A mathematical analysis of the effects of Hebbian learning rules on the dynamics and structure of discrete-time random recurrent neural networks

Benoit Siri, Hugues Berry, Bruno Cessac, Bruno Delord, Mathias Quoy

A mathematical analysis of the effects of Hebbian learning rules on the dynamics and structure of discrete-time random recurrent neural networks

Benoît Siri,[1] Hugues Berry,[1,*] Bruno Cessac,[2,3] Bruno Delord,[4] and Mathias Quoy[5]

[1] *Team Alchemy, INRIA, Parc Club Orsay Université, 4 rue J Monod, 91893 Orsay Cedex - France*
[2] *Institut Non Linéaire de Nice, UMR 6618 CNRS-Université de Nice, 1361 route des Lucioles, 06560 Valbonne, France*
[3] *Team Odyssee, INRIA, 2004 Route des Lucioles, 06902 Sophia Antipolis, France*
[4] *ANIM, U742 INSERM - Université P.M. Curie, 9 quai Saint-Bernard, 75005 Paris, France*
[5] *ETIS, UMR 8051 CNRS-Université de Cergy-Pontoise-ENSEA, 6 avenue du Ponceau, BP 44, 95014 Cergy-Pontoise Cedex, France*

The analysis of learning recurrent neural networks is challenging, because neuron activity and learning dynamics are mutually coupled: neuron activity depends on the synaptic weight network, which itself varies non trivially under the influence of neuron activity. Understanding this interwoven evolution demands adapted theoretical tools. In this article, we present a mathematical analysis of the effects of Hebbian learning in random recurrent neural networks. Using theoretical tools from dynamical systems and graph theory, we study a generic "Hebb-like" learning rule that can include passive forgetting and different time scales for neuron activity and learning dynamics. We first show that the classical structural statistics from the so-called "complex networks" field (degree distribution, mean-shortest path, clustering index, modularity) do not provide useful insights for the characterization of the coupling between neuron dynamics and network evolution. Instead, this coupling can be analyzed more efficiently by the study of Jacobian matrices, which introduce both a structural and a dynamical point view on the neural network evolution. In this way, we show that "Hebb-like" learning leads to a reduction of the complexity of the dynamics manifested by a systematic decay of the largest Lyapunov exponent. This effect is caused by a contraction of the spectral radius of Jacobian matrices, induced either by passive forgetting or by saturation of the neurons. As a consequence learning drives the system from chaos to a steady state through a sequence of bifurcations. We show that the network sensitivity to the input pattern is maximal at the "edge of chaos". We also emphasize the role of feedback circuits in the Jacobian matrices and the link to cooperative systems.

I. INTRODUCTION

The mathematical study of learning effects (or more generally synaptic plasticity) in neural networks is a difficult task because the dynamics of the neurons depends on the synaptic weight network, that itself evolves non trivially under the influence of neuron dynamics. Understanding this mutual coupling between neuron dynamics and network structure (and its effects on the computational efficiency of the neural network) is a key problem in computational neuroscience and necessitates new analytical approaches.

In recent years, the related field of dynamical systems interacting on complex networks has attracted vast interest. Most studies have focused on the influence of network structure on the global dynamics (for a review, see [9]). In particular, much effort has been devoted to the relationships between node synchronization and the classical statistical quantifiers of complex networks (degree distribution, average clustering index, mean shortest path, motifs, modularity...) [26, 37, 39]. The core idea was that the impact of network topology on global dynamics might be prominent, so that these structural statistics may be good indicators of global dynamics. This assumption proved however largely wrong and some of the related studies yielded contradictory results [32, 39]. Actually, synchronization properties cannot be systematically deduced from topology statistics but may be inferred from the spectrum of the network [3]. Most of these studies have considered diffusive coupling between the nodes [27]. In this case, the adjacency matrix has real nonnegative eigenvalues, and global properties, such as stability of the synchronized states [4] can easily be inferred from its spectral properties. Unfortunately, the coupling between neurons (synaptic weights) in neural networks is rarely diffusive, the corresponding matrix is not symmetric and may contain positive and negative elements. Hence these results are not directly applicable to neural networks.

Discrete-time random recurrent neural networks (RRNNs) with local learning rules are known to display a rich variety of dynamical behaviors, including fixed points, limit cycle oscillations, quasi periodicity and deterministic

*Corresponding author, hugues.berry@inria.fr

chaos [20], that are suspected to be similar to experimental behaviors observed in the olfactory bulb [22, 48]. It is also known that the application of local learning rules reduces the dynamics of chaotic RRNNs to simpler attractors that are specific of the learned input pattern [18]. This phenomenon endows RRNNs with associative memory properties, but remains poorly understood.

Setting up tools from dynamical systems and graph theory, we proceed here to a mathematical analysis of the coupled dynamics between neurons and network structure in the case of RRNNs equipped with biologically-realistic "Hebb-like" learning rules. Our choice for "Hebb-like" rule implementations, though directly derived from classical textbooks on neural networks [33], cannot pretend to cover the entire range of dynamical complexity and variety observed in biological neural networks. However, it allows a thorough mathematical analysis of the effects of learning backed up by numerical simulations, thus yielding a step forward in the study of these complex dynamical systems. These effects are threefold: (i) structural effects ; (ii) dynamical effects ; (iii) "functional" effects. These points will be developed in details in the next sections, but we would like to present here a brief summary.

1. **Structural effects.** There is a first, evident, effect of Hebbian learning: a rewiring of the neural network. However, this rewiring is not some random process where edges would be selected or removed independently of the history. Instead, it results from a complex process where edges are potentiated or depressed according to the neuron dynamics, which is itself depending on the input. The question is therefore whether one can nevertheless extract some general characteristics of the network structure evolution. In this paper we show that Hebbian learning:

   (a) increases the number of positive feedback loops (IV A).

   (b) produces in some situations, a connectivity structure reminiscent of "small-world" networks(III A).

2. **Dynamical effects.** Since neuron dynamics depends on the values of the synaptic weights, one may expect an infinite variety of dynamics when these values are changing. This is actually not the case, and standard structural stability results in dynamical systems theory show, on the opposite, that one can classify the dynamics into a few different *regimes*. A particularly prominent example concerns the (idealistic) situation where the synaptic weights are *independent random variables*. In this case the observed dynamical regimes are: fixed point, periodic, quasi periodic, chaotic. The choice of independent random synaptic weights corresponds to a situation where no prior information is known about the synaptic network structure and, from an information theoretic point of view this amounts to maximizing an entropy. Remarkably enough, this has a direct correspondence with the dynamics: having random independent synapses typically results in a chaotic dynamics, provided that neuron reactivity is sufficiently large. Chaos is characterized both by a positive Lyapunov exponent and a positive entropy. Presenting an input/stimulus to the neural network will result in a quantitative change of the dynamics, whereby the neural network adapts its evolution to this input. But typically, a (weak) input, presented to a network with chaotic dynamics, will not change the chaotic nature of the dynamics (structural stability of hyperbolic attractors [35]). Thus, observing a trajectory of the network before and after input presentation, will not allow to infer information about this input, unless one observes the system on an arbitrary large time scale. On the opposite, the expected effect of Hebbian learning should be to reduce the entropy of the system when the input is present. In this paper we shall actually show rigorously this effect. Hebbian learning results in a reduction of the dynamical complexity measured by the decay of the largest Lyapunov exponent (see Fig. 4 below for an illustration).

3. **Functional effects.** Learning should result in the acquisition of a new ability. The network after learning should be able to recognize a learned input, while this was not necessarily the case before learning. In the present context, recognition is manifested by a drastic change in the dynamics whenever a learned input is presented. Moreover, this effect must be robust and selective. This was indeed observed in [18] for a model similar to (1,4). In this paper we explain this effect on theoretical ground and relate it to bifurcations induced by learning.

Hence, the paper is organized as follows. We present the model and the chosen generic framework for neuron dynamics and learning rules in section II. The following sections are devoted to the analysis of the model. In section III, we study the mutual coupling on the sole basis of its graph-structural effects and show that, albeit slight changes are observed, this approach fails short of yielding causal explanations of the effects of learning in the system. The next section (IV) shows that, taking into account information about dynamics, these effects can be apprehended by mathematical tools from dynamical systems and graph theory. These analytical results are confirmed by thorough numerical simulations. Then we show functional effects (V) related to network sensitivity to the learned pattern. Finally, in the last section (VI), we discuss arguments in favor of the extension of most of our mathematical results to learning rules that do not exactly conform to the generic framework we study here.

II. THE MODEL

We consider firing-rate recurrent neural networks with N neurons and discrete-time dynamics, where learning may occur on a different (slower) time scale than neuron dynamics. Synaptic weights are thus constant for $\tau \geq 1$ consecutive dynamics steps, which defines a "learning epoch". The weights are then updated and a new learning epoch begins. We denote by t the update index of neuron states (neuron dynamics) inside a learning epoch, while T indicates the update index of synaptic weights (learning dynamics). Let $x_i^{(T)}(t) \in [0,1]$ be the mean firing rate of neuron i, at time t within the learning epoch T. Call $\mathbf{x}^{(T)}(t)$ the vector $\left(x_i^{(T)}(t)\right)_{i=1}^N$. Denote by $\mathbf{F}$ the function $\mathbf{F}: \mathbb{R}^N \to \mathbb{R}^N$ such that $F_i(\mathbf{x}) = f(x_i)$. Let $\mathcal{W}^{(T)}$ be the matrix of synaptic weights at the T-th learning epoch and $\boldsymbol{\xi}$ the vector $(\xi_i)_{i=1}^N$. Then the discrete time neuron dynamics (1) writes:

$$x_i^{(T)}(t+1) = f\left(\sum_{j=1}^N W_{ij}^{(T)} x_j^{(T)}(t) + \xi_i\right). \tag{1}$$

or using vectorial notation:

$$\mathbf{x}^{(T)}(t+1) = \mathbf{F}\left[\mathcal{W}^{(T)}\mathbf{x}^{(T)}(t) + \boldsymbol{\xi}\right] = \mathbf{F}\left[\mathbf{u}^{(T)}(t)\right], \tag{2}$$

where:

$$\mathbf{u}^{(T)}(t) = \mathcal{W}^{(T)}\mathbf{x}^{(T)}(t) + \boldsymbol{\xi}, \tag{3}$$

is the vector of components $u_i^{(T)}(t) = \sum_{j=1}^N W_{ij}^{(T)} x_j^{(T)} + \xi_i$. $u_i^{(T)}(t)$ is called "the local field (or the synaptic potential) for neuron i, at neuron time t and learning epoch T".

Here, f is a sigmoidal transfer function (e.g. $f(x) = 1/2\,(1 + \tanh(gx))$). The output gain g tunes the nonlinearity of the function and mimics the reactivity of the neuron. ξ_i is a (time constant) external input applied to neuron i. The vector $\boldsymbol{\xi} = (\xi_i)_{i=1}^N$ is the "pattern" to be learned. $W_{ij}^{(T)}$ represents the weight of the synapse from presynaptic neuron j to postsynaptic neuron i during learning epoch T. The initial weights $W_{ij}^{(1)}$ are randomly and *independently* sampled from a Gaussian law with mean 0 and variance $1/N$. Hence, the synaptic weight matrix $\mathcal{W}^{(T)} = \left(W_{ij}^{(T)}\right)_{i,j=1}^N$ typically contains positive (excitation), negative (inhibition) or null (no synapse) elements and is asymmetric ($W_{ij}^{(T)} \neq W_{ji}^{(T)}$). At the end of one learning epoch, the neuron dynamics indices are reset, and $x_i^{(T+1)}(0) = x_i^{(T)}(\tau), \forall i$.

Our aim is to study the mutual coupling between neuron dynamics (eq. 1) and the structure of the network, that evolves through a given learning rule. Our interest here is in learning rules that conform to Hebb's postulate [29]. More specifically, we define the following generic formulation [33]:

$$W_{ij}^{(T+1)} = \lambda W_{ij}^{(T)} + \frac{\alpha}{N}\Gamma_{ij}^{(T)}, \tag{4}$$

or

$$\mathcal{W}^{(T+1)} = \lambda \mathcal{W}^{(T)} + \frac{\alpha}{N}\Gamma^{(T)}, \tag{5}$$

where $\Gamma^{(T)}$ is the matrix of $\Gamma_{ij}^{(T)}$'s and α is the learning rate.

This equation deserves further comments. The first term in the right-hand side (RHS) member accounts for passive forgetting, i.e. $\lambda \in [0,1]$ is the forgetting rate. If $\lambda < 1$ and $\Gamma_{ij} = 0$ (i.e. both pre- and postsynaptic neurons are silent, see below), eq.(4) leads to an exponential decay of the synaptic weights (hence passive forgetting), with a characteristic rate $\frac{1}{|\log(\lambda)|}$. Note that there is no forgetting when $\lambda = 1$. The second term in the RHS member of eq.(4) generically accounts for activity-dependent plasticity, i.e. the effects of the pre- and postsynaptic neuron firing rates. We focus here on learning rules where this term depends on the *history* of activities[1], i.e.

$$\Gamma_{ij}^{(T)} = h(\tilde{x}_i^{(T)}, \tilde{x}_j^{(T)}), \tag{6}$$

[1] As a matter of fact, note that $\Gamma_{ij}^{(T)}$ is a function of the trajectories $\tilde{x}_i^{(T)}, \tilde{x}_j^{(T)}$, which depend on $\mathcal{W}^{(T)}$, which depends on $\Gamma_{ij}^{(T-1)}$...

where $\tilde{x}_i^{(T)} = \left\{x_i^{(T)}(t)\right\}_{t=1}^{\tau}$ is the trajectory of neuron i firing rate. In the present paper, as a simple example, we shall associate to the history of neuron i rate an activity index $m_i^{(T)}$:

$$m_i^{(T)} = \frac{1}{\tau}\sum_{t=1}^{\tau} x_i^{(T)}(t) - d_i, \tag{7}$$

where $d_i \in [0,1]$ is a threshold and h will be a function of $m_i^{(T)}, m_j^{(T)}$. The neuron is considered active during learning epoch T whenever $m_i^{(T)} > 0$, and silent else. d_i does not need to be explicitly defined in the mathematical study. In numerical simulations however, d_i has to be explicit and we set $d_i = 0.50, \quad \forall i$.

