

Local learning rules and bifurcations in the global dynamics of random recurrent neural networks*

Benoit Siri¹, Hugues Berry¹, Bruno Cessac^{2,3}, Bruno Delord⁴, and Mathias Quoy⁵

¹ Team Alchemy, INRIA, Orsay, France hugues.berry@inria.fr

² INLN and Université de Nice, Nice, France

³ Team Odyssee, INRIA, Sophia-Antipolis, France

⁴ ANIM, Inserm U742 & UPMC, Paris, France,

⁵ ETIS, UMR 8051, Université de Cergy-Pontoise - ENSEA, Cergy-Pontoise France

Summary. In this article, we investigate the effects of biologically-plausible local learning rules (Hebbian rules) on the global dynamics of firing-rate random recurrent neural networks. In these models, learning collapses the initially chaotic dynamics onto simpler attractors (limit cycle or fixed point) that are specific of the learned pattern. This phenomenon provides these networks with associative memory properties but remains poorly understood. Focusing on the spectral properties of the Jacobian and weight matrices, we show here that learning brings global dynamics near a bifurcation point where network sensitivity to the input pattern is maximal. We thus contribute to a theoretical framework to uncover the relations between dynamics, structure and function in these networks.

1.1 Introduction

In recent years, a large amount of work concerning dynamical systems interacting on complex networks has focused on the influence of network topology on the global dynamics (for a review, see [4]). In particular, many studies have been devoted to the relationships between node synchronization and the classical statistical quantifiers of complex networks (degree distribution, average clustering index, mean shortest path...) [9, 15, 13]. The main idea was that the impact of network topology on the global dynamics might be prominent, so that these structural statistics may be good indicators of the global dynamics. This assumption proved however largely wrong and many of the related studies yielded contradictory results [15, 12]. Actually, synchronization

* This work was supported by a grant of the French National Research Agency, project JC05_63935 “ASTICO”

properties cannot be systematically deduced from topology statistics but may be inferred from the spectrum of the network [1]. Accordingly, many studies have considered diffusive coupling of the nodes [10]. In this case, the adjacency matrix has real nonnegative eigenvalues, and global properties, such as stability of the synchronized states [2], can easily be inferred from its spectral properties.

In this framework, neural networks with complex topology are particularly interesting because the dynamics of the neurons (the network nodes) depends on synaptic weights (the network links), that themselves vary over time as a function of the neuron dynamics. Understanding this mutual coupling between dynamics and topology and its effects on the computations made by the network is a key problem in computational neuroscience, that could benefit the multidisciplinary tools of complex systems approaches. Unfortunately, the coupling between neurons (synaptic weight) is rarely diffusive, so that the corresponding matrix is not symmetric and may contain positive and negative elements. The spectral characteristics of such systems and their relations to global dynamical properties remain to be understood.

In this paper, we wish to study the specific case of random recurrent neural networks (RRNNs). These network models display a rich variety of dynamical behaviors, including fixed points, limit cycle oscillations, quasiperiodicity and deterministic chaos [7], that are suspected to be similar to experimental behaviors observed in the olfactory bulb [17, 8]. It is known that the application of biologically-plausible local learning rules (Hebbian rules) reduces the dynamics of chaotic RRNNs to simpler attractors that are specific of the learned input pattern [6]. This phenomenon endows RRNNs with associative memory properties, but remains poorly understood. We propose here a theoretical and simulation study of the effects of these local rules on the global dynamics change during learning. To this aim, we mostly focus on the spectral properties of the weight and Jacobian matrices.

