

LEARNING-INDUCED TOPOLOGICAL EFFECTS ON DYNAMICS IN NEURAL NETWORKS

B. Siri
Alchemy
INRIA
Orsay, France
benoit.siri@inria.fr

H. Berry
Alchemy
INRIA
Orsay, France
hugues.berry@inria.fr
M. Quoy
ETIS,
CNRS UMR8051 & UCP-ENSEA
Cergy-Pontoise, France
quoy@ensea.fr

B. Cessac
Institut Non Linéaire de Nice
CNRS UMR6618 & UNSA
Nice, France
bruno.cessac@inln.cnrs.fr

B. Delord
ANIM
Inserm U742 & UPMC
Paris, France
bruno.delord@snv.jussieu.fr
O. Temam
Alchemy
INRIA
Orsay, France
olivier.temam@inria.fr

RESUME

Dans cet article nous relierons des processus dynamiques complexes, comme l'apprentissage, aux modifications structurelles qu'ils induisent dans le cadre des réseaux de neurones récurrents aléatoires (RRNNs). Dans ces réseaux, l'apprentissage modifie la dynamique progressivement, réduisant le chaos d'origine à un point-fixe attractif spécifique de la modalité apprise. En appliquant une règle simple d'apprentissage Hebbienne, nous montrons que la réduction de la dynamique s'accompagne de modifications des boucles locales basées sur les liens synaptiques. De plus, l'apprentissage Hebbien semble lier fortement les neurones actifs entre eux, tout en préservant de courts chemins de connection entre les neurones. Nous observons donc, en conséquence des modifications des boucles locales, une réorganisation des synapses les plus fortes en un réseau "petit-monde" (ou "small-world", SW). Ces résultats apportent des éclairages nouveaux sur les bases structurelles qui sous-tendent le traitement de l'information (i.e. reconnaissance de patterns) dans les RRNNs. De plus, ils ouvrent des perspectives sur les relations entre l'apprentissage Hebbien et l'architecture petit-monde.

KEY WORDS:

Neural networks, chaos, dynamics, learning, structure.

1 Introduction

Neural networks show amazing abilities for information storage and processing, and stimulus-dependent activity shaping. While the relationships between network structure and dynamics in non-neural networks have recently motivated a great amount of work (see [1] for review), they are especially meaningful in neural networks. Actually, learning in neural networks implies that activity guides the way synapses evolve; but the resulting connectivity structure in turn raises a new dynamical regime. This modification cycle becomes even more complex if the considered basic architecture is not feed-forward but includes recurrent synap-

tic links, like in cortical structures. We investigate here this kind of cross-interactions between dynamics and connectivity structure (topology) in chaotic random recurrent neural networks (RRNNs) during learning.

In the absence of learning, chaotic dynamics in RRNNs has been extensively studied using statistical physics, especially mean-field theory. As the non-linearity of the neuron responses or the variance of inputs increase the dynamics gets more and more complex until reaching chaos (via a quasi-periodicity route). During learning we observe that chaotic networks undergo a reverse behavior (dynamics reduction) [2]. However, as learning suppresses the statistical conditions that allow to use mean-field theory, the mechanisms responsive for this dynamics reduction cannot be studied directly with such tools. Yet, the understanding of this dynamics reduction could lead to new insights about several forms of information processing. For instance, the global dynamical behavior of RRNNs is highly similar to that exhibited by the olfactory bulb. Experimental works showed that the activity of this area is spontaneously chaotic. When a learned odor is presented, the dynamics changes and collapses onto a stable limit cycle, specific for this odor [3]. In this paper, we present analysis methods that allow us to study the dynamics-structure relationships in this framework.

2 The model

Let us consider N fully connected neurons (here $N = 500$). The activity of each neuron is denoted by $x_i \in [0, 1]$, for $i = 1..N$. These activities are updated using

$$x_i(t+1) = f \left(\sum_{j=1}^N w_{ij}(t)x_j(t) + I_i \right) \quad \forall i = 1 \dots N \quad (1)$$

In Eq. 1, $f(x)$ is a sigmoidal transfer function ($f(x) = (1 + \tanh(gx))/2$) (here $g = 10.0$), w_{ij} is the weight of the link from neuron j to neuron i , and I_i is a time-independent

input applied to neuron i . The weight matrix \mathbf{W} is initialized with a normal distribution : $\mathbf{W}(0) = \mathcal{N}(0, J^2/N)$, here with $J = 1$. Thus, weights can be positive, negative or null and are asymmetric ($w_{ij} \neq w_{ji}$). Finally, the initial activity vector $\mathbf{X}(0)$ is random with uniform distribution between 0 and 1.

