Mean field system of a two-layer neural model in a diffusive regime

Xavier Erny*

Abstract

We study a model of interacting neurons. This neural system is composed of two layers of neurons such that the neurons of the first layer send their spikes to the neurons of the second one: if N is the number of neurons of the first layer, at each spiking time of the first layer, every neuron of both layers receives an amount of potential of the form $U/\sqrt{N}$, where U is a centered random variable. This kind of structure of neurons can model a part of the structure of the visual cortex: the first layer represents the primary visual cortex V1 and the second one the visual area V2. The model consists of two stochastic processes, one modelling the membrane potential of the neurons of the first layer, and the other the membrane potential of the neurons of the second one. We prove the convergence of these processes as the number of neurons N goes to infinity and obtain a convergence speed. The proofs rely on similar arguments as those used in [17]: the convergence speed of the semigroups of the processes is obtained from the convergence speed of their infinitesimal generators using a Trotter-Kato formula, and from the regularity of the limit semigroup.

1 Introduction

1.1 Model

The model we study in this paper is a mean-field model of interacting neurons. Such models have been studied widely in the literature to model the dynamics of neural networks (see e.g. [20, 2, 32, 10] and the references therein). A natural way of modelling the activity of the neurons is to use point processes to represent their spike trains. In this context, Hawkes processes have called attention as a natural model for the spike trains of neurons. These processes have been introduced in [24] to model the aftershocks of earthquakes. By definition, Hawkes processes are mutually exciting point processes, therefore they can be naturally used to model the mutually exciting/inhibiting behavior of

---

*Université Paris-Saclay, CNRS, Univ Evry, Laboratoire de Mathématiques et Modélisation d'Evry, 91037, Evry, France
   E-mail: xav.erny@orange.fr

neurons (see [35, 22, 36]). Some authors have also managed to model the repolarization of the neurons using Hawkes processes with variable length memory (see [21, 25, 14]). Finally "age dependent Hawkes processes" introduced in [7] and further studied in [8, 5] can be used to make the spike rate of a neuron depend on the delay since its last spike, allowing to consider for example a refractory period for the neurons.

This paper aims to study a model of two layers of neurons such that:

- the neurons of the first layer send their spikes to the neurons of both the first and second layers,
- the spikes of the neurons of the second layer do not affect these two layers.

We note N the number of neurons of the first layer. The idea of our modelling is to study the "averaged effect" of the first layer on a single neuron of the second one. In the following, we model this averaged effect through the averaged value of the membrane potentials of the neurons of the first layer. Let us note X_t^N the averaged value of these potential at time t, and Y_t^N the membrane potential of one particular neuron of the second layer at time t.

We assume that the dynamics of the processes X^N and Y^N are given by the following SDEs:

$$X_t^N = X_0^N - \alpha_1 \int_0^t X_s^N ds + \frac{1}{\sqrt{N}} \sum_{j=1}^N \int_{[0,t] \times \mathbb{R}_+ \times \mathbb{R}} u \cdot \mathbb{1}_{\left\{z \leq f_1(X_{s-}^N)\right\}} d\pi^{(1),j}(s,z,u) \quad (1.1)$$

$$- \frac{1}{N} \sum_{j=1}^N \int_{[0,t] \times \mathbb{R}_+ \times \mathbb{R}} X_{s-}^N \mathbb{1}_{\left\{z \leq f_1(X_{s-}^N)\right\}} d\pi^{(1),j}(s,z,u),$$

$$Y_t^N = Y_0^N - \alpha_2 \int_0^t Y_s^N ds + \frac{1}{\sqrt{N}} \sum_{j=1}^N \int_{[0,t] \times \mathbb{R}_+ \times \mathbb{R}} u \cdot \mathbb{1}_{\left\{z \leq f_1(X_{s-}^N)\right\}} d\pi^{(1),j}(s,z,u)$$

$$- \int_{[0,t] \times \mathbb{R}_+} Y_{s-}^N \mathbb{1}_{\left\{z \leq f_2(Y_{s-}^N)\right\}} d\pi^{(2)}(s,z),$$

where X_{t-}^N (resp. Y_{t-}^N) denotes the limit from the left of X^N (resp. Y^N) at time t. Above, α_1, α_2 are positive parameters (resp. f_1, f_2 are non-negative functions) of the model that represent the leakage rate characteristics (resp. the spike rate functions) of respectively the neurons of the first layer and the neurons of the second one. In (1.1), $\pi^{(1),j}$ are i.i.d. Poisson measures with intensity $ds \cdot dz \cdot d\nu(u)$ (ν is a centered probability measure on $\mathbb{R}$), and $\pi^{(2)}$ a Poisson measure with Lebesgue intensity independent of the other Poisson measures.

Remark 1.1. Note that it is possible to rewrite equation (1.1) in a more compact form:

$$X_t^N = X_0^N - \alpha_1 \int_0^t X_s^N ds + \frac{1}{\sqrt{N}} \int_{[0,t] \times \mathbb{R}_+ \times \mathbb{R}} u \cdot \mathbb{1}_{\left\{z \leq f_1(X_{s-}^N)\right\}} d\pi^{(1)}(s,z,u)$$

$$- \frac{1}{N} \int_{[0,t] \times \mathbb{R}_+ \times \mathbb{R}} X_{s-}^N \mathbb{1}_{\left\{z \leq f_1(X_{s-}^N)\right\}} d\pi^{(1)}(s,z,u),$$

$$Y_t^N = Y_0^N - \alpha_2 \int_0^t Y_s^N ds + \frac{1}{\sqrt{N}} \int_{[0,t] \times \mathbb{R}_+ \times \mathbb{R}} u \cdot \mathbb{1}_{\left\{z \leq f_1(X_{s-}^N)\right\}} d\pi^{(1)}(s,z,u)$$

$$- \int_{[0,t] \times \mathbb{R}_+} Y_{s-}^N \mathbb{1}_{\left\{z \leq f_2(Y_{s-}^N)\right\}} d\pi^{(2)}(s,z),$$

where $\pi^{(1)}$ is a Poisson measure with intensity $N \cdot ds \cdot dz \cdot d\nu(u)$. It is formally possible to define $\pi^{(1)}$ as the union of the $\pi^{(1),j}$ ($1 \leq j \leq N$). In the rest of the paper, we use the form (1.1) because it is more understandable for the model.

In (1.1), the counting processes

$$P_t^{(1),N,i} := \int_{[0,t] \times \mathbb{R}_+ \times \mathbb{R}} \mathbb{1}_{\left\{z \le f_1(X_{s-}^N)\right\}} d\pi^{(1),i}(s, z, u), \quad 1 \le i \le N,$$

and

$$P_t^{(2),N} := \int_{[0,t] \times \mathbb{R}_+} \mathbb{1}_{\left\{z \le f_2(X_{s-}^N)\right\}} d\pi^{(2)}(s, z),$$

model respectively the spike trains of the neuron i of the first layer ($1 \le i \le N$) and that of the particular neuron of the second one. In this model, the form of the synaptic weights is $u/\sqrt{N}$ where u is the third variable of the Poisson measures $\pi^{(1),i}$ ($1 \le i \le N$). These variables can be seen as random variables that are i.i.d. ν−distributed, and whence centered. Note that it is necessary to consider centered synaptic weights because of the normalization $N^{-1/2}$, and in order to make appear these centered variables u, we need to write the point processes $P^{(1),N,i}$ as thinnings of Poisson measures in (1.1).

Let us explain in other words the dynamics of the equations (1.1). The second term of both SDEs describes the effect of the spikes of the neurons of the first layer on both layers. The last term in the SDE of Y^N (i.e. the fourth line of (1.1)) models the repolarization of the potential of the particular neuron of the second layer: right after the neuron emits a spike, its potential goes fast (here it jumps immediately) to its resting value that we assume to be zero. The last term of the SDE of X^N (i.e. the second line of (1.1)) should be interpreted as an averaged repolarization. Indeed, the neurons of the first layers fire at rate $N \cdot f_1(X_{t-}^N)$. So X_t^N should be interpreted as the averaged potential of the neurons of the first layer. And, according to our model, when a neuron of the first layer sends a spike, its potential should jump to zero. As a consequence, if we assume that this potential was "close" to the averaged value X_{t-}^N before the jump, then, right after the spike, the quantity lost by the averaged potential is $N^{-1}X_{t-}^N$.

The reason why we only consider the effect of the first layer of neurons through their averaged potential X^N is to guarantee our model (X^N, Y^N) to be two-dimensional, and not $(N+1)$−dimensional. In this second frame, the technics used in the proofs would fail. This if for the same reason that we assume our neural networks to be complete graphs, since it implies that the dynamics of all the neurons are similar.

The main result of this paper (Theorem 1.5) is the convergence in distribution as N goes to infinity (with an explicit convergence speed) of the sequence of the two-dimensional Markov processes $(X^N, Y^N)_N$ (defined in (1.1)) to the following limit process

$$\bar{X}_t = \bar{X}_0 - \alpha_1 \int_0^t \bar{X}_s ds + \sigma \int_0^t \sqrt{f_1(\bar{X}_s)} dW_s - \int_0^t \bar{X}_s f_1(\bar{X}_s) ds, \tag{1.2}$$

$$\bar{Y}_t = \bar{Y}_0 - \alpha_2 \int_0^t \bar{Y}_s ds + \sigma \int_0^t \sqrt{f_1(\bar{X}_s)} dW_s - \int_{[0,t] \times \mathbb{R}_+} \bar{Y}_{s-} \mathbb{1}_{\left\{z \le f_2(\bar{Y}_{s-})\right\}} d\bar{\pi}(s, z),$$

where W is a standard one-dimensional Brownian motion, $\bar{\pi}$ is a Poisson measure on $\mathbb{R}_+^2$ with Lebesgue intensity independent of W, and σ^2 is the variance of the probability measure ν.

Remark 1.2. Note that Theorem 1.5 has a practical interest if one wants to simulate the trajectories of the process (X^N, Y^N). According to this result, the limit process $(\bar{X}, \bar{Y})$ is an approximation of (X^N, Y^N), and it is easier to simulate. Indeed, to simulate (X^N, Y^N) one has to compute the trajectories of the N Poisson measures $\pi^{(1),i}$ ($1 \le i \le N$), which can be computationally hard if N is large. So it can be easier to simulate a single Brownian motion W instead of those Poisson measures.

