Distributed synaptic weights in a LIF neural network and learning rules

Benoît Perthame, Delphine Salort, Gilles Wainrib

Distributed synaptic weights in a LIF neural network and learning rules

Benoît Perthame* Delphine Salort† Gilles Wainrib‡

June 17, 2017

Abstract

Leaky integrate-and-fire (LIF) models are mean-field limits, with a large number of neurons, used to describe neural networks. We consider inhomogeneous networks structured by a connectivity parameter (strengths of the synaptic weights) with the effect of processing the input current with different intensities.

We first study the properties of the network activity depending on the distribution of synaptic weights and in particular its discrimination capacity. Then, we consider simple learning rules and determine the synaptic weight distribution it generates. We outline the role of noise as a selection principle and the capacity to memorized a learned signal.

Key words: Neural networks; Learning rules; Fokker-Planck equation; Integrate and Fire;
Mathematics Subject Classification (2010): 35Q84; 68T05; 82C32; 92B20;

1 Introduction

Learning and memory are essential cognitive functions which are supported by the subtle mechanisms of synaptic plasticity [26]. Since the seminal work of Hebb [22], one of the key challenges in theoretical neuroscience and artificial intelligence is to understand the consequences of various learning rules on the organization of neural networks and on the way they process and memorize information.

Despite several theoretical works on this topic [21, 15, 20, 19, 18] it remains difficult to investigate learning processes using macroscopic models with infinite number of neurons because these usually assume a form of homogeneity, either under uniform synaptic weights assumptions leading to McKean-Vlasov limits [13, 2, 16, 17, 3] or under random connectivity models leading to dynamical spin-glass limits [28, 1]. In contrast, when studying synaptic plasticity and learning, one needs to describe all the connections between each pair of neurons, hence breaking the homogeneity usually necessary to derive such macroscopic limits.

*Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7598, Laboratoire Jacques-Louis Lions, Inria Équipe MAMBA, 4, place Jussieu 75005, Paris, France, Email: benoit.perthame@upmc.fr

†Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7238 , Laboratoire de Biologie Computationnelle et quantitative, 4, place Jussieu 75005, Paris, France, Email: delphine.salort@upmc.fr

‡Ecole Normale Superieure France, Département d'Informatique, équipe DATA, Paris, France, Email: gilles.wainrib@ens.fr

In this article, we propose a way to circumvent this difficulty by introducing a mathematical model describing a macroscopic population of leaky integrate-and-fire neurons (LIF in short), which are interacting through a mean-field variable and where learning rules are governed by the mean activity of the network. More precisely, in contrast with previous works on such questions, we consider a model where neuronal subpopulations interact with the mean-field through a heterogeneous distribution of synaptic weights values: in other words, each subpopulation sees the mean-field through a different lens. Instead of considering activity-dependent changes in the pairwise synaptic weights between neurons, we consider that each subpopulation receives a weighed version of the overall network activity, and that the associated weights can be dynamically modified according to a specific rule. Based on this heterogeneous model, we are therefore able to integrate a learning rule term in the equation, for the first time in the class of macroscopic LIF equations. One particular instance of such learning rule corresponds to the idea of Hebbian learning, in the sense that the connection between a given subpopulation and the mean-field is strengthen if both have a correlated activity and is weakened otherwise.

The introduction of this mathematical framework supports the investigation of several questions inspired by the seminal work of Hopfield [23], which are answered, at least partially, in this article:

1. For a given pattern of steady-state neural activity, can we always find a heterogeneous synaptic weight distribution that generates such activity?

2. What is the equilibrium synaptic weight distribution according to the learning rule and to the external signal?

3. Is the system able to remember which external signal was presented during the learning phase?

We present in Section 2 the different mathematical models that we consider in order to study the effect of a mean field learning rule on coupled neural networks governed by Noisy LIF models structured by a connectivity parameter. In Section 3, we introduce some material about the possible stationary solutions of these models without learning rule, in particular we study existence and uniqueness. This material will be use throughout the paper. Next, we focus our study on qualitative properties on the input-output map and on the learning rules. In Section 4, we prove that our model can produce a large class of output signals by an appropriate choice of the distribution of connectivity. In Section 5, we show the ability of our model to differentiate different inputs via a discrimination property: we prove that we cannot obtain the same output signal considering two different input signals and, given two different inputs, we give an estimate of the difference between the two possible output signals associated to two different inputs. The last sections are devoted to the study with the learning rule. In Section 6, we address the full system with learning and we describe possible equilibrium connectivity distributions and prove non-uniqueness. We propose in Section 7 a selection principle by adding noise in the learning stage. In Section 8, we describe the ability of such system to memorize learned signals by discriminating easily new incoming inputs after learning.

2 Mathematical models of a mean field learning rule

It is standard to describe homogeneous neural networks by LIF models. In order to study the impact of heterogeneity and of mean field learning rules, we introduce a mathematical model for coupled neural networks. To this end, we firstly explain the equations which describe the activity of a macroscopic population of LIF neural networks when they interact through they mean activity. Secondly, we present how mean field learning rules may be included in these equations.

2.1 Structured Noisy LIF model

We consider a heterogeneous population of homogeneous neural networks structured by their synaptic weights $w \in (-\infty, +\infty)$, negative sign stands for inhibitory neurons and positive sign for excitatory neurons. We have chosen to use a signed parameter σ, instead of a system of two equations for excitatory and inhibitory neurones, because this leads to a simpler formalism and avoids boundary conditions in $w = 0$. We assume that each homogeneous subpopulation, with synaptic weight w, is governed by the classical mean field noisy integrate and fire equation, widely used for large neural networks [7, 6, 5]. Moreover, we assume that the subpopulations interact via the total firing rate $\bar{N}(t)$ defined as the mean activity for all the subpopulations. Setting v the value of the action potential, $V_F > V_R$ the values of the firing and reset potentials, we consider the classical equation

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial v}\left[\big(-v + I(t,w) + w\sigma(\bar{N}(t))\big)p\right] - a\frac{\partial^2 p}{\partial v^2} = N(w,t)\delta(v - V_R), \tag{1}$$

with the boundary and initial conditions

$$p(V_F, w, t) = 0, \qquad p(-\infty, w, t) = 0, \qquad p(v, w, 0) = p^0(v, w). \tag{2}$$

The solution $p(v, w, t)$ defines the probability to find a neuron at potential v with a synaptic weight w, the coefficient a represents the synaptic noise which we assume to be constant and $I(t, w)$ is the input signal which strength is possibly modulated by the synaptic weights. The subnetwork activity $N(w, t)$ and total activity $\bar{N}(t)$ are defined as

$$N(w,t) := -a\frac{\partial p}{\partial v}(V_F, w, t) \geq 0, \qquad \bar{N}(t) = \int_{-\infty}^{\infty} N(w,t)dw. \tag{3}$$

The function $\sigma(\cdot)$ represents the response of the network to the total activity. We use either $\sigma(N) = N$, or the following class with saturation

$$\sigma \in \mathcal{C}^2(\mathbb{R}^+; \mathbb{R}^+), \qquad \sigma_M = \max \sigma(\cdot) < \infty, \qquad \sigma' \geq 0. \tag{4}$$

We recall some mathematical properties of distributional solutions of Equation (1)–(3) which were studied in [8, 10, 11]. There, some existence, uniqueness and long time behaviour results are established for distributional solutions. For excitatory networks, solutions can blow-up in finite time, as discovered in [8], when $\sigma(N) = N$. The saturation assumption (4) prevents blow-up, a phenomena which also appears if one uses the activity dependent noise $a := a_0 + a_1 N$, (*see* [11]), or if a refractory state is included [9]. In the inhibitory case, blow-up never occurs when noise is independent of the activity [10, 11] and solutions are globally bounded. This holds even for $\sigma(N) = N$ and assumption (4) is not fundamental in the inhibitory case. Then, the main open problem which remains is to prove the long time convergence; only small perturbations of the linear case are treated so far.

The initial data $p^0(v, w) \geq 0$ is a probability density and a basic property of the above LIF model is that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{V_F} p(v,w,t)\,dvdw = \int_{-\infty}^{\infty}\int_{-\infty}^{V_F} p^0(v,w)\,dvdw = 1. \tag{5}$$

We finally define the probability density of neural subnetworks with synaptic weight w by

$$H(w,t) = \int_{-\infty}^{V_F} p(v,w,t)\,dv, \qquad \int_{-\infty}^{\infty} H(w,t)dw = 1. \tag{6}$$

Let us mention that, so far, the function H is independent of time because in Equation (1), the distribution of synaptic weights is fixed. Moreover, with this distribution H, an input signal $I(w)$ is stored as a *normalized output signal* which we define thanks to the network activity as

$$S(w) = \frac{N(w)}{\bar{N}} \geq 0, \qquad \int_{-\infty}^{+\infty} S(w)dw = 1. \tag{7}$$

The normalization is due to the size of the network normalized by $\int H(w)dw = 1$ which induces a limitation of possible outputs.
Let us now include some learning rules that may modify this distribution.

