Hebb rule results in a deformation of the basins of attractions: this can be done for example overlaying Hebb to another interaction structure as a diluted [45, 46] or a hierarchical [47, 48, 49, 50] structure.

In order to describe the overall state of the system, one introduces the macroscopic observable $\mathbf{m}$, also called *Mattis magnetization*, that is a vector of length P, whose μ-th component represents the overlap between the spin configuration and the μ-th pattern:

$$m_{\mu}(\boldsymbol{\sigma}|\boldsymbol{\xi}) := m_{\mu} = \frac{1}{N} \sum_{i=1}^{N} \xi_i^{\mu} \sigma_i \ \in [-1, 1]. \tag{2.6}$$

Notice that m_{μ} can also be written as $m_{\mu} = 2d_H(\boldsymbol{\sigma}, \boldsymbol{\xi}) - 1$, where $d_H(x, y)$ is the Hamming distance between the strings x and y. The Mattis magnetization plays as the *order parameter* for the system as a certain arrangement for $\mathbf{m} = (m_1, m_2, ..., m_P)$ can be associated to a global state for the system; in particular, the retrieval of a pattern μ corresponds to $m_{\mu} \neq 0$. It is worth underlying that, in the thermodynamic limit, (2.4) yields to mutually orthogonal patterns in such a way that only one pattern at a time can be retrieved exactly, i.e. $m_{\mu} = 1$. Now, combining (2.3), (2.5) and (2.6) we can state that the Hamiltonian for the Hopfield model equipped with N Ising neurons and P patterns is defined as

$$H_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) = -\frac{1}{N} \sum_{1 \leqslant i < j \leqslant N} \sum_{\mu=1}^{P} \xi_i^{\mu} \xi_j^{\mu} \sigma_i \sigma_j = -N \sum_{\mu=1}^{P} m_{\mu}^2 + \frac{P}{N}. \tag{2.7}$$

One can also account for an external field $\mathbf{h} = (h_1, h_2, ..., h_P)$ which biases the neuron configurations toward the retrieval of the μ-th pattern with a relative magnitude h_{μ}, leading to

$$H_N(\boldsymbol{\sigma}|\boldsymbol{\xi}, \mathbf{h}) = -N \sum_{\mu=1}^{P} m_{\mu}^2 - N \sum_{\mu=1}^{P} m_{\mu} h_{\mu}, \tag{2.8}$$

where we neglected the unnecessary constant term P/N. The partition function $Z_N(\beta, \mathbf{h}|\boldsymbol{\xi})$ for the Hopfield network then reads as

$$Z_N(\beta, \mathbf{h}|\boldsymbol{\xi}) = \sum_{\boldsymbol{\sigma}} \exp\left[\beta N \left(\sum_{\mu=1}^{P} m_{\mu}^2 + \sum_{\mu=1}^{P} h_{\mu} m_{\mu}\right)\right]. \tag{2.9}$$

Hereafter, whenever suitable, we will drop the dependence on $\boldsymbol{\xi}$ and on $\mathbf{h}$ in order to lighten the notation. We also define the expectation of the observable $O(\boldsymbol{\sigma})$, that is a function of the state $\boldsymbol{\sigma}$, as the average with respect to the equilibrium (or Boltzmann-Gibbs) distribution (2.2):

$$\langle O \rangle_N = \sum_{\boldsymbol{\sigma}} O(\boldsymbol{\sigma}) p_{\text{eq}}(\boldsymbol{\sigma}) = \frac{1}{Z_N(\beta)} \sum_{\boldsymbol{\sigma}} O(\boldsymbol{\sigma}) e^{-\beta H_N(\boldsymbol{\sigma})}. \tag{2.10}$$

In particular, in the following, we shall be interested in the expectation $\langle \mathbf{m} \rangle_N$ which, in a statistical-mechanics framework, can be obtained by extremizing the intensive pressure (or, equivalently, the free energy[2]) of the model. More precisely, recalling that the intensive pressure at finite size N is defined as

$$f_N(\beta, \mathbf{h}|\boldsymbol{\xi}) := \frac{1}{N} \ln Z_N(\beta, \mathbf{h}|\boldsymbol{\xi}), \tag{2.11}$$

and denoting the thermodynamic limit as $f(\beta, \mathbf{h}|\boldsymbol{\xi}) = \lim_{N \to \infty} f_N(\beta, \mathbf{h}|\boldsymbol{\xi})$, it holds

$$\langle \mathbf{m} \rangle = \lim_{N \to \infty} \langle \mathbf{m} \rangle_N = \lim_{N \to \infty} \frac{1}{\beta} \nabla_{\boldsymbol{\lambda}} f_N(\beta, \mathbf{h} + \boldsymbol{\lambda}|\boldsymbol{\xi})|_{\boldsymbol{\lambda}=0} = \frac{1}{\beta} \nabla_{\boldsymbol{\lambda}} f(\beta, \mathbf{h} + \boldsymbol{\lambda}|\boldsymbol{\xi})|_{\boldsymbol{\lambda}=0}. \tag{2.12}$$

---

[2]The intensive pressure, here denoted as $f_N(\beta)$, is strictly related to the (possibly more familiar) free energy $\tilde{f}_N(\beta)$ by $f_N(\beta) = -\beta \tilde{f}_N(\beta)$. Therefore, the existence and the uniqueness of $\tilde{f}_N(\beta)$ also ensure the existence and uniqueness of $f_N(\beta)$ and vice versa, while the positive convexity of $\tilde{f}_N(\beta)$ ensures the negative convexity of $f_N(\beta)$. As a result, the thermodynamic equilibrium can be detected as a minimum for the free energy or as a maximum for the pressure.

For the Hopfield model (2.8), in the low-load regime, the limiting free energy $f(\beta, \mathbf{h}|\boldsymbol{\xi})$ is proved to exist and to be selfaveraging over the patterns' noise. Moreover, one can write a.s.

$$f(\beta, \mathbf{h}) = \lim_{N\to\infty} \frac{1}{N} \ln Z_N(\beta, \mathbf{h}|\boldsymbol{\xi}) = \sup_{\mathbf{M}} \left[\ln 2 + \langle \ln \cosh(\beta\boldsymbol{\xi}(\mathbf{M} + \mathbf{h})) \rangle_{\boldsymbol{\xi}} - \frac{1}{2} \sum_{\mu=1}^{P} M_\mu^2 \right], \qquad (2.13)$$

where $\langle \, \cdot \, \rangle_{\boldsymbol{\xi}}$ means the average over the patterns $\boldsymbol{\xi}$. By extremizing the previous expression for the pressure with respect to the trial magnetization $\mathbf{M}$, we get the following *self-consistent* equations

$$M_\mu = \langle \xi^\mu \tanh[\beta\boldsymbol{\xi}(\mathbf{M} + \mathbf{h})] \rangle_{\boldsymbol{\xi}}, \ \mu = 1, ..., P, \qquad (2.14)$$

or, in vectorial notation,

$$\mathbf{M} = \langle \boldsymbol{\xi} \tanh[\beta\boldsymbol{\xi}(\mathbf{M} + \mathbf{h})] \rangle_{\boldsymbol{\xi}}. \qquad (2.15)$$

Using equations (2.12) and (2.13) it holds that the solution of (2.15) coincides with $\langle \mathbf{m} \rangle$ and gives the behavior of the magnetizations with respect to the system parameters. In particular, posing $\mathbf{h} = 0$, one can see that as $T > 1$ the only solution is $M_\mu = 0$, $\forall \mu = 1, ..., P$, that is, noise prevails and the system is not able to retrieve, while as $T < 1$ the system exhibits a symmetry breaking: equations (2.15) have multiple solutions, each one related to a different way of taking the limit $\mathbf{h} \to 0$ and there are solutions with (at least) one non-null component of the magnetization, where the corresponding pattern is retrieved. At $T = 1$ the system exhibits a continuous phase transition.

The phase diagram becomes richer at high load (i.e, when P grows linearly with N) and the equation (2.13) does not hold anymore. In this case it is well known the existence of another phase transition from the retrieval phase to a spin-glass phase, where the system can freeze on thermodynamic states that are not correlated with any pattern. The occurrence of such new phase depends on the number of stored patterns: when $\alpha = P/N$ exceeds a critical capacity the system is not able to retrieve anymore. The study of the phase diagram and the value of the critical capacity have been intensively investigated in the literature with approximated techniques from statistical physics [4, 5, 6]. Conversely, rigorous results are sparser and partial [51, 52, 53, 54, 55, 56].

3. THE MULTI-SPECIES HOPFIELD MODEL

In this section we introduce a multi-species Hopfield model which generalizes the standard model introduced in the previous section by allowing for neurons belonging to different groups (or species) characterized by different inter-group and intra-group couplings.

Let $\nu \geq 2$ be the number of species, each made of N_a neurons, $a = 1, ..., \nu$. We denote with $\sigma_i^a \in \{-1, 1\}$, $i = 1, ..., N_a$, the state of the i-th neuron in the a-th group. The overall number of neurons is $N = \sum_{a=1}^{\nu} N_a$ and, in the thermodynamic limit $N \to \infty$, we define the parameters

$$\alpha_a := \lim_{N\to\infty} \frac{N_a}{N} \in (0, 1), \ \text{with } a = 1, ..., \nu \qquad (3.1)$$

Without loss of generality we assume always $\alpha_a = N_a/N$, also before the limit. We associate to each group a a set of P patterns $\boldsymbol{\xi}^{\mu,a} = \{\xi_i^{\mu,a}\}_{i=1}^{N_a}$, $\mu = 1, ..., P$, drawn according to (2.4) and a set of P *Mattis magnetizations*

$$m_a^\mu := m_a^\mu(\boldsymbol{\sigma}|\boldsymbol{\xi}) = \frac{1}{N_a} \sum_{i=1}^{N_a} \xi_i^{\mu,a} \sigma_i^a \qquad \forall \mu = 1, \dots, P. \qquad (3.2)$$

In other terms, m_a^μ represents the normalized overlap between the μ-th *pattern* $\boldsymbol{\xi}^{\mu,a}$, related to the a-th group, and the neuron configuration $\boldsymbol{\sigma}^a$ of the same group.