Then, if we simulate the limit process $(\bar{X}, \bar{Y})$ instead of the "real" one (X^N, Y^N) (i.e. the one that actually corresponds to the neural model), we make an error. The

interesting point of Theorem 1.5 is to quantify this error (i.e. the convergence speed). Let us mention that this error is of order $N^{-1/2}$, which can be considered "small" if the first layer represents for example the primary visual cortex V1, since the number of neurons N in V1 is estimated to be of order 10^8 according to [30].

1.2 Biological motivation

This multi-layers structure is a natural model for the visual cortex, which is composed of five interacting and functionally distinct layers of neurons (see Chapter 5 of [26] for a detailed description of the structure of the visual cortex). The two-layer model of this paper can be seen as a model of the primary visual cortex V1 (the first layer) and the visual area V2 (the second layer). A more interesting model could consist in considering five layers to represent V1-V5 and more complex interactions between these layers. As we have mentioned earlier, to simplify the model we only consider an averaged potential X^N for the neurons of the first layer. We also assume that all the neurons of the second layer have the same dynamics, and that is why we just study one particular neuron of the second layer.

In our model, we assume our first layer to be a complete graph (i.e. each neuron is connected by a synapse to any other neuron). As it is explained above, this property is important in our proofs, but it is a drawback of the biological model since a neuron has only a limited number of synaptic connections compared to the number of neurons of the network. It is possible to extend this model to the case where the first layer is divided into a fix number k of sublayers. In this case, each sublayer has to be assumed to be a complete graph and one can consider any interaction structure between the sublayers. In this new model, assuming that the number k of sublayers does not depend on N, the results of the paper would hold true, and the techniques used to prove the results would be the same (the main difference is that X^N would be a $k-$dimensional process instead of an one-dimensional one). This model may be interesting to model the six functionally distinct layers of V1, and even the four sublayers of the fourth one of V1 (i.e. taking $k = 9$) (see e.g. Chapter 5 of [26]). In order to ease the readability of the proofs, we prefer not to study the "multi-sublayers model". Let us also mention that for similar reasons, it is possible to extend our model to study a fix number of neurons of the second layer and not only one (provided that this number does not depend on N).

Let us end this section by commenting the form of the synaptic weights $u/\sqrt{N}$ where the u are random variables. These centered synaptic weights model the fact that we consider "balanced networks" (i.e. neural networks where the excitatory inputs and the inhibitory ones are balanced). Experiments support this hypothesis of balanced synaptic connections as a natural model (see [38, 23] and the references therein). Let us also mention that, the experiments done in [4] support the assumptions that the order of magnitude of the synaptic weights is close to the theoretical value (i.e. the square root of the number of presynaptic neurons, since the interactions are balanced), which is $\sqrt{N}$ with our assumptions.

1.3 Mathematical techniques

Let us give a brief overview of the mathematical studies of balanced network: [19] have proved a large deviation principle for a discrete-time model with correlated synaptic weights, in [39] the authors have studied a continuous-time model of ODEs parametrized by random synaptic weights where at the limit a Gaussian field is created by these weights, [37] have studied the exsitence of a balanced state for the large scale limit of a similar model, and [34] have proved the convergence of Hawkes processes in distribution to a limit process that is solution to a stochastic convolution equation (also

known as stochastic Volterra equation of convolution type). The form of the model of [34] is close to the one we consider here. In these papers, the synaptic weight of a given synapse is a random variable and each time a neuron sends a spike to another neuron, the potential of this other neuron always receives the same random variable from this synapse. In our model (similarly as in [17, 14]), at each spiking time, we make a choice of random synaptic weight independently of the previous ones. This implies that the role of a synapse (i.e. excitatory or inhibitory) can change. This property is inconsistent with biological evidence, but compared to the model of [34], it allows us to obtain a convergence speed in addition to the convergence in distribution of the model, and it also allows us to consider more general distributions than just Rademacher distribution for the law of the synaptic weights.

Let us also note that this kind of multi-layers model has already been studied in [11] and at Section 5 of [15]. In both these references, the number of layers is not limited to two (but does not depend on N) and every membrane potential is modeled by a specific process without the "repolarization" effect in the equations. Consequently the dimension of the $N-$neurons model is N, and as we have claimed previously, the techniques of this paper would not work on these models. In [11], the authors have obtained a L^1-convergence with a convergence speed in $N^{-1/2}$. The processes they have studied are Hawkes processes and the order of synaptic weight is N^{-1}. The model in [15] is closer to (1.1) since the synaptic weight is also centered and of order $N^{-1/2}$. The advantage of the model of the current paper is that we obtain a convergence speed of order $N^{-1/2}$ whereas in [15], we have just proved a convergence in distribution and in [16] (where the results hold true in a multi-layers framework) the order of the convergence speed is $N^{-1/10}$. Note that it is interesting to have a good convergence speed for simulation purpose (see Remark 1.2).

The question of the convergence of system with this kind of repolarization effect has already been studied previously (in a one-layer framework): see [10, 1, 13] for models with synaptic weights of order N^{-1}, and [14] for synaptic weights of order $N^{-1/2}$. The novelty of this article consists in obtaining a convergence speed for the convergence in distribution where the order of the synaptic weights is $N^{-1/2}$.

We can also remark that (1.1) is a generalization of the model of [17] where we have studied the convergence of the one-dimensional process X^N in a simplified framework (i.e. without the last term in the SDE of X^N). The interesting points of the model of this paper is that it is more relevant for the modelling of a particular structure of neural network (e.g. the layers V1 and V2 of the virtual cortex), and that we can still prove similar results.

In order to quantify the convergence speed of the distribution of (X^N, Y^N) to the distribution of $(\bar{X}, \bar{Y})$, we rely on their semigroups, respectively denoted by $(P_t^N)_t$ and $(\bar{P}_t)_t$. By definition the semigroups are families of operators that can be defined on the space of continuous and bounded functions by the following: for g continuous and bounded, and $x, y \in \mathbb{R}$,

$$P_t^N g(x, y) := \mathbb{E}_{(x,y)} \left[g(X_t^N, Y_t^N) \right],$$

where $\mathbb{E}_{(x,y)}$ is the expectation related to the probability measure $\mathbb{P}_{(x,y)}$ under which the initial value of the process (X^N, Y^N) is given by the deterministic value $(x, y) = (X_0^N, Y_0^N)$.

To obtain this convergence speed, we use similar arguments as in [17]. We consider the infinitesimal generators A^N and $\bar{A}$ of respectively P^N and $\bar{P}$. These infinitesimal generators are operators that can be defined as the derivatives of their respective semigroups:

$$A^N g(x, y) = \lim_{t \to 0} \frac{1}{t} \left(P_t^N g(x, y) - g(x, y) \right).$$

The notion of generators can slightly differ from a reference to another for technical reasons. This is why we define precisely the notion that we use in this paper at the beginning of Section 2.1 (it is the same notion that we have introduced and studied in Appendix A of [17]).

We begin by proving the convergence of these generators and by obtaining a convergence speed. This convergence speed depends on the derivatives of the test-function g. Then, we deduce a convergence speed for the semigroups using the following formula:

$$\bar{P}_t g(x,y) - P_t^N g(x,y) = \int_0^t P_{t-s}^N \left(\bar{A} - A^N \right) \bar{P}_s g(x,y) ds.$$

This kind of techniques used to obtain the convergence of the semigroups from the convergence of the generators are similar to the results of the Section 1.6 of [18]. In particular, the formula above is proved in Lemma 1.6.2 of [18] with stronger assumptions on the generators compared to the result of this paper and [17].

To be able to use (2.1) to deduce a convergence speed for the semigroups from the one for the generators, we need to control the differentiability of the function $(x,y) \mapsto \bar{P}_s g(x,y)$. So, one of the main step of the proof of our main result consists in studying the regularity of the limit semigroup $(\bar{P}_t)_t$. For this purpose, we need to study the regularity of the related stochastic flow (see Section 1.4 for the notion of stochastic flow). Compared to the proofs of [17], there are two additional technical difficulties that we handle at Section 2.2:

- The trajectories of the process $(\bar{X}, \bar{Y})$ are not continuous. Whence, the regularity of its semigroup is harder to study.

- The jump times of the process $(\bar{X}, \bar{Y})$ occur at the rate $f_2(\bar{Y}_{t-})$. This implies that this rate depends on the initial condition of the process. The regularity of this dependency is hard to study directly.

To overcome the first point, we study the regularity of the stochastic flow between the jump times (see Proposition 2.7). Let us remark that [3] have already used the idea of working between the jump times of the flow to study its regularity and that of its semigroup. For the second point, we introduce an auxiliary limit process $(\tilde{X}, \tilde{Y})$ for which the rate of the jumps is constant. Then, we deduce the regularity of the semigroup $(\bar{P}_t)_t$ from the regularity of the flow of $(\tilde{X}, \tilde{Y})$ using Girsanov's theorem for jump processes that gives the explicit Radon-Nikodym derivative between the processes $(\bar{X}, \bar{Y})$ and $(\tilde{X}, \tilde{Y})$ (see Proposition 2.4).

1.4 Notation

Throughout the paper, we use the following notation:

- For $n, d \in \mathbb{N}^*$, we note $C_b^n(\mathbb{R}^d)$ the set of $\mathbb{R}-$valued functions defined on $\mathbb{R}^d$ that are C^n, bounded and such that all their partial derivatives up to order n are bounded.

- We call $\beta \in \mathbb{N}^2$ a multi-index and we note $|\beta| = \beta_1 + \beta_2$. If $g : \mathbb{R}^2 \to \mathbb{R}$ is C^n and if β is a multi-index such that $|\beta| \leq n$, we write $\partial_\beta g$ the partial derivative of g w.r.t. β. For the first order partial derivatives, we write instead $\partial_1 g$ and $\partial_2 g$, and for the second order derivatives, we write $\partial_{i,j}^2 g$ ($i,j \in \{1,2\}$). When there is no possible confusion, we may use the classical notation $\partial_x g(x,y) = \partial_1 g(x,y)$.

