

Parallel retrieval of correlated patterns: From Hopfield networks to Boltzmann machines

Elena Agliari^{a,*}, Adriano Barra^{b,c}, Andrea De Antoni^d, Andrea Galluzzi^b

^a *Università di Parma, Dipartimento di Fisica and INFN Gruppo di Parma, Italy*

^b *La Sapienza Università di Roma, Dipartimento di Fisica, Italy*

^c *GNFM Gruppo di Roma, Italy*

^d *School I & C, École Polytechnique Fédérale de Lausanne, Switzerland*

ARTICLE INFO

Article history:

Received 22 May 2012

Received in revised form 19 November 2012

Accepted 20 November 2012

Keywords:

Neural networks
Boltzmann machines
Pattern retrieval
Multitasking networks
Correlated patterns

ABSTRACT

In this work, we first revise some extensions of the standard Hopfield model in the low storage limit, namely the correlated attractor case and the multitasking case recently introduced by the authors. The former case is based on a modification of the Hebbian prescription, which induces a coupling between consecutive patterns and this effect is tuned by a parameter a . In the latter case, dilution is introduced in pattern entries, in such a way that a fraction d of them is blank. Then, we merge these two extensions to obtain a system able to retrieve several patterns in parallel and the quality of retrieval, encoded by the set of Mattis magnetizations $\{m^\mu\}$, is reminiscent of the correlation among patterns. By tuning the parameters d and a , qualitatively different outputs emerge, ranging from highly hierarchical to symmetric. The investigations are accomplished by means of both numerical simulations and statistical mechanics analysis, properly adapting a novel technique originally developed for spin glasses, i.e. the Hamilton–Jacobi interpolation, with excellent agreement. Finally, we show the thermodynamical equivalence of this associative network with a (restricted) Boltzmann machine and study its stochastic dynamics to obtain even a dynamical picture, perfectly consistent with the static scenario earlier discussed.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

In the past century, the seminal works by Minsky and Papert (Minsky & Papert, 1969), Turing (Turing, 1950) and von Neumann (Heims, 1980) set the basis of modern artificial intelligence and, remarkably, established a link between robotics and information theory (Kinchin, 1957). Another fundamental contribution in this sense was achieved by Hopfield (Hopfield, 1982), who, beyond offering a simple mathematical prescription for the Hebbian rule for learning (Hopfield, 1982), also pointed out that artificial neural networks can be embedded in a statistical mechanics framework. The latter was rigorously settled by Amit, Gutfreund and Sompolinsky (AGS) (Amit, 1989), ultimately reinforcing the bridge between cybernetics and information theory (Jaynes, 1957), given the deep connection between the latter and statistical mechanics (Kinchin, 1949, 1957).

As a second-order result, artificial intelligence, whose development had been mainly due to mathematicians and engineers,

became accessible to theoretical physicists too: in particular, when Hopfield published his celebrated paper, the statistical mechanics of disordered systems (mainly spin glasses (Mézard, Parisi, & Virasoro, 1987)) had just reached its maturity and served as a theoretical laboratory where AGS, as well as many others, gave rise to the mathematical backbone of these associative networks.

In a nutshell, the standard Hopfield model can be described by a two body mean-field Hamiltonian (a Lyapunov cost function (Amit, 1989)), which somehow interpolates between the one describing ferromagnetism, already introduced by Curie and Weiss (CW) (Barra, 2008), and the one describing spin-glasses developed by Sherrington and Kirkpatrick (SK) (Mézard et al., 1987). Its dichotomic variables (initially termed “spins” in the original CW or SK theories) are here promoted to perform as binary neurons (some “on/off” exasperations of more standard integrate-and-fire models (Gerstner, 2002)) and the interaction matrix (called the synaptic matrix in this context) assumes a (symmetrized) Hebbian form where information, represented as a set of patterns (namely vectors of ± 1 random entries), is stored. One of the main goals achieved by the statistical mechanics analysis of this model is a clear picture where memory is no longer thought of as statically stored into a confined region (somehow similar to hard disks),

* Corresponding author.

E-mail address: elena.agliari@fis.unipr.it (E. Agliari).

but it is spread over the non-linear retroactive synaptic loops merging neurons themselves. Furthermore, it has been offering a methodology where puzzling questions, such as the memory capacity of the network or its stability under the presence of noise, could finally be consistently formulated.

The success of the statistical-mechanics analysis of neural networks is confirmed by the fact that several variations on the theme followed and many scientific journals dedicated to this very subject arose. For instance, Amit, Cugliandolo, Griniatsly and Tsodyks (Cugliandolo, 1993; Cugliandolo & Tsodyks, 1994; Griniasty, Tsodyks, & Amit, 1993) considered a simple modification of the Hebbian prescription, able to capture the spatial correlation between attractors observed experimentally as a consequence of a proper learning. More precisely, a scalar correlation parameter a is introduced and when its value overcomes a threshold (whose value contains valuable physics as we will explain), the retrieval of a given pattern induces the simultaneous retrieval of its most-correlated counterparts, in some hierarchical way, hence bypassing the standard single retrieval of the original framework (the so-called “pure state”). This route has been further paved in Fukai, Kimoto, Doi, and Okada (1999), Gandolfo, Laanait, Messenger, and Ruiz (1999) and Shcherbina and Tirozzi (1994).

In another extension, proposed by some of the authors of the present paper (Agliari et al., submitted for publication; Agliari, Barra, Galluzzi, Guerra, & Moauro, in press), the hypothesis of strictly non-zero pattern entries is relaxed in such a way that a fraction d of entries is blank. This is shown to imply retrieval of a given pattern without exhausting all the neurons and, following thermodynamic prescriptions (free energy minimizations), the remaining free neurons arrange cooperatively to retrieve further patterns, again in a hierarchical fashion. As a result, the network is able to perform a parallel retrieval of uncorrelated patterns.

Here we consider a Hopfield network exhibiting both correlated patterns and diluted pattern entries, and we study its equilibrium properties through statistical mechanics and Monte Carlo simulations, focusing on the low-storage regime. The analytical investigation is accomplished through a novel mathematical methodology, i.e., the Hamilton–Jacobi technique (developed earlier in Barra, 2008; Genovese & Barra, 2009; Guerra, 2001), which is also carefully explained. The emerging behavior of the system is found to depend qualitatively on a and on d , and we can distinguish different kinds of fixed points, corresponding to the so-called pure-state or to hierarchical states referred to as “correlated”, “parallel” or “dense”. In particular, hierarchy among patterns is stronger for small degree of dilution, while at large d the hierarchy is smoother.

Moreover, we consider the equivalence between the Hopfield model and a class of Boltzmann machines (Bengio, 2009) developed in Barra, Bernacchia, Santucci, and Contucci (2012) and Barra, Guerra, and Genovese (2010) and we show that this equivalence is rather robust and can be established also for the correlated and diluted Hopfield studied here. Interestingly, this approach allows the investigation of dynamic properties of the model which are as well discussed.

The paper is organized as follows. In Section 2, starting from the low-storage Hopfield model, we revise, quickly and pedagogically, the three extensions (and relative phase diagrams) of interest, namely the high-storage case (tuned by a scalar parameter α), the correlated case (tuned by a scalar parameter a) and the parallel case (tuned by a scalar parameter d). In Section 3, we move to the general scenario and we present our main results both theoretical and numerical. Then, in Section 4, we analyze the system from the perspective of Boltzmann machines. Finally, Section 5 is devoted to a summary and a discussion of results. The technical details of our investigations are all collected in the [Appendices](#).

2. Modelization

Here, we briefly describe the main features of the conventional Hopfield model (for an extensive treatment see, e.g., Amit, 1989; Coolen, Kühn, & Sollich, 2005).

Let us consider a network of N neurons. Each neuron σ_i can take two states, namely, $\sigma_i = +1$ (fire) and $\sigma_i = -1$ (quiescent). Neuronal states are given by the set of variables $\sigma = (\sigma_1, \dots, \sigma_N)$. Each neuron is located on a complete graph and the synaptic connection between two arbitrary neurons, say, σ_i and σ_j , is defined by the following Hebb rule:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu, \quad (1)$$

where $\xi^\mu = (\xi_1, \dots, \xi_N)$ denotes the set of memorized patterns, each specified by a label $\mu = 1, \dots, P$. The entries are usually dichotomic, i.e., $\xi_i^\mu \in \{+1, -1\}$, chosen randomly and independently with equal probability, namely, for any i and μ ,

$$P(\xi_i^\mu) = \frac{1}{2}(\delta_{\xi_i^\mu -1} + \delta_{\xi_i^\mu +1}), \quad (2)$$

where the Kronecker δ_x equals 1 iff $x = 0$, otherwise it is zero. Patterns are usually assumed as quenched, that is, the performance of the network is analyzed keeping the synaptic values fixed.

The Hamiltonian describing this system is

$$H(\sigma, \xi) = - \sum_{i=1}^N \sum_{i>j=1}^N J_{ij} \sigma_i \sigma_j = - \frac{1}{2N} \sum_{\substack{i,j=1 \\ i \neq j}}^N \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j, \quad (3)$$

so that the field insisting on spin i is

$$h_i(\sigma, \xi) = \sum_{\substack{j=1 \\ j \neq i}}^N J_{ij} \sigma_j. \quad (4)$$

The evolution of the system is ruled by a stochastic dynamics, according to which the probability that the activity of a neuron i assumes the value σ_i is

$$P(\sigma_i; \sigma, \xi, \beta) = \frac{1}{2} [1 + \tanh(\beta h_i(\sigma_i))], \quad (5)$$

where β tunes the level of noise such that for $\beta \rightarrow 0$ the system behaves completely randomly, while for $\beta \rightarrow \infty$ it becomes noiseless and deterministic; note that the noiseless limit of Eq. (5) is $\sigma_i(t+1) = \text{sign}[h_i(t)]$.

The main feature of the model described by Eqs. (3) and (5) is its ability to work as an associative memory. More precisely, the patterns are said to be memorized if each of the network configurations $\sigma_i = \xi_i^\mu$ for $i = 1, \dots, N$, for every one of the P patterns labeled by μ , is a fixed point of the dynamics. Introducing the overlap m^μ between the state of neurons σ and one of the patterns ξ^μ , as

$$m^\mu = \frac{1}{N} (\sigma \cdot \xi^\mu) = \frac{1}{N} \sum_i \sigma_i \xi_i^\mu, \quad (6)$$

a pattern μ is said to be retrieved if, in the thermodynamic limit, $m^\mu = \mathcal{O}(1)$.