1.2 Model

We consider firing-rate recurrent neural networks with discrete time dynamics. Our model considers that learning dynamics may be slower than neuron dynamics. Synaptic weights are thus constant for τ consecutive dynamics steps, which defines a learning epoch. We denote by t the update index of neuron states (dynamics) inside a learning epoch while T indicates the update index of synaptic weights (learning). Let $x_i^{(T)}(t)$ be the mean firing rate of neuron i at dynamics time t of the learning epoch T . Its dynamics is given by

$$\begin{cases} x_i^{(T)}(t+1) = f\left(\sum_{j=1}^N W_{ij}^{(T)} x_j^{(T)}(t) + \theta_i + \xi_i\right) \\ x_i^{(T+1)}(0) = x_i^{(T)}(\tau) \end{cases} \quad (1.1)$$

where f is a sigmoidal function (e.g. $f(x) = 1/2(1 + \tanh(gx))$). The output gain g tunes the nonlinearity of the function. θ_i is the threshold of the neuron

response and ξ_i a (time constant) external input applied at neuron i . $W_{ij}^{(T)}$ represents the weight of the synapse from neuron j to neuron i at learning epoch T . The initial weights $W_{ij}^{(1)}$ are randomly sampled from a Gaussian law with 0 mean and variance $1/N$, where N is the number of neurons. Hence the synaptic weight matrix \mathcal{W} contains positive (excitation), negative (inhibition) or null (no synapse) elements and is asymmetric ($W_{ij}^{(T)} \neq W_{ji}^{(T)}$).

We consider here learning rules that conform to Hebb's postulate about learning in biological neurons [11]. Hence, we are interested in generic rules where learning occurs via local (i.e. depending on the pre- and post-synaptic neurons only) modifications of synaptic weights. We also assume that no synaptic change takes place if the postsynaptic neuron is silent, except for exponential decay which corresponds to passive forgetting.

Defining the activity of neuron i , $m_i^{(T)} = \bar{x}_i^{(T)} - d_i$, where $\bar{x}_i^{(T)}$ is the time-averaged firing rate of neuron i during learning epoch T , and d_i is the associated activity threshold (i.e. neuron i is said silent during learning epoch T whenever $m_i^{(T)} \leq 0$ and active else), the learning rule we studied here writes:

$$W_{ij}^{(T+1)} = \lambda W_{ij}^{(T)} + \frac{\alpha}{N} m_i^{(T)} m_j^{(T)} H(m_i^{(T)}) \quad (1.2)$$

where α is the learning rate, H the Heaviside function ($H(x) = 1$ if $x > 0$ and 0 otherwise) and λ is the rate of passive forgetting. In the following we set $d_i = |g/\theta_i|$. Finally, we assume that weights cannot change sign and are hardly bounded to 0.

1.3 Results

1.3.1 Dynamics reduction

Here, we set the parameters so that the spontaneous (initial) dynamics is chaotic. Previous simulation results have shown that Hebbian learning leads to a simplification of the dynamics from the initial strange attractor to a quasiperiodic one, and, as learning goes on, a limit cycle and finally a fixed point [6]. In fact, the application of our learning rule for as few as 5 learning epochs is enough for the largest Lyapunov exponent L_1 to decrease and become negative (not shown). We give here theoretical and experimental explanations of this dynamic reduction. For concision, we only present the main theoretical results. Thorough presentation and detailed proofs will be given elsewhere (Siri *et al.*, manuscript in preparation).

A bound on the largest Lyapunov exponent at learning epoch T , $L_1^{(T)}$, can easily be found. Let ρ_T be the natural invariant measure (Sinai-Ruelle-Bowen measure) at learning epoch T and $\langle \phi \rangle_T = \int \phi d\rho_T$ the average of a function ϕ with respect to ρ_T , then

$$L_1^{(T)} \leq \log(\|\mathcal{W}^{(T)}\|) + \left\langle \log(\max_i f'(u_i)) \right\rangle_T, \quad (1.3)$$

where $u_i = \sum_{j=1}^N W_{ij}^{(T)} x_j^{(T)}(t) + \theta_i + \xi_i$ is the local field of neuron i .

This shows that L_1 is expected to decrease whenever the spectrum radius of the weight matrix \mathcal{W} shrinks or $\log(\max_i f'(u_i))$ decreases. The latter term in fact becomes negative when all neurons are (on average) saturated. Note that this effect is related to amplification/saturation and can be addressed in the context of linear response theory [5]. It follows from this result that if learning increases the saturation level of neurons or shrinks \mathcal{W} spectrum radius, then the result will be a decay of $L_1^{(T)}$.