In our model, learning is performed by a rule inspired by Hebb's postulate, which expresses the fact that synaptic links between two neurons are strengthened when they are co-active. The implementation used here is straightforward and translates this postulate into :

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \cdot x_i(t+1)x_j(t) \quad (2)$$

with $\alpha = 10^{-2}$. We add the constraint that a weight cannot change its sign. Note that with this implementation, weights cannot decrease and positive ones can grow unbounded. A neuron cannot be linked to itself ($w_{ii} = 0$).

The pattern we use in all our experiments consists in a binary cross pattern of $(22 \times 22 + 16)$ units which is directly mapped onto the network. We set either $I_i = 0.5$ for activated units or $I_i = 0.0$ for inactive ones. Once the pattern is applied, the network learns it by repeating a two-step learning phase : 1) the network dynamics evolves according to Eq. 1 during T transitory steps (no synaptic change occurs at this point), 2) then we apply the learning rule Eq. 2 for weight evolution. Here we used $T = 20$, although other positive integer values could fit.

3 Results

3.1 Behavior and function

RRNNs learn to associate an input to a specific dynamical regime. Let us picture a typical learning simulation. It consists of 3 stages. During the first stage, we inspect the behavior of the network with and without applied pattern (keeping weights at their initial values). The second stage is the learning phase, which is an iteration of 81 learning steps as described in 2. In the third stage we stop learning and inspect again the network with and without the learned pattern. A pattern is learned if the spontaneous dynamics settles on an attractor of lower complexity when the pattern is applied. Fig. 1 A and B show attractors for the first stage of this simulation. Without pattern (A), the dynamics settles on a strange attractor. Dynamics is thus chaotic. By applying a pattern (B), we observe a slight shift of the average activity but the general regime remains chaotic. Learning drives the dynamics toward the attractor shown in (C). This regime is quite simpler than the original dynamics and describes regularities in the network activity : we observe a limit cycle. When removing the pattern (D), the dynamics changes and settles on a different attractor. This clearly outlines the differences between pattern-influenced and no-pattern dynamical regimes. Therefore the consequence of learning resides in two phenomena : first, attractors after learning are simpler than the initial chaotic ones, second, the resulting attractor is specific to the learned pattern.

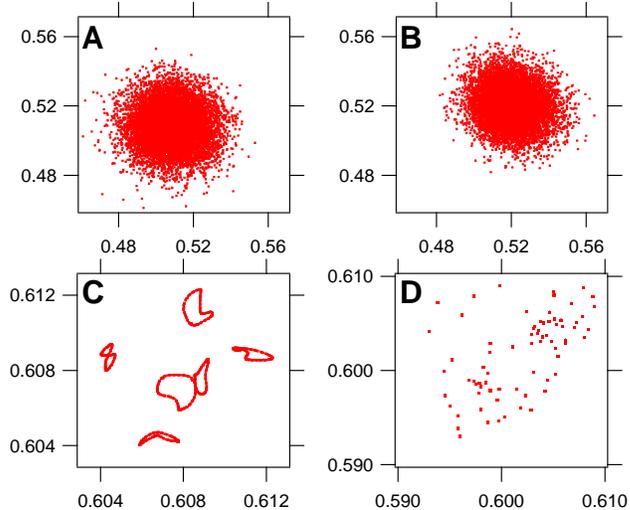


Figure 1. A typical simulation illustrating how the dynamics supports the network function. The panels present the attractors obtained by plotting $x(t+1)$ vs $x(t)$ at different stages of the simulation. 2000 steps are plotted for each figure. See text for details.