2.2 Models with mean field learning rules.

Next, we introduce some learning rules in order to modulate the distribution of synaptic weights H and allow the network to recognize some given input signals I by choosing an appropriate heterogeneous synaptic weight distribution H adapted to the signal I. To this end, we have chosen learning rules inspired from the seminal Hebbian rule which essentially consists in assuming that the strength of weights w_{ij} between two neurons i and j increases when the two neurons have high activity simultaneously. For M neurons in interactions, the classical Hebbian rule relates the weights to the activity N_i of the neuron i

$$\frac{d}{dt}w_{ij} = k_{ij}N_iN_j - w_{ij}, \quad 1 \leq i,\ j \leq M.$$

In our context, we assume that the subnetworks interact only via the total firing rate $\bar{N}$, with synaptic weights described with a single parameter w, not a matrix. Hence, we cannot generalize directly the Hebbian learning rule and we give the following interpretation. All the subnetworks parametrized by w may modulate their intrinsic synaptic weight w with respect to a function Φ which depends on the intrinsic activity $N(w)$ of the network parametrized by w and of the total activity of the network $\bar{N}$. Then, the proposed generalization of the Hebbian rule consists in choosing

$$\Phi(N(w), \bar{N}) = \bar{N}N(w)K(w), \tag{8}$$

where $K(\cdot)$ represents the learning strength of the subnetwork with synaptic weight w.

Adding the above choice of learning rule, we obtain the following equation

$$\begin{cases} \frac{\partial p}{\partial t} + \frac{\partial}{\partial v}\left[\big(-v + I(w) + w\sigma(\bar{N}(t))\big)p\right] + \varepsilon\frac{\partial}{\partial w}\left[\big(\Phi - w\big)p\right] - a\frac{\partial^2 p}{\partial v^2} = N(w,t)\delta(v - V_R), \\ N(w,t) := -a\frac{\partial p}{\partial v}(V_F, w, t) \geq 0, \qquad \bar{N}(t) = \displaystyle\int_{-\infty}^{\infty} N(w,t)dw, \end{cases} \tag{9}$$

with the boundary and initial conditions

$$p(V_F, w, t) = 0, \qquad p(-\infty, w, t) = 0, \qquad p(v, \pm\infty, t) = 0, \qquad p(v, w, 0) = p^0(v, w).$$

Here, ε stands for a time scale which takes into account that learning is slower than the normal activity of the network, and $\Phi = \Phi\big(N(w,t), \bar{N}(t)\big)$ represents the learning rule. Notice that a desirable property is that the flus $\Phi - w$ is inward which occurs for instance when Φ is bounded or sub-linear at infinity. Several direct extensions of the Hebbian rule are possible for example

$$\Phi = N(w,t)\int K(w,w')N(w',t)dw',$$

or inspired from STDP rule (spike timing dependent plasticity, see [21] for instance), where post- and pre-synaptic spike times are compared, we may choose

$$\Phi\big(N(w,t),\bar{N}(t)\big) = \Phi\big(\big(N(w,t) *_t g\bar{N}(t) - N(w,t)\bar{N}(t) *_t g\big).$$

Here $g(t) = 0$ for $t < 0$ and $g(t) = e^{-t/\tau}$ for $t > 0$.

3 Stationary solution of the nonlinear problem without learning rule

Throughout this paper we use some material about the possible stationary solutions of Equations (1)–(3) submitted to a given input signal $I(w)$. These stationary states are defined through the equation

$$\frac{\partial}{\partial v}\left[\big(-v + I(w) + w\sigma(\bar{N})\big)P(v,w)\right] - a\frac{\partial^2 P(v,w)}{\partial v^2} = N(w)\delta(v - V_R), \tag{10}$$

with the boundary conditions

$$P(V_F,w) = 0, \quad P(-\infty,w) = 0, \quad \text{and} \quad N(w) = -a\frac{\partial P}{\partial v}(V_F,w) \geq 0. \tag{11}$$

We recall that the nonlinearity is driven by the network total activity defined as

$$\bar{N} = \int N(w)dw, \tag{12}$$

and that we assume a normalization (5) which is written

$$\int_{-\infty}^{V_F} P(v,w)dv = H(w), \qquad \int H(w)dw = 1. \tag{13}$$

Our first result is the

Theorem 3.1 (Existence of stationary states) *We assume* (4)*, give the input signal* $I \in L^\infty(\mathbb{R})$ *and the synaptic weight distribution* $H(w)$ *normalized to* 1 *such that there exists* $\varepsilon > 0$ *with*

$$\int_0^\infty w^2 H(w)dw < \infty. \tag{14}$$

Then, there is at least one solution of (10)–(13)*.*
In the case of inhibitory network, that is supp$(H) \subset (-\infty, 0]$*, for all* $\sigma \in \mathcal{C}^2$*, the solution is unique.*

Let us mention that for a single w, semi-explicit formula are available, see [8] for instance, and the stationary states are not necessarily unique in the excitatory case.

Proof. Our approach is to solve the nonlinear problem using a fixed point argument on the value $\bar{N}$. Being given $\bar{N}$, $I(w)$ and $H(w)$, we consider the linear problem where w is a parameter (we do not repeat the boundary conditions)

$$\begin{cases} \frac{\partial}{\partial v}\left[\big(-v + I(w) + w\sigma\big(\bar{N}\big)\big)Q_{\bar{N},I}\right] - a\frac{\partial^2 Q_{\bar{N},I}}{\partial v^2} = N_{\bar{N},I}(w)\delta(v - V_R), \\ \\ Q_{\bar{N},I}(V_F,w) = 0, \quad N_{\bar{N},I}(w) = -a\frac{\partial Q_{\bar{N},I}(V_F,w)}{\partial v}, \quad \int_{-\infty}^{V_F} Q_{\bar{N},I}(v,w)dv = 1. \end{cases} \tag{15}$$

Let $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ defined by

$$\psi(\bar{N}) = \int_{-\infty}^{+\infty} N_{\bar{N},I}(w) H(w) dw.$$

Then, we obtain a solution of (10)–(12) if and only if $\bar{N}$ is a fixed point of the application Ψ, that is

$$\int_{-\infty}^{+\infty} N_{\bar{N},I}(w) H(w) dw = \bar{N}, \qquad \text{and then } P(v,w) = H(w) Q_{\bar{N},I}(v,w). \tag{16}$$

To prove the existence of such a fixed point, we need a careful analysis of the mapping $\bar{N} \mapsto N_{\bar{N},I}(w)$, which we perform in the next subsection, adding many properties that will be used later on. The conclusion of the proof is given afterwards.

3.1 Main properties of $N_{\bar{N},I}(w)$ in (15)

Being given $\bar{N}$, the linear stationary state Equation (15), because it is solved w by w, is a standard equation and solutions form a one dimensional vector space (the eigenspace for the eigenvalue 0) according to the Krein-Rutman theorem [14]. Uniqueness is enforced thanks to the normalization as a probability.

Integrating Equation (15), we obtain that a solution satisfies

$$\big(-v + I(w) + w\sigma(\bar{N}) \big) Q_{\bar{N},I}(v,w) - a \frac{\partial Q_{\bar{N},I}(v,w)}{\partial v} = \begin{cases} 0 & \text{for } \; v < V_R, \\ N_{\bar{N},I}(w) & \text{for } \; v > V_R, \end{cases} \tag{17}$$

with $N_{\bar{N},I}(w)$ to be found such that $\int_{-\infty}^{V_F} Q_{\bar{N},I}(v,w) dv = 1$. Hence, the solution is explicitly given by

$$Q_{\bar{N},I}(v,w) = \begin{cases} \frac{1}{Z_{\bar{N},I}(w)} e^{-\frac{(v - I(w) - w\sigma(\bar{N}))^2}{2a}} \displaystyle\int_{V_R}^{V_F} e^{\frac{(v' - I(w) - w\sigma(\bar{N}))^2}{2a}} dv' & \text{for } \; v < V_R, \\ \frac{1}{Z_{\bar{N},I}(w)} e^{-\frac{(v - I(w) - w\sigma(\bar{N}))^2}{2a}} \displaystyle\int_{v}^{V_F} e^{\frac{(v' - I(w) - w\sigma(\bar{N}))^2}{2a}} dv' & \text{for } \; v > V_R, \end{cases} \tag{18}$$

with

$$Z_{\bar{N},I}(w) = \int_{-\infty}^{V_F} \int_{v' = \max(V_R, v)}^{V_F} e^{\frac{(v' - I(w) - w\sigma(\bar{N}))^2 - (v - I(w) - w\sigma(\bar{N}))^2}{2a}} dv' dv$$

which is also written under the more convenient form

$$Z_{\bar{N},I}(w) = \int_{-\infty}^{V_F} \int_{v' = \max(V_R, v)}^{V_F} e^{\frac{(v' - v).[v' + v - 2I(w) - 2w\sigma(\bar{N})]}{2a}} dv' dv. \tag{19}$$

Because of the boundary condition $Q_{\bar{N},I}(V_F, w) = 0$, an immediate consequence of (17) is the relation

$$N_{\bar{N},I}(w) = \frac{a}{Z_{\bar{N},I}(w)} = -a \frac{\partial Q_{\bar{N},I}(V_F, w)}{\partial v}. \tag{20}$$

Throughout the paper, we use properties of the function $Z_{\bar{N},I}(w)$ which we state in the following lemma.

