The Journal of Mathematical Neuroscience
a SpringerOpen Journal

**RESEARCH**                                                    Open Access

# Multiscale analysis of slow-fast neuronal learning models with noise

**Mathieu Galtier · Gilles Wainrib**

**Abstract** This paper deals with the application of temporal averaging methods to recurrent networks of noisy neurons undergoing a slow and unsupervised modification of their connectivity matrix called learning. Three time-scales arise for these models: (i) the fast neuronal dynamics, (ii) the intermediate external input to the system, and (iii) the slow learning mechanisms. Based on this time-scale separation, we apply an extension of the mathematical theory of stochastic averaging with periodic forcing in order to derive a reduced deterministic model for the connectivity dynamics. We focus on a class of models where the activity is linear to understand the specificity of several learning rules (Hebbian, trace or anti-symmetric learning). In a weakly connected regime, we study the equilibrium connectivity which gathers the entire 'knowledge' of the network about the inputs. We develop an asymptotic method to approximate this equilibrium. We show that the symmetric part of the connectivity post-learning encodes the correlation structure of the inputs, whereas the anti-symmetric part corresponds to the cross correlation between the inputs and their time derivative. Moreover, the time-scales ratio appears as an important parameter revealing temporal correlations.

M. Galtier (✉)
NeuroMathComp Project Team, INRIA/ENS Paris, 23 avenue d'Italie, Paris, 75013, France
e-mail: m.galtier@jacobs-university.de

M. Galtier
School of Engineering and Science, Jacobs University Bremen gGmbH, College Ring 1, P.O. Box 750 561, Bremen, 28725, Germany

G. Wainrib
Laboratoire Analyse Géométrie et Applications, Université Paris 13, 99 avenue Jean-Baptiste Clément, Villetaneuse, Paris, France
e-mail: wainrib@math.univ-paris13.fr

## 1 Introduction

Complex systems are made of a large number of interacting elements leading to non-trivial behaviors. They arise in various areas of research such as biology, social sciences, physics or communication networks. In particular in neuroscience, the nervous system is composed of billions of interconnected neurons interacting with their environment. Two specific features of this class of complex systems are that (i) external inputs and (ii) internal sources of random fluctuations influence their dynamics. Their theoretical understanding is a great challenge and involves high-dimensional non-linear mathematical models integrating non-autonomous and stochastic perturbations.

Modeling these systems gives rise to many different scales both in space and in time. In particular, learning processes in the brain involve three time-scales: from neuronal activity (fast), external stimulation (intermediate) to synaptic plasticity (slow). Here, fast time-scale corresponds to a few milliseconds and slow time-scale to minutes/hour, and intermediate time-scale generally ranges between fast and slow scales, although some stimuli may be faster than neuronal activity time-scale (*e.g.*, submilliseconds auditory signals [1]). The separation of these time-scales is an important and useful property in their study. Indeed, multiscale methods appear particularly relevant to handle and simplify such complex systems.

First, stochastic averaging principle [2, 3] is a powerful tool to analyze the impact of noise on slow-fast dynamical systems. This method relies on approximating the fast dynamics by its quasi-stationary measure and averaging the slow evolution with respect to this measure. In the asymptotic regime of perfect time-scale separation, this leads to a slow reduced system whose analysis enables a better understanding of the original stochastic model.

Second, periodic averaging theory [4], which has been originally developed for celestial mechanics, is particularly relevant to study the effect of fast deterministic and periodic perturbations (external input) on dynamical systems. This method also leads to a reduced model where the external perturbation is time-averaged.

It seems appropriate to gather these two methods to address our case of a noisy and input-driven slow-fast dynamical system. This combined approach provides a novel way to understand the interactions between the three time-scales relevant in our models. More precisely, we will consider the following class of multiscale stochastic differential equations (SDEs), with $\epsilon_1, \epsilon_2 > 0$ two small parameters

$$\begin{cases} d\mathbf{v}^\epsilon = \frac{1}{\epsilon_1}[F(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon, \mathbf{u}(\frac{t}{\epsilon_2}))]\, dt + \frac{1}{\sqrt{\epsilon_1}} \boldsymbol{\Sigma}\, dB(t), \\ d\mathbf{w}^\epsilon = G(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon)\, dt, \end{cases} \tag{1}$$

where $\mathbf{v}^\epsilon \in \mathbb{R}^p$ represents the fast activity of the individual elements, $\mathbf{w}^\epsilon \in \mathbb{R}^q$ represents the connectivity weights that vary slowly due to plasticity, and $\mathbf{u}(t) \in \mathbb{R}^p$ rep-

resents the value of the external input at time $t$. Random perturbations are included in the form of a diffusion term, and $(B(t))$ is a standard Brownian motion.

We are interested in the double limit $\epsilon_1 \to 0$ and $\epsilon_2 \to 0$ to describe the evolution of the slow variable $\mathbf{w}$ in the asymptotic regime where both the variable $\mathbf{v}$ and the external input are much faster than $\mathbf{w}$. This asymptotic regime corresponds to the study of a neuronal network in which both the external input $\mathbf{u}$ and the neuronal activity $\mathbf{v}$ operate on a faster time-scale than the slow plasticity-driven evolution of synaptic weights $\mathbf{W}$. To account for the possible difference of time-scales between $\mathbf{v}$ and the input, we introduce the time-scale ratio $\mu = \epsilon_1/\epsilon_2 \in [0, \infty]$. In the interesting case where $\mu \in (0, \infty)$, one needs to understand the long-time behavior of the rescaled periodically forced SDE for any $\mathbf{w}_0$ fixed

$$d\mathbf{v} = F(\mathbf{v}, \mathbf{w}_0, \mu t)\, dt + \mathbf{\Sigma}(\mathbf{v}, \mathbf{w}_0)\, dB(t).$$

Recently, in an important contribution [5], a precise understanding of the long-time behavior of such processes has been obtained using methods from partial differential equations. In particular, conditions ensuring the existence of a periodic family of probability measures to which the law of $\mathbf{v}$ converges as time grows have been identified, together with a sharp estimation of the speed of mixing. These results are at the heart of the extension of the classical stochastic averaging principle [2] to the case of periodically forced slow-fast SDEs [6]. As a result, we obtain a reduced equation describing the slow evolution of variable $\mathbf{w}$ in the form of an ordinary differential equation,

$$\frac{d\mathbf{w}}{dt} = \bar{G}(\mathbf{w}),$$

where $\bar{G}$ is constructed as an average of $G$ with respect to a specific probability measure, as explained in Section 2.

This paper first introduces the appropriate mathematical framework and then focuses on applying these multiscale methods to learning neural networks.

The individual elements of these networks are neurons or populations of neurons. A common assumption at the basis of mathematical neuroscience [7] is to model their behavior by a stochastic differential equation which is made of four different contributions: (i) an intrinsic dynamics term, (ii) a communication term, (iii) a term for the external input, and (iv) a stochastic term for the intrinsic variability. Assuming that their activity is represented by the fast variable $\mathbf{v} \in \mathbb{R}^n$, the first equation of system (1) is a generic representation of a neural network (function $F$ corresponds to the first three terms contributing to the dynamics). In the literature, the level of non-linearity of the function $F$ ranges from a linear (or almost-linear) system to spiking neuron dynamics [8], yet the structure of the system is universal.

These neurons are interconnected through a connectivity matrix which represents the strength of the synapses connecting the real neurons together. The slow modification of the connectivity between the neurons is commonly thought to be the essence of learning. Unsupervised learning rules update the connectivity exclusively based on the value of the activity variable. Therefore, this mechanism is represented by the slow equation above, where $\mathbf{w} \in \mathbb{R}^{n \times n}$ is the connectivity matrix and $G$ is the learning rule. Probably the most famous of these rules is the Hebbian learning rule

introduced in [9]. It says that if both neurons A and B are active at the same time, then the synapses from A to B and B to A should be strengthened proportionally to the product of the activity of A and B. There are many different variations of this correlation-based principle which can be found in [10, 11]. Another recent, unsupervised, biologically motivated learning rule is the spike-timing-dependent plasticity (STDP) reviewed in [12]. It is similar to Hebbian learning except that it focuses on causation instead of correlation and that it occurs on a faster time-scale. Both of these types of rule correspond to $G$ being quadratic in $\mathbf{v}$.

Previous literature about dynamic learning networks is thick, yet we take a significantly different approach to understand the problem. An historical focus was the understanding of feedforward deterministic networks [13–15]. Another approach consisted in precomputing the connectivity of a recurrent network according to the principles underlying the Hebbian rule [16]. Actually, most of current research in the field is focused on STDP and is based on the precise times of the spikes, making them explicit in computations [17–20]. Our approach is different from the others regarding at least one of the following points: (i) we consider recurrent networks, (ii) we study the evolution of the coupled system activity/connectivity, and (iii) we consider bounded dynamical systems for the activity without asking them to be spiking. Besides, our approach is a rigorous mathematical analysis in a field where most results rely heavily on heuristic arguments and numerical simulations. To our knowledge, this is the first time such models expressed in a slow-fast SDE formalism are analyzed using temporal averaging principles.

The purpose of this application is to understand what the network learns from the exposition to time-dependent inputs. In other words, we are interested in the evolution of the connectivity variable, which evolves on a slow time-scale, under the influence of the external input and some noise added on the fast variable. More precisely, we intend to explicitly compute the equilibrium connectivities of such systems. This final matrix corresponds to the knowledge the network has extracted from the inputs. Although the derivation of the results is mathematically tough for untrained readers, we have tried to extract widely understandable conclusions from our mathematical results and we believe this paper brings novel elements to the debate about the role and mechanisms of learning in large scale networks.

Although the averaging method is a generic principle, we have made significant assumptions to keep the analysis of the averaged system mathematically tractable. In particular, we will assume that the activity evolves according to a linear stochastic differential equation. This is not very realistic when modeling individual neurons, but it seems more reasonable to model populations of neurons; see Chapter 11 of [7].

The paper is organized as follows. Section 2 is devoted to introducing the temporal averaging theory. Theorem 2.2 is the main result of this section. It provides the technical tool to tackle learning neural networks. Section 3 corresponds to application of the mathematical tools developed in the previous section onto the models of learning neural networks. A generic model is described and three different particular models of increasing complexity are analyzed. First, Hebbian learning, then trace-learning, and finally STDP learning are analyzed for linear activities. Finally, Section 4 is a discussion of the consequences of the previous results from the viewpoint of their biological interpretation.

## 2 Averaging principles: theory

In this section, we present multiscale theoretical results concerning stochastic averaging of periodically forced SDEs (Section 2.3). These results combine ideas from singular perturbations, classical periodic averaging and stochastic averaging principles. Therefore, we recall briefly, in Sections 2.1 and 2.2, several basic features of these principles, providing several examples that are closely related to the application developed in Section 3.

### 2.1 Periodic averaging principle

We present here an example of a slow-fast ordinary differential equation perturbed by a fast external periodic input. We have chosen this example since it readily illustrates many ideas that will be developed in the following sections. In particular, this example shows how the ratio between the time-scale separation of the system and the time-scale of the input appears as a new crucial parameter.

*Example 2.1* Consider the following linear time-inhomogeneous dynamical system with $\epsilon_1, \epsilon_2 > 0$ two parameters:

$$\frac{dv^\epsilon}{dt} = \frac{1}{\epsilon_1}\left(-v^\epsilon + \sin\left(\frac{t}{\epsilon_2}\right)\right),$$

$$\frac{dw^\epsilon}{dt} = -w^\epsilon + v^{\epsilon 2}.$$

This system is particularly handy since one can solve analytically the first ordinary differential equation, that is,

$$v(t) = \frac{1}{1+\mu^2}\left(\sin\left(\frac{t}{\epsilon_2}\right) - \mu\cos\left(\frac{t}{\epsilon_2}\right)\right) + v_0 e^{-\frac{t}{\epsilon_1}},$$

where we have introduced the *time-scales ratio*

$$\mu := \frac{\epsilon_1}{\epsilon_2}.$$

In this system, one can distinguish various asymptotic regimes when $\epsilon_1$ and $\epsilon_2$ are small according to the asymptotic value of $\mu$:

- Regime 1: Slow input $\mu = 0$:
    First, if $\epsilon_1 \to 0$ and $\epsilon_2$ is fixed, then $v(t)$ is close to $\sin(\frac{t}{\epsilon_2})$, and from *geometric singular perturbation theory* [21, 22] one can approximate the slow variable $w^\epsilon$ by the solution of

$$\frac{dw}{dt} = -w + \left(\sin\left(\frac{t}{\epsilon_2}\right)\right)^2.$$

    Now taking the limit $\epsilon_2 \to 0$ and applying the classical *averaging principle* [4] for periodically driven differential equations, one can approximate $w^\epsilon$ by the solution

of

$$\frac{dw}{dt} = -w + \frac{1}{2},$$

since $\frac{1}{2\pi} \int_0^{2\pi} \sin(s)^2 \, ds = \frac{1}{2}$.

- Regime 2: Fast input $\mu = \infty$:

  If $\epsilon_2 \to 0$ and $\epsilon_1$ is fixed, then the classical averaging principle implies that $v^\epsilon$ is close to the solution of

$$\frac{dv}{dt} = -\frac{v}{\epsilon_1},$$

  so that $w^\epsilon$ can be approximated by

$$\frac{dw}{dt} = -w + \left(v_0 e^{-t/\epsilon_1}\right)^2,$$

  and when $\epsilon_1 \to 0$, one does not recover the same asymptotic behavior as in Regime 1.

- Regime 3: Time-scales matching $0 < \mu < \infty$:

  Now consider the intermediate case where $\epsilon_1$ is asymptotically proportional to $\epsilon_2$. In this case, $v^\epsilon$ can be approximated on the fast time-scale $t/\epsilon_1$ by the periodic solution $\bar{v}_\mu(t) = \frac{1}{1+\mu^2}(\sin(\mu t) - \mu \cos(\mu t))$ of $\frac{dv}{dt} = -v + \sin(\mu t)$. As a consequence, $w^\epsilon$ will be close to the solution of

$$\frac{dw}{dt} = -w + \frac{1}{2(1 + \mu^2)},$$

since $\frac{1}{2\pi} \int_0^{2\pi} \bar{v}_\mu(t/\mu)^2 \, dt = \frac{1}{2(1+\mu^2)}$.

Thus, we have seen in this example that

1. the two limits $\epsilon_1 \to 0$ and $\epsilon_2 \to 0$ do not commute,
2. the ratio $\mu$ between the internal time-scale separation $\epsilon_1$ and the input time-scale $\epsilon_2$ is a key parameter in the study of slow-fast systems subject to a time-dependent perturbation.

## 2.2  Stochastic averaging principle

Time-scales separation is a key property to investigate the dynamical behavior of non-linear multiscale systems, with techniques ranging from averaging principles to geometric singular perturbation theory. This property appears to be also crucial to understanding the impact of noise. Instead of carrying a small noise analysis, a multiscale approach based on the *stochastic averaging principle* [2] can be a powerful tool to unravel subtle interplays between noise properties and non-linearities. More precisely, consider a system of SDEs in $\mathbb{R}^{p+q}$:

$$d\mathbf{v}_t^\epsilon = \frac{1}{\epsilon} F\left(\mathbf{v}_t^\epsilon, \mathbf{w}_t^\epsilon\right) dt + \frac{1}{\sqrt{\epsilon}} \mathbf{\Sigma}\left(\mathbf{v}_t^\epsilon, \mathbf{w}_t^\epsilon\right) \cdot dB(t),$$

$$d\mathbf{w}_t^\epsilon = G\left(\mathbf{v}_t^\epsilon, \mathbf{w}_t^\epsilon\right) dt,$$

with initial conditions $\mathbf{v}^\epsilon(0) = \mathbf{v}_0$, $\mathbf{w}^\epsilon(0) = \mathbf{w}_0$, and where $\mathbf{w}^\epsilon \in \mathbb{R}^q$ is called the slow variable, $\mathbf{v}^\epsilon \in \mathbb{R}^p$ is the fast variable, with $F$, $G$, $\mathbf{\Sigma}$ smooth functions ensuring the existence and uniqueness for the solution $(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon)$, and $B(t)$ a $p$-dimensional standard Brownian motion, defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Time-scale separation in encoded in the small parameter $\epsilon$, which denotes in this section a single positive real number.

In order to approximate the behavior of $(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon)$ for small $\epsilon$, the idea is to average out the equation for the slow variable with respect to the stationary distribution of the fast one. More precisely, one first assumes that for each $\mathbf{w} \in \mathbb{R}^q$ fixed, the *frozen* fast SDE,

$$d\mathbf{v}_t = F(\mathbf{v}_t, \mathbf{w})\, dt + \mathbf{\Sigma}(\mathbf{v}_t, \mathbf{w}) \cdot dB(t),$$

admits a unique invariant measure, denoted $\rho^\mathbf{w}(d\mathbf{v})$. Then, one defines the averaged drift vector field $\bar{G}$

$$\bar{G}(\mathbf{w}) := \int_{\mathbb{R}^m} G(\mathbf{v}, \mathbf{w}) \rho^\mathbf{w}(d\mathbf{v}) \tag{2}$$

and $\mathbf{w}$ the solution of $\frac{d\mathbf{w}}{dt} = \bar{G}(\mathbf{w})$ with the initial condition $\mathbf{w}(0) = y_0$. Under some dissipativity assumptions, the stochastic averaging principle [2] states:

**Theorem 2.1** *For any $\delta > 0$ and $T > 0$,*

$$\lim_{\epsilon \to 0} \mathbb{P}\left[ \sup_{t \in [0,T]} \left\| \mathbf{w}_t^\epsilon - \mathbf{w}_t \right\|^2 > \delta \right] = 0. \tag{3}$$

As a consequence, analyzing the behavior of the deterministic solution $\mathbf{w}$ can help to understand useful features of the stochastic process $(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon)$.

*Example 2.2* In this example we consider a similar system as in Example 2.1, but with a noise term instead of the periodic perturbation. Namely, we consider $(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon)$ the solution of the system of SDEs,

$$dv^\epsilon = -\frac{1}{\epsilon} v^\epsilon\, dt + \frac{\sigma}{\sqrt{\epsilon}}\, dB(t),$$

$$dw^\epsilon = \left( -w^\epsilon + \left( v^\epsilon \right)^2 \right) dt,$$

with $\epsilon > 0$ a small parameter and $\sigma > 0$ a positive constant. From Theorem 2.1, the stochastic slow variable $w^\epsilon$ can be approximated in the sense of (3) by the deterministic solution $w$ of

$$\frac{dw}{dt} = \int_{v \in \mathbb{R}} \left( -w + v^2 \right) \rho(dv),$$

where $\rho(dv)$ is the stationary measure of the linear diffusion process,

$$dv = -v\, dt + \sigma\, dB(t),$$

that is,

$$\rho(dv) = \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{v^2}{\sigma^2}}.$$

Consequently, $w^\epsilon$ can be approximated in the limit $\epsilon \to 0$ by the solution of

$$\frac{dw}{dt} = -w + \frac{\sigma^2}{2}.$$

Applying (3) leads to the following result: for any $T > 0$ and $\delta > 0$,

$$\lim_{\epsilon \to 0} \mathbb{P}\left[ \sup_{t\in[0,T]} \left| w_t^\epsilon - \left(y_0 - \frac{\sigma^2}{2}\right)e^{-t} + \frac{\sigma^2}{2}\right|^2 > \delta\right] = 0.$$

Interestingly, the asymptotic behavior of $w^\epsilon$ for small $\epsilon$ is characterized by a deterministic trajectory that depends on the strength $\sigma$ of the noise applied to the system. Thus, the stochastic averaging principle appears particularly interesting when unraveling the impact of noise strength on slow-fast systems.

Many other results have been developed since, extending the set-up to the case where the slow variable has a diffusion component or to infinite-dimensional settings for instance, and also refining the convergence study, providing *homogenization* results concerning the limit of $\epsilon^{-1/2}(w^\epsilon - w)$ or establishing large deviation principles (see [23] for a recent monograph). However, fewer results are available in the case of non-homogeneous SDEs, that is, when the system is perturbed by an external time-dependent signal. This setting is of particular interest in the framework of stochastic learning models, and we present the main relevant mathematical results in the following section.

### 2.3 Double averaging principle

Combining ideas of periodic and stochastic averaging introduced previously, we present here theoretical results concerning multiscale SDEs driven by an external time-periodic input. Consider $(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon)$ the solution of

$$d\mathbf{v}^\epsilon = \frac{1}{\epsilon_1}\left[ F\left(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon, \frac{t}{\epsilon_2}\right)\right] dt + \frac{1}{\sqrt{\epsilon_1}}\mathbf{\Sigma}\left(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon\right) \cdot dB(t),$$

$$d\mathbf{w}^\epsilon = G\left(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon\right) dt,$$

(4)

with $t \to F(\mathbf{v}, \mathbf{w}, t) \in \mathbb{R}^p$ a $\tau$-periodic function and $\epsilon = (\epsilon_1, \epsilon_2) \in \mathbb{R}_+^2$. Parameter $\epsilon_1$ represents the internal time-scale separation and $\epsilon_2$ the input time-scale. We consider the case where both $\epsilon_1$ and $\epsilon_2$ are small, that is, a strong time-scale separation between the fast variable $\mathbf{v}^\epsilon \in \mathbb{R}^p$ and the slow one $\mathbf{w}^\epsilon \in \mathbb{R}^q$, and a fast periodic modulation of the fast drift $F(\mathbf{v}, \mathbf{w}, \cdot)$.

We further denote $\mathbf{z} = (\mathbf{v}, \mathbf{w})$.

**Definition 2.1** We define the asymptotic time-scale ratio

$$\mu := \lim_{|\epsilon| \to 0} \frac{\epsilon_1}{\epsilon_2}. \tag{5}$$

Accordingly, we denote $\lim_{|\epsilon| \to 0}^{\mu}$ the distinguished limit when $\epsilon_1 \to 0$, $\epsilon_2 \to 0$ with $\epsilon_1/\epsilon_2 \to \mu$.

The following assumption is made to ensure existence and uniqueness of a strong solution to system (4). In the following, $\langle \mathbf{z}_1, \mathbf{z}_2 \rangle$ will denote the usual scalar product for vectors.

**Assumption 2.1** Existence and uniqueness of a strong solution

(i) The functions $F$, $G$, and $\boldsymbol{\Sigma}$ are locally Lipschitz continuous in the space variable $\mathbf{z}$. More precisely, for any $R > 0$, there exists a constant $\alpha_R$ such that

$$\left\| F(\mathbf{z}) - F(\mathbf{z}') \right\| \le \alpha_R \|\mathbf{z} - \mathbf{z}'\| \quad \text{for any } \mathbf{z}, \mathbf{z}' \in \mathbb{R}^{p+q} \text{ with } \|\mathbf{z}\| \le R \text{ and } \|\mathbf{z}'\| \le R.$$

(ii) There exists a constant $R > 0$ such that

$$\sup_{\|\mathbf{z}\| > R, t > 0} \frac{\langle (F(\mathbf{z}, t), G(\mathbf{z})), \mathbf{z} \rangle}{\|\mathbf{z}\|^2} < 0.$$

To control the asymptotic behavior of the fast variable, one further assumes the following.

**Assumption 2.2** Asymptotic behavior of the fast process:

(i) The diffusion matrix $\boldsymbol{\Sigma}$ is bounded

$$\exists M_{\boldsymbol{\Sigma}} > 0 \quad \text{s.t. } \forall \mathbf{z}, \left\| \boldsymbol{\Sigma}(\mathbf{z}) \right\| < M_{\boldsymbol{\Sigma}}$$

and uniformly non-degenerate

$$\exists \eta_0 > 0 \quad \text{s.t. } \forall \mathbf{v}, \mathbf{z} \big\langle \boldsymbol{\Sigma}(\mathbf{z}) \cdot \boldsymbol{\Sigma}(\mathbf{z})' \mathbf{v}, \mathbf{v} \big\rangle \ge \eta_0 \|\mathbf{v}\|^2.$$

(ii) There exists $r_0 < 0$ such that for all $t \ge 0$ and for all $\mathbf{z}, \mathbf{x} \in \mathbb{R}^{p+q}$,

$$\big\langle \nabla_{\mathbf{z}} F(\mathbf{z}, t) \cdot \mathbf{x}, \mathbf{x} \big\rangle \le r_0 \|\mathbf{x}\|^2.$$

According to the value of $\mu \in \{0, \mathbb{R}_+^*, \infty\}$, the stochastic averaging principle is based on a description of the asymptotic behavior of various rescaled fast frozen processes. More precisely, under Assumptions 2.1 and 2.2, one can deduce that:

- For any fixed $\mathbf{w}_0 \in \mathbb{R}^q$ and $t_0 > 0$ fixed, the law of the rescaled time-homogeneous frozen process,

$$d\mathbf{v} = F(\mathbf{v}, \mathbf{w}_0, t_0) \, dt + \boldsymbol{\Sigma}(\mathbf{v}, \mathbf{w}_0) \, dB(t),$$

converges exponentially fast to a unique invariant probability measure denoted by $\rho^{\mathbf{w}_0, t_0}(d\mathbf{v})$.

- For any fixed $\mathbf{w}_0 \in \mathbb{R}^q$, there exists a $\frac{\tau}{\mu}$-periodic evolution system of measures $\nu_\mu^{\mathbf{w}_0}(t, d\mathbf{v})$, different from $\rho^{\mathbf{w}_0,t}(d\mathbf{v})$ above, such that the law of the rescaled time-inhomogeneous frozen process,

$$d\mathbf{v} = F(\mathbf{v}, \mathbf{w}_0, \mu t)\, dt + \mathbf{\Sigma}(\mathbf{v}, \mathbf{w}_0)\, dB(t), \tag{6}$$

  converges exponentially fast towards $\nu_\mu^{\mathbf{w}_0}(t, \cdot)$, uniformly with respect to $\mathbf{w}_0$ (*cf.* the Appendix Theorem A.1).
- For any fixed $\mathbf{w}_0 \in \mathbb{R}^q$, the law of the rescaled time-homogeneous frozen process,

$$d\mathbf{v} = \bar{F}(\mathbf{v}, \mathbf{w}_0)\, dt + \mathbf{\Sigma}(\mathbf{v}, \mathbf{w}_0)\, dB(t),$$

  where $\bar{F}(\mathbf{v}, \mathbf{w}_0) := \tau^{-1} \int_0^\tau F(\mathbf{v}, \mathbf{w}_0, t)\, dt$, converges exponentially fast towards a unique invariant probability measure denoted by $\bar{\rho}^{\mathbf{w}_0}(d\mathbf{v})$.

According to the value of $\mu$, we introduce a vector field $\bar{G}_\mu$ which will play a role similar to $\bar{G}$ introduced in equation (2).

**Definition 2.2** We define $\bar{G}_\mu : \mathbb{R}^q \to \mathbb{R}^q$ as follows. In the time-scale matching case, that is, when $0 < \mu < \infty$, then

$$\bar{G}_\mu(\mathbf{w}) := \left(\frac{\tau}{\mu}\right)^{-1} \int_0^{\frac{\tau}{\mu}} \int_{\mathbf{v} \in \mathbf{R}^p} G(\mathbf{v}, \mathbf{w}) \nu_\mu^{\mathbf{w}}(t, d\mathbf{v})\, dt. \tag{7}$$

*Notation* We may denote the periodic system of measures $\nu_\mu^{\mathbf{w}}(t, d\mathbf{v})$ associated with (6) by $\nu_\mu^{\mathbf{w}}[F, \mathbf{\Sigma}](t, d\mathbf{v})$ to emphasize its relationship with $F$ and $\mathbf{\Sigma}$. Accordingly, we may denote $\bar{G}_\mu(\mathbf{w})$ by $\bar{G}_\mu^{[F, \mathbf{\Sigma}]}(\mathbf{w})$.

We are now able to present our main mathematical result. Extending Theorem 2.1, the following theorem describes the asymptotic behavior of the slow variable $\mathbf{w}^\epsilon$ when $\epsilon \to 0$ with $\epsilon_1/\epsilon_2 \to \mu$. We refer to [6] for more details about the full mathematical proof of this result.