Given the definition (6), the Hamiltonian (3) can also be written as

$$H(\sigma, \xi) = -N \sum_{\mu=1}^P (m^\mu)^2 + P = -Nm^2 + P, \quad (7)$$

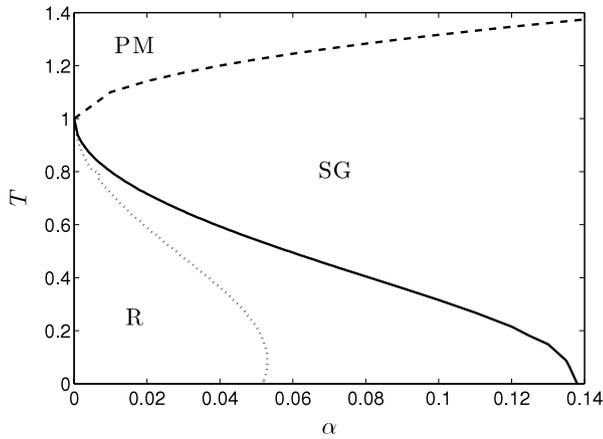


Fig. 1. At a high level of noise the system is ergodic (PM) and no retrieval can be accomplished ($m_\mu = 0, \forall \mu$). By decreasing the noise level below a critical temperature (dashed line), keeping α strictly positive, one enters a “spin-glass” phase (SG), where there is no retrieval ($m_\mu = 0$), yet the system is no longer full-ergodic. Now, if the number of patterns is small enough ($\alpha < 0.138$), by further decreasing the level of noise, one eventually crosses a line (solid curve), below which the system develops $2P$ metastable retrieval states, each can be separately retrieved with a macroscopic overlap ($m_\mu \neq 0$). Finally, when α is small enough ($\alpha < 0.05$), a further transition occurs at a critical temperature (dotted line), such that below this line the retrieval states become global minima (R).

and, similarly,

$$h_i(\sigma, \xi) = \sum_{\mu=1}^P \xi_i^\mu m^\mu - \frac{P}{N} \sigma_i. \quad (8)$$

The analytical investigation of the system is usually accomplished in the thermodynamic limit $N \rightarrow \infty$, consistently with the fact that real networks are comprised of a very large number of neurons. Dealing with this limit, it is convenient to specify the relative number of stored patterns, namely P/N and to define the ratio $\alpha = \lim_{N \rightarrow \infty} P/N$. The case $\alpha = 0$, corresponding to a number P of stored patterns scaling sub-linearly with respect to the amount of performing neurons N , is often referred to as “low storage”. Conversely, the case of finite α is often referred to as “high storage”.

The overall behavior of the system is ruled by the parameters $T \equiv 1/\beta$ (fast noise) and α (slow noise) and it can be summarized by means of the phase diagram shown in Fig. 1. Notice that for $\alpha = 0$, the so-called pure-state ansatz

$$\mathbf{m} = (m, 0, \dots, 0), \quad (9)$$

always corresponds to a stable solution for $T < 1$; the order in the entries is purely conventional and here we assume that the first pattern is the one stimulated.

3. Generalizations

The Hebbian coupling in Eq. (1) can be generalized in order to include possibly more complex combinations among patterns; for instance, we can write

$$J_{ij} = \frac{1}{N} \sum_{\mu, \nu=1}^{P, P} \xi_i^\mu X_{\mu\nu} \xi_j^\nu, \quad (10)$$

where \mathbf{X} is a symmetric matrix; of course, by taking \mathbf{X} equal to the identity matrix we recover Eq. (1). A particular example of generalized Hebbian kernel was introduced in Griniasty et al. (1993), and further investigated in Cugliandolo (1993) and Cugliandolo and Tsodyks (1994), as

$$\mathbf{X} = \begin{pmatrix} 1 & a & 0 & \dots & a \\ a & 1 & a & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a & 0 & \dots & a & 1 \end{pmatrix}. \quad (11)$$

In this way the coupling between two arbitrary neurons turns out to be

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^P [\xi_i^\mu \xi_j^\mu + a(\xi_i^{\mu+1} \xi_j^\mu + \xi_i^{\mu-1} \xi_j^\mu)]. \quad (12)$$

Hence, each memorized pattern, meant as a cyclic sequence, couples the consecutive patterns with a strength a , in addition to the usual auto-associative term.

This modification of the Hopfield model was proposed in Griniasty et al. (1993) to capture some basic experimental features about coding in the temporal cortex of the monkey (Miyashita, 1988; Miyashita & Chang, 1988): a temporal correlation among visual stimuli can evoke a neuronal activity displaying spatial correlation. Indeed, the synaptic matrix (12) is able to reproduce this experimental feature in both low (Cugliandolo, 1993; Griniasty et al., 1993) and high (Cugliandolo & Tsodyks, 1994) storage regimes.

For the former case, one derives the mean-field equations determining the attractors, which, since the matrix is symmetric, are simple fixed points. In the limit of a large network, they read off as (Griniasty et al., 1993)

$$m^\mu = \left\langle \xi^\mu \tanh \left(\beta \sum_{\mu=1}^P m^\mu [\xi_i^\mu + a(\xi_i^{\mu+1} + \xi_i^{\mu-1})] \right) \right\rangle_\xi, \quad (13)$$

where $\langle \cdot \rangle_\xi$ means an average over the quenched distribution of patterns.

In Griniasty et al. (1993), the previous equation was solved by starting from a pure pattern state and iterating until convergence. In the noiseless case, where the hyperbolic tangent can be replaced by the sign function, the pure state ansatz is still a fixed point of the dynamics if $a \in [0, 1/2)$, while if $a \in (1/2, 1]$, the system evolves to an attractor characterized by the Mattis magnetizations (assuming $P \geq 10$, see Appendix A)

$$\mathbf{m} = \frac{1}{2^7} (77, 51, 13, 3, 1, 0, \dots, 0, \dots, 0, 1, 3, 13, 51), \quad (14)$$

namely, the overlap with the pattern used as stimulus is the largest and the overlap with the neighboring patterns in the stored sequence decays symmetrically until vanishing at a distance of 5. Some insights into these results can be found in Appendix A.

In the presence of noise, one can distinguish four different regimes according to the value of the parameters a and T . The overall behavior of the system is summarized in the plot of Fig. 2. A similar phase diagram, as a function of α and a , was drawn in Cugliandolo and Tsodyks (1994) for the high-storage regime.

A further generalization can be implemented in order to account for the fact that the pattern distribution may not be uniform or that pattern entries may possibly be blank. For instance, in the latter case one may replace Eq. (2) by

$$P(\xi_i^\mu) = \frac{1-d}{2} \delta_{\xi_i^\mu - 1} + \frac{1-d}{2} \delta_{\xi_i^\mu + 1} + d \delta_{\xi_i^\mu}, \quad (15)$$

where d encodes the degree of dilution of pattern entries. This kind of extension has strong biological motivations, too. In fact, the distribution in Eq. (2) necessarily implies that the retrieval of a unique pattern does employ all the available neurons, so that no resources are left for further tasks. Conversely, with Eq. (15) the retrieval of one pattern still allows available neurons which can be used to recall other patterns. The resulting network is therefore able to process several patterns simultaneously. The behavior of this system is deeply investigated in Agliari et al. (submitted for publication, in press), as far as the low storage regime is concerned.

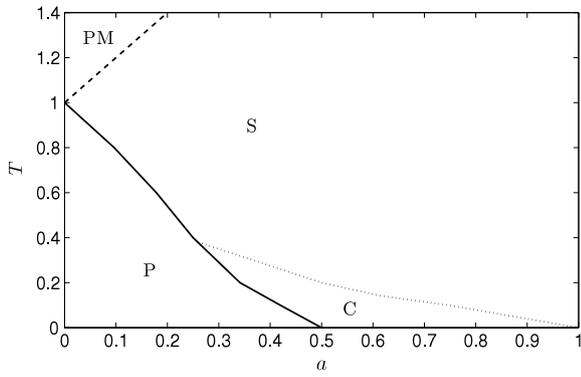


Fig. 2. Phase diagram for the correlated model with low storage ($P = 13$), as originally reported in Cugliandolo (1993). At a high level of noise the system is ergodic (PM) and it eventually reaches a state with $m^\mu = 0, \forall \mu$. At smaller temperatures (below the dashed line), the system evolves to a so-called symmetric state (S), characterized by, approximately, $m^\mu = m \neq 0, \forall \mu$. Then, if a is small enough, by further reducing the temperature (below the solid line), the network behaves as a Hopfield network and the pure state retrieval (P) can be recovered. On the other hand, if a is larger, as the temperature is reduced, correlated attractors (C) appear according to Eq. (14). Then, if the temperature is further lowered, the system recovers the Hopfield-like regime. If $a > 1/2$, the pure state regime is no longer achievable.

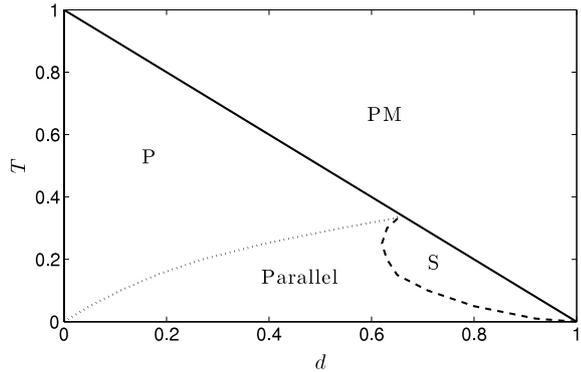


Fig. 3. At high levels of noise the system is ergodic (PM) and below the temperature $T = 1 - d$ (continuous line) it can develop a pure state retrieval (P) or a symmetric retrieval (S), according to whether the dilution is small or large, respectively. At small temperatures and intermediate degree of dilution the system can develop a parallel (Parallel) retrieval, according to Eq. (16). The continuous line works for any value of P , while the dotted and dashed lines were obtained numerically for the case $P = 3$.

In particular, it was shown both analytically (via density of states analysis) and numerically (via Monte Carlo simulations), that the system evolves to an equilibrium state where several patterns are contemporary retrieved; in the noiseless limit $T = 0$ and for not too large d , the equilibrium state is characterized by a hierarchical overlap

$$\mathbf{m} = (1 - d)(1, d, d^2, \dots, 0), \quad (16)$$

hereafter referred to as “parallel ansatz”, while, in the presence of noise, one can distinguish different phases as shown by the diagram in Fig. 3.

To summarize, both generalizations discussed above, i.e. Eqs. (12) and (15), induce the break-down of the pure-state ansatz and allow the retrieval of multiple patterns without falling in spurious states. In the following, we merge such generalizations and consider a system exhibiting both correlation among patterns and dilution in pattern entries.