Lemma 3.2 *Given $I \in L^\infty(\mathbb{R})$ and $\bar N > 0$, the unique solution $Q_{\bar N,I}(v,w) > 0$ of Equation (15), defined by (18)-(19), satisfies the following estimates. There is a constant $C(a, V_R, V_F)$ such that*

$$C \min\left(\frac{1}{\|I\|_{L^\infty} + |w|_+\sigma(\bar N)}, (V_F - V_R)\right)^2 \le Z_{\bar N,I}(w) \le Ce^{\frac{\left(\|I\|_{L^\infty}+|w|_-\sigma(\bar N)\right)^2}{a}}, \tag{21}$$

$$\lim_{w\to+\infty} Z_{\bar N,I}(w) = 0, \quad \lim_{w\to-\infty} Z_{\bar N,I}(w) = +\infty, \quad \inf_{\bar N} Z_{\bar N,I}(w) > 0, \quad \forall w \in \mathbb{R}, \tag{22}$$

$$\forall w \le 0, \quad \partial_{\bar N} Z_{\bar N,I}(w) \ge 0 \quad \textit{and} \quad \forall w \ge 0, \quad \partial_{\bar N} Z_{\bar N,I}(w) \le 0, \tag{23}$$

$$\partial_w Z_{\bar N,I}(w) \le 0, \quad \textit{if } I'(\cdot) \ge 0. \tag{24}$$

Moreover, when $\sigma(N) = N$, the following estimates holds

$$\lim_{-w\bar N\to+\infty} Z_{\bar N,I}(w) = +\infty, \tag{25}$$

$$\partial^2_{\bar N\bar N} Z_{\bar N,I}(w) > 0, \quad \forall w \le 0. \tag{26}$$

Proof of Lemma 3.2. We first prove inequality (21). We set $A = \|I\|_{L^\infty} + |w|_-\sigma(\bar N)$ where $|w|_- = -\min(0,w)$. As

$$e^{\frac{2(v'-v)(-I(w)-w\sigma(\bar N))}{2a}} \le e^{\frac{2(V_F-v)A}{2a}}, \quad \text{for } V_F \ge v' \ge v,$$

we obtain, with formula (19), that there exists a constant C such that

$$Z_{\bar N,I}(w) \le \int_{-\infty}^{V_F}\int_{v'=\max(V_R,v)}^{V_F} e^{\frac{|v'|^2-|v|^2+2(V_F-v)A}{2a}}\,dv'dv \le C\int_{-\infty}^{V_F} e^{\frac{-|v|^2+2(V_F-v)A}{2a}}\,dv.$$

Therefore, we conclude the upper bound with

$$Z_{\bar N,I}(w) \le Ce^{\frac{2V_FA}{2a}}\int_{-\infty}^{V_F} e^{\frac{-|v|^2-2vA}{2a}}\,dv = Ce^{\frac{V_F^2+A^2+A^2}{2a}}\int_{-\infty}^{V_F} e^{\frac{-|v|^2-2vA-A^2}{2a}}\,dv.$$

For the lower bound, we set

$$|w|_+ = \max(0,w) \text{ and } \widetilde A = \|I\|_{L^\infty} + |w|_+\sigma(\bar N),$$

and conclude that

$$e^{-\frac{(V_F-v)\widetilde A}{a}} \le e^{\frac{2(v'-v)(-I(w)-w\sigma(\bar N))}{2a}} \quad \text{for } V_F \ge v' \ge v.$$

Inserting this lower bound in formula (19), we deduce that there exists a constant $C(a, V_R, V_F)$ such that

$$Z_{\bar N,I}(w) \ge \int_{-\infty}^{V_F}(V_F-\max(V_R,v))e^{\frac{-|v|^2-2(V_F-v)\tilde A}{2a}}dv \ge C\int_{V_R}^{V_F}(V_F-v)e^{\frac{-(V_F-v)\tilde A}{a}}dv = C\int_0^{V_F-V_R} ze^{\frac{-z\tilde A}{a}}dz.$$

We deduce that there exists a constant $C(a, V_R, V_F)$ such that

$$Z_{\bar N,I}(w) \ge C\min(\frac{1}{\widetilde A^2}, (V_F - V_R)^2),$$

which ends the proof of estimate (21).
Next, we prove the inequality (22). We have, for all $w \in \mathbb{R}$,

$$\begin{aligned}\int_{-\infty}^{V_F}\int_{v'=\max(V_R,v)}^{V_F} & e^{\frac{(v'-v).(v'+v-2\|I\|_{L^\infty}-2w\sigma(\bar N))}{2a}}dv'dv \le Z_{\bar N,I}(w)\\ &\le \int_{-\infty}^{V_F}\int_{v'=\max(V_R,v)}^{V_F} e^{\frac{(v'-v).(v'+v+2\|I\|_{L^\infty}-2w\sigma(\bar N))}{2a}}dv'dv.\end{aligned} \tag{27}$$

Because $\sigma > 0$ for $\bar N \neq 0$, we have almost everywhere in v and v'

$$\lim_{w\to-\infty} e^{-2w\sigma(\bar N)(v'-v)} = +\infty \quad \text{and} \quad \lim_{w\to+\infty} e^{-2w\sigma(\bar N)(v'-v)} = 0.$$

To prove the estimate (23), we differentiating the explicit formula of $Z_{\bar N,I}(w)$ with respect to $\bar N$. We obtain that

$$\partial_{\bar N} Z_{\bar N,I}(w) = -w\sigma'(\bar N)\int_{v=-\infty}^{V_F}\int_{v'=\max(V_R,v)}^{V_F} \frac{v'-v}{a} e^{\frac{(v'-I(w)-w\sigma(\bar N))^2}{2a}} e^{-\frac{(v-I(w)-w\sigma(\bar N))^2}{2a}} dv\,dv',$$

which directly gives (23).

To prove (24), we observe that, when $I'(w) \ge 0$,

$$\partial_w Z_{\bar N,I}(w) = -(I'(w)+\sigma(\bar N))\int_{v=-\infty}^{V_F}\int_{v'=\max(V_R,v)}^{V_F} \frac{v'-v}{a} e^{\frac{(v'-v)(v'+v-2I(w)-2w\sigma(\bar N))}{2a}} dv\,dv' \le 0.$$

Next, when $\sigma(N) = N$, we have

$$Z_{\bar N,I}(w) \ge \int_{v=w\bar N}^{w\bar N+1} e^{-\frac{(v-I(w)-w\bar N)^2}{2a}} \left(\int_{v'=\max(v,V_R)}^{V_F} e^{\frac{(v'-I(w)-w\bar N)^2}{2a}} dv'\right) dv.$$

Therefore, we obtain estimate (25) because for $v \in (w\bar N, w\bar N+1)$, we have

$$e^{-\frac{(v-I(w)-w\bar N)^2}{2a}} \ge e^{-\frac{(\|I\|_{L^\infty}+2)^2}{2a}}, \quad \text{so that} \quad Z_{\bar N,I}(w) \ge e^{-\frac{(\|I\|_{L^\infty}+2)^2}{2a}} \int_{v'=V_R}^{V_F} e^{\frac{(v'-I(w)-w\bar N)^2}{2a}} dv'.$$

Finally, the inequality (26) follows from

$$\partial^2_{\bar N\bar N} Z_{\bar N,I} = \frac{w^2}{a^2}\int_{v=-\infty}^{V_F}\int_{v'=\max(v,V_R)}^{V_F} (v'-v)^2 e^{-\frac{(v-I(w)-w\sigma(\bar N))^2}{2a}} e^{\frac{(v'-I(w)-w\sigma(\bar N))^2}{2a}} dv'dv > 0.$$

The proof of Lemma 3.2 is complete.

□

3.2 Conclusion of the proof of Theorem 3.1

We come back to the fixed point equation (16). With the above notations, it is restated, using the quantity $Z_{\bar N,I}(w)$ defined by (19), as a fixed point of the function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$

$$\psi(\bar N) := a\int_{-\infty}^{+\infty} \frac{H(w)}{Z_{\bar N,I}(w)} dw = \bar N. \tag{28}$$

We have $\psi(0) > 0$.
In the inhibitory case, when $\mathrm{supp}(H) \subset \mathbb{R}^-$, we have $\psi'(\cdot) < 0$ thanks to (23), and thus there is a unique fixed point.
When excitatory weights are considered, as

$$\lim_{w\to+\infty} \frac{1}{\bar{Z}_{\bar{N},I}(w)} = 0,$$

we have to impose an additional assumption on H to control $\psi(\cdot)$. For this purpose, using estimate (21), thanks to the assumption (14), we know that $\psi(\cdot)$ remains bounded and hence, there is at least one fixed point by continuity. □

4 Output signals induced from a synaptic weight distribution

As a first property of the network properties, we aim at identifying which possible steady states output activities $N(w)$ or signal $S(w)$ can be generated by the network by varying the synaptic weight distribution H.