The Hamiltonian $H_{N,\nu}(\boldsymbol{\sigma})$ describing the multi-species Hopfield model reads as

$$H_{N,\nu}(\boldsymbol{\sigma}|\boldsymbol{\alpha}, \mathbf{k}, \boldsymbol{\xi}) = -\frac{1}{2N} \sum_{a=1}^{\nu} \sum_{i,j=1}^{N_a} k_a \sum_{\mu=1}^{P} \xi_i^{\mu,a} \xi_j^{\mu,a} \sigma_i^a \sigma_j^a - \frac{1}{N} \sum_{(a,b)}^{\nu} \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} \sum_{\mu=1}^{P} \xi_i^{\mu,a} \xi_j^{\mu,b} \sigma_i^a \sigma_j^b \qquad (3.3)$$

$$= -\frac{N}{2} \left(\sum_{\mu=1}^{P} \sum_{a=1}^{\nu} k_a \alpha_a^2 (m_a^\mu)^2 + \sum_{\mu=1}^{P} \sum_{a,b=1}^{\nu} \alpha_a \alpha_b m_a^\mu m_b^\mu \right), \qquad (3.4)$$

where the first term is due to interactions between neurons belonging to the same group, while the second term is due to interactions between neurons belonging to different groups; also, $k_a \in (0, 1)$, $\forall a \in \{1, \dots, \nu\}$, are introduced to tune the magnitude of the intra-group respect to the inter-group interactions. Clearly, setting $\nu = 1$, the second terms in (3.3) and in (3.4) vanish and we recover the standard Hopfield model (2.7), i.e. $H_{N,1}(\boldsymbol{\sigma}) \equiv H_N(\boldsymbol{\sigma})$. Also in the homogeneous case, i.e. $k_a = 1$,

$\forall a \in \{1, \ldots, \nu\}$, we recover the standard Hopfield model for $N = \sum_{a=1}^{\nu} N_a$ neurons and patterns $\boldsymbol{\xi}^{\mu} = (\boldsymbol{\xi}^{\mu,1}, \ldots, \boldsymbol{\xi}^{\mu,\nu})$, each one obtained by concatenating the ν patterns $\boldsymbol{\xi}^{\mu,a}$, i.e. $H_{N,\nu}(\boldsymbol{\sigma}|\boldsymbol{k}=\mathbf{1}) \equiv H_N(\boldsymbol{\sigma})$.

Again, we can introduce an external field h_a^{μ}, $\forall a = 1, \ldots, \nu$ and $\forall \mu = 1, \ldots, P$, which forces the neurons of the a-th group $\boldsymbol{\sigma}^a$ to retrieve of the μ-th pattern $\boldsymbol{\xi}^{\mu,a}$, as

$$H_{N,\nu}(\boldsymbol{\sigma}|\boldsymbol{\alpha},\boldsymbol{k},\mathbf{h},\boldsymbol{\xi}) = -\frac{N}{2}\left(\sum_{\mu=1}^{P}\sum_{a=1}^{\nu} k_a \alpha_a^2 (m_a^{\mu})^2 + \sum_{\mu=1}^{P}\sum_{a,b=1}^{\nu} \alpha_a \alpha_b m_a^{\mu} m_b^{\mu}\right) - N\sum_{\mu=1}^{P}\sum_{a=1}^{\nu}\alpha_a m_a^{\mu} h_a^{\mu}. \quad (3.5)$$

Again, setting $\nu = 1$ or $\boldsymbol{k} = \mathbf{1}$ (and homogeneous fields $h_a^{\mu} = h^{\mu}$ for each species), we recover (2.8). As before we shall drop the dependence on the system parameters whenever unnecessary.

In (3.4) we exploited the definition (3.2) to write $H_{N,\nu}(\boldsymbol{\sigma})$ in terms of the Mattis magnetizations: notice that the free field Hamiltonian is a quadratic form of $\mathbf{m}$ and can be written as

$$H_{N,\nu}(\boldsymbol{\sigma}) = -\frac{N}{2}\sum_{\mu=1}^{P}(\mathbf{m}^{\mu}, \mathbf{J}\mathbf{m}^{\mu}) = -\frac{N}{2}\mathbf{m}^{\mathrm{T}}\,\mathbf{J}\,\mathbf{m}, \quad (3.6)$$

where $\mathbf{m}^{\mu} = (m_1^{\mu}, \ldots, m_{\nu}^{\mu})$, $\mathbf{m}$ is the $\nu \times P$ matrix whose columns are $\mathbf{m}^{\mu}$ and $\mathbf{J}$ is the interaction matrix with entries

$$\mathbf{J}_{ab} = \begin{cases} k_a \alpha_a^2 & \text{if} \quad a = b \\ \alpha_a \alpha_b & \text{if} \quad a \neq b. \end{cases} \quad (3.7)$$

Analogously to the standard Hopfield model, we define the partition function and the pressure of the model as

$$Z_{N,\nu}(\beta,\boldsymbol{\alpha},\boldsymbol{k},\boldsymbol{h}|\boldsymbol{\xi}) = \sum_{\boldsymbol{\sigma}}\exp\{-\beta H_{N,\nu}(\boldsymbol{\sigma})\}; \quad f_{N,\nu}(\beta,\boldsymbol{\alpha},\boldsymbol{k},\boldsymbol{h}|\boldsymbol{\xi}) = \frac{1}{N}\ln Z_{N,\nu}(\beta,\boldsymbol{\alpha},\boldsymbol{k},\boldsymbol{h}|\boldsymbol{\xi}). \quad (3.8)$$

The interaction matrix (3.7) has, in general, not definite sign for small values of k_a, meaning higher disassortativity in the interactions among groups. This lack of convexity makes the analysis of the free energy highly non-trivial, since the rigorous techniques typically adopted strongly rely on the control of the order parameters fluctuations, the latter being an easy task using convexity properties (see e.g., [57, 58]). The main result of this paper is the following

Theorem 1. *Let $c > 1 + \dfrac{\nu - 1}{k^*}$, where $k^* = \min\{k_a\}_{a=1}^{\nu}$ and $\boldsymbol{J}^c = c\ \boldsymbol{diag}(k_1\alpha_1^2, \ldots, k_\nu\alpha_\nu^2) - \boldsymbol{J}$. The thermodynamic limit of the pressure of the multi-species Hopfield model described by the Hamiltonian (3.5) is given a.s. by*

$$\begin{aligned} f_{\nu}(\beta,\boldsymbol{\alpha},\boldsymbol{k},\boldsymbol{h}) &= \lim_{N\to\infty}\frac{1}{N}\ln Z_{N,\nu}(\beta,\boldsymbol{\alpha},\boldsymbol{k},\boldsymbol{h}|\boldsymbol{\xi}) \\ &= \sup_{\boldsymbol{m}\in\mathbb{R}^{\nu\times P}}\left(\beta\frac{\boldsymbol{m}^T\boldsymbol{J}^c\boldsymbol{m}}{2} + \sum_{a=1}^{\nu}\alpha_a f(\beta c k_a \alpha_a\ ,\beta\boldsymbol{h}_a\ -(\beta(\boldsymbol{A}^{-1}\boldsymbol{J}^c\boldsymbol{m})_a))\right). \end{aligned} \quad (3.9)$$

where $f(t,\boldsymbol{y})$ represents the pressure of the standard Hopfield model at a given temperature t and with a suitable external field $\boldsymbol{y}$, and $\mathbf{A} = \boldsymbol{diag}(\alpha_1, \ldots, \alpha_\nu)$. The optimal order parameters $\boldsymbol{m}$, are the solution of the self-consistent equations

$$\boldsymbol{m}_a = \left\langle \boldsymbol{\xi}\tanh\left[\beta\boldsymbol{\xi}\cdot(\boldsymbol{A}^{-1}\boldsymbol{J}\boldsymbol{m}+\boldsymbol{h})_a\right]\right\rangle_{\boldsymbol{\xi}\in\{-1,1\}^P}. \quad (3.10)$$

As a consistency check, by setting $k_a = 1$, $\boldsymbol{h}_a = \boldsymbol{h}$ and posing $M^{\mu} = \sum_{a=1}^{\nu} m_a^{\mu}\alpha_a$ for any $a = 1, \ldots, \nu$ and $\mu = 1, \ldots, P$, we get

$$\boldsymbol{M} = \sum_{a=1}^{\nu}\alpha_a\left\langle \boldsymbol{\xi}\tanh[\beta\boldsymbol{\xi}\cdot(\boldsymbol{M}+\boldsymbol{h})]\right\rangle_{\boldsymbol{\xi}} = \left\langle \boldsymbol{\xi}\tanh[\beta\boldsymbol{\xi}\cdot(\boldsymbol{M}+\boldsymbol{h})]\right\rangle_{\boldsymbol{\xi}}, \quad (3.11)$$

which are the standard Hopfield magnetizations (2.15) as expected. In general, plugging the solution of (3.10) into (3.9) we get

$$f_{\nu}(\beta,\boldsymbol{\alpha},\boldsymbol{k},\boldsymbol{h}) = \ln 2 + \sum_{a=1}^{\nu}\alpha_a\left\langle\log\cosh(\beta\boldsymbol{\xi}\cdot(\mathbf{A}^{-1}\mathbf{J}\mathbf{m}+\boldsymbol{h})_a)\right\rangle_{\boldsymbol{\xi}} - \beta\frac{\mathbf{m}^T\mathbf{J}\mathbf{m}}{2}. \quad (3.12)$$

Taking the derivative of (3.12) with respect to the external fields, one can see that the solution of (3.10) has the meaning of the averaged Mattis magnetization in the thermodynamic limit. It is important to

note that (3.12) does not depend on c as it should. In fact, c is just needed to get the correct convexity of the variational principle (3.9), i.e. it folds the trial free energy leaving its extremal point unchanged.