The effect of the parameter λ is to allow the system forget its initial synaptic structure. It gives the possibility to “rewire” entirely the network in a time scale proportional to $\frac{1}{|\log(\lambda)|}$. Indeed, it can be shown that

$$\mathcal{W}^{(T+1)} = \lambda^T \mathcal{W}^{(1)} + \frac{\alpha}{N} \sum_{n=1}^T \lambda^{T-n} \mathcal{G}^{(n)}. \quad (1.4)$$

where $\mathcal{W}^{(1)}$ is the initial synaptic matrix and $\mathcal{G}^{(n)}$ is the matrix whose elements are given by $G_{ij}^{(n)} = m_i^{(n)} m_j^{(n)} H(m_i^{(n)})$. If $\lambda = 1$, a direct corollary of this equation is that $\mathcal{W}^{(T)}$ will diverge. On the opposite, if $\lambda < 1$, the “direct” contribution of $\mathcal{W}^{(1)}$, to $\mathcal{W}^{(T+1)}$ decays exponentially fast. In this case, one thus expects to reach a stationary regime where synaptic weights do not evolve any more. The existence of a such a stationary distribution for the synaptic weights can be shown to be provided by the sufficient condition

$$\mathcal{W}^{(K)} = \frac{\alpha}{N(1-\lambda)} \Gamma \quad (1.5)$$

at some learning epoch K , with $\Gamma = \text{cste}$. Assuming this condition is satisfied and let $s_1^{(T)}$ be the eigenvalue of $\mathcal{W}^{(T)}$ with largest norm, we have

$$|s_1^{(T+1)}| \leq \lambda^T \|\mathcal{W}^{(1)}\| + \frac{\alpha}{N} \frac{1-\lambda^T}{1-\lambda} C \leq \lambda^T \|\mathcal{W}^{(1)}\| + \frac{\alpha}{N} \frac{1}{1-\lambda} C \quad (1.6)$$

where C is a constant.

Thus this equation predicts that the Hebbian learning rule should contract the spectral radius of the weight matrix (i.e. λ^T vanishes when $\lambda < 1$). We also note that, as $\frac{\alpha}{N}$ is small, the contraction should be dominated by $\lambda^T \|\mathcal{W}^{(1)}\|$. Figure 1.1 shows the time evolution of $s_1^{(T)}$ during simulations. This figure first confirms that learning comes up with a decrease of $|s_1^{(T)}|$ but also demonstrates that the initial dynamics is mainly dictated by λ^T . Both results are in excellent agreement with predictions of eq. 1.6.

Hence, these theoretical considerations show that dynamics reduction in this system is mainly due to the effect of forgetting, that reduces the spectral

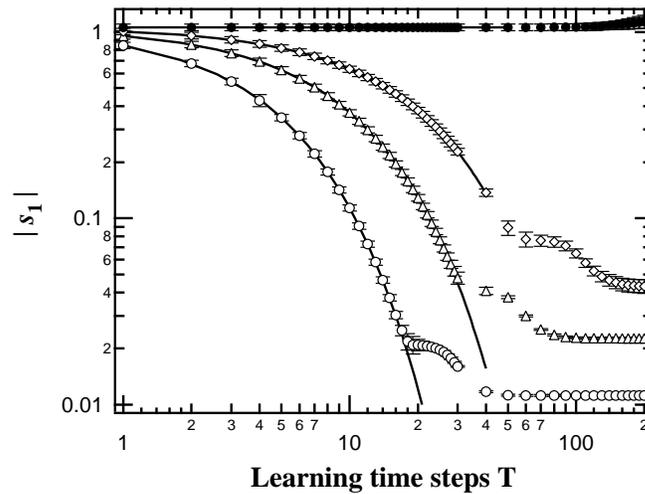


Fig. 1.1. Hebbian learning rule induces contraction of the weight matrix spectral radius. The evolution during learning of the norm of \mathcal{W} largest eigenvalue, $|s_1^{(T)}|$ is plotted on a log-log scale for $\lambda = 0.80$ (open circles), 0.90 (open triangles), 0.95 (open diamonds) or 1.00 (close circles). Each value is an average over 50 realizations with different initial conditions (initial weights and neuron states). Bars are standard deviations. Full lines are plots of exponential decreases with equation $g(T) = |s_1^{(1)}|\lambda^T$. Other parameters were: $N = 100$, $\alpha = 10^{-3}$, $g = 10$ and $\theta_i = 0.15$, $\xi_i = 0.010 \sin(2\pi i/N) \cos(8\pi i/N) \forall i = 1 \dots N$

radius of the synaptic weight matrix, which, in turn yields reduction of the largest Lyapunov exponent. Note however that, even without forgetting ($\lambda = 1$), dynamics reduction can be caused by any Hebbian rule that increases saturation of the neurons (eq. 1.3).