3.2 Dynamics reduction : structural factors

We now turn our attention to structural factors that are responsive for this dynamics reduction. To understand this effect we study the way local interaction loops change during learning. Here we call n -loop a non-null circuit in synaptic links, connecting n neurons such as $\{w_{21}, w_{32}, \dots, w_{n(n-1)}, w_{1n}\}$. The weight of a loop is $\pi_l = (\prod_{i=1}^{n-1} w_{(i+1)i}) \times w_{1n}$, $i \in \{1 \dots n\}$, i.e. the product of the links that compose it. We say that loop l is positive if $\pi_l > 0$ and negative if not. A great deal of work has been devoted to studies of the influence of the network composition, in terms of these local loops, on the network global behavior [4]. Note that a continuous time version of Eq. 1 considered on a network constituted by only one negative loop generically undergoes a Hopf bifurcation, giving rise to oscillations when the synaptic strength \mathbf{W} (or the gain g) increases. On the one hand, when several negative loops are competing together and when \mathbf{W} (or g) increases, we are first expecting synchronization phases leading to quasiperiodicity and frequency locking, and then chaos, by the Ruelle-Takens scenario [5]. This is exactly what is observed in the dynamical system Eq. 1. On the other hand, positive loops lead to so-called cooperative systems [6] which are convergent (they only have fixed-points and not more complex attractors). We are thus expecting that a decay in the negative loops naturally leads to a reduction of the dynamics complexity.

For computational reasons we will only consider here 2- and 3-loops. Looking at our rule Eq. 2, a straight-

3.3 Global organization during learning

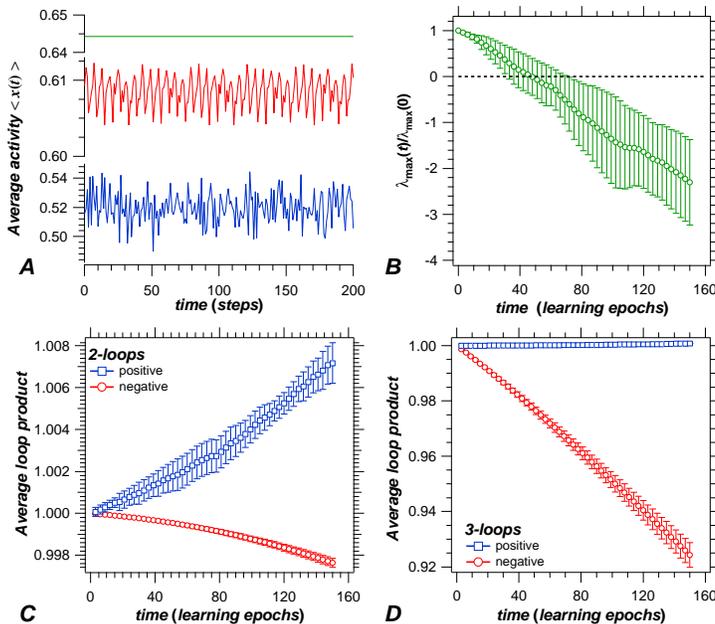


Figure 2. Dynamics and local loop content changes during learning with rule Eq. 2. In (B–D), the variables shown are normalized by their values before learning and are averages over 50 different initial conditions. Bars represent standard deviation. (A) Dynamics of the RRNNs before learning (bottom); as soon as $\lambda_{max} = 0$ (when the network leaves chaotic behavior, middle); or after 150 learning epochs (top). Here dynamics is $x(t) = \frac{1}{N} \sum_{i=1}^N x_i(t)$ for a single realization of the initial synaptic strengths. (B) Decrease of λ_{max} during learning (normalized by $\lambda_{max}(0)$). (C) Evolution of the average π_l for positive (squares) and negative (circles) 2-loops. (D) Evolution of the average loop product for positive (squares) and negative (circles) 3-loops (made of 3 negative links only)

forward *a priori* analysis could bring us some intuition on what is happening. Consider two neurons i and j , both active at time t ($x_i(t) \sim 1$ and $x_j(t) \sim 1$). According to the rule, the link binding them will be positively incremented, whatever its sign. Hence inhibitory synapses will vanish and excitatory synapses will strengthen. As learning goes on, inhibitory links will eventually vanish, so that negative loops made of only inhibitory links will vanish as well.

Fig. 2 shows the dynamics at different stages of learning, the evolution of the maximal Lyapunov exponent and the average value of π_l ($\bar{\pi}_l$) for 2- and 3-loops through time. Explicit decrease of $\bar{\pi}_l$ for negative loops and increase for positive loops are observed, while λ_{max} progressively shrinks toward negative values. Taken together, those results suggest that rule Eq. 2 brings dynamics stability by silencing negative loops. However, our simulations showed that negative 3-loops composed of only one negative link keep growing due to the rise of the two positive links (not shown).

