Exponential ergodicity for diffusions with jumps driven by a Hawkes process

Charlotte Dion, Sarah Lemler, Eva Löcherbach

Exponential ergodicity for diffusions with jumps driven by a Hawkes process

Charlotte Dion[(1)] and Sarah Lemler[(2)] and Eva Löcherbach[(3)]

[(1)] Sorbonne Université, UMR CNRS 8001, LPSM, 75005 Paris, France
[(2)] École CentraleSupélec, MICS, 91190 Gif-sur-Yvette, France
[(3)] Université Paris 1, SAMM EA 4543, 75013 Paris, France

Abstract

In this paper, we introduce a new class of processes which are diffusions with jumps, where the jumps are driven by a multivariate linear Hawkes process, and study their long-time behavior. In the case of exponential memory kernels for the underlying Hawkes process, we establish conditions for the positive Harris recurrence of the couple (X, λ), where X denotes the diffusion process and λ the stochastic intensity of the driving Hawkes. As a direct consequence of the Harris recurrence, we obtain the ergodic theorem for X. Furthermore, we provide sufficient conditions under which the process is exponentially $\beta-$mixing. This paper is the foundation for a second paper [11] in which we carry a statistical study of diffusions driven by Hawkes jumps, with a view towards applications in neuroscience.

AMS Classification: 60J35, 60F99, 60H99, 60G55

Keywords: Diffusions with jumps; Linear Hawkes process, Piecewise deterministic Markov processes (PDMP), Ergodicity; Foster-Lyapunov drift criteria; β-mixing.

1 Introduction

1.1 Motivation

In this work, we introduce a new class of continuous time processes $X = (X_t)_{t\geq 0}$ with jumps driven by a multivariate Hawkes process N, satisfying the following equation

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s + \int_0^t a(X_{s-}) \sum_{j=1}^{M} dN_s^{(j)}, \tag{1.1}$$

where W is the standard one-dimensional Brownian motion and N is a multivariate Hawkes process having M components with intensity process λ. N is supposed to be independent from W. We shall specify the dynamics of N later on, in particular, the precise form of the stochastic intensity process λ is given in (2.1) below.

The jumps of the Hawkes process impact the dynamic of the diffusion process. In this model, the structure of the jumps is different from classical jump-diffusion processes with Lévy jumps. Indeed:

1) the intensity of the jump process N does not depend on the dynamic of X,
2) the Hawkes process has a special structure of time dependency: the state of the intensity at time t depends on all the past,
3) the multidimensional nature of the Hawkes process can be interpreted as the influence of M subjects communicating one with each other and impacting the dynamic of X along time.

Before entering the theoretical study section, we present two examples of applications to allow the reader to understand the interest of this model.

The first example is the study of the behavior of a set of neurons. Indeed, neurons are communicating through the emission of electrical signals, thus forming a network. When a neuron sends a signal to another neuron, its membrane potential increases sharply and we say that the neuron spikes at this time. If we are looking at one fixed neuron at the heart of this network, its activity between two spikes can be designed by the present model. Indeed, during this interval the membrane potential can be described by a diffusion process influenced by the spikes coming from the neurons around modelled by the Hawkes process. Hawkes processes have been studied recently in the context of neurosciences [see 33, 18] to model the interactions between the neurons within their sequences of spikes. The new model we propose allows to use both informations: the continuous membrane potential for a fixed neuron together with the spikes of several other neurons around it.

The second example comes from financial mathematics where many models have been proposed for the stochastic volatility of a stock price. Again, our jump-diffusion model could be adapted here. In this case, the economic news (about the company that we study or about other companies impacting the market) can be well described by a Hawkes process. There is a large literature in the financial side that studied Hawkes processes, among others recently [36, 23, 3], which motivates also the present work.

We have just detailed two examples but we hope that our model can have various other applications in different areas.

The aim of the present article is to study the longtime behavior of the couple of processes (X, λ). In particular we prove the positive Harris recurrence of (X, λ), implying ergodicity, and we establish the speed of $\beta-$mixing. These are classical probabilistic properties, which are the key ingredients for the theoretical study of estimators of the model parameters. The estimation of the coefficients of the process will be the issue of coming works.