1.3.2 Learning and bifurcation

The jacobian matrix of the system \mathcal{DF}_x in our case reads

$$DF_{ij} = \frac{\partial f(u_i)}{\partial x_j} = f'(u_i)W_{ij} \quad (1.7)$$

Let $\mu_1^{(T)}$ be the eigenvalue of the Jacobian matrix with largest norm. It can be proven that

$$|\mu_1^{(T+1)}| \leq \max_i f'(u_i^{(T+1)}) \left[\lambda^T \|\mathcal{W}^{(1)}\| + \frac{\alpha}{N} \frac{1 - \lambda^T}{1 - \lambda} C \right] \quad (1.8)$$

One sees that the spectrum of $\mathcal{DF}_x^{(T)}$ is contracted by two effects: the contraction of the spectrum of \mathcal{W} and the decay of $\max_i f'(u_i^{(T+1)})$, once again related to saturation effects. Note that if $\lambda = 1$ (no forgetting) and $\mathcal{W}^{(T)}$

diverges then $\mathbf{u}^{(T)}$ diverges as well, leading $\max_i f'(u_i^{(T+1)})$ to converge toward 0. Hence, even in the absence of forgetting, saturation effects can lead to a decrease of $|\mu_1^{(T+1)}|$. Figure 1.2 shows the time evolution of $|\mu_1^{(T)}|$ during learning with $\lambda < 1$. This figure confirms that the highest eigenvalue indeed decreases during learning with a rate that depends on λ .

However, the main effect of learning is to make the network sensitive to the learned pattern. This implies that removal of the pattern should lead to a significant change in the networks dynamics. Defining \mathbf{x} as the dynamics obtained when the input pattern ξ is applied to the network:

$$\mathbf{x}^{(T)}(t+1) = f(\mathcal{W}\mathbf{x}^{(T)}(t) + \theta + \xi) \quad (1.9)$$

and \mathbf{x}' the dynamics obtained without pattern application:

$$\mathbf{x}'^{(T)}(t+1) = f(\mathcal{W}\mathbf{x}'^{(T)}(t) + \theta) \quad (1.10)$$

Sensitivity to the input pattern may be expressed as the amplitude of the network response obtained when ξ is removed. In other words, significant changes occur in the dynamics if $\delta\mathbf{x} = \mathbf{x}' - \mathbf{x}$ is large when ξ is removed. It is possible to show that

$$(\mathcal{I} - \mathcal{DF}_{\mathbf{x}}) \delta\mathbf{x} = -\Lambda(\mathbf{u})\xi \quad (1.11)$$

In this equation, \mathcal{I} is the unit matrix and $\Lambda(\mathbf{u})$ is the matrix defined by $\Lambda(\mathbf{u})_{ij} = f'(u_i)\delta_{ij}$ where $\delta_{ij} = 1$ if $i = j$, and 0 else. If 1 is not an eigenvalue of $\mathcal{DF}_{\mathbf{x}}$, then one may express $\delta\mathbf{x}$ as a function of ξ . Therefore, if ξ is small, $\delta\mathbf{x}$ will be small too. In this case, removing the pattern does not lead to a significant change in the dynamics. In other words, when the system is not close to a bifurcation $\delta\mathbf{x}$ remains proportional to ξ and is negligible when ξ is small. Similarly, if 1 is an eigenvalue of $\mathcal{DF}_{\mathbf{x}}$, but the projection of ξ in the corresponding direction vanishes, we are in the same case as before.