Let’s have a look at the global network organization arising with learning. Learning changes the network architecture away from its original configuration. In order to capture the main features of this evolution, we inspect the synapse network through two indices used in complex network study : the clustering index $\langle C \rangle$ and the mean shortest path $\langle l \rangle$. $\langle l \rangle$ reports the average minimal number of links needed to connect any pair of nodes. Let d_{ij} be the shortest number of links to reach node i from j , then $\langle l \rangle$ is the average of d_{ij} over all pairs of nodes. $\langle C \rangle$ denotes the probability that if node i is linked to both j and k , then j and k are themselves interconnected. Formally, it is expressed as the average over all nodes of $C_i = 1/(N^2 - N) \sum_{i,j,k} a_{ij}a_{ik}a_{jk}$, where $a_{ij} = 1$ if a link exists between i and j (else 0).

Following Watts and Strogatz’s definition [7], a high $\langle C \rangle$ combined with a low $\langle l \rangle$ define a “small-world” (SW) network. This topology, widely found in nature (including many neural networks), lies somewhere between totally random networks and regular ones (e.g. lattices). Broadly speaking, SW networks combine good local communication properties (high $\langle C \rangle$, i.e. densely interconnected local neighborhoods, like in regular networks) with good global ones (short $\langle l \rangle$, like in totally random networks). The effects of this topology on the dynamics it underlays are in many cases dramatic (see [1] for review).

In our case, since dynamics is supported by the strongest synapses, we apply a threshold ϵ to extract the matrix $\mathbf{S}(t)$, where $s_{ij}(t) = \Theta(|w_{ij}(t)| - \epsilon)$, with $\Theta(\cdot)$ the Heavyside step function. Hence, $s_{ij}(t) = 1$ if the absolute value of the link from j to i is stronger than the threshold ϵ . \mathbf{S} thus gives us the sparse network of the strongest synapses. To get a good overview of the different weight layers in networks we thresholded \mathbf{W} with several ϵ . We can inspect this way the organization at long learning times. Fig. 3.(A) and (B) present evolution of normalized $\langle C \rangle$ and $\langle l \rangle$ respectively, after long-time learning. They show an increased proportion of local interconnection density compared to random networks, while the distance between two neurons remains as short as in the initial random network. In other words, networks show a high cluster index while their mean shortest path remains low. Thus by definition, the Hebbian rule used here organizes the networks as a SW one. This effect was already reported in [8]. Note that $\langle C \rangle$ rises as the threshold increases, suggesting the presence of a dense core of hard-linked neurons.

However, significant differences between the learning-induced structure and the initial one cannot be seen at early learning time, since learning gain per step is negligible compared to the initial weight distribution. To circumvent this problem we apply the same thresholding operation to the synaptic modification matrices $\delta\mathbf{W}(t)$, where $\delta w_{ij}(t) = w_{ij}(t) - w_{ij}(t - 1)$. We obtain thus $\delta\mathbf{S}(t)$, which allows us to observe global organization