1.2 State of the art

On the one hand time-homogeneous diffusion processes are well known Markov processes that have been widely studied. Let us cite for example [22] which has initiated the work on Markovian properties on diffusion processes or the complete works [21, 32] in which all properties of diffusion are summarized. Ergodicity and the exponential decrease of the uniformly strong mixing coefficient have been established in the pioneering papers [37, 24]. Then, jump-diffusion processes have been introduced, for example in the context of risk management [see 36], either driven by a Poisson process or more generally by a Lévy process [see 1, for a review]. Their properties have been studied from a probabilistic point of view by [27].

On the other hand, Hawkes processes [17] have been studied a lot lately. These point processes generalize the Poisson processes by introducing what is commonly called "self-excitation": Present jumps are able to trigger future jumps, and in this way, correlations between successive jump events are introduced. [9] gives a precise synthesis of earlier results published by Hawkes. Important probabilistic results for these mutually exciting point processes, such as stability, limit theorems and large population limits, have been obtained in [6, 10, 2, 4].

The present paper is devoted to the particular case of linear Hawkes processes having exponential memory kernels. These processes are of special interest, since the associated stochastic intensity processes are Markovian. The linear structure of the intensity process implies that a clustering representation of the Hawkes process [see 16, 34] is available, and this is the reason why linear Hawkes processes are the most studied ones. Linear Hawkes processes are generalized Poisson cluster processes with clusters that are finite branching processes described by an offspring intensity measure that has support in $\mathbb{R}^+$ [see 31]. For such processes, ergodicity has been established under mild conditions in [6]. However, the mixing property obtained there is stated without any control on the tails of the coupling time such that no precise control on the speed of convergence to equilibrium is available. [34] exploits the cluster representation to obtain properties in the case of kernel functions having bounded support. In the present paper we wish however to deal with kernel functions having unbounded support such that we can not rely on these results neither. More recently, [7] has obtained a control of the τ-mixing coefficient in the inhibitory case – but this does not imply the $\beta-$mixing property of the process.

In this paper, we prove the exponential β-mixing property for the couple (X, λ) in the case of exponential kernel functions having unbounded support.

1.3 Main contributions and plan of the paper

More precisely, this paper is devoted to a general study of the longtime behavior of a jump diffusion process, where the jumps are driven by a multivariate linear Hawkes process having exponential memory kernels. This study is separated into two parts, first, a study of the autonomous Hawkes process and its long time behavior in a general frame, second, a study of the couple (X, λ), relying on the classical regeneration technique ([28, 29]).

Using a contraction argument, Theorem 3.2 establishes the exponential rate of convergence to equilibrium of a sub-critical linear Hawkes process having exponential memory kernels. This convergence holds for the Wasserstein $1-$distance, the memory kernels do not need to be exponential. Our proof is inspired by the recent article [13].

We then study the long time behavior of the couple (X, λ). We do this in a Markovian framework, that is, we suppose that the intensity process λ is a Markov process itself. This is why we restrict attention to exponential memory kernels in this part of the paper. In this case, λ is a multidimensional piecewise deterministic Markov process (PDMP), and our proof of the Harris recurrence of (X, λ) follows a well-known scheme, see e.g. [28, 29].

Firstly, we establish a local Döblin lower bound for the transition semigroup of the process (X, λ), based on a lower bound of the joint transition density of (X, λ). This is the content of Theorem 4.4. Let us briefly describe the difficulties related to this result. If we suppose that the underlying diffusion part is uniformly elliptic, e.g., $\sigma^2(\cdot)$ is uniformly lower-bounded, then under mild assumptions on the drift part, classical results on transition densities of diffusion processes imply that such lower bounds always hold for the diffusion part of X (more precisely, we rely on density estimates obtained by [14]). So obviously, the difficulty comes from the jump part of the process, that is, from the PDMP λ. Indeed, the only randomness present in the evolution of λ is given by the random jump times – the jump heights of λ being completely deterministic. To be able to make use of this randomness, we apply an integration by parts formula with respect to these jump times. Since each jump time can at most generate one direction of noise and since λ is $M-$dimensional, we have to iterate this integration by parts procedure M times. Once the jump part of the process possesses some part of Lebesgue density, we have to ensure that this density is preserved by the stochastic flow. Our proof, rather technical, relies on fine support properties of (X, λ) and provides, as a by-product, a simulation algorithm for the process.