In the last section we will present the proof of Theorem 1 based on the introduction of a suitable perturbation, induced by c, in such a way that the resulting Hamiltonian becomes a convex form of the **m**'s. This way we will be able to map the problem into a mechanical framework in such a way that the free energy can be obtained as a solution of a Hamilton-Jacobi equation. Before proving the Theorem (see Sec. 5), in the next Section we present some special cases of the Multi-species Hopfield mode.

4. Special cases

In this section we show two particular cases for the multi-species Hopfield model (3.3), namely the bi-directional associative memory (BAM) model and the three-layer Restricted Boltzmann Machine (RBM), two interesting generalization of the Hopfield model introduced in the context of, respectively, biological neural networks and artificial neural networks and Deep Learning.

4.1. **The BAM model.** As we mentioned in Sec. 2, the Hopfield model reproduces the ability of the human brain to recall information from corrupted or partial data, yet human memory has many other characteristics. Among these is the ability to recall information coupled to one another: for example, if we remember the surname of someone we know, listening to that surname, even in the presence of some errors, we will probably be able to recall both the surname and the name of this person. The Hopfield model does not show this kind of behavior. The bi-directional associative memory (BAM) model, introduced by Kosko [18] in 1988, is the simplest network with this property: it is a neural network capable of storing and associating pairs of data, encoded by pairs of binary strings, of length N and M respectively, using a network of $N + M$ neurons, as we are going to review briefly.

In this model one distinguishes two different types of neurons: the *input neurons* σ_i, labelled by $i \in \{1, ..., N\}$, and the *output neurons* σ_i, labelled by $i \in \{1 + N, ..., M + N\}$. The connections between the input and output neurons depend on the P pairs of patterns $(\boldsymbol{\xi}, \boldsymbol{\eta})$ with N and M components, respectively. Each component ξ_i^μ, $i = 1, ..., N, \mu = 1, ..., P$ and η_i^μ, $i = N + 1, ..., N + M, \mu = 1, ..., P$ are independent random variables and identically distributed with zero average. The synaptic matrix of the interactions between the spin $i \in \{1, ..., N\}$ and the spin $j \in \{N + 1, ..., N + M\}$ is given by

$$J_{ij} = \begin{cases} \frac{1}{\sqrt{MN}} \sum_{\mu=1}^{P} \xi_i^\mu \eta_j^\mu & \text{if } i \in \{1, ..., N\} \text{ and } j \in \{N + 1, ..., N + M\}. \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

Each set of neuron can be associated to an order parameter, referred to as **m** and **n**, respectively, whose μ-th entry quantifies the overlap between the spin configuration and the related patterns, that is

$$m_\mu = \frac{1}{N} \sum_{i=1}^{N} \sigma_i \xi_i^\mu, \ n_\mu = \frac{1}{M} \sum_{i=N+1}^{N+M} \sigma_i \eta_i^\mu \tag{4.2}$$

The Hamiltonian describing this system reads as

$$H_{N,M}(\boldsymbol{\sigma}|\boldsymbol{\xi}, \boldsymbol{\eta}) = \frac{1}{\sqrt{NM}} \sum_{\mu=1}^{P} \sum_{i=1}^{N} \sum_{j=1}^{M} \xi_i^\mu \eta_j^\mu \sigma_i \sigma_j = \sqrt{MN} \sum_{\mu=1}^{P} m_\mu n_\mu, \tag{4.3}$$

and the dynamics (2.1) can be adopted straightforwardly to make the system relax toward an equilibrium configuration described by the related Boltzmann-Gibbs distribution. The recall phase occurs analogously as for the Hopfield model: if, given a certain initial point, the spin configuration converges to a state such that $\sigma_i = \xi_i^\mu \ \forall i = 1, ..., N$ and $\sigma_i = \eta_i^\mu \ \forall i = N + 1, ..., N + M$, the information, codified by the strings $\boldsymbol{\xi}^\mu$ and $\boldsymbol{\eta}^\mu$, is retrieved. The BAM model was investigated extensively by [19, 20] addressing the low-storage (i.e., P/M and P/N both vanishing in the thermodynamic limit) as well as the high-load (i.e., P/M and P/N both finite in the thermodynamic limit) regimes. In particular, in the former case they found that the order parameters fulfil the following self-consistent equations

$$\mathbf{m} = \langle \boldsymbol{\xi} \tanh(\beta \sqrt{\tfrac{M}{N}} \ \boldsymbol{\xi} \cdot \mathbf{n}) \rangle_{\boldsymbol{\xi}}, \tag{4.4}$$

$$\mathbf{n} = \langle \boldsymbol{\eta} \tanh(\beta \sqrt{\tfrac{N}{M}} \ \boldsymbol{\eta} \cdot \mathbf{m}) \rangle_{\boldsymbol{\eta}}, \tag{4.5}$$
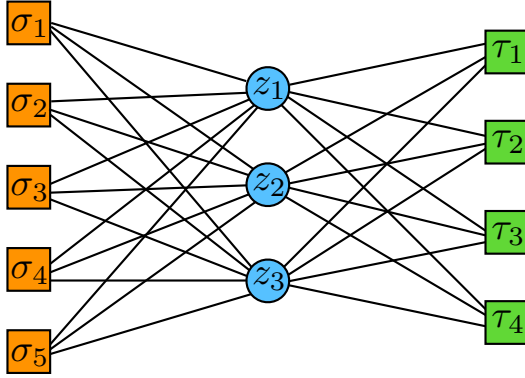
FIGURE 1. Sketch of a RBM with three layers. The visible layer (on the left) is made of $N = 5$ spins (σ_i, $i = 1, ..., N$), the output layer (on the right) is made of $M = 4$ spins (σ_i, $i = 1, ..., N$), and the hidden layer (in the middle) is made of $P = 3$ spins (z_i, $i = 1, ..., P$). In our analysis we assumed boolean visible and output neurons ($\boldsymbol{\sigma} \in \{-1, +1\}^N$ and $\boldsymbol{\tau} \in \{-1, +1\}^M$) while Gaussian neurons ($\boldsymbol{z} \in \mathbb{R}^P$) in the hidden layer.

and, in the latter case, they derive an expression for the critical capacity, namely the maximum number of patterns that the network is able to safely handle: this quantity is shown to depend on the ratio M/N and it is maximum for $M = N$[3]. Further results can be found in [59, 60, 61, 62, 63, 64].

The Hamiltonian (4.3) can be recast in the general description provided by (3.3), by setting $\nu = 2$ and $k_1 = k_2 = 0$. Otherwise stated, we are considering two groups where interactions only occur between neurons belonging to different groups. Moreover, $\boldsymbol{\xi}^\mu = \boldsymbol{\xi}^{\mu,1}$ and $\boldsymbol{\eta}^\mu = \boldsymbol{\xi}^{\mu,2}$; $\boldsymbol{\sigma}^1 = \{\sigma_1, ..., \sigma_N\}$ and $\boldsymbol{\sigma}^2 = \{\sigma_{N+1}, ..., \sigma_{N+M}\}$; $\alpha_1 = \frac{N}{N+M}$ and $\alpha_2 = \frac{M}{N+M}$. By setting these parameters in the self-consistent equations (3.10), simply rescaling the temperature as $\beta \to \beta/\sqrt{\alpha_1 \alpha_2}$ because of the different normalizations in (4.3) and (3.4) we exactly recover (4.4) and (4.5).

4.2. **Three-layer Restricted Boltzmann Machines.** In Sec. 2 we presented the Hopfield network as a standard model for associative memory, while the basic model for neural networks learning processes is the Boltzmann machine (BM). Remarkably, retrieval and learning represent two complementary aspects of the cognitive process. In particular, the Hebb rule (2.5) can be shown to emerge from a learning process accomplished through a BM [11, 31]. The recent success of Deep Learning in a wide range of different applicative sciences poses the attention on a specific class of BMs, the Deep Boltzmann Machines [1]. They are composed of multiple layers of units (neurons), where units belonging to any two consecutive layers are connected. The elementary component of this architecture is thus a particular two layers BM, also referred to as Restricted Boltzmann Machine (RBM).

More precisely, let us consider a group of N binary units $\sigma_i \in \{-1, +1\}$, $i \in \{1, ..., N\}$, which represent the neurons of the first layer (also called visible), a group of units z_μ, $\mu \in \{1, ..., P\}$, which represent the neurons of the second layer (also called hidden), and a set of weights w_i^μ associated to the links connecting the neurons σ_i and z_μ. For the scope of this section we consider the hidden units to be real variables, distributed with a Gaussian prior. All the results can be suitably generalized to the case of different priors [12, 13]. We can then define the Hamiltonian of the restricted Boltzmann machine as

$$H_N(\boldsymbol{\sigma}, \mathbf{z}|\mathbf{w}) = -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{\mu=1}^{P} w_i^\mu \sigma_i z_\mu. \qquad (4.6)$$

The associated Gibbs distribution is used typically to fit some data distribution over the visible layer, optimizing over the weights. Thus a RBM can be used as a parametric probabilistic model in an unsupervised learning framework [66]. For a given realization of the weights the related intensive pressure

---

[3]The critical threshold for the BAM model occurs to be lower than that for the standard Hopfield model and this result may be explained by noticing that storing a pair of patterns corresponds to doubling the information stored per pattern and therefore the total information stored in the $(N + M)^2/2$ synaptic connections.

of a RBM reads as

$$f_N(\beta|\mathbf{w}) = \frac{1}{N}\ln\sum_{\boldsymbol{\sigma}}\prod_{\mu}\int_{-\infty}^{\infty}dz_{\mu}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z_{\mu}^2}{2}\right)\exp\left(\frac{\beta}{\sqrt{N}}\sum_{\mu}\sum_{i}w_i^{\mu}\sigma_i z_{\mu}\right). \tag{4.7}$$

Since there are no weights between neurons belonging to the same layer, we can marginalize over the z variables by evaluating the Gaussian integral obtaining

$$f_N(\beta|\mathbf{w}) = \frac{1}{N}\ln\sum_{\boldsymbol{\sigma}}\exp\left\{\frac{\beta^2}{2N}\sum_{i,j=1}^{N}\left(\sum_{\mu=1}^{P}w_i^{\mu}w_j^{\mu}\right)\sigma_i\sigma_j\right\}.$$

Comparing this result with the pressure of the standard Hopfield model (2.7), we can see that the two expressions are equivalent (apart for the value of the temperature) identifying the weights of the RBM $\mathbf{w}^{\mu}$ with the patterns $\boldsymbol{\xi}^{\mu}$ of the Hopfield network. This mapping was used in [12, 13] for studying the performances of RBM's in learning processes using the knowledge of the Hopfield model phase diagram. Hereafter, we use the same argument to show that a 3-layer RBM is equivalent to a 2-species Hopfield model. This is in agreement with a recent finding by M.Mezard [65] that has shown, on the other hand, how a 2-layers RBM with combinatorial weights can be seen as a 3-layers RBM.