**Theorem 2.2** *Let* $\mu \in (0, \infty)$. *If* $\mathbf{w}$ *is the solution of*

$$\frac{d\mathbf{w}}{dt} = \bar{G}_\mu(\mathbf{w}) \quad \text{with } \mathbf{w}(0) = \mathbf{w}^\epsilon(0), \tag{8}$$

*then the following convergence result holds, for all $T > 0$ and $\delta > 0$:*

$$\lim_{|\epsilon| \to 0}^{\mu} \mathbb{P}\left[\sup_{t \in [0,T]} \left|\mathbf{w}_t^\epsilon - \mathbf{w}_t\right|^2 > \delta\right] = 0.$$

*Remark 2.1*

1. The extremal cases $\mu = 0$ and $\mu = \infty$ are not covered in full rigor by Theorem 2.2. However, the study of the sequential limits $\epsilon_1 \to 0$ followed by $\epsilon_2 \to 0$ or $\epsilon_2 \to 0$ followed by $\epsilon_1 \to 0$ can be deduced from an appropriate combination of classical periodic and stochastic averaging theorems:

- Slow input: If one considers the case where the limit $\epsilon_1 \to 0$ is taken *first*, so that from Theorem 2.1 with fast variable $\mathbf{v}^\epsilon$ and slow variables $\mathbf{w}^\epsilon$ and $t$ (with the trivial equation $\dot{t} = 1$), $\mathbf{w}^\epsilon$ is close in probability on finite time-intervals to the solution of the following inhomogeneous ordinary differential equation:

$$\frac{d\tilde{\mathbf{w}}}{dt} = \int_{\mathbf{v} \in \mathbb{R}^p} G(\mathbf{v}, \tilde{\mathbf{w}}) \rho^{\tilde{\mathbf{w}}, t/\epsilon_2}(d\mathbf{v}) := \tilde{G}(\tilde{\mathbf{w}}, t/\epsilon_2).$$

Then taking the limit $\epsilon_2 \to 0$, one can apply the deterministic averaging principle to the fast periodic vector field $\tilde{G}(\mathbf{w}, t/\epsilon_2)$, so that $\tilde{\mathbf{w}}$ converges when $\epsilon_2 \to 0$ to the solution of

$$\frac{d\mathbf{w}}{dt} = \tau^{-1} \int_0^\tau \tilde{G}(\mathbf{v}, \mathbf{w}) \, dt = \bar{G}_0(\mathbf{w}),$$

where

$$\bar{G}_0(\mathbf{w}) := \tau^{-1} \int_0^\tau \int_{\mathbf{v} \in \mathbf{R}^p} G(\mathbf{v}, \mathbf{w}) \rho^{\mathbf{w}, t}(d\mathbf{v}) \, dt.$$

- Fast input: If the limit $\epsilon_2 \to 0$ is taken first, one first has to perform a classical averaging of the periodic drift $F(\mathbf{v}, \mathbf{w}, t/\epsilon_2)$ leading to the homogeneous system of SDEs (4), but with $\bar{F}(\mathbf{v}, \mathbf{w})$ instead of $F(\mathbf{v}, \mathbf{w}, t/\epsilon_2)$. Then, an application of Theorem 2.1 on this system gives an averaged vector field

$$\bar{G}_\infty(\mathbf{w}) := \int_{\mathbf{v} \in \mathbf{R}^p} G(\mathbf{v}, \mathbf{w}) \bar{\rho}^{\mathbf{w}}(d\mathbf{v}).$$

2. To study the extremal cases $\mu = 0$ and $\mu = \infty$ in full generality, one would need to consider all the possible relationships between $\epsilon_1$ and $\epsilon_2$, not only the linear one as in the present article, but also of the type $\epsilon_1 = \epsilon_2^\alpha$ for example. In this case, we believe that the regime $\alpha < 1$ converges to the same limit as taking the limit $\epsilon_2$ first and the regime $\alpha > 1$ corresponds to taking the limit $\epsilon_1$ first. The intermediate regime $\alpha = 1$ seems to be the only one for which the limit cannot be obtained by combining classical averaging principles. Therefore, the present article is focused on this case, in which the averaged system depends explicitly on the scaling parameter $\mu$. Moreover, in terms of applications, this parameter can have a relatively easy interpretation in terms of the ratio of time-scales between intrinsic neuronal activity and typical stimulus time-scales in a given situation. Although the zeroth order limit (*i.e.*, the averaged system) seems to depend only on the position of $\alpha$ with respect to 1, it seems reasonable to expect that the fluctuations around the limit would depend on the precise value of $\alpha$. This is a difficult question which may deserve further analysis.

   The case $0 < \mu < \infty$ is already very rich in the sense that it combines simultaneously both the periodic and stochastic averaging principles in a new way that cannot be recovered by sequential applications of those principles. A particular role is played by the frozen periodically-forced SDE (6). The equivalent of the quasi-stationary measure $\rho^{\mathbf{w}}$ of Theorem 2.1 is given by the asymptotically periodic behavior of equation (6), represented by the periodic family of measures $\nu_\mu^{\mathbf{w}}(t, d\mathbf{v})$.

3. By a rescaling of the frozen process (6), one deduces the following *scaling* relationships:

$$\nu_\mu^{\mathbf{w}}[F, \mathbf{\Sigma}](t, d\mathbf{v}) = \nu_1^{\mathbf{w}}\left[\frac{F}{\mu}, \frac{\mathbf{\Sigma}}{\sqrt{\mu}}\right](\mu t, d\mathbf{v})$$

and

$$\bar{G}_\mu^{[F,\mathbf{\Sigma}]}(\mathbf{w}) = \bar{G}_1^{[\frac{F}{\mu}, \frac{\mathbf{\Sigma}}{\sqrt{\mu}}]}(\mathbf{w}).$$

Therefore, if one knows, in the case $\mu = 1$, the averaged vector field associated with the fast process generated by a drift $F$ and a diffusion coefficient $\sigma$, denoted $\bar{G}_1[F, \mathbf{\Sigma}]$, it is possible to deduce $\bar{G}_\mu$ in the general case $\mu \in (0, \infty)$ with a change $F \to \mu F$ and $\mathbf{\Sigma} \to \sqrt{\mu}\mathbf{\Sigma}$.

4. It seems reasonable to expect that the above result is still valid when considering ergodic, but not necessarily periodic, time dependency of the function $F(\mathbf{v}, \mathbf{w}, \cdot)$. In equation (7), instead of integrating $\nu_\mu^{\mathbf{w}}(t, d\mathbf{v})$ over one period, one should integrate it with respect to an ergodic stationary measure. However, this extension requires non-trivial technical improvements of [5] which are beyond the scope of this paper.

### 2.3.1 Case of a fast linear SDE with periodic input

We present here an elementary case where one can compute explicitly the quasi-stationary time-periodic family of measures $\nu_\mu^{\mathbf{w}}(t, x)$, when the equation for the fast variable is linear. Namely, we consider $\mathbf{v} \in \mathbb{R}^p$ the solution of

$$d\mathbf{v}(t) = \left(-\mathbf{A} \cdot \mathbf{v}(t) + \mathbf{u}(\mu t)\right) dt + \mathbf{\Sigma} \cdot dB(t),$$

with initial condition $\mathbf{v}(0) = \mathbf{v}_0 \in \mathbb{R}^p$, and where $\mathbf{A} \in \mathbb{R}^{p \times p}$ is a matrix whose eigenvalues have positive real parts and $\mathbf{u}(\cdot)$ is a $\tau$-periodic function.

We are interested in the large time behavior of the law of $\mathbf{v}(t)$, which is a time-inhomogeneous Ornstein-Uhlenbeck process. From [5] we know that its law converges to a $\tau$-periodic family of probability measures $\nu(t, d\mathbf{v})$. Due to the linearity in the previous equation, $\nu(t, d\mathbf{v})$ is Gaussian with a time-dependent mean and a constant covariance matrix

$$\nu(t, d\mathbf{v}) = \mathcal{N}_{\bar{\mathbf{v}}(t), \mathbf{Q}}(d\mathbf{v}),$$

where $\bar{\mathbf{v}}$ is the $\frac{\tau}{\mu}$-periodic attractor of $\frac{d\bar{\mathbf{v}}}{dt} = -\mathbf{A} \cdot \bar{\mathbf{v}}(t) + \mathbf{u}(\mu t)$, *i.e.*,

$$\bar{\mathbf{v}}(t) = \int_{-\infty}^t e^{-A(t-s)} \mathbf{u}(\mu s) \, ds,$$

and $\mathbf{Q}$ is the unique solution of the Lyapunov equation

$$\mathbf{A} \cdot \mathbf{Q} + \mathbf{Q} \cdot \mathbf{A}' + \mathbf{\Sigma} \cdot \mathbf{\Sigma}' = 0. \tag{9}$$

Indeed, if one denotes $\mathbf{c}(t) = \mathbf{v}(t) - \bar{\mathbf{v}}(t)$, then $\mathbf{c}(t)$ is a solution of the classical homogeneous Ornstein-Uhlenbeck equation

$$d\mathbf{c}(t) = -\mathbf{A}\mathbf{c}(t)\,dt + \mathbf{\Sigma}\,dB(t),$$

whose stationary distribution is known to be a centered Gaussian measure with the covariance matrix $\mathbf{Q}$ solution of (9); see Chapter 3.2 of [24]. Notice that if $\mathbf{A}$ is self-adjoint with respect to $(\mathbf{\Sigma} \cdot \mathbf{\Sigma}')^{-1}$ (*i.e.*, $\mathbf{A} \cdot (\mathbf{\Sigma} \cdot \mathbf{\Sigma}') = (\mathbf{\Sigma} \cdot \mathbf{\Sigma}') \cdot \mathbf{A}'$), then the solution is $\mathbf{Q} = \frac{\mathbf{A}^{-1} \cdot (\mathbf{\Sigma} \cdot \mathbf{\Sigma}')}{2} = \frac{(\mathbf{\Sigma} \cdot \mathbf{\Sigma}') \cdot \mathbf{A}'^{-1}}{2}$, which will be used in Appendix B.2.

Hence, in the linear case, the averaged vector field of equation (7) becomes

$$\bar{G}_\mu(\mathbf{w}) := \left(\frac{\tau}{\mu}\right)^{-1} \int_0^{\frac{\tau}{\mu}} \int_{\mathbf{v} \in \mathbb{R}^p} G\big(\bar{\mathbf{v}}(t) + \mathbf{v}, \mathbf{w}\big)\mathcal{N}_{0,\mathbf{Q}}(d\mathbf{v})\,dt, \qquad (10)$$

where $\mathcal{N}_{\mathbf{x},\mathbf{Q}}$ is the probability density function of the Gaussian law with mean $\mathbf{x} \in \mathbb{R}^q$ and covariance $\mathbf{Q} \in \mathbb{R}^{p \times p}$.

Therefore, due to the linearity of the fast SDE, the periodic system of measure $\nu$ is just a constant Gaussian distribution shifted by a periodic function of time $\mathbf{v}(t)$. In case $G$ is quadratic in $\mathbf{v}$, this remark implies that one can perform independently the integral over time and over $\mathbb{R}^p$ in formula (10) (noting that the crossed term has a zero average). In this case, contributions from the periodic input and from noise appear in the averaged vector field in an additive way.

*Example 2.3* In this last example, we consider a combination between Example 2.1 and Example 2.2, namely we consider the following system of periodically forced SDEs:

$$dv^\epsilon = \frac{1}{\epsilon_1}\left[-v^\epsilon + \sin\left(\frac{t}{\epsilon_2}\right)\right]dt + \frac{\sigma}{\sqrt{\epsilon_1}}\,dB(t),$$

$$dw^\epsilon = \big(-w^\epsilon + (v^\epsilon)^2\big)\,dt.$$

As in Example 2.1 and as shown above, the behavior of this system when both $\epsilon_1$ and $\epsilon_2$ are small depends on the parameter $\mu$ defined in (5). More precisely, we have the following three regimes:

- Regime 1: slow input:

$$\bar{G}_0(w) = -w + \frac{\sigma^2}{2} + \frac{1}{2}.$$

- Regime 2: fast input:

$$\bar{G}_\infty(w) = -w + \frac{\sigma^2}{2}.$$

- Regime 3: time-scale matching:

$$\bar{G}_\mu(w) = -w + \frac{\sigma^2}{2} + \frac{1}{2(1+\mu^2)}.$$

### 2.4 Truncation and asymptotic well-posedness

In some cases, Assumptions 2.1-2.2 may not be satisfied on the entire phase space $\mathbb{R}^p \times \mathbb{R}^q$, but only on a subset. Such situations will appear in Section 3 when considering learning models. We introduce here a more refined set of assumptions ensuring that Theorem 2.2 still applies.

Let us start with an example, namely the following bi-dimensional system with white noise input:

$$\begin{cases} dv^\epsilon = \frac{1}{\epsilon}(-lv^\epsilon + w^\epsilon v^\epsilon)\, dt + \frac{\sigma}{\sqrt{\epsilon}}\, dB(t), \\ dw^\epsilon = (-\kappa w^\epsilon + (v^\epsilon)^2)\, dt, \end{cases} \tag{11}$$

with $\epsilon > 0, \sigma > 0, l > 0, \mu > 0$.

For the fast drift $-(l - w)v$ to be non-explosive, it is necessary to have $w < l - \alpha$ with $\alpha > 0$ for all time. The concern about this system comes from the fact that the slow variable $w$ may reach $l$ due to the fluctuations captured in the term $v^2$, for instance, if $\kappa$ is not large enough. Such a system may have exponentially growing trajectories. However, we claim that for small enough $\epsilon$, $w^\epsilon$ will remain close to its averaged limit $w$ for a very long time, and if this limit remains below $l - \alpha$, then $w^\epsilon$ can be considered as well-posed in the asymptotic limit $\epsilon \to 0$. To make this argument more rigorous, we suggest the following definition.

**Definition 2.3** A stochastic differential equation with a given initial condition is asymptotically well posed in probability if for the given initial condition,

1. a unique solution exists until a random time $\tau_\epsilon$
2. for all $T > 0$,

$$\lim_{\epsilon \to 0} \mathbb{P}[\tau_\epsilon \geq T] = 1.$$

We give in the following proposition sufficient conditions for system (4) to be asymptotically well posed in probability and to satisfy conclusions of Theorem 2.2.

Let us introduce the following set of additional assumptions.

**Assumption 2.3** Moment conditions:
  (i) There exists $p > 2$ such that

$$\text{for any } T > 0, \quad \sup_\epsilon \mathbb{E}\left[ \sup_{0 \leq t \leq T} \left\| \mathbf{v}_t^\epsilon \right\|^p + \left\| \mathbf{w}_t^\epsilon \right\|^p \right] < \infty.$$

  (ii) For any $T > 0$ and any bounded subset $K$ of $\mathbb{R}^q$,

$$\sup_{\epsilon_1 > 0, \epsilon_2 > 0, \mathbf{w} \in K} \mathbb{E}\left[ \sup_{0 \leq t \leq T} \left\| G(\mathbf{v}_t^\epsilon, \mathbf{w}) \right\|^2 \right] < \infty.$$

*Remark 2.2* This last set of assumptions will be satisfied in all the applications of Section 3 since we consider linear models with additive noise for the equation of **v**, implying this variable to be Gaussian and the function $G$ only involves quadratic

moments of **v**; therefore, the moment conditions (i) and (ii) will be satisfied without any difficulty. Moreover, if one considers non-linear models for the variable **v**, then the Gaussian property may be lost; however, adding sigmoidal non-linearity has, in general, the effect of bounding the dynamics, thus making these moment assumptions reasonable to check in most models of interest.

**Property 2.3** *If there exists a subset $\mathcal{E}$ of $\mathbb{R}^q$ such that*

1. *The functions $F$, $G$, $\Sigma$ satisfy Assumptions 2.1-2.3 restricted on $\mathbb{R}^p \times \mathcal{E}$.*
2. *$\mathcal{E}$ is invariant under the flow of $\bar{G}_\mu$, as defined in (7).*

*Then for any initial condition $\mathbf{w}_0 \in \mathcal{E}$, system (4) is asymptotically well posed in probability and $\mathbf{w}^\epsilon$ satisfies the conclusion of Theorem 2.2.*

*Proof* See Appendix A.2.                                                                          □

Here, we show that it applies to system (11). First, with $\mathcal{E}_\alpha = \{w \in \mathbb{R}, w < l - \alpha\}$, for some $\alpha \in \,]0, l[$, it is possible to show that Assumptions 2.1-2.2 are satisfied on $\mathbb{R}^p \times \mathcal{E}_\alpha$. Then, as a special case of (10), we obtain the following averaged system:

$$\frac{dw}{dt} = -\kappa w + \frac{\sigma^2}{2(l - w)} := \bar{G}(w).$$

It remains to check that the solution of this system satisfies

$$\exists \alpha > 0, \quad \text{such that } w(0) < l - \alpha \quad \Rightarrow \quad \forall t > 0, \quad w(t) < l - \alpha,$$

that is, the subset $\mathcal{E}_\alpha$ is invariant under the flow of $\bar{G}$.

This property is satisfied as soon as

$$\eta := \frac{2\sigma^2}{\kappa l^2} < 1.$$

Indeed, one can show that $\bar{G}(w) = 0$ admits two solutions iff $\eta < 1$,

$$w_\pm = \frac{l}{2}(1 \pm \sqrt{1 - \eta}) \in (0, l),$$

and that $w_-$ is stable whereas $w_+$ is unstable. Thus, if $w(0) < l - \alpha$ with $\alpha = l - w_+ > 0$, then $w(t) < l - \alpha$ for all $t > 0$. In fact, the invariance property is true for all $\alpha \in \,]l - w_-, l - w_+[$.

# 3 Averaging learning neural networks

In this section, we apply the temporal averaging methods derived in Section 2 on models of unsupervised learning neural networks. First, we design a generic learning model and show that one can define formally an averaged system with equation (7). However, going beyond the mere definition of the averaged system seems very

difficult and we only manage to get explicit results for simple systems where the fast activity dynamics is linear. In the three last subsections, we push the analysis for three examples of increasing complexity.

In the following, we always consider that the initial connectivity is 0. This is an arbitrary choice but without consequences, because we focus on the regime where there is a single globally stable equilibrium point (see Section 3.2.3).

## 3.1 A generic learning neural network

We now introduce a large class of stochastic neuronal networks with learning models. They are defined as coupled systems describing the simultaneous evolution of the activity of $n \in \mathbb{N}$ neurons and the connectivity between them. We define $\mathbf{v} \in \mathbb{R}^n$, the *activity field* of the network, and $\mathbf{W} \in \mathbb{R}^{n \times n}$, the *connectivity matrix*.

Each neuron variable $\mathbf{v}_i$ is assumed to follow the SDE

$$d\mathbf{v}_i = \left( f_i(\mathbf{v}_i) + \mathbf{u}_i \right) dt + \mathbf{\Sigma} \cdot dB_i(t),$$

where the function $f_i$ characterizes the intrinsic non-linear dynamical behavior of neuron $i$ and $\mathbf{u}_i$ is the input received by neuron $i$. The stochastic term $\mathbf{\Sigma} \cdot dB_i(t)$ is added to account for internal sources of noise. In terms of notations, $(B(t))_{t \geq 0}$ is a standard $n$-dimensional Brownian motion, $\mathbf{\Sigma}$ is an $n \times n$ matrix, possibly function of $\mathbf{v}$ or other variables, and $\mathbf{\Sigma} \cdot dB_i(t)$ denotes the $i$th component of the vector $\mathbf{\Sigma} \cdot dB(t)$.

The input $\mathbf{u}_i$ to neuron $i$ has mainly two components: the external input $\mathbf{u}_i^{\text{ext}}$ and the input coming from other neurons in the network $\mathbf{u}_i^{\text{syn}}$. The latter is *a priori* a complex combination of post-synaptic potentials coming from many other neurons. The coefficient $\mathbf{W}_{ij}$ of the connectivity matrix accounts for the strength of a synapse $j \to i$. Note that neurons can be connected to themselves, *i.e.*, $\mathbf{W}_{ii}$ is not necessarily null. Thus, we can write

$$\mathbf{u}_i^{\text{syn}} := \mathcal{S}\left( \sum_{j=1}^n \mathbf{W}_{ij} \mathcal{H}(\mathbf{v}_i, \mathbf{v}_j) \right),$$

where $\mathcal{S} : \mathbb{R} \to \mathbb{R}$ and $\mathcal{H}$ is a function taking the history of $\mathbf{v}_i$ and $\mathbf{v}_j$ and returning a real for each time $t$ (to take convolutions into account). In practical cases, they are often taken to be sigmoidal functions. We abusively redefine $\mathcal{S}$ and $\mathcal{H}$ as vector valued operators corresponding to the element-wise application of their real counterparts. We also define the function $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^n$ such that $\mathcal{F}(\mathbf{v})_i = f_i(\mathbf{v}_i)$. Together with a slow generic learning rule, this leads to defining a *stochastic learning model* as the following system of SDEs.

**Definition 3.1**

$$\begin{cases} d\mathbf{v}^\epsilon = \frac{1}{\epsilon}[\mathcal{F}(\mathbf{v}^\epsilon) + \mathcal{S}(\mathbf{W}^\epsilon \cdot \mathcal{H}(\mathbf{v}^\epsilon)) + \mathbf{u}^{\text{ext}}(t)] dt + \frac{1}{\sqrt{\epsilon}} \Sigma(\mathbf{v}^\epsilon, \mathbf{W}^\epsilon) \cdot dB(t), \\ d\mathbf{W}^\epsilon = G(\mathbf{W}^\epsilon, \mathbf{v}^\epsilon) dt. \end{cases}$$

Before applying the general theory of Section 2, let us make several comments about this generic model of neural network with learning. This model is a non-autonomous, stochastic, non-linear slow-fast system.

In order to apply Theorem 2.2, one needs Assumptions 2.1, 2.2, and 2.3 to be satisfied, restricting the space of possible functions $\mathcal{S}$, $\mathcal{H}$, $\mathbf{F}$, $\mathbf{\Sigma}$, and $G$. In particular, Assumption 2.2(ii) seems rather restrictive since it excludes systems with multiple equilibria and suggests that the general theory is more suited to deal with rate-based networks. However, one should keep in mind that these assumptions are only sufficient, and that the double averaging principle may work as well in systems which do not satisfy readily those assumptions.

As we will show in Section 3.3, a particular form of history-dependence can be taken into account, to a certain extent. Indeed, for instance, if the function $\mathcal{F}$ is actually a functional of the past trajectory of variable $\mathbf{v}^\epsilon$ which can be expressed as the solution of an additional SDE, then it may be possible to include a certain form of history-dependence. However, purely time-delayed systems do not enter the scope of this theory, although it might be possible to derive an analogous averaging method in this framework.

The noise term can be purely additive or set by a particular function $\mathbf{\Sigma}(\mathbf{v}, \mathbf{W})$ as long as it satisfies Assumption 2.2(i), meaning that it must be uniformly non-degenerate.

In the following subsection, we apply the averaging theory to various combinations of neuronal network models, embodied by choices of functions $\mathcal{S}$, $\mathcal{H}$, $\mathbf{F}$, $\mathbf{\Sigma}$, and various learning rules, embodied by a choice of the function $G$. We will also analyze the obtained averaged system, describing the slow dynamics of the connectivity matrix in the limit of perfect time-scale separation and, in particular, study the convergence of this averaged system to an equilibrium point.

## 3.2 Symmetric Hebbian learning

One of the simplest, yet non-trivial, stochastic learning models is obtained when considering

- A linear model for neuronal activity, namely $f_i(\mathbf{v}_i) = -l\mathbf{v}_i$ with $l$ a positive constant.
- A linear model for the synaptic transmission, namely $\mathcal{S}(\mathbf{v}_i) = \mathbf{v}_i$ and $\mathcal{H}(\mathbf{v}_i, \mathbf{v}_j) = \mathbf{v}_j$.
- A constant diffusion matrix $\mathbf{\Sigma}$ (additive noise) proportional to the identity $\mathbf{\Sigma} = \sigma Id$ (spatially uncorrelated noise).
- A Hebbian learning rule with linear decay, namely $G_{ij}(\mathbf{W}, \mathbf{v}) = -\kappa \mathbf{W}_{ij} + \mathbf{v}_i \mathbf{v}_j$. Actually, it corresponds to the tensor product: $\{\mathbf{v} \otimes \mathbf{v}\}_{ij} = \mathbf{v}_i \mathbf{v}_j$.

This model can be written as follows:

$$\begin{cases} d\mathbf{v}^\epsilon = \frac{1}{\epsilon_1}(-\mathbf{L} \cdot \mathbf{v}^\epsilon + \mathbf{W}^\epsilon \cdot \mathbf{v}^\epsilon + \mathbf{u}(\frac{t}{\epsilon_2}))\,dt + \frac{\sigma}{\sqrt{\epsilon_1}}\,dB(t), \\ \frac{d\mathbf{W}^\epsilon}{dt} = G(\mathbf{v}^\epsilon, \mathbf{W}^\epsilon) = -\kappa \mathbf{W}^\epsilon + \mathbf{v}^\epsilon \otimes \mathbf{v}^\epsilon, \end{cases} \quad (12)$$

where neurons are assumed to have the same decay constant: $\mathbf{L} = l I_d$; $\mathbf{u}$ is a periodic continuous input (it replaces $\mathbf{u}^{\text{ext}}$ in the previous section); $\sigma, \epsilon_1, \epsilon_2, \kappa \in \mathbb{R}_+$ with $\epsilon_1, \epsilon_2 \ll 1$ and $B(t)$ is $n$-dimensional Brownian noise.

The first question that arises is about the well-posedness of the system: What is the definition interval of the solutions of system (12)? Do they explode in finite time? At first sight, it seems there may be a runaway of the solution if the largest real part among the eigenvalues of $\mathbf{W}$ grows bigger than $l$. In fact, it turns out this scenario can be avoided if the following assumption linking the parameters of the system is satisfied.

**Assumption 3.1** There exists $p \in \,]0, 1[$ such that

$$\left( \frac{\sigma^2 l}{2p(1-p)} + \frac{u_m^2}{p(1-p)^2} \right) < \kappa l^3,$$

where $u_m = \sup_{t \in \mathbb{R}_+} \|\mathbf{u}(t)\|_2$.

It corresponds to making sure the external (*i.e.*, $u_m$) or internal (*i.e.*, $\sigma$) excitations are not too large compared to the decay mechanism (represented by $\kappa$ and $l$). Note that if $p \in \,]0, 1[$, $u_m$ and $d$ are fixed, it is sufficient to increase $\kappa$ or $l$ for this assumption to be satisfied.

Under this assumption, the space

$$E_p = \left\{ \mathbf{W} \in \mathbb{R}^{n \times n} : \mathbf{W} \text{ is symmetric, } \mathbf{W} \geq 0 \text{ and } \mathbf{W} < p\mathbf{L} \right\}$$

is invariant by the flow of the averaged system $\bar{G}$, where $\mathbf{W} \geq 0$ means $\mathbf{W}$ is semi-definite positive and $\mathbf{W} < p\mathbf{L}$ means $p\mathbf{L} - \mathbf{W}$ is definite positive. Therefore, the averaged system is defined and bounded on $\mathbb{R}_+$. The slow/fast system being asymptotically close to the averaged system, it is therefore asymptotically well-defined in probability. This is summarized in the following theorem.

**Theorem 3.1** *If Assumption* 3.1 *is verified for* $p \in \,]0, 1[$, *then system* (12) *is asymptotically well posed in probability and the connectivity matrix* $\mathbf{W}^\epsilon$, *the solution of system* (12), *converges to* $\mathbf{W}$ *in the sense that for all* $\delta, T > 0$,

$$\lim_{\epsilon \to 0}^{\mu} \mathbb{P} \left[ \sup_{t \in [0,T]} \left| \mathbf{W}_t^\epsilon - \mathbf{W}_t \right|^2 > \delta \right] = 0,$$

*where* $\mathbf{W}$ *is the deterministic solution of*

$$\frac{d\mathbf{W}_{ij}}{dt} = \bar{G}(\mathbf{W})_{ij} = \underbrace{-\kappa \mathbf{W}_{ij}}_{decay} + \underbrace{\frac{\mu}{\tau} \int_0^{\frac{\tau}{\mu}} \bar{\mathbf{v}}_i(s)\bar{\mathbf{v}}_j(s)\, ds}_{correlation} + \underbrace{\frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})_{ij}^{-1}}_{noise}, \qquad (13)$$

*where* $\bar{\mathbf{v}}(t)$ *is the* $\frac{\tau}{\mu}$-*periodic attractor of* $\frac{d\bar{\mathbf{v}}}{dt} = (\mathbf{W} - \mathbf{L}) \cdot \bar{\mathbf{v}} + \mathbf{u}(\mu t)$, *where* $\mathbf{W} \in \mathbb{R}^{n \times n}$ *is supposed to be fixed.*

*Proof* See Theorem B.1 in Appendix B.2. □

In the following, we focus on the averaged system described by (13). Its right-hand side is made of three terms: a linear and homogeneous decay, a correlation term, and a noise term. The last two terms are made explicit in the following.

### 3.2.1 Noise term

As seen in Section 2, in the linear case, the noise term $\mathbf{Q}$ is the unique solution of the Lyapunov equation (9) with $\mathbf{A} = \mathbf{W} - \mathbf{L}$ and $\Sigma = \sigma Id$. Because the noise is spatially uncorrelated and identical for each neuron and also because the connectivity is symmetric, observe that $\mathbf{Q} = \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})^{-1}$ is the unique solution of the system.

In more complicated cases, the computation of this term appears to be much more difficult as we will see in Section 3.4.

### 3.2.2 Correlation term

This term corresponds to the auto-correlation of neuronal activity. It is only implicitly defined; thus, this section is devoted to finding an explicit form depending only on the parameters $l$, $\mu$, $\tau$, the connectivity $\mathbf{W}$, and the inputs $\mathbf{u}$. Actually, one can perform an expansion of this term with respect to a small parameter corresponding to a *weakly connected expansion*. Most terms vanish if the connectivity $\mathbf{W}$ is small compared to the strength of the intrinsic decaying dynamics of neurons $l$.

The auto-correlation term of a $\frac{\tau}{\mu}$-periodic function can be rewritten as

$$\left\{\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}'\right\}_{ij} = \int_0^{\frac{\tau}{\mu}} \bar{\mathbf{v}}_i(s)\bar{\mathbf{v}}_j(s)\, ds.$$

With this notation, it is simple to think of $\mathbf{v}$ as a 'semi-continuous matrix' of $\mathbb{R}^{n \times [0, \frac{\tau}{\mu}[}$. Hence, the operator '$\cdot$' can be though of as a matrix multiplication. Similarly, the transpose operator turns a matrix $\bar{\mathbf{v}} \in \mathbb{R}^{n \times [0, \frac{\tau}{\mu}[}$ into a matrix $\bar{\mathbf{v}}' \in \mathbb{R}^{[0, \frac{\tau}{\mu}[ \times n}$. See Appendix B.1 for details about the notations.

It is common knowledge, see [17] for instance, that this term gathers information about the correlation of the inputs. Indeed, if we assume that the input is sufficiently slow, then $\bar{\mathbf{v}}$ has enough time to converge to $\mathbf{u}(t)$ for all $t \in [0, +\infty[$. Therefore, in the first order $\bar{\mathbf{v}}(t) \simeq (\mathbf{W} - \mathbf{L})^{-1} \cdot \mathbf{u}(t)$. This leads to $\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}' \simeq (\mathbf{W} - \mathbf{L})^{-1} \cdot \mathbf{u} \cdot \mathbf{u}' \cdot (\mathbf{W}' - \mathbf{L})^{-1}$. In the weakly connected regime, one can assume that $\mathbf{W} - \mathbf{L} \simeq -\mathbf{L}$ leading to $\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}' \simeq \frac{1}{l^2} \mathbf{u} \cdot \mathbf{u}'$ which is the auto-correlation of the inputs.