Definition (7) actually encompasses several cases. If $\tau = 1$, weight changes depend only on the instantaneous firing rates, while if $\tau \gg 1$, weight changes depend on the mean value of the firing rate, averaged over a time window of duration τ in the learning epoch. In many aspects the former case can be considered as genuine plasticity, while the latter may be related to meta-plasticity [1]. In this paper, we set $\tau \to \infty$ for the mathematical analysis, and $\tau = 10^4$ in the numerical simulations.

Explicit definition of the function h in eq.(6) is constrained by Hebb's postulate for plasticity. But this postulate is somewhat loosely defined, so that many implementations are possible in our framework. Our choice is guided by the following points:

1. $h > 0$ whenever i (post-synaptic neuron) is active and j (pre-synaptic neuron) is active, analogous to the conditions that prevail to long-term potentiation (LTP) of synaptic connections.

2. $h < 0$ whenever i is inactive and j active, similar to homosynaptic long-term depression (LTD).

3. $h = 0$ whenever the presynaptic neuron j is inactive. This point is often considered as a corollary to Hebb's rule [33]. Moreover, it renders the learning rule asymmetric. Hence, it excludes the possibility that dynamics changes induced by learning could be due to weight symmetrization. Note however that it formally excludes heterosynaptic LTD [6], that would correspond to $h < 0$ for i active and j inactive. In fact, most of the results presented in the paper are still valid whenever heterosynaptic LTD is included (see section VI A for a discussion).

Although these settings are sufficient for mathematical analysis, the h function has to be more precisely defined for numerical simulations. Hence, for the simulations, we set

$$\Gamma_{ij}^{(T)} = m_i^{(T)} m_j^{(T)} H\left(m_j^{(T)}\right), \tag{8}$$

where $H(x)$ is the Heaviside function ($H(x) = 1$ for $x > 0$, 0 else). To summarize, the learning rule used in the following numerical simulations is

$$W_{ij}^{(T+1)} = \lambda W_{ij}^{(T)} + \frac{\alpha}{N} m_i^{(T)} m_j^{(T)} H\left(m_j^{(T)}\right). \tag{9}$$

Finally, in the simulations, we forbid that weights change their sign, as well as self-connections ($W_{ii}^{(T)} = 0, \quad \forall i, \forall T$). Note however that these setups do not have a significant impact on the results presented in this work.

The vectorial notation for eq. (8), useful in the sequel, is:

$$\Gamma^{(T)} = \mathbf{m}^{(T)} \left[\mathbf{m}^{(T)} H(\mathbf{m}^{(T)})\right]^{+}, \tag{10}$$

where $\mathbf{m}^{(T)} = \left(m_i^{(T)}\right)_{i=1}^{N}$, $H(\mathbf{m}^{(T)}) = \left(H(m_i^{(T)})\right)_{i=1}^{N}$, and $+$ denotes the transpose.

Hence, the set of synaptic weights at time $T+1$ and the dynamics of the corresponding neurons are a function of the *whole history* of the system. In this way, we address a very untypical and complex type of dynamical systems where the flow at time t is a function of the past *trajectory* and not only a function of the previous state. (In the context of stochastic processes, such systems are called "chains with complete connections" by opposition to (generalized) Markov processes). This induces rich properties such as a wide learning-induced *variability* in the neuron response to a given stimulus, with the same set of initial synaptic weights, simply by changing the initial conditions for the neurons.

Equations (1,4) define a dynamical system where two distinct processes (neuron dynamics and synaptic network evolution) with two distinct time scales, are involved. It results in a complex interwoven evolution where the neuron dynamics depends on the synaptic graph structure and synapses evolve according to neuron activity. On general grounds, this process has a memory that is a priori infinite and the state of the neural network depends on the past history. This results in a high variability of histories.

III. STRUCTURAL VIEWPOINT

This section aims to study the changes in the network structure induced by the learning rule (4). The network structure can be captured by at least three different matrices. The most natural one is the weight matrix $\mathcal{W}$ that defines the synaptic weight network. From the matrix $\mathcal{W}$ an adjacency matrix can be extracted (to be defined below) . We shall see that these matrices do not provide enough information to analyze the interplay between neuron dynamics and synaptic weight dynamics. On the opposite, we shall see that some general properties of Jacobian matrices allow to connect topological properties of the synaptic graph (e.g. existence of feedback loops) to dynamical properties.

A. Adjacency matrix $\mathcal{A}$

The adjacency matrix of a graph gives information about its connectivity. More specifically, its elements $a_{ij} = 1$ if neuron j is presynaptically connected to neuron i, and 0 else. The interest of its study originates from two points. First, when considering a real biological neural network, information about connectivity (i.e. which neuron is connected to which other) is usually much easier to obtain experimentally than information about the synaptic weights. Accordingly, the experimental data available about real neural network structure is usually described at the adjacency level [49, 50, 56]. Hence comparison between our theoretical/simulation results and real biological networks is much easier at this level. Second, the growing field of study of the so-called "complex networks" has set up several useful statistical tools to quantify real-world partially random networks [9]. But most of these tools are mature only when considering the adjacency matrix (see section III B). We thus studied the connectivity structure using these "complex networks" tools. The adjacency matrix can be obtained directly from the weight matrix $\mathcal{W}$. In our case however, the latter is initially fully connected and contains positive and negative elements. We thus had to adapt some of the "complex networks" tools to our case.

Thresholding. Note first that we are more interested here in the strength of the connection between two neurons, than in its inhibitory/excitatory nature. To facilitate comparison, we chose here to apply a simple relative thresholding method that consists in keeping only the absolute values of the θ percent highest weights (again, in absolute value) from the $N(N-1)$ connections in $\mathcal{W}$, yielding the matrix $\mathcal{C}(\theta)$. Hence gradual decrease of θ enables to gradually isolate the adjacency network formed by the most active weights only. The adjacency matrix $\mathcal{E}(\theta)$ is defined as $e_{ij}(\theta) = H(c_{ij}(\theta))$ and is asymmetric (i.e. $e_{ij}(\theta)$ can be different from $e_{ji}(\theta)$). Unfortunately, most of the "complex network" tools proposed up to now are defined for symmetrical adjacency matrices. As a final step we thus symmetrize $\mathcal{E}(\theta)$, i.e. we build the matrix $\mathcal{A}(\theta)$ whose elements are given by $a_{ij}(\theta) = a_{ji}(\theta) = \max(e_{ij}(\theta), e_{ji}(\theta))$. Thus $a_{ij}(\theta)$ indicates whether i and j are connected by a synapse with a large weight compared to the rest of the network. We limit the range of θ values to ensure that no neuron gets disconnected from the network. Hence, while the weight network contains N nodes and $K = N(N-1)$ unidirectional links, the adjacency network built upon $\mathcal{A}(\theta)$ contains $k(\theta) \leq K$ bidirectional links and N connected nodes.

Clustering index. The clustering index C is a statistical quantifier of the network structure and reflects the degree of "cliquishness" or local clustering in the network [55]. It expresses the probability that two nodes connected to a third one are also connected together and thus can be interpreted as the density of triangular subgraphs in the network. The clustering index of a neuron i in the thresholded symmetrical adjacency matrix $\mathcal{A}(\theta)$, $C_i(\theta)$ is given by

$$C_i(\theta) = \frac{1}{k_i(\theta)\,(k_i(\theta)-1)} \sum_{j,h} a_{ij}(\theta) a_{ih}(\theta) a_{jh}(\theta), \tag{11}$$

where $k_i(\theta) = \sum_{j=1}^{N} a_{ij}(\theta)$ is the degree of neuron i (the number of neurons to which i is connected). The clustering index $C(\theta)$ is the average over the network nodes

$$C(\theta) = \frac{1}{N} \sum_{i=1}^{N} C_i(\theta). \tag{12}$$

Mean shortest path. Let $d(i,j)$ be the shortest path (in number of links/synapses) between neuron i and j in $\mathcal{A}(\theta)$, then the mean shortest path (MSP) is its average over the $N(N-1)$ nonidentical neurons pairs

$$MSP(\theta) = \frac{1}{N(N-1)} \sum_{i,j} d(i,j). \tag{13}$$

Normalization. These two quantifiers are usually informative only when compared to similar measures obtained from reference uncorrelated random networks [55]. If the adjacency matrix $\mathcal{A}(\theta)$ contains N neurons and $k(\theta)$ bidirectional links, we build a reference uncorrelated network as follows. We start with N unconnected neurons. We then choose uniformly at random a pair of non-identical neurons (i,j) and connect them bidirectionally. The latter step is iterated until the number of non-identical bidirectional links is $k(\theta)$. We then compute the clustering index and mean shortest path of the resulting reference network, and average the obtained values over 15 realizations of the reference network, yielding the reference values $C_{rand}(\theta)$ and $MSP_{rand}(\theta)$.

Simulation results. Figure 1A & B show simulation results for the evolution of $C^{(T)}(\theta)/C^{(T)}_{rand}(\theta)$ and $MSP^{(T)}(\theta)/MSP^{(T)}_{rand}(\theta)$ during learning. The distribution of the initial weights over the network being (uncorrelated) random, the resulting adjacency matrix $\mathcal{A}^{(1)}(\theta)$ is essentially an uncorrelated random network, i.e. one expects $C^{(1)}(\theta)/C^{(1)}_{rand}(\theta) \approx 1$ and $MSP^{(1)}(\theta)/MSP^{(1)}_{rand}(\theta) \approx 1, \quad \forall\theta$. This is confirmed in fig. 1: during the approximately first 100 learning epochs, both network statistics remain around 1. From the point of view of these two statistics, the corresponding adjacency matrices are essentially undistinguishable from uncorrelated random networks. The situation however changes for longer learning epochs. For $T \gtrsim 100$, the relative MSP remains essentially 1 for all thresholds θ (fig. 1B). Hence, the average minimal number of synapses linking any two neurons in the network remains low, even when only large synapses are considered.

Conversely, the clustering index (fig. 1A) increases at $T > 100$ for the stronger synapses and reaches a stable value that is somewhat higher than in reference uncorrelated random networks. For instance, the adjacency matrix formed by the highest 30% synapses displays a clustering index that is 20% higher than the reference value. Hence, in the adjacency matrices obtained at long learning epochs for the strong synapses, the probability that the neighbors of a given neuron are themselves interconnected is slightly higher than if these strong synapses were laid at random over the network. In other terms, the learning rule yields correlations among the largest synapses at long learning epochs.

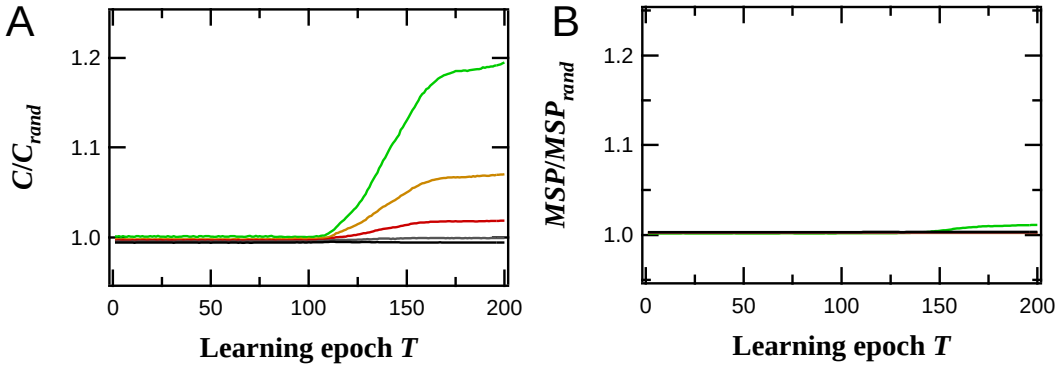

FIG. 1: Evolution of the normalized structural statistics during learning with rule eq.(9). Values are averages over 50 different realizations of the network (random initial firing rates and synaptic weights). The value of the threshold θ ranges, from top to bottom in each panel, from 30% to 50% (by 5% increments). (A) Normalized clustering index $C^{(T)}(\theta)/C^{(T)}_{rand}(\theta)$. (B) Normalized mean-shortest path $MSP^{(T)}(\theta)/MSP^{(T)}_{rand}(\theta)$. For comparison purpose, the scales on the y-axes have been set identical. Other parameters were: $\lambda = 0.90$, $\alpha = 5 \times 10^{-3}$, $g = 10$, $\xi_i = 0.010 \sin(2\pi i/N) \cos(8\pi i/N)$ $\forall i = 1 \ldots N$ and $N = 100$.

In the literature related to "complex networks", networks with a larger clustering index but a similar MSP with respect to a comparable uncorrelated random network, are referred to as *small-world* networks [2]. Note however that

[2] We use here the definition that is commonly found in the recent "complex network" literature, i.e. small-world networks are networks

no quantitative criterion exists to decide whether a given increase of the clustering index is large enough for the network to be small-world. We thus interpret our result as indicative of the emergence of a "weakly small-world" property. Hence, the learning rule eq.(9) slightly changes the distribution of the large synapses over the weight network, so that these large synapses organize as a "weakly small-world" network. Note that we have checked that this low but significant increase of the clustering index for the large synapses does not depend on the size of the network N.