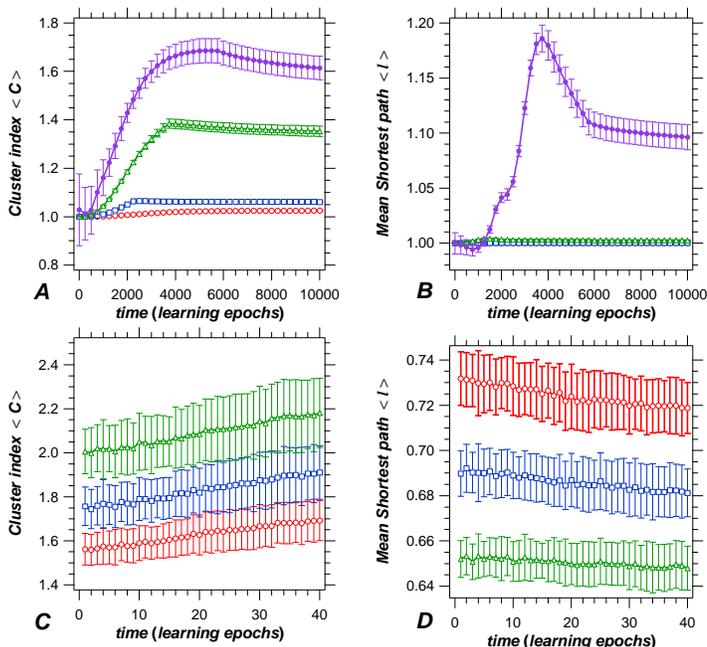


Figure 3. Network structure changes during learning with rule Eq. 2. All values are normalized by the value that would be obtained with a comparable random network (with the same N and number of synapses), averaged over 50 different initial conditions. Evolution of clustering coefficient (A) or mean shortest path (B) of the network of strong synapses at long learning times. Thresholds $\epsilon = 0.01$ (open circles), 0.05 (squares), 0.08 (triangles) or 0.12 (full circles). (C) and (D) present the same measurements, applied to the network of fast learning synapses ($\delta\mathbf{W}$) during reduction of the dynamics (see Fig. 2). Here, thresholds are $\epsilon = 10^{-9}$ (circles), 10^{-8} (squares) or 10^{-7} (triangles). Note that computation of the mean shortest path in (B & D) is obtained by averaging over neuron pairs that can be connected by a path. Bars are standard deviation.

at the same time scale as that of the reduction of the dynamics. Fig.3.(C) and (D) show the analysis at early learning time when inspecting $\delta\mathbf{W}$ s. As soon as learning begins, SW organization quickly appears. After sufficient learning iterations, this effect eventually becomes visible in the \mathbf{W} weight matrix. This phenomenon can be related to the local loop evolution described in 3.2. Indeed, our learning rule favors positive loops because of the increasing weight of excitatory links. In the case of the positive 3-loops composed of three excitatory links, this makes active neurons strongly assembled together, thus increasing $\langle C \rangle$.

4 Conclusion and future work

Our aim in the present paper was to study how the structure, dynamics and function are related in RRNNs evolving with a simple Hebbian learning. We show that this specific in-

terplay between structure and dynamics enables the system to perform pattern recognition tasks. We also show that the learning rule Eq. 2 modifies the local loop content of the network, increasing the weights of positive loops and decreasing those of negative ones. This is likely to explain the reduction of the network dynamics complexity from chaos to regular limit cycle oscillations and fixed-point stability. Finally, because the rule favors positive 3-loops, we show that the global structure of the network is progressively enriched with triplets of interconnected, strongly excitatory neurons. Because the mean shortest path length remains stable, strong synapses distribute on the RRNN structure as a SW network.

A useful property of the learning rule studied (Eq. 2) resides in its simplicity, which facilitates its analysis. Whether the conclusions formulated here can be generalized to other, more complex learning rules, is currently under study. One interesting rule known as the covariance rule, for instance, compares the unit activities to their average values over a period T ($x_i(t) - \bar{x}_i$). Most of the conclusions presented in this paper for rule Eq. 2 (including SW organization) are also valid with this rule. However, interpretation of the local loop content is much more difficult and will be the subject of future research. Conversely, other possible implementations behave in a very different way, so that the SW structure reported below is not always observed. Here again, understanding the relationships between dynamics and structure modification in this case will be the subject of our future work.

References

- [1] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., D.U., H., Complex networks: structure and dynamics, *Phys. Rep.*, 2006, 424, 175–308.
- [2] Daucé, E., Quoy, M., Cessac, B., Doyon, B., Samuelides, M., Self-organization and pattern-induced reduction of dynamics in recurrent networks, *Neural Networks*, 1998, 11, 521–533.
- [3] Skarda, C., Freeman, W., How brains make chaos in order to make sense of the world, *Behav. Brain Sci.*, 1987, 10, 161–195.
- [4] Gouzé, J., Positive and negative circuits in dynamical systems, *J. Biol. Syst.*, 1998, 6, 11–15.
- [5] Ruelle D., T. F., On the nature of turbulence, *Comm. Math. Phys.*, 1971, 20, 167–192.
- [6] Hirsch, M. W., Convergent activation dynamics in continuous time networks, *Neural Networks*, 1989, 2, 331–349.
- [7] Watts, D., Strogatz, S., Collective dynamics of “small-world” networks, *Nature*, 1998, 393, 440–442.
- [8] Berry, H., Quoy, M., Structure and dynamics of random recurrent neural networks, *Adaptive Behavior*, 2006, in press.