Actually, without the assumption of a slow input, lagged correlations of the input appear in the averaged system. Before giving the expression of these temporal correlations, we need to introduce some notations. First, define the convolution filter $g_{l/\mu} : t \mapsto \frac{l}{\mu} e^{-\frac{l}{\mu}t} H(t)$, where $H$ is the Heaviside function. This family of functions is displayed for different values of $\frac{l}{\mu}$ in Figure 4(a). Note that $g_{l/\mu} \to \delta_0$ when $\frac{l}{\mu} \to +\infty$, where $\delta_0$ is the Dirac distribution centered at the origin. In this asymptotic regime, the convolution filter and its iterates $g_{l/\mu} * \cdots * g_{l/\mu}$ are equal to the identity.

We also define the filtered correlation of the inputs $\mathbf{C}^{k,p} \in \mathbb{R}^{n \times n}$ by

$$\mathbf{C}^{k,q} \overset{\text{def}}{=} \frac{1}{u_m^2 \tau}\left(\mathbf{u} * g_{l/\mu}^{(k+1)}\right) \cdot \left(\mathbf{u} * g_{l/\mu}^{(q+1)}\right)',$$

where $g_{l/\mu}^{(k+1)} = g_{l/\mu} * \cdots * g_{l/\mu}$ is the $k$th convolution of $g_{l/\mu}$ with itself and $u_m = \sup_{t \in \mathbb{R}_+} \|\mathbf{u}(t)\|_2$. This is the correlation matrix of the inputs filtered by two different

**Fig. 1** This shows the $(k, q)$-temporal profiles with $\frac{l}{\mu} = 1$, *i.e.*, the functions $g_1^{(k+1)} * g_1'^{(q+1)}$ for $q = 0$ and $k$ ranging from 0 to 6. For $k = q = 0$, the temporal profile is even and this also occurs to be true for any $k = q$. When $k > q$, the function reaches its maximum for strictly positive values that grow with the difference $k - q$. Besides, the temporal profiles are flattened when $k + q$ increases.
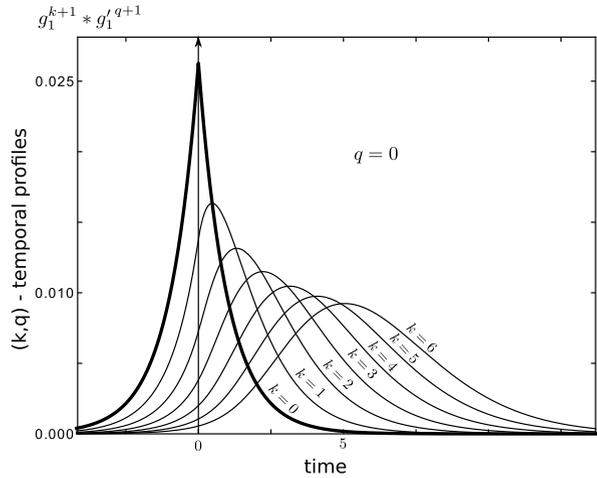


functions. It is easy to show that this is similar to computing the cross-correlation of the inputs with the inputs filtered by another function,

$$\mathbf{C}^{k,q} = \frac{1}{u_m^2 \tau} \left( \mathbf{u} * \left( g_{l/\mu}^{(k+1)} * g_{l/\mu}^{(q+1)'} \right) \right) \cdot \mathbf{u}'$$

$$= \frac{1}{u_m^2 \tau} \mathbf{u} \cdot \left( \mathbf{u} * \left( g_{l/\mu}^{(k+1)} * g_{l/\mu}^{(q+1)'} \right) \right)', \tag{14}$$

which motivates the definition of the $(k, p)$-temporal profile $g_{l/\mu}^{(k+1)} * g_{l/\mu}'^{(q+1)}$, where $(g_{l/\mu}')^{(k)}(t) = (g_{l/\mu}^{(k)})'(t) = g_{l/\mu}^{(k)}(-t)$. This notation is deliberately similar to that of the transpose operator we use in the proofs. These functions are shown in Figure 1. We have not found a way to make them explicit; therefore, the following remarks are simply based on numerical illustrations. When $k = q$, the temporal profiles are centered. The larger the difference $k - q$, the larger the center of the bell. The larger the sum $k + q$, the larger the standard deviation. This motivates the idea that $\mathbf{C}^{k,p}$ can be thought of as the $k - q$ lagged correlation of the inputs. One can also say that $\mathbf{C}^{10,10}$ is more blurred than $\mathbf{C}^{0,0}$ in the sense that the inputs are temporally integrated over a 'wider' window in the first case.

Observe that $g_{l/\mu}^{(k+1)}(t) = \frac{l^{k+1}}{\mu^{k+1}k!} t^k e^{-\frac{l}{\mu}t} H(t)$. Therefore, $\|g_{l/\mu}^{(k+1)}\|_1 = \frac{\Gamma(k+1)}{k!} = 1$. Thanks to Young's inequality for convolutions, which says that $\|\mathbf{u} * g_{l/\mu}^{(k)}\|_2 \leq \|\mathbf{u}\|_2 \|g_{l/\mu}^{(k)}\|_1$, it can be proved that $\|\mathbf{C}^{k,q}\|_2 \leq 1$.

We intend to express the correlation term as an infinite converging sum involving these filtered correlations. In this perspective, we use a result we have proved in [25] to write the solution of a general class of non-autonomous linear systems (*e.g.*, $\frac{d\bar{\mathbf{v}}}{dt} = (\mathbf{W} - \mathbf{L}) \cdot \bar{\mathbf{v}} + \mathbf{u}(t)$) as an infinite sum, in the case $\mu = 1$.

**Lemma 3.2** *If $\bar{\mathbf{v}}$ is the solution, with zero as initial condition, of $\frac{d\bar{\mathbf{v}}}{dt} = (\mathbf{W} - \mathbf{L}) \cdot \bar{\mathbf{v}} + \mathbf{u}(t)$ it can be written by the sum below which converges if $\mathbf{W}$ is in $E_p$ for $p \in \,]0,1[$.*

$$\bar{\mathbf{v}} = \sum_{k=0}^{+\infty} \frac{\mathbf{W}^k}{l^{k+1}} \cdot \mathbf{u} * g_l^{(k+1)},$$

*where $g_l : t \mapsto le^{-lt}H(t)$.*

*Proof* See Lemma B.2 in Appendix B.2.                                                                  $\square$

This is a decomposition of the solution of a linear differential system on the basis of operators where the spatial and temporal parts are decoupled. This important step in a detailed study of the averaged equation cannot be achieved easily in models with non-linear activity. Everything is now set up to introduce the explicit expansion of the correlation we are using in what follows. Indeed, we use the previous result to rewrite the correlation term as follows.

**Property 3.3** *The correlation term can be written*

$$\frac{\mu}{\tau} \bar{\mathbf{v}} \cdot \bar{\mathbf{v}}' = \frac{u_m^2}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{W}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q}.$$

*Proof* See Theorem B.3 in Appendix B.2.                                                                $\square$

This infinite sum of convolved filters is reminiscent of a property of Hawkes processes that have a linear input-output gain [26].

The speed of inputs characterized by $\mu$ only appears in the temporal profiles $g_{l/\mu}^{(k)} * g_{l/\mu}'^{(q)}$. In particular, if the inputs are much slower than neuronal activity time-scale, *i.e.*, $\mu = 0$, then $g_{+\infty} = \delta_0$ and $\mathbf{u} * g_{+\infty} = \mathbf{u}$. Therefore, $\mathbf{C}^{k,q} = \mathbf{C}^{0,0}$ and the sums in the formula of Property 3.3 are separable, leading to $\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}' = (\mathbf{L} - \mathbf{W})^{-1} \cdot \mathbf{u} \cdot \mathbf{u}' \cdot (\mathbf{L} - \mathbf{W}')^{-1}$, which corresponds to the heuristic result previously explained.

Therefore, the averaged equation can be explicitly rewritten

$$\frac{d\mathbf{W}}{dt} = \bar{G}(\mathbf{W}) = -\kappa \mathbf{W} + \frac{u_m^2}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{W}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q} + \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})^{-1}. \quad (15)$$

In Figure 2, we illustrate this result by comparing, for different $\epsilon = \epsilon_1 = \epsilon_2$ (*i.e.*, we choose $\mu = 1$ in this example), the stochastic system and its averaged version. The above decomposition has been used as the basis for numerical computation of trajectories of the averaged system.

### 3.2.3 Global stability of the equilibrium point

Now that we have found an explicit formulation for the averaged system, it is natural to study its dynamics. Actually, we prove in the following that if the connectivity $\mathbf{W}$
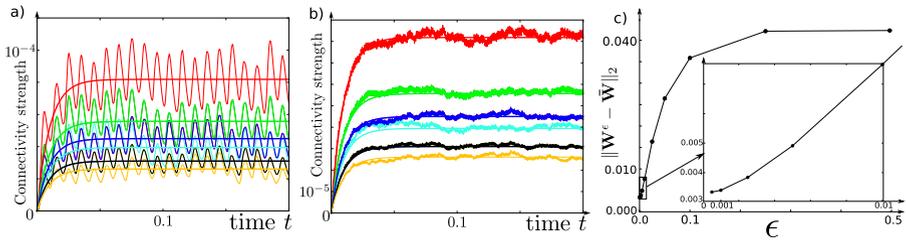
**Fig. 2** The first two figures, (**a**) and (**b**), represent the evolution of the connectivity for original stochastic system (12), superimposed with averaged system (13), for two different values of $\epsilon$: respectively $\epsilon = 0.01$ and $\epsilon = 0.001$, where we have chosen $\epsilon = \epsilon_1 = \epsilon_2$. Each color corresponds to the weight of an edge in a network made of $n = 3$ neurons. As expected, it seems that the smaller $\epsilon$, the better the approximation. This can be seen in the picture (**c**) where we have plotted the precision on the $y$-axis and $\epsilon$ on the $x$-axis. The parameters used here are $l = 12$, $\mu = 1$, $\kappa = 100$, $\sigma = 0.05$. The inputs have a random (but frozen) spatial structure and evolve according to a sinusoidal function.

is kept smaller than $\frac{l}{3}$, *i.e.*, Assumption 3.1 is verified with $p \leq \frac{1}{3}$, then the dynamics is trivial: the system converges to a single equilibrium point. Indeed, under the previous assumption, the system can be written $\bar{G}(\mathbf{W}) = -\kappa \mathbf{W} + F(\mathbf{W})$, where $F$ is a contraction operator on $E_{\frac{1}{3}}$. Therefore, one can prove the uniqueness of the fixed point with the Banach fixed point argument and exhibit an energy function for the system.

**Theorem 3.4** *If Assumption 3.1 is verified for $p \leq \frac{1}{3}$, then there is a unique equilibrium point in the invariant subset $E_p$ which is globally, asymptotically stable.*

*Proof* See Theorem B.4 in Appendix B.2.                                    □

The fact that the equilibrium point is unique means that the 'knowledge' of the network about its environment (corresponding by hypothesis to the connectivity) eventually is unique. For a given input and any initial condition, the network can only converge to the same 'knowledge' or 'understanding' of this input.

### 3.2.4 Explicit expansion of the equilibrium point

When the network is weakly connected, the high-order terms in expansion (15) may be neglected. In this section, we follow this idea and find an explicit expansion for the equilibrium connectivity where the strength of the connectivity is the small parameter enabling the expansion. The weaker the connectivity, the more terms can be neglected in the expansion.

In fact, it is not natural to speak about a weakly connected learning network since the connectivity is a variable. However, we are able to identify a *weak connectivity index* which controls the strength of the connectivity. We say the connectivity is weak when it is negligible compared to the intrinsic leak term, *i.e.*, $\frac{\|\|\mathbf{W}\|\|}{l}$ is small. We show in the Appendix that this weak connectivity index depends only on the parameters of

the network and can be written

$$\tilde{p} = \frac{u_m^2}{\kappa l^3} + \frac{\sigma^2}{2\kappa l^2}.$$

In the asymptotic regime $\tilde{p} \to 0$, we have $\frac{W}{\tilde{p}l} = \mathcal{O}(1)$. This index is the 'small' parameter needed to perform the expansion. We also define $\lambda = \frac{\sigma^2 l}{2u_m^2}$, which has information about the way $\tilde{p}$ is converging to zero. In fact, it is the ratio of the two terms of $\tilde{p}$.

With these, we can prove that the equilibrium connectivity $\mathbf{W}^*$ has the following asymptotic expansion in $\tilde{p}$.

**Theorem 3.5**

$$\mathbf{W}^* = \frac{\tilde{p}l}{1+\lambda}\big(\lambda + \mathbf{C}^{0,0}\big) + \frac{\tilde{p}^2 l}{(1+\lambda)^2}\big(\lambda^2 + \lambda\big(\mathbf{C}^{0,0} + \mathbf{C}^{1,0} + \mathbf{C}^{0,1}\big)$$
$$+ \mathbf{C}^{0,0} \cdot \mathbf{C}^{1,0} + \mathbf{C}^{0,1} \cdot \mathbf{C}^{0,0}\big) + \mathcal{O}\big(\tilde{p}^3\big).$$

*Proof* See Theorem B.5 in Appendix B.2.                                      □

At the first order, the final connectivity is $\mathbf{C}^{0,0}$, the filtered correlation of the inputs convolved with a bell-shaped centered temporal profile. In the case of Figure 3, this is quite a good approximation of the final connectivity.

Not only the spatial correlation is encoded in the weights, but there is also some information about the temporal correlation, *i.e.*, two successive but orthogonal events occurring in the inputs will be wired in the connectivity although they do not appear in the spatial correlations; see Figure 3 for an example.

### 3.3 Trace learning: band-pass filter effect

In this section, we study an improvement of the learning model by adding a certain form of history dependence in the system and explain the way it changes the results of the previous section. Given that Theorem 2.2 only applies to an instantaneous process, we will only be able to treat the history-dependent systems which can be reformulated as instantaneous processes. Actually, this class of systems contains models which are biologically more relevant than the previous model and which will exhibit interesting additional functional behaviors. In particular, this covers the following features:

- Trace learning.

    It is likely that a biological learning rule will integrate the activity over a short time. As Földiàk suggested in [27], it makes sense to consider the learning equation as being

$$\frac{d\mathbf{W}^\epsilon}{dt} = -\kappa \mathbf{W}^\epsilon + \big(\mathbf{v}^\epsilon * g_1\big) \otimes \big(\mathbf{v}^\epsilon * g_1\big),$$

    where $*$ is the convolution and $g_1 : t \in \mathbb{R} \mapsto \beta_1 e^{-\beta_1 t} H(t)$. Rolls and Deco numerically show [15] that the temporal convolution, leading to a spatio-temporal
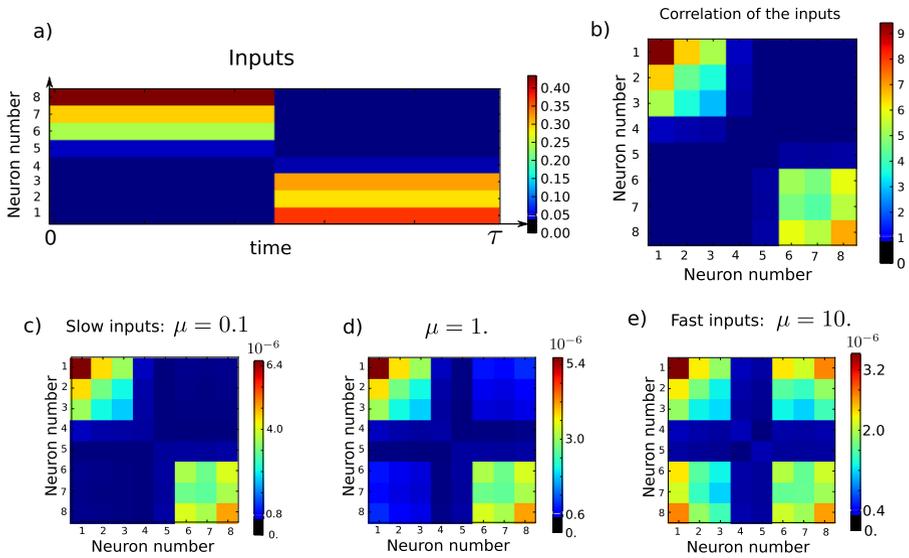
**Fig. 3** (**a**) shows the temporal evolution of the input to a $n = 8$ neurons network. It is made of two spatially random patterns that are shown alternatively. (**b**) shows the correlation matrix of the inputs. The off-diagonal terms are null because the two patterns are spatially orthogonal. (**c**), (**d**), and (**e**) represent the first order of Theorem 3.5 expansion for different $\mu$. Actually, this approximation is quite good since the percentage of error between the averaged system and the first order, computed by error $= \frac{\|\mathbf{W} - \text{order 1}\|_1}{\|\mathbf{W}\|_1}$, have an order of magnitude of $10^{-4}\%$ for the three figures. These figures make it possible to observe the role of $\mu$. If $\mu$ is small, *i.e.*, the inputs are slow, then the transient can be neglected and the learned connectivity is roughly the correlation of the inputs; see (a). If $\mu$ increases, *i.e.*, the inputs are faster, then the connectivity starts to encode a link between the two patterns that were flashed circularly and elicited responses that did not fade away when the other pattern appeared. The temporal structure of the inputs is also learned when $\mu$ is large. The parameters used in this figure are $\epsilon = 0.001, l = 12, \kappa = 100, \sigma = 0.02$.

learning, makes it possible to perform invariant object recognition. Besides, trace learning appears to be the symmetric part of the biological STDP rule that we detail in Section 3.4.

- Damped oscillatory neurons.

    Many neurons have an oscillatory behavior. Although we cannot take this into account in a linear model, we can model a neuron by a damped oscillator, which also introduces a new important time-scale in the system. Adding adaptation to neuronal dynamics is an elementary way to implement this idea. This corresponds to modeling a single neuron without inputs by the equivalent formulations

$$\begin{cases} \frac{dv^\epsilon}{dt} = -lz^\epsilon, \\ \frac{dz^\epsilon}{dt} = \beta_2(v^\epsilon - z^\epsilon) \end{cases} \Leftrightarrow \begin{cases} \frac{dv^\epsilon}{dt} = -lv^\epsilon * g_2, \\ \text{where } g_2(t) = \beta_2 e^{-\beta_2 t} H(t). \end{cases}$$

- Dynamic synapses.

    The electro-chemical process of synaptic communication is very complicated and non-linear. Yet, one of the features of synaptic communication we can take into account in a linear model is the shape of the post-synaptic potentials. In this

section, we consider that each synapse is a linear filter whose finite impulse response (*i.e.*, the post-synaptic potential) has the shape $g_3(t) = \beta_3 e^{-\beta_3 t} H(t)$. This is a common assumption which, for instance, is at the basis of traditional rate based models; see Chapter 11 of [7].

For mathematical tractability, we assume in the following that $\beta = \beta_1 = \beta_2 = \beta_3 \in \mathbb{R}_+$ such that $g_\beta = g_1 = g_2 = g_3$, *i.e.*, the time-scales of the neurons, those of the synapses and those of the learning windows are the same. Actually, there is a large variety of temporal scales of neurons, synapses, and learning windows, which makes this assumption not absurd. Besides, in many brain areas, examples of these time constants are in the same range ($\simeq 10$ ms). Yet, investigating the impact of breaking this assumption would be necessary to model better biological networks. This leads to the following system:

$$\begin{cases} d\mathbf{v}^\epsilon = \frac{1}{\epsilon_1}((\mathbf{W}^\epsilon - \mathbf{L}) \cdot \mathbf{v}^\epsilon * g_\beta + \mathbf{u}(\frac{t}{\epsilon_2})) \, dt + \frac{\sigma}{\sqrt{\epsilon_1}} dB(t), \\ \frac{d\mathbf{W}^\epsilon}{dt} = -\kappa \mathbf{W}^\epsilon + (\mathbf{v}^\epsilon * g_\beta) \otimes (\mathbf{v}^\epsilon * g_\beta), \end{cases} \quad (16)$$

where the notations are the same as in Section 3.2. The behavior of a single neuron will be oscillatory damped if $\Delta = \sqrt{1 - 4\frac{l}{\beta}}$ is a pure imaginary number, *i.e.*, $4l > \beta$. This is the regime on which we focus. Actually, the Hebbian linear case of Section 3.2 corresponds to $\beta = +\infty$ in this delayed system.

To comply with the hypotheses of Theorem 2.2 (*i.e.*, no dependence of the history of the process), we can add a variable $\mathbf{z}$ to the system which takes care of integrating the variable $\mathbf{v}$ over an exponential window. It leads to the equivalent system (in the limit $\sigma_\mathbf{z} \to 0$)

$$\begin{cases} d\binom{\mathbf{v}^\epsilon}{\mathbf{z}^\epsilon} = \frac{1}{\epsilon_1}\left[\binom{0 \quad \mathbf{W}-\mathbf{L}}{\beta \quad -\beta}\binom{\mathbf{v}^\epsilon}{\mathbf{z}^\epsilon} + \binom{\mathbf{u}(\frac{t}{\epsilon_2})}{0}\right] dt + \binom{\frac{\sigma}{\sqrt{\epsilon_1}} dB(t)}{\frac{\sigma_\mathbf{z}}{\sqrt{\epsilon_1}} dB(t)}, \\ \frac{d\mathbf{W}^\epsilon}{dt} = -\kappa \mathbf{W}^\epsilon + \mathbf{z}^\epsilon \otimes \mathbf{z}^\epsilon. \end{cases}$$

This trick makes it possible to deal with some history-based processes where the dependence on the past is exponential.

It turns out most of the results of Section 3.2 remain true for system (16) as detailed in the following. The existence of the solution on $\mathbb{R}_+$ is proved in Theorem B.6. The computations show that in the averaged system, the noise term remains identical, whereas the correlation term is to be replaced by $\frac{\mu}{\tau}(\bar{\mathbf{v}} * g_\beta) \cdot (\bar{\mathbf{v}} * g_\beta)'$. Besides, Lemma 3.2 can be extended to our delayed system by changing only the temporal filters; see Lemma B.7. Together with Lemma C.3, this proves the result of Theorem B.8.

$$\frac{\mu}{\tau}(\bar{\mathbf{v}} * g_\beta) \cdot (\bar{\mathbf{v}} * g_\beta)' = \frac{u_m^2 \|v\|_1^2}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{W}^k}{(l/\|v\|_1)^k} \cdot \tilde{\mathbf{C}}^{k,q} \cdot \frac{\mathbf{W}'^q}{(l/\|v\|_1)^q},$$

where

$$\tilde{\mathbf{C}}^{k,q} = \frac{1}{u_m^2 \tau \|v\|_1^{k+q+2}} (\mathbf{u} * v^{(k+1)}) \cdot (\mathbf{u} * v^{(q+1)})',$$
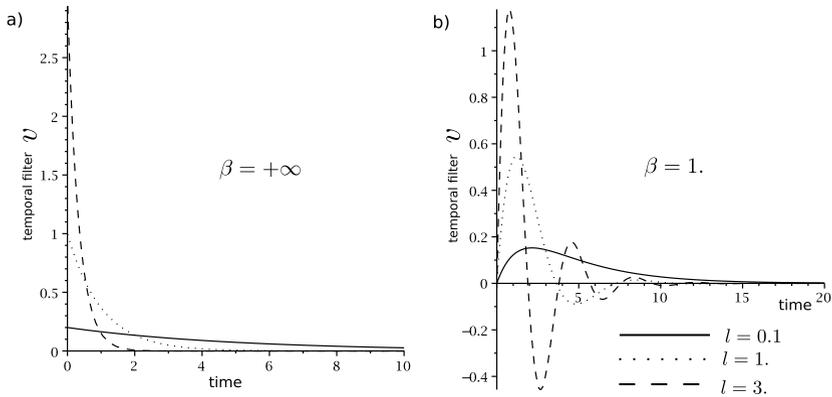
**Fig. 4** These represent the temporal filter $v : t \mapsto v(t)$ for different parameters. (**a**) When $\beta = +\infty$, we are in the Hebbian linear case of Appendix B.2. The temporal filters are just decaying exponentials which averaged the signal over a past window. (**b**) When the dynamics of the neurons and synapse are oscillatory damped, some oscillations appear in the temporal filters. The number of oscillations depends on $\Delta$. If $\Delta$ is real, then there are no oscillations as in the previous case. However, when $\Delta$ becomes a pure imaginary number, it creates a few oscillations which are even more numerous if $|\Delta|$ increases.

where $v : t \to \frac{l}{\mu\Delta}(e^{-\frac{\beta}{2\mu}(1-\Delta)t} - e^{-\frac{\beta}{2\mu}(1+\Delta)t})H(t)$. Observe that applying Young's inequality to convolutions leads to $\|\tilde{\mathbf{C}}^{k,q}\|_2 \leq 1$. Actually, Lemma C.3 shows that $v^{(k)} = v_k : t \mapsto \frac{\sqrt{\pi\beta}}{k!}e^{-\frac{\beta}{2}t}(\frac{t}{|\Delta|})^{k+\frac{1}{2}}J_{k+\frac{1}{2}}(\frac{\beta|\Delta|}{2}t)H(t)$, where $J_n(z)$ is the Bessel function of the first kind. The value of the L1 norm of $v$ is computed in Appendix C.3. It leads to $\|v\|_1 = \coth(\frac{\pi}{2\Delta})$ if $\Delta$ is a pure imaginary number and $\|v\|_1 = 1$ else.

Therefore, the averaged system can be rewritten

$$\frac{d\mathbf{W}}{dt} = \bar{G}(\mathbf{W}) = -\kappa\mathbf{W} + \frac{u_m^2\|v\|_1^2}{l^2}\sum_{k,q=0}^{+\infty}\frac{\mathbf{W}^k}{(l/\|v\|_1)^k}\cdot\tilde{\mathbf{C}}^{k,q}\cdot\frac{\mathbf{W}'^q}{(l/\|v\|_1)^q} + \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{W})^{-1}.$$

As before, the existence and uniqueness of a globally attractive equilibrium point is guaranteed if Assumption 3.1 is verified for $p \leq \frac{1}{2\|v\|_1^3+1}$; see Theorem B.9.

Besides, the weakly connected expansion of the equilibrium point we did in Section 3.2.4 can be derived in this case (see Theorem B.10). At the first order, this leads to the equilibrium connectivity

$$\mathbf{W}^* = \frac{\tilde{p}l}{1+\lambda}(\lambda + \|v\|_1^2\tilde{\mathbf{C}}^{0,0}) + \mathcal{O}(\tilde{p}^2\|v\|_1).$$

The second order is given in Theorem B.10. The main difference with the Hebbian linear case is the shape of the temporal filters. Actually, the temporal filters have an oscillatory damped behavior if $\Delta$ is purely imaginary. These two cases are compared in Figure 4.

These oscillatory damped filters have the effect of amplifying a particular frequency of the input signal. As shown in Figure 5, if $\Delta$ is a pure imaginary number, then $\mathbf{D}^{0,0}$ is the cross-correlation of the band-pass filtered inputs with themselves.

**Fig. 5** This is the spectral profile $|\widehat{v * v'}|(\xi)$ for $\beta = 1$ and $l \in ]0, 2]$, where $\widehat{v * v'}$ denotes the Fourier transform of $v * v'$. When $4l < \beta$, the filter reaches its maximum for the null frequency, but if $l$ increases beyond $\frac{\beta}{4}$, the filter becomes a band-pass filter with long tails in $\frac{1}{\xi^2}$.
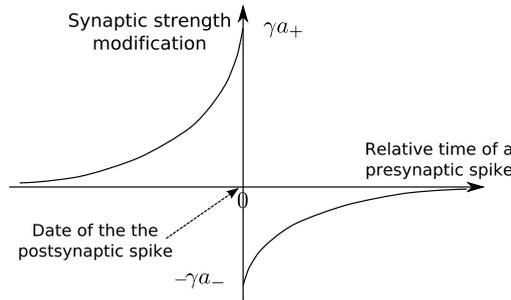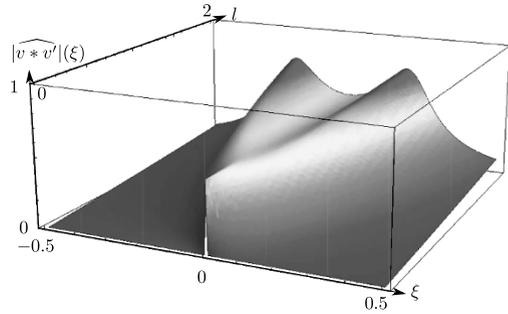




**Fig. 6** This figure represents the synapse strength modification when the post-synaptic neuron emits a spike. The $y$-axis corresponds to an additive or multiplicative update of the connectivity. For instance, in [28], this is $\frac{\Delta \mathbf{W}_{ij}}{\mathbf{W}_{ij}}$ for the negative part of the curve. However, we assume an additive update in this paper. The $x$-axis is the time at which a pre-synaptic spike reaches the synapse, relatively to the time of post-synaptic time chosen to be 0.

This band-pass filter effect can also be observed in the higher-order terms of the weakly connected expansion. This suggests that the biophysical oscillatory behavior of neurons and synapses leads to selecting the corresponding frequency of the inputs and performing the same computation as for the Hebbian linear case of the previous section: computing the correlation of the (filtered) inputs.

### 3.4 Asymmetric 'STDP' learning with correlated noise

Here, we extend the results to temporally asymmetric learning rules and spatially correlated noise. We consider a learning rule that is similar to the spike-timing-dependent plasticity (STDP) which is closer to biological experiments than the previous Hebbian rules. It has been observed that the strength of the connection between two neurons depends mainly on the difference between the time of the spikes emitted by each neuron as shown in Figure 6; see [12].