- For $g \in C_b^n(\mathbb{R}^2)$, we note

$$||g||_{n,\infty} = \sum_{|\beta| \leq n} ||\partial_\beta g||_\infty.$$

- Given a process X (resp. a multi-dimensional process (X, Y)) that is solution to some SDE, we note $X^{(x)}$ (resp. $(X^{(x,y)}, Y^{(x,y)})$) the process solution to the SDE w.r.t. the initial condition $X_0 = x$ (resp. $(X_0, Y_0) = (x, y)$). The function $x \mapsto X^{(x)}$ (resp. $(x, y) \mapsto (X^{(x,y)}, Y^{(x,y)})$) is called the stochastic flow of the process X (resp. (X, Y)). By definition, a stochastic flow of a Markov process represents its trajectories as functions of its initial condition (see Definition 1.1.1 of [29] for a more general definition). In this paper, we speak about stochastic flow instead of the trajectories of the process itself when we want to emphasize the fact that it is a function of the initial value, and to use the same terminology as [29].
- In all the paper, C denotes an arbitrary positive constant that is independent of any parameters except the parameters of the model. If a constant depends on some parameter θ, we write C_θ instead. The value of such constants can change from line to line in an equation.

1.5 Assumptions and main result

To begin with, let us state the conditions we need to study the model.

Assumptions 1.3.

(i) The function f_1 is $C_b^1(\mathbb{R})$, positive, and, for some constant $C > 0$, for any $x \geq 0$, $|f_1'(x)| \leq C/(1 + |x|)$,

(ii) the function f_2 is non-negative and belongs to $C_b^3(\mathbb{R})$,

(iii) the function $\sqrt{f_1}$ belongs to $C_b^4(\mathbb{R})$,

(iv) the probability measure ν is centered and admits a finite third moment. We note σ^2 its variance.

Remark 1.4. Any sigmoid function satisfies the conditions imposed by Assump. $1.3.(i) - (iii)$ on the functions f_1 and f_2. This kind of function has been shown to be natural for neural models (see [40]).

Under Assumption $1.3.(i) - (ii)$, the strong well-posedness of the SDEs (1.1) and (1.2) is a classical result since these two conditions guarantee that the coefficients are Lipschitz continuous and the jump rates bounded (see Theorem IV.9.1 of [27]).

The boundedness of f_1 and the assumption on f_1' are used to prove the well-posedness of the limit process (1.2) and to simplify some technical details, since it implies that the function $x \mapsto x f_1(x)$ is Lipschitz continuous. On the other hand, the boundedness of f_2 is important to control the Radon-Nikodym derivative of the limit process $(\bar{X}, \bar{Y})$ with respect to the auxiliary process $(\tilde{X}, \tilde{Y})$ that we introduce at Section 2.2 (see Proposition 2.4 for the form of this derivative).

Condition (iii) allows to prove that the stochastic flow of the auxiliary process is C^3 between its jump times (see Proposition 2.6). And condition (iv) allows us to obtain a convergence speed for the generators (see Lemma 2.3).

Under the additional assumption that (X_0^N, Y_0^N) converges in distribution to $(\bar{X}_0, \bar{Y}_0)$, it is known that the process (X^N, Y^N) converges in distribution to $(\bar{X}, \bar{Y})$ in Skorokhod topology (see Theorem IX.4.21 of [28]). The main result of this paper consists in obtaining the convergence speed of the semigroups P^N to the semigroup $\bar{P}$ of these processes (recalling that the processes (X^N, Y^N) and $(\bar{X}, \bar{Y})$ are defined respectively at (1.1) and (1.2), and that a semigroup can be defined as $P_t^N g(x, y) := \mathbb{E}[g(X_t^N, Y_t^N)|(X_0^N, Y_0^N) = (x, y)])$.

Theorem 1.5. Grant Assumptions $1.3.(i) - (iv)$. For all $t \geq 0$, $N \in \mathbb{N}^*$, $g \in C_b^3(\mathbb{R}^2)$ and $x, y \in \mathbb{R}$,

$$\left| \bar{P}_t g(x, y) - P_t^N g(x, y) \right| \leq C_t \cdot N^{-1/2} \cdot ||g||_{3,\infty} + C_t \cdot N^{-1} \left(1 + x^2 \right) \cdot ||g||_{2,\infty}.$$

Figure 1 shows two simulations of the process Y^N with $N = 100$ and $N = 1000$.
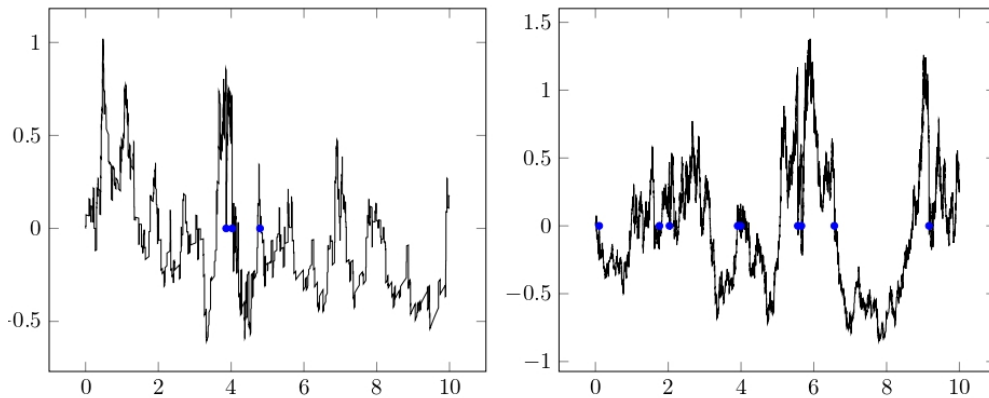
Figure 1: Simulation of trajectories of $(Y_t^N)_{0 \le t \le 10}$ with $X_0^N = 0$, $Y_0^N = 0$, $\alpha_1 = \alpha_2 = 1$, $\nu = \mathcal{N}(0,1)$, $f_1(x) = f_2(x) = 4/(1 + \exp(2(1-x)))$, $N = 100$ (left panel) and $N = 1000$ (right panel). The blue points are the reset jump times of the process Y^N.

2 Proof of Theorem 1.5

We prove Theorem 1.5 in three steps. In Section 2.1, we obtain an explicit convergence speed for the convergence of the generators A^N of (X^N, Y^N) to the generator $\bar{A}$ of $(\bar{X}, \bar{Y})$ (see Lemma 2.3). Then, we want to use this result using the following formula

$$\bar{P}_t g(x,y) - P_t^N g(x,y) = \int_0^t P_{t-s}^N \left(\bar{A} - A^N \right) \bar{P}_s g(x,y) ds, \qquad (2.1)$$

to deduce the convergence speed for the semigroups P^N and $\bar{P}$. The problem is that our bound for the convergence speed of $A^N g(x,y) - \bar{A} g(x,y)$ depends on the derivatives of the function g. So we need a control on the derivatives of the limit semigroup $(\bar{P}_t)_t$. This is done at Section 2.2 (Proposition 2.8). Finally, at Section 2.3, we apply the formula (2.1) to conclude the proof.

2.1 Convergence of the infinitesimal generators

One can note that the processes $(\bar{X}, \bar{Y})$ and (X^N, Y^N) ($N \in \mathbb{N}^*$) are Markov processes. In the following, we note $\bar{P}$ and P^N ($N \in \mathbb{N}^*$) their respective semigroups, and $\bar{A}$ and A^N ($N \in \mathbb{N}^*$) their respective infinitesimal generators. Usually, if Z is a $\mathbb{R}^2-$valued Markov process, we define its semigroup P by: for any $t \ge 0, z \in \mathbb{R}^2$, and any continuous and bounded function g,

$$P_t g(z) = \mathbb{E}_z \left[g(Z_t) \right],$$

where the index z indicates that we consider the law under which $Z_0 = z$.

The definition of the semigroup of a Markov process coincides in general to the one above, but the notion of infinitesimal generators can change from a reference to another. Here we consider a notion of extended generators (similar as the one used in [33, 9]). If Z is a $\mathbb{R}^2-$valued Markov process (whose generator will be denoted by A), we note $\mathcal{D}'(A)$ the set of continuous and bounded functions g for which it exists a function h_g that satisfies, for all $z \in \mathbb{R}^2, t \ge 0$,

$$\mathbb{E}_z \int_0^t |h_g(Z_s)| ds < \infty \quad \text{and} \quad \mathbb{E}_z \left[g(Z_t) \right] - g(z) = \mathbb{E}_z \int_0^t h_g(Z_s) ds.$$

If this function h_g exists, we note it $Ag := h_g$. Then A is an operator called the (extended) generator of Z. Note that the expression of a generator does not depend on the different definitions, what can change from a definition to another is the domain on which the generator is defined.

Example 2.1. Let X be a Markov process solution to the following one-dimensional SDE

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t + \int_{\mathbb{R}_+} \Phi(X_{t-}, u)\mathbb{1}_{\{z \leq f(X_{t-})\}} d\pi(t, z, u),$$

where W is a standard Brownian motion and π a Poisson measure on $\mathbb{R}_+^2 \times \mathbb{R}$ with intensity $dt \cdot dz \cdot d\nu(u)$. Then, Ito's formula allows to prove (provided some sufficient control on the functions b, σ, f, Φ) that $\mathcal{D}'(A)$ contains $C_b^2(\mathbb{R})$ and that

$$Ag(x) = b(x)g'(x) + \frac{1}{2}\sigma(x)^2 g''(x) + f(x) \int_{\mathbb{R}} [g(x + \Phi(x, u)) - g(x)] d\nu(u).$$

In this example we consider a one-dimensional SDE to keep things simple, but this can be generalized to any finite-dimensional SDE.

We refer to Appendix A of [17] for the study of some basic properties of this notion of generators (in particular for Proposition 5.6 where (2.1) is proved for another model).

Let us give the form of the generators A^N and $\bar{A}$ of the processes (X^N, Y^N) and $(\bar{X}, \bar{Y})$. The following lemma is a direct consequence of Ito's formula and of the form of the SDEs (1.1) and (1.2).