4. General case

Considering a low-storage regime with constant P , the general case with $a \in [0, 1]$ and $d \in [0, 1]$ can be visualized as a square

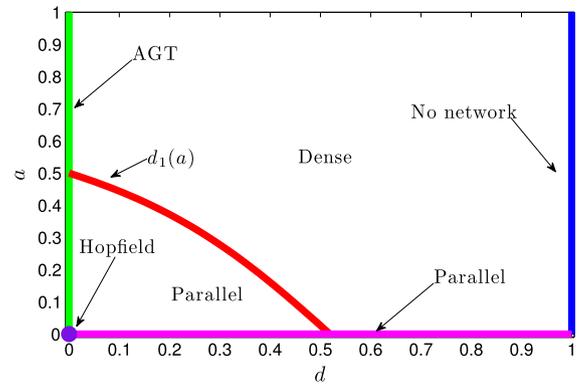


Fig. 4. Schematic representation of the general model considered in the low-storage regime ($\alpha = 0$) and zero noise ($T = 0$). According to the value of the parameters a (degree of correlation) and d (degree of dilution) the system can display different kinds of behavior: when $d = a = 0$ the standard Hopfield model (Hopfield, 1982) is recovered; when $a > 0$ and $d = 0$ the extension proposed by Amit, Griniasty and Tsodyks (AGT) (Griniasty et al., 1993) is recovered, while when $a = 0$ and $d > 0$ the extension proposed in Agliari et al. (in press) and leading to a parallel retrieval is recovered. When $d = 1$ the system degenerates into a set of independent spins as no link survives. When $a > 0$ and $d > 0$ the system can still exhibit parallel retrieval provided that both dilution and correlation are small enough; this is demarcated by the red curve corresponding to Eq. (20). For larger values of a and d the system relaxes to a stable state where Mattis magnetizations are “dense”. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(see Fig. 4), where vertices and nodes correspond to either already-known or trivial cases, while the bulk will be discussed in the following.

First, we notice that the coupling distribution is still normal with average $\langle J \rangle_\xi = 0$ and variance $\langle J^2 \rangle_\xi = (1 + 2a^2)(1 - d)^2 / (2P)$. The last result can be realized easily by considering a random walk of length P : The walker is endowed with a waiting probability d and at each unit time it performs three steps, one of length 1 and two of length a .

Moreover, as shown in Appendix C, the self-consistency equations found in Agliari et al. (in press) and Griniasty et al. (1993) can be properly extended to the case $d \neq 0$ as

$$\mathbf{m} = \langle \xi \tanh(\beta \xi \cdot \mathbf{X} \mathbf{m}) \rangle_\xi, \quad (17)$$

where \mathbf{X} is the matrix inducing the correlation (see Eq. (11)) and the brackets $\langle \cdot \rangle_\xi$ now mean an average over the possible realizations of dilution too.

4.1. Free-noise system: $T = 0$

The numerical solutions of the self-consistency equation (17) are shown in Figs. 5 and 6, as functions of d and a ; several choices of P are also compared. Let us focus on the case $P = 5$ (see Fig. 5) for a detailed description of the system performance.

When $a < 1/2$, the parallel ansatz (16) works up to a critical dilution $d_1(a)$, above which the gap between magnetizations, i.e. $|m^\mu - m^\nu|$, drops abruptly and, for $d > d_1(a)$, all magnetizations are close and decrease monotonically to zero. To see this, let us reshuffle the ansatz in (16), so as to account for the hierarchy induced by correlation, that is,

$$\mathbf{m} = (1 - d)(1, d, d^3, d^4, d^2), \quad (18)$$

which can be straightforwardly extended to any arbitrary P . Given the state (18), the field insisting on σ_i is

$$\begin{aligned} h_i &= \sum_{\mu=1}^P [\xi_i^\mu + a(\xi_i^{\mu-1} + \xi_i^{\mu+1})] m^\mu \\ &= (1 - d) \{ \xi_i^1 [1 + ad(1 + d)] + \xi_i^2 [d + a(1 + d^3)] \} \end{aligned}$$

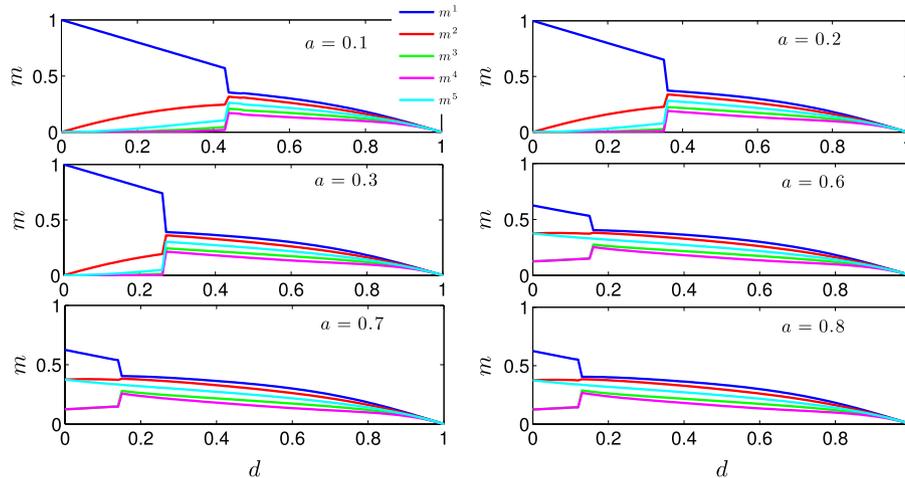


Fig. 5. Magnetization m versus degree of dilution for fixed $P = 5$ and $T = 0.0001$; magnetizations related to different patterns are shown in different colors. Several values of a are considered, as specified in each panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

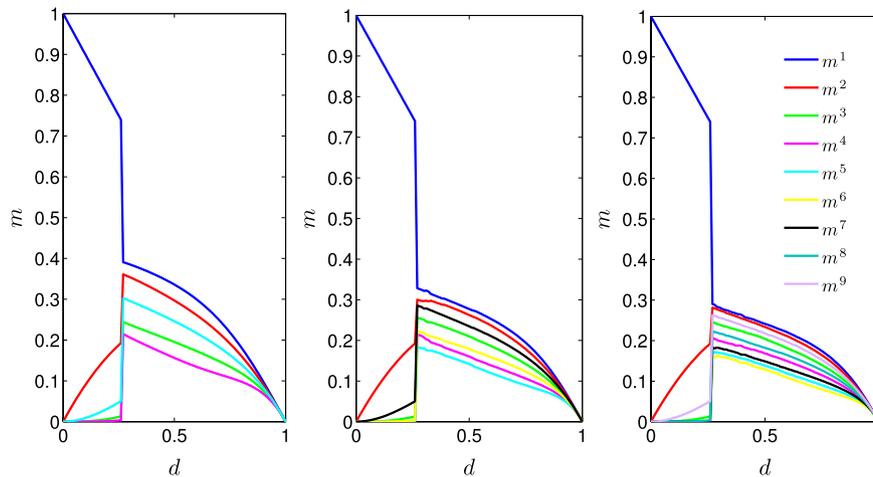


Fig. 6. Magnetization m versus degree of dilution for fixed $a = 0.3$ and $T = 0.0001$. Several values of P are considered for comparison: $P = 5$ (leftmost panel), $P = 7$ (central panel) and $P = 9$ (rightmost panel). Magnetizations related to different patterns are shown in different colors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\begin{aligned} &+ \xi_i^3[d^3 + ad(1 + d^3)] \\ &+ \xi_i^4[d^4 + ad^2(1 + d)] + \xi_i^5[d^2 + a(1 + d^4)]. \end{aligned} \quad (19)$$

A signal-to-noise analysis suggests that this state is stable only for small degrees of dilution. In fact, there exist configurations (e.g., $\xi_i^1 \neq 0$ and $\xi_i^1 = -\xi_i^\mu$, for any $\mu > 1$) possibly giving rise to a misalignment between σ_i and ξ_i^1 , with consequent reduction of m^1 . This can occur only for $d > d_1(a)$, $d_1(a)$ being the root of the equation $a = (1 - d - d^2 - d^3 - d^4)/[2(1 + d^3 + d^4)]$, as confirmed numerically (see Fig. 5). In general, for arbitrary P , one has

$$a = (1 - 2d + d^p)/[2(1 - d + d^3 - d^p)], \quad (20)$$

which is plotted in Fig. 4.

As $d \geq d_1(a)$, the magnetic configuration corresponding to Eq. (18) undergoes an updating where a fraction of the spins aligned with ξ^1 flips to agree mostly with ξ^2 and ξ^5 , and partly also with ξ^3 and ξ^4 ; as a result, m^1 is reduced, while the other magnetizations are increased. Analogously, a fraction of the spins aligned with ξ^2 is unstable and flips so to align mostly with ξ^5 ; consequently, there is a second-order correction which is upwards for m^5 (and to a less extent for m^1 , m^3 and m^4) and downwards for m^2 . Similar arguments apply for higher-order corrections.

At large values of dilution it is convenient to start from a different ansatz, namely from the symmetric state

$$\mathbf{m} = \tilde{m}(1, 1, 1, 1, 1). \quad (21)$$

This is expected to work properly when dilution is so large that the signal on any arbitrary spin σ_i stems from only one pattern, i.e., $\xi_i^\mu \neq 0$ and $\xi_i^\nu = 0, \forall \nu \neq \mu$. This approximately corresponds to $d > 1 - 1/P$. The related magnetization is therefore $\tilde{m} = d^4(1 - d)$. Now, reducing d , we can imagine, for simplicity, that each spin σ_i feels a signal from just two different patterns, say ξ^1 and ξ^2 . The prevailing pattern, say ξ^1 , will increase the related magnetization and vice versa. This induces the breakdown of the symmetric condition so that m^1 grows larger, followed by m^2 , m^5 , and so on. The gap between magnetizations corresponds to the amount of spins which have broken the local symmetry, that is $d^3(1 - d)^2$. Thus, magnetizations differ by the same amount and this configuration is stable for large enough dilutions. By further thickening non-null entries, each spin has to manage a more complex signal and higher order corrections arise. For instance, one finds $m^1 = d^4(1 - d) + 4d^3(1 - d)^2 + 2d^2(1 - d)^3$, and similarly for $m^{\mu > 1}$. This picture is consistent with numerical data and, for large enough values of d , it is independent of a (see Fig. 5). Notice that in this regime of high dilution hierarchy effects are smoothed down, that is, magnetizations are close and we refer to this kind of state as “dense”.