We prove that any nonnegative normalized output signal $S \in L^1(\mathbb{R})$, with fast decay at $-\infty$, can be, up to a multiplicative constant, reproduced by a stationary state of Equation (10) for a well chosen synaptic weight distribution H.

Theorem 4.1 (Relation output signal to synaptic weights) *We assume* (4) *and give the input signal* $I \in \mathcal{C}^1_{\mathrm{b}}(\mathbb{R})$ *and the output signal* $S \geq 0$ *normalized with* $\int_{-\infty}^{+\infty} S(w) = 1$ *and satisfyning* $\int_{-\infty}^{0} e^{-\gamma w} S(w) < \infty$ *with* $\gamma > \sigma_M \frac{V_F - V_R}{a}$. *Then, we can find a synaptic weight distribution* $H(w)$ *normalized to* 1, *such that* (7) *holds true for a solution of* (10)–(13).

In the case of inhibitory signal, that means supp$(S) \subset (-\infty, 0]$, *for all* $\sigma \in \mathcal{C}^2$, *the synaptic weight distribution* H *and* $\bar{N}$ *are unique.*

Proof. Consider an output normalized signal $S(w) \geq 0$. Using the notations of section 3, the relations (7) are reduced to building a distribution $H(w)$ normalized to 1, such that the following relations hold

$$N(w) = H(w) N_{\bar{N},I}(w), \qquad H(w) = \bar{N} \frac{S(w)}{N_{\bar{N},I}(w)}, \tag{29}$$

and the fixed point condition (16) is automatically satisfied thanks to the normalization of S.

In other words, we look for $\bar{N}$ such that

$$H(w) = \bar{N} \frac{Z_{\bar{N},I}(w)}{a} S(w) \quad \text{ and } \quad \int_{-\infty}^{+\infty} H(w) dw = 1.$$

Let us mention that $H \in L^1$ because of estimate (22) and the integrability condition on $S(\cdot)$.

These conditions are reduced to achieve the value a by the mapping $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ defined by

$$\psi(\bar{N}) = \bar{N} \int_{-\infty}^{+\infty} S(w) Z_{\bar{N},I}(w) dw = a.$$

We have obviously $\psi(0) = 0$. Moreover, using the last estimate of (22), we obtain that

$$\psi(\bar{N}) \geq \bar{N} \left(\int_{-\infty}^{+\infty} S(w) \inf_{\bar{N}} Z_{\bar{N},I}(w) dw \right) \text{ with } \int_{-\infty}^{+\infty} S(w) \inf_{\bar{N}} Z_{\bar{N},I}(w) dw > 0$$

and so $\psi(+\infty) = +\infty$ which implies existence of the activity $\bar{N}$ satisfying the desired nonlinearity.

In the inhibitory case, we notice that ψ is increasing as a consequence of (23) and uniqueness follows.

The proof of Theorem 4.1 is complete. □

5 Discrimination property

A desired property of the input-ouput map is to be able to discriminate between signals. In our language, it is to say that two different input signals $I(w)$ will generate two different network activities $N(w)$. To state a more precise result, we need the notation, for two bounded input currents I and J,

$$L(w) = e^{\frac{-(\|I\|_{L^\infty} + |w| - \sigma_M)^2 - (\|J\|_{L^\infty} + |w| - \sigma_M)^2}{a}}.$$

The discrimination property is a consequence of the functional inequality

$$\int |N_I(w) - N_J(w)| dw \geq C \int L(w) H(w) |I(w) - J(w)| dw, \tag{30}$$

for two solutions of (10)–(13). The network under consideration has this property.

Theorem 5.1 (Discrimination property) *We consider two bounded input currents I and J. Being given normalized synaptic weights $H(w)$ such that $\int_0^{+\infty} wH(w)dw < +\infty$. We define*

$$\nu = \|\sigma'\|_\infty \int |w| L(w) H(w) dw < +\infty$$

Then, the discrimination inequality (30) *holds true with a positive constant*

$$C = C(\nu, a, V_F, V_R, \|I\|_\infty, \|J\|_\infty, \sigma_M).$$

Proof. We denote by $\bar{M}$ a total activity obtained via Equations (10)–(13) stemming from the current J (that is the input current is given by J), and by $\bar{N}$ a total activity obtained via Equations (10)–(13) stemming from the current I.

We have

$$N_I(w) - N_J(w) = H(w)[N_{\bar{N},I}(w) - N_{\overline{M},J}(w)] = H(w)[\frac{a}{Z_{\bar{N},I}(w)} - \frac{a}{Z_{\bar{M},J}(w)}].$$

Therefore using the upper bound (21), we find that there exists a constant C such that

$$|N_I(w) - N_J(w)| = aH(w) \frac{|Z_{\bar{N},I}(w) - Z_{\bar{M},J}(w)|}{Z_{\bar{N},I}(w) Z_{\bar{M},J}(w)} \geq CL(w)H(w)|Z_{\bar{N},I}(w) - Z_{\bar{M},J}(w)|. \tag{31}$$

To go further, we set $\alpha = I(w) + w\sigma(\bar{N})$, $\beta = J(w) + w\sigma(\bar{M})$, and write, using (19),

$$\begin{aligned} Z_{\bar{N},I}(w) - Z_{\bar{M},J}(w) &= \int_{-\infty}^{V_F} \int_{v'=\max(V_R,v)}^{V_F} \left[e^{\frac{(v'-v).(v'+v-2\alpha)}{2a}} - e^{\frac{(v'-v).(v'+v-2\beta)}{2a}} \right] dv'dv \\ &= -\int_{-\infty}^{V_F} \int_{v'=\max(V_R,v)}^{V_F} \int_{\alpha}^{\beta} \frac{v'-v}{a} e^{\frac{(v'-v).(v'+v-2\gamma)}{2a}} d\gamma dv'dv. \end{aligned}$$

We assume, without lose of generality, that $\beta > \alpha$. Then, we have

$$\begin{aligned} \left| Z_{\bar{N},I}(w) - Z_{\bar{M},J}(w) \right| &\geq \int_{-\infty}^{V_F} \int_{v'=V_R}^{V_F} \int_{\alpha}^{\beta} \frac{v'-v}{a} e^{\frac{(v'-v).(v'+v-2\gamma)}{2a}} dv'dvd\gamma \\ &\geq \int_{-\infty}^{V_R-a} \int_{v'=V_R}^{V_F} \int_{\alpha}^{\beta} \frac{v'-v}{a} e^{\frac{(v'-v).(v'+v-2\gamma)}{2a}} dv'dvd\gamma \end{aligned}$$

As for all $v \in (-\infty, V_R - a)$ and $v' \in (V_R, V_F)$, it holds $\frac{v'-v}{a} \geq 1$, we obtain that, with

$$\gamma_\infty = \|I\|_\infty + \|J\|_\infty + R\sigma_M,$$

$$\begin{aligned} \left| Z_{\bar{N},I}(w) - Z_{\bar{M},J}(w) \right| &\geq \int_{\alpha}^{\beta} \int_{-\infty}^{V_R-a} \int_{v'=V_R}^{V_F} e^{\frac{(v'-v)(v'+v-2\gamma_\infty)}{2a}} dv'dvd\gamma \\ &\geq C(a, V_R, V_F, \gamma_\infty)\ (\alpha - \beta). \end{aligned}$$

We can now go back to (31) and obtain

$$\begin{aligned} |N_I(w) - N_J(w)| &\geq CL(w)H(w)|I(w) + w\sigma(\bar{N}) - J(w) - w\sigma(\bar{M})| \\ &\geq CL(w)H(w)\big[|I(w) - J(w)| - |w|\,\|\sigma'\|_\infty |\bar{N} - \bar{M}|\big]. \end{aligned}$$

As a consequence, with our definition of ν, we find

$$\int |N_I(w) - N_J(w)|dw \geq C \int L(w)H(w)|I(w) - J(w)|dw - C\nu|\bar{N} - \bar{M}|$$

and because

$$|\bar{N} - \bar{M}| = \left| \int N_I(w) - N_J(w)dw \right| \leq \int |N_I(w) - N_J(w)|\, dw$$

we finally obtain

$$\int |N_I(w) - N_J(w)|dw \geq \frac{C}{1 + C\nu} \int L(w)H(w)|I(w) - J(w)|dw.$$

Theorem 5.1 is proved. □

6 Synaptic weight distribution stemming from a learning rule

We now study how a learning rule defines a specific synaptic weight distribution. We assume that our network is submitted to the simplified Hebbian learning rule (8) with a function $K(\cdot)$ which is piecewise $\mathcal{C}^1$, discontinuous at 0 and satisfies

$$wK(w) > 0 \text{ for } w \neq 0, \quad 0 < K_m \leq |K(\cdot)| \leq K_M \text{ a.e. .} \tag{32}$$