Lemma 2.2. *Grant Assumptions 1.3.$(i) - (ii)$. $\mathcal{D}'(A^N)$ contains $C_b^1(\mathbb{R}^2)$, and for $g \in C_b^1(\mathbb{R}^2)$ and $(x, y) \in \mathbb{R}^2$,*

$$\begin{aligned}
A^N g(x, y) = &- \alpha_1 x \partial_1 g(x, y) - \alpha_2 y \partial_2 g(x, y) \\
&+ N f_1(x) \int_{\mathbb{R}} \left[g\left(x + \frac{u}{\sqrt{N}} - \frac{x}{N}, y + \frac{u}{\sqrt{N}}\right) - g(x, y) \right] d\nu(u) \\
&+ f_2(y) \left[g(x, 0) - g(x, y)\right].
\end{aligned}$$

$\mathcal{D}'(\bar{A})$ contains $C_b^2(\mathbb{R}^2)$, and for $g \in C_b^2(\mathbb{R}^2)$ and $(x, y) \in \mathbb{R}^2$,

$$\begin{aligned}
\bar{A}g(x, y) = &- \alpha_1 x \partial_1 g(x, y) - x f_1(x) \partial_1 g(x, y) - \alpha_2 y \partial_2 g(x, y) \\
&+ \frac{\sigma^2}{2} f_1(x) \sum_{i,j=1}^{2} \partial_{i,j}^2 g(x, y) + f_2(y) \left[g(x, 0) - g(x, y)\right].
\end{aligned}$$

We have the following control on the convergence of the generators.

Lemma 2.3. *Grant Assumptions 1.3.$(i) - (ii)$ and Assumption 1.3.(iv). For $g \in C_b^3(\mathbb{R}^2)$ and $(x, y) \in \mathbb{R}^2$,*

$$\left|A^N g(x, y) - \bar{A}g(x, y)\right| \leq \frac{f_1(x)}{6N^{1/2}} \left(\int_{\mathbb{R}} |u|^3 d\nu(u)\right) \sum_{|\beta|=3} ||\partial_\beta g||_\infty + \frac{1}{2}x^2 f_1(x) N^{-1} \left|\left|\partial_{11}^2 g\right|\right|_\infty.$$

Proof. Recalling that ν is a centered probability measure and that σ^2 is its variance,

$$
A^N g(x,y) - \bar{A}g(x,y)
$$

$$
= \int_{\mathbb{R}} N f_1(x) \left[g\left(x + \frac{u}{\sqrt{N}} - \frac{x}{N}, y + \frac{u}{\sqrt{N}} \right) - g(x,y) \right] d\nu(u)
$$

$$
+ x f_1(x) \partial_1 g(x,y) - \frac{\sigma^2}{2} f_1(x) \sum_{i,j=1}^{2} \partial_{i,j}^2 g(x,y)
$$

$$
= N f_1(x) \int_{\mathbb{R}} \left[g\left(x + \frac{u}{\sqrt{N}} - \frac{x}{N}, y + \frac{u}{\sqrt{N}} \right) - g(x,y) - \frac{u}{\sqrt{N}} \sum_{i=1}^{2} \partial_i g(x,y) + \frac{x}{N} \partial_1 g(x,y) \right.
$$

$$
\left. - \frac{u^2}{2N} \sum_{i,j=1}^{2} \partial_{i,j}^2 g(x,y) \right] d\nu(u).
$$

In order to make appear the second-order term of the Taylor polynomial of $g(x + u/\sqrt{N} - x/N, y + u/\sqrt{N})$, let us write

$$
\frac{u^2}{2N} \sum_{i,j=1}^{2} \partial_{i,j}^2 g(x,y)
$$

$$
= \frac{1}{2}\left(\frac{u}{\sqrt{N}} - \frac{x}{N} \right)^2 \partial_{11}^2 g(x,y) + \frac{u^2}{2N} \partial_{22}^2 g(x,y) + \frac{u}{\sqrt{N}} \left(\frac{u}{\sqrt{N}} - \frac{x}{N} \right) \partial_{12}^2 g(x,y)
$$

$$
+ \frac{ux}{N\sqrt{N}} \partial_{11}^2 g(x,y) - \frac{x^2}{2N^2} \partial_{11}^2 g(x,y) + \frac{ux}{N\sqrt{N}} \partial_{12}^2 g(x,y).
$$

Therefore, since $\int_{\mathbb{R}} u\, d\nu(u) = 0$,

$$
A^N g(x,y) - \bar{A}g(x,y)
$$

$$
= N f_1(x) \int_{\mathbb{R}} \left[g\left(x + \frac{u}{\sqrt{N}} - \frac{x}{N}, y + \frac{u}{\sqrt{N}} \right) - g(x,y) - \frac{u}{\sqrt{N}} \sum_{i=1}^{2} \partial_i g(x,y) + \frac{x}{N} \partial_1 g(x,y) \right.
$$

$$
\left. - \left(\frac{1}{2}\left(\frac{u}{\sqrt{N}} - \frac{x}{N} \right)^2 \partial_{11}^2 g(x,y) + \frac{u^2}{2N} \partial_{22}^2 g(x,y) + \frac{u}{\sqrt{N}} \left(\frac{u}{\sqrt{N}} - \frac{x}{N} \right) \partial_{12}^2 g(x,y) \right) \right] d\nu(u)
$$

$$
+ \frac{x^2}{2N} f_1(x) \partial_{11}^2 g(x,y).
$$

As a consequence

$$
\left| A^N g(x,y) - \bar{A}g(x,y) \right|
$$

$$
\leq N f_1(x) \int_{\mathbb{R}} \left[\left| g\left(x + \frac{u}{\sqrt{N}} - \frac{x}{N}, y + \frac{u}{\sqrt{N}} \right) - g(x,y) \right. \right.
$$

$$
- \frac{u}{\sqrt{N}} \sum_{i=1}^{2} \partial_i g(x,y) + \frac{x}{N} f_1(x) \partial_1 g(x,y) - \frac{1}{2}\left(\frac{u}{\sqrt{N}} - \frac{x}{N} \right)^2 \partial_{11}^2 g(x,y)
$$

$$
\left. \left. - \frac{u^2}{2N} \partial_{22}^2 g(x,y) - \frac{u}{\sqrt{N}} \left(\frac{u}{\sqrt{N}} - \frac{x}{N} \right) \partial_{12}^2 g(x,y) \right| \right] d\nu(u)
$$

$$
+ \frac{x^2}{2N} f_1(x) |\partial_{11}^2 g(x,y)|.
$$

Then Taylor-Lagrange's inequality gives the result. $\square$

2.2 Regularity of the limit semigroup and stochastic flow

In this section, we prove the regularity of the semigroup of the limit process $(\bar{X}, \bar{Y})$ w.r.t. its initial condition and a control of its derivatives. However, there are additional technical difficulties in this proof compared to [17]. We rely on Theorem 1.4.1 of [29] which guarantees the differentiability of stochastic flows under some conditions. Let us recall that the stochastic flow of $(\bar{X}, \bar{Y})$ is the family of functions $(x, y) \mapsto (\bar{X}_t^{(x)}, \bar{Y}_t^{(x,y)})$ ($t \geq 0$). A well-known kind of flow is called "Brownian flow". By definition, a Brownian flow is a stochastic flow that is almost surely continuous w.r.t. the time-parameter t with independent increments w.r.t. the time-parameter.

The main difficulty is that the trajectories of the process $(\bar{X}, \bar{Y})$ are not continuous. Hence, the stochastic flow of this process is not a Brownian flow and Theorem 1.4.1 of [29] cannot be applied. So, we study the stochastic flow between its jump times. Then we face another difficulty: the jump times of the stochastic flow depends on the initial condition, and we do not want to study the regularity of this dependency.

This is why we introduce the following auxiliary limit process $(\tilde{X}, \tilde{Y})$. The only difference with the limit process $(\bar{X}, \bar{Y})$ is that the stochastic intensity of the jump term in the SDE of $\tilde{Y}$ is constant. In other words, the occurrences of the jump times of the process $\tilde{Y}$ are distributed as a homogeneous Poisson process.

$$\tilde{X}_t = \tilde{X}_0 - \alpha_1 \int_0^t \tilde{X}_s ds + \sigma \int_0^t \sqrt{f_1(\tilde{X}_s)} dW_s - \int_0^t \tilde{X}_s f_1(\tilde{X}_s) ds,$$

$$\tilde{Y}_t = \tilde{Y}_0 - \alpha_2 \int_0^t \tilde{Y}_s ds + \sigma \int_0^t \sqrt{f_1(\tilde{X}_s)} dW_s - \int_{[0,t] \times [0, ||f_2||_\infty]} \tilde{Y}_{s-} d\tilde{\pi}^1(s, z),$$

where $\tilde{\pi}^1$ is a Poisson measure on $\mathbb{R}_+^2$ with Lebesgue intensity.

Let us note $N_t = \tilde{\pi}^1([0,t] \times ||f_2||_\infty)$ the number of jumps of the process $\tilde{Y}$ before time t. By definition of $\tilde{\pi}^1$, $(N_t)_{t \geq 0}$ is a homogeneous Poisson process with rate $||f_2||_\infty$. We also note T_n ($n \geq 1$) the jump times of the process $\tilde{Y}$.

The next proposition is a consequence of Girsanov's theorem for jump processes. It gives the Radon-Nikodym derivative between the distribution of $(\bar{X}_t, \bar{Y}_t)$ and that of $(\tilde{X}_t, \tilde{Y}_t)$, for any $t \geq 0$. This result is a straightforward application of Theorem $VI.T3$ of [6] (see also of Theorem 3.5 [31]), its proof is therefore omitted.