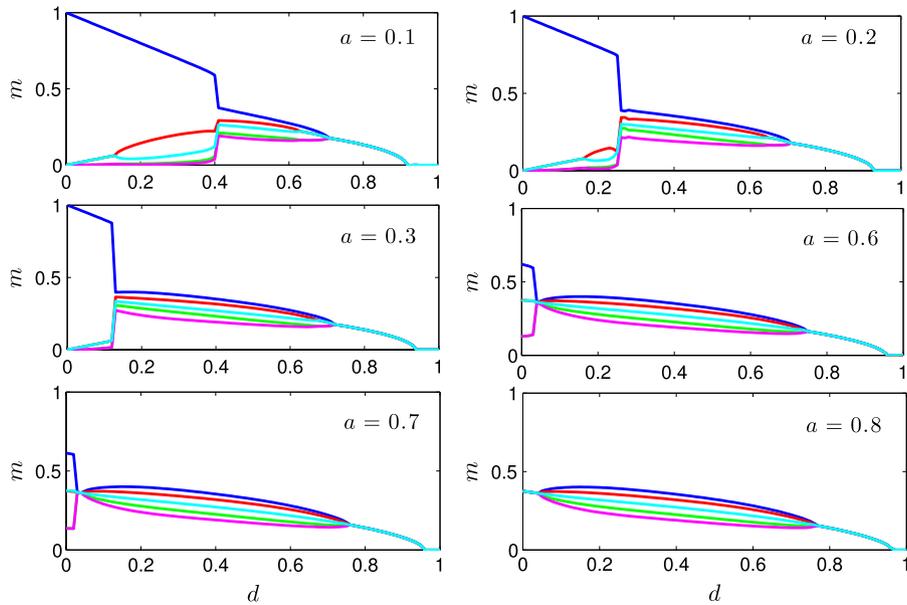


Fig. 7. Magnetization m versus degree of dilution for fixed $P = 5$ and $T = 0.1$; magnetizations related to different patterns are shown in different colors. Several values of a are considered, as specified in each panel. The legend is the same as in Fig. 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

When $a > 1/2$, the parallel ansatz in Eq. (18) is no longer successful at small d , in fact, correlation effects prevail and one should rather consider a perturbed version of the correlated ansatz (14), that is,

$$\mathbf{m} = (1-d) \frac{1}{8} (5, 3, 1, 1, 3). \quad (22)$$

We use (22) as the initial state for our numerical calculations finding, as fixed point, $m^1 = (1-d)5/8$, $m^2 = (1-d^2)3/8$, $m^3 = m^4 = (1+d)1/8$, $m^5 = (1-d+d^2)3/8$. This state works up to a critical dilution $d_2(a)$, where, again there is the establishment of a situation with magnetizations close to and monotonically decreasing to zero. This scenario is analogous to the one describe above and, basically, $d_2(a)$ marks the onset of the region where dilution effects prevail. The threshold value d_2 is slowly decreasing with a .

4.2. Noisy system: $T > 0$

The noisy case gives rise to a very rich phenomenology, as evidenced by the plots shown in Fig. 7.

In the range of temperatures considered, i.e. $T \leq 0.1$, we found that, when $d < d_1(a, T)$ and $a < a_1(T)$, the parallel ansatz (18) works; in general, $d_1(a, T)$ decreases with T and with a , consistently with what we found in the noiseless case (see Fig. 4). Moreover, $a_1(T)$ also decreases with T , consistently with the case $d = 0$ (Cugliandolo & Tsodyks, 1994) (see Fig. 2): from a_1 onwards correlation effects get non-negligible. For larger values of a , namely $a_1(T) < a < a_2(T)$, the perturbed correlated ansatz (22) works, while for $a > a_2(T)$ correlation effects are so important that a symmetric state emerges. Again, we underline the consistency with the case $d = 0$ (Cugliandolo & Tsodyks, 1994): the region $a_1(T) < a < a_2(T)$ corresponds to an intermediate degree of correlation which yields a hierarchical state, while $a > a_2(T)$ corresponds to a high degree of correlation which induces a symmetric state (see Fig. 2).

As for the region of high dilution, we notice that when d is close to 1 the paramagnetic state $\mathbf{m} = (0, 0, 0, \dots, 0)$ emerges. In fact, as long as the signal $(1-d) + 2a(1-d)$ is smaller than noise T , no retrieval can be accomplished, therefore, the condition

$$d < 1 - T/(1+2a) \quad (23)$$

must be fulfilled for $m^\mu > 0$ to hold. The system then relaxes to a symmetric state which lasts up to intermediate dilution, where a state with “dense” magnetizations, analogous to the one described in Section 4.1, emerges.

4.3. Monte Carlo simulations

The model was analyzed also via Monte Carlo simulations, which were implemented to determine the equilibrium values of the order parameter associated to the following Hopfield-like Hamiltonian

$$H = - \sum_{i < j} \sigma_i \sigma_j J_{ij} = \frac{-1}{2N} \sum_j \sigma_j \times \sum_{\mu} [\xi_i^{\mu} \xi_j^{\mu} + a(\xi_i^{\mu+1} \xi_j^{\mu} + \xi_i^{\mu-1} \xi_j^{\mu})], \quad (24)$$

where the coupling encodes correlation among patterns according to Eq. (12), and pattern entries are extracted according to Eq. (15).

The dynamical microscopic variables evolve under the stochastic Glauber dynamic (Glauber, 1963)

$$\sigma_i(t + \delta t) = \text{sign}\{\tanh[\beta h_i(\sigma(t))] + \eta_i(t)\}, \quad (25)$$

where the fields $h_i = \sum_j J_{ij} \sigma_j(t)$ represent the post-synaptic potentials of the neurons. The independent random numbers $\eta_i(t)$, distributed uniformly in $[0, 1]$, provide the dynamics with a source of stochasticity. The parameter $\beta = 1/T$ controls the influence of the noise on the microscopic variables σ_i . In the limit $T \rightarrow 0$, namely $\beta \rightarrow \infty$, the process becomes deterministic and the system evolves according to $\sigma_i(t + \delta t) = \text{sign}(h_i)$.

In general, simulations were carried out using graphs consisting of 10^4 “neurons” and averaging on statistical samples composed of 10^2 realizations. For each realization of the pattern set $\{\xi^{\mu}\}_{\mu=1,\dots,P}$, the equilibrium values of Mattis magnetizations were determined as a function of d and the degree of dilution in pattern entries is incremented in steps of $\Delta d = 0.01$, by sequentially setting equal to zero the entries of the P vectors, in agreement with the distribution (15).

Overall, there is a very good agreement between results from MC simulations, from numerical solution of self-consistent equations and from analytical investigations (see Fig. 8).

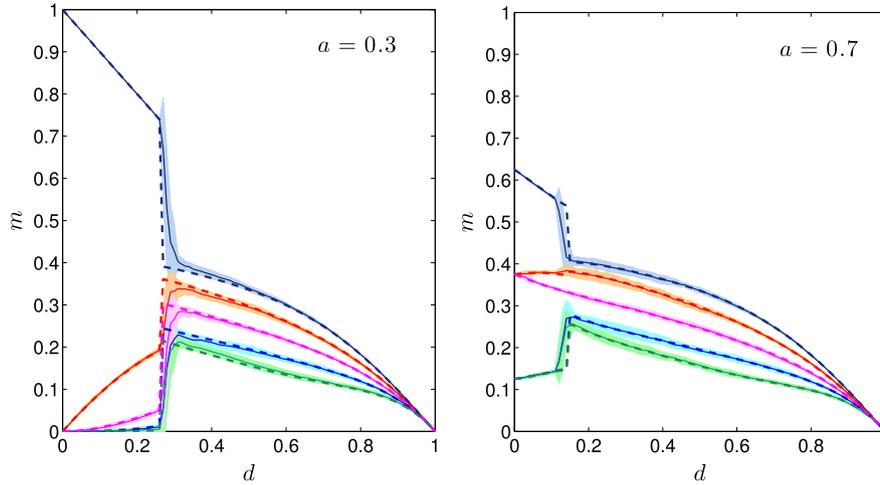


Fig. 8. Magnetization m versus degree of dilution for fixed $P = 5$, $T = 0.0001$ and $a = 0.3$ (left panel) or $a = 0.7$ (right panel). Results from numerical solution of Eq. (17) (dashed, thick line) and Monte Carlo simulations (solid, thin lines) with associated error (shadows) are compared showing, overall, a very good agreement. The legend is the same as in Fig. 5.

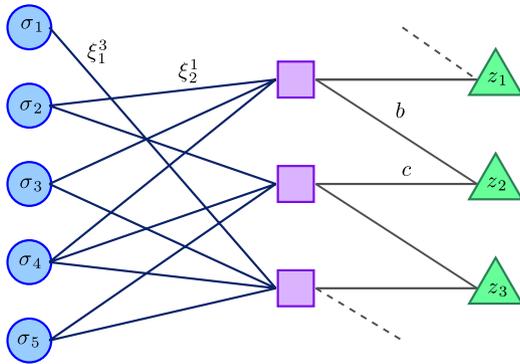


Fig. 9. Schematic representation of a hybrid BM, with $N = 5$ visible nodes (\circ) and $P = 3$ hidden nodes (\triangle). The number of boxes (\square) is P as well. The average number of links stemming from visible units is 2, due to dilution. The link between the i -th visible unit and the μ -th box is ξ_i^μ ; the link between the μ -th box and the μ -th [$(\mu + 1)$ -th] hidden unit is c [b].

5. Extended Boltzmann machine

It is possible to get a deeper insight into the behavior of the system from the perspective of Boltzmann machines (BMs), exploiting the approach first introduced in Barra et al. (2012). In particular, it was shown that a “hybrid” BM characterized by a bipartite topology (where the two parties are made up by N visible units σ_i and by P hidden units z_μ , respectively), after a marginalization over the (analog) hidden units, turns out to be (thermodynamically) equivalent to a Hopfield network. In this equivalence the N visible units play the role of neurons and the link connecting σ_i to z_μ is associated to a weight ξ_i^μ . The term “hybrid” refers to the choice of the variables associated to units: the visible units are binary ($\sigma_i \in \{-1, +1\}$), as in a Restricted Boltzmann Machine, while the hidden ones are analog ($z_\mu \in \mathbb{R}$), as in a Restricted Diffusion Network.

As we are going to show, this picture can be extended to include also the correlation among attractors and the dilution in pattern entries. More precisely, we introduce an additional layer made up by P “boxes”, which switches the signal ξ_i^μ on the two hidden variables z_μ and $z_{\mu+1}$ (see Fig. 9).

Such boxes do not correspond to any dynamical variable, but they retain a structural function as they properly organize the interactions between the two “active” layers: The binary layer is linked to boxes by a synaptic matrix ξ , the boxes are in turn connected to the analog layer by a “connection matrix” that we

call $\tilde{\mathbf{X}}$. The synaptic matrix ξ is $P \times N$ dimensional, each row ξ^μ being a stored pattern. A link between the discrete neuron σ_i and the μ -th box is drawn with weight ξ_i^μ , which takes a value in the alphabet $\{-1, 0, 1\}$ following a proper probability distribution. A null weight corresponds to a lack of link, that is, we are introducing a random dilution in the left of the structure. On the other hand, the matrix $\tilde{\mathbf{X}}$ is $P \times P$ dimensional and meant to recover the correlation among the stored patterns. Here, we choose ξ according to Eq. (15) and $\tilde{\mathbf{X}}$ such as to recover (Cugliandolo, 1993; Cugliandolo & Tsodyks, 1994; Griniasty et al., 1993), namely

$$\tilde{X}_{\mu,v} = c\delta_{\mu,v} + b\delta_{\mu,v-1}, \quad (26)$$

where c and b are parameters tuning the strength of correlation between consecutive pattern entries (*vide infra*). More complex and intriguing choices of $\tilde{\mathbf{X}}$ could be implemented, possibly related to a major adherence to biology.