Assuming that the decay time of the positive and negative parts of Figure 6 are equal, we approximate this function by $t \mapsto a_+ g_\gamma(-t) - a_- g_\gamma(t)$, where $g_\gamma(t) = \gamma e^{-\gamma t} H(t)$. Actually, this corresponds to $\dot{\mathbf{W}}^\epsilon_{ij} = -\kappa \mathbf{W}^\epsilon_{ij} + a_+ \mathbf{v}_i (\mathbf{v}^\epsilon_j * g_\gamma) - a_- (\mathbf{v}^\epsilon_i * g_\gamma) \mathbf{v}^\epsilon_j$. If the neuron has a spiking behavior, then the term $a_+ \mathbf{v}^\epsilon_i(t)(\mathbf{v}^\epsilon_j * g_\gamma)(t)$ is significant when the post-synaptic neuron $i$ is spiking at time $t$, and then it counts

the number of previous spikes from the pre-synaptic neuron $j$ that might have caused the post-synaptic spike. This calculus is weighted by an exponentially decaying function. This accounts for the left part of Figure 6. The last term $-a_-(\mathbf{v}_i^\epsilon * g_\gamma)\mathbf{v}_j^\epsilon$ takes the opposite perspective. It is significant when the pre-synaptic neuron $j$ is spiking and counts the number of previous spikes from the post-synaptic neuron $i$ that are not likely to have been caused by the pre-synaptic neuron. The computation is also weighted by the mirrored function of an exponentially decaying function. This accounts for the right part of Figure 6. This leads to the coupled system

$$\begin{cases} d\mathbf{v}^\epsilon = \frac{1}{\epsilon_1}(f(\mathbf{v}^\epsilon) + \mathbf{W} \cdot \mathbf{v}^\epsilon + \mathbf{u}(\frac{t}{\epsilon_2}))\,dt + \frac{1}{\sqrt{\epsilon_1}}\mathbf{\Sigma} \cdot dB(t), \\ \frac{d\mathbf{W}^\epsilon}{dt} = G(\mathbf{v}^\epsilon, \mathbf{W}^\epsilon) = -\kappa\mathbf{W}^\epsilon + a_+\mathbf{v}^\epsilon \otimes (\mathbf{v}^\epsilon * g_\gamma) - a_-(\mathbf{v}^\epsilon * g_\gamma) \otimes \mathbf{v}^\epsilon, \end{cases} \tag{17}$$

where the non-linear intrinsic dynamics of the neurons is represented by $f$. Indeed, the term $\{a_+\mathbf{v}^\epsilon(t) \otimes (\mathbf{v}^\epsilon * g_\gamma)(t)\}_{ij} = a_+\mathbf{v}_i^\epsilon(t)(\mathbf{v}^\epsilon * g_\gamma)_j(t)$ is negligible when the neuron is quiet and maximal at the top of the spikes emitted by neuron $i$. Therefore, it records the value of the pre-synaptic membrane potential weighted by the function $g_\gamma$ when the post-synaptic neuron spikes. This accounts for the positive part of Figure 6. Similarly, the negative part corresponds to $-a_-(\mathbf{v}^\epsilon * g_\gamma) \otimes \mathbf{v}^\epsilon$.

Actually, this formulation is valid for any non-linear activity with correlated noise. However, studying the role of STDP in spiking networks is beyond the scope of this paper since we are only able to have explicit results for models with linear activity. Therefore, we will assume that the activity is linear while keeping the learning rule as it was derived in the spiking case, *i.e.*, we assume $f(\mathbf{v}) = -l\mathbf{v} = -\mathbf{L} \cdot \mathbf{v}$ in the system above.

We also use the trick of adding additional variables to get rid of the history-dependency. This reads

$$\begin{cases} d\begin{pmatrix} \mathbf{v}^\epsilon \\ \mathbf{z}^\epsilon \end{pmatrix} = \frac{1}{\epsilon_1}\left[\begin{pmatrix} \mathbf{W}-\mathbf{L} & 0 \\ \gamma & -\gamma \end{pmatrix}\begin{pmatrix} \mathbf{v}^\epsilon \\ \mathbf{z}^\epsilon \end{pmatrix} + \begin{pmatrix} \mathbf{u}(\frac{t}{\epsilon_2}) \\ 0 \end{pmatrix}\right]dt + \begin{pmatrix} \frac{\sigma}{\sqrt{\epsilon_1}}dB(t) \\ \frac{\sigma_\mathbf{z}}{\sqrt{\epsilon_1}}dB(t) \end{pmatrix}, \\ \frac{d\mathbf{W}^\epsilon}{dt} = -\kappa\mathbf{W}^\epsilon + a_+\mathbf{v}^\epsilon \otimes \mathbf{z}^\epsilon - a_-\mathbf{z}^\epsilon \otimes \mathbf{v}^\epsilon. \end{cases}$$

In this framework, the method exposed in Section 3.2 holds with small changes. First, the well-posedness assumption becomes

**Assumption 3.2** There exists $p \in ]0, 1[$ such that

$$\frac{|a_+| + |a_-|}{p(1-p)}\left(\frac{s^2\gamma}{2(1+\gamma/l-p)} + \frac{u_m^2}{(1-p)}\right) < \kappa l^3,$$

where $s^2$ is the maximal eigenvalue of $\mathbf{\Sigma} \cdot \mathbf{\Sigma}'$.

Under this assumption, the system is asymptotically well posed in probability (Theorem B.11). And we show the averaged system is

$$\frac{d\mathbf{W}}{dt} = \bar{G}(\mathbf{W}) = -\kappa\mathbf{W} + \frac{u_m^2(|a_+| + |a_-|)}{l^2}\sum_{k,q=0}^{+\infty}\frac{\mathbf{W}^k}{l^k} \cdot \mathbf{D}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q} + \mathbf{Q}, \tag{18}$$

where we have used Theorem B.12 to expand the correlation term. The noise term $\mathbf{Q}$ is equal to $\mathbf{Q}_{11} \cdot (\mathbf{L} + \gamma - \mathbf{W}')^{-1}$, where $\mathbf{Q}_{11}$ is the unique solution of the Lyapunov equation $(\mathbf{W} - \mathbf{L}) \cdot \mathbf{Q}_{11} + \mathbf{Q}_{11} \cdot (\mathbf{W}' - \mathbf{L}) + \mathbf{\Sigma} \cdot \mathbf{\Sigma}' = 0$. Lemma D.1 gives a solution for this equation which leads to $\mathbf{Q} = \gamma \sum_{k=0}^{+\infty} \mathbf{W}^k \cdot \mathbf{\Sigma} \cdot \mathbf{\Sigma}' \cdot (2\mathbf{L} - \mathbf{W}')^{-(k+1)} \cdot (\mathbf{L} + \gamma - \mathbf{W}')^{-1}$. In equation (18), the correlation matrices $\mathbf{D}^{k,q}$ are given by

$$\mathbf{D}^{k,q} = \frac{1}{u_m^2 \tau (|a_+| + |a_-|)} \left( \mathbf{u} * g_{l/\mu}^{(k+1)} * \left( a_+ g_\gamma' - a_- g_\gamma \right) \right) \cdot \left( \mathbf{u} * g_{l/\mu}^{(q+1)} \right)'.$$

According to Theorem B.13, the system is also globally asymptotically convergent to a single equilibrium, which we study in the following.

We perform a weakly connected expansion of the equilibrium connectivity of system (18). As shown in Theorem B.14, the first order of the expansion is

$$\mathbf{W}^* = \frac{\tilde{p}l}{1 + \lambda} \left( \lambda(\alpha_+ - \alpha_-) \frac{\mathbf{\Sigma} \cdot \mathbf{\Sigma}'}{d} + \mathbf{D}^{0,0} \right) + \mathcal{O}(\tilde{p}^2).$$

Writing $\mathbf{D}^{0,0} = \frac{1}{u_m^2 \tau (|a_+| + |a_-|)} (\mathbf{S} + \mathbf{A})$, where $\mathbf{S}$ is symmetric and $\mathbf{A}$ is skew-symmetric, leads to

$$\mathbf{S} = \frac{a_+ - a_-}{2} \mathbf{u} * g_{l/\mu} * \left( g_\gamma' + g_\gamma \right) \cdot (\mathbf{u} * g_{l/\mu})',$$

$$\mathbf{A} = \frac{a_+ + a_-}{2} \mathbf{u} * g_{l/\mu} * \left( g_\gamma' - g_\gamma \right) \cdot (\mathbf{u} * g_{l/\mu})'.$$

According to Lemma C.1, the symmetric part is very similar to the trace learning case in Section 3.3. Applying Lemma C.2 leads to

$$\mathbf{S} = (a_+ - a_-)(\mathbf{u} * g_{l/\mu} * g_\gamma) \cdot (\mathbf{u} * g_{l/\mu} * g_\gamma)',$$

$$\mathbf{A} = \frac{a_+ + a_-}{\gamma} \left( \frac{d\mathbf{u}}{dt} * g_{l/\mu} * g_\gamma \right) \cdot (\mathbf{u} * g_{l/\mu} * g_\gamma)'. \tag{19}$$

Therefore, the STDP learning rule simply adds an antisymmetric part to the final connectivity keeping the symmetric part as the Hebbian case. Besides, the antisymmetric part corresponds to computing the cross-correlation of the inputs with its derivative. For high-order terms, this remains true although the temporal profiles are different from the first order. These results are in line with previous works underlying the similarity between STDP learning and differential Hebbian learning, where $G(\mathbf{v}) \sim \dot{\mathbf{v}} \otimes \mathbf{v}$; see [29].

Figure 7 shows an example of purely antisymmetric STDP learning, *i.e.*, $a_+ = a_-$. The final connectivity matrix is therefore antisymmetric as shown in Figure 7(b) and the noise has no impact on learning. It shows the network finally approximates the connectivity given in (19).

# 4 Discussion

We have applied temporal averaging methods on slow/fast systems modeling the learning mechanisms occurring in linear stochastic neural networks. When we make
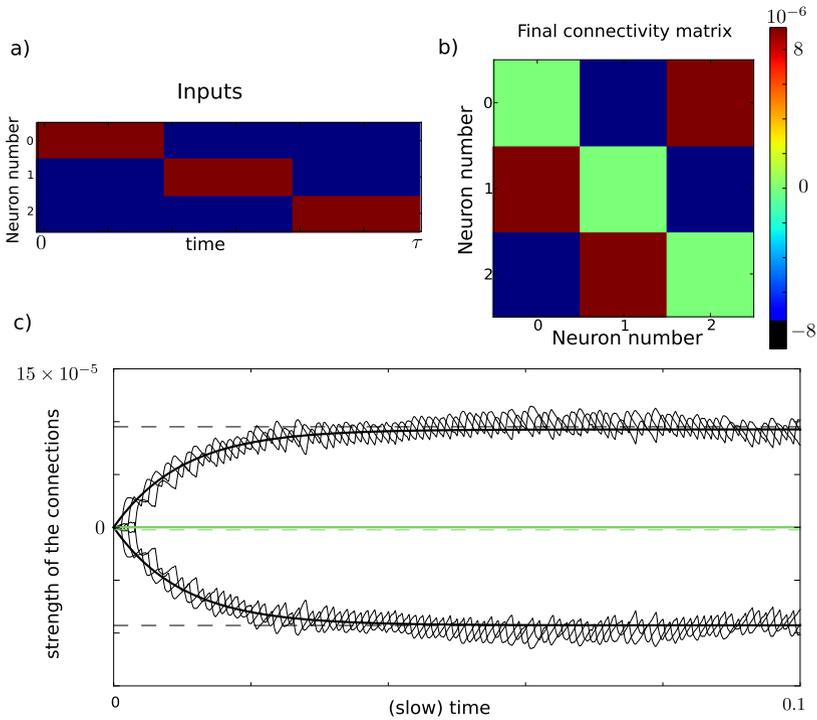
**Fig. 7** Antisymmetric STDP learning for a network of $n = 3$ neurons. (**a**) Temporal evolution of the inputs to the network. The three neurons are successively and periodically excited. The *red color* corresponds to an excitation of 1 and the *blue* to no excitation. (**b**) Equilibrium connectivity. The matrix is antisymmetric and shows that neurons excite one of their neighbors and are inhibited by the other. (**c**) Temporal evolution of the connectivity strength. The colors correspond to those of (b). The connectivity of system (17) corresponds to the *plain thin oscillatory curves*. The connectivity of the averaged system (18) (with $k, q < 4$) corresponds to the *plain thick lines*. Note that each curve corresponds to the superposition of three connections which remain equal through learning. The *dashed curves* correspond to the antisymmetric part in (19). The parameters chosen for this simulation were $l = 10$, $\kappa = 100$, $\gamma = 3$, $a_+ = a_- = 1$, $\tau = 3$, $\sigma = 0.001$, $\mu = 1$, $\epsilon = 0.001$. The system was simulated on the fast time-scale during $T = 10,000$ time steps of size $dt = 0.01$.

sure the connectivity remains small, the dynamics of the averaged system appears to be simple: the connectivity always converges to a unique equilibrium point. Then, we performed a weakly connected expansion of this final connectivity whose terms are combinations of the noise covariance and the lagged correlations of the inputs: the first-order term is simply the sum of the noise covariance and the correlation of the inputs.

- As opposed to the former input/ouput vision of the neurons, we have considered the membrane potential **v** to be the solution of a dynamical system. The consequence of this modeling choice is that not only the spatial correlations, but also the temporal correlations are learned. Due to the fact we take the transients into account, the activity never converges but it lives between the representation of the inputs. Therefore, it links concepts together.

The parameter $\mu$ is the ratio of the time-scales between the inputs and the activity variable. If $\mu = 0$, the inputs are infinitely slow and the activity variable has enough time to converge towards its equilibrium point. When $\mu$ grows, the dynamics becomes more and more transient, it has no time to converge. Therefore, if the inputs are extremely slow, the network only learns the spatial correlation of the inputs. If the inputs are fast, it also learns the temporal correlations. This is illustrated in Figure 3.

This suggests that learning associations between concepts, for instance, learning words in two different languages, may be optimized by presenting two words to be associated circularly with a certain frequency. Indeed, increasing the frequency (with a fixed duration of exposition to each word) amounts to increasing $\mu$. Therefore, the network learns better the temporal correlations of the inputs and thus strengthens the link between these two concepts.

- According to the model of resonator neuron [30], Section 3.3 suggests that neurons and synapses with a preferred frequency of oscillation will preferably extract the correlation of the inputs filtered by a band pass filter centered on the intrinsic frequency of the neurons.

  Actually, it has been observed that the auditory cortex is tonotopically organized, *i.e.*, the neurons are arranged by frequency [31]. It is traditionally thought that this is achieved thanks to a particular connectivity between the neurons. We exhibit here another mechanism to select this frequency which is solely based on the parameters of the neurons: a network with a lot of different neurons whose intrinsic frequencies are uniformly spread is likely to perform a Fourier-like operation: decomposing the signal by frequency.

  In particular, this emphasizes the fact that the network does not treat space and time similarly. Roughly speaking, associating several pictures and associating several sounds are therefore two different tasks which involve different mechanisms.

- In this paper, the original hierarchy of the network has been neglected: the network is made of neurons which receive external inputs. A natural way to include a hierarchical structure (with layers for instance), without changing the setup of the paper, is therefore to remove the external input to some neurons. However, according to Theorem 3.5 (and its extensions Theorems B.10 and B.14), one can see that these neurons will be disconnected from the others at the first order (if the noise is spatially uncorrelated). Linear activities imply that the high level neurons disconnect from others, which is a problem. In fact, one can observe that the second-order term in Theorem 3.5 is not null if the noise matrix $\boldsymbol{\Sigma}$ is not diagonal. In fact, this is the noise between neurons which will recruit the high level neurons to build connections from and to them.

  It is likely that a significant part of noise in the brain is locally induced, *e.g.*, local perturbations due to blood vessels or local chemical signals. In a way, the neurons close to each other share their noise and it seems reasonable to choose the matrix $\boldsymbol{\Sigma}$ so that it reflects the biological proximity between neurons. In fact, $\boldsymbol{\Sigma}$ specifies the original structure of the network and makes it possible for close-by neurons to recruit each other.

  Another idea to address hierarchy in networks would be to replace the synaptic decay term $-\kappa \mathbf{W}$ by another homeostatic term [32] which would enforce the emergence of a strong hierarchical structure.

- It is also interesting to observe that most of the noise contribution to the equilibrium connectivity for STDP learning (see Theorem B.14) vanishes if the learning is purely skew-symmetric, *i.e.*, $a_+ = a_-$. In fact, it is only the symmetric part of learning, *i.e.*, the Hebbian mechanism, that writes the noise in the connectivity.
- We have shown that there is a natural analogous STDP learning for spiking neurons in our case of linear neurons. This asymmetric rule converges to a final connectivity which can be decomposed into symmetric and skew-symmetric parts. The first one is similar to the symmetric Hebbian learning case, emphasizing that the STDP is nothing more than an asymmetric Hebbian-like learning rule. The skew-symmetric part of the final connectivity is the cross-correlation between the inputs and their derivatives.

  This has an interesting signification when looking at the spontaneous activity of the network post-learning. In fact, if we assume that the inputs are generated by an autonomous system $\frac{d\mathbf{u}}{dt} = \zeta(\mathbf{u})$, then according to the bottom equation in formula (19), the spontaneous activity is governed by

$$d\mathbf{v} = \left( \zeta(\mathbf{u}) \cdot \mathbf{u}' \cdot \mathbf{v} - l\mathbf{v} \right) dt + \mathbf{\Sigma} \cdot dB(t).$$

In a way, the noise terms generate random patterns which tend to be forgotten by the network due to the leak term $-l\mathbf{v}$. The only drift is due to $\zeta(\mathbf{u}) \cdot \mathbf{u}' \cdot \mathbf{v} \simeq \mathbf{E}_{\langle \mathbf{v}, \mathbf{u} \rangle}(\zeta(\mathbf{u}))$ which is the expectation of the vector field defining the dynamics of inputs with a measure being the scalar product between the activity variable and the inputs. In other words, if the activity is close to the inputs at a given time $t^* \in \mathbb{R}_+$, *i.e.*, $\langle \mathbf{v}, \mathbf{u}(t^*) \rangle$ is large, then the activity will evolve in the same direction as what this input would have done. The network has modeled the temporal structure of the inputs. The spontaneous activity predicts and replays the inputs the network has learned.

There are still numerous challenges to carry on in this direction.

First, it seems natural to look for an application of these mathematical methods to more realistic models. The two main limitations of the class of models we study in Section 3 are (i) the activity variable is governed by a linear equation and (ii) all the neurons are assumed to be identical. The mathematical analysis in this paper was made possible by the assumption that the neural network has a linear dynamics, which does not reflect the intrinsic non-linear behavior of the neurons. However, the cornerstone of the application of temporal averaging methods to a learning neural network, namely Property 3.3, is similar to the behavior of Poisson processes [26] which has useful applications for learning neural networks [19, 20]. This suggests that the dynamics studied in this paper might be quite similar to some non-linear network models. Studying more rigorously the extension of the present theory to non-linear and heterogeneous models is the next step toward a better modeling of biologically plausible neural networks.

Second, we have shown that the equilibrium connectivity was made of a symmetric and antisymmetric term. In terms of statistical analysis of data sets, the symmetric part corresponds to classical correlation matrices. However, the antisymmetric part suggests a way to improve the purely correlation-based approach used in many statistical analyses (*e.g.*, PCA) toward a causality-oriented framework which might be better suited to deal with dynamical data.

## Competing interests

## Authors' contributions

## Appendix A: Stochastic and periodic averaging

### A.1 Long-time behavior of inhomogeneous Markov processes

In order to construct the averaged vector field $\bar{G}_\mu(\mathbf{w})$ in the time-scale matching case ($0 < \mu < \infty$), one needs to understand properly the long-time behavior of the rescaled inhomogeneous frozen process

$$d\mathbf{v} = F(\mathbf{v}, \mathbf{w}_0, \mu t)\, dt + \mathbf{\Sigma}(\mathbf{v}, \mathbf{w}_0)\, dB(t). \tag{20}$$

Under regularity and dissipativity conditions, [5] proves the following general result about the asymptotic behavior of the solution of

$$dX_t = b(X_t, t)\, dt + \sigma(X_t, t)\, dB(t), \quad t > s,$$
$$X_s = x,$$

where $t \to b(x, t)$ and $t \to \sigma(x, t)$ are $\tau$-periodic.

The first point of the following theorem gives the definition of *evolution systems of measures*, which generalizes the notion of invariant measures in the case of inhomogeneous Markov processes. The exponential estimate of 2. in the following theorem is a key point to prove the averaging principle of Theorem 2.2.

**Theorem A.1** ([5])

1. *There exists a unique $\tau$-periodic family of probability measures $\{\mu(s, \cdot), s \in \mathbf{R}\}$ such that for all functions $\phi$ continuous and bounded,*

$$\int_{x \in \mathbf{R}^p} \mathbf{E}[\phi(X_t)]\mu(s, dx) = \int_{x \in \mathbf{R}^p} \phi(x)\mu(t, dx).$$

*Such a family is called evolution systems of measures.*

2. *Furthermore, under stronger dissipativity condition, the convergence of the law of X to $\mu$ is exponentially fast. More precisely, for any $r \in (1, +\infty)$, there exist $M > 0$ and $\omega < 0$ such that for all $\phi$ in the space of p-integrable functions with respect to $\mu(t, \cdot)$, $L^r(\mathbf{R}^p, \mu(t, \cdot))$,*

$$\int_{x \in \mathbf{R}^p} \left\| \mathbf{E}\big[\phi(X_t^{s,x})\big] - \int_{x' \in \mathbf{R}^p} \phi(x')\mu(t, dx') \right\|^r \mu(s, dx)$$

$$\leq M e^{\omega(t-s)} \int_{x \in \mathbf{R}^p} \|\phi(x)\|^r \mu(t, dx).$$

## A.2 Proof of Property 2.3

**Property A.2** *If there exists a smooth subset $\mathcal{E}$ of $\mathbb{R}^q$ such that*

1. *The functions $F$, $G$, $\mathbf{\Sigma}$ satisfy Assumptions 2.1-2.3 restricted on $\mathbb{R}^p \times \mathcal{E}$.*
2. *$\mathcal{E}$ is invariant under the flow of $\bar{G}_\mu$, as defined in (7).*

*Then for any initial condition $\mathbf{w}_0 \in \mathcal{E}$, system (4) is asymptotically well posed in probability and $\mathbf{w}^\epsilon$ satisfies the conclusion of Theorem 2.2.*

*Proof* The idea of the proof is to truncate the original system, replacing $G$ by a smooth truncation which coincides with $G$ on $\mathcal{E}$ and which is close to 0 outside $\mathcal{E}$. More precisely, for $\beta > 0$, we introduce $\psi_\beta : \mathbb{R}^q \to \mathbb{R}^q$ a regular function (locally Lipschitz) such that $\psi_\beta(\mathbf{w}) = 0$ if $\mathbf{w} \notin \mathcal{E}$ or $\mathbf{w} \in \partial \mathcal{E}$ and $\lim_{\beta \to 0} \psi_\beta(\mathbf{w}) = 1$ if $\mathbf{w} \in \mathcal{E} - \partial \mathcal{E}$. We define

$$\tilde{G}_\beta(\mathbf{v}, \mathbf{w}) = G(\mathbf{v}, \mathbf{w})\psi_\beta(\mathbf{w}).$$

Then, we introduce $(\tilde{\mathbf{v}}^{\epsilon,\beta}, \tilde{\mathbf{w}}^{\epsilon,\beta})$ the solution of the auxiliary system

$$d\tilde{\mathbf{v}}^{\epsilon,\beta} = \frac{1}{\epsilon_1}\left[ F\left(\tilde{\mathbf{v}}^{\epsilon,\beta}, \tilde{\mathbf{w}}^{\epsilon,\beta}, \frac{t}{\epsilon_2}\right)\right] dt + \frac{1}{\sqrt{\epsilon_1}} \mathbf{\Sigma}\big(\tilde{\mathbf{v}}^{\epsilon,\beta}, \tilde{\mathbf{w}}^{\epsilon,\beta}\big) dB(t),$$

$$d\tilde{\mathbf{w}}^{\epsilon,\beta} = \tilde{G}_\beta\big(\tilde{\mathbf{v}}^{\epsilon,\beta}, \tilde{\mathbf{w}}^{\epsilon,\beta}\big) dt$$

with the same initial condition as $(\mathbf{v}^\epsilon, \mathbf{w}^\epsilon)$.

Let $T, \delta, \eta > 0$ be three positive reals. Let us introduce a few more notations. We will need to consider a subset of $\mathcal{E}$ defined by

$$\mathcal{E}^\beta := \big\{\mathbf{w} \in \mathcal{E}; \|\psi_\beta(\mathbf{w}) - 1\| \leq (\eta)^{1/2}\delta\big\}.$$

We also introduce the following stopping times:

$$\tau_\epsilon := \inf\big\{t \geq 0; \mathbf{w}_t^\epsilon \in \partial \mathcal{E}\big\}, \qquad \tau_\epsilon^\beta := \inf\big\{t \geq 0; \mathbf{w}_t^\epsilon \in \partial \mathcal{E}^\beta\big\},$$

$$\tilde{\tau}_\epsilon := \inf\big\{t \geq 0; \tilde{\mathbf{w}}_t^{\epsilon,\beta} \in \partial \mathcal{E}\big\}, \qquad \tilde{\tau}_\epsilon^\beta := \inf\big\{t \geq 0; \tilde{\mathbf{w}}_t^{\epsilon,\beta} \in \partial \mathcal{E}^\beta\big\}.$$

Finally, we define $T_\epsilon := \min(T, \tau_\epsilon, \tilde{\tau}_\epsilon)$ and $T_\epsilon^\beta := \min(T, \tau_\epsilon^\beta, \tilde{\tau}_\epsilon^\beta)$.

Let us remark at this point that in order to prove that $\mathbb{P}[\tau_\epsilon \geq T] \to 1$ (which is our aim), it is sufficient to work on the bounded stopping time $\min(T, \tau_\epsilon)$, since $\mathbb{P}[\tau_\epsilon \geq T] = \mathbb{P}[\min(T, \tau_\epsilon) \geq T]$. In other words, the realizations of $\mathbf{w}^\epsilon$ which stay longer than $T$ inside $\mathcal{E}$ are not problematic. Therefore, we introduce $\hat{\tau}_\epsilon := \min(T, \tau_\epsilon)$.

Our first claim is that on finite time intervals $[0, T]$, $\tilde{\mathbf{w}}^{\epsilon, \beta}$ is a good approximation of $\mathbf{w}^\epsilon$ inside $\mathcal{E}$ as long as one chooses $\beta$ sufficiently small. To prove our claim, we proceed in two steps, first working inside $\mathcal{E}^\beta$ and then in $\mathcal{E} - \mathcal{E}^\beta$:

1. For any $\beta > 0$, one controls the difference between $\mathbf{w}^\epsilon$ and $\tilde{\mathbf{w}}^{\epsilon, \beta}$ on $\mathcal{E}^\beta$ since one controls the difference between the drifts. By an application of Lemma A.3 below (we need here the moment Assumption 2.3(i)), there exists a constant $C$ (which may depend on $T, \beta, \ldots$) such that

$$\mathbb{E}\left[\sup_{0 \leq t \leq T_\epsilon^\beta} \left\| \mathbf{w}_t^\epsilon - \tilde{\mathbf{w}}_t^{\epsilon, \beta} \right\|^2\right] \leq C \eta \delta^2. \tag{21}$$

We conclude by an application of the Markov inequality, implying

$$\mathbb{P}\left[\sup_{0 \leq t \leq T_\epsilon^\beta} \left\| \mathbf{w}_t^\epsilon - \tilde{\mathbf{w}}_t^{\epsilon, \beta} \right\| > \delta\right] \leq \frac{1}{\delta^2} \mathbb{E}\left[\sup_{0 \leq t \leq T_\epsilon^\beta} \left\| \mathbf{w}_t^\epsilon - \tilde{\mathbf{w}}_t^{\epsilon, \beta} \right\|^2\right] \leq C \eta. \tag{22}$$

2. One needs now to control the situation outside $\mathcal{E}^\beta$, that is, on $\mathcal{E} - \mathcal{E}^\beta$. The idea is that while one does not control the difference between $G$ and $\tilde{G}_\beta$ anymore, one can still choose $\beta$ sufficiently small such that $\mathcal{E}^\beta$ becomes arbitrary close to $\mathcal{E}$, hence implying that $\hat{\tau}_\epsilon$ and $T_\epsilon^\beta$ are arbitrary close with high probability, namely

$$\forall \theta, \lambda > 0, \exists \beta > 0, \quad \mathbb{P}\left[\tau_\epsilon - T_\epsilon^\beta > \lambda\right] < \theta. \tag{23}$$

With $\theta = (\delta \eta)^2$ and $\lambda = \delta \eta$, one obtains that for sufficiently small $\beta$,

$$\mathbb{P}\left[\hat{\tau}_\epsilon - T_\epsilon^\beta > \eta \delta\right] < (\eta \delta)^2. \tag{24}$$

Let us denote $S := \sup_{T_\epsilon^\beta \leq t \leq \hat{\tau}_\epsilon} \| \mathbf{w}_t^\epsilon - \tilde{\mathbf{w}}_t^{\epsilon, \beta} \|$. Then, one can split the calculus of $\mathbb{E}[S]$ according to the event $A = \{\hat{\tau}_\epsilon - T_\epsilon^\beta > \eta \delta\}$:

$$\begin{aligned}
\mathbb{E}[S] &= \mathbb{E}[S \mathbb{I}_A] + \mathbb{E}[S \mathbb{I}_{A^c}] \\
&\leq \left(2 K_G T \mathbb{P}[A]\right)^{1/2} + \left(2 K_G \mathbb{E}\left[\left(\hat{\tau}_\epsilon - T_\epsilon^\beta\right)^2 \mathbb{I}_{A^c}\right]\right)^{1/2} \\
&\leq C_2 \eta \delta,
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality and the moment Assumption 2.3(ii) (yielding the constant $K_G$) in the second line.