Proposition 2.4. *Grant Assumptions* $1.3.(i) - (ii)$. *Let* $N_t := \tilde{\pi}^1([0,t] \times [0, ||f_2||_\infty])$ *and* T_n *(*$n \in \mathbb{N}^*$*) be the jump times of* $(N_t)_t$. *For all* $g \in C_b(\mathbb{R}^2)$, *and,* $x, y \in \mathbb{R}$,

$$\mathbb{E}\left[g\left(\bar{X}_t^{(x)}, \bar{Y}_t^{(x,y)}\right)\right]$$

$$= \mathbb{E}\left[g\left(\tilde{X}_t^{(x)}, \tilde{Y}_t^{(x,y)}\right) \prod_{n=1}^{N_t} \frac{f_2\left(\tilde{Y}_{T_n}^{(x,y)}\right)}{||f_2||_\infty} \exp\left[-\int_0^t \left(f_2\left(\tilde{Y}_s^{(x,y)}\right) - ||f_2||_\infty\right) ds\right]\right].$$

In order to obtain a control on the regularity of the semigroup $\bar{P}$ of the limit process $(\bar{X}, \bar{Y})$, we rely on the previous proposition and on a control of the derivatives of the stochastic flow of the auxiliary process $(\tilde{X}, \tilde{Y})$. The next step is to prove this control.

Let us note $\varphi_t(x, y)$ ($t \in \mathbb{R}_+, (x, y) \in \mathbb{R}^2$) the stochastic flow related to the following $2-$dimensional SDE:

$$X_t = X_0 - \alpha_1 \int_0^t X_s ds + \sigma \int_0^t \sqrt{f_1(X_s)} dW_s - \int_0^t X_s f_1(X_s) ds, \qquad (2.2)$$

$$Y_t = Y_0 - \alpha_2 \int_0^t Y_s ds + \sigma \int_0^t \sqrt{f_1(X_s)} dW_s.$$

With our previous notation, it means that $\varphi_t(x, y) := (X_t^{(x)}, Y_t^{(x,y)})$, if (X, Y) is defined as the solution of the above $2-$dimensional SDE. One can note that $\varphi_t(x, y)$ is the flow of the process $(\tilde{X}, \tilde{Y})$ between its jump times. It is clear that $\varphi_t(x, y)$ is a stochastic flow by Definition 1.1.1 of [29]. It is even a Brownian flow since the process (X, Y) has independent increments and continuous trajectories.

Remark 2.5. Contrarily to the more general framework of [29], the stochastic flow $\varphi_t(x, y)$ depends only on one time parameter, because the process is time-homogeneous.

Proposition 2.6. *Grant Assumption 1.3.(i) and Assumption 1.3.(iii). Almost surely, for all $t \geq 0$, $(x, y) \mapsto \varphi_t(x, y)$ is a C^3-diffeomorphism.*

Proof. We rely on Theorem 1.4.1 of [29] and use the notation therein. Let us note (a, b) the local characteristics of the flow $\varphi_t(x, y)$. The notion of local characteristics is formally defined in Assumption 1 of [29, p. 8], and when we study the flow related to an SDE with a drift term and a Brownian term, the function b can be shown to be the coefficient of the drift term, and a the matrix of quadratic covariations of the Brownian term. Here the local characteristics of $\varphi_t(x, y)$ are given by

$$b(x, y) = \begin{pmatrix} -\alpha_1 x - x f_1(x) \\ -\alpha_2 y \end{pmatrix} \text{ and } a((x, y), (x', y')) = \sigma^2 \sqrt{f_1(x) f_1(x')} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

One can check that the Assumptions 1, 2 and 3 of [29, p. 8, 9, 15] are satisfied. Indeed, Assumptions 1 and 2 follows easily from the fact that, if (X, Y) is solution to (2.2), then

$$\mathbb{E}_{(x,y)} \left[(X_t)^2 + (Y_t)^2 \right] \leq C_t (1 + x^2 + y^2),$$

where the inequality above can be proved with the same reasoning as in the proof of Lemma 3.2. For Assumption 3, it is sufficient to remark that the functions $x \mapsto x f_1(x)$ and $x \mapsto \sqrt{f_1(x)}$ are Lipschitz continuous under Assumptions 1.3.(i) and 1.3.(iii).

Then, Theorem 1.4.1 of [29] allows to conclude the proof. $\qquad \square$

Proposition 2.7. *Grant Assumptions 1.3.$(i) - (iii)$. Almost surely, for all $t \geq 0$, $(x, y) \mapsto (\tilde{X}_t^{(x)}, \tilde{Y}_t^{(x,y)})$ is C^3 and, for all even $p \in \mathbb{N}^*$, for all multi-index $\beta \in \mathbb{N}^2$ such that $1 \leq |\beta| \leq 3$,*

$$\sup_{x \in \mathbb{R}} \mathbb{E} \left[\sup_{0 \leq s \leq t} \left(\partial_\beta \tilde{X}_s^{(x)} \right)^p \right] + \sup_{x,y \in \mathbb{R}} \mathbb{E} \left[\sup_{0 \leq s \leq t} \left(\partial_\beta \tilde{Y}_s^{(x,y)} \right)^p \right] < \infty.$$

Proof. We already know that, for all $t \geq 0$, $x \mapsto \tilde{X}_t^{(x)}$ is C^3 by Proposition 2.6. Indeed, by definition $\tilde{X}_t^{(x)}$ is the first coordinate of $\varphi_t(x, y)$.

Now, let us recall that T_n ($n \in \mathbb{N}^*$) denote the jump times of the process $(\tilde{X}^{(x)}, \tilde{Y}^{(x,y)})$ and that these jump times do not depend on the initial condition (x, y). For $t < T_1$,

$$(\tilde{X}_t^{(x)}, \tilde{Y}_t^{(x,y)}) = \varphi_t(x, y)$$

is C^3 w.r.t. (x, y) by Proposition 2.6. And, for $t \in [T_n, T_{n+1}[$ ($n \in \mathbb{N}^*$),

$$(\tilde{X}_t^{(x)}, \tilde{Y}_t^{(x,y)}) = \varphi_{t-T_n}(\tilde{X}_{T_n}^{(x)}, 0),$$

which is also C^3 w.r.t. (x, y) as a composition of the function $(x, y) \in \mathbb{R}^2 \mapsto \varphi_{t-T_n}(x, y)$ and the function $(x, y) \in \mathbb{R}^2 \mapsto (\tilde{X}_{T_n}^{(x)}, 0)$. This proves that, almost surely, for all $t \geq 0$, $(x, y) \mapsto (\tilde{X}_t^{(x)}, \tilde{Y}_t^{(x,y)})$ is C^3. Now, we prove the second part of the statement.

By definition, for all $t \geq 0$,

$$\tilde{X}_t^{(x)} = x - \alpha_1 \int_0^t \tilde{X}_s^{(x)} ds - \int_0^t \tilde{X}_s^{(x)} f_1(\tilde{X}_s^{(x)}) ds + \sigma \int_0^t \sqrt{f_1\left(\tilde{X}_s^{(x)}\right)} dW_s.$$

This implies

$$\partial_x \tilde{X}_t^{(x)} = 1 - \alpha_1 \int_0^t \partial_x \tilde{X}_s^{(x)} ds - \int_0^t \tilde{X}_s^{(x)} \partial_x \tilde{X}_s^{(x)} f_1'(\tilde{X}_s^{(x)}) ds - \int_0^t \partial_x \tilde{X}_s^{(x)} f_1(\tilde{X}_s^{(x)}) ds$$
$$+ \sigma \int_0^t \partial_x \tilde{X}_s^{(x)} \left(\sqrt{f_1} \right)' \left(\tilde{X}_s^{(x)} \right) dW_s.$$

We also know that

$$\partial_x \tilde{Y}_t^{(x,y)} = -\alpha_2 \int_0^t \partial_x \tilde{Y}_s^{(x,y)} ds + \sigma \int_0^t \partial_x \tilde{X}_s^{(x)} \cdot \sqrt{f_1 \left(\tilde{X}_s^{(x)} \right)} dW_s$$
$$- \int_0^t \int_{[0,||f_2||_\infty]} \partial_x \tilde{Y}_s^{(x,y)} d\tilde{\pi}(s,z).$$

Noticing that the initial condition of the process $(\partial_x \tilde{X}_t^{(x)}, \partial_x \tilde{Y}_t^{(x,y)})$ is $(1,0)$, Lemma 3.2 allows to prove the result of the lemma for the first order partial derivative w.r.t. x. The results for the other partial derivatives follow from the same reasoning. □

Proposition 2.8. *Grant Assumptions* 1.3.$(i) - (iii)$. *For all* $t \geq 0$ *and* $g \in C_b^3(\mathbb{R}^2)$, *the function* $(x,y) \in \mathbb{R}^2 \mapsto \bar{P}_t g(x,y)$ *is* C^3 *and for all multi-index* β *such that* $|\beta| = 3$,

$$\sup_{0 \leq s \leq t} \left|\left| \partial_\beta \bar{P}_s g \right|\right|_\infty \leq C_t ||g||_{3,\infty},$$

where $C_t > 0$ *does not depend on* g.

Proof. For the sake of readability, we do not prove the result for the multi-indexes β such that $|\beta| = 3$ but for those such that $|\beta| = 1$. This case is easier to prove and the techniques are exactly the same.

Recalling Proposition 2.4, we have

$$\bar{P}_t g(x,y) = \mathbb{E}\left[g\left(\tilde{X}_t^{(x)}, \tilde{Y}_t^{(x,y)} \right) \prod_{n=1}^{N_t} \frac{f_2\left(\tilde{Y}_{T_n}^{(x,y)} \right)}{||f_2||_\infty} \exp\left[-\int_0^t \left(f_2\left(\tilde{Y}_s^{(x,y)} \right) - ||f_2||_\infty \right) ds \right] \right].$$
$$(2.3)$$

To prove that the function $\bar{P}_t g$ is C^1, we rely on Lemma 3.1 which is a generalization of the classical result about the "differentiation under the integral sign" that uses Vitali theorem instead of the dominated convergence theorem.

By Proposition 2.7, we know that the expression in the expectation of the right-hand side of (2.3) is C^1 almost surely. In addition, one can control the expectation of the square of the derivatives of this expression thanks to Proposition 2.7 (see below for the explicit expression of the derivatives). More precisely, if we note $Z^{(x,y)}$ the random variable appearing in the expectation of (2.3), then Proposition 2.7 implies that for any $x', y' \in \mathbb{R}$,

$$\sup_{x \in \mathbb{R}} \int_\Omega \left| \partial_x Z^{(x,y')} \right|^2 d\mathbb{P}(\omega) < \infty \text{ and } \sup_{y \in \mathbb{R}} \int_\Omega \left| \partial_y Z^{(x',y)} \right|^2 d\mathbb{P}(\omega) < \infty.$$

This implies that the uniform integrability condition of Lemma 3.1 is satisfied.