The sign condition is just to impose that inhibitory and excitatory neurons may change weight but remain in the same status. We also give a bounded input signal. Our aim here is to study possible distributions of synaptic weights generated by the pair (K, I). We point out a specific difficulty which motivates the more thorough analysis in the next section,

The model we work on is the steady state equation

$$\begin{cases} \frac{\partial}{\partial v}\left[\big(-v + I(w) + w\sigma(\bar{N})\big)p\right] + \varepsilon\frac{\partial}{\partial w}\left[\big(K(w)N(w)\bar{N} - w\big)p\right] - a\frac{\partial^2 p}{\partial v^2} = N(w)\delta(v - V_R), \\ \\ N(w) = -a\frac{\partial p(V_F, w)}{\partial v}, \quad \bar{N} = \displaystyle\int_{-\infty}^{+\infty} N(w)dw, \quad H(w) = \int_{-\infty}^{V_F} p(v, w)dv, \quad \int_{-\infty}^{\infty} H(w)dw = 1, \end{cases} \tag{33}$$

with the boundary conditions

$$p(V_F, w) = 0, \qquad p(-\infty, w) = 0, \qquad p(v, \pm\infty) = 0. \tag{34}$$

We give an input signal $I(w)$ and the learning rule $K(w)$ which selects a distribution H. Which are the possible synaptic weights $H(w)$?

We show that many distributions of synaptic weights are possible and solutions of Equation (33) are far from unique. We recall the definition of $Q_{\bar{N},I}(w, v)$ and $N_{I,\bar{N}}(w)$ in Equation (15) and state the

Theorem 6.1 (Weight distribution induced by learning) *Let $I \in \mathcal{C}_b(\mathbb{R})$ and K satisfy (32). Then, there exists infinitely many solutions $\overline{P}(w, v) \geq 0$ of Equation (33) independent of ε. They are given by*

$$\overline{P}(v, w) = H_{\bar{N},A}(w)Q_{\bar{N},I}(w, v), \qquad \textit{with} \qquad H_{\bar{N},A}(w) = \frac{w}{\bar{N}K(w)N_{\bar{N},I}(w)}\mathbb{I}_A(w) \geq 0,$$

for some appropriate subsets $A \subset \mathbb{R}$ such that

$$\int_{-\infty}^{+\infty} H_{\bar{N},A}(w)dw = 1. \tag{35}$$

Notice that this non-uniqueness theorem yields the question to find an organizing principle which selects the synaptic weight among the large class built in the proof of Theorem 6.1. This is the topic of Section 7.

Proof of Theorem 6.1. The strategy of proof is as for Theorem 4.1 and we look for a fixed point for the total activity $\bar{N}$ to build a solution of the nonlinear Equation (33).

Therefore we fix a value $\bar{N} > 0$ and a bounded subset of $A \subset \mathbb{R}$. Let the synaptic weights $H_{\bar{N},A}(w)$ be defined by

$$H_{\bar{N},A}(w) = \frac{w}{\bar{N}K(w)N_{\bar{N},I}(w)}\mathbb{I}_A(w) \geq 0,$$

the sign being a consequence of the sign condition on K. One readily checks that, with $\overline{P}$ defined in Theorem 6.1, we have $(N(w, t)K(w)\bar{N} - w)\overline{P} = 0$ and thus $\overline{P}$ is indeed a solution of Equation (9).

It remains to solve the fixed point, that is to find $\bar{N}$ such that the following condition holds:

$$\int_{w=-\infty}^{+\infty} H_{\bar{N},A}(w)dw = 1 \quad \text{and} \quad \bar{N} = \int H_{\bar{N},A}(w)N_{I,\bar{N}}(w)dw.$$

That is also written, recalling the notation (20),

$$\int_A \frac{wZ_{\bar{N},I}(w)}{K(w)}dw = a\bar{N} \quad \text{and} \quad \bar{N}^2 = \int_A \frac{w}{K(w)}dw. \tag{36}$$

Because, in Theorem 6.1, the condition for the choice of the set A is the only constraint (35), to conclude its proof it is enough to give a specific construction in the inhibitory case, which is addressed in Proposition 6.2 and hence the proof of Theorem 6.1 is finished assuming Proposition 6.2. □

6.1 Solutions for learning with inhibitory weights only.

The proof of Theorem 6.1 can be concluded choosing the set A so as to select inhibitory weights only.

Proposition 6.2 (Inhibitory weights) *There exists infinitely many steady states $\overline{P}(v,w)$ of Equation (33) independent of ε which supports in the variable w are union of intervals of $(-\infty,0)$. In particular there is one of the form*

$$H_{\bar{N}}(w) = \frac{w}{K(w)\bar{N}N_{\bar{N},I}(w)}\mathbb{I}_{-\varphi\le w<0},$$

and it is unique in the case where $I'(w) \ge 0$ and where the saturation is neglected, that is $\sigma(N) = N$.

Proof of Proposition 6.2. We first observe that if the support of $\overline{P}(v,w)$ is equal to $A = (-\varphi(\bar{N}), 0)$, then, with the second relation in (36), necessarily $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ must be defined by

$$\bar{N}^2 = \int_{-\varphi(\bar{N})}^{0} -\frac{w}{K(w)}dw. \tag{37}$$

There exist $C_1 > 0$ and $C_2 > 0$ such that this function φ satisfies

$$\varphi(0) = 0, \qquad C_1 N \le \varphi(N) \le C_2 N, \qquad C_1 \le \varphi'(N) \le C_2. \tag{38}$$

The first two statements are immediate and the third one follows, differentiating (37) and using (32), from the identity

$$2\bar{N} = -\varphi'(\bar{N})\frac{\varphi(\bar{N})}{K(-\varphi(\bar{N}))}. \tag{39}$$

Secondly, the value $\bar{N} > 0$ has to satisfy

$$\Phi(\bar{N}) = \int_{-\varphi(\bar{N})}^{0} \frac{w}{K(w)}Z_{\bar{N},I}(w)dw = a\bar{N}. \tag{40}$$

Our goal is to prove that there exists a positive solution of Equation (40) and that it is unique when $\sigma(N) = N$. To do this, let us compute the first two derivatives of Φ. We have

$$\Phi'(\bar{N}) = -\varphi'(\bar{N})\frac{\varphi(\bar{N})}{K(-\varphi(\bar{N}))}Z_{\bar{N},I}(-\varphi(\bar{N})) + \int_{-\varphi(\bar{N})}^{0} \frac{w}{K(w)}\partial_{\bar{N}}Z_{\bar{N},I}(w)dw.$$

Using identity (39), we obtain that

$$\Phi'(\bar{N}) = 2\bar{N} Z_{\bar{N},I}(-\varphi(\bar{N})) + \int_{-\varphi(\bar{N})}^{0} \frac{w}{K(w)} \partial_{\bar{N}} Z_{\bar{N},I}(w) dw.$$

In particular, $\Phi'(0) = 0$. Using Lemma 3.2, we obtain that there exists $C > 0$ such that for all $w \leq 0$ and $\bar{N} > 0$,

$$Z_{\bar{N},I}(w) > C \quad \text{and} \quad \partial_{\bar{N}} Z_{\bar{N},I}(w) \geq 0.$$

Hence, there exists $C > 0$ such that

$$\Phi'(\bar{N}) \geq C\bar{N}.$$

As $\Phi'(0) = \Phi(0) = 0$, there exists at least one nonnegative solution to (32).

To prove that there is a unique one in the case where $\sigma(N) = N$, it suffices to show that $\Phi'' > 0$. Using (24), identity (39) and Lemma 3.2, we obtain that

$$\begin{aligned}\Phi''(\bar{N}) =& 2Z_{\bar{N},I}(-\varphi(\bar{N})) - 2\bar{N}\varphi'(\bar{N})\partial_w Z_{\bar{N},I}(-\varphi(\bar{N})) + 4\bar{N}\partial_{\bar{N}} Z_{\bar{N},I}(-\varphi(\bar{N})) \\ &+ \int_{-\varphi(\bar{N})}^{0} \frac{w}{K(w)} \partial_{\bar{N}\bar{N}} Z_{\bar{N},I}(w) dw > 0\end{aligned}$$

To prove that there exists infinitely many steady states of Equation (33), we notice that there exists infinitely many other choices than $\mathbb{I}_{0 \leq w \leq -\varphi(\bar{N})}$ of subintervals such that the same proof holds. An example is, given $w_0 \leq 0$, to consider the function $\widetilde{\varphi}(\bar{N})$ such that

$$H_{\bar{N}}(w) = \frac{-w}{\bar{N} N_{\bar{N},I}(w)} \mathbb{I}_{w_0 \geq w \geq -\widetilde{\varphi}(\bar{N})}, \qquad N(w) = N_{\bar{N},I}(w) H_{\bar{N}}(w)$$

where $\widetilde{\varphi}(\bar{N})$ is the value determined by

$$\int_{-\infty}^{0} H_{\bar{N}}(w) dw = 1.$$

The above argument applies directly and this concludes the proof of Proposition 6.2 and of Theorem 6.1. □

6.2 Non existence result for learning with excitatory weights only

One might try the same approach and try to find purely excitatory weights. This is not always possible and we have the

Proposition 6.3 (Excitatory weights) *We take for $w > 0$ a bounded signal I. When $\sigma(N) = \sigma_0 \frac{N}{1+N}$ with σ_0 large enough, there is no solution of (33) with a weight distribution under the form*

$$H_{\bar{N},A}(w) = \frac{w}{\bar{N} K(w) N_{\bar{N},I}(w)} \mathbb{I}_A(w), \qquad \text{with} \quad A = (0, \varphi).$$

This Proposition implies that, for the purely excitatory case, we may not have convergence of the solution of Equation (33) to a stationary state. As an example, in the situation of Proposition 6.3, to hope having convergence to a stationary state, we have to deal with an initial condition where the support of H and $(-\infty, 0)$ is non empty.