So, we deduce by the Markov inequality that $\sup_{T_\epsilon^\beta \leq t \leq \hat{\tau}_\epsilon} \| \mathbf{w}_t^\epsilon - \tilde{\mathbf{w}}_t^{\epsilon, \beta} \|$ is arbitrary small in probability.

From the combination of 1. and 2., we deduce that one can choose $\beta$ small enough such that

$$\mathbb{P}\left[\sup_{0 \leq t \leq T \wedge \tau_\epsilon} \left\| \mathbf{w}_t^\epsilon - \tilde{\mathbf{w}}_t^{\epsilon, \beta} \right\| > \delta\right] \leq (C_1 + C_2) \eta. \tag{25}$$

We can now proceed to the application of Theorem 2.2 to the truncated system. As $(\tilde{\mathbf{v}}^{\epsilon,\beta_0}, \tilde{\mathbf{w}}^{\epsilon,\beta_0})$ remains in $\mathbb{R}^p \times \mathcal{E}$, one can extend smoothly $F$ and $\Sigma$ outside $\mathcal{E}$ so that $(F, \Sigma)$ satisfies Assumptions 2.1-2.2. Therefore, one can apply Theorem 2.2 to the auxiliary system: for all $\delta, T > 0$,

$$\lim_{\epsilon \to 0}^{\mu} \mathbb{P}\left[\sup_{t \in [0,T]} \left\| \tilde{\mathbf{w}}_t^{\epsilon,\beta_0} - \mathbf{w}_t \right\| > \delta \right] = 0,$$

where $\mathbf{w}$ is defined by (8). As a consequence, there exists $\epsilon_0$ such that for all $\epsilon$ with $\|\epsilon\| < \epsilon_0$,

$$\mathbb{P}\left[\sup_{t \in [0,T]} \left\| \tilde{\mathbf{w}}_t^{\epsilon,\beta_0} - \mathbf{w}_t \right\| > \delta \right] < \eta.$$

Then, as $|\hat{\mathbf{w}}_t^{\epsilon} - \mathbf{w}_t| \leq |\hat{\mathbf{w}}_t^{\epsilon} - \tilde{\mathbf{w}}_t^{\epsilon,\beta_0}| + |\tilde{\mathbf{w}}_t^{\epsilon,\beta_0} - \mathbf{w}_t|$, one deduces that for all $\epsilon$ with $\|\epsilon\| < \epsilon_0$,

$$\mathbb{P}\left[\sup_{t \in [0,T]} \left\| \hat{\mathbf{w}}_t^{\epsilon} - \mathbf{w}_t \right\| > \delta \right] < (C_1 + C_2 + 1)\eta,$$

that is to say,

$$\lim_{\epsilon \to 0}^{\mu} \mathbb{P}\left[\sup_{t \in [0,T]} \left\| \hat{\mathbf{w}}_t^{\epsilon} - \mathbf{w}_t \right\| > \delta \right] = 0.$$

We know by assumption 2. of the statement of Property 2.3, for all $t \geq 0$, $\mathbf{w}_t \in \mathcal{E}$, so we conclude the proof by observing that for all $T > 0$,

$$\lim_{\epsilon \to 0} \mathbb{P}[\tau_\epsilon \geq T] = 1. \qquad \square$$

In the following lemma, we show that the solutions of two SDEs, whose drifts are close on a subset of the state space, remain close on a finite time interval. The difficulty here lies in the fact that we deal with only *locally Lipschitz* coefficients.

**Lemma A.3** *Suppose $x$ and $y$ are solutions, with identical initial conditions in $\mathcal{H} \subset \mathbb{R}^n$, of the following stochastic differential equations in $\mathbb{R}^n$:*

$$dx_t = a(x_t, t)\, dt + b(x_t, t)\, dB(t), \tag{26}$$

$$dy_t = h(y_t)a(y_t, t)\, dt + b(y_t, t)\, dB(t). \tag{27}$$

*Let $T > 0$ be a fixed time. We define*

$$\tau_{\mathcal{H}} = \min\big(\inf\{t \geq 0;\, x_t \in \partial H\}, \inf\{t \geq 0;\, y_t \in \partial H\}\big).$$

*We make the following assumptions*:

1. *Approximation assumption*:

$$\sup_{y \in \mathcal{H}} \left\| h(y) - 1 \right\| \leq \xi;$$

2. *Local Lipschitz assumption*: *for all $a, b \in \mathbb{R}^n$ with $\max(\|a\|, \|b\|) \leq R$, there exists a constant $C_R$ such that*

$$\left\| a(x, t) - a(y, t) \right\|^2 \leq C_R \|x - y\|^2;$$

3. *Boundedness assumption*: *there exists $p > 2$ and $A > 0$ such that*

$$\mathbb{E}\left[ \sup_{0 \leq t \leq T} \|x_t\|^p \right] \leq A \quad and \quad \mathbb{E}\left[ \sup_{0 \leq t \leq T} \|y_t\|^p \right] \leq A,$$

*and if $\|x\| \leq R$, then there exists $K_R$ such that $\|a(x)\| \leq K_R$.*

*Under the above assumptions, there exists a constant $C$ (depending on the quantities defined above, but not on $\xi$) such that*

$$\mathbb{E}\left[ \sup_{0 \leq t \leq \min(T, \tau_\mathcal{H})} \|x_t - y_t\|^2 \right] \leq C\xi^2. \tag{28}$$

*Proof* Although the Lipschitz constant is not bounded on $\mathcal{H}$, we can use the boundedness assumption to show that the probability of reaching a level $R$ before time $T$ will be very small for large $R$, and then use the classical strategy inside $\{\|x_t\| \leq R\}$ where everything works as if the coefficients were globally Lipschitz. A similar strategy is used in [33] to prove a strong convergence theorem for the Euler scheme without the global Lipschitz assumption. We adapt here the ideas of their proof to our setting.

Therefore, we introduce the following stopping times:

$$\theta_R := \inf\{t \geq 0; \|x_t\| \geq R\},$$

$$\theta_R^\beta := \inf\{t \geq 0; \|y_t\| \geq R\} \quad and \quad \rho := \min(\theta_R, \theta_R^\beta, \tau_\mathcal{H}).$$

We also denote $e(t) := x_t - y_t$.

Splitting the following expectation according to the value of $\rho$, and applying the Young inequality,

$$ab \leq \frac{d}{r}a^r + \frac{1}{qd^{q/r}}b^q \quad \text{for } r^{-1} + q^{-1} = 1 \text{ and any } a, b, d > 0,$$

we obtain, for any $d > 0$,

$$\mathbb{E}\left[ \sup_{0 \leq t \leq \min(T, \tau_\mathcal{H})} \|e(t)\|^2 \right]$$

$$\leq \mathbb{E}\left[ \sup_{0 \leq t \leq \min(T, \tau_\mathcal{H})} \|e(\min(t, \rho))\|^2 \right]$$

$$+ \frac{2d}{p}\mathbb{E}\left[ \sup_{0 \leq t \leq T} s\|e(t)\|^p \right] + \frac{1 - 2/p}{d^{2/(p-2)}}\mathbb{P}\left[\theta_R \leq T \text{ or } \theta_R^\beta \leq T\right].$$

Then we use the boundedness assumption and the Markov inequality to deduce that

$$\mathbb{P}\left[\theta_R \leq T \text{ or } \theta_R^\beta \leq T\right] \leq \frac{2A}{R^p} \quad \text{and} \quad \mathbb{E}\left[ \sup_{0 \leq t \leq T} \|e(t)\|^p \right] \leq 2^p A.$$

Now, we can focus on the supremum of the error before time $\rho$. We first apply the Cauchy-Schwarz inequality

$$\left\| e\big(\min(t,\rho)\big) \right\|^2 = \left\| \int_0^{\min(t,\rho)} \big(a(x_s,s) - h(y_s)a(y_s,s)\big)\, ds \right.$$

$$\left. + \int_0^{\min(t,\rho)} \big(b(x_s,s) - b(y_s,s)\big)\, dB(s) \right\|^2$$

$$\leq 2\bigg[ T \int_0^{\min(t,\rho)} \left\| a(x_s,s) - h(y_s)a(y_s,s) \right\|^2 ds$$

$$+ \left\| \int_0^{\min(t,\rho)} \big(b(x_s,s) - b(y_s,s)\big)\, dB(s) \right\|^2 \bigg].$$

Then, we use the local Lipschitz and the boundedness assumptions, together with the Doob inequality (the first inequality) to deal with the stochastic integral: for any $u > 0$,

$$\mathbb{E}\bigg[ \sup_{0 \leq t \leq u} \left\| e\big(\min(t,\rho)\big) \right\|^2 \bigg]$$

$$\leq 2\mathbb{E}\bigg[ T \int_0^{\min(u,\rho)} \left\| a(x_s,s) - h(y_s)a(y_s,s) \right\|^2 ds$$

$$+ 4 \int_0^{\min(u,\rho)} \left\| \sigma(x_s,s) - \sigma(y_s,s) \right\|^2 ds \bigg]$$

$$\leq 2\mathbb{E}\bigg[ TC_R \int_0^{\min(u,\rho)} \|x_s - y_s\|^2 ds + T^2 K_R^2 \xi^2 + 4C_R \int_0^{\min(u,\rho)} \|x_s - y_s\|^2 ds \bigg]$$

$$\leq 2C_R(T+4)\mathbb{E}\bigg[ \int_0^{\min(u,\rho)} \sup_{0 \leq r \leq s} \big\{ \|x_{\min(r,\rho)} - y_{\min(r,\rho)}\|^2 \big\}\, ds \bigg] + 2T^2 K_R^2 \xi^2.$$

We then apply the Gronwall lemma

$$\mathbb{E}\bigg[ \sup_{0 \leq t \leq T} \left\| e\big(\min(t,\rho)\big) \right\|^2 \bigg] \leq 2T^2 K_R^2 \xi^2 e^{2C_R(T+4)}. \tag{29}$$

Finally, we choose $d$ small enough such that

$$\frac{2^{p+1} d A}{p} \leq \xi^2,$$

and $R$ large enough such that

$$\frac{2A(p-2)}{R^p p d^{2/(p-2)}} \leq \xi^2$$

yielding

$$\mathbb{E}\left[\sup_{0 \le t \le \min(T, \tau_{\mathcal{H}})} \|x_t - y_t\|^2\right] \le \left(2 + 2T^2 K_R^2 e^{2C_R(T+4)}\right)\xi^2.$$

□

## Appendix B: Proofs of Section 3

### B.1 Notations and definitions

Throughout the paper, lower-case normal letters are constants, lower-case bold letters are vectors or vector-valued functions, and upper-case bold letters are matrices.

- $l, \kappa, \tau, \epsilon_1, \epsilon_2, \mu, \sigma^2, \beta, \gamma, a_\pm \in \mathbb{R}_+$ are parameters of the network. We also define $\Delta = \sqrt{1 - 4\frac{l}{\beta}}$ for Section 3.3 and $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$, a fixed noise matrix, for Section 3.4. We write $s^2 = \||\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}'\||$.
- $n \in \mathbb{N}$ is the number of neurons in the network.
- $\mathbf{v} \in C^1(\mathbb{R}_+, \mathbb{R}^n)$ is the field of membrane potential in the network.
- $\mathbf{u} \in C^1(\mathbb{R}_+, \mathbb{R}^n)$ is the field of inputs to the network. We write

$$u_m = \sup_{t \in \mathbb{R}_+} \|\mathbf{u}(t)\|_2.$$

- $\mathbf{v} \otimes \mathbf{u} \in C^1(\mathbb{R}_+, \mathbb{R}^{n \times n})$ is the tensor product between $\mathbf{u}$ and $\mathbf{v}$, which simply means $\{\mathbf{u} \otimes \mathbf{v}\}_{ij}(t) = \mathbf{u}_i(t)\mathbf{v}_j(t)$.
- $\mathbf{W} \in C^1(\mathbb{R}_+, \mathbb{R}^{n \times n})$ is the connectivity of the network. Throughout the paper, we assume $\mathbf{W}(0) = 0$.
- $\langle \mathbf{x}, \mathbf{y} \rangle$ is the scalar product between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
- $\|\mathbf{u}(t)\|_p$ for $p = 1, 2$ is the $L^p$ norm of $\mathbf{u}(t) \in \mathbb{R}^n$, *i.e.*, $\|\mathbf{u}(t)\|_p = (\sum_{i=1}^n |\mathbf{u}_i(t)|^p)^{\frac{1}{p}}$. And similarly for the connectivity matrices of $\mathbb{R}^{n \times n}$ with a double sum.
- $\||\mathbf{W}\|| = \sup_{\mathbf{x} \in \mathbb{C}^n, \|\mathbf{x}\| = 1} |\langle \mathbf{x}, \mathbf{W} \cdot \mathbf{x}\rangle| = \max_{i \in \{1,...,n\}}\{|\lambda_i| : \lambda_i \text{ is an eigenvalue of } \mathbf{W}\}$.
- $\mathbf{J}'$ is the transpose of the matrix $\mathbf{J} \in \mathbb{R}^{n \times n}$.
- $\mathbf{x} \cdot \mathbf{y}' \in \mathbb{R}^{n \times n}$ is the cross-correlation matrix of two compactly supported and differentiable functions from $\mathbb{R}$ to $\mathbb{R}^n$, *i.e.*,

$$\{\mathbf{x} \cdot \mathbf{y}'\}_{ij} = \int_{-\infty}^{+\infty} \mathbf{x}_i(t)\mathbf{y}_j(t)\, dt.$$

- $H$ is the Heaviside function, *i.e.*,

$$H(t) = \begin{cases} 0 & \text{if } t \le 0, \\ 1 & \text{if } t > 0. \end{cases}$$

- The real functions

$$g_\gamma : t \mapsto \gamma e^{-\gamma t} H(t),$$

$$v : t \mapsto \frac{l}{\mu \Delta} \left( e^{-\frac{\beta}{2\mu}(1-\Delta)t} - e^{-\frac{\beta}{2\mu}(1+\Delta)t} \right) H(t), \tag{30}$$

$$w : t \mapsto \frac{l}{2\mu \Delta} \left( (1+\Delta) e^{-\frac{\beta}{2\mu}(1-\Delta)t} - (1-\Delta) e^{-\frac{\beta}{2\mu}(1+\Delta)t} \right) H(t)$$

are integrable on $\mathbb{R}$.

### B.1.1 Notations for the Appendix

The computations involve a lot of convolutions and, for readability of the Appendix, we introduce some new notations. Indeed, we rewrite the time-convolution between **u** and $g$, an integrable function on $\mathbb{R}$,

$$\mathbf{u} * g = \mathbf{u} \cdot \mathcal{G}.$$

This suggests one should think of **v** as a semi-continuous matrix of $\mathbb{R}^{n \times \mathbb{R}}$ and of $\mathcal{G}_\gamma$ as a continuous matrix of $\mathbb{R}^{\mathbb{R} \times \mathbb{R}}$, such that $\mathbf{u}_{it} = u_i(t)$ and $\mathcal{G}_{st} = g(t-s)$. Indeed, in this framework the convolution with $g$ is nothing but the continuous matrix multiplication between **v** and a continuous Toeplitz matrix generated row by row by $g$. Hence, the operator '·' can be though of as a matrix multiplication.

Therefore, it is natural to define $(\mathbf{u} * g)' = (\mathbf{u} \cdot \mathcal{G})' = \mathcal{G}' \cdot \mathbf{u}'$, where $\mathcal{G}' \in \mathbb{R}^{\mathbb{R} \times \mathbb{R}}$ is the transpose of $\mathcal{G}$, *i.e.*, the continuous Toeplitz matrix generated row by row by $g(-\cdot) : t \mapsto g(-t)$ and $\mathbf{u}' \in \mathbb{R}^{\mathbb{R} \times n}$. Thus, for $g$ and $h$, two integrable functions on $\mathbb{R}$, we can rewrite

$$(\mathbf{x} * g) \cdot (\mathbf{y} * h)' = \mathbf{x} \cdot \mathcal{G} \cdot \mathcal{H}' \cdot \mathbf{y}',$$

where $\mathcal{G}$ and $\mathcal{H}$ are their associated continuous matrices. More generally, the bold curved letters $\mathcal{G}, \mathcal{V}, \mathcal{W}$ represent these continuous Toeplitz matrices which are well defined through their action as convolution operators with $g$, $v$, and $w$. The previous formulation naturally expresses the symmetry of relation (14).

### B.2 Hebbian learning with linear activity

In this part, we consider system (12).

### B.2.1 Application of temporal averaging theory

**Theorem B.1** *If Assumption 3.1 is verified for $p \in ]0, 1[$, then system (12) is asymptotically well posed in probability and the connectivity matrix $\mathbf{W}^\epsilon$, the solution of system (12), converges to $\mathbf{W}$, in the sense that for all $\delta, T > 0$,*

$$\overset{\mu}{\underset{\epsilon \to 0}{\lim}} \mathbb{P} \left[ \sup_{t \in [0,T]} \left| \mathbf{W}_t^\epsilon - \mathbf{W}_t \right|^2 > \delta \right] = 0,$$

*where* $\mathbf{W}$ *is the deterministic solution of*

$$\frac{d\mathbf{W}_{ij}}{dt} = \bar{G}(\mathbf{W})_{ij} = \underbrace{-\kappa \mathbf{W}_{ij}}_{decay} + \underbrace{\frac{\mu}{\tau} \int_0^{\frac{\tau}{\mu}} \bar{\mathbf{v}}_i(s)\bar{\mathbf{v}}_j(s)\,ds}_{correlation} + \underbrace{\frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})^{-1}_{ij}}_{noise},$$

*where* $\bar{\mathbf{v}}(t)$ *is the* $\frac{\tau}{\mu}$-*periodic attractor of* $\frac{d\bar{\mathbf{v}}}{dt} = (\mathbf{W} - \mathbf{L}) \cdot \bar{\mathbf{v}} + \mathbf{u}(\mu t)$, *where* $\mathbf{W} \in \mathbb{R}^{n \times n}$ *is supposed to be fixed.*

*Proof* We are going to apply Property 2.3. For $p \in\ ]0, 1[$, consider the space

$$E_p = \left\{ \mathbf{W} \in \mathbb{R}^{n \times n} : \mathbf{W} \text{ is symmetric}, \mathbf{W} \geq 0 \text{ and } \|\|\mathbf{W}\|\| < lp \right\}.$$

First, since $\mathbf{L} - \mathbf{W}$ is strictly positive for $\mathbf{W}$ in $E_p$, Assumptions 2.1-2.2 are satisfied on $\mathbb{R}^n \times E_p$. Then, we only need to compute the averaged vector field $\bar{G}$ and show that $E_p$ is invariant under the flow of $\bar{G}$.

1. Computation of the averaged vector field $\bar{G}$:

The fast variable is linear, the averaged vector field is given by (10). This reads

$$\bar{G}(\mathbf{W}) = \left(\frac{\tau}{\mu}\right)^{-1} \int_0^{\frac{\tau}{\mu}} \int_{\mathbf{x} \in \mathbb{R}^n} G(\bar{\mathbf{v}}(t) + \mathbf{x}, \mathbf{W}) \mathcal{N}_{0,\mathbf{Q}}(d\mathbf{x})\,dt,$$

where $\mathcal{N}_{\mathbf{v},\mathbf{Q}}$ is the probability density function of the Gaussian law with mean $\mathbf{v}$ and covariance $\mathbf{Q}$. And $\mathbf{Q}$ is the unique solution of (9), with $\mathbf{\Sigma} = \sigma Id$. This leads to $\mathbf{Q} = \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})^{-1}$.

Therefore,

$$\bar{G}(\mathbf{W}) = -\kappa \mathbf{W} + \frac{\mu}{\tau} \int_0^{\frac{\tau}{\mu}} \left( \int_{\mathbf{x} \in \mathbb{R}^n} (\bar{\mathbf{v}}(t) + \mathbf{x}) \otimes (\bar{\mathbf{v}}(t) + \mathbf{x}) \mathcal{N}_{0,\mathbf{Q}}(d\mathbf{x}) \right) dt$$

$$= -\kappa \mathbf{W} + \frac{\mu}{\tau} \int_0^{\frac{\tau}{\mu}} \bar{\mathbf{v}}(t) \otimes \bar{\mathbf{v}}(t)\,dt + \frac{\mu}{\tau} \int_0^{\frac{\tau}{\mu}} \left( \bar{\mathbf{v}}(t) \otimes \underbrace{\int_{\mathbf{x} \in \mathbb{R}^n} \mathbf{x} \mathcal{N}_{0,\mathbf{Q}}(d\mathbf{x})}_{\text{Expectation of } \mathcal{N}(0,Q)=0} \right) dt$$

$$+ \frac{\mu}{\tau} \int_0^{\frac{\tau}{\mu}} \left( \underbrace{\int_{\mathbf{x} \in \mathbb{R}^n} \mathbf{x} \mathcal{N}_{0,\mathbf{Q}}(d\mathbf{x})}_{\text{Expectation of } \mathcal{N}(0,Q)=0} \otimes \bar{\mathbf{v}}(t) \right) dt$$

$$+ \frac{\mu}{\tau} \int_0^{\frac{\tau}{\mu}} \left( \underbrace{\int_{\mathbf{x} \in \mathbb{R}^n} \mathbf{x} \otimes \mathbf{x} \mathcal{N}_{0,\mathbf{Q}}(d\mathbf{x})}_{\text{Covariance of } \mathcal{N}(0,Q)=Q} \right) dt$$

$$= -\kappa \mathbf{W} + \frac{\mu}{\tau} \int_0^{\frac{\tau}{\mu}} \bar{\mathbf{v}}(t) \otimes \bar{\mathbf{v}}(t)\,dt + \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})^{-1}.$$

The integral term in the equation above is the correlation matrix of the $\frac{\tau}{\mu}$-periodic function $\bar{\bar{\mathbf{v}}}$. To rewrite this term, we define $\bar{\mathbf{v}} \in \mathbb{R}^{n \times [0, \frac{\tau}{\mu}[}$ such that $\bar{\mathbf{v}}(i, t) = \bar{\mathbf{v}}(t)_i$. $\bar{\mathbf{v}}$

can be seen as a matrix gathering the history of $\bar{\mathbf{v}}$, *i.e.*, each column of $\bar{\mathbf{v}}$ corresponds to the vector $\bar{\mathbf{v}}(t)$ for a given $t \in [0, \frac{\tau}{\mu}[$. It turns out

$$\int_0^{\frac{\tau}{\mu}} \bar{\mathbf{v}}(t) \otimes \bar{\mathbf{v}}(t)\, dt = \bar{\mathbf{v}} \cdot \bar{\mathbf{v}}'.$$

Therefore,

$$\bar{G}(\mathbf{W}) = -\kappa \mathbf{W} + \frac{\mu}{\tau} \bar{\mathbf{v}} \cdot \bar{\mathbf{v}}' + \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})^{-1}.$$

According to the results in Section 2, the solutions of a differential system with such a right-hand side are close to that of the initial system (12). Hence, we focus exclusively on it and try to unveil the properties of its solutions which will be retrospectively extended to those of the initial system (12).

2. Invariance of $E_p$ under the flow of (13):

Here we assume that $\mathbf{W}(0) \in E_p$ and we want to prove that the trajectory of $\mathbf{W}$ is in $E_p$, too.

(a) Symmetry:

It is clear that each term in $\bar{G}$ is symmetric. Their sum is therefore symmetric and so is $\mathbf{W}(t)$.

(b) Inequality $\mathbf{W} \geq 0$:

The correlation term $\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}'$ is a Gramian matrix and is therefore positive. Because $\mathbf{L} - \mathbf{W}$ is assumed to be positive, therefore, its inverse is also positive. Therefore, if $\mathbf{e}_i$ is an eigenvector of $\mathbf{W} \geq 0$ associated with a null eigenvalue, then $\mathbf{e}_i' \cdot \bar{G}(\mathbf{W}) \cdot \mathbf{e}_i \geq 0$. Thus, the trajectories of (13) remain positive.

(c) Inequality $\|\|\mathbf{W}\|\| < lp$:

The argument here is that of the inward pointing subspace. We intend to find a condition under which the flow $\bar{G}$ is pointing inward the space $\{\mathbf{W} : \|\|\mathbf{W}\|\| < lp\}$. Roughly speaking, this will be done by defining a real valued function $g$ strictly negative on the subspace and positive outside and then showing that its gradient (or differential) on the border goes in the opposite direction of the flow, *i.e.*, $d_{\mathbf{W}} g(\bar{G}(\mathbf{W})) < 0$ for $\mathbf{W} \in g^{-1}(0)$.

For all $\mathbf{x} \in \mathbb{C}^n$ such that $\|\mathbf{x}\| = 1$, define a family of positive numbers $(\alpha_{\mathbf{x}})$ whose supremum is written $\alpha^*$ and a family of functions $(g^{\mathbf{x}})$ such that

$$g^{\mathbf{x}} : \begin{array}{l} \mathbb{R}^{n \times n} \to \mathbb{R}, \\ \mathbf{J} \mapsto \|\mathbf{J} \cdot \mathbf{x}\|^2 - \alpha_{\mathbf{x}}^2. \end{array}$$

Observe that the differential of $g^{\mathbf{x}}$ at $\mathbf{W}$ applied to $\mathbf{J}$ is $dg^{\mathbf{x}}_{\mathbf{W}}(\mathbf{J}) = \frac{1}{2}\langle \mathbf{W} \cdot \mathbf{x}, \mathbf{J} \cdot \mathbf{x} \rangle$. For $\mathbf{W} \in g^{\mathbf{x}-1}(0)$, *i.e.*, $\|\mathbf{W} \cdot \mathbf{x}\| = \alpha_{\mathbf{x}}$, compute

$$2dg^{\mathbf{x}}_{\mathbf{W}}\big(\bar{G}(\mathbf{W})\big) = -\kappa \underbrace{\langle \mathbf{W} \cdot \mathbf{x}, \mathbf{W} \cdot \mathbf{x} \rangle}_{=\alpha_{\mathbf{x}}^2} + \frac{\mu}{\tau} \underbrace{\langle \mathbf{W} \cdot \mathbf{x}, \bar{\mathbf{v}} \cdot \bar{\mathbf{v}}' \cdot \mathbf{x} \rangle}_{=A}$$

$$+ \underbrace{\left\langle \mathbf{W} \cdot \mathbf{x}, \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})^{-1} \cdot \mathbf{x} \right\rangle}_{=B}.$$

- Upper bound of $A$:

    Applying Cauchy-Schwarz leads to

$$|A| \leq \|\mathbf{W} \cdot \mathbf{x}\| \|\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}' \cdot \mathbf{x}\| \leq \alpha_{\mathbf{x}} \int_0^{\frac{\tau}{\mu}} \|\bar{\mathbf{v}}(s) \otimes \bar{\mathbf{v}}(s) \cdot \mathbf{x}\| \, ds$$

$$\leq \alpha_{\mathbf{x}} \int_0^{\frac{\tau}{\mu}} |\langle \bar{\mathbf{v}}(s), \mathbf{x} \rangle| \|\bar{\mathbf{v}}(s)\| \, ds \leq \alpha_{\mathbf{x}} \int_0^{\frac{\tau}{\mu}} \|\bar{\mathbf{v}}(s)\|^2 \, ds.$$

    However, for $t \geq 0$

$$\|\bar{\mathbf{v}}(t)\| \leq \int_{-\infty}^t \|e^{(\mathbf{W}-\mathbf{L})(t-s)} \cdot \mathbf{u}(\mu s)\| \, ds \leq u_m \int_{-\infty}^t e^{(\alpha^*-l)(t-s)} \, ds$$

$$\leq u_m e^{(\alpha^*-l)t} \left[ \frac{e^{-(\alpha^*-l)s}}{l-\alpha^*} \right]_{-\infty}^t = \frac{u_m}{l-\alpha^*}.$$

    Therefore, $A \leq \frac{\alpha_{\mathbf{x}} \tau u_m^2}{\mu(l-\alpha)^2}$.

- Upper bound of $B$:

    Observe that for $\mathbf{J}$ a positive definite matrix whose eigenvalues are the $\lambda_i$, then the spectrum of $\mathbf{J}^{-1}$ is $\{\frac{1}{\lambda_i}\}$. Therefore, $\|\|\mathbf{J}^{-1}\|\| = \frac{1}{\min(\lambda_i)}$. Therefore, if $\mathbf{J} = \mathbf{L} - \mathbf{W}$, then $\|\|\mathbf{J}^{-1}\|\| \leq \frac{1}{l-\alpha^*}$.