So Lemma 3.1 (applied twice since there are two coordinates) implies that $\bar{P}_t g$ is C^1. In addition, denoting by

$$L_t(x,y) := \prod_{n=1}^{N_t} \frac{f_2\left(\tilde{Y}_{T_n}^{(x,y)} \right)}{||f_2||_\infty} \exp\left[-\int_0^t \left(f_2\left(\tilde{Y}_s^{(x,y)} \right) - ||f_2||_\infty \right) ds \right],$$

we have (thanks to Lemma 3.1)

$$\partial_2 \left(\bar{P}_t g \right) (x, y) = \mathbb{E} \left[\partial_y \tilde{Y}_t^{(x,y)} \cdot \partial_2 g \left(\tilde{X}_t^{(x)}, \tilde{Y}_t^{(x,y)} \right) L_t(x, y) \right]$$
$$+ \mathbb{E} \left[g \left(\tilde{X}_t^{(x)}, \tilde{Y}_t^{(x,y)} \right) \sum_{n=1}^{N_t} \partial_y \tilde{Y}_{T_n}^{(x,y)} \frac{f_2' \left(\tilde{Y}_{T_n}^{(x,y)} \right)}{||f_2||_\infty} \prod_{k \neq n} \frac{f_2 \left(\tilde{Y}_{T_k}^{(x,y)} \right)}{||f_2||_\infty} \right.$$
$$\left. \times \exp \left(\int_0^t (||f_2||_\infty - f_2(\tilde{Y}_s^{(x,y)})) ds \right) \right]$$
$$- \mathbb{E} \left[\int_0^t \partial_y \tilde{Y}_s^{(x,y)} f_2' \left(\tilde{Y}_s^{(x,y)} \right) ds \cdot g \left(\tilde{X}_t^{(x)}, \tilde{Y}_t^{(x,y)} \right) L_t(x, y) \right].$$

As a consequence,

$$\left| \partial_2 \left(\bar{P}_t g \right) (x, y) \right| \leq ||\partial_2 g||_\infty e^{t||f_2||_\infty} \mathbb{E} \left[\left| \partial_y \tilde{Y}_t^{(x,y)} \right| \right]$$
$$+ ||g||_\infty \frac{||f_2'||_\infty}{||f_2||_\infty} e^{||f_2||_\infty t} \mathbb{E} \left[N_t \sup_{0 \leq s \leq t} \left| \partial_y \tilde{Y}_s^{(x,y)} \right| \right]$$
$$+ ||f_2'||_\infty \cdot ||g||_\infty e^{t||f_2||_\infty} t \mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \partial_y \tilde{Y}_s^{(x,y)} \right| \right].$$

The expressions of the first line and the third line above can be bounded uniformly in (x, y) thanks to Proposition 2.7. For the expression of the second line, we can use Cauchy-Schwarz's inequality for example (we can recall that N_t follows the Poisson distribution with parameter $t||f_2||_\infty$).

This implies that

$$\sup_{x,y \in \mathbb{R}} \sup_{0 \leq s \leq t} \left| \partial_2 \left(\bar{P}_s g \right) (x, y) \right| \leq C_t ||g||_{1,\infty}.$$

With similar computation, we can prove that

$$\sup_{x,y \in \mathbb{R}} \sup_{0 \leq s \leq t} \left| \partial_1 \left(\bar{P}_s g \right) (x, y) \right| \leq C_t ||g||_{1,\infty}.$$

So we have proved the result for the two multi-indexes β such that $|\beta| = 1$. Note that in the case for the multi-indexes β such that $|\beta| = 3$, we would need a control on higher moments of the processes $\tilde{X}, \tilde{Y}$ and their derivatives to control expectations of product of these processes. Indeed, the partial derivative w.r.t. the multi-index $(1, 2)$ would involve terms as

$$\mathbb{E} \left[\left(\sup_{0 \leq s \leq t} |\partial_x \tilde{Y}_s^{(x,y)}| \right) \left(\sup_{0 \leq s \leq t} |\partial_y \tilde{Y}_s^{(x,y)}| \right)^2 \right]$$

which can be controlled using Cauchy-Schwarz's inequality, recalling that Proposition 2.7 allows to bound every polynomial moment of the partial derivatives of the process $(\tilde{X}, \tilde{Y})$ w.r.t. its initial condition. $\qquad \square$

2.3 End of the proof of Theorem 1.5

With a similar proof to the one of Proposition 5.6 of [17], we can prove that, for all $t \geq 0$, $N \in \mathbb{N}^*$, $g \in C_b^3(\mathbb{R}^2)$, $x, y \in \mathbb{R}$,

$$\left(\bar{P}_t - P_t^N \right) g(x, y) = \int_0^t P_{t-s}^N \left(\bar{A} - A^N \right) \bar{P}_s g(x, y) ds.$$

Consequently,

$$\left| \left(\bar{P}_t - P_t^N \right) g(x,y) \right| \leq \int_0^t \mathbb{E}_{(x,y)} \left[\left| \left(\bar{A} - A^N \right) \bar{P}_s g(X_{t-s}^N, Y_{t-s}^N) \right| \right] ds$$

$$\leq C \cdot N^{-1/2} \int_0^t \mathbb{E}_{(x,y)} \left[f_1(X_{t-s}^N) \sum_{|\beta|=3} || \partial_\beta \left(\bar{P}_s g \right) ||_\infty \right] ds$$

$$+ C N^{-1} \int_0^t \mathbb{E}_{(x,y)} \left[\left(X_{t-s}^N \right)^2 f_1 \left(X_{t-s}^N \right) || \partial_{11}^2 (\bar{P}_s g) ||_\infty \right] ds$$

$$\leq C_t \cdot N^{-1/2} \cdot ||f_1||_\infty ||g||_{3,\infty} + C_t \cdot N^{-1} \cdot ||f_1||_\infty \left(1 + x^2 \right) ||g||_{2,\infty},$$

where we have used Lemma 2.3 to obtain the second inequality above, and Proposition 2.8 and Lemma 3.4 to obtain the last one. More precisely, Proposition 2.8 is used to control the partial derivatives of the function $\bar{P}_s g$, and Lemma 3.4 to control the second order moment of X_{t-s}^N.

3 Technical lemmas

The first lemma allows to differentiate under an integral sign. Contrary to the classical result that relies on the dominated convergence theorem, this result relies on Vitali convergence theorem. This result is a direct consequence of Lemma 6.1 of [12].

Lemma 3.1. *Let $F : (x, \omega) \in \mathbb{R} \times \Omega \mapsto \mathbb{R}$ be a measurable function, where $(\Omega, \mathbb{P})$ is a probability space. Assume that:*

- *$\mathbb{P}-$almost surely, $x \mapsto F(x, \omega)$ is C^1,*
- *the set of functions $\{ \omega \mapsto \partial_1 F(x, \omega) \ : \ x \in \mathbb{R} \}$ is uniformly integrable.*

Then, $\mathbb{P}-$almost surely, $x \mapsto \int_\Omega F(x, \omega) d\mathbb{P}(\omega)$ is C^1 and its derivative is

$$\int_\Omega \partial_1 F(x, \omega) d\mathbb{P}(\omega).$$

Proof. By Lemma 6.1 of [12], for any $T > 0$, the function $x \mapsto \int_\Omega F(x, \omega) d\mathbb{P}(\omega)$ is almost surely C^1 on $[-T, T]$ with derivative

$$\int_\Omega \partial_1 F(x, \omega) d\mathbb{P}(\omega).$$

As it holds true for any $T > 0$, the result is proved. $\square$

The next lemma allows to obtain a priori estimates on solutions of certain SDEs (in our model, it is used to control the polynomial moments of the derivatives of the limit process in Proposition 2.7). This result is not written to be general, because some specific properties of our SDEs allow some simplifications to obtain the result.

Lemma 3.2. *Let $(X_t, Y_t)_t$ be a solution of the following two-dimensional SDE*

$$dX_t = b_1(X_t, Y_t)dt + \varsigma_1(X_t, Y_t)dW_t,$$

$$dY_t = b_2(X_t, Y_t)dt + \varsigma_2(X_t, Y_t)dW_t - \int_{[0,+\infty[} Y_{t-} \mathbb{1}_{\{z \leq f(Y_{t-})\}} d\pi(z,t),$$

where $(W_t)_t$ is one-dimensional standard Brownian motion, π a Poisson measure on $\mathbb{R}_+^2$ with Lebesgue intensity and f a non-negative measurable function. Assume that the functions b_1, b_2, ς_1 and ς_2 are sublinear: for all $x, y \in \mathbb{R}$,

$$|b_1(x,y)| + |b_2(x,y)| + |\varsigma_1(x,y)| + |\varsigma_2(x,y)| \leq C(1 + |x| + |y|).$$

Then:

(i) *for any even $p \in \mathbb{N}$ such that X_0 and Y_0 both have a finite $p-$order moment, for all $t \geq 0$,*

$$\mathbb{E}\left[X_t^p\right] + \mathbb{E}\left[Y_t^p\right] \leq C_{t,p}(1 + \mathbb{E}\left[X_0^p\right] + \mathbb{E}\left[Y_0^p\right]),$$

with $C_{t,p} > 0$ some constant,

(ii) *if X_0 and Y_0 both have a finite $2p-$th order moment (for some even $p \in \mathbb{N}^*$), for all $t \geq 0$,*

$$\mathbb{E}\left[\sup_{0 \leq s \leq t} X_s^p\right] + \mathbb{E}\left[\sup_{0 \leq s \leq t} Y_s^p\right] \leq C_{t,p}\left(1 + \mathbb{E}\left[X_0^{2p}\right] + \mathbb{E}\left[Y_0^{2p}\right]\right),$$

with $C_{t,p} > 0$ another constant.