Proof. Using the same proof as for Proposition 6.2, we impose, still because of the conditions in (36),

$$\bar{N}^2 = \int_0^{\varphi(\bar{N})} \frac{w}{K(w)} dw.$$

and the properties (38) still hold. According to the other condition in (36), we examine the condition

$$a\bar{N} = \int_0^{\varphi(\bar{N})} \frac{w}{K(w)} Z_{\bar{N},I}(w) dw := \Phi(\bar{N}).$$

We notice that

$$\Phi'(\bar{N}) = \varphi'(\bar{N}) \frac{\varphi(\bar{N})}{K(\varphi(\bar{N}))} Z_{\bar{N},I}(\varphi(\bar{N})) + \int_0^{\bar{N}^2} \partial_{\bar{N}} Z_{\bar{N},I}(w) dw.$$

Since $\Phi'(0) = 0$, we expect that, if there was a fixed point N_0, then the first fixed point will be such that $\Phi'(N_0) \geq 1$. However at such a fixed point, we find, because $\partial_{\bar{N}} Z_{\bar{N},I}(w) < 0$ for $w > 0$,

$$\Phi'(N_0) \leq \varphi'(N0) \frac{\varphi(N_0)}{K_M} \int_{-\infty}^{V_F} \int_{v'=\max(V_R,v)}^{V_F} e^{\frac{(v'-v).[v'+v+2\|I\|_\infty - 2\sigma_0 \frac{N_0 \varphi(N_0^2)}{1+N_0})]}{2a}} dv' dv.$$

From the properties (38), we conclude that for σ_0 large enough, we have necessary

$$\Phi'(N_0) < 1,$$

which is a contradiction and concludes the proof. □

7 Selection of inhibitory synaptic weights by noise

In this section, we choose $K(w) = -1$ for $w \leq 0$ and we only consider inhibitory interconnections. In view of the result of Section 6, we try to find a selection principle for the synaptic weight distribution which would single out the choice $A = (-\varphi, 0)$ established in Proposition 6.2. Indeed, among the infinitely many steady states constructed in Theorem 6.1, numerical evidence, in Section 8, indicates that a unique stationary state is selected with $A = (-\sqrt{2}\,\bar{N}, 0)$.

Two difficulties occur, namely the selection of the set A and the uniqueness of the value $\bar{N}$ when solving the fixed point (36). As in the inhibitory case, blow-up does not occur for Leaky Integrate and Fire models [11], we simply assume that $\sigma\big(\bar{N}(t)\big) = \bar{N}(t)$.

A possible organizational principle can be noise, which is compatible with numerical diffusion in the observations of Section 6. Therefore, we heuristically study the stationary state of a modified equation with a Gaussian noise, of intensity $\nu > 0$, on the variable w. We use slow-fast limit in order to take into account that learning is on a slow scale compared to neural activity. Then, we may compute more easily the potential stationary states of the new equation given by

$$\begin{aligned} \varepsilon \tfrac{\partial p_\epsilon}{\partial t} + \tfrac{\partial}{\partial v}\big[\big(-v + I(w) + w\bar{N}_\epsilon(t)\big)p_\epsilon\big] + \quad & \varepsilon \tfrac{\partial}{\partial w}\big[\big(-N_\epsilon(w,t)\bar{N}_\epsilon(t) - w\big)p_\epsilon\big] \\ & -a\tfrac{\partial^2 p_\epsilon}{\partial v^2} - \varepsilon\nu \tfrac{\partial^2 p_\epsilon}{\partial w^2} = N_\epsilon(w,t)\delta(v - V_R), \end{aligned} \tag{41}$$

with boundary conditions adapted to the purpose of dealing only with $w \leq 0$,

$$p_\epsilon(V_F, w, t) = 0, \qquad p_\epsilon(-\infty, w, t) = 0, \qquad \left(N_\epsilon(w,t)\bar{N}_\epsilon(t) + w\right) p_\epsilon = -\nu \frac{\partial p_\epsilon}{\partial w}, \quad \text{at} \;\; w = 0. \tag{42}$$

Here $\varepsilon > 0$ represents the time scale of learning and thus vanishes in the limit of fast network adaptation vs slow learning.

In a first and formal step in our analysis, we consider the fast time scale $\varepsilon \to 0$. This yields the steady state Integrate and Fire density distribution as studied in Section 3. Since the synaptic weight distribution takes a value $\widetilde{H}(w)$ that changes according to the slow time scale, we fix it here and find

$$p^*[\widetilde{H}] = \widetilde{H}(w) Q_{\bar{N}[\widetilde{H}],I}(v, w)$$

with $Q_{\bar{N},I}$ defined through (15), (16). Then, $\bar{N}[\widetilde{H}]$ is solution of the fixed point equation

$$\bar{N}[\widetilde{H}] = \int_0^\infty \widetilde{H}(w) N_{\bar{N}[\widetilde{H}],I}(w) dw$$

with $N_{\bar{N}[\widetilde{H}],I}$ is defined by Equation (20).

In a second step, we can integrate in v Equation (41) and divide by ε. Recalling that, from (6), $\widetilde{H}_\epsilon(w,t) = \int_{-\infty}^{V_F} p_\epsilon(v,w,t) dv$, we obtain

$$\frac{\partial \widetilde{H}_\epsilon}{\partial t} + \frac{\partial}{\partial w}\left(\left(-\widetilde{N}_\epsilon(w)\bar{N}_\epsilon(t) - w\right)\widetilde{H}_\epsilon\right) = \nu \frac{\partial^2 \widetilde{H}_\epsilon}{\partial w^2}.$$

With the equilibrium of the first step, we find the limit

$$\frac{\partial \widetilde{H}}{\partial t} + \frac{\partial}{\partial w}\left(\left(-\widetilde{N}(w)\bar{N}(t) - w\right)\widetilde{H}\right) = \nu \frac{\partial^2 \widetilde{H}}{\partial w^2}, \quad w \leq 0 \tag{43}$$

where

$$\widetilde{N}(w,t) = N_{\bar{N}[\widetilde{H}(.,t)],I}\widetilde{H}(w,t), \quad N_{\bar{N}[\widetilde{H}(.,t)],I}(w,t) := -a\frac{\partial}{\partial v} Q_{\bar{N}[\widetilde{H}(.,t)]}(V_F, w), \quad \bar{N}(t) = \bar{N}[\widetilde{H}(.,t)], \tag{44}$$

and with the no-flux boundary condition

$$\left(\widetilde{N}(w,t)\bar{N}(t) + w\right)\widetilde{H} = -\nu \frac{\partial \widetilde{H}}{\partial w}, \quad \text{at} \;\; w = 0. \tag{45}$$

Notice that the form of $\widetilde{N}(w,t)$ makes that Equation (43) is nonlinear hyperbolic, closely related to first order scalar conservation laws, [12, 27]. Therefore, we may expect that discontinuities (shocks) can be formed and that noise selects indeed a specific solution, namely the entropy solution. This is stated in the following Theorem:

Theorem 7.1 (Small noise limit) *Assume that $I'(w) \geq 0$. As $\nu \to 0$, the steady state of Equation* (43)– (45) *converges to the unique steady state built in Proposition 6.2 and supported by the single interval* $[-A, 0]$.