    Using the previous observation and Cauchy-Schwarz leads to

$$|B| \leq \alpha_{\mathbf{x}} \frac{\sigma^2}{2} \|\|(\mathbf{L}-\mathbf{W})^{-1}\|\| \leq \frac{\alpha_{\mathbf{x}} \sigma^2}{2(l-\alpha^*)}.$$

Therefore, for $\alpha^* < l$

$$\frac{2 d g_{\mathbf{W}}^{\mathbf{x}}(\bar{G}(\mathbf{W}))}{\alpha_{\mathbf{x}}} \leq -\kappa \alpha_{\mathbf{x}} + \frac{u_m^2}{(l-\alpha^*)^2} + \frac{\sigma^2}{2(l-\alpha^*)}$$

$$= \frac{1}{(l-\alpha^*)^2} P(\alpha^*) + \kappa(\alpha^* - \alpha_{\mathbf{x}}),$$

where

$$P(\alpha) = -\kappa \alpha^3 + 2\kappa l \alpha^2 - \left(\kappa l^2 + \frac{\sigma^2}{2}\right) \alpha + \left(u_m^2 + \frac{l\sigma^2}{2}\right). \qquad (31)$$

Now write $\alpha^* = pl$ with $p \in \, ]0, 1[$. Equation (31) becomes

$$P(p) = -\kappa l^3 p(1-p)^2 + \frac{l\sigma^2}{2}(1-p) + u_m^2.$$

When there exists $p$ such that $P(p) < 0$ (which corresponds to Assumption 3.1), then their exists a ball of radius $pl$ on which the dynamics is pointing inward. It means any matrix $\mathbf{W}$ whose maximal eigenvalue is $\alpha^* = pl$ will see this eigenvalue (and those which are sufficiently close to it, *i.e.*, for which $\alpha^* - \alpha_{\mathbf{x}} > 0$ is

sufficiently small) decreasing along the trajectories of the system. Therefore, the space $E_p$ is invariant by the flow of the system iff Assumption 3.1 is satisfied.

The trajectories of system (13) with the initial condition in $E_p$ are defined on $\mathbb{R}_+$ and remain bounded. Indeed, if $\mathbf{W}(0) \in E_p$, the connectivity will stay in $E_p$, in particular $0 < \mathbf{L} - \mathbf{W} \leq \mathbf{L}$ along the trajectories, more precisely $\mathbf{L} - \mathbf{W}$ is a strictly positive constant since $p \in ]0, 1[$. Because $\bar{\mathbf{v}}$ is also bounded by $\frac{u_m}{l(1-p)}$, $\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}' + \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})^{-1}$ is bounded. The right-hand side of system (13) is the sum a bounded term and a linear term multiplied by a negative constant; therefore, the system remains bounded and it cannot explode in finite time: it is defined on $\mathbb{R}_+$. $\qquad\square$

### B.2.2 An expansion for the correlation term

We first state a useful lemma.

**Lemma B.2** *If $\bar{\mathbf{v}}$ is the solution, with zero as initial condition, of $\frac{d\bar{\mathbf{v}}}{dt} = (\mathbf{W} - \mathbf{L}) \cdot \bar{\mathbf{v}} + \mathbf{u}(t)$, it can be written by the sum below which converges if $\mathbf{W}$ is in $E_p$ for $p \in ]0, 1[$.*

$$\bar{\mathbf{v}} = \sum_{k=0}^{+\infty} \frac{\mathbf{W}^k}{l^{k+1}} \cdot \mathbf{u} * g_l^{(k+1)},$$

*where $g_l : t \mapsto l e^{-lt} H(t)$.*

*Proof* It can be proven as a trivial rewriting of the variation of parameters formula for linear systems. A more general approach, which extends to delayed systems, was developed by Galtier and Touboul [25]; see the first example for the proof of this lemma. $\qquad\square$

This is useful to find the next result.

**Property B.3** *The correlation term can be written*

$$\frac{\mu}{\tau}\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}' = \frac{u_m^2}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{W}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q}.$$

*Proof* We can use Lemma 3.2 with $\mu \neq 1$ and compute the cross product $\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}'$.

Therefore, consider $\mathbf{u}(\mu\cdot) : t \mapsto \mathbf{u}(\mu t)$ instead of $\mathbf{u}$. A change of variable shows that $(\mathbf{u}(\mu\cdot) * g_l^{(k)})(t) = \frac{1}{\mu}(\mathbf{u} * g_l^{(k)}(\frac{\cdot}{\mu}))(\mu t)$. Therefore,

$$\frac{\mu}{\tau}\{\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}'\}_{ij} = \frac{\mu}{\tau} \int_0^{\frac{\tau}{\mu}} \bar{\mathbf{v}}_i(t)\bar{\mathbf{v}}_j(t)\, dt = \frac{1}{\tau} \int_0^{\tau} \bar{\mathbf{v}}_i\left(\frac{s}{\mu}\right)\bar{\mathbf{v}}_j\left(\frac{s}{\mu}\right) ds$$

$$= \frac{1}{\tau} \int_0^{\tau} \left(\sum_{k=0}^{+\infty} \frac{\mathbf{W}^k}{l^{k+1}} \cdot \left(\mathbf{u}(\mu\cdot) * g_l^{(k+1)}\right)\left(\frac{s}{\mu}\right)\right)_i$$

$$\times \left( \sum_{q=0}^{+\infty} \frac{\mathbf{W}^q}{l^{q+1}} \cdot \big(\mathbf{u}(\mu \cdot) * g_l^{(q+1)}\big)\left(\frac{s}{\mu}\right) \right)_j \, ds$$

$$= \frac{1}{\tau} \int_0^\tau \left( \sum_{k=0}^{+\infty} \frac{\mathbf{W}^k}{l^{k+1}} \cdot \left( \mathbf{u} * \frac{g_l^{(k+1)}(\cdot/\mu)}{\mu} \right)(s) \right)_i$$

$$\times \left( \sum_{q=0}^{+\infty} \frac{\mathbf{W}^q}{l^{q+1}} \cdot \left( \mathbf{u} * \frac{g_l^{(q+1)}(\cdot/\mu)}{\mu} \right)(s) \right)_j \, ds$$

$$= \left\{ \frac{u_m^2}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{W}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q} \right\}_{ij} . \qquad \Box$$

### B.2.3 Global stability of the single equilibrium point

**Theorem B.4** *If Assumption* 3.1 *is verified for* $p \le \frac{1}{3}$, *then there is a unique equilibrium point in the invariant subset* $E_p$ *which is globally, asymptotically stable.*

*Proof* For this proof, define $F(\mathbf{W}) = \frac{u_m^2}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{W}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q} + \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})^{-1}$.

First, we compute the differential of $F$ and show it is a bounded operator. Second, we show it implies the existence and uniqueness of an equilibrium point under some condition. Then, we find an energy for the system which says the fixed point is a global attractor. Finally, we show the stability condition is the same as Assumption 3.1 for $p \le \frac{1}{3}$.

1. We compute the differential of each term in $F$:

- Formally write the second term $\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}'(\mathbf{W}) = \sum_{k,q=0}^{+\infty} \frac{\mathbf{W}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q}$. To find its differential, compute $\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}'(\mathbf{W} + \mathbf{J}) - \bar{\mathbf{v}} \cdot \bar{\mathbf{v}}'(\mathbf{W})$ and keep the terms at the first order in $\mathbf{J}$. Before computing the whole sum, observe that

$$(\mathbf{W} + \mathbf{J})^k \cdot \mathbf{C}^{k,q} \cdot (\mathbf{W} + \mathbf{J})'^q - \mathbf{W}^k \cdot \mathbf{C}^{k,q} \cdot \mathbf{W}'^q$$

$$= \sum_{m=0}^{k-1} \mathbf{W}^m \cdot \mathbf{J} \cdot \mathbf{W}^{k-1-m} \cdot \mathbf{C}^{k,q} \cdot \mathbf{W}'^q$$

$$+ \sum_{m=0}^{q-1} \mathbf{W}^k \cdot \mathbf{C}^{k,q} \cdot \mathbf{W}'^m \cdot \mathbf{J}' \cdot \mathbf{W}'^{q-1-m} + \mathcal{O}\big(\|\mathbf{J}\|^2\big).$$

This leads to

$$d\bar{\mathbf{v}} \cdot \bar{\mathbf{v}}'_{\mathbf{W}}(\mathbf{J}) = \frac{1}{l} \sum_{k,q=0}^{+\infty} \left( \sum_{m=0}^{k-1} \frac{\mathbf{W}^m}{l^m} \cdot \mathbf{J} \cdot \frac{\mathbf{W}^{k-1-m}}{l^{k-1-m}} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q} \right.$$

$$\left. + \sum_{l=0}^{q-1} \frac{\mathbf{W}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{W}'^m}{l^m} \cdot \mathbf{J}' \cdot \frac{\mathbf{W}'^{q-1-m}}{l^{q-1-m}} \right).$$

- Write $Q : \mathbf{W} \mapsto (\mathbf{L} - \mathbf{W})^{-1}$. We can write $(\mathbf{L} - \mathbf{W}) \cdot Q(\mathbf{W}) = Id$ and use the chain rule to compute the differential of $Q$ at $\mathbf{W}$, which gives $-\mathbf{J} \cdot Q(\mathbf{W}) + (\mathbf{L} - \mathbf{W}) \cdot dQ_{\mathbf{W}}(\mathbf{J}) = 0$. Therefore,

$$dQ_{\mathbf{W}}(\mathbf{J}) = (\mathbf{L} - \mathbf{W})^{-1} \cdot \mathbf{J} \cdot (\mathbf{L} - \mathbf{W})^{-1}.$$

The differential of $F$ at $\mathbf{W}$ is the sum of these two terms.

2. We want to compute the norm of $\|dF_{\mathbf{W}}(\mathbf{J})\|_2$ for $\|\mathbf{J}\|_2 = 1$. First, observe that for three square matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$,

$$\|\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}\|_2^2 = \sum_{i,j=1}^{n} B_{ij}^2 \|\mathbf{A} \cdot (\mathbf{e}_i \otimes \mathbf{e}_j) \cdot \mathbf{C}\|_2^2 \le \sum_{i,j=1}^{n} B_{ij}^2 \|\mathbf{A} \cdot \mathbf{e}_i\|_2^2 \|\mathbf{C} \cdot \mathbf{e}_j\|_2^2$$

$$\le \sum_{i,j=1}^{n} B_{ij}^2 \|\|\mathbf{A}\|\|^2 \|\|\mathbf{C}\|\|^2,$$

for $\mathbf{e}_i$ the vectors of the canonical basis of $\mathbb{R}^n$. This leads to $\|\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}\|_2 \le \|\mathbf{B}\|_2 \|\|\mathbf{A}\|\| \|\|\mathbf{C}\|\|$. Therefore, because $\|\|\mathbf{A}\|\| \le \|\mathbf{A}\|_2$,

$$\left\| \frac{\mathbf{W}^m}{l^m} \cdot \mathbf{J} \cdot \frac{\mathbf{W}^{k-1-m}}{l^{k-1-m}} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q} \right\|_2 \le \frac{\|\|\mathbf{W}\|\|^m}{l^m} \left\| \frac{\mathbf{W}^{k-1-m}}{l^{k-1-m}} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q} \right\|_2$$

$$\le u_m^2 \frac{\|\|\mathbf{W}\|\|^{k-1}}{l^{k-1}} \frac{\|\|\mathbf{W}\|\|^q}{l^q}.$$

Therefore,

$$\|dF_{\mathbf{W}}(\mathbf{J})\|_2 \le \frac{u_m^2}{l^3} \sum_{k,q=0}^{+\infty} \left( k \frac{\|\|\mathbf{W}\|\|^{k-1}}{l^{k-1}} \frac{\|\|\mathbf{W}\|\|^q}{l^q} + q \frac{\|\|\mathbf{W}\|\|^k}{l^k} \frac{\|\|\mathbf{W}\|\|^{q-1}}{l^{q-1}} \right)$$

$$+ \frac{\sigma^2}{2} \|\|(\mathbf{L} - \mathbf{W})^{-1}\|\|^2$$

$$\le \frac{2u_m^2}{l^3} \left( \sum_{k=0}^{+\infty} k p^{k-1} \right) \left( \sum_{q=0}^{+\infty} p^q \right) + \frac{\sigma^2}{2} \|\|(\mathbf{L} - \mathbf{W})^{-1}\|\|^2$$

$$\le \frac{2u_m^2}{l^3 (1-p)^3} + \frac{\sigma^2}{2l^2 (1-p)^2}.$$

This inequality is true for all $\mathbf{J}$ with $\|\mathbf{J}\|_2 = 1$; therefore, it is also true for the operator norm

$$\|\|dF_{\mathbf{W}}\|\| \le \frac{2u_m^2}{l^3 (1-p)^3} + \frac{\sigma^2}{2l^2 (1-p)^2}.$$

Therefore, $F$ is a $k$-Lipschitz operator where $k = \frac{2u_m^2}{l^3(1-p)^3} + \frac{\sigma^2}{2l^2(1-p)^2}$. This means $\|F(\mathbf{W}) - F(\mathbf{J})\|_2 \le k \|\mathbf{W} - \mathbf{J}\|_2$.

3. The equilibrium points of system (15) necessarily verify the equation $\mathbf{W} = \frac{1}{\kappa}F(\mathbf{W})$. If

$$\frac{2u_m^2}{(1-p)^3} + \frac{l\sigma^2}{2(1-p)^2} < \kappa l^3, \tag{32}$$

then $\frac{1}{\kappa}F$ is a contraction map from $E_p$ to itself. Therefore, the Banach fixed point theorem says that there is a unique fixed point which we write $\mathbf{W}^*$.

4. We now show that, under assumption (32), $\mathbf{W} \mapsto \|\mathbf{W} - \mathbf{W}^*\|_2^2$ is an energy function for the system $\frac{d\mathbf{W}}{dt'} = -\mathbf{W} + \frac{1}{\kappa}F(\mathbf{W})$ (which is a rescaled version of system (15)).

   Indeed, compute the derivative of this energy along the trajectories of the system

$$2\frac{d}{dt}\|\mathbf{W}(t) - \mathbf{W}^*\|_2^2 = \left\langle \mathbf{W} - \mathbf{W}^*, -\mathbf{W} + \frac{1}{\kappa}F(\mathbf{W})\right\rangle$$

$$= -\langle \mathbf{W} - \mathbf{W}^*, \mathbf{W} - \mathbf{W}^*\rangle + \left\langle \mathbf{W} - \mathbf{W}^*, \frac{1}{\kappa}F(\mathbf{W}) - \mathbf{W}^*\right\rangle$$

$$= -\|\mathbf{W} - \mathbf{W}^*\|_2^2 + \left\langle \mathbf{W} - \mathbf{W}^*, \frac{1}{\kappa}F(\mathbf{W}) - \frac{1}{\kappa}F(\mathbf{W}^*)\right\rangle$$

$$\leq -\|\mathbf{W} - \mathbf{W}^*\|_2^2 + \|\mathbf{W} - \mathbf{W}^*\|_2 \left\|\frac{1}{\kappa}F(\mathbf{W}) - \frac{1}{\kappa}F(\mathbf{W}^*)\right\|_2$$

$$\leq \frac{1}{\kappa l^3}\left(\frac{2u_m^2}{(1-p)^3} + \frac{l\sigma^2}{2(1-p)^2} - \kappa l^3\right)\|\mathbf{W} - \mathbf{W}^*\|_2^2 \leq 0.$$

   The energy is lower-bounded, takes its minimum for $\mathbf{W} = \mathbf{W}^*$ and the decreases along the trajectories of the system. Therefore, $\mathbf{W}^*$ is globally asymptotically stable if assumption (32) is verified.

5. Observe that if Assumption 3.1 is verified for $p \leq \frac{1}{3}$, then $\frac{1}{1-p} < \frac{2}{1-p} \leq \frac{1}{p}$. Therefore, Assumption 3.1 implies that (32) is also true. This concludes the proof. □

### B.2.4 Explicit expansion of the equilibrium point

Recall the notations $\tilde{p} = \frac{u_m^2}{\kappa l^3} + \frac{\sigma^2}{2\kappa l^2}$ and $\lambda = \frac{\sigma^2 l}{2u_m^2}$.

**Theorem B.5**

$$\mathbf{W} = \frac{\tilde{p}l}{1+\lambda}\left(\lambda + \mathbf{C}^{0,0}\right) + \frac{\tilde{p}^2 l}{(1+\lambda)^2}\left(\lambda^2 + \lambda\left(\mathbf{C}^{0,0} + \mathbf{C}^{1,0} + \mathbf{C}^{0,1}\right)\right.$$

$$\left. + \mathbf{C}^{0,0} \cdot \mathbf{C}^{1,0} + \mathbf{C}^{0,1} \cdot \mathbf{C}^{0,0}\right) + \mathcal{O}(\tilde{p}^3).$$

Actually, it is possible to compute recursively the $n$th term of the expansion above, although their complexity explodes.

*Proof* Define $p^*$ the smallest value in $]0, 1[$ such that Assumption 3.1 is valid. This implies

$$p^*\left((1-p^*)^2 + \frac{\sigma^2}{2\kappa l^2}\right) = \frac{u_m^2}{\kappa l^3} + \frac{\sigma^2}{2\kappa l^2}.$$

The weak connectivity index $\tilde{p}$ controls the ratio of the connection over the strength of intrinsic dynamics. Indeed, these two variables are of the same order because

$$\frac{p^*}{\tilde{p}} = \frac{1}{(1-p^*)^2 + \frac{\sigma^2}{2\kappa l^2}} = \mathcal{O}_{\tilde{p}\to 0}(1).$$

We want to approximate the equilibrium $\mathbf{W}^*$, *i.e.*, the solution of $\bar{G}(\mathbf{W}^*) = 0$, in the regime $\tilde{p} \ll 1$. Define $\mathbf{\Omega} = \frac{\mathbf{W}}{\tilde{p}l}$ such that $\|\|\mathbf{\Omega}\|\| = \mathcal{O}(1)$. We abusively write $\bar{G}(\mathbf{\Omega}) = \bar{G}(\tilde{p}l\mathbf{\Omega})$ such that

$$\bar{G}(\mathbf{\Omega}) = -\tilde{p}l\kappa\mathbf{\Omega} + \frac{u_m^2}{l^2}\sum_{k,q=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k \cdot \mathbf{C}^{k,q} \cdot (\tilde{p}\mathbf{\Omega})^q + \frac{\sigma^2}{2l}\sum_{k=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k.$$

Recalling $\lambda = \frac{\sigma^2 l}{2u_m^2}$ leads to

$$\bar{G}(\mathbf{\Omega}) = \left(\frac{u_m^2}{l^2} + \frac{\sigma^2}{2l}\right)\left(-\mathbf{\Omega} + \frac{1}{1+\lambda}\sum_{k,q=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k \cdot \mathbf{C}^{k,q} \cdot (\tilde{p}\mathbf{\Omega})^q + \frac{\lambda}{1+\lambda}\sum_{k=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k\right).$$

Now, we write a candidate $\mathbf{\Omega}^{(m)} = \sum_{a=0}^{m}\tilde{p}^a\mathbf{\Omega}_a$, then we chose the terms $\mathbf{\Omega}_a = \mathcal{O}(1)$ so that the first $m$th orders in $\bar{G}(\mathbf{\Omega}^{(m)})$ vanish. This implies that $\|\bar{G}(\mathbf{\Omega}^*) - \bar{G}(\mathbf{\Omega}^{(m)})\| = \mathcal{O}(\tilde{p}^{m+1})$, where $\mathbf{\Omega}^* = \frac{\mathbf{W}^*}{\tilde{p}l}$. Then, we use the fact that the minimal absolute value of the eigenvalues of $\bar{G}$ is larger than $\kappa - (\frac{2u_m^2}{l^3(1-p)^3} + \frac{\sigma^2}{2l^2(1-p)^2}) > 0$. Indeed, it means

$$\|\mathbf{W}^* - \mathbf{W}^{(m)}\| < \frac{1}{\kappa - (\frac{2u_m^2}{l^3(1-p)^3} + \frac{\sigma^2}{2l^2(1-p)^2})}\mathcal{O}(p^{m+1}) < \frac{1}{\kappa - (\frac{2u_m^2}{l^3} + \frac{\sigma^2}{2l^2})}\mathcal{O}(p^{m+1}),$$

*i.e.*, $\mathbf{\Omega}^{(m)} = \mathbf{\Omega}^* + \mathcal{O}(\tilde{p}^{m+1})$.

Thus, we need to find the $\mathbf{\Omega}_a$ such that the first $m$th orders in $\bar{G}(\mathbf{\Omega}^{(m)})$ vanish. Therefore, we need to expand all the terms in $\bar{G}(\mathbf{\Omega})$. The first term is obvious. In the following, we write the second term $F(\mathbf{\Omega})$ associated to the correlations and look for an explicit expression of the $\mathbf{F}_a$ such that $F(\mathbf{\Omega}) = \sum_{a=0}^{+\infty}\tilde{p}^a\mathbf{F}_a$. Second, we write the third term $Q(\mathbf{\Omega})$ associated to the noise and look for an explicit expression of the $\mathbf{Q}_a$ such that $Q(\mathbf{\Omega}) = \sum_{a=0}^{+\infty}\tilde{p}^a\mathbf{Q}_a$.

- Finding the $\mathbf{F}_a$:

  First, observe that

$$\mathbf{\Omega}^q = \sum_{i=0}^{+\infty}\tilde{p}^i\sum_{\eta\in\mathbb{N}^q,\sum_k\eta_k=i}\mathbf{\Omega}_{\eta_1} \cdot \mathbf{\Omega}_{\eta_2} \cdot \ldots \cdot \mathbf{\Omega}_{\eta_q}. \tag{33}$$

This leads to

$$F(\boldsymbol{\Omega}) = \frac{1}{1+\lambda} \sum_{\substack{k,q=0 \\ j \le i}}^{+\infty} \sum_{i,j=0}^{+\infty} \tilde{p}^{i+k+q} \sum_{\substack{\eta \in \mathbb{N}^j, \sum_n \eta_n = k \\ \theta \in \mathbb{N}^{i-j}, \sum_n \theta_n = q}} \boldsymbol{\Omega}_{\eta_1} \cdot \ldots \cdot \boldsymbol{\Omega}_{\eta_j}$$
$$\times \, \mathbf{C}^{j,i-j} \cdot \boldsymbol{\Omega}'_{\theta_1} \cdot \ldots \cdot \boldsymbol{\Omega}'_{\theta_{i-j}}.$$

The $a$th term in the power expansion in $\tilde{p}$ verifies $a = i + k + q$. More precisely, this reads

$$\mathbf{F}_a = \frac{1}{1+\lambda} \sum_{\substack{k,q,i=0 \\ a=i+k+q}}^{+\infty} \sum_{j=0}^{i} \sum_{\substack{\eta \in \mathbb{N}^j, \sum_n \eta_n = k \\ \theta \in \mathbb{N}^{i-j}, \sum_n \theta_n = q}} \boldsymbol{\Omega}_{\eta_1} \cdot \ldots \cdot \boldsymbol{\Omega}_{\eta_j} \cdot \mathbf{C}^{j,i-j} \cdot \boldsymbol{\Omega}'_{\theta_1} \cdot \ldots \cdot \boldsymbol{\Omega}'_{\theta_{i-j}}.$$

This equation is scary but it reduces to simple expressions for small $a \in \mathbb{N}$.

- Finding the $\mathbf{Q}_a$:

   Using equation (33) leads to

$$Q(\boldsymbol{\Omega}) = \frac{\lambda}{1+\lambda} \sum_{i,q=0}^{+\infty} \tilde{p}^{i+q} \sum_{\eta \in \mathbb{N}^q, \sum_k \eta_k = i} \boldsymbol{\Omega}_{\eta_1} \cdot \boldsymbol{\Omega}_{\eta_2} \cdot \ldots \cdot \boldsymbol{\Omega}_{\eta_q}.$$

The $a$th term in the power expansion in $\tilde{p}$ verifies $a = i + q$. More precisely, this reads

$$\mathbf{Q}_a = \frac{\lambda}{1+\lambda} \sum_{\substack{q,i=0 \\ a=i+q}}^{+\infty} \tilde{p}^{i+q} \sum_{\eta \in \mathbb{N}^q, \sum_k \eta_k = i} \boldsymbol{\Omega}_{\eta_1} \cdot \boldsymbol{\Omega}_{\eta_2} \cdot \ldots \cdot \boldsymbol{\Omega}_{\eta_q}.$$

Therefore,

| $a$ | $(1 + \frac{1}{\lambda})\mathbf{Q}_a$ | $(1+\lambda)\mathbf{F}_a$ |
|---|---|---|
| 0 | $Id$ | $\mathbf{C}^{0,0}$ |
| 1 | $\boldsymbol{\Omega}_0$ | $\boldsymbol{\Omega}_0 \cdot \mathbf{C}^{1,0} + \mathbf{C}^{0,1} \cdot \boldsymbol{\Omega}_0$ |
| 2 | $\boldsymbol{\Omega}_0^2 + \boldsymbol{\Omega}_1$ | $\boldsymbol{\Omega}_0^2 \cdot \mathbf{C}^{2,0} + \mathbf{C}^{0,2} \cdot \boldsymbol{\Omega}_0^2 + \boldsymbol{\Omega}_0 \cdot \mathbf{C}^{1,1} \cdot \boldsymbol{\Omega}_0 + \boldsymbol{\Omega}_1 \cdot \mathbf{C}^{1,0} + \mathbf{C}^{0,1} \cdot \boldsymbol{\Omega}_1$ |

Therefore, it is easy to compute $\boldsymbol{\Omega}_a = \mathbf{F}_a + \mathbf{Q}_a$ for $a \in \mathbb{N}$. By definition $\mathbf{W} = \tilde{p}l\boldsymbol{\Omega} = \tilde{p}l(\mathbf{F} + \mathbf{Q})$, which leads to the result. $\qquad\square$

### B.3 Trace learning with damped oscillators and dynamic synapses

**Theorem B.6** *If Assumption* 3.1 *is verified for* $p \in \,]0, 1[$, *then system* (16) *is asymptotically well posed in probability and the connectivity matrix* $\mathbf{W}^\epsilon$, *solution of system* (16), *converges to* $\mathbf{W}$ *in the sense that for all* $\delta, T > 0$,

$$\lim_{\epsilon \to 0} \overset{\mu}{\mathbb{P}} \Big[ \sup_{t \in [0,T]} \big| \mathbf{W}_t^\epsilon - \mathbf{W}_t \big|^2 > \delta \Big] = 0,$$

*where* $\mathbf{W}$ *is the deterministic solution of*

$$\frac{d\mathbf{W}_{ij}}{dt} = \bar{G}(\mathbf{W})_{ij} = \underbrace{-\kappa\mathbf{W}_{ij}}_{decay} + \underbrace{\frac{\mu}{\tau}\int_0^{\frac{\tau}{\mu}}(\bar{\mathbf{v}}_i * g_\beta)(s)(\bar{\mathbf{v}}_j * g_\beta)(s)\,ds}_{correlation} + \underbrace{\mathbf{Q}_{22}}_{noise},$$

*where* $\bar{\mathbf{v}}(t)$ *is the* $\frac{\tau}{\mu}$*-periodic attractor of* $\frac{d\bar{\mathbf{v}}}{dt} = (\mathbf{W} - \mathbf{L}) \cdot \bar{\mathbf{v}} * g_\beta + \mathbf{u}(\mu t)$, *where* $\mathbf{W} \in \mathbb{R}^{n \times n}$ *is supposed to be fixed. And* $\mathbf{Q}_{22}$ *is a noise matrix described below.*

*Proof* First, let us recall the instantaneous reformulation of (16)

$$\begin{cases} d\binom{\mathbf{v}^\epsilon}{\mathbf{z}^\epsilon} = \frac{1}{\epsilon_1}\left[\begin{pmatrix} 0 & \mathbf{W}-\mathbf{L} \\ \beta & -\beta \end{pmatrix}\binom{\mathbf{v}^\epsilon}{\mathbf{z}^\epsilon} + \binom{\mathbf{u}(\frac{t}{\epsilon_2})}{0}\right]dt + \binom{\frac{\sigma}{\sqrt{\epsilon_1}}dB(t)}{\frac{\sigma_\mathbf{z}}{\sqrt{\epsilon_1}}dB(t)}, \\ \frac{d\mathbf{W}^\epsilon}{dt} = -\kappa\mathbf{W}^\epsilon + \mathbf{z}^\epsilon \otimes \mathbf{z}^\epsilon. \end{cases}$$

Starting from this system, the structure of the proof of Theorem 3.1 remains unchanged. The correlation term is to be replaced by $\frac{\mu}{\tau}\bar{\mathbf{v}} \cdot \mathcal{G}_\beta \cdot \mathcal{G}'_\beta \cdot \bar{\mathbf{v}}'$. The noise term we are looking for is $\mathbf{Q}_{22}$ in the Lyapunov equation (see (9)) below

$$\begin{pmatrix} 0 & \mathbf{W}-\mathbf{L} \\ \beta & -\beta \end{pmatrix} \cdot \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}'_{12} \\ \mathbf{Q}_{12} & \mathbf{Q}_{22} \end{pmatrix} + \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}'_{12} \\ \mathbf{Q}_{12} & \mathbf{Q}_{22} \end{pmatrix} \cdot \begin{pmatrix} 0 & \beta \\ \mathbf{W}'-\mathbf{L} & -\beta \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma_\mathbf{z}^2 \end{pmatrix} = 0.$$

Because the learning rule is symmetric, then the space of symmetric matrices is invariant and we can restrict this section to the symmetric case. It is easy to show that this Lyapunov equation has a unique solution, because the sum of two eigenvalues of the drift matrix is never null (provided $\mathbf{W}$ stays in $E_p$). This leads to the system

$$\begin{cases} (\mathbf{W}-\mathbf{L}) \cdot \mathbf{Q}_{12} + \mathbf{Q}'_{12} \cdot (\mathbf{W}-\mathbf{L}) + \sigma^2 = 0, & \text{(a)} \\ \beta(\mathbf{Q}_{11} - \mathbf{Q}_{12}) + \mathbf{Q}_{22} \cdot (\mathbf{W}-\mathbf{L}) = 0, & \text{(b)} \\ \mathbf{Q}_{22} = \frac{\mathbf{Q}_{12}+\mathbf{Q}'_{12}}{2} + \frac{\sigma_\mathbf{z}^2}{2\beta}I_d. & \text{(c)} \end{cases}$$

One solution of equation (a) is $\mathbf{Q}_{12} = \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{W})^{-1}$. Equation (c) defines $\mathbf{Q}_{22}$ properly. Indeed, because $\mathbf{W}$ is symmetric, so is $\mathbf{Q}_{12}$ and $\mathbf{Q}_{22} = \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{W})^{-1} + \frac{\sigma_\mathbf{z}^2}{2\beta}I_d$. Similarly, equation (b) defines $\mathbf{Q}_{11}$ but it remains to be checked that this definition is that of a symmetric matrix. In fact, it works because $\mathbf{W}$ is assumed symmetric and the noise has no off-diagonal terms. Indeed, in this case, $\mathbf{Q}_{11} = \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{W})^{-1} + \frac{\sigma_\mathbf{z}^2}{2\beta} + \frac{\sigma_\mathbf{z}^2}{2\beta^2}(\mathbf{L}-\mathbf{W})$. This solution is thus the unique solution of the Lyapunov equation.