Again, the observed effect on the clustering index is indeed quite weak so that the pertinence of this effect is questionable (see section VI A for a discussion). As we shall see, the effects of Hebbian learning on neuron dynamics is quite a bit more prominent. Actually, anticipating the results on dynamical effects, one notes that the time scale for dynamical changes is definitely smaller than the time scale where network structuration starts. In any case, the previous indicators give no real clue about the mutual coupling between global dynamics and the network structure. Hence, in our case, the information offered by the classical statistics of the "complex networks" approaches fails short of explaining the dynamical effects of learning. In the following, we examine what information can be obtained if the structure is observed at the level of the weights or Jacobian matrices.

B. weight matrix $\mathcal{W}$

The weight matrix $\mathcal{W}$ is arguably the most natural object for the study of the structure of the synaptic networks. It also contains information about transport properties of neural fluxes in the network, that is not available in the mere adjacency matrix. This information can in some cases be studied using weighted extension of the same structural statistics as those used above. However, extensions of these statistics to weighted networks are mostly restricted to symmetric networks with positive weights [43]. Thus they cannot be directly applied to the case studied here. The transport properties of weighted networks may as well be questioned through the behavior of random walkers over the network (i.e. spectral properties of the Laplacian [53]). But again, these results can be easily interpreted only in the case of networks with positive (and sometimes symmetric) weights [17, 53]. Furthermore, these spectral studies implicitly assume the absence of a specific dynamics in each network node (i.e. no neuronal transfer function), which is a further strong limitation in our case. We conclude from these elements that if neuron dynamics is *not* taken into account, the analysis of the structure of the weight network can hardly bring valuable information about the effect of learning in our system. However, we show in section IV that if neuron dynamics is accounted for, the behavior of the weight matrix becomes a crucial element.

C. Jacobian matrices.

The Jacobian matrices (def. 15) are the key objects of our analysis. They indeed provide quite a bit more information than the mere synaptic graph, which does not provide any direct information on node dynamics and is a poor indicator of the effective graph structure. Indeed, from the dynamical point of view Jacobian matrices are related to stability properties, occurrence of bifurcations and Lyapunov exponents of a *non linear* dynamical system. These aspects are well known and are widely developed in the present paper. The *Jacobian matrix* of $\mathbf{F}$ at $\mathbf{x}$, denoted by $D\mathbf{F}_{\mathbf{X}}$, has components:

$$\frac{\partial F_i}{\partial x_j} = f'(\sum_{k=1}^{N} W_{ik}x_k + \xi_i)W_{ij} = f'(u_i)W_{ij}. \tag{14}$$

Thus it displays the following specific structure:

$$D\mathbf{F}_{\mathbf{X}} = \Lambda(\mathbf{u})\mathcal{W}, \tag{15}$$

with:

$$\Lambda_{ij}(\mathbf{u}) = f'(u_i)\delta_{ij}. \tag{16}$$

---

that display a high clustering index and a short mean shortest path with respect to comparable uncorrelated random networks [9]. Note however that former literature, especially in the graph theory domain, defines the small-world property as an at most logarithmic increase of the MSP with the number of nodes N and does not consider local clustering properties such as accounted for by the clustering index.

Note that $D\mathbf{F}_{\mathbf{X}}$ depends on $\mathbf{x}$, contrarily to $\mathcal{W}$. Another aspect developed here is connected to causal action of a neuron to another. Assume that we slightly perturb at time t the state of neuron j with a small perturbation (e.g. $x_j(t) \to x_j(t) + \delta_j(t)$). Then the effect of this change on neuron i, at time $t+1$ is given by $x_i(t+1) = f\left(\sum_{k=1}^{N} W_{ik}x_k(t) + \xi_i + W_{ij}\delta_j(t)\right)$. One can perform a Taylor expansion of this expression in powers of $W_{ij}\delta_j(t)$. To the linear order the effect is given by $f'(u_i(t))W_{ij}\delta_j(t)$ where $u_i(t) = \sum_{k=1}^{N} W_{ik}x_k(t) + \xi_i$ has already been defined above. More generally the effects of perturbations at the linear order are given by the Jacobian matrix $D\mathbf{F}_{\mathbf{X}}$, which depends on the state $\mathbf{x}$ of the neural network. To each Jacobian matrix $D\mathbf{F}_{\mathbf{X}}$ one can associate a graph, called "the graph of linear influences", such that there is an oriented edge $j \to i$ iff $\frac{\partial f(u_i)}{\partial x_j} \neq 0$. The edge is positive if $\frac{\partial f(u_i)}{\partial x_j} > 0$ and negative if $\frac{\partial f(u_i)}{\partial x_j} < 0$. An important remark is that this graph depends on the current state $\mathbf{x}$, contrarily to the weight matrix which is a constant inside a given learning epoch. This has important consequences. Indeed, in our case since $\frac{\partial F_i}{\partial x_j} = f'(u_i)W_{ij}$ (eq. 14) the edge $j \to i$ in the graph of linear influences can be very small even if the synaptic weight W_{ij} is large. It suffices that $|u_i|$ be large. This effect, due to the saturation of the transfer function f, is prominent in the subsequent studies.

We have now the following situation: "above" (in the tangent bundle) each point $\mathbf{x}$, there is graph. This graph contains *circuits or feedback loops*. If e is an edge, denote by $o(e)$ the origin of the edge and $t(e)$ its end. Then a circuit is a sequence of edges $e_1, ..., e_k$ such that $o(e_{i+1}) = t(e_i)$, $\forall i = 1...k-1$, and $t(e_k) = o(e_1)$. Such a circuit is positive (negative) if the product of its edges is positive (negative). A positive circuit basically yields (to the linear order) a positive feedback that induces an increase of the activity of the neurons in this circuit. Obviously, there is no exponential increase since rapidly nonlinear terms will saturate this effect. It is thus expected that positive loops enhance stability.

A particularly prominent example of this is well known in the framework of continuous time neural networks models and also in genetic networks. It is provided by so-called "cooperative systems". A dynamical system is called cooperative if $\frac{\partial f(u_i)}{\partial x_j}(\mathbf{x}) \geq 0, \forall i \neq j$. Therefore, in this case, all edges are positive edges[3], whatever the state of the neural network and all circuits are positive. Cooperative systems preserve the following partial order $\mathbf{x} \leq \mathbf{y} \Leftrightarrow x_i \leq y_i,\ i = 1 \dots N$. Thus $\mathbf{x}(\mathbf{0}) \leq \mathbf{y}(\mathbf{0}) \Rightarrow \mathbf{x}(\mathbf{t}) \leq \mathbf{y}(\mathbf{t}),\ \forall t > 0$ (this corresponds to the positive feedback discussed above). From these inequalities, Hirsch [31] proved that for a two dimensional cooperative dynamical system, any bounded trajectory converges to a fixed point. In larger dimension, one needs moreover a technical condition on the Jacobian matrix: it must be irreducible. Then Hirsch proved that the ω-limit set of almost every bounded trajectory is made of fixed points. Note that this result holds when f is nonlinear.

On the opposite, negative loops usually generate oscillations. For example, the second Thomas conjecture [52], proved by Gouzé [25] under the hypothesis that the sign of the Jacobian matrix elements do not depend on the state, states that "A negative loop is a necessary condition for a stable periodic behavior". In our model, negative loop will generate oscillations provided that the nonlinearity g is sufficiently large. This can be easily figured out by considering a system with 2 neurons. A necessary condition to have a Hopf bifurcation giving rise to oscillations is $W_{12}W_{21} < 0$, but the bifurcation occurs only when g is large enough.

In a generic situation, one has positive and negative loops. In a model such as eq.(1) the weight of a loop $k_1, k_2, \dots k_n, k_1$ is given by $\prod_{l=1}^{n} W_{k_{l+1}k_l} f'(u_{k_l})$, where $k_{n+1} = k_1$. Therefore, the weight of a loop is a product of a "topological" contribution (the product $\prod_{l=1}^{n} W_{k_{l+1}k_l}$) and of a dynamical contribution (the product $\prod_{l=1}^{n} f'(u_{k_l})$). This last term depends on the state of the neurons in the loop and evolves in time. Therefore, the weight of a loop depends on time (but its sign remains constant). When increasing g, the existence of negative loops generates a cascade of Hopf bifurcations leading the system to chaos via quasiperiodicity [12, 14, 20]. A detailed investigation of the conjugated effects of topology and dynamics in the context of linear response can be found in [15].

D. Conclusion

In the framework of the dynamical system studied in the present paper, the above considerations lead to the conclusion that the study of the network on the sole basis of its structural properties is not sufficient to apprehend the mutual coupling between neuron dynamics and network structure. The adjacency matrix (section III A) can be used to get very general information on the gross organization of the network under the influence of the Hebb-like

[3] More generally, there is a variable change which maps the initial dynamical system to a cooperative system with positive edges.

learning rule, but it fails short of providing causal relationships between the evolution of the dynamics and that of the structure. Furthermore, the topological statistics classically used in the complex networks approaches seem not to be powerful enough to account for the structural changes induced by our Hebb-like learning rule. More information could in principle be found in the weight matrix (section III B), but its analysis is in practice hardly feasible in our case, because it is asymmetric and contains negative weights. Finally, the structure of the network built on the Jacobian matrices (section III C) yield quite a bit more information as we now develop.

IV. DYNAMICAL VIEWPOINT

The properties of the dynamical system (2) depend on the parameter g, on the N inputs ξ_i and they also highly depend on the value of the $N(N-1)$ synaptic weights. Thus, a large number of parameters acts on the dynamics of this dynamical system. However, some general statements can be inferred from standard methods in dynamical systems theory. Due to the saturation of the sigmoidal transfer function[4], dynamics is contracting in some directions of the phase space. Thus, trajectories converge to some[5] *attractor*. Depending on the parameters, dynamics can also be contracting everywhere [12]. There may as well exist neutral directions. In this case, the dynamics is periodic or quasiperiodic. One can also have local expansion effects, resulting in a chaotic attractor. This typically arises when g is sufficiently large (depending on the spectral properties of $\mathcal{W}$). The situation is thus very different from Hopfield networks, where several distinct attractors coexist and can be reached selectively, depending on the initial condition in the activity space. In our case, different attractors may only be reached through a parameter change, i.e. a bifurcation. In particular, the input pattern $\boldsymbol{\xi}$ or the variance of the initial weight distribution may be used as such a bifurcation parameter [18].

Starting from spontaneous chaotic dynamics, previous simulation works have shown that the application of Hebbian learning leads to a simplification of the chaotic dynamics on a quasiperiodic attractor, a limit cycle and finally a fixed point [18]. This is exemplified in fig. 4B, which shows simulation results for the network-averaged neuron dynamics obtained at different learning epochs. The dynamics is initially chaotic ($T = 1$), then gradually settles onto a periodic limit cycle ($T = 5$ & 6), then on a fixed point attractor at longer learning epochs (see e.g. $T = 100$ in this figure). This modification of the global dynamics is a typical example of the reduction of the attractor complexity due to the mutual coupling between weight evolution and neuron dynamics. The aim of this section is to provide theoretical and experimental explanations of this reduction of complexity.

A. Feedback circuits.

As already mentioned in section III C, the study of feedback circuits in the Jacobian matrix could represent a convenient link between structural and dynamical aspects of the system. The definition and interpretation of feedback circuits in the framework of coupled dynamical systems has already been presented in section III C. We just recall here that globally speaking, positive feedback circuits tend to induce (fixed-point) stability, while negative ones rather promote oscillations (and when coupled, chaotic behaviors). We measured the evolution of these circuits during learning in numerical simulations via the weighted-fraction of positive circuits[6] in the Jacobian matrix, $R_n^{(T)}$, that we defined as

$$R_n^{(T)} = \frac{\sigma_n^{+(T)}}{|\sigma_n^{+(T)}| + |\sigma_n^{-(T)}|} \tag{17}$$

where $\sigma_n^{+(T)}$ (resp. $\sigma_n^{-(T)}$) is the sum of the weights of every positive (resp. negative) circuits of length n in the Jacobian network at learning epoch T. Hence $R_n^{(T)}$ ranges from 0 to 1. If its value is > 0.5, the positive circuits of length n are stronger (in total weight) than the negative ones. For computation time reason, we only computed here the weighted-fraction of positive circuits for circuits of length $n = 2$ and $n = 3$ (i.e. $R_2^{(T)}$ and $R_3^{(T)}$).

---

[4] This effect can also arise if the spectral radius of the weight matrix is sufficiently small, as we shall see.

[5] In the following discussion we are implicitly assuming that the system displays a single attractor. This can however be controlled (see [14] for details on the determination of the regions in the parameter space where only one attractor exists). If several attractors coexist they can be of different types (e.g. chaotic attractors coexisting with limit cycles [13]).

[6] This fraction does not depend on $\mathbf{x}$ since $f'(u_i) \geq 0$.

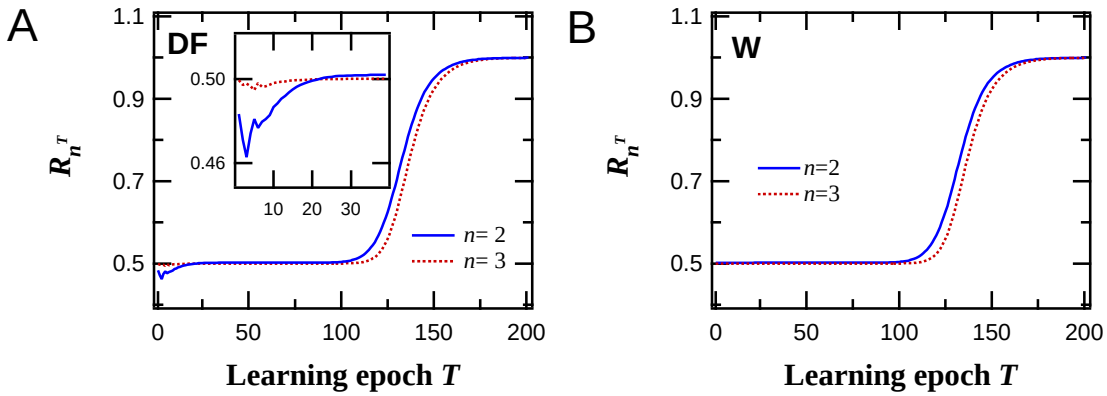

FIG. 2: Evolution of the weighted-fraction of positive circuits $R_n^{(T)}$ for loops in $D\mathbf{F}$ (A) and $\mathcal{W}$ (B) and circuit length $n = 2$ (thick line) and $n = 3$ (dotted line). The inset in (A) is a magnification of the first 40 learning epochs. Values are averages over 50 different networks using $\lambda = 0.90$. See text for definitions. All other parameters as in fig. 1.