Proof. After integrating the equation for the steady states of (43), and using the boundary condition at $w = 0$, we find that each stationary state $\widetilde{H}_\nu$ satisfies

$$\left(\widetilde{H}_\nu N_{\bar{N}[\widetilde{H}_\nu],I}(w)\bar{N} + w\right)\widetilde{H}_\nu = -\nu\frac{d\widetilde{H}_\nu}{dw}. \tag{46}$$

We first observe that solutions $\widetilde{H}_\nu$ of such an equation cannot vanish at a point because they are given by an exponential.
Next, we claim that there is $w_\nu < 0$ such that

$$\widetilde{H}'_\nu(w) > 0 \text{ for } w < w_\nu, \quad \text{and} \quad \widetilde{H}'_\nu(w) < 0 \text{ for } w > w_\nu. \tag{47}$$

Indeed, since $\widetilde{H}_\nu(0) > 0$, for w close to zero, we have $\widetilde{H}_\nu N_{\bar{N}[\widetilde{H}_\nu],I}(w)\bar{N} + w > 0$ and thus $\widetilde{H}'_\nu(w) < 0$. Because $\widetilde{H}_\nu$ is integrable on $(-\infty, 0)$, there has to be a largest value w_ν where

$$0 = \widetilde{H}_\nu N_{\bar{N}[\widetilde{H}_\nu],I}(w_\nu)\bar{N} + w_\nu \quad \text{and} \quad 0 = \widetilde{H}'_\nu(w_\nu).$$

Finally, on $(-\infty, w_\nu)$ we necessarily have $\widetilde{H}'_\nu(w) > 0$ because $\frac{-w}{N_{\bar{N}[\widetilde{H}_\nu],I}(w)}$ is a decreasing function thanks to the assumption $I'(w) > 0$ which implies $N'_{\bar{N}[\widetilde{H}_\nu],I} > 0$ using (24) because $N_{\bar{N}[\widetilde{H}_\nu],I} = \frac{a}{Z_{\bar{N}[\widetilde{H}_\nu],I}}$. Therefore, if there was a second crossing point $w_1 < w_\nu$, where $\widetilde{H}_\nu N_{\bar{N}[\widetilde{H}_\nu],I}(w_1)\bar{N} = w_1$, we should have both $\widetilde{H}'_\nu(w_1) < 0$ (to cross a decreasing function) and, the condition $\widetilde{H}'_\nu(w_1) = 0$ from (46). A contradiction which states (47).
From this property, that $\int_{-\infty}^{0}\widetilde{H}_\nu(w)dw = 1$ and the control from below and above of $N_{\bar{N}[\widetilde{H}_\nu],I}(w)$ using Lemma 3.2, we conclude that $\widetilde{H}_\nu$ is uniformly bounded and has the uniform decay $\exp(-\frac{w^2}{2\nu})$ as $w \to -\infty$. Therefore we may pass to the limit in (46) and conclude that the limit $\widetilde{H}$ satisfies either $\widetilde{H}(w) = 0$, or $\widetilde{H}(w)N_{\bar{N}[\widetilde{H}],I}(w)\bar{N} + w = 0$. From (47), this identifies the support of $\widetilde{H}$ as stated in Theorem 7.1. □

8 Learning, testing and pattern recognition

Based on numerical simulations, we illustrate the discrimination property stated in Section 5. We consider the following two-phase setting:

Learning phase

1. An heterogeneous input $I(w)$ is presented to the system, while the learning process is active. The chosen initial data is supported on inhibitory weights so as to avoid the complexity of excitatory cases and the learning rule is determined for the inhibitory weights by $-N(w)\bar{N}$, as in section 6, by taking $K(w) = -1$ if $w \leq 0$.

2. After some time, the synaptic weight distribution $H(w,t)$ converges to an equilibrium distribution $H^*_I(w)$, which depends on I.

Testing phase

1. The learning process is now switched off, and a new input $J(w)$ is presented to the system.

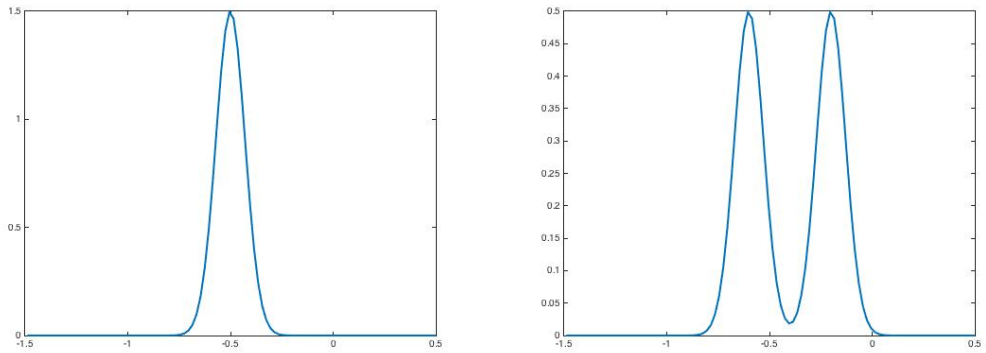

Figure 1: (Two input signals) We have used two input signals that we denote $I(w)$ on the left and $J(w)$ on the right.

2. After some time, the solution $p_J(v, w, t)$ reaches an equilibrium $p_J^*(v, w)$, which can be summarized by the output signal $N_J^*(w)$ which is the neural activity distribution across the heterogeneous populations.

The numerics has been performed using a finite difference method. For the Fokker-Planck equation on the potential, we use the Sharfetter-Gummel method [24]. For the transport equation on the weight variable, we use an upwind scheme [4, 25]. The matlab code is available on demand to one of the authors.

Then, from the mathematical analysis performed in previous sections, we know that the following "pattern recognition" property will be observed: *the system can detect whether the new input $J(w)$ is actually the same one that has been presented during the learning phase, i.e. $I(w)$: indeed, in this case, $N_J^*(w) = w\mathbf{1}_{[-A,0]}$ has a very specific shape.* A remarkable feature is that this specific shape does not depend upon the original input I that has been learned in the learning phase: it is an intrinsic property of the system. This is particularly interesting because it implies that detecting a learned pattern could be implemented by an external system which would be independent of the given pattern.

To illustrate this pattern recognition property we display in Figure 1 the two input signals we have used for the learning-testing set-up

$$I(w) = 1.5 * \exp\left(-\frac{(w+0.5)^2}{0.01}\right), \quad J(w) = .5 * \exp\left(-\frac{(w+0.2)^2}{0.01}\right) + .5 * \exp\left(-\frac{(w+0.6)^2}{0.01}\right).$$

After presentation of these input currents $I(w)$ and $J(w)$, the synaptic weight distribution converges to H_I^* that are displayed in Figure 2. The corresponding network activity $N(w)$ are shown in Figure 3. During the testing phase, learning is off and the system reacts differently according to the input it receives: if the new input is the same as the learned one, then the neural activity distributes according to the specific shape predicted by the theory and already shown in Figure 3, indicating that the network has recognized the learned pattern. Whereas if the new input is not the same, here we invert I and J as input currents, then the neural activity distributions have a very different shape Figure 4. This illustrates the discrimination property.

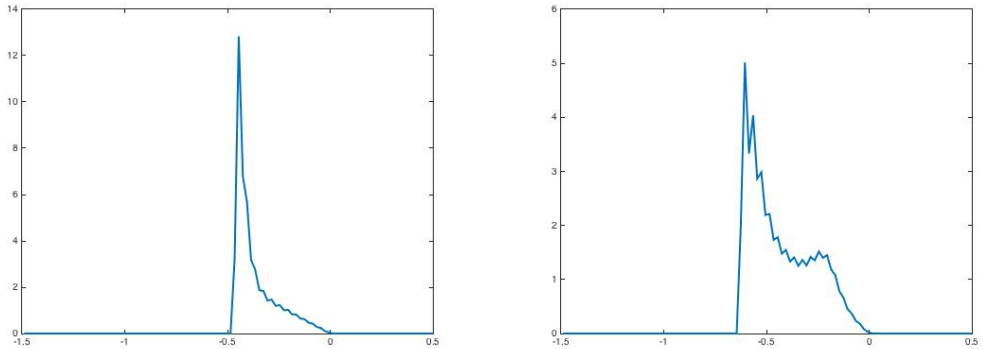

Figure 2: (Two connectivities) This figure displays the synaptic connectivities obtained after learning with the input signals of Figure 1.

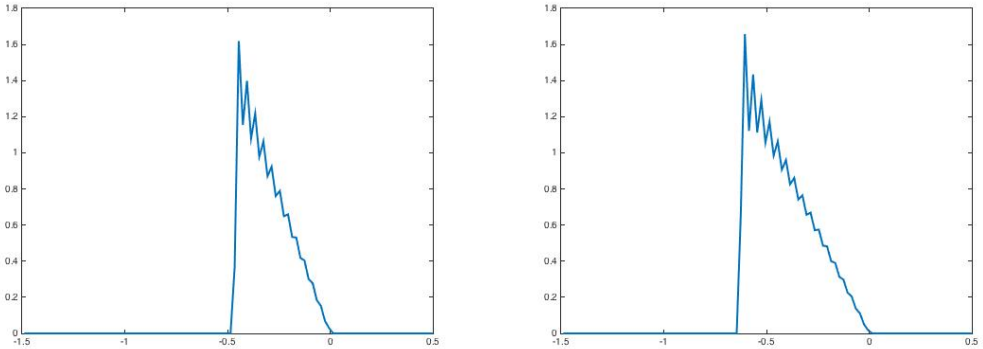

Figure 3: (Output with learned signal) This figure displays the netwotk activities $N(w)$ when learned with the two input signals of Figure 1, I on the left, J on the right.