Therefore,

$$\bar{G}(\mathbf{W}) = -\kappa\mathbf{W} + \frac{\mu}{\tau}\bar{\mathbf{v}} \cdot \mathcal{G}_\beta \cdot \mathcal{G}'_\beta \cdot \bar{\mathbf{v}}' + \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{W})^{-1} + \frac{\sigma_\mathbf{z}^2}{2\beta}I_d.$$

Thus, this application of Theorem 2.2 to the instantaneous system with $\sigma_\mathbf{z} \neq 0$, leads to the previous averaged equation. To recover the initial case (16), we can let $\sigma_\mathbf{z} \to 0$.

We see that the function $\bar{G}$ tends to

$$\bar{G}^*(\mathbf{W}) = -\kappa\mathbf{W} + \frac{\mu}{\tau}\bar{\mathbf{v}} \cdot \mathcal{G}_\beta \cdot \mathcal{G}'_\beta \cdot \bar{\mathbf{v}}' + \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{W})^{-1},$$

which we will rewrite $\bar{G}$ for simplicity in the following. Thus, this definition of $\bar{G}$ defines the averaged system for the original equation (16).

In the derivation of the condition under which $\|\|\mathbf{W}\|\|$ remain smaller than $lp$, the upper bound of the term $A$ changes as follows. Define $M \in \mathbb{R}_+$ so that $\|\bar{\mathbf{v}}(t)\| \le M$ for all $t > 0$. Because we assume $\bar{\mathbf{v}}(\mathbb{R}_-) = 0$, the variation of parameters formula for linear retarded differential equations with constant coefficients (see Chapter 6 of [34]) reads $\bar{\mathbf{v}}(t) = \int_0^t \mathbf{U}(t-s) \cdot \mathbf{u}(\mu s)\, ds$ where the resolvent $\mathbf{U}$ is the solution of $\dot{\mathbf{U}} = (\mathbf{W} - \mathbf{L}) \cdot (\mathbf{U} * g)$. We use Corollary 1.1 of Chapter 6 of [34], which is based on Grönwall's lemma, to claim that $\|\mathbf{U}(t-s)\| \le e^{(t-s)(\alpha^*-l)}$. Therefore,

$$\left\|\bar{\mathbf{v}}(t)\right\| \le \int_{-\infty}^t \left\|\mathbf{U}(t-s)\right\| \left\|u(\mu s)\right\| ds \le u_m \left[\frac{e^{(t-s)(\alpha^*-l)}}{l-\alpha^*}\right]_{-\infty}^t \le \frac{u_m}{l-\alpha^*} = M.$$

Then, we used Young's inequality for convolution to get $\|(\bar{\mathbf{v}} * g)(t)\|_2 \le \|\bar{\mathbf{v}}\|_2 \|g\|_1 = \|\bar{\mathbf{v}}\|_2$.

Therefore, the upper bound of $A$ remains unchanged.

Therefore, the polynomial $P$ remains the same and Assumption 3.1 is still relevant to this problem. □

**Lemma B.7** *If $\bar{\mathbf{v}}$ is the solution, with zero as initial condition, of $\frac{d\bar{\mathbf{v}}}{dt} = (\mathbf{W} - \mathbf{L}) \cdot \bar{\mathbf{v}} * g_\beta + \mathbf{u}(t)$, it can be written by the sum below which converges if $\mathbf{W}$ is in $E_p$ for $p \in ]0, 1[$.*

$$\bar{\mathbf{v}} = \sum_{k=0}^{+\infty} \frac{\mathbf{W}^k}{l^{k+1}} \cdot \mathbf{u} \cdot \tilde{\mathcal{W}} \cdot \tilde{\mathcal{V}}^k,$$

*where $\tilde{\mathcal{W}}$ and $\tilde{\mathcal{V}}$ are convolution operators respectively generated by the functions $\tilde{w}$ and $\tilde{v}$ detailed below*

$$\tilde{w} : t \mapsto \frac{l}{2\Delta}\left((1+\Delta)e^{-\frac{\beta}{2}(1-\Delta)t} - (1-\Delta)e^{-\frac{\beta}{2}(1+\Delta)t}\right)H(t),$$

$$\tilde{v} : t \mapsto \frac{l}{\Delta}\left(e^{-\frac{\beta}{2}(1-\Delta)t} - e^{-\frac{\beta}{2}(1+\Delta)t}\right)H(t),$$

*where $H$ is the Heaviside function, $\Delta = \sqrt{1 - \frac{4l}{\beta}}$. If $\Delta$ is a pure imaginary number, the expression above still holds with the hyperbolic functions sh and ch being turned into classical trigonometric functions sin and cos and $\Delta$ being replaced by its modulus.*

*If $\mathbf{W}$ is in $E_p$ for $p \in ]0, 1[$, then this expansion converges.*

*Proof* See the second example of [25]. □

Using Lemma C.3, on can redefine

$$\tilde{\mathbf{C}}^{k,q} = \frac{1}{u_m^2 \tau \|v\|_1^{k+q+2}} \mathbf{u} \cdot \mathcal{V}^{k+1} \cdot \left(\mathbf{u} \cdot \mathcal{V}^{q+1}\right)',$$

where $\mathcal{V}$ is the convolution operator generated by $v(t) = \frac{l}{\mu\Delta}(e^{-\frac{\beta}{2\mu}(1-\Delta)t} - e^{-\frac{\beta}{2\mu}(1+\Delta)t})H(t)$ (see Appendix C for details). Observe that applying Young's inequality for convolutions leads to $\|\tilde{\mathbf{C}}^{k,q}\|_2 \le 1$.

Therefore, we can rewrite Theorem 3.3 into

**Theorem B.8**

$$\frac{\mu}{\tau}\bar{\mathbf{v}} \cdot \mathcal{G}_\beta \cdot \mathcal{G}'_\beta \cdot \bar{\mathbf{v}} = \frac{u_m^2\|v\|_1^2}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{W}^k}{(l/\|v\|_1)^k} \cdot \tilde{\mathbf{C}}^{k,q} \cdot \frac{\mathbf{W}'^q}{(l/\|v\|_1)^q}.$$

*Proof* Similar to that of Theorem 3.3. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

**Theorem B.9** *If Assumption 3.1 is verified for $p \le \frac{1}{2\|v\|_1^3+1}$, there is a unique equilibrium point which is globally, asymptotically stable.*

*Proof* Similar to the proof of Theorem B.4. $\qquad\qquad\qquad\qquad\qquad\square$

With the same definitions for $\tilde{p} = \frac{u_m^2}{\kappa l^3} + \frac{\sigma^2}{2\kappa l^2}$ and $\lambda = \frac{\sigma^2 l}{2u_m^2}$, we can show

**Theorem B.10** *The weakly connected expansion of the equilibrium point is*

$$\mathbf{W}^* = \frac{\tilde{p}l}{1+\lambda}\left(\lambda + \|v\|_1^2\tilde{\mathbf{C}}^{0,0}\right)$$

$$+ \frac{\tilde{p}^2\|v\|_1 l}{(1+\lambda)^2}\left(\frac{\lambda^2}{\|v\|_1} + \lambda\left(\|v\|_1\tilde{\mathbf{C}}^{0,0} + \|v\|_1^2\tilde{\mathbf{C}}^{1,0} + \|v\|_1^2\tilde{\mathbf{C}}^{0,1}\right)\right.$$

$$\left. + \|v\|_1^4\tilde{\mathbf{C}}^{0,0} \cdot \tilde{\mathbf{C}}^{1,0} + \|v\|_1^4\tilde{\mathbf{C}}^{0,1} \cdot \tilde{\mathbf{C}}^{0,0}\right) + \mathcal{O}\left(\tilde{p}^3\|v\|_1^2\right).$$

*Proof* Define $\mathbf{\Omega} = \frac{\mathbf{W}}{\tilde{p}l}$ so that

$$\bar{G}(\mathbf{\Omega}) = \left(\frac{u_m^2}{l^2} + \frac{\sigma^2}{2l}\right)\left(-\mathbf{\Omega} + \frac{\|v\|_1^2}{1+\lambda}\sum_{k,q=0}^{+\infty}\left(\tilde{p}\|v\|_1\mathbf{\Omega}\right)^k \cdot \tilde{\mathbf{C}}^{k,q} \cdot \left(\tilde{p}\|v\|_1\mathbf{\Omega}\right)^q\right.$$

$$\left. + \frac{\lambda}{1+\lambda}\sum_{k=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k\right).$$

So, the expansion will be in orders of $\tilde{p}\|v\|_1$ with $\|v\|_1 \ge 1$.

Therefore,

| $a$ | $(1 + \frac{1}{\lambda})\mathbf{Q}_a$ | $\frac{1+\lambda}{\|v\|_1^2}\mathbf{F}_a$ |
|---|---|---|
| 0 | $Id$ | $\tilde{\mathbf{C}}^{0,0}$ |
| 1 | $\frac{\boldsymbol{\Omega}_0}{\|v\|_1}$ | $\boldsymbol{\Omega}_0 \cdot \tilde{\mathbf{C}}^{1,0} + \tilde{\mathbf{C}}^{0,1} \cdot \boldsymbol{\Omega}_0$ |
| 2 | $\frac{\boldsymbol{\Omega}_0^2 + \boldsymbol{\Omega}_1}{\|v\|_1^2}$ | $\boldsymbol{\Omega}_0^2 \cdot \tilde{\mathbf{C}}^{2,0} + \tilde{\mathbf{C}}^{0,2} \cdot \boldsymbol{\Omega}_0^2 + \boldsymbol{\Omega}_0 \cdot \tilde{\mathbf{C}}^{1,1} \cdot \boldsymbol{\Omega}_0 + \boldsymbol{\Omega}_1 \cdot \tilde{\mathbf{C}}^{1,0} + \tilde{\mathbf{C}}^{0,1} \cdot \boldsymbol{\Omega}_1$ |

Actually, it is possible to compute recursively the $n$th terms, although their complexity explodes. Therefore, it is easy to compute $\boldsymbol{\Omega}_a = \mathbf{F}_a + \mathbf{Q}_a$ for $a \in \mathbb{N}$. By definition $\mathbf{W} = \tilde{p}l\boldsymbol{\Omega} = \tilde{p}l(\mathbf{F} + \mathbf{Q})$, which leads to the result. $\qquad\square$

### B.4 STDP learning with linear neurons and correlated noise

Consider the following $n$-dimensional stochastic differential system:

$$
\begin{cases}
d\mathbf{v}^\epsilon = \frac{1}{\epsilon_1}(-\mathbf{L} \cdot \mathbf{v}^\epsilon + \mathbf{W} \cdot \mathbf{v}^\epsilon + \mathbf{u}(\frac{t}{\epsilon_2}))\,dt + \frac{1}{\sqrt{\epsilon_1}}\boldsymbol{\Sigma} \cdot dB(t), \\
\frac{d\mathbf{W}^\epsilon}{dt} = G(\mathbf{v}^\epsilon, \mathbf{W}^\epsilon) = -\kappa\mathbf{W}^\epsilon + a_+\mathbf{v}^\epsilon \otimes (\mathbf{v}^\epsilon * g_\gamma) - a_-(\mathbf{v}^\epsilon * g_\gamma) \otimes \mathbf{v}^\epsilon,
\end{cases}
$$

where $\mathbf{u}$ is a continuous input in $\mathbb{R}^n$, $l, \epsilon_1, \epsilon_2, \kappa \in \mathbb{R}_+$, $a_+, a_- \in \mathbb{R}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ and $B(t)$ is $n$-dimensional Brownian noise, and for all $\gamma > 0$, $g_\gamma : t \mapsto \gamma e^{-\gamma t} H(t)$ where $H$ is the Heaviside function. Recall the well-posedness Assumption 3.2

**Assumption B.1** There exists $p \in ]0, 1[$ such that

$$
\frac{|a_+| + |a_-|}{p(1 - p)}\left(\frac{s^2\gamma}{2(1 + \gamma/l - p)} + \frac{u_m^2}{(1 - p)}\right) < \kappa l^3.
$$

**Theorem B.11** *If Assumption 3.2 is verified for $p \in ]0, 1[$, then system (17) is asymptotically well posed in probability and the connectivity matrix $\mathbf{W}^\epsilon$, the solution of system (17), converges to $\mathbf{W}$ in the sense that for all $\delta, T > 0$,*

$$
\overset{\mu}{\lim_{\epsilon \to 0}} \mathbb{P}\left[\sup_{t \in [0,T]} |\mathbf{W}_t^\epsilon - \mathbf{W}_t|^2 > \delta\right] = 0,
$$

*where $\mathbf{W}$ is the deterministic solution of*

$$
\frac{d\mathbf{W}_{ij}}{dt} = \bar{G}(\mathbf{W})_{ij}
$$

$$
= \underbrace{-\kappa\mathbf{W}_{ij}}_{decay} + \underbrace{\frac{\mu}{\tau}\int_0^{\frac{\tau}{\mu}} a_+\bar{\mathbf{v}}_i(s)(\bar{\mathbf{v}}_j * g_\gamma)(s) - a_-(\bar{\mathbf{v}}_i * g_\gamma)(s)\bar{\mathbf{v}}_j(s)\,ds}_{correlation} + \underbrace{\mathbf{Q}_{12}}_{noise},
$$

*where $\bar{\mathbf{v}}(t)$ is the $\frac{\tau}{\mu}$-periodic attractor of $\frac{d\bar{\mathbf{v}}}{dt} = (\mathbf{W} - \mathbf{L}) \cdot \bar{\mathbf{v}} + \mathbf{u}(\mu t)$, where $\mathbf{W} \in \mathbb{R}^{n \times n}$ is supposed to be fixed. And $\mathbf{Q}_{12}$ is described below.*

*Proof* We recall the instantaneous reformulation of the original system (17)

$$\begin{cases} d\begin{pmatrix} \mathbf{v}^{\epsilon} \\ \mathbf{z}^{\epsilon} \end{pmatrix} = \frac{1}{\epsilon_1}\left[ \begin{pmatrix} \mathbf{W}-\mathbf{L} & 0 \\ \gamma & -\gamma \end{pmatrix}\begin{pmatrix} \mathbf{v}^{\epsilon} \\ \mathbf{z}^{\epsilon} \end{pmatrix} + \begin{pmatrix} \mathbf{u}(\frac{t}{\epsilon_2}) \\ 0 \end{pmatrix} \right] dt + \begin{pmatrix} \frac{\sigma}{\sqrt{\epsilon_1}} dB(t) \\ \frac{\sigma_{\mathbf{z}}}{\sqrt{\epsilon_1}} dB(t) \end{pmatrix}, \\ \frac{d\mathbf{W}^{\epsilon}}{dt} = -\kappa \mathbf{W}^{\epsilon} + a_+ \mathbf{v}^{\epsilon} \otimes \mathbf{z}^{\epsilon} - a_- \mathbf{z}^{\epsilon} \otimes \mathbf{v}^{\epsilon}. \end{cases}$$

With this linear expression, the structure of the proof of Theorem 3.1 remains unchanged. The correlation term is to be replaced by $\frac{\mu}{\tau}(a_+\bar{\mathbf{v}} \cdot \mathcal{G}'_\gamma \cdot \bar{\mathbf{v}}' + a_-\bar{\mathbf{v}} \cdot \mathcal{G}_\gamma \cdot \bar{\mathbf{v}}')$. The noise term we are looking for is $\mathbf{Q}_{12}$ in the Lyapunov equation (see (9)) below

$$\begin{pmatrix} \mathbf{W}-\mathbf{L} & 0 \\ \gamma & -\gamma \end{pmatrix} \cdot \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}'_{12} \\ \mathbf{Q}_{12} & \mathbf{Q}_{22} \end{pmatrix}$$

$$+ \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}'_{12} \\ \mathbf{Q}_{12} & \mathbf{Q}_{22} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{W}'-\mathbf{L} & \gamma \\ 0 & -\gamma \end{pmatrix} + \begin{pmatrix} \mathbf{\Sigma} \cdot \mathbf{\Sigma}' & 0 \\ 0 & \frac{\sigma_{\mathbf{z}}^2}{2} Id \end{pmatrix} = 0.$$

This leads to the system

$$\begin{cases} (\mathbf{W}-\mathbf{L}) \cdot \mathbf{Q}_{11} + \mathbf{Q}_{11} \cdot (\mathbf{W}'-\mathbf{L}) + \mathbf{\Sigma} \cdot \mathbf{\Sigma}' = 0, & \text{(a)} \\ \gamma(\mathbf{Q}_{11} - \mathbf{Q}_{12}) + \mathbf{Q}_{12} \cdot (\mathbf{W}'-\mathbf{L}) = 0, & \text{(b)} \\ \mathbf{Q}_{22} = \frac{\mathbf{Q}_{12}+\mathbf{Q}'_{12}}{2} + \frac{\sigma_{\mathbf{z}}^2}{2} Id. & \text{(c)} \end{cases} \quad (34)$$

The matrix $\mathbf{Q}_{11}$ is the solution of a Lyapunov equation (see equation (a)). Lemma D.1 gives an explicit solution: $\mathbf{Q}_{11} = \sum_{k=0}^{+\infty} \mathbf{W}^k \cdot \mathbf{\Sigma} \cdot \mathbf{\Sigma}' \cdot (2\mathbf{L}-\mathbf{W}')^{-(k+1)}$. Equation (b) leads to

$$\mathbf{Q}_{12} = \gamma \mathbf{Q}_{11} \cdot (\mathbf{L}+\gamma-\mathbf{W}')^{-1} = \gamma \sum_{k=0}^{+\infty} \mathbf{W}^k \cdot \mathbf{\Sigma} \cdot \mathbf{\Sigma}' \cdot (2\mathbf{L}-\mathbf{W}')^{-(k+1)} \cdot (\mathbf{L}+\gamma-\mathbf{W}')^{-1}.$$

We see that it does not depend on $\sigma_{\mathbf{z}}$, which, once Theorem 2.2 is applied, can be considered null so that the average system $\bar{G}$ corresponds to the original system (17). Therefore,

$$\bar{G}(\mathbf{W}) = -\kappa\mathbf{W} + \frac{\mu}{\tau}\left(a_+\bar{\mathbf{v}} \cdot \mathcal{G}'_\gamma \cdot \bar{\mathbf{v}}' - a_-\bar{\mathbf{v}} \cdot \mathcal{G}_\gamma \cdot \bar{\mathbf{v}}'\right) + a_+\mathbf{Q}'_{12} - a_-\mathbf{Q}_{12}. \quad (35)$$

We show that for $\mathbf{W}$ already in $E_p$, it will stay forever in $E_p$:

1. Inequality $\mathbf{W} \geq 0$:

   Decomposing the connectivity as $\mathbf{W} = \mathbf{S} + i\mathbf{A}$ leads to $\langle X, \mathbf{W} \cdot X \rangle = \langle X, \mathbf{S} \cdot X \rangle + i\langle X, \mathbf{A} \cdot X \rangle$. By hermiticity of $\mathbf{S}$ and $\mathbf{A}$, $\langle X, \mathbf{S} \cdot X \rangle$ and $\langle X, \mathbf{A} \cdot X \rangle$ are real numbers. This means we only have to show that the eigenvalues of $\mathbf{S}$ remain positive along the dynamics. Taking the symmetric part of equation (35) leads to

$$\frac{d\mathbf{S}}{dt} = -\kappa\mathbf{S} + \frac{\mu(a_+ - a_-)}{2\tau}\bar{\mathbf{v}} \cdot (\mathcal{G}_\gamma + \mathcal{G}'_\gamma) \cdot \bar{\mathbf{v}}' + (a_+ - a_-)\mathbf{Q}_{22}.$$

   Suppose we take an initial condition $\mathbf{S}_0 > 0$. It is clear that if $\bar{\mathbf{v}} \cdot (\mathcal{G} + \mathcal{G}') \cdot \bar{\mathbf{v}}'$ and $\mathbf{Q}_{22}$ are always positive, then $\mathbf{S}$ will remain positive. This would prove the result. Therefore, focus on

- Proving $\bar{\mathbf{v}} \cdot (\mathcal{G}_\gamma + \mathcal{G}'_\gamma) \cdot \bar{\mathbf{v}}' \geq 0$:

    According to the first point of Lemma C.1, $\mathcal{G}_\gamma + \mathcal{G}'_\gamma = 2\mathcal{G}_\gamma \cdot \mathcal{G}'_\gamma$. Therefore, $\bar{\mathbf{v}} \cdot (\mathcal{G}_\gamma + \mathcal{G}'_\gamma) \cdot \bar{\mathbf{v}}' = 2\bar{\mathbf{v}} \cdot \mathcal{G}_\gamma \cdot (\bar{\mathbf{v}} \cdot \mathcal{G}_\gamma)'$ is a Gramian matrix and therefore it is positive.

- Proving $\mathbf{Q}_{22} \geq 0$:

    $\mathbf{Q}_{22}$ is the covariance matrix of the random value $\mathbf{z}$, therefore, it is positive-semi-definite.

2. Inequality $\||\mathbf{W}\|| < lp$:

    For all $\mathbf{x} \in \mathbf{C}^n$ such that $\|\mathbf{x}\| = 1$, define a family of positive numbers $(\alpha_{\mathbf{x}})$ whose supremum is written $\alpha^*$ and a family of functions $(g^{\mathbf{x}})$ such that

$$g^{\mathbf{x}} : \mathbf{J} \to \langle \mathbf{x}, \mathbf{J} \cdot \mathbf{x} \rangle - \alpha_{\mathbf{x}}.$$

Because $g$ is linear, $dg^{\mathbf{x}}_{\mathbf{W}}(\mathbf{J}) = \langle \mathbf{x}, \mathbf{J} \cdot \mathbf{x} \rangle$. For $\mathbf{W} \in g^{\mathbf{x}-1}(0)$, *i.e.*, $\langle \mathbf{x}, \mathbf{W} \cdot \mathbf{x} \rangle = \alpha_{\mathbf{x}}$, compute

$$dg^{\mathbf{x}}_{\mathbf{W}}\big(G^\mu(\mathbf{W})\big) = -\kappa \underbrace{\langle \mathbf{x}, \mathbf{W} \cdot \mathbf{x} \rangle}_{=\alpha_{\mathbf{x}}} + \frac{\mu}{\tau} \underbrace{\big\langle \mathbf{x}, \bar{\mathbf{v}} \cdot \big(a_+\mathcal{G}_\gamma - a_-\mathcal{G}'_\gamma\big) \cdot \bar{\mathbf{v}}' \cdot \mathbf{x} \big\rangle}_{=A}$$
$$+ \big(|a_+| + |a_-|\big) \underbrace{\langle \mathbf{x}, \mathbf{Q}_{12} \cdot \mathbf{x} \rangle}_{=B}.$$

- Upper bound of $A$:

    Cauchy-Schwarz leads to

$$|A| \leq |a_+| \big\| \bar{\mathbf{v}} \cdot \mathcal{G}_\gamma \cdot \bar{\mathbf{v}}' \cdot \mathbf{x} \big\| + |a_-| \big\| \bar{\mathbf{v}} \cdot \mathcal{G}'_\gamma \cdot \bar{\mathbf{v}}' \cdot \mathbf{x} \big\|.$$

As before, we can use Young's inequality for convolutions to find an upper bound of $A$ which reads

$$A \leq \frac{\tau u_m^2 (|a_+| + |a_-|)}{(l - \alpha^*)^2}.$$

- Upper bound of $B$:

    According to Proposition 11.9.3 of [35] the solution of the Lyapunov equation (a) in system (34) can be rewritten

$$\mathbf{Q}_{11} = \int_0^{+\infty} e^{-t(\mathbf{L} - \mathbf{W})} \cdot \mathbf{\Sigma} \cdot \mathbf{\Sigma}^t \cdot e^{-t(\mathbf{L} - \mathbf{W}')} \, dt,$$

because $(\mathbf{W} - \mathbf{L}) \oplus (\mathbf{W} - \mathbf{L})$ is not singular due to the fact $\mathbf{W} \in E_p$.

Observe that for $\mathbf{A}$ a positive matrix whose eigenvalues are the $\lambda_i$, then the spectrum of $e^{-\mathbf{A}}$ is $\{e^{-\lambda_i} : i = 1, \ldots, n\}$. Therefore, $\||e^{-\mathbf{A}}\|| = e^{-\min(|\lambda_i|)}$. Therefore, if $\mathbf{A} = \mathbf{L} - \mathbf{W}$, then $\||e^{-\mathbf{A}}\|| \leq e^{\alpha^* - l}$. This leads to

$$\||\mathbf{Q}_{11}\|| \leq s^2 \int_0^{+\infty} e^{2(\alpha^* - l)t} \, dt = s^2 \left[ \frac{e^{2(\alpha^* - l)t}}{2(\alpha^* - l)} \right]_0^{+\infty} = \frac{s^2}{2(l - \alpha^*)}.$$

Then we apply the same arguments to say that

$$|B| = \|\mathbf{Q}_{12}\| \leq \|\mathbf{Q}_{11}\| \|\gamma(\mathbf{L} + \gamma - \mathbf{W})^{-1}\|$$

$$\leq \frac{s^2 \gamma}{2(l - \alpha^*)(l + \gamma - \alpha^*)}.$$

The rest of the proof is identical to the Hebbian case. Assumption 3.1 is changed to Assumption 3.2 for $E_p$ to be invariant by the flow $\bar{G}$. ☐

Define

$$\mathbf{D}^{k,q} = \frac{1}{u_m^2 \tau (|a_+| + |a_-|)} \mathbf{u} \cdot \mathcal{G}_{l/\mu}^{k+1} \cdot \left(a_+ \mathcal{G}_\gamma' - a_- \mathcal{G}_\gamma\right) \cdot \mathcal{G}_{l/\mu}'^{k+1} \cdot \mathbf{u}',$$

such that $\|\mathbf{D}^{k,q}\|_2 \leq 1$.
In this framework, one can prove

**Theorem B.12** *The correlation term can be written*

$$\frac{\mu}{\tau}\left(a_+ \bar{\mathbf{v}} \cdot \mathcal{G}_\gamma' \cdot \bar{\mathbf{v}}' - a_- \bar{\mathbf{v}} \cdot \mathcal{G}_\gamma \cdot \bar{\mathbf{v}}'\right)$$

$$= \frac{u_m^2(|a_+| + |a_-|)}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{W}^k}{l^k} \cdot \mathbf{D}^{k,q} \cdot \frac{\mathbf{W}'^q}{l^q}.$$

*Proof* Similar to that of Theorem 3.3. ☐

**Theorem B.13** *If Assumption 3.2 is verified for $p \leq \frac{1}{3}$, there is a unique equilibrium point which is globally, asymptotically stable.*

*Proof* Similar to the previous case. ☐

Now, we proceed as before by defining

$$\tilde{p} = \frac{|a_+| + |a_-|}{\kappa l^3}\left(\frac{s^2}{2(\frac{1}{l} + \frac{1}{\gamma})} + u_m^2\right) \quad \text{and} \quad \lambda = \frac{s^2}{2u_m^2(\frac{1}{l} + \frac{1}{\gamma})}.$$

**Theorem B.14**

$$\mathbf{W} = \frac{\tilde{p}l}{1 + \lambda}\left(\lambda(\alpha_+ - \alpha_-)\frac{\mathbf{\Sigma} \cdot \mathbf{\Sigma}'}{d} + \mathbf{D}^{0,0}\right)$$

$$+ \frac{\tilde{p}^2 l}{(1 + \lambda)^2}\left(\lambda^2(a_+ - a_-)^2\left(1 + \frac{1}{1 + \gamma/l}\right)\frac{\mathbf{\Sigma} \cdot \mathbf{\Sigma}'^2}{d^2}\right.$$

$$+ \lambda\left(\frac{a_+ - a_-}{2}\right)\left((\mathbf{D}^{0,0} + 2\mathbf{D}^{0,1}) \cdot \frac{\mathbf{\Sigma} \cdot \mathbf{\Sigma}'}{d} + \frac{\mathbf{\Sigma} \cdot \mathbf{\Sigma}'}{d} \cdot (\mathbf{D}^{0,0} + 2\mathbf{D}^{1,0})\right)$$

$$+ \frac{\lambda}{1+\gamma/l}\left(a_+ \mathbf{D}^{0,0}\cdot\frac{\mathbf{\Sigma}\cdot\mathbf{\Sigma}'}{d} - a_-\frac{\mathbf{\Sigma}\cdot\mathbf{\Sigma}'}{d}\cdot\mathbf{D}^{0,0}\right)$$

$$+ \mathbf{D}^{0,0}\cdot\mathbf{D}^{1,0} + \mathbf{D}^{1,0}\cdot\mathbf{D}^{0,0}\bigg) + \mathcal{O}(\tilde{p}^3).$$