Measurements of $R_2^{(T)}$ and $R_3^{(T)}$ during numerical simulations are presented in figure 2A. The initial value for the length-2 circuits $R_2^{(1)}$ is less than 0.5 ($R_2^{(1)} \approx 0.47$, see fig. 2A, inset), indicating a slight initial imbalance in favor of negative circuits over positive ones. According to the aforementioned theoretical considerations about feedback circuits, this would yield a tendency toward complex oscillatory dynamics, and may be considered another viewpoint to explain the initial chaotic dynamics. Note however that the initial imbalance in circuits of length-3 is much more modest, $R_3^{(1)} \approx 0.496$.

When the learning rule eq.(9) is applied, $R_2^{(T)}$ increases and converges to 0.50 in 10 to 20 learning epochs. Hence, one expects the corresponding dynamics attractors to become less chaotic and more periodic (if not fixed point). This is exactly the behavior observed in the simulations (fig. 4B). Hence, the study of the feedback circuits in the Jacobian matrix offers an explanation to the reduction of dynamics induced by learning at short learning epochs.

Upon further learning, $R_2^{(T)}$ and $R_3^{(T)}$ remain constant at 0.5 up to $T \approx 100$ learning epochs. Thus, these quantities do not detect variations in the balance between positive and negative circuits for $20 < T < 100$. However, at longer times ($T > 100$), $R_2^{(T)}$ and $R_3^{(T)}$ both increase abruptly and rapidly reach ≈ 1.0. Hence, at long learning epochs, the system rapidly switches to a state where the weight hold by negative circuits is essentially negligible with respect to the weight hold by positive circuits.

Note that the time course of these indicators for $T > 20$ closely follows the time course of the relative clustering index (fig. 1A). The causal relation between these two phenomena is however not obvious. Note also that our learning rule eq.(9) conserves the initial balance between positive and negative weights. Hence the relative decrease of the weights on the negative circuits is not caused by a specific decrease of the negative weights, but by preferential silencing of those negative weights implicated in negative circuits.

As explained in section III B, the form of the Jacobian matrix in our system implies that the sign of a feedback circuit is given by the sign of the weights along the circuit. Figure 2B shows the evolution of the weighted-fraction for circuits computed in $\mathcal{W}$, i.e. we compute here the weight of the feedback circuit $e_1, \ldots, e_k$ as the product of the *synaptic* weights of its edges. It is clear from this figure that the evolution of the weighted-fraction of positive circuits in $\mathcal{W}$ fails short of detecting the initial imbalance observed in the feedback circuits of $D\mathbf{F}$. However, the evolution of this fraction at long times $20 < T < 100$ is identical to that measured in $D\mathbf{F}$. Thus in our system, the weighted-fraction of positive circuits in $\mathcal{W}$ is not sensitive enough to detect the initial subtle changes of the dynamics but accounts the emergence of the fixed point dynamical regime at long times. Thus it may provide a link between purely structural and purely dynamical viewpoints.

B. Entropy reduction.

1. Evolution of the weight matrix.

From eq. (5) it is easy to show by recurrence that:

$$\mathcal{W}^{(T+1)} = \lambda^T \mathcal{W}^{(1)} + \frac{\alpha}{N} \sum_{n=1}^{T} \lambda^{T-n} \Gamma^{(n)}. \tag{18}$$

The evolution of the weight matrix under the influence of the generic learning rule eq.(4) originates thus from two additive contributions. If $\lambda < 1$, the "direct" contribution of $\mathcal{W}^{(1)}$ to $\mathcal{W}^{(T+1)}$ (the first term in the RHS member) decays exponentially fast. Hence the effect of λ is to allow the system to forget its initial synaptic structure. It gives the possibility to "rewire" entirely the network in a time scale proportional to $\frac{1}{|\log(\lambda)|}$. The second RHS term of eq.(18) corresponds to the new synaptic structure that emerges with learning and replaces the initial one (that fades away exponentially fast). Importantly, this second term includes contributions from each previous matrices $\Gamma^{(n)}, \quad \forall n \leq T$ (with an exponentially decreasing contribution λ^{T-n}). Hence, the emerging weight structure will depend on *the whole history of the neuron dynamics.*

Note also that if $\lambda = 1$ then this term may diverge, leading to a divergence of $\mathcal{W}^{(T)}$. Therefore, in this case, one has to add an artificial cut-off to avoid this unphysical divergence. If $\lambda < 1$, one expects, on the opposite, to reach a stationary regime where synaptic weights do not evolve anymore. Both matrices $\mathcal{W}^{(T)}$ and $\Gamma^{(T)}$ are expected to stabilize at long learning epochs to constant values: $\lim_{T\to\infty} \mathcal{W}^{(T)} = \mathcal{W}^{(\infty)} = \text{cste}$ and $\lim_{T\to\infty} \Gamma^{(T)} = \Gamma^{(\infty)} = \text{cste}$. This means that, if $\lambda < 1$, the dynamics will settle at long learning epochs onto a stable attractor that will not be modified by further learning. The existence of such a stationary distribution is provided by the sufficient condition:

$$\mathcal{W}^{(\infty)} = \frac{\alpha}{N(1-\lambda)} \Gamma^{(\infty)}. \tag{19}$$

We show in appendix B that, assuming moderate assumptions on h (eq. 6), $\|\Gamma^{(T)}\|$ can be upper-bounded $\forall T$ by a constant C, so that $\|\mathcal{W}^{(\infty)}\| \leq \alpha C / (N(1-\lambda))$. From eq.(18), an upper bound for the norm of $\mathcal{W}^{(T)}$ is trivially found:

$$\|\mathcal{W}^{(T+1)}\| \leq \lambda^T \|\mathcal{W}^{(1)}\| + \frac{\alpha}{N} \sum_{n=1}^{T} \lambda^{T-n} \|\Gamma^{(n)}\|, \tag{20}$$

where $\|\|$ is the operator norm (induced e.g. by Euclidean norm). Hence,

$$\|\mathcal{W}^{(T+1)}\| \leq \lambda^T \|\mathcal{W}^{(1)}\| + \frac{\alpha}{N} \frac{1-\lambda^T}{1-\lambda} C \leq \lambda^T \|\mathcal{W}^{(1)}\| + \frac{\alpha}{N} \frac{1}{1-\lambda} C. \tag{21}$$

This result shows that the major effect of the "Hebb-like" learning rule we study consists in an exponentially fast contraction of the norm of the weight matrix ($\alpha C / (N(1-\lambda))$ is a constant), which is due to the term λ, i.e. to passive forgetting ($\lambda < 1$).

Numerical simulations are in agreement with these analytical results. Note that the analytical results need not to be "confirmed" by numerical simulations since they are rigorous. But they only provide an upper bound that can be rough, while numerics allows us to be more precise. Actually, we did not measure the norm of the weight matrix during simulations, but the evolution of its spectral radius $|s_1^{(T)}|$. Let $s_i^{(T)}$ be the eigenvalues of $\mathcal{W}^{(T)}$, ordered such that $|s_1^{(T)}| \geq |s_2^{(T)}| \geq \ldots \geq s_i^{(T)} \geq \ldots$. Since $|s_1^{(T)}|$, the spectral radius of $\mathcal{W}^{(T)}$, is smaller than $\|\mathcal{W}^{(T)}\|$ one has from eq.(21):

$$|s_1^{(T+1)}| \leq \lambda^T \|\mathcal{W}^{(1)}\| + \frac{\alpha}{N} \frac{1}{1-\lambda} C. \tag{22}$$

This equation predicts an exponentially fast contraction of the spectral radius with time, controlled by the forgetting rate λ. Figure 3 shows the evolution of the spectral radius of $\mathcal{W}^{(T)}$ for different values of λ during numerical simulations (open symbols). The results confirm this decay of the spectral radius. Moreover, we also plot on this figure (full lines) exponential decays according to the first RHS member of eq.(22), i.e. $g(T) = |s_1^{(1)}| \lambda^T$. The almost perfect agreement with the measurements tells us that the bound above is not so bad.

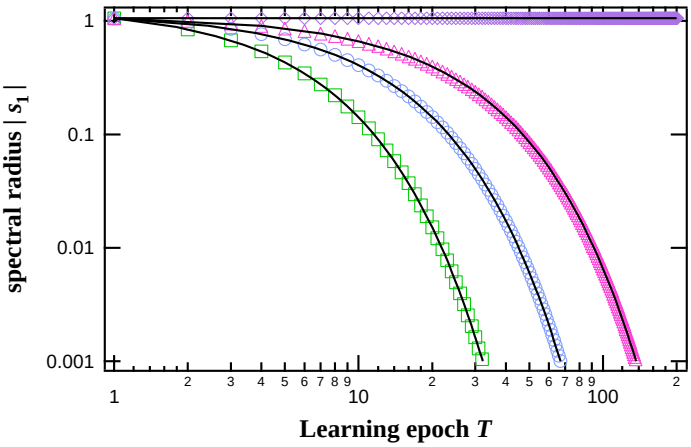

FIG. 3: The Hebb-like learning rule eq.(9) contracts the spectral radius of $\mathcal{W}$. The evolution during learning of the norm of $\mathcal{W}$ largest eigenvalue, $|s_1^{(T)}|$ is plotted on a log-log scale for, from bottom to top, $\lambda = 0.80$ (squares), 0.90 (circles), 0.95 (triangles) or 1.00 (diamonds). Each value is an average over 50 realizations with different initial conditions (initial weights and neuron states). Standard deviations are smaller than the symbols. Black full lines are plots of exponential decreases with equation $g(T) = |s_1^{(1)}|\lambda^T$. All other parameters as in fig. 1

2. *Jacobian matrices.*

Fix $\mathbf{x} \in [0,1]^N$. As indicated above, the Jacobian matrix has in our case a specific form $D\mathbf{F}_{\mathbf{x}}^{(T)} = \Lambda(\mathbf{u}^{(T)})\mathcal{W}^{(T)}$, where $\mathbf{u}^{(T)} = \mathcal{W}^{(T)}\mathbf{x} + \boldsymbol{\xi}$. A bound for the spectral radius of $D\mathbf{F}_{\mathbf{x}}^{(T)}$ can thus easily be derived. Call $\mu_i^{(T)}(\mathbf{x})$ the eigenvalues of $D\mathbf{F}_{\mathbf{x}}^{(T)}$ ordered such that $|\mu_1^{(T)}(\mathbf{x})| \geq |\mu_2^{(T)}(\mathbf{x})| \geq \ldots \geq |\mu_i^{(T)}(\mathbf{x})| \geq \ldots$. One has, $\forall \mathbf{x}$:

$$|\mu_1^{(T)}(\mathbf{x})| \leq \|D\mathbf{F}_{\mathbf{x}}^{(T)}\| \leq \|\Lambda(\mathbf{u}^{(T)})\|\|\mathcal{W}^{(T)}\|. \tag{23}$$

Since $\|\Lambda(\mathbf{u}^{(T)})\| = \max_i f'(u_i^{(T)})$ (Λ is diagonal and $f' > 0$), one finally gets

$$|\mu_1^{(T)}(\mathbf{x})| \leq \max_i f'(u_i^{(T)})\|\mathcal{W}^{(T)}\|. \tag{24}$$

Therefore, the spectrum of $D\mathbf{F}_{\mathbf{x}}^{(T)}$ can be contracted by two effects: the contraction of the spectrum of $\mathcal{W}^{(T)}$ and/or the decay of $\max_i f'(u_i)$ related to saturation effects. Indeed, $f'(u_i)$ is small if x_i is saturated to 0 or 1 (i.e. $|u_i|$ is large), but large whenever $|u_i|$ is intermediate, i.e. falls into the central, pseudo-linear part of the sigmoid $f(u_i)$. We have already evidenced below that $\lambda < 1$ yields a decrease of $\|\mathcal{W}^{(T)}\|$. Note that even if $\lambda = 1$ (no passive forgetting) and $\mathcal{W}^{(T)}$ diverges, then $\mathbf{u}^{(T)}$ diverges as well, leading $\max_i f'(u_i^{(T)})$ to vanish, thus decreasing anyway the spectral radius of the Jacobian matrix. Hence, if the initial value of $|\mu_1^{(T)}(\mathbf{x})|$ is larger[7] than 1, eq.(24) predicts that it may decrease down to a value < 1. We are dealing here with discrete time dynamical systems, so that the value $|\mu_1^{(T)}(\mathbf{x})| = 1$ locates a *bifurcation* of the dynamical system. Hence, eq.(24) opens up the possibility that learning drives the system through bifurcations. Figure 6 presents numerical simulations illustrating this.

One may however argue that eq. (24) depends on $\mathbf{x}$ and does not give any indication on the *typical* behavior of the dynamical system. This is the topic of the next section.

3. *A bound on the maximal Lyapunov exponent.*

In a "chaotic" system, the trajectory of a point $\mathbf{x}$ in the phase space is typically such that a small perturbation about $\mathbf{x}$ is either amplified or contracted according to the direction of this perturbation. The rate of contraction/expansion as

[7] In the limit $N \to \infty$, and in the case of random independent identically distributed weights with 0 mean and variance $\frac{1}{N}$, $|\mu_1^{(T)}(\mathbf{x})|$ converges almost surely to a value proportional to g, the proportionality factor depending on the explicit form of f [12, 24]

well as the corresponding eigendirections are respectively given by the eigenvalues and eigenvectors of $D\mathbf{F}_{\mathbf{x}}$. Therefore, they depend on $\mathbf{x}$ and are varying along the trajectory of $\mathbf{x}$. In the present case, the previous section has shown that these variations are basically due to variations in the local fields (synaptic current) received by each neuron, which induce variations in the saturation of the transfer function f and changes in its derivative f'. This saturation effects explains *why* it is not sufficient to consider the mere synaptic graph structure to infer information about dynamics. Nevertheless, one step further in the analysis would be to characterize the behavior of *typical* trajectories and how saturation effects act, *on average*, on dynamics. This information is provided by the computation of the largest Lyapunov exponent (see appendix for a definition).