Let us consider a RBM made of three layers: let σ_i, $i \in \{1,\ldots,N\}$, the neurons belonging to the first (visible) layer, z_{μ}, $\mu \in \{1,\ldots,P\}$, the neurons belonging to the hidden layer and τ_k, $k \in \{1,\ldots,M\}$, the neurons belonging to the third (output) layer, as sketched in Fig. 1. We focus on the regime $P \ll N, M$ and we take $\sigma_i, \tau_k \in \{-1, +1\}$ and $z_{\mu} \in \mathbb{R}$ as Gaussian variables for simplicity. This kind of structure is often referred to as an autoencoder. The Boltzmann-Gibbs distribution associated to this system is

$$p_{M,N}(\boldsymbol{\sigma},\boldsymbol{\tau},\mathbf{z}|\boldsymbol{\xi},\boldsymbol{\eta}) = \frac{1}{Z_{M,N}}\prod_{\mu}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z_{\mu}^2}{2}\right)\exp\left\{\frac{\sqrt{\beta}}{\sqrt{N}}\sum_{i=1}^{N}\sum_{\mu=1}^{P}\xi_i^{\mu}\sigma_i z_{\mu} + \frac{\sqrt{\beta}}{\sqrt{M}}\sum_{k=1}^{M}\sum_{\mu=1}^{P}\eta_k^{\mu}\tau_k z_{\mu}\right\}. \tag{4.8}$$

Since neurons belonging to the same layer are not interacting each other, we can marginalize the joint probability (4.8) over z exploiting the Gaussian integration as

$$\begin{aligned}p_{M,N}(\boldsymbol{\sigma},\boldsymbol{\tau}|\boldsymbol{\xi},\boldsymbol{\eta}) &= \frac{1}{Z_{M,N}}\exp\left\{\frac{\beta}{2N}\sum_{\mu=1}^{P}\sum_{i,j=1}^{N,N}\xi_i^{\mu}\xi_j^{\mu}\sigma_i\sigma_j + \frac{\beta}{2M}\sum_{\mu=1}^{P}\sum_{k,l=1}^{M,M}\eta_k^{\mu}\eta_l^{\mu}\tau_k\tau_l\right.\\ &+\left. \frac{\beta}{\sqrt{NM}}\sum_{\mu=1}^{P}\sum_{i,k=1}^{N,M}\xi_i^{\mu}\eta_k^{\mu}\sigma_i\tau_k\right\}. \end{aligned} \tag{4.9}$$

Comparing (4.9) with (2.2) one can see that the 3-layers RBM considered here is equivalent to a 2-species Hopfiled model with Hamiltonian

$$H_{M,N}(\boldsymbol{\sigma},\boldsymbol{\tau}) = -\left(\frac{1}{2N}\sum_{\mu=1}^{P}\sum_{i,j=1}^{N,N}\xi_i^{\mu}\xi_j^{\mu}\sigma_i\sigma_j + \frac{1}{2M}\sum_{\mu=1}^{P}\sum_{k,l=1}^{M,M}\eta_k^{\mu}\eta_l^{\mu}\tau_k\tau_l + \frac{1}{\sqrt{NM}}\sum_{\mu=1}^{P}\sum_{i,k=1}^{N,M}\xi_i^{\mu}\eta_k^{\mu}\sigma_i\tau_k\right). \tag{4.10}$$

Introducing the two sets of order parameters

$$m_{\mu} = \frac{1}{N}\sum_{i=1}^{N}\xi_i^{\mu}\sigma_i, \quad n_{\mu} = \frac{1}{M}\sum_{k=1}^{M}\eta_k^{\mu}\tau_k. \tag{4.11}$$

the expression (4.10) can be recast as

$$H_{M,N}(\mathbf{m},\mathbf{n}) = -\frac{N}{2}\sum_{\mu=1}^{P}m_{\mu}^2 - \frac{M}{2}\sum_{\mu=1}^{P}n_{\mu}^2 - \sqrt{NM}\sum_{\mu=1}^{P}m_{\mu}n_{\mu}, \tag{4.12}$$

which corresponds to a multi-species Hopfield model (3.3) as long as we set $\nu = 2$, $\beta \to \beta(M+N)/\sqrt{MN}$, $k_1 = \sqrt{M/N}$ and $k_2 = \sqrt{N/M}$. The solution of this system can therefore be derived from (3.9) by properly setting the parameters. To fix ideas we assume, without loss of generality, that $M = \gamma N$, where $\gamma \in \mathbb{R}^+$. Thus, recalling (3.10), we find

$$\mathbf{m} = \left\langle \boldsymbol{\xi}\tanh\left[\beta\sum_{\mu=1}^{P}(m_{\mu} + \sqrt{\gamma}n_{\mu})\cdot\boldsymbol{\xi}\right]\right\rangle_{\boldsymbol{\xi}}, \tag{4.13}$$
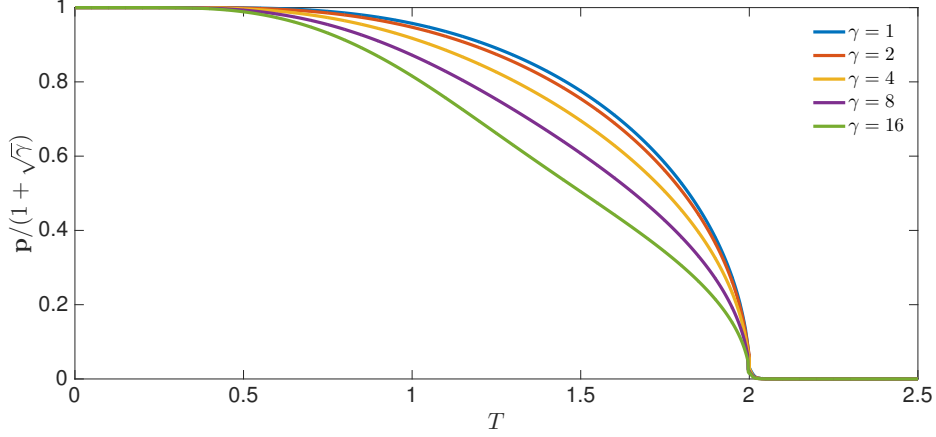
FIGURE 2. This plot shows the behavior of the normalized overlap $p_\mu/(1+\sqrt{\gamma})$ related to the retrieved pattern, as a function of the temperature T, obtained by solving the self-consistent equation (4.15) for $P = 2$. In this case one pattern, say $\mu = 1$ to fix ideas, is retrieved and the related overlap follows the outline shown in this figure; the remaining pattern $\mu = 2$ is not retrieved and the related overlap is null for any T. Several values of the parameter γ are considered and shown in different colors, as explained by the legend.

$$\mathbf{n} = \left\langle \boldsymbol{\eta} \tanh \left[\frac{\beta}{\sqrt{\gamma}} \sum_{\mu=1}^{P} (m_\mu + \sqrt{\gamma}\, n_\mu) \cdot \boldsymbol{\eta} \right] \right\rangle_{\boldsymbol{\eta}} . \tag{4.14}$$

The structure of (4.13) and (4.14) suggests that we can linearly combine $\mathbf{m}$ and $\mathbf{n}$ to get a unique order parameter, referred to as $\mathbf{p}$ and whose μ-th component reads as $p_\mu = m_\mu + \sqrt{\gamma}\, n_\mu$, $\mu = 1, ..., P$. As a result, (4.13) and (4.14) can be recast as

$$\mathbf{p} = \langle \boldsymbol{\xi} \tanh(\beta \boldsymbol{\xi} \cdot \mathbf{p}) \rangle_{\boldsymbol{\xi}} + \sqrt{\gamma} \left\langle \boldsymbol{\eta} \tanh \left(\frac{\beta}{\sqrt{\gamma}} \boldsymbol{\eta} \cdot \mathbf{p} \right) \right\rangle_{\boldsymbol{\eta}} \tag{4.15}$$