To be more rigorous, call $\Delta^{(T)}[\phi]$ the variation of the (time) average value of ϕ induced by pattern removal and λ_k and \mathbf{v}_k the eigenvalues and eigenvectors of $\mathcal{W}^{(T)}\Lambda(\mathbf{u}^{*(T)})$, ordered such that $|\lambda_N| \leq |\lambda_{N-1}| \leq |\lambda_1| < 1$. Then one can derive from eq.(1.11):

$$\Delta^{(T)}[\mathbf{u}] = -\sum_{k=1}^N \frac{(\mathbf{v}_k, \xi)}{1 - \lambda_k} \mathbf{v}_k \quad (1.12)$$

where $(,)$ denotes the inner product. It is clear from this equation that $\Delta[\mathbf{u}]$ diverges whenever $(\mathbf{v}_k, \xi) > 0$ (thus $\xi \neq 0$) and $\lambda_1 = 1$. It can be shown that $\lambda_k = \mu_k$ (i.e. the eigenvalues of $\mathcal{W}^{(T)}\Lambda(\mathbf{u}^{*(T)})$ are also the eigenvalues of $\mathcal{DF} = \Lambda(\mathbf{u}^{*(T)})\mathcal{W}^{(T)}$). Thus, eq.(1.12) actually predicts that if 1 is an eigenvalue of $\mathcal{DF}_{\mathbf{x}}$, and neither ξ nor its projection in the corresponding direction is 0, then $\Delta[\mathbf{u}]$ is not small anymore and depends on the nonlinear term $\Lambda(\mathbf{u})$. This case is illustrated in figure 1.2. In this figure, we measure the following quantity:

$$\Delta^{(T)}[A] = \frac{1}{N} \sqrt{\sum_{i=1}^N \left(\langle A_{ii}^{(T)} \rangle - \langle A'_{ii} \rangle \right)^2} \quad (1.13)$$

where

$$\langle A_{ii}^{(T)} \rangle = \lim_{\tau \rightarrow \infty} \sum_{t=1}^{\tau} A_{ii}^{(T)}(t) \quad (1.14)$$

and A' stands for A without pattern ($\xi_i(t) = 0 \forall i$). Hence $\Delta[A]$ measures the change in neuron excitability when the input pattern is withdrawn. A_{ii} is maximal when the local field of neuron i , u_i , falls in the central part of the sigmoid transfer function f (i.e. when $f'(u_i)$ is maximal), therefore when i is not saturated and is the most sensitive to fluctuation in its inputs. Note however that when $f'(u_i)$ is maximal A_{ii} is maximal, but it does not necessary entail that $\Delta[A]$ is maximal.

Hence, the effect of removing the pattern is predicted to be maximal when

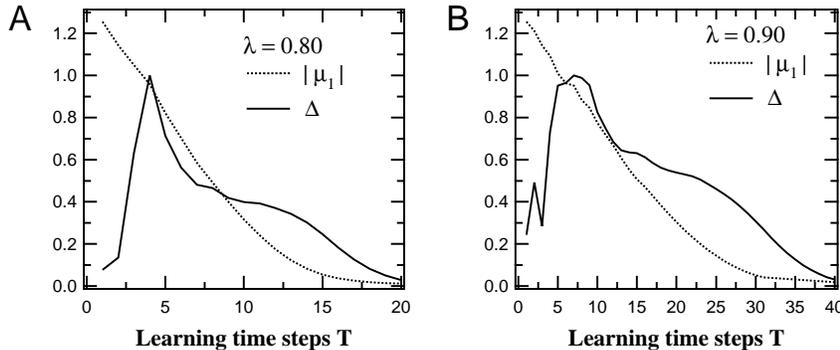


Fig. 1.2. The network sensitivity to the input pattern is maximal close to a bifurcation. The evolution during learning of the norm of $\mathcal{D}\mathcal{F}_x$ largest eigenvalue, $|\mu_1^{(T)}|$ (dotted line) is plotted together with the sensitivity measure Δ (full line) for $\lambda = 0.80$ (A) or 0.90 (B). The values of Δ 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. 1.1

$|\mu_1^{(T)}|$ (but not necessarily $s_1^{(T)}$) is close to 1. This effect is clearly visible from Fig 1.2, from where it can be seen that the maximal value of Δ is indeed obtained for learning epochs that yield $|\mu_1^{(T)}|$ values close to 1.