In the present setting, the largest Lyapunov exponent, $L_1^{(T)}$ depends on the learning epoch T. The Lyapunov exponent $L_1^{(1)}$, before starting learning, can be computed exactly in the thermodynamic limit $N \to \infty$, using the fact that the W_{ij}'s are i.i.d. random variables [13], and one can show that it is positive provided g is sufficiently large. But because the weights deviate from i.i.d. random distribution under the influence of Hebb-like learning, the evolution of $L_1^{(T)}$ cannot be computed analytically as soon as $T > 1$. However, the following theorem (proven in appendix C) yields a useful upper-bound:

Theorem 1

$$L_1^{(T)} \leq \log(\|\mathcal{W}^{(T)}\|) + \left\langle \log(\max_i f'(u_i)) \right\rangle^{(T)}. \tag{25}$$

where $\langle \log(\max_i f'(u_i)) \rangle^{(T)}$ *denotes the time average of* $\log(\max_i f'(u_i))$*, in the learning epoch* T *(see appendix for details).*

This theorem emphasizes the two main effects that may contribute to a decrease of $L_1^{(T)}$. The first term in the RHS member states that $L_1^{(T)}$ will decrease if the norm of the weight matrix $\|\mathcal{W}^{(T)}\|$ decreases during learning. The second term is related to the saturation of neurons, as expected. However, the main difference with eq. (24) is that we have now an information on how saturation effects act *on average* on dynamics, via $\log(f')$. The second term in the RHS member is positive if some neurons have an average $\log(f')$ larger than 1 (that is, they are mainly dominated by amplification effects corresponding to the central part of the sigmoid) and becomes negative when all neurons are on average saturated. Note that this contraction/expansion effect can be addressed in the context of linear response theory [15].

In any case, it follows from this result that if learning increases the saturation level of neurons or decreases the norm of the weight matrix $\|\mathcal{W}^{(T)}\|$, then the result will be a decay of $L_1^{(T)}$, thus a possible transition from chaotic to simpler attractors. A canonical measure of dynamical complexity is the Kolmogorov-Sinai (KS) entropy which is bounded from above by the sum of positive Lyapunov exponents. Therefore, if the largest Lyapunov exponent decreases, KS entropy decreases.

On numerical grounds we observe the following. Fig. 4A shows measurements of $L_1^{(T)}$ during numerical simulations with different values of the passive forgetting rate λ. Its initial value is positive because we start our simulations with chaotic networks ($L_1^{(1)} \approx 0.21 \pm 0.10$). The Hebb-like learning rule eq.(9) indeed leads to a rapid decay of $L_1^{(T)}$, whose rate depends on λ. Hence $L_1^{(T)}$ shifts quickly to negative values, confirming the decrease of the dynamical complexity that could be inferred from visual inspection of fig. 4B.

To conclude, there is a systematic decay of $L_1^{(T)}$ induced by passive forgetting and a dynamically induced decay occurring whenever the level of saturation (measured e.g. by $-\langle \log(\max_i f'(u_i)) \rangle_T$) of the neurons increases.

C. Neuron activity.

We now present analytical results concerning the evolution of individual neuron activity. The learning rule (5) will result in a variation of the attractor structure which can be slight (e.g. away from a bifurcation) or sharp. These variations can be measured by changes in the average value of some relevant observable such as neuron activity. (More generally, learning will induce a variation in the SRB measure $\rho^{(T)}$, see appendix). Let $\delta\rho^{(T+1)}(\mathbf{x})$ be the variation of the average activity $\mathbf{x}$ between learning epoch T and $T+1$. By definition (see appendix):

$$\delta\rho^{(T+1)}(\mathbf{x}) = \langle \mathbf{x} \rangle^{(T+1)} - \langle \mathbf{x} \rangle^{(T)}. \tag{26}$$

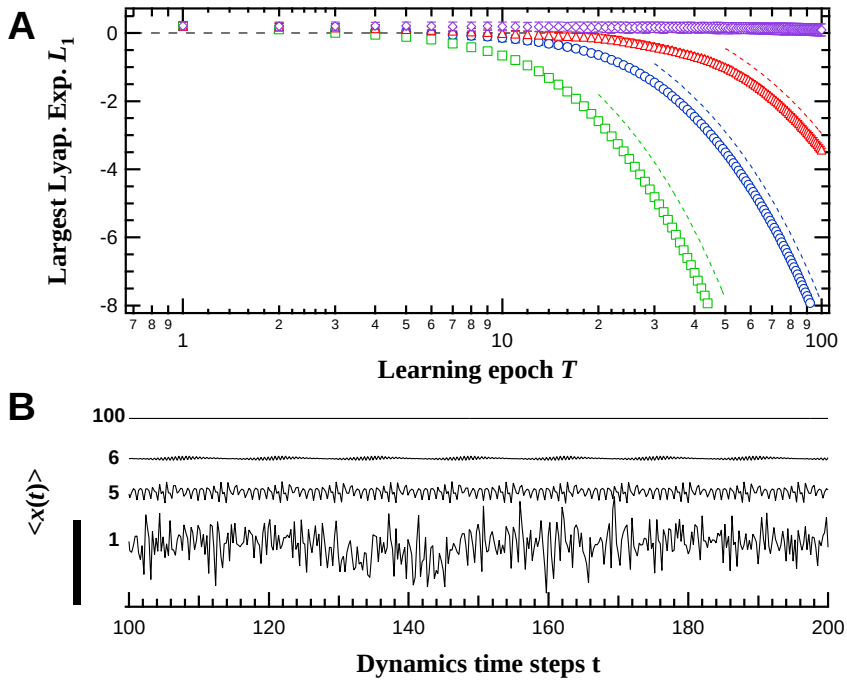

FIG. 4: The "Hebb-like" learning rule eq.(9) induces reduction of the dynamics complexity from chaotic to periodic and fixed point. (A) Evolution of the largest Lyapunov exponent L_1 during 100 learning epochs for, from bottom to top, $\lambda = 0.80$ (squares), 0.90 (circles), 0.95 (triangles) or 1.00 (diamonds). Each value is an average over 50 realizations with different initial conditions (initial weights and neuron states). Bars are standard deviations (and are mostly smaller than symbol size). The dashed lines illustrate λ decays of form $g(T) \propto T\log(\lambda)$ (see text). (B) Examples of network dynamics when learning is stopped at epoch (from bottom to top) $T = 1$ (initial conditions), 5, 6 or 100. These curves show the network-averaged state $\left\langle x^{(T)}(t)\right\rangle = 1/N\sum_{i=1}^{N} x_i^{(T)}(t)$ and are shifted on the y-axis for clarity. The height of the vertical bar represents an amplitude of 0.1. $N = 100$ and all other parameters are as in fig. 1.

We show in appendix D that the average value of the neuron local field, $\mathbf{u}$, at learning epoch T depends on four additive terms:

$$\langle \mathbf{u}\rangle^{(T+1)} = \lambda^T \langle \mathbf{u}\rangle^{(1)} + (1-\lambda^T)\boldsymbol{\xi} + \lambda \sum_{n=1}^{T} \lambda^{T-n} \mathcal{W}^{(n)} \delta\rho^{(n+1)}(\mathbf{x}) + \frac{\alpha}{N}\sum_{n=1}^{T} \lambda^{T-n}\Gamma^{(n)} \langle \mathbf{x}\rangle^{(n+1)}\,. \tag{27}$$

Provided that $\lambda < 1$, as $T \to +\infty$, time averages of observables converge to a constant. So that $\delta\rho^{(T)}(\mathbf{x}) \to 0$ and $\lim_{T\to+\infty} \langle \mathbf{x}\rangle^{(T)} = \langle \mathbf{x}\rangle^{(\infty)}$. Therefore, asymptotically:

$$\langle \mathbf{u}\rangle^{(\infty)} = \boldsymbol{\xi} + \mathbf{H}^{(\infty)}, \tag{28}$$

where:

$$\mathbf{H}^{(\infty)} = \frac{\alpha}{N(1-\lambda)}\Gamma^{(\infty)} \langle \mathbf{x}\rangle^{(\infty)}\,. \tag{29}$$

Therefore, the asymptotic local field is the sum of the *stimulus* (input pattern) plus an additional vector $\mathbf{H}^{(\infty)}$ which accounts for the history of the system.

This last term has an interesting structure in the case of the learning rule (10). Indeed, in this case:

$$\mathbf{H}^{(\infty)} = \frac{\alpha}{N(1-\lambda)}\mathbf{m}^{(\infty)}\left[\mathbf{m}^{(\infty)} H(\mathbf{m}^{(\infty)})\right]^{+} \langle \mathbf{x}\rangle^{(\infty)}\,,$$

so that:

$$H_i^{(\infty)} = \frac{\alpha}{N(1-\lambda)} \eta m_i^{(\infty)} \tag{30}$$

where :

$$\eta = \sum_{j, m_j^{(\infty)} > 0} m_j^{(\infty)} x_j^{(\infty)} = \sum_{j, x_j^{(\infty)} > d_j} (x_j^{(\infty)} - d_j) x_j^{(\infty)}, \tag{31}$$

can be interpreted as an *order parameter*. A large positive η corresponds to a system where neurons are mainly saturated to 1, while a small η corresponds to neuron whose average activity is close to d_i.

Note that η is related to a set of self-consistent equations. Indeed, since $x_i = f(u_i)$ one has:

$$u_i^{\infty} = \xi_i + \frac{\alpha}{N(1-\lambda)} \eta \left[\langle f(u_i) \rangle^{(\infty)} - d_i \right] \tag{32}$$

In the case where this constant asymptotic attractor is a fixed point (i.e. the attractor with smallest complexity), one has:

$$u_i^* = \xi_i + \frac{\alpha}{N(1-\lambda)} \eta (f(u_i^*) - d_i), \tag{33}$$

where $\mathbf{u}^*$ and $\mathbf{x}^*$ denote the values of $\mathbf{u}$ and $\mathbf{x}$, respectively, on the fixed point attractor. This simple case illustrates a very interesting point. We have a set of N nonlinear self-consistent equations (32) including a local term (u_i^{∞}) and a global term η. Assume that we slightly perturb the system, for example by removing the stimulus ξ_i at some point i. Then either the corresponding equation (32) is away from a bifurcation point and this only results in a slight change in u_i^*, or a bifurcation (typically saddle-node) occurs. This results in a big change in u_i^* *which may in turn result in a big change in* η. This change may then induce drastic activity change in the neurons $j \neq i$ by some avalanche-like mechanism. On practical grounds this means that the presentation (the removal) of some part of the input pattern may induce a drastic change of the dynamics.

Equations (28), (29) characterize the asymptotic regime $T \to \infty$ which is not the most exciting one according e.g. to fig 4,6. On intermediate time scales, eq. (27) must be used. It shows that the local field $\mathbf{u}$ contains a component which is the input pattern plus an additional term that can be weak or not. Figure 5 shows numerical simulations of the evolution of the local field $\mathbf{u}$ during learning. Clearly, while the initial values are random, the local field (thin full line) shows a marked tendency to converge to the input pattern (thick dashed line) after as soon as 10 learning epochs. The convergence is complete after ≈ 60 learning epochs. There remains nevertheless still an additional term corresponding to $\mathbf{H}^{(\infty)}$.

The fact that $\langle \mathbf{u} \rangle^{(T)}$ contains a component equal to $\boldsymbol{\xi}$ has a deep impact on the dynamics, as we now develop.

V. FUNCTIONAL VIEWPOINT

The basic function of RRNNs is to learn a specific pattern $\boldsymbol{\xi}$. In our terms, a pattern is learned when the dynamics settles onto a limit cycle associated to the input pattern. This reminds Freeman's way of defining a recognized odor by rabbits [22, 23]. This learning procedure is selective (i.e. only the chosen pattern is learned) and robust (i.e. a noisy version of the learned pattern should lead to an attractor similar to the one reached after presentation of the learned pattern) [18]. Furthermore, an important corollary is that removal of the learned pattern after learning should lead to a significative change in the network dynamics. We now proceed to an analysis of this latter property.