The dynamics of the hidden and visible layers are quite different. As explained in Barra et al. (2012), the activity in the analog layer follows a Ornstein–Uhlenbeck (OU) diffusion process as

$$\tau \dot{z}_\mu = -z_\mu + \beta \varphi_\mu + \sqrt{2\tau} \zeta_\mu(t), \quad (27)$$

where $-z_\mu$ represents a leakage term, φ_μ denotes the input due to the state of the visible layer, ζ_μ is a white Gaussian noise with zero mean and covariance $\langle \zeta_\mu(t) \zeta_\nu(t') \rangle = \delta_{\mu,\nu} \delta(t - t')$, τ is the typical timescale and β tunes the strength of the input fluctuations. In vector notation the field on the analog layer is $\varphi = \tilde{\mathbf{X}} \cdot \xi \cdot \sigma / \sqrt{N}$, or, more explicitly,

$$\varphi_\mu = \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{v=1}^P \tilde{X}_{\mu,v} \xi_i^v \sigma_i = \frac{1}{\sqrt{N}} \sum_{i=1}^N (c \xi_i^\mu + b \xi_i^{\mu+1}) \sigma_i. \quad (28)$$

The activity in the digital layer follows a Glauber dynamics as

$$\tau' \langle \dot{m}^\mu \rangle_\xi = -\langle m^\mu \rangle_\xi + \left\langle \xi^\mu \tanh \left[\beta \frac{1}{\sqrt{N}} \phi \right] \right\rangle_\xi, \quad (29)$$

where the interaction with the hidden layer is encoded by $\phi = \tilde{\mathbf{X}} \cdot \xi \cdot \mathbf{z}$, that is,

$$\phi_i = \sum_{\mu=1}^P \sum_{v=1}^P \tilde{X}_{\mu,v} \xi_i^v z_\mu = \sum_{\mu=1}^N (c \xi_i^\mu + b \xi_i^{\mu+1}) z_\mu. \quad (30)$$

The timescale of the analog dynamics (27) is assumed to be much faster than that of the digital one (29), that is $\tau' \gg \tau$.

Since all the interactions are symmetric, it is possible to describe this system through a Hamiltonian formulation: from the OU process of Eq. (27) we can write

$$\tau \dot{z}_\mu = -\partial_{z_\mu} \tilde{H}(\mathbf{z}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \tilde{\mathbf{X}}),$$

being

$$\tilde{H}(\mathbf{z}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \tilde{\mathbf{X}}) = z^{2/2} - \beta \sum_{\mu=1}^P \varphi_\mu z_\mu. \quad (31)$$

The partition function $Z_N(\beta; \boldsymbol{\xi}, \tilde{\mathbf{X}})$ for such a system then reads off as

$$Z_N(\beta; \boldsymbol{\xi}, \tilde{\mathbf{X}}) = \sum_{\boldsymbol{\sigma}} \int \prod_{\mu=1}^P d\mu(z_\mu) e^{-\tilde{H}(\mathbf{z}, \boldsymbol{\sigma}, \boldsymbol{\xi}, \tilde{\mathbf{X}})}, \quad (32)$$

where $d\mu(z_\mu)$ is the Gaussian weight obtained by integrating the leakage term in the OU equation.

Now, by performing the Gaussian integration, we get

$$\begin{aligned} Z_{N,P}(\beta; \boldsymbol{\xi}, \tilde{\mathbf{X}}) &= (2\pi)^{P/2} \sum_{\boldsymbol{\sigma}} e^{\frac{\beta^2}{2} \sum_{\mu=1}^P \varphi_\mu^2} \\ &= (2\pi)^{P/2} \sum_{\boldsymbol{\sigma}} e^{-\frac{\beta^2}{2} H(\boldsymbol{\sigma}, \boldsymbol{\xi}, \tilde{\mathbf{X}})}, \end{aligned} \quad (33)$$

where

$$H(\boldsymbol{\sigma}, \boldsymbol{\xi}, \tilde{\mathbf{X}}) = -\varphi^2 = -\frac{1}{N} \boldsymbol{\sigma}^T \cdot \boldsymbol{\xi}^T \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \boldsymbol{\xi} \cdot \boldsymbol{\sigma}, \quad (34)$$

which corresponds to a Hopfield model with patterns $\tilde{\boldsymbol{\xi}} = \tilde{\mathbf{X}} \cdot \boldsymbol{\xi}$, under the shift $\beta^2 \rightarrow \beta$. We then call $\mathbf{X} = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ the correlation matrix which is obviously symmetric, so that the interactions between the σ s are symmetric, leading to an equilibrium scenario. Using Eq. (26), the matrix \mathbf{X} is

$$X_{\mu,\nu} = (c^2 + b^2) \delta_{\mu,\nu} + cb(\delta_{\mu,\nu+1} + \delta_{\mu-1,\nu}), \quad (35)$$

and we can fix $b^2 + c^2 = 1$ and $bc = a$, to recover the coupling in Eq. (12). It is easy to see that $a \leq 1/2$, as long as $b, c \in \mathbb{R}$. In general, with some algebra, we get

$$c = \pm \frac{1}{2} (\sqrt{1+2a} \pm \sqrt{1-2a}), \quad (36)$$

$$b = \pm \frac{1}{2} (\sqrt{1+2a} \mp \sqrt{1-2a}), \quad (37)$$

therefore, the product $\tilde{\mathbf{X}} \cdot \boldsymbol{\xi}$ appearing in both fields φ_μ (see Eq. (28)) and ϕ_μ (see Eq. (30)), turns out to be

$$\begin{aligned} (\tilde{\mathbf{X}} \cdot \boldsymbol{\xi})_{\mu,i} &= \pm \frac{1}{2} [\sqrt{1+2a} (\xi_i^\mu + \xi_i^{\mu+1}) \\ &\quad \pm \sqrt{1-2a} (\xi_i^\mu - \xi_i^{\mu+1})]. \end{aligned} \quad (38)$$

Thus, when $a \leq 1/2$, $(\tilde{\mathbf{X}} \cdot \boldsymbol{\xi})_{\mu,i} \in \mathbb{R}, \forall \mu, i$, while for $a > 1/2$, $(\tilde{\mathbf{X}} \cdot \boldsymbol{\xi})_{\mu,i}$ can be either real or pure imaginary, according to whether the μ -th entry and the following $\mu + 1$ -th are aligned or not.

Having described the behavior of the fields, we can now deepen our investigation on the dynamics of the Boltzmann machine underlying our generalized Hopfield model.

Let us write down explicitly the two coupled stochastic Langevin equations (namely one OU process for the hidden layer, and one Glauber process for the Hopfield neurons) as

$$\tau \dot{z}_\mu = -z_\mu + \frac{\beta}{\sqrt{N}} \sum_i (c \xi_i^\mu + b \xi_i^{\mu+1}) \sigma_i \quad (39)$$

$$\begin{aligned} \tau' \langle \dot{m}_\mu \rangle &= -\langle m_\mu \rangle_\xi \\ &\quad + \left\langle \xi^\mu \tanh \left[\beta \frac{1}{\sqrt{N}} \sum_\nu z_\nu (c \xi^\nu + b \xi^{\nu+1}) \right] \right\rangle_\xi. \end{aligned} \quad (40)$$

Note that by assuming thermalization of the fastest variables with respect to the dynamical evolution of the magnetizations, namely requiring $\dot{z}_\mu = 0$, we can use Eq. (39) to explicit the term z_ν in the argument of the hyperbolic tangent in Eq. (40), hence recovering the self-consistencies of Eq. (17) (see also Appendix C).

Assuming that the two time scales belong to two distinct time sectors, it is possible to proceed in the opposite way, that is

$$\begin{aligned} \langle \dot{m}_\mu \rangle &\cong 0 \Rightarrow \langle m_\mu \rangle \\ &= \left\langle \xi^\mu \tanh \left[\beta \frac{1}{\sqrt{N}} \sum_\nu z_\nu (c \xi^\nu + b \xi^{\nu+1}) \right] \right\rangle_\xi. \end{aligned} \quad (41)$$

For the sake of simplicity let us deal with the $P = 2$ case, being the generalization to the straightforward case $P > 2$. Linearization implies

$$\begin{aligned} \tau \dot{z}_1 &= -z_1 + \langle (c \xi^1 + b \xi^2) \beta^2 [z_1 (c \xi^1 + b \xi^2) \\ &\quad + z_2 (c \xi^2 + b \xi^1)] \rangle_\xi, \end{aligned} \quad (42)$$

$$\begin{aligned} \tau \dot{z}_2 &= -z_2 + \langle (c \xi^2 + b \xi^1) \beta^2 [z_1 (c \xi^1 + b \xi^2) \\ &\quad + z_2 (c \xi^2 + b \xi^1)] \rangle_\xi, \end{aligned} \quad (43)$$

which, recalling that $c^2 + b^2 = 1$ and $cb = a$, turn out to be

$$\tau \dot{z}_1 = z_1 [-1 + (1-d)\beta^2] + z_2 [2a(1-d)\beta^2], \quad (44)$$

$$\tau \dot{z}_2 = z_2 [-1 + (1-d)\beta^2] + z_1 [2a(1-d)\beta^2]. \quad (45)$$

It is convenient to rotate the plane variables z_1, z_2 and define

$$x = z_1 + z_2 \quad (46)$$

$$y = z_1 - z_2, \quad (47)$$

such that Eqs. (42) and (43) can be restated as

$$\dot{x} = -x \left[\frac{1}{\tau} - \frac{(1-d)\beta^2(c+b)^2}{\tau} \right] \quad (48)$$

$$\dot{y} = -y \left[\frac{1}{\tau} - \frac{(1-d)\beta^2(c-b)^2}{\tau} \right], \quad (49)$$

which, in terms of the parameter a , are

$$\dot{x} = -x \left[\frac{1}{\tau} - \frac{(1-d)\beta^2(1+2a)}{\tau} \right] \quad (50)$$

$$\dot{y} = -y \left[\frac{1}{\tau} - \frac{(1-d)\beta^2(1-2a)}{\tau} \right], \quad (51)$$

whose solution is

$$\begin{cases} x(t) = x(0)e^{Y_x t} = x(0) \exp \left[\frac{-t}{\tau} (1 - (1-d)\beta^2(1+2a)) \right] \\ y(t) = y(0)e^{Y_y t} = y(0) \exp \left[\frac{-t}{\tau} (1 - (1-d)\beta^2(1-2a)) \right]. \end{cases} \quad (52)$$

The Lyapunov exponents of the dynamical system Y_x, Y_y turn out to be

$$Y_x = -\frac{1}{\tau} [1 - \beta^2(1-d)(1+2a)], \quad (53)$$

$$Y_y = -\frac{1}{\tau} [1 - \beta^2(1-d)(1-2a)]. \quad (54)$$

This dynamic scenario can be summarized as follows: if the noise level is high ($\beta \ll 1$), the dynamics is basically quenched on its fixed points $x = 0, y = 0$ and the corresponding Hopfield model is in the ergodic phase. If the noise level is reduced below the critical threshold, then two behaviors may appear: if $a \leq 1/2$ both x and y increase, which means that only one z variable is moving away

from its trivial equilibrium state (this corresponds to a retrieval of a single pattern in the generalized Hopfield counterpart); if $a > 1/2$, x increases while y points to zero, which means that both the variables z_1, z_2 are moving away from their trivial equilibrium values (this corresponds to a correlated retrieval in the generalized Hopfield counterpart).