This self-consistent equation was solved numerical and results are shown in Fig.. 2. These plots suggest a phase transition at $T_c = 2$, such that, as $T > T_c$, the magnetization vanishes and retrieval is no longer achievable. In fact let us notice that

$$\mathbf{p}^2 = \left\langle \boldsymbol{\xi} \tanh(\beta \boldsymbol{\xi} \cdot \mathbf{p}) \right\rangle_{\boldsymbol{\xi}} \cdot \mathbf{p} + \sqrt{\gamma} \left\langle \boldsymbol{\eta} \tanh \left(\frac{\beta}{\sqrt{\gamma}} \boldsymbol{\eta} \cdot \mathbf{p} \right) \right\rangle_{\boldsymbol{\eta}} \cdot \mathbf{p} \tag{4.16}$$

$$= \left\langle \boldsymbol{\xi} \cdot \mathbf{p} \tanh(\beta \boldsymbol{\xi} \cdot \mathbf{p}) \right\rangle_{\boldsymbol{\xi}} + \sqrt{\gamma} \left\langle \boldsymbol{\eta} \cdot \mathbf{p} \tanh \left(\frac{\beta}{\sqrt{\gamma}} \boldsymbol{\eta} \cdot \mathbf{p} \right) \right\rangle_{\boldsymbol{\eta}},$$

and, recalling that $|\tanh(z)| \le |z|$, we get

$$\mathbf{p}^2 \le \beta \big(\langle (\boldsymbol{\xi} \cdot \mathbf{p})^2 \rangle_{\boldsymbol{\xi}} + \langle (\boldsymbol{\eta} \cdot \mathbf{p})^2 \rangle_{\boldsymbol{\eta}} \big) = 2\beta \langle (\boldsymbol{\lambda} \cdot \mathbf{p})^2 \rangle_{\boldsymbol{\lambda}} = 2\beta \mathbf{p}^2. \tag{4.17}$$

We can therefore conclude that as $\beta < \frac{1}{2}$, namely as $T > 2$, the only equilibrium solution for the problem is $\mathbf{p} = \mathbf{0}$. We want to check further that the critical temperature of the model is $T_c = 2$. Since

$$\langle \xi^\mu \rangle_{\boldsymbol{\xi}} = 0 \quad \text{e} \quad \langle \xi^\mu \xi^\nu \rangle_{\boldsymbol{\xi}} = \delta_{\mu\nu} \quad \langle \xi^\mu \xi^\nu \xi^\rho \xi^\sigma \rangle_{\boldsymbol{\xi}} = \delta_{\mu\nu}\delta_{\rho\sigma} + \delta_{\mu\rho}\delta_{\nu\sigma} + \delta_{\mu\sigma}\delta_{\rho\nu} - 2\delta_{\mu\rho}\delta_{\rho\nu}\delta_{\nu\sigma},$$

expanding (4.15) around $p_\mu = 0$ we get

$$p_\mu = \beta p_\mu + \beta^3 p_\mu \big(-\mathbf{p}^2 + \frac{2}{3} p_\mu^2 \big) + \beta p_\mu + \frac{\beta^3}{\gamma} p_\mu \big(-\mathbf{p}^2 + \frac{2}{3} p_\mu^2 \big).$$

Asking for a pure state retrieval, namely $\mathbf{p}^2 = p_\mu^2$, we get

$$p_\mu(2\beta - 1) - \beta^3 p_\mu \left(1 + \frac{1}{\gamma} \right) \frac{1}{3} p_\mu^2 = 0.$$

9

Replacing $\beta = 1/T$, the non trivial solutions are

$$p_\mu = \pm \sqrt{\frac{3(T-2)}{1+\frac{1}{\gamma}}}, \tag{4.18}$$

i.e. the critical temperature is $T_c = 2$. Moreover, one can see that p_μ scales as $\sqrt{T-2}$ and therefore, the critical exponent turns out to be $1/2$, which suggests that the phase transition is of the second order.

5. Solution of the multi-species Hopfield model

We investigate the model introduced in the previous section focusing on the low-load regime, namely assuming $\lim_{N\to\infty} P/N = 0$. As underlined in Section 3, the Hamiltonian (3.3) under study is a quadratic form with non-definite sign. However, we can "adjust" it by inserting a suitable, additional term that enhances the diagonal entries of the matrix $\mathbf{J}$ (3.7), in such a way that the contributions due to intra-group interactions are prevailing and the matrix gets positive definite. Being c a real, strictly positive parameter, we define

$$\mathbf{J}^c = c \, \mathbf{diag}(k_1\alpha_1^2, ..., k_\nu\alpha_\nu^2) - \mathbf{J}. \tag{5.1}$$

Thus it holds

Lemma 1. *As soon as $c > 1 + \dfrac{\nu-1}{k^*}$, where $k^* = \min\{k_a\}_{a=1}^\nu$, $\mathbf{J}^c$ is positive definite. Then, the matrix $\mathbf{J}^c$ can be written as $\mathbf{J}^c = \mathbf{P}^T\mathbf{P}$, where $\mathbf{P}$ is the matrix whose rows are given by the normalized eigenvectors of $\mathbf{J}^c$.*

The proof of this lemma can be found in Appendix A.

Proof. of Theorem 1.

Let $c > 1 + \dfrac{\nu-1}{k^*}$ and let us transform the order parameters $\mathbf{m}$ according to the rotation defined by $\mathbf{P}$ as in Lemma (1), namely $\mathbf{m}' = \mathbf{Pm}$, or, more explicitly, $m_a^{\mu\prime} = \sum_{k=1}^\nu P_{ak}m_k^\mu$. Following [27, 58, 67, 68, 69] let us also introduce the interpolating function $\Phi_{N,\nu}(t,\mathbf{x})$, where the variables $t \in \mathbb{R}^+$ and $\mathbf{x} \in \mathbb{R}^{\nu \times P}$ are meant, respectively, as generalized time and space:

$$\Phi_{N,\nu}(t,\mathbf{x}) = \frac{1}{N}\ln\sum_{\boldsymbol{\sigma}}\exp\left\{\frac{Nt}{2}\left(\sum_{\mu=1}^P\sum_{a=1}^\nu k_a\alpha_a^2(m_a^\mu)^2 + \sum_{\mu=1}^P\sum_{a\neq b}^{\nu,\nu}\alpha_a\alpha_b m_a^\mu m_b^\mu\right)\right.$$
$$\left. + N\frac{(\beta-t)c}{2}\sum_{\mu=1}^P\sum_{a=1}^\nu k_a\alpha_a^2 m_a^2 + N\sum_{\mu=1}^P\sum_{a=1}^\nu x_a^\mu m_a^{\mu\prime}\right\}. \tag{5.2}$$

In order to lighten the notation we also introduce the interpolating *Boltzmann factor* $B_{N,\nu}(t,\mathbf{x})$ such that $\Phi_{N,\nu}(t,\mathbf{x}) = \frac{1}{N}\ln\sum_{\boldsymbol{\sigma}}B_{N,\nu}(t,\mathbf{x})$ and the average performed with respect to $B_{N,\nu}(t,\mathbf{x})$ shall be denoted as $\langle\cdot\rangle_{(t,\mathbf{x})}$. The interpolating partition function $Z_{N,\nu}(t,\mathbf{x})$ is defined analogously. Notice that, when $t = \beta$ and $\mathbf{x}(\mathbf{h}) = \beta(\mathbf{P}^T)^{-1}\mathbf{Ah}$, we recover the original model described by (3.5), i.e.

$$f_\nu(\beta,\boldsymbol{\alpha},\boldsymbol{k},\mathbf{h}) = \lim_{N\to\infty}\Phi_{N,\nu}(\beta,\beta(\mathbf{P}^T)^{-1}\mathbf{Ah}), \tag{5.3}$$

and the average $\langle\cdot\rangle_{(t,\mathbf{x})}$ recovers the expectation over the standard Boltzmann-Gibbs measure.
We now show that $\Phi_{N,\nu}(t,\mathbf{x})$ fulfills an Hamilton-Jacobi equation; to this task we first evaluate the derivate of (5.2) with respect to t

$$\partial_t\Phi_{N,\nu}(t,\mathbf{x}) = \frac{1}{2Z_{N,\nu}(t,\mathbf{x})}\sum_{\boldsymbol{\sigma}}\left[(1-c)\sum_{\mu=1}^P\sum_{a=1}^\nu k_a\alpha_a^2 m_a^{\mu 2} + \sum_{\mu=1}^P\sum_{a\neq b}^{\nu,\nu}\alpha_a\alpha_b m_a^\mu m_b^\mu\right]B_{N,\nu}(t,\mathbf{x}) \tag{5.4}$$

$$= -\frac{1}{2}\langle\mathbf{m},\mathbf{J}^c\mathbf{m}\rangle = -\frac{1}{2}\langle\mathbf{Pm},\mathbf{Pm}\rangle = -\frac{1}{2}\sum_{a=1}^\nu\sum_{\mu=1}^P\langle(m_a^{\mu\prime})^2\rangle, \tag{5.5}$$

and the derivative with respect to x_a^μ , $a = 1, \ldots, \nu$:

$$\partial_{x_a^\mu}\Phi_{N,\nu}(t,\mathbf{x}) = \frac{1}{Z_{N,\nu}(t,\mathbf{x})}\sum_{\boldsymbol{\sigma}}m_a^{\mu\prime}B_{N,\nu}(t,\mathbf{x}) = \langle m_a^{\mu\prime}\rangle. \tag{5.6}$$

Combining (5.5)-(5.8), we see that the following Hamilton-Jacobi equation for $\Phi_{N,\nu}(t, \mathbf{x})$ holds by construction

$$\partial_t \Phi_{N,\nu}(t, \mathbf{x}) + \frac{1}{2} |\nabla \Phi_{N,\nu}(t, \mathbf{x})|^2 + V_N(t, \mathbf{x}) = 0. \tag{5.7}$$

where we have introduced the potential $V_N(t, \mathbf{x})$ as the sum of fluctuations of the rotated magnetizations with respect to the average $\langle \cdot \rangle_{(t,\mathbf{x})}$ as

$$V_N(t, \mathbf{x}) = \frac{1}{2N} \Delta \Phi_{N,\nu}(t, \mathbf{x}) = \frac{1}{2} \sum_{a=1}^{\nu} \sum_{\mu=1}^{P} [\langle (m_a^{\mu\prime})^2 \rangle - \langle m_a^{\mu\prime} \rangle^2]. \tag{5.8}$$