Label by $\mathbf{x}$ (resp. $\mathbf{u}$) the neuron firing rate (resp. local field) obtained when the (time constant) input patter $\boldsymbol{\xi}$ is applied to the network

$$\mathbf{x}^{(T)}(t+1) = f\left(\mathbf{u}^{(T)}(t)\right), \quad \mathbf{u}^{(T)}(t) = \mathcal{W}^{(T)} \mathbf{x}^{(T)}(t) + \boldsymbol{\xi} \tag{34}$$

and by $\mathbf{x}'$ (resp. $\mathbf{u}'$) the corresponding quantities when $\boldsymbol{\xi}$ is removed

$$\mathbf{x}'^{(T)}(t+1) = f\left(\mathbf{u}'^{(T)}(t)\right), \quad \mathbf{u}'^{(T)}(t) = \mathcal{W}^{(T)} \mathbf{x}'^{(T)}(t) \tag{35}$$

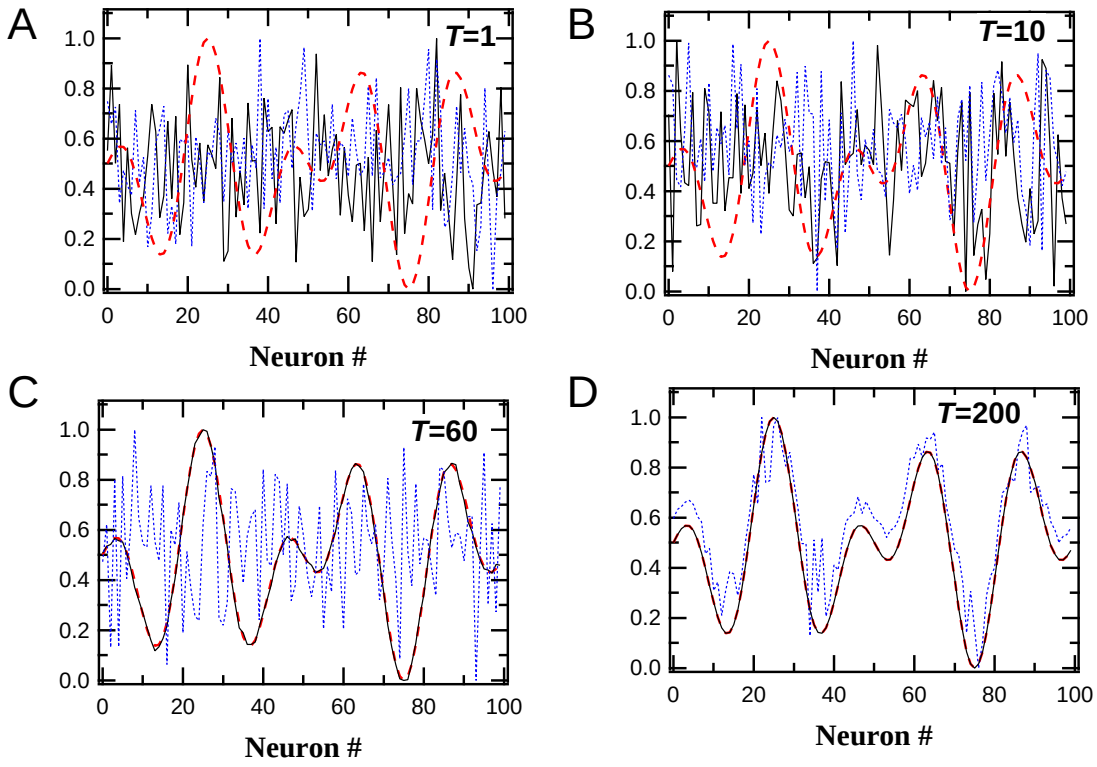

FIG. 5: The local field $\mathbf{u} = \mathcal{W}\mathbf{x} + \boldsymbol{\xi}$ (thin full line) and the real part of the first eigenvector of the Jacobian matrix (thin dotted line) converge to the input pattern $\boldsymbol{\xi}$ (thick dashed line) at intermediate-to-long learning epochs. Snapshot are presented at $T = 1$ (A, initial conditions), $T = 10$ (B), $T = 60$ (C) and $T = 200$ (D) learning epochs. Each curve plots averages over 50 realizations (standard deviations are omitted for clarity), and every vector has been normalized to $[0, 1]$ for clarity. All other parameters as in fig. 1

The removal of $\boldsymbol{\xi}$ will change the attractor structure and the average value of any observable ϕ (though the amplitude of this change depends on ϕ). More precisely call:

$$\Delta^{(T)}[\phi] = \langle \phi(\mathbf{x}') \rangle^{(T)} - \langle \phi(\mathbf{x}) \rangle^{(T)} \tag{36}$$

where $\langle \phi(\mathbf{x}') \rangle^{(T)}$ is the (time) average value of ϕ without $\boldsymbol{\xi}$ and $\langle \phi(\mathbf{x}) \rangle^{(T)}$ the average value with $\boldsymbol{\xi}$. There are two cases.

In the first case, the system is away from a bifurcation point and removal will result in a variation of $\Delta^{(T)}[\phi]$ that remains proportional to $\boldsymbol{\xi}$ provided $\boldsymbol{\xi}$ is sufficiently small. We emphasize that this result holds even if dynamics is chaotic and can be rigorously proven in the context of uniformly hyperbolic systems. This is the linear response theory developed by Ruelle [42]. In the present context, one finds for example that the variation of the average value of $\mathbf{u}$ is given by [15],[16]:

$$\Delta^{(T)}[\mathbf{u}] = -\chi^{(T)}\boldsymbol{\xi} \tag{37}$$

where $\chi^{(T)}$ is the matrix[8]:

[8] The convergence of this series is discussed in [15, 16, 42]. Note that a similar formula can be written for an arbitrary observable ϕ, but is more cumbersome.

$$\chi_{ij}^{(T)} = \mathcal{I} + \sum_{n=1}^{+\infty} \sum_{\gamma_{ij}(n)} \prod_{l=1}^{n} W_{k_l k_{l-1}} \left\langle \prod_{l=1}^{n} f'(u_{k_{l-1}}(l-1)) \right\rangle^{(T)} \tag{38}$$

where the sum $\sum_{\gamma_{ij}(n)}$ holds on every possible path $\gamma_{ij}(n)$ of length n, connecting neuron $k_0 = j$ to neuron $k_n = i$, in n steps.

Note therefore that $\Delta^{(T)}[\mathbf{u}] = -\boldsymbol{\xi} - M^{(T)}\boldsymbol{\xi}$ where the matrix $M^{(T)}$ integrates dynamical effects. A slight variation of u_i at $t = 0$ implies a reorganization of the dynamics which results in a complex formula for the variation of $\langle \mathbf{u} \rangle^{(T)}$, even if the dominant term is $\boldsymbol{\xi}$, as expected. More precisely, as emphasized several times above, one remarks that each path in the sum $\sum_{\gamma_{ij}(n)}$ is weighted by the product of a *topological* contribution depending only on the weights W_{ij} and on a *dynamical* contribution. The weight of a path $\gamma_{ij}(\tau)$ depends on the average value of $\left\langle \prod_{l=1}^{n} f'(u_{k_{l-1}}(l-1)) \right\rangle^{(T)}$ thus on *correlations* between the state of saturation of the units $k_0, \ldots, k_{n-1}$ at times $0, \ldots, n-1$. This formula shows how the effects of a pattern removal are complex when dealing with a chaotic dynamics.

On the opposite, close to a bifurcation point this variation is typically not proportional to $\boldsymbol{\xi}$ and may lead to drastic changes in the dynamics. From the analysis above, we therefore expect pattern removal to have a maximal effect at "the edge of chaos", namely when the (average) value of the spectral radius of $D\mathbf{F}_{\mathbf{X}}$ is close to 1. However, in this case, one cannot use the linear response formula above (the series diverges). One can however use it *close* to the edge of chaos and investigate how it diverges when approaching the critical point. This is however a huge work that will be the topic of a forthcoming paper. In the present one, we shall focus on the case where dynamics has converged to a stable fixed point $\mathbf{u}^{*(T)}$ (resp. $\mathbf{x}^{*(T)}$) (namely, when $L_1^{(T)} < 0$, see e.g. Fig. 4). Note that this point depends nevertheless on the learning epoch. In this case, eq. (37) reduces to:

$$\Delta^{(T)}[\mathbf{u}] = -\sum_{n=0}^{\infty} \left(\mathcal{W}^{(T)} \Lambda(\mathbf{u}^*) \right)^n \boldsymbol{\xi} \tag{39}$$

Calling $\lambda_k, \mathbf{v}_k$ the eigenvalues and eigenvectors of $\mathcal{W}^{(T)}\Lambda(\mathbf{u}^{*(T)})$, ordered such that $|\lambda_N| \leq |\lambda_{N-1}| \leq |\lambda_1| < 1$ one obtains:

$$\Delta^{(T)}[\mathbf{u}] = -\sum_{k=1}^{N} \frac{(\mathbf{v}_k, \boldsymbol{\xi})}{1 - \lambda_k} \mathbf{v}_k \tag{40}$$

where $(\ ,\ )$ denotes the inner product. Actually, this result can be easily found without using linear response, by a simple Taylor expansion (see appendix E). As a matter of fact, the RHS diverges if $\lambda_1 = 1$ and if $(\mathbf{v}_1, \boldsymbol{\xi}) > 0$.

From this analysis, we therefore expect pattern removal to have a maximal effect at "the edge of chaos", namely when the value of the spectral radius[9] of $D\mathbf{F}_{\mathbf{X}}$ is close to 1. As mentioned above, the effects will be however more or less prominent according to the choice of the observable ϕ. We found numerically that the effects where particularly prominent with the following quantity:

$$\Delta^{(T)}[\Lambda] = \frac{1}{N} \sqrt{\sum_{i=1}^{N} \left(\langle \Lambda_{ii}(\mathbf{u}) \rangle^{(T)} - \langle \Lambda_{ii}(\mathbf{u}') \rangle^{(T)} \right)^2} \tag{41}$$

Indeed, recall that $\Lambda_{ii} = f'(u_i)$ is maximal when the local field of i falls in the central pseudo-linear part of the transfer function, hence when neuron i is the most sensitive to its input. Hence $\Delta^{(T)}[\Lambda]$ measures how neuron excitability is modified when the pattern is removed. The evolution of $\Delta^{(T)}[\Lambda]$ during learning with rule eq.(9) is

[9] There is a subtlety here. We have $D\mathbf{F}_{\mathbf{X}} = \Lambda(\mathbf{u})\mathcal{W}$, while in formula (40) we consider the eigenvalues of $\mathcal{W}\Lambda(\mathbf{u})$. However, if $\lambda_k, \mathbf{v}_k$ are eigenvalues and eigenvectors of $\mathcal{W}\Lambda(u)$ then $\Lambda(u)\mathcal{W}\Lambda(u)\mathbf{v}_k = D\mathbf{F}_{\mathbf{X}}\Lambda(u)\mathbf{v}_k = \lambda_k \Lambda(u)\mathbf{v}_k$. Therefore, $\lambda_k, \Lambda(u)\mathbf{v}_k$ are eigenvalues and eigenvectors of $D\mathbf{F}_{\mathbf{X}}$.

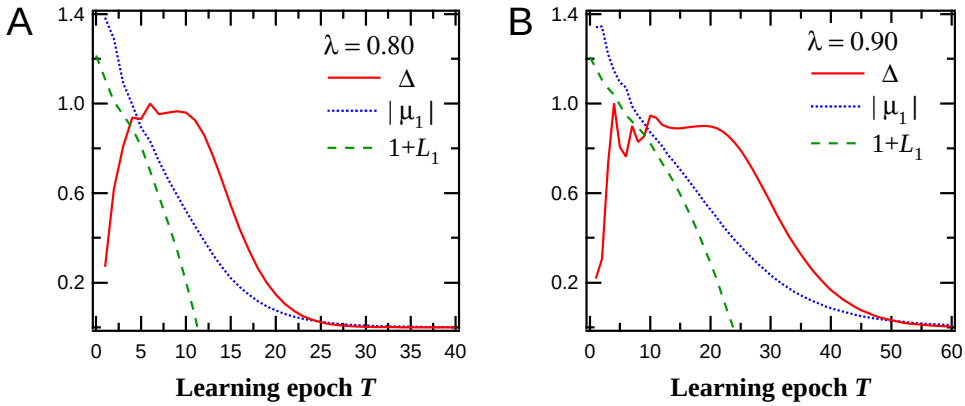

FIG. 6: The network sensitivity to the input pattern is maximal close to a bifurcation. The evolution of the average value for the spectral radius of $D\mathbf{F}_{\mathbf{x}}^{(T)}$ during learning (dotted line) is plotted together with the sensitivity measure $\Delta^{(T)}[\Lambda]$ (full line) for $\lambda = 0.80$ (A) or 0.90 (B). The panels also display the corresponding evolution of the largest Lyapunov exponent L_1, plotted as $1.0 + L_1$ for obvious comparison purpose (dashed line). The values of $\Delta^{(T)}[\Lambda]$ are normalized to the $[0-1]$ range for comparison purposes. Each value is an average over 50 realizations (standard deviations are omitted for clarity). All other parameters were as in fig. 1

shown on fig. 6 (full lines) for two values of the passive forgetting rate λ. $\Delta^{(T)}[\Lambda]$ is found to increase to a maximum at early learning epochs, and vanishes afterwards. Interestingly, comparison with the decay of the leading eigenvalue of the Jacobian matrix, μ_1 (dotted lines) shows that the maximal values of $\Delta^{(T)}[\Lambda]$ are obtained when $|\mu_1|$ is close to 1. Hence these numerical simulations confirm 02that sensitivity to pattern removal is maximal when the leading eigenvalue is close to 1. Hence, *"Hebb-like" learning drives the global dynamics through a bifurcation, in the neighborhood of which sensitivity to the input pattern is maximal.* We think this property is crucial regarding the memory properties of RRNNs, that must be able to detect whether a learned pattern is applied to the network, i.e. whose response must be very different when a learned pattern is applied and when no learned pattern is applied. This property is obtained here by the fact that pattern removal drives the system from one side of a bifurcation manifold to the other, hence yielding major changes in the network dynamics. We also note that this property is obtained at the frontier where the chaotic strange attractor begins to destabilize ($|\mu_1| = 1$), hence at the so-called "edge of chaos".

We showed in section IV B that our "Hebb-like" learning rule contracts the spectral radius of $D\mathbf{F}_{\mathbf{X}}, \forall \mathbf{x}$, so that the latter crosses the value 1 at some learning epoch. So, at some point in the learning process, 1 is ensured to be an eigenvalue of $D\mathbf{F}_{\mathbf{X}}$. The evolution of v_1, the eigenvector associated to the leading eigenvalue of the Jacobian matrix μ_1, is less obvious. We plot on fig. 5 (dotted lines) the evolution of its real part during numerical simulations (actually, its imaginary part vanishes after just a couple of learning epochs). It is clear from the numerical simulation results presented in this figure that the possibility of a vanishing projection of the input pattern $\boldsymbol{\xi}$ (thick dashed line) on v_1 can be ruled out (the two vectors are not orthogonal). The tendency is even opposite, i.e. v_1 is found to align on the input pattern at long learning epochs ($T \gtrsim 100$).