Switching to the original variables we get

$$z_1(t) = e^{-\frac{t}{\tau}[1-(1-d)\beta]} \left\{ z_1(0) \cosh \left[\frac{t(1-d)\beta 2a}{\tau} \right] + z_2(0) \sinh \left[\frac{t(1-d)\beta 2a}{\tau} \right] \right\},$$

$$z_2(t) = e^{-\frac{t}{\tau}[1-(1-d)\beta]} \left\{ z_1(0) \sinh \left[\frac{t(1-d)\beta 2a}{\tau} \right] + z_2(0) \cosh \left[\frac{t(1-d)\beta 2a}{\tau} \right] \right\}.$$

Again, Lyapunov exponents describe a dynamics in agreement with the statistical mechanics findings.

6. Discussion

While technology becomes more and more automatized, our need for a systemic description of cybernetics, able to go over the pure mechanistic approach, gets more urgent. Among the several ways tried in this sense, neural networks, with their feedback loops among neurons, the multitude of their stable states and their stability under attacks (being the latter noise, dilution or various perturbations), seem definitely promising and worth being further investigated.

Along this line, in this work we considered a complex perturbation of the paradigmatic Hopfield model, by assuming a correlation among patterns of stored information *and* dilution in pattern entries. First, we reviewed and deepened both the limiting cases, corresponding to a Hopfield model with correlated attractors (introduced and developed by Amit, Cugliandolo, Griniatsly and Tsodsky (Cugliandolo, 1993; Cugliandolo & Tsodyks, 1994; Griniasty et al., 1993)) and to a Hopfield model with diluted patterns (introduced by some of us Agliari et al., submitted for publication, in press). The general case, displaying a correlation parameter $a > 0$ and a degree of dilution $d > 0$, has been analyzed from different perspectives obtaining a consistent and broad description. In particular, we showed that the system exhibits a very rich behavior depending qualitatively on a , on d and on the noise T : in the phase space there are regions where the pure-state ansatz is recovered, and others where several patterns can be retrieved simultaneously and such parallel retrieval can be highly hierarchical or rather homogeneous or even symmetric.

Further, recalling that interactions among spins are symmetric and therefore a Hamiltonian description is always achievable, we can look at the system as the result of marginalization of a suitable (restricted) Boltzmann machine made by two layers (a visible, digital layer built by the Hopfield neurons and a hidden, analog layer made by continuous variables) interconnected by a passive layer of bridges allowing for pattern correlations. In this way the dynamics of the system can be addressed as well.

Acknowledgments

This work is supported by FIRB grant RBFRO8EKEV.

La Sapienza Università di Roma and INFN are acknowledged too for partial financial support.

The authors are grateful to Ton Coolen and Francesco Moauro for useful discussions.

Appendix A

In this appendix we provide some insights into the shape of the attractors emerging for the correlated model in the noiseless case. We recall for consistency the coupling

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^P [\xi_i^\mu \xi_j^\mu + a(\xi_i^{\mu+1} \xi_j^\mu + \xi_i^{\mu-1} \xi_j^\mu)], \quad (55)$$

where the pattern matrix ξ is quenched. Due to the definition above, magnetizations are expected to reach a hierarchical structure, where the largest one, say m^1 , corresponds to the stimulus and the remaining are symmetrically decreasing, say

$$m^1 \geq m^2 = m^P \geq \dots \geq m^{(P+1)/2} = m^{(P+1)/2+1}, \quad (56)$$

where we assumed P as odd. The distance between the pattern μ and the stimulated pattern is $k(\mu, P) = \min[\mu - 1, P - (\mu - 1)]$.

Moreover, each pattern μ determines a field h^μ , which tends to align the i -th spin with ξ_i^μ . The field reads off as

$$h^\mu = m^\mu + a(m^{\mu+1} + m^{\mu-1}). \quad (57)$$

At zero fast noise we have that

$$\sigma_i = \text{sign}(\varphi_i) = \text{sign} \left[\sum_{\mu=1}^P (\xi_i^\mu h^\mu) \right]. \quad (58)$$

Due to Eqs. (56) and (57), the first pattern is likely to be associated to a large field and therefore to determine the sign of the overall sum appearing in Eq. (57). On the other hand, patterns with μ close to $(P + 1)/2$ are unlikely to give an effective contribution to φ_i and therefore to align the corresponding spins. Indeed, the field $h_\nu \xi_i^\nu$ may determine the sign of φ_i for special arrangements of the patterns μ corresponding to smaller distance, i.e. $k(\mu, P) < k(\nu, P)$. More precisely, their configuration must be staggered, i.e., under gauge symmetry, $\xi_i^1 = +1$, $\xi_i^2 = \xi_i^P = -1$, $\xi_i^3 = \xi_i^{P-1} = +1$, \dots , $\xi_i^{\nu-1} = \xi_i^{\nu+3}$. By counting such configurations one gets m^ν .

With some abuse of language, in the following we will denote with m_k the Mattis magnetization corresponding to patterns at a distance k from the first one. For simplicity, we also assume P small such that $m_k \neq 0, \forall k$.

Then, it is easy to see that, over the 2^P possible pattern configurations, those which effectively contribute to $m_{(P-1)/2}$ are only 4. In fact, it must be $\xi^{(P+1)/2} = \xi^{(P+1)/2+1} = +1(-1)$ and all the remaining must be staggered; therefore, $m_{(P-1)/2} = 4/2^P = 2^{2-P}$.

As for $m_{(P-3)/2}$, contributions come from configurations where the patterns corresponding to $\mu < (P - 1)/2$ are staggered. Such configurations are 2^4 , but we need to exclude those which are actually ruled by the farthest patterns, which are 4, hence, the overall contribution is $16 - 4 = 12$ and $m_{(P-3)/2} = 12/2^P = 3 \times 2^{2-P}$.

We can proceed analogously for the following contributions. In general, by denoting with c_k the k -th contribution, one has the following recursive expression

$$c_{k-1} = 2^{2k} - c_k \quad (59)$$

with $c_{(P-1)/2} = 4$ and $k < (P - 1)/2$. For the last contribution, one has $c_{k-1} = 2^{2k-1} - c_k$, because the last pattern has no “twin”.

Applying this result we get

$$\mathbf{m} = \frac{1}{2}(1, 1, 1), \quad \text{for } P = 3 \quad (60)$$

$$\mathbf{m} = \frac{1}{8}(5, 3, 1, 1, 3), \quad \text{for } P = 5 \quad (61)$$

$$\mathbf{m} = \frac{1}{32}(19, 13, 3, 1, 1, 3, 13), \quad \text{for } P = 7 \quad (62)$$

$$\mathbf{m} = \frac{1}{128}(77, 51, 13, 3, 1, 1, 3, 13, 51), \quad \text{for } P = 9, \quad (63)$$

consistently with Cugliandolo (1993) and Cugliandolo and Tsodyks (1994).

Let us now consider the case $P = 11$; following the previous machinery we get $\mathbf{m} = \frac{1}{256}(307, 205, 51, 13, 3, 1, 1, 3, 13, 51, 205)$. However, such a state is not stable over the whole range of a . In fact, by requiring that the field due to the farthest pattern is larger than the field generated by the staggered configuration of patterns we get

$$2(2a - 1)(-m_6/2 + m_5 - m_4 + m_3 - m_2) \leq m_1 \quad (64)$$

which implies $a < 23/42 \approx 0.54$. From that value of a , the previous state is replaced by a stable state corresponding to $\mathbf{m} = \frac{1}{128}(77, 51, 13, 3, 1, 0, 0, 1, 3, 13, 51)$, which is always stable. Similarly, for $P = 13$, we get a state for \mathbf{m} with $m_i > 0, \forall i$, which is stable only when $a < 85/164 \approx 0.518$, for larger values of a this is replaced by the state found for $P = 11$ and then for the state found for $P = 9$.

All these results have been quantitatively confirmed numerically. We finally notice that for the arguments presented here there is no need for the low storage hypothesis.

Appendix B

In this appendix we want to show that the model is well behaved, namely, that its intensive free energy has a thermodynamic limit that exists and is unique: Despite it may look as a redundant check, we stress that the thermodynamic limit of the high storage Hopfield model (e.g. the $\alpha > 0$ case) is still lacking, hence rigorous results on its possible variants still deserve some interest.

To obtain the desired result, our approach follows two steps: first we show, via annealing, that the intensive free energy is bounded in the system size, then we show that it is also super-additive. As a consequence of these two results the statement straightforwardly follows Ruelle (1999).

Remember that $F_N(\beta, a, d) = N^{-1} \mathbb{E} \ln Z_N(\beta, a, d)$, where

$$Z_N(\beta, a, d) = \sum_{\sigma} \exp(-\beta H_N(\sigma; \xi))$$

is the partition function. Annealing the free energy consists in considering the following bound

$$F_N(\beta, a, d) = \left\langle \frac{1}{N} \log \sum_{\sigma} e^{-\beta H_N(\sigma; \xi)} \right\rangle \quad (65)$$

$$\leq \frac{1}{N} \log \left\langle \sum_{\sigma} e^{-\beta H_N(\sigma; \xi)} \right\rangle \quad (66)$$

$$\leq \frac{1}{N} \log \sum_{\sigma} e^{-\beta(H_N(\sigma; \xi))}, \quad (67)$$

where in the last line we used Jensen inequality.