Otherwise stated, $\Phi_{N,\nu}(t, \mathbf{x})$ plays as the action potential for a particle moving in $\mathbb{R}^{\nu \times P}$ in the presence of a potential $V_N(t, \mathbf{x})$. In this mechanical analogy, $\partial_{x_a^\mu} \Phi_{N,\nu}(x, t)$ plays as the velocity of the particle. As $N \to \infty$, the solution of this PDE can be shown to approach the unique viscosity solution of the free Hamilton-Jacobi equation, given by the Hopf-Lax formula (see for instance [70, 71])

$$\Phi_\nu(t, \mathbf{x}) := \lim_{N \to \infty} \Phi_{N,\nu}(t, \mathbf{x}) = \min_{\mathbf{y} \in \mathbb{R}^{\nu \times P}} \left(-\frac{(\mathbf{x} - \mathbf{y})^2}{2t} + \Phi_{N,\nu}(0, \mathbf{y}) \right). \tag{5.9}$$

Now, since as $N \to \infty$ the motion is free, we can relate the position $\mathbf{y}$ at time $t = 0$ to the position $\mathbf{x}$ at the arbitrary time t by $\mathbf{y} = \mathbf{x} - t\,\mathbf{m}'$, that is

$$\Phi_\nu(t, \mathbf{x}) = \min_{\mathbf{m}' \in \mathbb{R}^{\nu \times P}} \left(-\frac{t(\mathbf{m}')^2}{2} + \Phi_{N,\nu}(0, \mathbf{x} - t\,\mathbf{m}') \right). \tag{5.10}$$

We can calculate explicitly $\Phi_{N,\nu}(0, \mathbf{y})$ from (5.2) as

$$\Phi_{N,\nu}(0, \mathbf{y}) = \frac{1}{N} \ln \sum_{\boldsymbol{\sigma}} \exp \left\{ \frac{N}{2} \sum_{a=1}^{\nu} \sum_{\mu=1}^{P} c\beta k_a \alpha_a^2 \,(m_a^\mu)^2 + N \sum_{a=1}^{\nu} \sum_{\mu=1}^{P} y_a^\mu (\mathbf{P}\mathbf{m}^\mu)_a \right\}$$

$$= \sum_{a=1}^{\nu} \frac{\alpha_a}{N_a} \ln \sum_{\boldsymbol{\sigma}_a} \exp \left\{ \frac{N_a}{2} \sum_{\mu=1}^{P} c\beta k_a \alpha_a \,(m_a^\mu)^2 + N_a \sum_{\mu=1}^{P} (\mathbf{A}^{-1} \mathbf{P}^T \boldsymbol{y}^\mu)_a m_a^\mu \right\}, \tag{5.11}$$

where in the last line we highlighted the contribution stemming from each group. In fact, (5.11) turns out to be the linear combination of the pressures of standard Hopfield models of different sizes, at different temperatures and different external fields. This can be seen recalling (2.9) and (2.11). In particular, in this configuration, the a-th term corresponds to a Hopfield model with inverse temperature $\beta c k_a \alpha_a$ and external field $((\mathbf{A})^{-1} \mathbf{P}^T \mathbf{y})_a$ Thus, in the thermodynamic limit, it holds

$$\Phi_\nu(0, \mathbf{x} - t\mathbf{m}') = \sum_{a=1}^{\nu} \alpha_a f(c\beta k_a \alpha_a \,, (\mathbf{A}^{-1} \mathbf{P}^T \mathbf{x})_a - t(\mathbf{A}^{-1} \mathbf{P}^T \mathbf{m}')_a). \tag{5.12}$$

Plugging (5.12) into (5.10), we get an expression for $\Phi_\nu(t, \boldsymbol{x})$ and, setting $t = \beta$ and $\mathbf{x}(\mathbf{h}) = \beta (\mathbf{P}^T)^{-1} \mathbf{A} \mathbf{h}$, we get the variational principle for the free energy (3.8)

$$f_\nu(\beta, \boldsymbol{\alpha}, \boldsymbol{\xi}, h) = \Phi_\nu(\beta, \mathbf{x}(\mathbf{h})) = \sup_{\mathbf{m}' \in \mathbb{R}^{\nu \times P}} \left[\frac{\beta(\mathbf{m}')^2}{2} + \sum_{a=1}^{\nu} \alpha_a f(\beta c k_a \alpha_a \,, \beta h_a \, - \beta (\mathbf{A}^{-1} \mathbf{P}^T \mathbf{m}')_a) \right]. \tag{5.13}$$

Coming back $\mathbf{m}' \to \mathbf{m}$, we can rewrite (5.13) as

$$f_\nu(\beta, \alpha, h) = \sup_{\mathbf{m} \in \mathbb{R}^{\nu \times P}} \left(\frac{\beta(\mathbf{m}, \mathbf{J}^c \mathbf{m})}{2} + \sum_{a=1}^{\nu} \alpha_a f(\beta c k_a \alpha_a \,, \beta h_a \, - (\beta (\mathbf{A}^{-1} \mathbf{J}^c \mathbf{m}^\mu)_a)) \right). \tag{5.14}$$

as stated in Theorem 1.

By differentiating the argument of (5.14) with respect to $\mathbf{m}$ we get

$$\mathbf{J}^c (\mathbf{m} - \mathbf{M}) = 0 \implies \mathbf{m} = \mathbf{M}, \tag{5.15}$$

since $\mathbf{J}^c$ is positive definite and where

$$\mathbf{M}_a = \nabla_{\boldsymbol{x}} f(\beta c k_a \alpha_a, \boldsymbol{x}) \big|_{\mathbf{x} = \beta \mathbf{h}_a - \beta (\mathbf{A}^{-1} \mathbf{J}^c \mathbf{m})_a}. \tag{5.16}$$

Otherwise stated, $\boldsymbol{M}_a$ is the expected magnetization for a standard Hopfield model with size N_a, set at temperature $(\beta c k_a \alpha_a)^{-1}$ and with external fields $\beta \mathbf{h}_a - \beta (\mathbf{A}^{-1} \mathbf{J}^c \mathbf{m})_a$. Recalling the self-consistent

equation (2.15) for the magnetization of the standard Hopfield model we get that $\boldsymbol{M}_a$ is solution of the following equation

$$
\begin{aligned}
\boldsymbol{M}_a &= \left\langle \boldsymbol{\xi} \tanh\left(\beta c k_a \alpha_a \ \boldsymbol{\xi} \cdot \boldsymbol{M}_a + \beta \boldsymbol{\xi} \cdot \boldsymbol{h}_a - \beta \ \boldsymbol{\xi} \cdot (\mathbf{A}^{-1}\mathbf{J}^c \mathbf{m})_a \right) \right\rangle_{\boldsymbol{\xi}} \\
&= \left\langle \boldsymbol{\xi} \tanh\left(\beta c k_a \alpha_a \boldsymbol{\xi} \cdot (\boldsymbol{M}_a - \boldsymbol{m}_a) + \beta \boldsymbol{\xi} \cdot (\mathbf{A}^{-1}\mathbf{J}\mathbf{m})_a + \beta \ \boldsymbol{\xi} \cdot \boldsymbol{h}_a \right) \right\rangle_{\boldsymbol{\xi}}.
\end{aligned}
\tag{5.17}
$$

Thus from (5.15) we get (3.10).

$\square$

6. Conclusions

In this work we considered multi-species Hopfield model in which different groups of neurons interact through a Hebbian-like coupling structure with different intensities. The model is completely solvable even in the non-convex region, where inter-groups couplings are stronger than those intra-groups. A variational principle for the free energy is introduced, whose solutions define also the possible thermodynamic states, described in terms of the Mattis magnetizations. The strategy generalizes the one introduced in [27] and is based on a suitable Hamiltonian convexification, together with an interpolating procedure, that allows the mapping with a mechanic problem, solved through an effective Hamilton-Jacobi equation.

The introduced model contains both the ingredients of the recent Deep Boltzmann machines: non-convexity and Hebbian structure. Actually the latter results from the partial marginalization over some groups of neurons in a non-convex structure and is at the root of the Deep Boltzmann Machine ability to store patterns of informations, with different levels of correlations. In fact we have shown two special cases of the model: the BAM, a two layer neural networks with Hebbian interactions, able to store couples of patterns and a 3-layers RBM with hidden gaussian units.

The analysis of this paper was performed in the low load regime, i.e. when the Hebbian structure is composed of a sub-extensive (with respect to the system size) number of patterns. This facilitates a lot the analysis because only the ferromagnetic nature of the interaction counts and the glassy states due to patterns interferences are not thermodynamic. The high load regime is still an open challenge: rigorous results for the standard (convex) Hopfield model lack, since rigorous results for the (non-convex) multi-species SK model are still incomplete. Non-convexity appears in many problems of optimization in statistical inference, from community detection in networks to low rank matrix factorization [72, 73] thus an effort in this direction is necessary.

Anyway one could start using non-rigorous statistical-physics techniques [4, 74] to study how the critical load of the standard Hopfield model changes in terms of the number of groups, their relative sizes, the shape of the intensity matrix and, not last, the type of the neurons (Boolean, Gaussian, etc.), being the last closely related to the activation function used in practical applications.