VI. DISCUSSION & CONCLUSION

A. Generalization to other learning rules

The coupled dynamical system studied in the present paper (eqs.(1) and (4)) displays several properties that allow rigorous mathematical study. However, many of the obtained results remain valid when some of these properties are relaxed. We give here a brief overview of the arguments related to this point. As already stated in the introduction, we do not pretend to encompass the spectrum of complexity and richness of biological learning rules. However, the properties of the studied learning rule are not unrealistic from a biological point of view.

A property of the learning rule eq.(4) is passive forgetting ($\lambda < 1$). From a biological perspective, this property is motivated by the fact that synaptic plasticity at the single synapse level is not permanent. Indeed, modifications

of the synaptic weights are locally embodied by modifications of the protein content or post-translational state in the synapse. Both are subject to a possibly rapid molecular turnover, and may thus rapidly vanish in the absence of plasticity-inducing stimuli, such as pre- and postsynaptic activity. Accordingly, several experimental works about the lifespan of cellular memory storage have evidence short durations. For instance, some forms of presynaptic plasticity at Schaffer collateral-CA1 synapses have been demonstrated to fade away (after induction) with a characteristic time of 20 mn [54] or even 20 sec [11]. This could be accounted for in our model by $\lambda \ll 1$, in which case the entirety of our analytical results apply.

Most studies about long-term plasticity forms have evidenced longer cellular memory time constants, ranging from hours to days [19, 30, 40]. This would correspond in our model to higher λ values, if not to $\lambda = 1$. In this case, the contraction of the spectral radius of the weight matrix can be less prominent. More precisely, its effect will reveal on longer time scales. As a matter of fact, the question is not so much to know what is "the value of λ" in real neural networks, but how the characteristic time scale $\frac{1}{|\log(\lambda)|}$ compares to other time scales in the system. In previous studies, we have considered Hebb-like learning rules devoid of passive forgetting (i.e. with $\lambda = 1$) [7, 47]. Numerical simulations of these rules have clearly evidenced a reduction of the attractor complexity during learning, in agreement with our present analytical results. In this case, the reduction of the attractor complexity is provoked by an increase of the average saturation level of the neurons.

Another assumption of our generic "Hebb-like" rule eq.(4) is that $\Gamma_{ij} = 0$ whenever the presynaptic neuron is silent. As already mentioned section II, an interpretation of this assumption is that the learning rule excludes heterosynaptic LTD. Our analytical results on the contraction of the spectral radius may remain valid when heterosynaptic LTD is accounted for, but this requires an extension of the model definition and further mathematical developments. This statement is supported by numerical simulations. We ran numerical simulations using a variant of eq.(4) in which the Heavyside term (that forbids heterosynaptic LTD in eq. 4) was omitted. The simulation results (not shown) were in agreement with all the analytical results supported here, including those on spectral radius contraction.

B. Small-world structure

The results we obtained concerning the application of the classical structural statistics from the complex networks field (clustering index, mean shortest path) to study the structural changes induced by learning, deserves further comment. There is emerging experimental evidence that numerous brain anatomical and functional connectivity networks at several length scales share a common small world connectivity topology (for a recent review, see [5]). Quantifications of the physical [44] or functional [8] connectivity of neuronal networks grown in vitro demonstrated small-world organizations. Direct quantifications of the anatomical connectivity of *Caenorhabditis elegans* full neural system [55] or, at larger scale, cortico-cortical area connectivity maps in macaque, cat [50] and more recently human [28], also concluded to the same type of connectivity. Moreover, quantitative studies of *functional* human brain connectivity based on MEG [51], EEG [38] or fMRI data [2, 21] also concluded to such a "small-world" organization. Current hypothesis for the frequent observation of small-world connectivity in real biological networks state that they may result from natural evolution, because small-world networks may be good trade-offs to minimize wiring length while preserving low energy costs and high local integration of the neurons [5, 34].

An alternative attractive hypothesis would be that small-world networks emerge spontaneously from neural networks subject to Hebbian learning. In fact, this small-world phenomenon may be explained by the conjunction of the basic properties of the "Hebb-like" learning rule and the possibility of network rewiring (basically due in our case to the total connectivity). Indeed, if neurons i and j are each highly correlated to a third neuron k, then both W_{ik} and W_{jk} will increase (in absolute value) by virtue of the principle that "synapses between neurons whose activities are correlated are strengthened". But, in this case, i and j will as well be correlated, so that the synapse between i and j will be strengthened. This leads naturally to clustered structures and thus to small-world correlations in the adjacency matrices.

In favor of this possibility, small-world connectivity has recently been shown to emerge spontaneously from spiking neuron networks with STDP plasticity and total connectivity [45] or with correlation-based rewiring [36]. Likewise, using the dynamical system studied in the present paper, but with $\lambda = 1$ (eq.9) and $d_i = 0$ (eq.7), we previously evidenced the emergence of a "strong" (i.e. $C/C_{rand} > 1.60$) small-world connectivity during learning [7]. Our present results however indicate that these kinds of interpretation should be taken with great care. For instance, with the parameter set used in the present paper, network connectivity, even at long learning epochs, deviates only slightly from its initial uncorrelated random organization. Hence, emergence of small-world connectivity, even in computational models of neural networks (i.e. not to speak about real neural networks), may be restricted to certain areas of the learning rule parameter space.

C. Mesoscopic level

We deal here with neuron networks of several thousands of neurons. One of our ideas here it is that there is an intermediate level between the individual neuron, or the very small circuit, and the whole brain. This intermediate, "mesoscopic", level provides an interesting entry into the explanation of the combination of various features or actions. The combination of several RRNN's has not yet be done and will be the subject of a future study.

D. Conclusion

Here, our aim was to study a family of neural networks with mutual coupling between neuron activity and synaptic structure and for which the mathematical study of the effects of "Hebb-like" learning could be accessible. Our major contributions can be summarized as follows. We show that in learning recurrent neural networks, the classical structural statistics from complex networks are not sufficient to explain the causal relationships between the evolution of the network weight structure and its dynamics. Conversely, we show that a dynamical point of view (i.e. employing analysis tools from dynamical systems and graph theory) is able to give a thorough mathematical account of the system behavior. Hence, the main message of this study would be that the behavior of these systems can be understood if the starting point of the analysis is the dynamics point of view (i.e. the Jacobian matrix) rather than the structural one (i.e. the weights or adjacency matrix). Whereas a large amount of work has been devoted to the mathematical study of neural network dynamics in the absence of learning, rigorous mathematical results in the case where neuron dynamics is mutually coupled to synaptic weight changes are much more difficult to obtain. The present paper might thus be considered a step forward to the analytical study of these systems.

Acknowledgments

This work was supported by a grant of the French National Research Agency, project JC05_63935 "ASTICO".

APPENDIX A: DEFINITIONS.

When dealing with chaotic systems, one is faced with the necessity to defining indicators measuring dynamical complexity. There are basically two families of indicators: one is based on topological properties (e.g. topological entropy), the other is based statistical properties (e.g. Lyapunov exponents or Kolmogorov-Sinai entropy). The latter family is more thoroughly understood than the former, since it can be addressed numerically or experimentally by *time averages* of relevant observables along typical trajectories of the dynamical system. However, to this aim, one has to assume a strong ergodic property: the time average of observables, along trajectories corresponding to initial conditions drawn at random with respect to a probability distribution having a density (with respect to the Lebesgue measure), is constant (it does not depend on the initial condition). This property is far from being evident. Actually, we are not able to prove it in the present context. On mathematical grounds, it corresponds to the following assumption.

Assumption 1 *Call μ_L is the Lebesgue measure on $[0,1]^N$ and let $\boldsymbol{F}^{*t}\mu_L$ the image of μ_L under $\boldsymbol{F}^t$. We assume that the following limit exists:*

$$\rho^{(T)} = \lim_{\tau\to\infty} \frac{1}{\tau}\sum_{t=1}^{\tau} \boldsymbol{F}^{*t}\mu_L \tag{A1}$$

where the probability measure $\rho^{(T)}$ is called "the Sinai-Ruelle-Bowen (SRB) measure at learning epoch T" [10, 41, 46]. Under this assumption the following holds. Let $\phi : [0,1]^N \to \mathbb{R}^N$ be some suitable (measurable) function. Then the time average:

$$\bar{\phi}[\boldsymbol{x}^{(T)}(0)] \stackrel{\text{def}}{=} \lim_{\tau\to\infty} \frac{1}{\tau}\sum_{t=1}^{\tau} \phi(\boldsymbol{x}^{(T)}(t)), \tag{A2}$$

where $\boldsymbol{x}(t) = \boldsymbol{F}^t(\boldsymbol{x})$, is equal to the ensemble average:

$$\langle\phi\rangle^{(T)} \stackrel{\text{def}}{=} \int_{[0,1]^N} \phi(\boldsymbol{x})\rho^{(T)}(d\boldsymbol{x}), \tag{A3}$$

for Lebesgue-almost every initial condition $\boldsymbol{x}^{(T)}(0)$.

In other words, time average and ensemble average are identical on practical grounds. The use of $\rho^{(T)}$ is required to prove the mathematical results below while time average is what we use for numerical simulations.

Note that in doing so, we have constructed a family of probability distributions $\rho^{(T)}$ that depends on the *time epoch* T. $\rho^{(T)}$ provides statistical information about the attractor structure. A prominent example is the maximal Lyapunov exponent. Let $\mathbf{x} \in [0,1]^N$, $\mathbf{v} \in \mathbb{R}^N$ and ρ be an SRB measure. Then, the largest Lyapunov exponent is given by:

$$L_1^{(T)} = \lim_{t\to\infty} \lim_{\|\mathbf{v}\|\to 0} \frac{1}{t} \log\left(\frac{\|D\mathbf{F}_{\mathbf{X}}^t \mathbf{v}\|}{\|\mathbf{v}\|}\right) \tag{A4}$$

Its value is constant for $\rho^{(T)}$ almost every $\mathbf{x}$. (Note indeed that the LHS does not depend on $\mathbf{x}, \mathbf{v}$, while the RHS does. This is a direct consequence of the assumption that $\rho^{(T)}$ is an SRB measure).

APPENDIX B: ASYMPTOTIC BEHAVIORS

In the specific learning rule eq.(9) used in our numerical simulations, $\Gamma_{ij} = m_i m_j H(m_j)$. Thus

$$\begin{aligned}
\|\Gamma\| &= \sup_x \frac{\|\Gamma\mathbf{x}\|}{\|\mathbf{x}\|} && \text{(B1)}\\
&= \sup_x \frac{\|\mathbf{m}\left[\mathbf{m}H(\mathbf{m})\right]^\top \mathbf{x}\|}{\|\mathbf{x}\|} && \text{(B2)}\\
&\leq \|\mathbf{m}\|\,\|\left[\mathbf{m}H(\mathbf{m})\right]^\top\| && \text{(B3)}\\
&\leq \left(\sum_{i=1}^N m_i^2\right)^{1/2} \left(\sum_{j=1,m_j>0}^N m_j^2\right)^{1/2} && \text{(B4)}\\
&\leq \sqrt{N}\sqrt{N}\phi^{1/2} && \text{(B5)}\\
&\leq N\sqrt{\phi} && \text{(B6)}
\end{aligned}$$

where $[\mathbf{v}]^\top$ denotes the transpose of vector $\mathbf{v}$, $\sum_{j=1,m_j>0}$ denotes a sum restricted to the active neurons and ϕ is the fraction of active neurons. Hence

$$\|\Gamma^{(T)}\| \leq N\sqrt{\phi^{(T)}} \tag{B7}$$

If (as observed in our numerical simulations) $\phi^{(T)}$ tends to a stationary value $\phi^{(\infty)}$ then

$$\|\Gamma^{(T)}\| \leq N\sqrt{\phi^{(\infty)}} \tag{B8}$$

Hence Γ is bounded in the specific case of eq.(9) by a constant $C = N\sqrt{\phi^{(\infty)}}$.

More generally, $\|\Gamma\|$ is bounded provided that the function h in (6) is bounded as well.

APPENDIX C: PROOF OF THEOREM 1

Let $\mathbf{v}, \mathbf{x} \in \mathbb{R}^N$. Denote by $\mathbf{x}(t) = \mathbf{F}^t(\mathbf{x})$, and $\mathbf{v}(t) = D\mathbf{F}_{\mathbf{x}(t)}.D\mathbf{F}_{\mathbf{X}}^{t-1}.\mathbf{v}$, $\mathbf{v}(0) = \mathbf{v}$. From the chain rule:

$$\frac{\|D\mathbf{F}_{\mathbf{X}}^t \mathbf{v}\|}{\|\mathbf{v}\|} = \frac{\|D\mathbf{F}_{\mathbf{X}(t)}\mathbf{v}(t-1)\|}{\|\mathbf{v}(t-1)\|} \frac{\|\mathbf{v}(t-1)\|}{\|\mathbf{v}\|}$$

$$= \frac{\|D\mathbf{F}_{\mathbf{X}(t)}\mathbf{v}(t-1)\|}{\|\mathbf{v}(t-1)\|}\frac{\|D\mathbf{F}_{\mathbf{X}(t-1)}\mathbf{v}(t-2)\|}{\|\mathbf{v}(t-2)\|}\cdots\frac{\|D\mathbf{F}_{\mathbf{X}(1)}\mathbf{v}\|}{\|\mathbf{v}\|}$$

Therefore:

$$L_1^{(T)} = \lim_{t\to\infty}\lim_{\|\mathbf{v}\|\to 0}\frac{1}{t}\sum_{n=1}^{t}\log\left(\frac{\|D\mathbf{F}_{\mathbf{X}(n)}\mathbf{v}(n-1)\|}{\|\mathbf{v}(n-1)\|}\right).$$

Since $\|A\mathbf{v}\| \leq \|A\|\|\mathbf{v}\|$:

$$L_1^{(T)} \leq \lim_{t\to\infty}\frac{1}{t}\sum_{n=1}^{t}\log\left(\|D\mathbf{F}_{\mathbf{X}(n)}\|\right) = \langle\log\left(\|D\mathbf{F}_{\mathbf{X}}\|\right)\rangle^{(T)} \ \ \rho^{(T)} - \text{almost surely.}$$

But since $D\mathbf{F}_{\mathbf{X}} = \Lambda(\mathbf{u})\mathcal{W}$, we have $\|D\mathbf{F}_{\mathbf{X}}\| \leq \|\mathcal{W}\|\|\Lambda(\mathbf{u})\| \leq \|\mathcal{W}\|\max_i(f'(u_i))$.