*Proof* First, we need to work on the noise term $\mathbf{Q} = a_+\mathbf{Q}'_{12} + a_-\mathbf{Q}_{12}$. Recall $\mathbf{Q}_{11}$ is the solution of the Lyapunov equation $(\mathbf{L}-\mathbf{W})\cdot\mathbf{Q}_{11} + \mathbf{Q}_{11}\cdot(\mathbf{L}-\mathbf{W})' + \mathbf{\Sigma}\cdot\mathbf{\Sigma}' = 0$. Lemma D.1 says that

$$\mathbf{Q}_{11} = \sum_{k=0}^{+\infty}\mathbf{W}^k\cdot\mathbf{\Sigma}\cdot\mathbf{\Sigma}'\cdot(2\mathbf{L}-\mathbf{W}')^{-(k+1)}$$

is a well-defined solution. We now use the fact that $(2\mathbf{L}-\mathbf{W}')^{-(k+1)} = \frac{1}{(2l)^{k+1}}\times$ $\sum_{n=0}^{+\infty}\binom{n+k}{n}\frac{\mathbf{W}'^n}{(2l)^n}$ to show that

$$\mathbf{Q}_{11} = \sum_{k,n=0}^{+\infty}\frac{1}{(2l)^{k+n+1}}\binom{n+k}{n}\mathbf{W}^k\cdot\mathbf{\Sigma}\cdot\mathbf{\Sigma}'\cdot\mathbf{W}'^n$$

and therefore

$$\mathbf{Q}_{12} = \frac{\gamma}{2l(l+\gamma)}\sum_{k,n,q=0}^{+\infty}\frac{1}{2^{k+n}(1+\gamma/l)^q}\binom{n+k}{n}\frac{\mathbf{W}^k}{l^k}\cdot\mathbf{\Sigma}\cdot\mathbf{\Sigma}'\cdot\frac{\mathbf{W}'^{n+q}}{l^{n+q}}.$$

Thus, writing $\alpha_\pm = \frac{a_\pm}{|a_+|+|a_-|}$ and $c_{k,n,q} = \frac{1}{2^{k+n}(1+\gamma/l)^q}\binom{n+k}{n}$, the noise term is

$$\mathbf{Q} = \frac{d(|a_+|+|a_-|)}{2l^2(\frac{1}{l}+\frac{1}{\gamma})}\sum_{k,n,q=0}^{+\infty}c_{k,n,q}\bigg(\alpha_+\frac{\mathbf{W}^{n+q}}{l^{n+q}}\cdot\frac{\mathbf{\Sigma}\cdot\mathbf{\Sigma}'}{d}\cdot\frac{\mathbf{W}'^k}{l^k}$$

$$-\alpha_-\frac{\mathbf{W}^k}{l^k}\cdot\frac{\mathbf{\Sigma}\cdot\mathbf{\Sigma}'}{d}\cdot\frac{\mathbf{W}'^{n+q}}{l^{n+q}}\bigg).$$

Define $\mathbf{\Omega} = \frac{\mathbf{W}}{\tilde{p}l}$ such that $\|\!|\mathbf{\Omega}|\!\| = \mathcal{O}(1)$. We improperly write $\bar{G}(\mathbf{\Omega}) = \bar{G}(\tilde{p}l\mathbf{\Omega})$ such that

$$\bar{G}(\mathbf{\Omega}) = -\tilde{p}l\kappa\mathbf{\Omega} + \frac{u_m^2(|a_+|+|a_-|)}{l^2}\sum_{k,q=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k\cdot\mathbf{D}^{k,q}\cdot(\tilde{p}\mathbf{\Omega})^q$$

$$+ \frac{d(|a_+|+|a_-|)}{2l^2(\frac{1}{l}+\frac{1}{\gamma})}\sum_{k,n,q=0}^{+\infty}c_{k,n,q}\bigg(\alpha_+(\tilde{p}\mathbf{\Omega})^{n+q}\cdot\frac{\mathbf{\Sigma}\cdot\mathbf{\Sigma}'}{d}\cdot(\tilde{p}\mathbf{\Omega}')^k$$

$$-\alpha_-(\tilde{p}\mathbf{\Omega})^k\cdot\frac{\mathbf{\Sigma}\cdot\mathbf{\Sigma}'}{d}\cdot(\tilde{p}\mathbf{\Omega}')^{n+q}\bigg).$$

This leads to

$$\bar{G}(\boldsymbol{\Omega}) = \left( \frac{u_m^2(|a_+| + |a_-|)}{l^2} + \frac{d(|a_+| + |a_-|)}{2l^2(\frac{1}{l} + \frac{1}{\gamma})} \right)$$

$$\times \left[ -\boldsymbol{\Omega} + \overbrace{\frac{1}{1+\lambda} \sum_{k,q=0}^{+\infty} (\tilde{p}\boldsymbol{\Omega})^k \cdot \mathbf{D}^{k,q} \cdot (\tilde{p}\boldsymbol{\Omega})^q}^{\tilde{\mathbf{F}}} + \frac{\lambda}{1+\lambda} \right.$$

$$\left. \times \underbrace{\sum_{k,n,q=0}^{+\infty} c_{k,n,q} \left( \alpha_+ (\tilde{p}\boldsymbol{\Omega})^{n+q} \cdot \frac{\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}'}{d} \cdot (\tilde{p}\boldsymbol{\Omega}')^k - \alpha_- (\tilde{p}\boldsymbol{\Omega})^k \cdot \frac{\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}'}{d} \cdot (\tilde{p}\boldsymbol{\Omega}')^{n+q} \right)}_{\tilde{\mathbf{Q}}} \right].$$

We are looking for $\mathbf{F}_a$ and $\mathbf{Q}_a$ in the expansions $\tilde{\mathbf{F}} = \sum_{a=0}^{+\infty} \mathbf{F}_a \tilde{p}^a$ and $\tilde{\mathbf{Q}} = \sum_{a=0}^{+\infty} \mathbf{Q}_a \tilde{p}^a$. Recall

$$\boldsymbol{\Omega}^p = \sum_{i=0}^{+\infty} \tilde{p}^i \sum_{\eta \in \mathbb{N}^p, \sum_k \eta_k = i} \boldsymbol{\Omega}_{\eta_1} \cdot \boldsymbol{\Omega}_{\eta_2} \cdot \ldots \cdot \boldsymbol{\Omega}_{\eta_p}.$$

Therefore,

$$\tilde{\mathbf{Q}} = \sum_{k,n,q,i,j=0}^{+\infty} c_{k,n,q} \tilde{p}^{k+n+q+i+j} \sum_{\substack{\eta \in \mathbb{N}^k, \sum_m \eta_m = i \\ \theta \in \mathbb{N}^{n+q}, \sum_m \theta_m = j}} \alpha_+ \boldsymbol{\Omega}_{\eta_1} \cdot \ldots \cdot \boldsymbol{\Omega}_{\eta_{n+q}}$$

$$\times \frac{\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}'}{d} \cdot \boldsymbol{\Omega}'_{\theta_1} \cdot \ldots \cdot \boldsymbol{\Omega}'_{\theta_k} - \alpha_- \boldsymbol{\Omega}_{\eta_1} \cdot \ldots \cdot \boldsymbol{\Omega}_{\eta_k} \cdot \frac{\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}'}{d} \cdot \boldsymbol{\Omega}'_{\theta_1} \cdot \ldots \cdot \boldsymbol{\Omega}'_{\theta_{n+q}}.$$

Leading to

$$\mathbf{Q}_a = \sum_{\substack{k,n,i,j=0 \\ a=k+n+q+i+j}}^{+\infty} c_{k,n,q} \tilde{p}^{k+n+q+i+j} \sum_{\substack{\eta \in \mathbb{N}^k, \sum_m \eta_m = i \\ \theta \in \mathbb{N}^{n+q}, \sum_m \theta_m = j}} \alpha_+ \boldsymbol{\Omega}_{\eta_1} \cdot \ldots \cdot \boldsymbol{\Omega}_{\eta_{n+q}}$$

$$\times \frac{\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}'}{d} \cdot \boldsymbol{\Omega}'_{\theta_1} \cdot \ldots \cdot \boldsymbol{\Omega}'_{\theta_k} - \alpha_- \boldsymbol{\Omega}_{\eta_1} \cdot \ldots \cdot \boldsymbol{\Omega}_{\eta_k} \cdot \frac{\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}'}{d} \cdot \boldsymbol{\Omega}'_{\theta_1} \cdot \ldots \cdot \boldsymbol{\Omega}'_{\theta_{n+q}}.$$

This equation is scary but it reduces to simple expressions for small $a \in \mathbb{N}$.

| $a$ | $\mathbf{Q}_a$ | $\mathbf{F}_a$ |
|---|---|---|
| 0 | $(\alpha_+ - \alpha_-)\frac{\boldsymbol{\Sigma}\cdot\boldsymbol{\Sigma}'}{d}$ | $\mathbf{D}^{0,0}$ |
| 1 | $\frac{\alpha_+ - \alpha_-}{2}(\boldsymbol{\Omega}_0 \cdot \frac{\boldsymbol{\Sigma}\cdot\boldsymbol{\Sigma}'}{d} + \frac{\boldsymbol{\Sigma}\cdot\boldsymbol{\Sigma}'}{d} \cdot \boldsymbol{\Omega}'_0) + \frac{1}{1+\gamma/l}(\alpha_+\boldsymbol{\Omega}_0 \cdot \frac{\boldsymbol{\Sigma}\cdot\boldsymbol{\Sigma}'}{d} - \alpha_- \frac{\boldsymbol{\Sigma}\cdot\boldsymbol{\Sigma}'}{d} \cdot \boldsymbol{\Omega}'_0)$ | $\boldsymbol{\Omega}_0 \cdot \mathbf{D}^{1,0} + \mathbf{D}^{0,1} \cdot \boldsymbol{\Omega}'_0$ |

Recall that $\mathbf{W} = \tilde{p}l\boldsymbol{\Omega} = \tilde{p}l(\frac{1}{1+\lambda}\tilde{\mathbf{F}} + \frac{\lambda}{1+\lambda}\tilde{\mathbf{Q}})$ to get the result. $\qquad\square$

# Appendix C: Properties of the convolution operators $\mathcal{G}_\gamma$, $\mathcal{W}$, and $\mathcal{V}$

Recall $\mathcal{G}_\gamma$, $\mathcal{W}$, and $\mathcal{V}$ are convolution operators respectively generated by $g_\gamma$, $v$, and $w$ defined in (30). Their Fourier transforms are respectively

$$\hat{g}_\gamma : \xi \mapsto \frac{\gamma}{\gamma + 2i\pi\xi},$$

$$\hat{v} : \xi \mapsto \frac{4\beta}{(\beta(1+\Delta) + 4i\pi\mu\xi)(\beta(1-\Delta) + 4i\pi\mu\xi)},$$

$$\hat{w} : \xi \mapsto \frac{4\beta + 8i\pi\mu\xi}{(\beta(1+\Delta) + 4i\pi\mu\xi)(\beta(1-\Delta) + 4i\pi\mu\xi)}.$$

## C.1 Algebraic properties

**Lemma C.1**

$$\frac{\mathcal{G}_\gamma + \mathcal{G}'_\gamma}{2} = \mathcal{G}_\gamma \cdot \mathcal{G}'_\gamma.$$

*Proof* Compute

$$
\begin{aligned}
(\mathcal{G}_\gamma \cdot \mathcal{G}'_\gamma)_{xy} &= \gamma^2 \int_{-\infty}^{+\infty} e^{-\gamma(x-z)} H(x-z) e^{-\gamma(y-z)} H(y-z)\, dz \\
&= \gamma^2 e^{-\gamma(x+y)} \int_{-\infty}^{\min(x,y)} e^{2\gamma z}\, dz = \gamma^2 e^{-\gamma(y+x)} \left[ \frac{e^{2\gamma z}}{2\gamma} \right]_{-\infty}^{\min(x,y)} \\
&= \frac{\gamma}{2} e^{-\gamma(y+x-2\min(x,y))}.
\end{aligned}
$$

Therefore, if $y \geq x$, then $(\mathcal{G}_\gamma \cdot \mathcal{G}'_\gamma)_{xy} = \frac{\gamma}{2} e^{-\gamma(y-x)}$, and if $x \geq y$, then $(\mathcal{G}_\gamma \cdot \mathcal{G}'_\gamma)_{xy} = \frac{\gamma}{2} e^{-\gamma(x-y)}$. The result follows. $\square$

**Lemma C.2**

$$\mathcal{G}'_\gamma - \mathcal{G}_\gamma = \frac{1}{\gamma}\mathcal{D} \cdot (\mathcal{G}'_\gamma + \mathcal{G}_\gamma),$$

*where $\mathcal{D}$ is the time-differentiation operator, i.e., $(\boldsymbol{\mathcal{X}} \cdot \mathcal{D})(t) = \frac{d\boldsymbol{\mathcal{X}}}{dt}(t)$.*

*Proof* $\mathcal{G}_\gamma$ and $\mathcal{G}'_\gamma$ are two convolution operators respectively generated by $g_\gamma : t \mapsto \gamma e^{-\gamma t} H(t)$ and $g'_\gamma : t \mapsto \gamma e^{\gamma t} H(-t)$. The Fourier transform of $g_\gamma$ is $\hat{h}(\xi) = \frac{\gamma}{\gamma + 2i\pi\xi}$. Therefore, the Fourier transform of $g'_\gamma - g_\gamma$ is

$$
\begin{aligned}
\widehat{g'_\gamma - g_\gamma}(\xi) &= \frac{\gamma}{\gamma - 2i\pi\xi} - \frac{\gamma}{\gamma + 2i\pi\xi} = \frac{2i\pi\xi}{\gamma} \frac{2\gamma^2}{\gamma^2 + 4\pi^2\xi^2} \\
&= \frac{2i\pi\xi}{\gamma}\left( \frac{\gamma}{\gamma + 2i\pi\xi} + \frac{\gamma}{\gamma - 2i\pi\xi} \right) = \frac{2i\pi\xi}{\gamma}\left( \widehat{g'_\gamma + g_\gamma}(\xi) \right).
\end{aligned}
$$

Because $\frac{d\hat{f}}{dt}(\xi) = 2i\pi\xi\,\hat{f}$, taking the inverse Fourier transform of $\widehat{g'_\gamma - g_\gamma}(\xi)$ gives the result.　　　　　　　　　　　　　　　　　　　　　　　　　　　□

**Lemma C.3**

$$\mathcal{W}\cdot\mathcal{V}^k\cdot\mathcal{G}_{\beta/\mu} = \mathcal{V}^{k+1}.$$

*Besides, if $\Delta \in i\mathbb{R}$, $\mathcal{V}^k$ is a convolution operator generated by*

$$v_k : t \mapsto \frac{\sqrt{\pi\beta}}{k!}e^{-\frac{\beta}{2}t}\left(\frac{t}{|\Delta|}\right)^{k+\frac{1}{2}}J_{k+\frac{1}{2}}\left(\frac{\beta|\Delta|}{2}t\right)H(t),$$

*where $J_n(z)$ is the Bessel function of the first kind. If $\Delta \in \mathbb{R}$, the formula above holds if one replaces $J_n(z)$ by $I_n(z)$, the modified Bessel function of the first kind.*

*Proof* We want to compute $\mathcal{W}\cdot\mathcal{V}^k\cdot\mathcal{G}_{\beta/\mu}$. Compute the Fourier transform of $w * v_k * g_{\beta/\mu}$, where $v_k$ is the result of k convolutions of $v$ with itself

$$\widehat{w * v_k * g_{\frac{\beta}{\mu}}}(\xi) = \hat{w}(\xi)\hat{g}_{\frac{\beta}{\mu}}(\xi)\hat{v}_k(\xi)$$

$$= \left(\frac{\beta}{(\frac{\beta(1+\Delta)}{2} + 2i\pi\mu\xi)(\frac{\beta(1-\Delta)}{2} + 2i\pi\mu\xi)}\right)^{k+1} = \hat{v}^{k+1}(\xi).$$

This proves the first result.

Then observe that

$$v_{k+1}(t) = \beta^{k+1}\mathcal{F}^{-1}\left(\xi \mapsto \frac{1}{(\frac{\beta(1+\Delta)}{2} + 2i\pi\mu\xi)^{k+1}}\right)$$

$$* \mathcal{F}^{-1}\left(\xi \mapsto \frac{1}{(\frac{\beta(1-\Delta)}{2} + 2i\pi\mu\xi)^{k+1}}\right)(t)$$

$$= \beta^{k+1}\left(s \mapsto \frac{s^k}{k!}e^{-\frac{\beta(1+\Delta)}{2}s}H(s)\right) * \left(s \mapsto \frac{s^k}{k!}e^{-\frac{\beta(1-\Delta)}{2}s}H(s)\right)(t)$$

$$= \frac{\beta^{k+1}}{k!^2}e^{-\frac{\beta(1-\Delta)}{2}t}\int_0^t s^k(t-s)^k e^{-\beta\Delta s}\,ds\,H(t).$$

The last integral can be analytically computed with the help of Bessel functions. In fact, it gives different results depending on the nature of $\Delta$ (whether it is real or imaginary).

- If $\Delta \in \mathbb{R}$, then defining $I_n(z)$, the modified Bessel function of the first kind, leads to

$$\int_0^t e^{-\beta\Delta s}s^k(t-s)^k\,ds = \sqrt{\pi}e^{-\frac{\beta\Delta}{2}t}k!\left(\frac{t}{\beta\Delta}\right)^{k+\frac{1}{2}}I_{k+\frac{1}{2}}\left(\frac{\beta\Delta}{2}t\right).$$

- If $\Delta \in i\mathbb{R}$, then defining $J_n(z)$, the Bessel function of the first kind, leads to

$$\int_0^t e^{-\beta \Delta s} s^k (t-s)^k \, ds = \sqrt{\pi} e^{-\frac{\beta \Delta}{2} t} k! \left(\frac{t}{\beta |\Delta|}\right)^{k+\frac{1}{2}} \left(\frac{\beta |\Delta|}{2} t\right).$$

This concludes the proof.    $\square$

### C.2 Signed integral

1. $\int_{-\infty}^{+\infty} g_\gamma(t) \, dt = \gamma \frac{0-1}{-\gamma} = 1.$
2. For $\Delta = \sqrt{1 - \frac{4l}{\beta}} \in \mathbb{C}$, compute

$$
\begin{aligned}
\int_{-\infty}^{+\infty} v(t) \, dt &= \frac{l}{\Delta \mu} \left( \int_0^{+\infty} e^{-\frac{\beta}{2\mu}(1-\Delta)t} \, dt - \int_0^{+\infty} e^{-\frac{\beta}{2\mu}(1-\Delta)t} \, dt \right) \\
&= \frac{l}{\Delta \mu} \left( \frac{0-1}{-\frac{\beta}{2\mu}(1-\Delta)} - \frac{0-1}{-\frac{\beta}{2\mu}(1+\Delta)} \right) \\
&= \frac{2l}{\Delta \beta} \frac{1+\Delta-(1-\Delta)}{1-\Delta^2} = \frac{4l}{\beta} \frac{\beta}{4l} = 1.
\end{aligned}
$$

3. Similarly,

$$
\begin{aligned}
\int_{-\infty}^{+\infty} w(t) \, dt &= \frac{l}{2\Delta \mu} \left( (1+\Delta) \int_0^{+\infty} e^{-\frac{\beta}{2\mu}(1-\Delta)t} \, dt \right. \\
&\qquad \left. - (1-\Delta) \int_0^{+\infty} e^{-\frac{\beta}{2\mu}(1-\Delta)t} \, dt \right) \\
&= \frac{l}{\Delta \beta} \frac{(1+\Delta)^2 - (1-\Delta)^2}{1-\Delta^2} \\
&= \frac{l}{\Delta \beta} \frac{4\Delta \beta}{4l} = 1.
\end{aligned}
$$

### C.3 L1 norm

- For $4l \le \beta$, i.e., $\Delta = \sqrt{1 - \frac{4l}{\beta}} \in \mathbb{R}_+$, then

  1. $g_\gamma(t) > 0$ and $\|g_\gamma\|_1 = \int_{\mathbb{R}} g_\gamma(t) \, dt = 1.$
  2. $v(t) = \frac{2l}{\Delta \mu} e^{-\frac{\beta}{2\mu} t} sh(\frac{\beta \Delta}{2\mu} t) H(t) \ge 0$ and $\|v\|_1 = \int_{\mathbb{R}} v(t) \, dt = 1.$
  3. $w(t) = \frac{l}{\Delta \mu} e^{-\frac{\beta}{2\mu} t} (sh(\frac{\beta \Delta}{2\mu} t) + \Delta ch(\frac{\beta \Delta}{2\mu} t)) H(t) \ge 0$ and $\|u\|_1 = \int_{\mathbb{R}} u(t) \, dt = 1.$

- For $4l > \beta$, i.e., $\Delta$ is a pure imaginary, we rewrite $\Delta = \sqrt{\frac{4l}{\beta} - 1}$ and observe that

  1. $g_\gamma(t) > 0$ and $\|g_\gamma\|_1 = \int_{\mathbb{R}} g_\gamma(t) \, dt = 1.$

2. $v(t) = \frac{2l}{\Delta\mu} e^{-\frac{\beta}{2\mu}t} \sin(\frac{\beta\Delta}{2\mu}t) H(t)$ which changes sign on $\mathbb{R}_+$. Therefore,

$$\|v\|_1 = \frac{2l}{\Delta\mu} \int_0^{+\infty} e^{-\frac{\beta}{2\mu}t} \left|\sin\left(\frac{\beta\Delta}{2\mu}t\right)\right| dt = \frac{4l}{\Delta^2\beta} \int_0^{+\infty} e^{-\frac{s}{\Delta}} \left|\sin(s)\right| ds$$

$$= \frac{4l}{\Delta^2\beta} \sum_{k=0}^{+\infty} (-1)^k \int_{k\pi}^{(k+1)\pi} e^{-\frac{s}{\Delta}} \sin(s)\, ds$$

$$= \frac{4l}{\Delta^2\beta} \sum_{k=0}^{+\infty} (-1)^k \frac{(-1)^k e^{-\frac{k\pi}{\Delta}} - (-1)^{k+1} e^{-\frac{(k+1)\pi}{\Delta}}}{1 + \frac{1}{\Delta^2}}$$

$$= \left(1 + e^{-\frac{\pi}{\Delta}}\right) \sum_{k=0}^{+\infty} e^{-\frac{k\pi}{\Delta}} = \frac{1 + e^{-\frac{\pi}{\Delta}}}{1 - e^{-\frac{\pi}{\Delta}}} = \coth\left(\frac{\pi}{2\Delta}\right).$$

3. $w(t) = \frac{l}{\Delta\mu} e^{-\frac{\beta}{2\mu}t} (\sin(\frac{\beta\Delta}{2\mu}t) + \Delta\cos(\frac{\beta\Delta}{2\mu}t)) H(t)$ which also changes sign on $\mathbb{R}_+$. We have not found a way to compute $\|w\|_1$ and write the result elegantly.

## Appendix D:  Solution of a Lyapunov equation

**Lemma D.1** *The solution of the following Lyapunov equation*

$$(\mathbf{L} - \mathbf{W}) \cdot \mathbf{X} + \mathbf{X} \cdot (\mathbf{L} - \mathbf{W})' + \mathbf{D} = 0,$$

*where $\mathbf{L} = lId$ is*

$$\mathbf{X} = -\sum_{k=0}^{+\infty} \mathbf{W}^k \cdot \mathbf{D} \cdot (2\mathbf{L} - \mathbf{W}')^{-(k+1)}. \tag{36}$$

*Proof* First, observe that if $\{|\lambda| : \lambda$ eigenvalue of $\mathbf{W}\} \in ]0, l[$ and $\mathbf{W} > 0$, then $\||\mathbf{W}|\| \||(2\mathbf{L} - \mathbf{W})^{-1}|\| < 1$. Therefore, $\mathbf{X}$ is well defined by equation (36).

Observe that $(2\mathbf{L} - \mathbf{W}')^{-1} \cdot (\mathbf{L} - \mathbf{W})' = I_d - \mathbf{L} \cdot (2\mathbf{L} - \mathbf{W}')^{-1}$. Assuming $\mathbf{X}$ is defined by equation (36), then based on the fact $\mathbf{L}$ commutes with any matrix (because it is a scalar matrix),

$$(\mathbf{L} - \mathbf{W}) \cdot \mathbf{X} + \mathbf{X} \cdot (\mathbf{L} - \mathbf{W})'$$

$$= -\left(\mathbf{L} \cdot \sum_{k=0}^{+\infty} \mathbf{W}^k \cdot \mathbf{D} \cdot (2\mathbf{L} - \mathbf{W}')^{-(k+1)} - \mathbf{W} \cdot \sum_{k=0}^{+\infty} \mathbf{W}^k \cdot \mathbf{D} \cdot (2\mathbf{L} - \mathbf{W}')^{-(k+1)}\right.$$

$$\left. + \sum_{k=0}^{+\infty} \mathbf{W}^k \cdot \mathbf{D} \cdot (2\mathbf{L} - \mathbf{W}')^{-k} - \mathbf{L} \cdot \sum_{k=0}^{+\infty} \mathbf{W}^k \cdot \mathbf{D} \cdot (2\mathbf{L} - \mathbf{W}')^{-(k+1)}\right)$$

$$= -\left(\sum_{k=0}^{+\infty} \mathbf{W}^k \cdot \mathbf{D} \cdot (2\mathbf{L} - \mathbf{W}')^{-k} - \sum_{k=0}^{+\infty} \mathbf{W}^{k+1} \cdot \mathbf{D} \cdot (2\mathbf{L} - \mathbf{W}')^{-(k+1)}\right) = -\mathbf{D}. \qquad \square$$

# References

1. Kandel E, Schwartz J, Jessell T, et al.: *Principles of Neural Science*, vol 4. New York: McGraw-Hill; 2000.
2. Khas'minskii R: **The averaging principle for stochastic differential equations**. *Probl Pereda Inf* 1968, **4**(2):86-87.
3. Pavliotis G, Stuart A: *Multiscale Methods: Averaging and Homogenization*, vol 53. Berlin: Springer; 2008.
4. Arnold V, Levi M: *Geometrical Methods in the Theory of Ordinary Differential Equations*, vol 250. Berlin: Springer; 1988.
5. Lorenzi L, Lunardi A, Zamboni A: **Asymptotic behavior in time periodic parabolic problems with unbounded coefficients**. *J Differ Equ* 2010, **249**(12):3377-3418.
6. Wainrib G: (2012). **Double averaging principle for periodically forced slow-fast stochastic systems**. *Electron Commun Probab*. In revision.
7. Ermentrout G, Terman D: *Mathematical Foundations of Neuroscience*, vol 35. Berlin: Springer; 2010.
8. Izhikevich E: *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Cambridge: MIT Press; 2007.
9. Hebb D: *The Organisation of Behaviour*; 1949.
10. Dayan P, Abbott L: *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge: MIT Press; 2001.
11. Gerstner W, Kistler W: *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge: Cambridge University Press; 2002.
12. Caporale N, Dan Y: **Spike timing-dependent plasticity: a Hebbian learning rule**. *Annu Rev Neurosci* 2008, **31**:25-46.
13. Oja E: **Simplified neuron model as a principal component analyzer**. *J Math Biol* 1982, **15**(3):267-273.
14. Miller K, MacKay D: **The role of constraints in Hebbian learning**. *Neural Comput* 1994, **6**(1):100-126.
15. Rolls E, Deco G: *Computational Neuroscience of Vision*, chapter 8. London: Oxford University Press; 2002.
16. Hopfield JJ: **Hopfield network**. *Scholarpedia* 2007, **2**(5):1977.
17. Gerstner W, Kistler W: **Mathematical formulations of Hebbian learning**. *Biol Cybern* 2002, **87**(5):404-415.
18. Morrison A, Diesmann M, Gerstner W: **Phenomenological models of synaptic plasticity based on spike timing**. *Biol Cybern* 2008, **98**(6):459-478.
19. Burkitt A, Gilson M, van Hemmen J: **Spike-timing-dependent plasticity for neurons with recurrent connections**. *Biol Cybern* 2007, **96**(5):533-546.
20. Gilson M, Burkitt A, Grayden D, Thomas D, van Hemmen J: **Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks iv**. *Biol Cybern* 2009, **101**(5):427-444.
21. Fenichel N: **Geometric singular perturbation theory for ordinary differential equations**. *J Differ Equ* 1979, **31**(1):53-98.
22. O'Malley R: *Singular Perturbation Methods for Ordinary Differential Equations*. New York: Springer; 1991.
23. Kifer Y: *Large Deviations and Adiabatic Transitions for Dynamical Systems and Markov Processes in Fully Coupled Averaging*, vol 944. Providence: Am. Math. Soc.; 2009.
24. Risken H: *The Fokker-Planck Equation: Methods of Solution and Applications*, vol 18. Berlin: Springer; 1996.
25. Galtier M, Touboul J: **On an explicit representation of the solution of linear stochastic partial differential equations with delays**. *C R Math* 2012, **350**:167-172.
26. Hawkes A: **Point spectra of some mutually exciting point processes**. *J R Stat Soc, Ser B, Methodol* 1971, **33**:438-443.
27. Földiák P: **Learning invariance from transformation sequences**. *Neural Comput* 1991, **3**(2):194-200.
28. Bi G, Poo M: **Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type**. *J Neurosci* 1998, **18**(24):10464.
29. Xie X, Seung H: **Spike-based learning rules and stabilization of persistent neural activity**. *Adv Neural Inf Process Syst* 2000, **12**:199-205.

30. Izhikevich E: **Resonate-and-fire neurons**. *Neural Netw* 2001, **14**(6-7):883-894.
31. Romani G, Williamson S, Kaufman L: **Tonotopic organization of the human auditory cortex**. *Science* 1982, **216**(4552):1339.
32. Abbott L, Nelson S, et al.: **Synaptic plasticity: taming the beast**. *Nat Neurosci* 2000, **3**:1178-1183.
33. Higham DJ, Mao X, Stuart AM: **Strong convergence of Euler-type methods for nonlinear stochastic differential equations**. *SIAM J Numer Anal* 2002, **40**(3):1041-1063.
34. Hale J, Lunel S: *Introduction to Functional Differential Equations*. Berlin: Springer; 1993.
35. Bernstein D: *Matrix Mathematics: Theory, Facts, and Formulas*. Princeton: Princeton University Press; 2009.