As we are dealing with discrete-time systems, this value of $|\mu_1^{(T)}| = 1$ corresponds to a bifurcation in the dynamics. Therefore, this proves that Hebbian learning in our case leads the system to a part of the parameter space where removing the input pattern makes the system cross a critical boundary. The way of tuning the parameters in order to promote this behavior is still under investigation.

1.4 Conclusion

In the present paper, we study the impact of Hebbian learning rules with passive forgetting and show that the observed dynamics reduction is caused either by forgetting-induced rewiring or by neuron saturation. These phenomena both shrink the spectral radius of the weight matrix, which, in turn decreases the largest Lyapunov exponent. Moreover, these very mechanisms contract the spectral radius of the Jacobian matrix, and we demonstrate that network sensitivity to the learned pattern is maximal when the norm of the leading eigenvalue of the Jacobian matrix is close to 1. In other words, learning drives the network through a bifurcation boundary and sensitivity to the pattern is maximal when pattern removal drives the system from one side of this bifurcation boundary to the other.

However, when learning goes on, the dynamics is driven away from this critical boundary, therefore pattern sensitivity eventually vanishes. This effect outlines the limits of such Hebbian learning rules in RRNNs. An interesting development would be to modify the rule with mechanisms granting that, once reached, the system remains in the “edge of chaos”. To find such mechanisms, we are currently looking at biology, and more specifically at homeostatic plasticity (see [18, 14] for reviews). Homeostatic plasticity has already been studied in recurrent neural networks and has been found to greatly enhance neural response [19]. Such a mechanism could provide our model with precise tuning of the synaptic matrix spectrum and neuron activity, thus may regulate learning and stabilize dynamics close to the edge of chaos.

On the other hand, our previous studies demonstrated that Hebbian learning also provokes the emergence of a small world structure for the corresponding adjacency network [16, 3]. Future works will focus on the relationships between these two effects.

References

1. F. Atay, T. Biyikoğlu, and J. Jost. Network synchronization : Spectral versus statistical properties. *Physica D*, 224:35–41, 2006.
2. M. Barahona and L. Pecora. Synchronization in small-world systems. *Phys. Rev. Lett.*, 89:054101, 2002.
3. H. Berry and M. Quoy. Structure and dynamics of random recurrent neural networks. *Adaptive Behavior*, 14:129–137, 2006.
4. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. Complex networks : Structure and dynamics. *Physics Reports*, (424):175–308, 2006.
5. B. Cessac and J. Sepulchre. Stable resonances and signal propagation in a chaotic network of coupled units. *Phys. Rev. E*, 70:056111, 2004.
6. 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.
7. 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.

8. W. Freeman. Simulation of chaotic eeg pattern with a dynamic model of the olfactory system. *Biol. Cyber.*, 56:139–150, 1987.
9. G. Grinstein and R. Linsker. Synchronous neural activity in scale-free network models versus random network models. *PNAS*, 28(102):9948–9953, 2005.
10. H. Hasegawa. Synchronisations in small-world networks of spiking neurons : Diffusive versus sigmoid couplings. *Phys. Rev. E.*, 72:056139, 2005.
11. D. Hebb. *The Organization of Behaviour*. John Wiley & Sons, New-York, 1948.
12. H. Hong, B. Kim, M. Choi, and H. Park. Factors that predict better synchronizability on complex networks. *Phys. Rev. E*, 65:067105, 2002.
13. 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, 2000.
14. E. Marder and J. Goaillard. Variability, compensation and homeostasis in neuron and network function. *Nature reviews Neurosciences*, 7:563–574, 2006.
15. 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.
16. 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.
17. C. Skarda and W. Freeman. How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10:161–195, 1987.
18. G. Turrigiano, K. Leslie, N. Desai, L. Rutherford, and S. Nelson. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391:892–896, 1998.
19. H. Williams and J. Noble. Homeostatic plasticity improves signal propagation in continuous-time recurrent neural networks. *BioSystems*, 87:252–259, 2007.