APPENDIX D: LOCAL FIELDS

Fix $\mathbf{x}$ and the time epoch T. Set $\mathbf{u} = \mathcal{W}^{(T)}\mathbf{x} + \boldsymbol{\xi}$. The average of $\mathbf{u}$, $\langle\mathbf{u}\rangle^{(T)}$ is defined either by the time average (A2) or by the ensemble average (A3). However, since $\mathcal{W}^{(T)}$ is constant during a given learning epoch one has:

$$\langle\mathbf{u}\rangle^{(T)} = \mathcal{W}^{(T)}\langle\mathbf{x}\rangle^{(T)} + \boldsymbol{\xi}, \quad \forall T. \tag{D1}$$

Therefore:

$$\langle\mathbf{u}\rangle^{(T+1)} = \mathcal{W}^{(T+1)}\langle\mathbf{x}\rangle^{(T+1)} + \boldsymbol{\xi} = (\lambda\mathcal{W}^{(T)} + \frac{\alpha}{N}\Gamma^{(T)})(\langle\mathbf{x}\rangle^{(T)} + \delta\rho^{(T+1)}(\mathbf{x}) + \boldsymbol{\xi},$$

where $\delta\rho^{(T+1)}(\mathbf{x}) \stackrel{\text{def}}{=} \langle\mathbf{x}\rangle^{(T+1)} - \langle\mathbf{x}\rangle^{(T)}$ is the difference of the average value of $\mathbf{x}$ between learning epochs $T+1$ and T.

Thus:

$$\langle\mathbf{u}\rangle^{(T+1)} = \lambda\langle\mathbf{u}\rangle^{(T)} + (1-\lambda)\boldsymbol{\xi} + \lambda\mathcal{W}^{(T)}\delta\rho^{(T+1)}(\mathbf{x}) + \frac{\alpha}{N}\Gamma^{(T)}\langle\mathbf{x}\rangle^{(T+1)},$$

and by recurrence:

$$\langle\mathbf{u}\rangle^{(T+1)} = \lambda^T\langle\mathbf{u}\rangle^{(1)} + (1-\lambda^T)\boldsymbol{\xi} + \lambda\sum_{n=1}^{T}\lambda^{T-n}\mathcal{W}^{(n)}\delta\rho^{(n+1)}(\mathbf{x}) + \frac{\alpha}{N}\sum_{n=1}^{T}\lambda^{T-n}\Gamma^{(n)}\langle\mathbf{x}\rangle^{(n+1)} \tag{D2}$$

APPENDIX E: PROOF OF EQ.(40)

Call $\mathbf{u}^{*(T)}$ ($\mathbf{u}'^{*(T)}$) the fixed point (for the variable $\mathbf{u}$) with (without) $\boldsymbol{\xi}$. We have:

$$\mathbf{u}'^{*(T)} = \mathcal{W}\mathbf{F}(\mathbf{u}'^{*(T)})$$

and:

$$\mathbf{u}^{*(T)} = \mathcal{W}\mathbf{F}(\mathbf{u}^{*(T)}) + \boldsymbol{\xi}$$

Therefore:

$$\mathbf{u}'^{*(T)} - \mathbf{u}^{*(T)} = \delta\mathbf{u}^{(T)} = \mathcal{W}\left[\mathbf{F}(\mathbf{u}^{*(T)} + \delta\mathbf{u}^{(T)}) - \mathbf{F}(\mathbf{u}^{*(T)})\right] - \boldsymbol{\xi}.$$

A series expansion yields, to the linear order:

$$(\mathcal{I} - \mathcal{W}\Lambda(\mathbf{u}^{(T)}))\delta\mathbf{u}^{(T)} = -\boldsymbol{\xi}$$

Decomposing on the eigenbasis $\mathbf{v}_k$ of $\mathcal{W}\Lambda(\mathbf{u}^{(T)})$ we obtain:

$$(1 - \lambda_k)(\delta\mathbf{u}^{(T)}, \mathbf{v}_k) = -(\boldsymbol{\xi}, \mathbf{v}_k) \tag{E1}$$

which corresponds to eq. (40) *provided* $|\lambda_k| < 1$ (ensuring that the matrix $\mathcal{I} - \mathcal{W}\Lambda(\mathbf{u}^{(T)})$ is invertible).

---

[1] W. C. Abraham and M. F. Bear. Metaplasticity: the plasticity of synaptic plasticity. *Trends Neurosci.*, 19:126–130, 1996.

[2] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.*, 26:63–72, 2006.

[3] F. Atay, T. Biyiko**u**glu, and J. Jost. Network synchronization : Spectral versus statistical properties. *Physica D*, 224:35–41, 2006.

[4] M. Barahona and L. Pecora. Synchronization in small-world systems. *Phys. Rev. Lett.*, 89:054101, 2002.

[5] D. Bassett and E. Bullmore. Small-world brain networks. *The Neuroscientist*, 12:512–523, 2006.

[6] M. Bear and W. Abraham. Long-term depression in hippocampus. *Annu. Rev. Neurosci.*, 19:437–462, 1996.

[7] H. Berry and M. Quoy. Structure and dynamics of random recurrent neural networks. *Adaptive Behavior*, 14:129–137, 2006.

[8] L. Bettencourt, G. Stephens, M. Ham, and G. Gross. Functional structure of cortical neuronal networks grown in vitro. *Phys. Rev. E*, 75:021915, 2007.

[9] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex networks : Structure and dynamics. *Physics Reports*, 424:175–308, 2006.

[10] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470. Berlin: Springer-Verlag, 1975.

[11] D. Brager, X. Cai, and S. Thompson. Activity-dependent activation of presynaptic protein kinase c mediates post-tetanic potentiation. *Nature Neurosci.*, 6:551–552, 2003.

[12] B. Cessac. Absolute stability criteria for random asymmetric neural networks. *J. of Physics A*, 27:927–930, 1994.

[13] B. Cessac. Increase in complexity in random neural networks. *J. de Physique*, 5:409–432, 1995.

[14] B. Cessac, B. Doyon, M. Quoy, and M. Samuelides. Mean-field equations, bifurcation map and route to chaos in discrete time neural networks. *Physica D*, 74:24–44, 1994.

[15] B. Cessac and J. Sepulchre. Stable resonances and signal propagation in a chaotic network of coupled units. *Phys. Rev. E*, 70:056111, 2004.

[16] B. Cessac and J. Sepulchre. Transmitting a signal by amplitude modulation in a chaotic network. *Chaos*, 16:013104, 2006.

[17] L. d. F. Costa, O. Sporns, L. Antiqueira, M. d. G. V. Nunes, and O. N. Oliveira, Jr. Correlations between structure and dynamics in complex networks. *ArXiv Physics e-prints*, 2006.

[18] E. Dauce, M. Quoy, B. Cessac, B. Doyon, and M. Samuelides. Self-organization and dynamics reduction in recurrent networks: stimulus presentation and learning. *Neural Networks*, 11:521–533, 1998.

[19] V. Doyere, M. Errington, S. Laroche, and T. Bliss. Low-frequency trains of paired stimuli induce long-term depression in area ca1 but not in dentate gyrus of the intact rat. *Hippocampus*, 6:52–57, 1996.

[20] B. Doyon, B. Cessac, M. Quoy, and M. Samuelides. Chaos in neural networks with random connectivity. *Int. Journ. of Bif. and Chaos*, 3(2):279–291, 1993.

[21] V. Eguiluz, D. Chialvo, G. Cecchi, and A. Apkarian. Scale-free brain functional networks. *Physical Review Letters*, 94:018102, 2005.

[22] W. Freeman. Simulation of chaotic eeg pattern with a dynamic model of the olfactory system. *Biol. Cyber.*, 56:139–150, 1987.

[23] W. Freeman, Y. Yao, and B. Burke. Central pattern generating and recognizing in olfactory bulb: a correlation learning rule. *Neur. Networks*, 1:277–288, 1988.

[24] V. Girko. Circular law. *Theor. Prob. Appl*, 29:694–706, 1984.

[25] J. Gouzé. Positive and negative circuits in dynamical systems. *Journ. Biol. Syst.*, 6(1):11–15, 1998.

[26] G. Grinstein and R. Linsker. Synchronous neural activity in scale-free network models versus random network models. *PNAS*, 28(102):9948–9953, 2005.

[27] H. Hasegawa. Synchronisations in small-world networks of spiking neurons : Diffusive versus sigmoid couplings. *Phys. Rev. E.*, 72:056139, 2005.

[28] Y. He, Z. Chen, and A. Evans. Small-world anatomical networks in the human brain revealed by cortical thickness from mri. *Cerebral Cortex*, 2007. Advance access published online January 4.

[29] D. Hebb. *The Organization of Behaviour*. John Wiley & Sons, New-York, 1948.

[30] A. Heynen, E. Quinlan, D. Bae, and M. Bear. Bidirectional, activity-dependent regulation of glutamate receptors in the adult hippocampus in vivo. *Neuron*, 28:527–536, 2000.
[31] M. Hirsch. Convergent activation dynamics in continuous time networks. *Neur. Networks*, 2:331–349, 1989.
[32] H. Hong, B. Kim, M. Choi, and H. Park. Factors that predict better synchronizability on complex networks. *Phys. Rev. E*, 65:067105, 2002.
[33] F. Hoppensteadt and E. Izhikevich. *Weakly Connected Neural Networks*. Springer Verlag, 1997.
[34] M. Kaiser and C. Hilgetag. Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Comput. Biol.*, 2:e95, 2006.
[35] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*. Kluwer, 1998.
[36] H. F. Kwok, P. Jurica, A. Raffone, and C. van Leeuwen. Robust emergence of small-world structure in networks of spiking neurons. *Cogn Neurodyn*, 1:39–51, 2007.
[37] L. F. Lago-Fernández, R. Huerta, F. Corbacho, and J. A. Sigüenza. Fast response and temporal coherent oscillations in small-world networks. *Phys. Rev. Lett.*, 84:2758–2761, 200.
[38] S. Micheloyannis, E. Pachou, C. Stam, M. Vourkas, S. Erimaki, and V. Tsirka. Using graph theoretical analysis of multi channel eeg to evaluate the neural efficiency hypothesis. *Neurosci. Lett*, 402:273–277, 2006.
[39] T. Nishikawa, A. E. Motter, Y. C. Lai, and F. C. Hoppensteadt. Heterogeneity in oscillator networks : are smaller worlds easier to synchronize ? *Phys. Rev. Lett.*, 91, 2003.
[40] R. Racine, N. Milgram, and S. Hafner. Long-term potentiation phenomena in the rat limbic forebrain. *Brain Res.*, 260:217–231, 1983.
[41] D. Ruelle. *Thermodynamic formalism*. Reading, Massachusetts: Addison-Wesley, 1978.
[42] D. Ruelle. Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics. *Journ. Stat. Phys.*, 95:393–468, 1999.
[43] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertész. Generalizations of the clustering coefficient to weighted complex networks. *Phys. Rev. E*, 75:027105, 2007.
[44] O. Shefi, I. Golding, R. Segev, E. Ben-Jacob, and A. Ayali. Morphological characterization of in vitro neuronal networks. *Phys. Rev. E*, 66:021905, 2002.
[45] C. W. Shin and S. Kim. Self-organized criticality and scale-free properties in emergent functional neural networks. *Phys. Rev. E*, 74:045101, 2006.
[46] Y. G. Sinai. Equilibrium states and the ergodic theory of anosov diffeomorphisms. *Lect. Notes.in Math.*, 27(4):21–69, 1972.
[47] B. Siri, H. Berry, B. Cessac, B. Delord, and M. Quoy. Topological and dynamical structures induced by hebbian learning in random neural networks. In *International Conference on Complex Systems*, Boston, june 2006.
[48] C. Skarda and W. Freeman. How brains make chaos in order to mahe sense of the world. *Behavioral and Brain Sciences*, 10:161–195, 1987.
[49] O. Sporns, D. chialvo, M. Kaiser, and C. Hilgetag. Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8(9):418–425, 2004.
[50] O. Sporns and J. Zwi. The small world of the cerebral cortex. *Neuroinformatics*, 2:145–162, 2004.
[51] C. Stam. Functional connectivity patterns of human magnetoencephalographic recordings: a 'small-world' network? *Neurosci. Lett*, 355:25–28, 2004.
[52] R. Thomas. *On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations*, chapter Numerical methods in the study of critical phenomena, pages 180–193. Springer-Verlag in Synergetics, 1981.
[53] D. Volchenkov and P. Blanchard. Random walks along the streets and canals in compact cities: Spectral analysis, dynamical modularity, information, and statistical mechanics. *Phys. Rev. E*, 75:026104, 2007.
[54] A. Volianskis and M. Jensen. Transient and sustained types of long-term potentiation in the ca1 area of the rat hippocampus. *J. Physiol.*, 550:459–492, 2003.
[55] D. Watts and S. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393:440–442, 1998.
[56] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. The structure of the nervous system of the nematode Caenorhabditis elegans. *Phil. Trans. R. Soc. of Lond. B*, 314:1–340, 1986.