Proof. **Step 1.** Let us prove the point (i). By Ito's formula, for all $t \geq 0$,

$$
\begin{aligned}
X_t^p =& X_0^p + p \int_0^t X_s^{p-1} b_1(X_s, Y_s) ds + \frac{p(p-1)}{2} \int_0^t X_s^{p-2} \varsigma_1(X_s, Y_s)^2 ds \\
& + p \int_0^t X_s^{p-1} \varsigma_1(X_s, Y_s) dW_s,
\end{aligned}
\tag{3.1}
$$

$$
\begin{aligned}
Y_t^p =& Y_0^p + p \int_0^t Y_s^{p-1} b_2(X_s, Y_s) ds + \frac{p(p-1)}{2} \int_0^t Y_s^{p-2} \varsigma_2(X_s, Y_s)^2 ds \\
& + p \int_0^t Y_s^{p-1} \varsigma_2(X_s, Y_s) dW_s - \int_{[0,t] \times [0,+\infty[} Y_{s-}^p \mathbb{1}_{\{z \leq f(Y_{s-})\}} d\pi(s, z) \\
\leq& Y_0^p + p \int_0^t Y_s^{p-1} b_2(X_s, Y_s) ds + \frac{p(p-1)}{2} \int_0^t Y_s^{p-2} \varsigma_2(X_s, Y_s)^2 ds \\
& + p \int_0^t Y_s^{p-1} \varsigma_2(X_s, Y_s) dW_s.
\end{aligned}
\tag{3.2}
$$

Now for $M > 0$, let us introduce the following stopping time

$$\tau_M := \inf\{t > 0 \ : \ |X_t| > M \text{ or } |Y_t| > M\},$$

and the function

$$u_{M,p}(t) := \mathbb{E}\left[(X_{t \wedge \tau_M})^p + (Y_{t \wedge \tau_M})^p\right].$$

Thanks to (3.1), we have for all $t \geq 0$,

$$
\begin{aligned}
\mathbb{E}\left[(X_{t \wedge \tau_M})^p\right] &= \mathbb{E}\left[X_0^p\right] + \mathbb{E}\left[\int_0^{t \wedge \tau_M} \left(p X_s^{p-1} b_1(X_s, Y_s) + \frac{p(p-1)}{2} X_s^{p-2} \varsigma_1(X_s, Y_s)^2\right) ds\right] \\
&\leq \mathbb{E}\left[X_0^p\right] + C \int_0^t \mathbb{E}\left[|X_{s \wedge \tau_M}|^{p-1} \cdot |b_1(X_{s \wedge \tau_M}, Y_{s \wedge \tau_M})| + X_{s \wedge \tau_M}^{p-2} \varsigma_1(X_{s \wedge \tau_M}, Y_{s \wedge \tau_M})^2\right] ds \\
&\leq \mathbb{E}\left[X_0^p\right] + C \int_0^t \left(1 + u_{M,p}(s)\right) ds,
\end{aligned}
$$

where we have used the sublinearity of the functions b_1 and ς_1 to obtain the last line. With the same reasoning, we also have using (3.2) for all $t \geq 0$,

$$\mathbb{E}\left[(Y_{t \wedge \tau_M})^p\right] \leq \mathbb{E}\left[Y_0^p\right] + C \int_0^t \left(1 + u_{M,p}(s)\right) ds.$$

So, for all $t \geq 0$,

$$u_{M,p}(t) \leq \mathbb{E}\left[X_0^p\right] + \mathbb{E}\left[Y_0^p\right] + Ct + C \int_0^t u_{M,p}(s) ds.$$

Then, as the function $u_{M,p}$ is locally integrable (thanks to the stopping time τ_M), Grönwall's lemma implies that for all $t \geq 0$,

$$u_{M,p}(t) \leq \left(\mathbb{E}\left[X_0^p\right] + \mathbb{E}\left[Y_0^p\right] + Ct\right) e^{Ct} \leq C_t(1 + \mathbb{E}\left[X_0^p\right] + \mathbb{E}\left[Y_0^p\right]), \tag{3.3}$$

where $C_t > 0$ does not depend on M.

Now let us prove that τ_M goes to infinity almost surely as M goes to infinity. Firstly, as $(\tau_M)_M$ is non-decreasing w.r.t. M, we know that τ_M has an almost sure limit τ as M goes to infinity. We have, for all $t > 0$, by Markov's inequality and thanks to (3.3),

$$\mathbb{P}\left(\tau_M \leq t\right) \leq \mathbb{P}\left(X_{t \wedge \tau_M}^p + Y_{t \wedge \tau_M}^p \geq M^p\right) \leq \frac{1}{M^p} u_{M,p}(t) \xrightarrow[M \to \infty]{} 0.$$

Consequently, for all $t > 0$,

$$\mathbb{P}\left(\tau \leq t\right) \leq \mathbb{P}\left(\bigcap_{M \in \mathbb{N}^*} \{\tau_M \leq t\}\right) = \lim_{M \to \infty} \mathbb{P}\left(\tau_M \leq t\right) = 0.$$

Hence,

$$\mathbb{P}\left(\tau < \infty\right) \leq \mathbb{P}\left(\bigcup_{t \in \mathbb{N}^*} \{\tau \leq t\}\right) = 0.$$

This means that τ_M goes to infinity almost surely as M goes to infinity.

Finally, applying Fatou's lemma to (3.3) gives that for all $t \geq 0$,

$$\mathbb{E}\left[X_t^p + Y_t^p\right] \leq C_t(1 + \mathbb{E}\left[X_0^p\right] + \mathbb{E}\left[Y_0^p\right]).$$

Step 2. Let us prove the point (ii). By (3.1) and Burkhölder-Davis-Gundy's inequality,

$$\mathbb{E}\left[\sup_{0 \leq s \leq t} X_s^p\right] \leq \mathbb{E}\left[X_0\right]^p + p \int_0^t \mathbb{E}\left[|X_s^{p-1} b_1(X_s, Y_s)|\right] ds$$

$$+ \frac{p(p-1)}{2} \int_0^t \mathbb{E}\left[X_s^{p-2} \varsigma_1(X_s, Y_s)^2\right] ds + Cp \left(\int_0^t \mathbb{E}\left[X_s^{2p-2} \varsigma_1(X_s, Y_s)^2\right] ds\right)^{1/2}$$

$$\leq \mathbb{E}\left[X_0^2\right] + C \int_0^t \left(1 + \mathbb{E}\left[X_s^p\right] + \mathbb{E}\left[Y_s^p\right]\right) ds$$

$$+ C \left(\int_0^t \left(1 + \mathbb{E}\left[X_s^{2p}\right] + \mathbb{E}\left[Y_s^{2p}\right]\right) ds\right)^{1/2}$$

$$\leq \mathbb{E}\left[X_0^p\right] + C \int_0^t \left(1 + \mathbb{E}\left[X_s^{2p}\right] + \mathbb{E}\left[Y_s^{2p}\right]\right) ds$$

$$\leq C_t \left(1 + \mathbb{E}\left[X_0^{2p}\right] + \mathbb{E}\left[Y_0^{2p}\right]\right).$$

The result for the process Y follows from the same computation using (3.2) instead of (3.1). $\qquad\square$

Remark 3.3. In the proof of Lemma 3.2, we use the stopping times τ_M to guarantee the function $u_{M,p}$ to be locally integrable. Otherwise we would not have the right to apply Grönwall's lemma.

The following lemma is similar to the previous one: it allows to control the moments of the process (X^N, Y^N) instead of the moments of the limit process $(\bar{X}, \bar{Y})$. An important point of the lemma is that, the desired estimates are uniform in N (provided that the initial conditions (X_0^N, Y_0^N) admit finite moments uniformly bounded in N). This lemma is used in the proof of Theorem 1.5 at Section 2.3. In the proof of Lemma 3.4 below, we do not introduce the stopping times τ_M as in the proof of Lemma 3.2 for the sake of readability.

Lemma 3.4. *Grant Assumption 1.3.(i) and Assumption 1.3.(iv). Let X^N be the solution of the first equation of (1.1). Assume that each X_0^N has a finite second order moment. Then, for any $N \in \mathbb{N}^*$ and $t \geq 0$,*

$$\sup_{s \leq t} \mathbb{E}\left[\left(X_s^N\right)^2\right] \leq C_t \left(1 + \mathbb{E}\left[(X_0^N)^2\right]\right),$$

where $C_t > 0$ does not depend on N.

Proof. By Ito's formula,

$$(X_t^N)^2 = (X_0^N)^2 - 2\alpha_1 \int_0^t (X_s^N)^2 ds$$
$$+ \sum_{j=1}^N \int_{[0,t] \times \mathbb{R}_+ \times \mathbb{R}} \left[\left(X_{s-}^N + \frac{u}{\sqrt{N}} - \frac{X_{s-}^N}{N}\right)^2 - (X_{s-}^N)^2\right] \mathbb{1}_{\left\{z \leq f_1(X_{s-}^N)\right\}} d\pi^{(1),j}(s,z,u)$$
$$\leq (X_0^N)^2$$
$$+ \sum_{j=1}^N \int_{[0,t] \times \mathbb{R}_+ \times \mathbb{R}} \left[\left(X_{s-}^N + \frac{u}{\sqrt{N}} - \frac{X_{s-}^N}{N}\right)^2 - (X_{s-}^N)^2\right] \mathbb{1}_{\left\{z \leq f_1(X_{s-}^N)\right\}} d\pi^{(1),j}(s,z,u)$$
$$\leq (X_0^N)^2 + \sum_{j=1}^N \int_{[0,t] \times \mathbb{R}_+ \times \mathbb{R}} \left[2\left(\frac{u}{\sqrt{N}} - \frac{X_{s-}^N}{N}\right) X_{s-}^N + \left(\frac{u}{\sqrt{N}} - \frac{X_{s-}^N}{N}\right)^2\right]$$
$$\times \mathbb{1}_{\left\{z \leq f_1(X_{s-}^N)\right\}} d\pi^{(1),j}(s,z,u).$$

Taking expectation, we obtain

$$\mathbb{E}\left[(X_t^N)^2\right] \leq \mathbb{E}\left[(X_0^N)^2\right]$$
$$+ N||f_1||_\infty \int_0^t \int_{\mathbb{R}} \mathbb{E}\left[2\left(\frac{u}{\sqrt{N}} - \frac{X_s^N}{N}\right) X_s^N + \left(\frac{u}{\sqrt{N}} - \frac{X_s^N}{N}\right)^2\right] d\nu(u) ds.$$