As a result we get

$$Z_N(\beta, a, d) = \sum_{\sigma} e^{\frac{\beta N}{2} \sum_{\mu} \left\{ (m_{\mu}^2(\sigma)) + a[(m_{\mu}(\sigma)m_{\mu+1}(\sigma)) + (m_{\mu}(\sigma)m_{\mu-1}(\sigma))] \right\}} \quad (68)$$

$$\leq 2^N e^{\frac{P}{2}(N-1)\beta(1+2a)(1-d)}, \quad (69)$$

by which the annealed free energy bound reads off as

$$F_N(\beta, a, d) \leq \ln 2 + \frac{P}{2}\beta(1+2a)(1-d) \left(1 + \frac{1}{N} \right), \quad (70)$$

such that the annealed free energy is $F_A(\beta, a, d) = \ln 2 + P\beta(1+2a)(1-d)/2$.

Let us move over toward proving the super-additivity property and consider two systems independent of each other and with respect to the original N -neurons model, and made of respectively by N_1 and N_2 neurons, such that $N = N_1 + N_2$.

In complete analogy with the original system we can introduce

$$m_{\mu}^{(1)} = \frac{1}{N_1} \sum_i^{N_1} \xi_i^{\mu} \sigma_i^{(1)}, \quad m_{\mu}^{(2)} = \frac{1}{N_2} \sum_i^{N_2} \xi_i^{\mu} \sigma_i^{(2)},$$

and note that the original Mattis magnetizations are linear combinations of the sub-systems counterparts such that

$$m_{\mu} = \frac{N_1}{N} m_{\mu}^{(1)} + \frac{N_2}{N} m_{\mu}^{(2)}.$$

Since the function $x \rightarrow x^2$ is convex (and the translation $x \rightarrow \tilde{x}x$ innocent) we have

$$\begin{aligned} Z_N(\beta, a, d) &\leq \sum_{\sigma} e^{\beta N_1 \sum_{\mu} \left\{ m_{\mu}^{(1),2}(\sigma) + a[m_{\mu}^{(1)}(\sigma)m_{\mu+1}^{(1)}(\sigma) + m_{\mu}^{(1)}(\sigma)m_{\mu-1}^{(1)}(\sigma)] \right\}} \\ &\quad \cdot e^{\beta N_2 \sum_{\mu} \left\{ m_{\mu}^{(2),2}(\sigma) + a[m_{\mu}^{(2)}(\sigma)m_{\mu+1}^{(2)}(\sigma) + m_{\mu}^{(2)}(\sigma)m_{\mu-1}^{(2)}(\sigma)] \right\}} \\ &= Z_{N_1}(\beta, a, d) Z_{N_2}(\beta, a, d), \end{aligned} \quad (71)$$

by which the free energy density $F_N(\beta, a, d)$ is shown to be sub-additive as

$$NF_N(\beta, a, d) \geq N_1 F_{N_1}(\beta, a, d) + N_2 F_{N_2}(\beta, a, d).$$

As the free energy density is sub-additive and it is limited (and this is an obvious consequence of the annealed bound), the infinite volume limit exists and is unique and equal to its sup over the system size $\lim_{N \rightarrow \infty} F_N(\beta, a, d) = \sup_N F_N(\beta, a, d) = F(\beta, a, d)$.

Appendix C

In this appendix we outline the statistical mechanics calculations that brought the self-consistency used in the text (Eq. (13)). Our calculations are based on the Hamilton–Jacobi interpolation technique (Barra, 2008; Genovese & Barra, 2009; Guerra, 2001). This appendix aims two different targets. From one side it outlines the physics of the model and describes it through the self-consistent equation; from the other side it develops a novel mathematical technique able to solve this kind of statistical mechanics problem.

In a nutshell the idea is to think of β as a “time variable” and to introduce P fictitious axes x_{μ} , meant as “space variables”, then, within an Hamilton–Jacobi framework, the free energy with respect to these Euclidean coordinates is shown to play the role of the Principal Hamilton Function, whose solution can then be extrapolated from classical mechanics.

Our generalization of the Hopfield model is described by the Hamiltonian:

$$H_N(\sigma, \xi) = -\frac{1}{2N} \sum_{i,j} \sigma_i \sigma_j \sum_{\mu,v} \xi_i^{\mu} X_{\mu,v} \xi_j^v, \quad (72)$$

as discussed in the text (see Sections 2 and 3).

The N -neuron partition function $Z_N(\beta, a, d)$ and the free energy $F(\beta, a, d)$ can be written as

$$Z_N(\beta, a, d) = \sum_{\sigma} \exp[-\beta H_N(\sigma, \xi)], \quad (73)$$

$$F(\beta, a, d) = \lim_{N \rightarrow \infty} \frac{1}{N} \langle \log Z_N(\beta, a, d) \rangle, \quad (74)$$

where $\langle \cdot \rangle$ again denotes the full averages over both the distribution of the quenched patterns ξ and the Boltzmann weight (for the sake of clarity, let us stress that the factor $B(\beta, a, d) = \exp[-\beta H_N(\sigma, \xi)]$ is termed the Boltzmann factor).

As anticipated, the idea of the Hamilton–Jacobi interpolation is to enlarge the “space of the parameters” by introducing a $P + 1$ Euclidean structure (where P dimensions are of space type and mirrors the P Mattis magnetization, while the remaining one is of time type and mirrors the temperature dependence) and to find a general solution for the free energy in this space thanks to techniques stemming from classical mechanics. The statistical mechanics free energy will then be simply this extended free energy evaluated in a particular point of this larger space. Analogously, the average $\langle \cdot \rangle_{(\mathbf{x}, t)}$ extends the ones earlier introduced by accounting for this generalized Boltzmann factor and will be denoted by $\langle \cdot \rangle$, wherever evaluated in the sense of statistical mechanics.

The “Euclidean” free energy for N neurons, namely $F_N(t, \mathbf{x})$, can then be written in vectorial terms as

$$F_N(t, \mathbf{x}) = \frac{1}{N} \left\langle \ln \left[\sum_{\sigma} \exp \left[\frac{-t}{2N} (\xi\sigma, X\xi\sigma) + (\mathbf{x}, \xi\sigma) \right] \right] \right\rangle. \quad (75)$$

The matrix X can be diagonalized through $X = U^\dagger D U$, where U and U^\dagger are (unitary) rotation matrices and D is the diagonal expression, such that

$$\begin{aligned} F_N(t, \mathbf{x}) &= \frac{1}{N} \left\langle \ln \sum_{\sigma} \exp \left[-\frac{t}{2N} (\xi\sigma, X\xi\sigma) + (\mathbf{x}, \xi\sigma) \right] \right\rangle, \\ &= \frac{1}{N} \left\langle \ln \sum_{\sigma} \exp \left[-\frac{t}{2N} (\sqrt{D}U\xi\sigma, \sqrt{D}U\xi\sigma) \right. \right. \\ &\quad \left. \left. + (\sqrt{D}^{-1}U^\dagger\mathbf{x}, \sqrt{D}U\xi\sigma) \right] \right\rangle, \end{aligned} \quad (76)$$

as $(\xi\sigma, X\xi\sigma) = (\xi\sigma, U^\dagger D U \xi\sigma) = (\sqrt{D}U\xi\sigma, \sqrt{D}U\xi\sigma)$ and $(\mathbf{x}, \xi\sigma) = (\mathbf{x}, U^\dagger U \xi\sigma) = (\sqrt{D}^{-1}U^\dagger\mathbf{x}, \sqrt{D}U\xi\sigma)$. If we switch to the new variables $\tilde{\xi} = \sqrt{D}U\xi$ and $\tilde{\mathbf{x}} = \sqrt{D}^{-1}U^\dagger\mathbf{x}$ we can write the Euclidean free energy in a canonical form as

$$F_N(t, \mathbf{x}) = \frac{1}{N} \left\langle \ln \sum_{\sigma} \exp \left(-\frac{t}{2N} (\tilde{\xi}\sigma, \tilde{\xi}\sigma) + (\tilde{\mathbf{x}}, \tilde{\xi}\sigma) \right) \right\rangle \quad (77)$$

$$\begin{aligned} &= \frac{1}{N} \left\langle \ln \sum_{\sigma} \exp \left(\sum_{ij} \sigma_i \sigma_j \sum_{\mu} \tilde{\xi}_i^{\mu} \tilde{\xi}_j^{\mu} \right. \right. \\ &\quad \left. \left. + \sum_{\mu} \tilde{\mathbf{x}}_{\mu} \sum_j \tilde{\xi}_j^{\mu} \sigma_j \right) \right\rangle. \end{aligned} \quad (78)$$

Thus, we write the (\mathbf{x}, t) -dependent Boltzmann factor as

$$B_N(\mathbf{x}, t) = \exp \left[\frac{-t}{2N} (\tilde{\xi}\sigma, \tilde{\xi}\sigma) + (\tilde{\mathbf{x}}, \tilde{\xi}\sigma) \right], \quad (79)$$

remembering that $B_N(\mathbf{x}, t)$ matches the classical statistical mechanics factor for $t = -\beta$ and $x_{\mu} = 0 \forall \mu$, as even a visual check can immediately confirm.

Now, let us consider the derivative of the free energy with respect to each dimension (i.e., t, \tilde{x}_{μ}):

$$\partial_t F_N(t, \mathbf{x}) = -\frac{1}{2} \sum_{\mu=1}^P \langle (\tilde{m}_{\mu})^2 \rangle_{(\tilde{\mathbf{x}}, t)}, \quad (80)$$

$$\partial_{\tilde{x}_{\mu}} F_N(t, \tilde{\mathbf{x}}) = \langle \tilde{m}_{\mu} \rangle_{(\tilde{\mathbf{x}}, t)}. \quad (81)$$

We notice that the free energy implicitly acts as a Principal Hamilton Action if we introduce the potential

$$V_N(t, \tilde{\mathbf{x}}) = \frac{1}{2} \sum_{\mu=1}^P [\langle (\tilde{m}_{\mu})^2 \rangle - \langle \tilde{m}_{\mu} \rangle^2]. \quad (82)$$

In fact, we can write the Hamilton–Jacobi equation for the F_N action as

$$\partial_t F_N(t, \tilde{\mathbf{x}}) + \frac{1}{2} \sum_{\mu} \left(\partial_{\tilde{x}_{\mu}} F_N(t, \tilde{\mathbf{x}}) \right)^2 + V_N(t, \tilde{\mathbf{x}}) = 0. \quad (83)$$

Interestingly, the potential is the sum of the variances of the order parameters and we know from the Central Limit Theorem argument that in the thermodynamic limit they must vanish, one by one, and, consequently, $\lim_{N \rightarrow \infty} V_N(t, \tilde{\mathbf{x}}) = 0$. Such a self-averaging property plays a key role in our approach as, in the thermodynamic limit, the motion turns out to be free. Moreover, as shown in [Appendix A](#), the limit

$$F(t, \tilde{\mathbf{x}}) = \lim_{N \rightarrow \infty} F_N(t, \tilde{\mathbf{x}}), \quad (84)$$

exists, and $F(t, \tilde{\mathbf{x}})$ can then be obtained solving the free-field Hamilton–Jacobi problem as

$$\partial_t F(t, \tilde{\mathbf{x}}) + \frac{1}{2} \sum_{\mu} \left(\partial_{\tilde{x}_{\mu}} F(t, \tilde{\mathbf{x}}) \right)^2 = 0. \quad (85)$$