Before proving Lemma 1 we define the matrix $\tilde{\mathbf{J}}^c$ as

$$\tilde{\mathbf{J}}^c = c \, \mathbf{diag}(k^* \alpha_1^2, ..., k^* \alpha_\nu^2) - \mathbf{J}, \tag{A.1}$$

where k^* is defined as in Lemma (1), then it holds the following

Lemma 2. *The quadratic form stemming from (A.1) is positive definite if $c > 1 + \dfrac{\nu - 1}{k^*}$. Moreover, we get $\tilde{\mathbf{J}}^c = \tilde{\mathbf{P}}^T \tilde{\mathbf{P}}$, where the rows of $\tilde{\mathbf{P}}$ are the linearly independent vectors given by*

$$\boldsymbol{v}^1 = \frac{\sqrt{k^*(c-1) - (\nu - 1)}}{\sqrt{\nu}}(\alpha_1, ... , \alpha_\nu)$$

$$\boldsymbol{v}^2 = \frac{\sqrt{k^*(c-1) + 1}}{\sqrt{2}}(\alpha_1, -\alpha_2, 0, ... , 0)$$

$$\vdots$$

$$\boldsymbol{v}^a = \frac{\sqrt{k^*(c-1) + 1}}{\sqrt{a(a-1)}}(\alpha_1, ... , \alpha_{a-1}, \alpha_a - a\alpha_a, 0, ..., 0)$$

$$\vdots$$

$$\boldsymbol{v}^\nu = \frac{\sqrt{k^*(c-1) + 1}}{\sqrt{\nu(\nu - 1)}}(\alpha_1, ... , \alpha_{\nu-1}, \alpha_\nu - \nu\alpha_\nu).$$

Proof. Let $\mathbf{A} = \mathbf{diag}(\alpha_1, \ldots, \alpha_\nu)$ and define

$$\tilde{\mathbf{T}}^c_{ab} := \begin{cases} (c-1)k^* & \text{se} \quad a = b \\[2mm] -1 & \text{se} \quad a \neq b. \end{cases} \tag{A.2}$$

such that we can write $\tilde{\mathbf{J}}^c = \mathbf{A}\tilde{\mathbf{T}}\mathbf{A}$ We now derive the eigenvectors of $\tilde{\mathbf{T}}^c$. In order for $\mathbf{w}$ to be eigenvector with eigenvalue λ the following homogeneous system must be fulfilled

$$(\tilde{\mathbf{T}}^c - \lambda \mathbb{I}_\nu)\mathbf{w} = 0. \tag{A.3}$$

Focusing on the i-th component

$$((\tilde{\mathbf{T}}^c - \lambda \mathbb{I}_\nu)\mathbf{w})_i = -\sum_{a \neq i} w_a + ((c-1)k^* - \lambda)w_i = -\sum_{a=1}^\nu w_a + ((c-1)k^* - \lambda + 1)w_i.$$

If $\sum_{a=1}^\nu w_a \neq 0$ then $\mathbf{w}^1 = \nu^{-1/2}(1, \ldots, 1)$ is eigenvector with eigenvalue $\lambda = (c-1)k^* + 1 - \nu$, being positive if $c > 1 + (\nu - 1)/k^*$.

The remaining eigenvectors live in the subspace of dimension $\nu - 1$, characterized by $\sum_{a=1}^\nu w_a = 0$ and are related to the same eigenvalue $\lambda = (c-1)k^* + 1$. Normalizing, we get that the base of eigenvectors of the subspace orthogonal to the subspace $\mathbb{R}\mathbf{w}^1$ is

$$\mathbf{w}^2 = \frac{\sqrt{k^*(c-1) + 1}}{\sqrt{2}}(1, -1, 0, \ldots, 0)$$

$$\vdots$$

$$\mathbf{w}^a = \frac{\sqrt{k^*(c-1) + 1}}{\sqrt{a(a-1)}}(1, \ldots, 1, 1-a, 0, \ldots, 0)$$

$$\vdots$$

$$\mathbf{w}^a = \frac{\sqrt{k^*(c-1) + 1}}{\sqrt{\nu(\nu - 1)}}(1, \ldots, 1, 1-\nu).$$

Let $\mathbf{P}'$ the matrix of dimension $\nu \times \nu$, whose rows are the vectors $\mathbf{w}_a$, $a = 1, \ldots, \nu$. The matrix $\mathbf{P}'$ is such that ${\mathbf{P}'}^T \mathbf{P}' = \tilde{\mathbf{T}}^c$, therefore if we pose $\tilde{\mathbf{P}} = \mathbf{P}' \mathbf{A}$ we get

$$\tilde{\mathbf{P}}^T \tilde{\mathbf{P}} = (\mathbf{P}' \mathbf{A})^T (\mathbf{P}' \mathbf{A}) = \mathbf{A} {\mathbf{P}'}^T \mathbf{P}' \mathbf{A} = \mathbf{A} \tilde{\mathbf{T}}^c \mathbf{A} = \tilde{\mathbf{J}}^c.$$

Finally, we notice that the rows of $\tilde{\mathbf{P}}$ are

$$\mathbf{v}^1 = \frac{\sqrt{(c-1)k^* - \nu + 1}}{\sqrt{\nu}} (\alpha_1, \ldots, \alpha_\nu)$$

$$\mathbf{v}^2 = \frac{\sqrt{k^*(c-1) + 1}}{\sqrt{2}} (\alpha_1, -\alpha_2, 0, \ldots, 0)$$

$$\vdots$$

$$\mathbf{v}^a = \frac{\sqrt{k^*(c-1) + 1}}{\sqrt{a(a-1)}} (\alpha_1, \ldots, \alpha_{a-1}, \alpha_a - a\alpha_a, 0, \ldots, 0)$$

$$\vdots$$

$$\mathbf{v}^\nu = \frac{\sqrt{k^*(c-1) + 1}}{\sqrt{\nu(\nu-1)}} (\alpha_1, \ldots, \alpha_{\nu-1}, \alpha_\nu - \nu\alpha_\nu).$$

$\square$

We are now ready to prove Lemma (1).

Proof. We want to prove that if the condition $c > 1 + \dfrac{\nu - 1}{k^*}$ holds, then (5.1) is positive definite. Along the proof we will exploit the fact that (A.1) is positive definite and we shall consider the diagonal matrix given by the difference between (5.1) and (A.1):

$$\mathbf{J}^c - \tilde{\mathbf{J}}^c = \begin{cases} (c-1)(k_a - k^*)\alpha_a^2 & \text{se} \quad a = b \\ \\ 0 & \text{se} \quad a \neq b. \end{cases} \tag{A.4}$$

If $c > 1 + \dfrac{\nu - 1}{k^*}$ then $\mathbf{J}^c - \tilde{\mathbf{J}}^c > 0$. If $\mathbf{u} \in \mathbb{R}^\nu$ is an arbitrary vector

$$\mathbf{u}^T \big(\mathbf{J}^c - \tilde{\mathbf{J}}^c \big) \mathbf{u} \geq 0 \tag{A.5}$$

from which we get

$$\mathbf{u}^T \mathbf{J}^c \mathbf{u} > \mathbf{u}^T \tilde{\mathbf{J}}^c \mathbf{u} > 0. \tag{A.6}$$

The last inequality is ensured by Lemma (2) and this concludes the proof. $\square$

REFERENCES

[1] Y. Bengio, Y. LeCun, G. Hinton. Deep Learning. *Nature*, 521: 436-444 (2015).

[2] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, Google book (2016).

[3] D.J. Amit, *Modeling brain function: The world of attractor neural networks*, Cambridge University Press, (1992).

[4] A.C.C. Coolen, R. Kühn, P. Sollich, *Theory of Neural Information Processing Systems*, Oxford Press, Oxford, (2005).

[5] D.J. Amit, H. Gutfreund, H. Sompolinsky, *Spin Glass model of neural networks*, Phys. Rev. A **32**, 1007-1018, (1985).

[6] D.J. Amit, H. Gutfreund, H. Sompolinsky, *Storing infinite numbers of patterns in a spin glass model of neural networks*, Phys. Rev. Lett. **55**, 1530-1533, (1985).

[7] D.H. Hackley, G.E. Hinton, T.J. Sejnowski, *A learning alghoritm for Boltzmann machines*, Cognitive Science **9**(1):147, (1985).

[8] R. Salakhutdinov, G.E. Hinton, *Deep Boltzmann machines*, AISTATS **1**, 3 (2009).

[9] G.E. Hinton, S. Osindero, Y.W. Teh, *A fast algorithm for deep belief nets*, Neural Comp. **18**, 1527-1554, (2006).

[10] H. Larocelle, M. Mandel, R. Pascanu, Y. Bengio, *Learning algorithms for the classification restricted Boltzmann machine.* J. Mach. Learn. **13**, 643-669, (2012).

[11] A. Barra, A. Bernacchia, E. Santucci, P. Contucci, *On the equivalence of Hopfield networks and Boltzmann machines*, Neur. Net. **34**:1-9 (2012).

[12] A. Barra, G. Genovese, P. Sollich and D. Tantari, *Phase transitions in Restricted Boltzmann Machines with generic priors*, Phys. Rev. E **96**(4):042156 (2017).

[13] A. Barra, G. Genovese, P. Sollich, D. Tantari, *Phase diagram of restricted Boltzmann machines and generalized Hopfield networks with arbitrary priors*, Phys. Rev. E **97**(2):022310 (2018).

[14] J. Tubiana, R. Monasson, *Emergence of Compositional Representations in Restricted Boltzmann Machines*, Phys. Rev. Lett. **118**, 138301 (2017).

[15] H. Huang, *Statistical mechanics of unsupervised feature learning in a restricted Boltzmann machine with binary synapses*, Journal of Statistical Mechanics: Theory and Experiment 2017(5):053302 (2017).

[16] H. Huang, *Role of zero synapses in unsupervised feature learning*, J. Phys. A, **51**(8),08LT01 (2018)

[17] O.D. Hebb, *The organization of behaviour: a neuropsychological theory*, Pshyc. Press (1949).

[18] B. Kosko, *Bidirectional associative memories*, IEEE Transactions on Systems, man, and Cybernetics, **18**(1):49–60 (1988).

[19] J. Kurchan, L. Peliti, and M. Saber, *A statistical investigation of bidirectional associative memories (BAM)*, Journal de Physique I, 4(11):1627–1639 (1994).

[20] H. Englisch, V. Mastropietro, B. Tirozzi, *The BAM storage capacity*, Journal de Physique I, 5(1):85–96 (1995).

[21] A. Barra, P. Contucci, E. Mingione, D. Tantari, *Multi-species mean field spin glasses. Rigorous results*, Annales Henri Poincaré, **16**, 691–708 (2015).