Now let us expand the term within the expectation

$$2\left(\frac{u}{\sqrt{N}} - \frac{X_s^N}{N}\right) X_s^N + \left(\frac{u}{\sqrt{N}} - \frac{X_s^N}{N}\right)^2 = 2\frac{u}{\sqrt{N}} X_s^N - 2\frac{(X_s^N)^2}{N} + \frac{u^2}{N} + \frac{(X_s^N)^2}{N^2} - 2\frac{u X_s^N}{N\sqrt{N}}.$$

Recalling that $\int u d\nu(u) = 0$ and that $\sigma^2 := \int u^2 d\nu(u)$, we have

$$\int_{\mathbb{R}} \mathbb{E}\left[2\left(\frac{u}{\sqrt{N}} - \frac{X_s^N}{N}\right) X_s^N + \left(\frac{u}{\sqrt{N}} - \frac{X_s^N}{N}\right)^2\right] d\nu(u)$$
$$= -\frac{2}{N} \mathbb{E}\left[(X_s^N)^2\right] + \frac{\sigma^2}{N} + \frac{1}{N^2} \mathbb{E}\left[(X_s^N)^2\right]$$
$$\leq \frac{\sigma^2}{N} + \frac{1}{N^2} \mathbb{E}\left[(X_s^N)^2\right].$$

Consequently,

$$\mathbb{E}\left[(X_t^N)^2\right] \leq \mathbb{E}\left[(X_0^N)^2\right] + \sigma^2 ||f_1||_\infty t + \frac{1}{N} \int_0^t \mathbb{E}\left[(X_s^N)^2\right] ds.$$

And Grönwall's lemma implies that

$$\mathbb{E}\left[(X_t^N)^2\right] \leq \left(\mathbb{E}\left[(X_0^N)^2\right] + \sigma^2 ||f_1||_\infty t\right) e^{t/N} \leq \left(\mathbb{E}\left[(X_0^N)^2\right] + \sigma^2 ||f_1||_\infty t\right) e^t,$$

which proves the result. □

References

[1] Luisa Andreis, Paolo Dai Pra, and Markus Fischer. *McKean–Vlasov limit for interacting systems with simultaneous jumps*, Stoch. Anal. Appl. **36** (2018), 960–995. doi. MR-3925147

[2] Javier Baladron, Diego Fasoli, Olivier Faugeras, and Jonathan Touboul. *Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons*, J. Math. Neurosci. **2** (2012), Art. 10, 50, doi. MR-2974499

[3] Vlad Bally, Dan Goreac, and Victor Rabiet. *Regularity and Stability for the Semigroup of Jump Diffusions with State-Dependent Intensity*, Ann. Appl. Probab. **28** (2018), 3028 – 3074. doi. MR-3847980

[4] Jérémie Barral and Alex D reyes. *Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics*, Nat. Neurosci. **19** (2016) 1690–1696. doi.

[5] Mads Bonde Raad, Susanne Ditlevsen, and Eva Löcherbach. *Stability and mean-field limits of age dependent Hawkes processes*, Ann. Instit. Henri Poincaré Probab. Stat. **56** (2020) 1958–1990. doi. MR-4116713

[6] Pierre Brémaud. *Point Processes and Queues, Martingale Dynamics*, Springer-Verlag (1981).

[7] Julien Chevallier. *Mean-field limit of generalized Hawkes processes*, Stochastic Process. Appl. **127** (2017) 3870–3912. doi. MR-3718099

[8] Julien Chevallier. *Fluctuations for mean-field interacting age-dependent Hawkes processes*, Electron. J. Probab. **22** (2017) 1–49. doi. MR-3646068

[9] Mark Davis. *Markov Models and Optimization*, First ed. London New York: Chapman & Hall (1993). doi.

[10] Anna De Masi, Antonio Galves, Eva Löcherbach, and Errico Presutti. *Hydrodynamic Limit for Interacting Neurons*, J. Stat. Phys. **158** (2015) 866–902. doi. MR-3311484

[11] Susanne Ditlevsen and Eva Löcherbach. *Multi-class Oscillating Systems of Interacting Neurons*, Stochastic Process. Appl. **127** (2017) 1840–1869. doi. MR-3646433

[12] Nathaniel Eldredge. *Strong hypercontractivity and strong logarithmic Sobolev inequalities for log-subharmonic functions on stratified Lie groups*, Nonlinear Anal. **168** (2018) 1–26. doi. MR-3759466

[13] Xavier Erny. *Well-posedness and propagation of chaos for McKean-Vlasov equations with jumps and locally Lipschitz coefficients*, Stochastic Process. Appl. **150** (2022) 192–214, doi. MR-4423077.

[14] Xavier Erny, Eva Löcherbach, and Dasha Loukianova. *Conditional propagation of chaos for mean field systems of interacting neurons*, Electron. J. Probab. **26** (2021) 1–25. doi. MR-4235471

[15] Xavier Erny, Eva Löcherbach, and Dasha Loukianova. *White-noise driven conditional McKean-Vlasov limits for systems of particles with simultaneous and random jumps*, Probab. Theory Related Fields, **183** (2022) 1027–1073, doi. MR-4453321.

[16] Xavier Erny, Eva Löcherbach, and Dasha Loukianova. *Strong error bounds for the convergence to its mean field limit for systems of interacting neurons in a diffusive scaling*, (2021) arXiv:2111.05213 [math].

[17] Xavier Erny, Eva Löcherbach, and Dasha Loukianova. *Mean field limits for interacting Hawkes processes in a diffusive regime*, Bernoulli **28** (2022) 125–149. doi. MR-4337700

[18] Stewart Ethier and Thomas Kurtz. *Markov Processes. Characterization and Convergence*, Wiley Series In Probability And Statistics (2005). doi.

[19] Olivier Faugeras and James MacLaurin. *Asymptotic Description of Neural Networks with Correlated Synaptic Weights*, Entropy **17** (2015) 4701–4743. doi.

[20] Olivier Faugeras, Jonathan Touboul, and Bruno Cessac. *A constructive mean-field analysis of multi population neural networks with random synaptic weights and stochastic inputs*, Front. Comput. Neurosci. **3** (2009). doi.

[21] Nicolas Fournier and Eva Löcherbach. *On a toy model of interacting neurons*, Ann. Inst. Henri Poincaré Probab. Stat. **52** (2016) 1844–1876. doi. MR-3573298

[22] Sonja Grün and Stefan Rotter. *Analysis of Parallel Spike Trains*, Springer Series in Computational Neuroscience (2010). doi.

[23] Bilal Haider, Alvaro Duque, Andrea Hasenstaub, and David McCormick. *Neocortical Network Activity In Vivo Is Generated through a Dynamic Balance of Excitation and Inhibition*, J. Neurosci. **26** (2006) 4535–4545. doi.

[24] Alan Hawkes. *Spectra of some self-exciting and mutually exciting point processes*, Biometrika **58** (1971) 83–90. doi. MR-278410

[25] Pierre Hodara and Eva Löcherbach. *Hawkes processes with variable length memory and an infinite number of components*, Adv. in Appl. Probab. **49** (2017) 84–107. doi. MR-3631217

[26] David Hubel. *Eye, Brain, and Vision*, Second ed. Scientific American Library (1995).

[27] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic Differential Equations and Diffusion Processes*, Second ed. North-Holland Publishing Company (1989).

[28] Jean Jacod and Albert Shiryaev. *Limit Theorems for Stochastic Processes*, Second ed. Springer-Verlag BerlinHeidelberg NewYork (2003).

[29] Hiroshi Kunita. *Lectures on Stochastic Flows and Applications for the Indian Institute Of Science Bangalore*, Tata Institute Of Fundamental Research Springer-Verlag (1986).

[30] Geneviève Leuba and Rudolf Kraftsik. *Changes in volume, surface estimate, three-dimensional shape and total number of neurons of the human primary visual cortex from midgestation until old age*. Anat. Embryol. **190** (1994) 351–366. doi.

[31] Eva Löcherbach. *Likelihood Ratio Processes for Markovian Particle Systems with Killing and Jumps*, Stat. Inference Stoch. Process. **5** (2002) 153–177. doi. MR-1917290

[32] Eric Luçon and Wilhelm Stannat. *Mean field limit for disordered diffusions with singular interactions*, Ann. Appl. Probab. **24** (2014) 1946–1993. doi. MR-3226169

[33] Sean Meyn and Richard Tweedie. *Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes*, Adv. in Appl. Probab. **25** (1993) 518–548. doi. MR-1234295

[34] Peter Pfaffelhuber, Stefan Rotter, and Jakob Stiefel. *Mean-field limits for non-linear Hawkes processes with excitation and inhibition*, Stochastic Process. Appl. **153** (2022) 57–78, doi. MR-4463083

[35] Jonathan Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan Litke, Eduardo Chichilnisky, and Eero Simoncelli. *Spatio-temporal correlations and visual signalling in a complete neuronal population*, Nature **454** (2008) 995–999. doi.

[36] Patricia Reynaud-Bouret, Vincent Rivoirard, Franck Grammont, and Christine Tuleau-Malot. *Goodness-of-Fit Tests and Nonparametric Adaptive Estimation for Spike Train Analysis*, J. Math. Neurosci. **4** (2014) 3 – 330325. doi. MR-3197017

[37] Robert Rosenbaum and Brent Doiron. *Balanced Networks of Spiking Neurons with Spatially Dependent Recurrent Connections*, Phys. Rev. X **4** (2014). doi.

[38] Yousheng Shu, Andrea Hasenstaub, and David McCormick. *Turning on and off recurrent balanced cortical activity*, Nature **423** (2003) 288–293. doi

[39] Haim Sompolinsky, Andrea Crisanti, and Hans-Jürgen Sommers *Chaos in Random Neural Networks*, Phys. Rev. Lett. **61** (1988) 259–262. doi. MR-949871

[40] Andrei Velichko and Petr Boriskov. *Oscillator Circuit for Spike Neural Network with Sigmoid Like Activation Function and Firing Rate Coding*, IEEE Trans. Circuits Syst. II: Express Br. **67** (2020) 3477–3481. doi