9 Conclusions and perspectives

We have introduced a novel mathematical framework to study learning mechanisms in macroscopic models of spiking neuronal networks by considering plasticity between neural subpopulation and the overall mean-field activity. When ignoring the learning rule, we have characterized the synaptic weight distribution which generates a given output signal, and we have shown a discrimination property. When the learning rule is activated, we have studied the multiple synaptic weight equilibria of the global coupled system with learning. A selection by noise selects a unique equilibria which is also observed numerically. Furthermore, we have investigated the ability of such models to perform pattern recognition tasks.

The class of models studied in this article are subject to several limitations and mainly that the network is coupled via a global activity and not by pairwise interactions. A related limitation is that stability and convergence to a unique equilibrium point depend on the excitatory/inhibitory nature of the synaptic weight as it does for the noisy integrate-and-fire network model. Because we have targeted mathematically proved results, we had to assume that the input signal is time independent, which is a restriction in the theory.

To further extend our study, one should investigate other learning rules. A possible extension is to

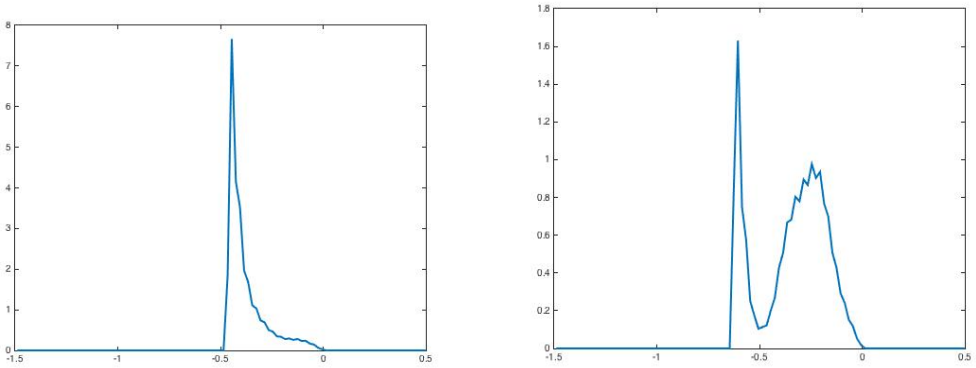

Figure 4: (Output with the other signal) With the synaptic weights of the learned signal I, we display the activty for the input signal J (left), and for learned signal J we display the activty for the input signal I (left).

use pairwise connections, leading to the following extension of our system

$$\begin{cases} \frac{\partial p}{\partial t} + \frac{\partial}{\partial v}\Big[\big(-v + I(w) + \int_{-\infty}^{\infty} C(w,w',t)N(w',t)dw'\big)p\Big] - a\frac{\partial^2 p}{\partial v^2} = N(w,t)\delta(v - V_R), \\ N(w,t) := -a\frac{\partial p}{\partial v}(V_F, w, t) \geq 0, \qquad \bar{N}(t) = \displaystyle\int_{-\infty}^{\infty} N(w,t)dw, \end{cases} \tag{48}$$

$$\frac{\partial}{\partial t}C(w,w',t) = K(w,w')N(w,t)N(w',t) - C(w,w',t).$$

In closer connection with biological mechanisms such as spike-timing dependent plasticity, which may also be integrated in the model with convolution operators. Other models of neuronal dynamics, beyond spiking models, such as rate models or coupled oscillator systems, could also be studied and compared within the proposed formalism.

Finally, to make the link with the fields of pattern recognition and machine learning deeper, further questions can be considered, for instance to quantify the discrimination ability between two signals or to evaluate the number and complexity of attractors, possibly dynamic, which can be stored into the synaptic weight distribution.

Acknowledgment: BP and DS are supported by the french "ANR blanche" project Kibord: ANR-13-BS01-0004.

References

[1] G. Ben Arous and A. Guionnet, *Large deviations for langevin spin glass dynamics*, Probability Theory and Related Fields, 102 (1995), pp. 455–509.

[2] L. Bertini, G. Giacomin, and K. Pakdaman, *Dynamical aspects of mean field plane rotators and the kuramoto model*, Journal of Statistical Physics, 138 (2010), pp. 270–290.

[3] M. Bossy, O. Faugeras, and D. Talay, *Clarification and complement to "Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons"*, J. Math. Neurosci., 5 (2015), pp. Art. 19, 23.

[4] F. Bouchut, *Non linear stability of finite volume methods for hyperbolic conservation laws and well balanced schemes for sources*, Birkhaüser-Verlag, 2004.

[5] R. Brette and W. Gerstner, *Adaptive exponential integrate-and-fire model as an effective description of neural activity*, Journal of neurophysiology, 94 (2005), pp. 3637–3642.

[6] N. Brunel, *Dynamics of sparsely connected networks of excitatory and inhibitory spiking networks*, J. Comp. Neurosci., 8 (2000), pp. 183–208.

[7] N. Brunel and V. Hakim, *Fast global oscillations in networks of integrate-and-fire neurons with long firing rates*, Neural Computation, 11 (1999), pp. 1621–1671.

[8] M. J. Cáceres, J. A. Carrillo, and B. Perthame, *Analysis of nonlinear noisy integrate & fire neuron models: blow-up and steady states*, Journal of Mathematical Neuroscience, 1-7 (2011).

[9] M. J. Cáceres and B. Perthame, *Beyond blow-up in excitatory integrate and fire neuronal networks: refractory period and spontaneous activity*, J. Theoret. Biol., 350 (2014), pp. 81–89.

[10] J. A. Carrillo, M. d. M. González, M. P. Gualdani, and M. E. Schonbek, *Classical solutions for a nonlinear fokker-planck equation arising in computational neuroscience*, Comm. in Partial Differential Equations, 38 (2013), pp. 385–409.

[11] J. A. Carrillo, B. Perthame, D. Salort, and D. Smets, *Qualitative properties of solutions for the noisy integrate and fire model in computational neuroscience*, Nonlinearity, 28 (2015), pp. 3365–3388.

[12] C. Dafermos, *Hyperbolic conservation laws in continuum physics*, in Grundlehren der Mathematischen Wissenschaften, vol. 325, Springer-Verlag, Berlin, 2000.

[13] P. Dai Pra and F. den Hollander, *Mckean-vlasov limit for interacting random processes in random media*, Journal of statistical physics, 84 (1996), pp. 735–772.

[14] R. Dautray and J.-L. Lions, *Mathematical analysis and numerical methods for science and technology. Vol. 6*, Springer-Verlag, Berlin, 1993. Evolution problems. II, With the collaboration of Claude Bardos, Michel Cessenat, Alain Kavenoky, Patrick Lascaux, Bertrand Mercier, Olivier Pironneau, Bruno Scheurer and Rémi Sentis, Translated from the French by Alan Craig.

[15] P. Dayan and L. F. Abbott, *Theoretical neuroscience*, vol. 806, Cambridge, MA: MIT Press, 2001.

[16] A. De Masi, A. Galves, E. Löcherbach, and E. Presutti, *Hydrodynamic limit for interacting neurons*, J. Stat. Phys., 158 (2015), pp. 866–902.

[17] F. Delarue, J. Inglis, S. Rubenthaler, E. Tanré, et al., *Global solvability of a networked integrate-and-fire model of mckean–vlasov type*, The Annals of Applied Probability, 25 (2015), pp. 2096–2133.

[18] M. Galtier and G. Wainrib, *Multiscale analysis of slow-fast neuronal learning models with noise*, The Journal of Mathematical Neuroscience, 2 (2012), pp. 1–64.

[19] M. N. Galtier and G. Wainrib, *A biological gradient descent for prediction through a combination of stdp and homeostatic plasticity*, Neural computation, 25 (2013), pp. 2815–2832.

[20] W. Gerstner and W. Kistler, *Mathematical formulations of hebbian learning*, Biological cybernetics, 87 (2002), pp. 404–415.

[21] W. Gerstner and W. M. Kistler, *Spiking neuron models: Single neurons, populations, plasticity*, Cambridge university press, 2002.

[22] D. O. Hebb, *The organization of behavior: A neuropsychological approach*, John Wiley & Sons, 1949.

[23] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proceedings of the national academy of sciences, 79 (1982), pp. 2554–2558.

[24] A. Jüngel, *Transport equations for semiconductors*, in Lecture Notes in Physics, vol. 773, Springer-Verlag, Berlin Heidelberg, 2009.

[25] R. J. LeVeque, *Finite volume methods for hyperbolic problems*, Cambridge University Press, 2002.

[26] S. Martin, P. Grimwood, and R. Morris, *Synaptic plasticity and memory: an evaluation of the hypothesis*, Annual review of neuroscience, 23 (2000), pp. 649–711.

[27] D. Serre, *Systems of conservation laws, I*, Cambridge University Press, 1999.

[28] H. Sompolinsky, A. Crisanti, and H. Sommers, *Chaos in random neural networks*, Physical Review Letters, 61 (1988), p. 259.