[22] A. Barra, G. Genovese, F. Guerra, *Equilibrium statistical mechanics of bipartite spin systems*, J. Phys. A: Math. Theor. **44**, 245002 (2011).

[23] A. Barra, A. Galluzzi, F. Guerra, A. Pizzoferrato, D. Tantari, *Mean field bipartite spin models treated with mechanical techniques*, Eur. Phys. J. B, **87**(3):74, (2014).

[24] D. Panchenko, *The free energy in a multi-species Sherrington-Kirkpatrick model*, The Annals of Probability, 43(6):3494–3513 (2015).

[25] G. Genovese, D. Tantari, *Overlap synchronisation in multipartite random energy models*, J. Stat. Phys., **169**(6), 1162-1170 (2017).

[26] P. Contucci, M.Fedele, C.Vernia, *Inverse problem robustness for multi-species mean field spin models*, J. Phys. A **46**, 065001 (2013).

[27] G. Genovese, D. Tantari, *Non-convex multipartite ferromagnets*, J. Stat. Phys. **163**(3):492–513 (2016).

[28] E. Agliari, A. Barra, A. Galluzzi, D. Tantari, F. Tavani, *A walk in the statistical mechanical formulation of neural networks - Alternative routes to Hebb prescription*, NCTA2014: Neural computation theory & application, **7**, 210–217 (2014).

[29] W.S. McCulloch, W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, Bull. Math. Biophys. **5**, 115-133, (1943).

[30] E.J. Gardner, D. J. Wallace, N. Stroud, *Training with noise and the storage of correlated patterns in a neural network model*, J. Phys. A **22**(12):2019, (1989).

[31] E. Agliari, A. Barra, A. De Antoni, A. Galluzzi, *Parallel retrieval of correlated patterns: From Hopfield networks to Boltzmann machines*, Neural Networks **38**:52-63 (2013).

[32] H. Gutfreund, *Neural networks with hierarchically correlated patterns*, Phys. Rev. A, **37**(2):570 (1988).

[33] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, F. Moauro, *Multitasking associative networks*, Phys. Rev. Lett. **109**, 268101, (2012).

[34] P. Sollich, D. Tantari, A. Annibale, A. Barra, *Extensive parallel processing on scale free networks*, Phys. Rev. Lett. **113**, 238106, (2014).

[35] E .Agliari, A. Annibale, A. Barra, A.C.C. Coolen, D. Tantari, *Immune networks: multitasking capabilities near saturation*, J. Phys. A, **46**, 415003 (2013).

[36] E. Agliari, A. Annibale, A. Barra, A.C.C. Coolen, D. Tantari, *Immune networks: multi-tasking capabilities at medium load*, J. Phys. A, **46**, 335101 (2013).

[37] E. Agliari, A. Annibale, A. Barra, A.C.C. Coolen, D. Tantari, *Retrieving infinite numbers of patterns in a spin-glass model of immune networks*, Europhys. Let. **117** (2), 28003 (2017).

[38] E. Agliari, A. Barra, A. Galluzzi, and M. Isopi, *Multitasking attractor networks with neuronal threshold noise*, Neur. Net., **49**:19–29 (2014).

[39] A. Barra, G. Genovese, F. Guerra, *The replica symmetric approximation of the analogical neural network*, J. Stat. Phys. **140**(4):784-796 (2010).

[40] A. Barra, G. Genovese, F. Guerra, D. Tantari, *How glassy are neural networks?*, J. Stat. Mech., **2012**(07):P07009 (2012).

[41] A. Barra, F. Guerra, *About the ergodic regime in the analogical Hopfield neural networks: Moments of the partition function*, J. Math. Phys. 49, 125217, (2008).

[42] A. Barra, G. Genovese, F. Guerra, D. Tantari, *About a solvable mean field model of a Gaussian spin glass*, J. Phys. A, **47**(15):155002 (2014).

[43] G. Genovese, D. Tantari, *Legendre Duality of Spherical and Gaussian Spin Glasses*, Mathematical Physics, Analysis and Geometry, **18**:10 (2015).

[44] E. Agliari, A. Barra, G. Del Ferraro, F. Guerra, D. Tantari, *Anergy in self-directed B lymphocytes: A statistical mechanics perspective*, J. Theor. Biol. **375**, 21-31 (2015).

[45] H. Sompolinsky, Phys. Rev. A, *Neural networks with nonlinear synapses and a static noise*, **34**, 2571(R) (1986).

[46] B. Wemmenhove, A. C. C. Coolen, *Finite connectivity attractor neural networks*, J. Phys. A **36**, 9617 (2003).

[47] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani, *Retrieval capabilities of hierarchical networks: From Dyson to Hopfield*, Phys. Rev. Lett. **114**, 028103, (2015).

[48] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani, *Hierarchical neural networks perform both serial and parallel processing*, Neur. Net. **66**, 22-35 (2015).

[49] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani, *Metastable states in the hierarchical Dyson model drive parallel processing in the hierarchical Hopfield network*, J. Phys. A, **48**(1):015001 (2014).

[50] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, F. Tavani, *Topological properties of hierarchical networks*, Phys. Rev. E **91**(6), 062807 (2015).

[51] L. Pastur, M. Shcherbina, B.Tirozzi, *The replica-symmetric solution without replica trick for the Hopfield model*, J. Stat. Phys., **74**(5):1161–1183 (1994).

[52] M. Talagrand, *Rigorous results for the Hopfield model with many patterns*, Prob. Theory and Rel. Fields, **110**(2):177-275 (1998).

[53] M. Talagrand, *Exponential inequalities and convergence of moments in the replica-symmetric regime of the Hopfield model*, Ann. Prob. **28**(4):1393-1469, (2000).

[54] A. Bovier, V. Gayrard, P. Picco, *Gibbs states of the Hopfield model with extensively many patterns*, J. Stat. Phys. **79**, 395-414 (1995).

[55] A. Bovier, V. Gayrard, *The retrieval phase of the Hopfield model, A rigorous analysis of the overlap distribution*, Prob. Theor. Rel. Fields **107**, 61-98 (1995).

[56] A. Bovier, V. Gayrard, *Hopfield models as generalized random mean field models*, Progress in Probability Vol. 41, (A. Bovier and P. Picco, Editors) Birkauser, Boston, (1997).

[57] A. Barra, *The mean field Ising model trough interpolating techniques*, J. Stat. Phys. **132**(5):787-809 (2008)

[58] F. Guerra, *Sum rules for the free energy in the mean field spin glass model*, Fields Inst. Comm. **30**:161 (2001).

[59] X. Liao, J. Yu. Qualitative analysis of Bi-directional Associative Memory with time delay. *Int. J. Circ. Theor. Appl.*, 26(3): 219-229 (1998).

[60] J. Cao, M. Xiao, *Stability and Hopf Bifurcation in a Simplified BAM Neural Network With Two Time Delays*, IEEE Transactions on Neural Networks, 18(2): 416-430 (2007).

[61] J. Cao, L. Wang, *Exponential stability and periodic oscillatory solution in BAM networks with delays*, IEEE Transactions on Neural Networks, 13(2): 457-463 (2007).

[62] J. Cao, *Global asymptotic stability of delayed bi-directional associative memory neural networks*, Applied Mathematics and Computation, 142(2-3): 333-339 (2003).

[63] J. Cao, Y. Wan, *Matrix measure strategies for stability and synchronization of inertial BAM neural network with time delays*, Neur. Net., **53**: 165-172 (2014).

[64] J.H. Park, C.H.Park, O.M. Kwon, S.M. Leed, *A new stability criterion for bidirectional associative memory neural networks of neutral-type*, Applied Mathematics and Computation, **199**(2):716-722 (2008).

[65] M. Mezard, *Mean-field message-passing equations in the Hopfield model and its generalizations*, Phys. Rev. E **95**(2), 022117 (2017).

[66] M. Gabrié, E. W. Tramel, F. Krzakala, *Training restricted Boltzmann machine via the Thouless-Anderson-Palmer free energy*, Advances in Neural Information Processing Systems **1**, 640-648, (2015).

[67] A. Barra, A. Di Biasio, F. Guerra, *Replica symmetry breaking in mean-field spin glasses through the Hamilton Jacobi technique*, J. Stat. Mech. 2010(09), P09006, (2010).

[68] A. Barra, G. Dal Ferraro, D. Tantari, *Mean field spin glasses treated with PDE techniques*, Eur. Phys. J. B, **86**(7):332 (2013).

[69] G. Genovese, A. Barra, *A mechanical approach to mean field spin models*, J. Math. Phys., **50**(5):053303 (2009).

[70] L. Evans, *Partial differential equations (graduate studies in mathematics*, vol. 19, American Mathematical Society, Providence (1998).

[71] P. Cannarsa, C. Sinestrari, *Semiconcave functions*, Hamilton-Jacobi equations, and optimal control, Boston, Birkhauser (2004).

[72] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, L. Zdeborova, *Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula* Advances in Neural Information Processing Systems, 424-432, (2015).

[73] Y. Kabashima, F. Krzakala, M. Mèzard, A. Sakata, L. Zdeborova, *Phase transitions and sample complexity in Bayes-optimal matrix factorization*, IEEE Transactions on Information Theory 62 (7), 4228-4265 (2016).

[74] M. Mézard, G. Parisi, M. A. Virasoro, *Spin glass theory and beyond*, World Scientific, Singapore, (1987).

ELENA AGLIARI: DIPARTIMENTO DI MATEMATICA, SAPIENZA UNIVERSITÀ DI ROMA, ITALY.
E-mail address: `elena.agliari@mat.uniroma1.it`

DANILA MIGLIOZZI: DIPARTIMENTO DI MATEMATICA, SAPIENZA UNIVERSITÀ DI ROMA, ITALY.
E-mail address: `danila.migliozzi@yahoo.it`

DANIELE TANTARI: SCUOLA NORMALE SUPERIORE, PISA, ITALY.
E-mail address: `daniele.tantari@sns.it`