From standard arguments of classical mechanics, it is simple to show that the solution for the Principal Hamilton Function, i.e. the free energy, is the integral of the Lagrangian over time plus the initial condition (which has the great advantage of being a trivial one-body calculation as $t = 0$ decouples the neurons). More explicitly,

$$F(t, \tilde{\mathbf{x}}) = F(t_0, \tilde{\mathbf{x}}_0) + \int_0^t dt' \mathcal{L}(t', \tilde{\mathbf{x}}), \quad (86)$$

where the Lagrangian can be written as

$$\mathcal{L}(t, \tilde{\mathbf{x}}) = \frac{1}{2} \sum_{\mu=1}^P \left(\partial_{\tilde{x}_{\mu}} F(t, \tilde{\mathbf{x}}) \right)^2 = \frac{1}{2} \sum_{\mu} \langle \tilde{m}_{\mu} \rangle^2. \quad (87)$$

Having neglected the potential, the motion must be constrained in straight hyperplanes, and the Cauchy problem is

$$\begin{cases} t_0 = 0 \\ \tilde{x}_{\mu} = \tilde{x}_{\mu}^0 + t \langle \tilde{m}_{\mu} \rangle. \end{cases} \quad (88)$$

We can now write the solution more explicitly as

$$\begin{aligned}
 F(t, \tilde{\mathbf{x}}) &= F(0, \tilde{\mathbf{x}}_0) + \int dt' \mathcal{L}(t', \tilde{\mathbf{x}}) \\
 &= \frac{t}{2} \sum_{\mu} \langle \tilde{m}_{\mu} \rangle^2 + \lim_{N \rightarrow \infty} \frac{1}{N} \\
 &\quad \times \sum_j \ln \left[\sum_{\sigma} \exp \left(\sigma_j \sum_{\mu} \tilde{x}_{\mu}^0 \tilde{\xi}_{\mu}^j \right) \right] \\
 &= \ln 2 + \frac{t}{2} \sum_{\mu} \langle \tilde{m}_{\mu} \rangle^2 \\
 &\quad + \left\langle \ln \left[\cosh \left[\sum_{\mu} (\tilde{x}_{\mu} - t \langle \tilde{m}_{\mu} \rangle) \tilde{\xi}_{\mu} \right] \right] \right\rangle. \quad (89)
 \end{aligned}$$

As a consequence, the free energy of this generalization of the Hopfield model can be written by choosing $t = -\beta$ and $\tilde{x}_{\mu} = 0$ for all the spatial dimensions, so as to have

$$\begin{aligned}
 F(\beta, a, d) &= \ln 2 - \frac{\beta}{2} \sum_{\mu} \langle \tilde{m}_{\mu} \rangle^2 \\
 &\quad + \left\langle \ln \left[\cosh \left[\beta \sum_{\mu} \langle \tilde{m}_{\mu} \rangle \tilde{\xi}_{\mu} \right] \right] \right\rangle. \quad (90)
 \end{aligned}$$

We can proceed to extremization, namely $\partial_{\tilde{m}_{\mu}} F(\beta, a, d) = 0$ to get

$$\langle \tilde{m}_{\mu} \rangle = \left\langle \tilde{\xi}_{\mu} \tanh \left[\beta \sum_{\mu} \langle \tilde{m}_{\mu} \rangle \tilde{\xi}_{\mu} \right] \right\rangle, \quad (91)$$

which, turning to the original variables, can be written as

$$F(\beta, a, d) = \ln 2 - \frac{\beta}{2} \sum_{\mu} \langle m_{\mu}^2 \rangle + \langle \ln \cosh \beta(\xi, X m) \rangle, \quad (92)$$

$$\langle m_{\mu} \rangle = \langle \xi_{\mu} \tanh [\beta(\xi, X m)] \rangle, \quad (93)$$

which are the equations that have been used throughout the text. For the sake of clarity, the expression for the Mattis magnetizations

$$m_{\mu} = \left\langle \xi_{\mu} \tanh \left[\frac{\beta}{P} \sum_v z_v (c \xi^v + b \xi^{v+1}) \right] \right\rangle \quad (94)$$

is written extensively for $P = 2$, namely

$$\begin{aligned}
 m_1 &= \frac{(1-d)^2}{2} \tanh \left[\frac{\beta}{2} (z_1 + z_2)(c + b) \right] \\
 &\quad + \frac{(1-d)^2}{2} \tanh \left[\frac{\beta}{2} (z_1 - z_2)(c - b) \right] \\
 &\quad + d(1-d) \tanh \left[\frac{\beta}{2} (z_1 c + z_2 b) \right], \quad (95)
 \end{aligned}$$

$$\begin{aligned}
 m_2 &= \frac{(1-d)^2}{2} \tanh \left[\frac{\beta}{2} (z_1 + z_2)(c + b) \right] \\
 &\quad + \frac{(1-d)^2}{2} \tanh \left[\frac{\beta}{2} (z_1 - z_2)(c - b) \right] + d(1-d) \tanh \\
 &\quad \times \left[\frac{\beta (z_1 + z_2)(c + b) - (z_1 - z_2)(b - c)}{2} \right], \quad (96)
 \end{aligned}$$

and for $P = 3$, namely

$$\begin{aligned}
 m_1 &= d^2(1-d) \tanh [\beta (m_1 + a(m_2 + m_3))] \\
 &\quad + \frac{d(1-d)^2}{2} \tanh [\beta (m_1 + m_3 + a(m_1 + 2m_2 + m_3))]
 \end{aligned}$$

$$\begin{aligned}
 &+ \frac{d(1-d)^2}{2} \tanh [\beta (m_1 - m_3 + a(m_3 - m_1))] \\
 &+ \frac{d(1-d)^2}{2} \tanh [\beta (m_1 + m_2 + a(m_1 + m_2 + 2m_3))] \\
 &+ \frac{d(1-d)^2}{2} \tanh [\beta (m_1 - m_2 + a(m_2 - m_1))] \\
 &+ \frac{(1-d)^3}{4} \tanh [\beta (m_1 - m_2 - m_3 - 2am_1)] \\
 &+ \frac{(1-d)^3}{4} \tanh [\beta (m_1 + m_2 - m_3 + 2am_3)] \\
 &+ \frac{(1-d)^3}{4} \tanh [\beta (m_1 - m_2 + m_3 + 2am_2)] \\
 &+ \frac{(1-d)^3}{4} \tanh \left[\beta (m_1 + m_2 + m_3 \right. \\
 &\quad \left. + 2a(m_1 + m_2 + m_3)) \right],
 \end{aligned}$$

as for m_2 and m_3 , they can be obtained through direct permutation $m_1 \rightarrow m_2 \rightarrow m_3 \rightarrow m_1$.

References

- Agliari, E., Barra, A., Bartolucci, S., Galluzzi, A., Guerra, F., & Moauro, F. Parallel processing in immune networks. *Physical Review E* (submitted for publication) available at [arXiv:1202.6326](https://arxiv.org/abs/1202.6326).
- Agliari, E., Barra, A., Galluzzi, A., Guerra, F., & Moauro, F. Multitasking associative networks. *Physical Review Letters* (in press) available at [arXiv:1111.5191](https://arxiv.org/abs/1111.5191).
- Amit, D. (1989). *Modeling brain function*. Cambridge University Press.
- Barra, A. (2008). The mean field ising model trough interpolating techniques. *Journal of Statistical Physics*, 132, 787.
- Barra, A., Bernacchia, A., Santucci, E., & Contucci, P. (2012). On the equivalence of hopfield networks and boltzmann machines. *Neural Networks*, 34, 1–9.
- Barra, A., Guerra, F., & Genovese, G. (2010). The replica symmetric behavior of the analogical neural network. *Journal of Statistical Physics*, 140(4), 784.
- Bengio, Y. (2009). Learning deep architectures for artificial intelligence. *Machine Learning*, 2, 127.
- Coolen, A., Kühn, R., & Sollich, P. (2005). *Theory of neural information processing systems*. Oxford University Press.
- Cugliandolo, L. (1993). Correlated attractors from uncorrelated stimuli. *Neural Computation*, 6, 220.
- Cugliandolo, L., & Tsodyks, M. (1994). Capacity of networks with correlated attractors. *Journal of Physics A: Mathematical and Theoretical*, 27, 741.
- Fukai, T., Kimoto, T., Doi, M., & Okada, M. (1999). Coexistence of uncorrelated and correlated attractors in a nonmonotonic neural network. *Journal of Physics A: Mathematical and General*, 32, 5551.
- Gandolfo, D., Laanait, L., Messenger, A., & Ruiz, J. (1999). Memory capacity in neural networks with spatial correlations between attractors. *Physica A*, 264, 305.
- Genovese, G., & Barra, A. (2009). A mechanical approach to mean field models. *Journal of Mathematical Physics*, 50, 053303.
- Gerstner, W. K. W. (2002). *Spiking neuron models: single neurons, populations, plasticity*. Cambridge University Press.
- Glauber, R. J. (1963). Time-dependent statistic of the ising model. *Journal of Mathematical Physics*, 4, 294.
- Griniasty, M., Tsodyks, M., & Amit, D. (1993). Conversion of temporal correlations between stimuli to spatial correlations between attractors. *Neural Computation*, 5, 1.
- Guerra, F. (2001). Sum rules for the free energy in the mean field spin glass model. *Fields Institute Communications*, 30, 161–170.
- Heims, S. J. (1980). *John von Neumann and Norbert Wiener: from mathematics to the technologies of life and death*. MIT Press.
- Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79, 2554–2558.
- Jaynes, E. (1957). Information theory and statistical mechanics. *Physical Review*, 106, 620.
- Kinchin, A. I. (1949). *Mathematical foundation of statistical mechanics*. Dover Publications.
- Kinchin, A. (1957). *Mathematical foundation of information theory*. Dover Publications.
- Mézard, M., Parisi, G., & Virasoro, M. A. (1987). *Spin glass theory and beyond*. Singapore: World Scientific.
- Minsky, M., & Papert, S. (1969). *Perceptrons*. MIT Press (enlarged edition, 1988).
- Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335, 817.
- Miyashita, Y., & Chang, H. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331, 68.
- Ruelle, D. (1999). *Statistical mechanics: rigorous results*. World Scientific.
- Shcherbina, M., & Tirozzi, B. (1994). Saddle point equations of a neural network with correlated attractors. *Journal de Physique I*, 4, 629.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 49, 433.