In a second step, a Foster-Lyapunov condition (see Proposition 4.5) ensures the control of the return times to some compact set K. We then use control arguments to deduce from this a precise control of

the hitting times of the set C where the transition densities are lower-bounded according to the local Döblin lower bound. Both results together imply the positive Harris recurrence of the process together with the ergodic theorem. This is our first main result, stated in Theorem 4.6.

Finally, following the steps of [27] where the ergodicity and exponential β-mixing bounds are established for diffusions with jumps driven by a Lévy process, we obtain the control of the $\beta-$mixing coefficient of (X, λ) in our second main theorem, Theorem 4.9.

Our paper is organised as follows. We introduce our model together with all necessary notation and assumptions in Section 2. Section 2.3 provides a simulation algorithm of the process which is interesting in its own right. In Section 3, we establish the exponential rate of convergence to equilibrium of the Hawkes process with respect to the Wasserstein distance. In Section 4, we prove the exponential ergodicity and the exponential $\beta-$mixing of the coupled process (X, λ).

2 Presentation of the model and first properties

2.1 The model

Throughout this paper we work on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F})$. We start by introducing the jump part of our process, that is, the multivariate Hawkes process.

This process is defined in terms of a collection of baseline intensities, which are positive constants $\xi_i \geq 0, 1 \leq i \leq M$, and in terms of interaction functions $h_{ij} : \mathbb{R}_+ \to \mathbb{R}_+, 1 \leq i, j \leq M$, which are measurable functions. Let moreover $n^{(i)}, 1 \leq i \leq M$, be discrete point measures on $\mathbb{R}_-$ satisfying that

$$\int_{\mathbb{R}_-} h_{ij}(t-s) n^{(j)}(ds) < \infty \text{ for all } t \geq 0.$$

We interpret them as *initial condition* of our process.

Definition 2.1. *A linear Hawkes process with parameters $(\xi_i, h_{ij})_{1 \leq i,j \leq M}$ and with initial condition $n^{(i)}, 1 \leq i \leq M$, is a multivariate counting process $N_t = (N_t^{(1)}, \ldots, N_t^{(M)}), t \geq 0$, such that*

1. *$\mathbb{P}-$almost surely, for all $i \neq j$, $N^{(i)}$ and $N^{(j)}$ never jump simultaneously,*
2. *for all $1 \leq i \leq M$, the compensator of $N_t^{(i)}$ is given by $\int_0^t \lambda_s^{(i)} ds$ where*

$$\lambda_t^{(i)} = \xi_i + \sum_{j=1}^{M} \int_0^{t-} h_{ij}(t-u) dN_u^{(j)} + \sum_{j=1}^{M} \int_{-\infty}^{0} h_{ij}(t-u) dn_u^{(j)}. \tag{2.1}$$

If the functions h_{ij} are locally integrable, the existence of the process $(N_t)_{t \geq 0}$ with the prescribed intensity (2.1) on finite time intervals follows from standard arguments, see e.g. [10].

We consider a jump diffusion process $X = (X_t)_{t \geq 0}$ taking values in $\mathbb{R}$, whose jumps are governed by N. More precisely, X is solution of the stochastic differential equation

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dW_s + \int_0^t a(X_{s^-}) \sum_{j=1}^{M} dN_s^{(j)}, \tag{2.2}$$

where X_0 is a random variable independent of W. Here W is the standard Brownian motion in dimension one, independent of the multivariate Hawkes process N_t having intensity (2.1).

To ensure that the above equation is well posed it is sufficient to impose Lipschitz conditions on the coefficients σ and b and to suppose that a is measurable. In the sequel we shall also need estimates on

the transition densities of the diffusion process underlying (2.2) [1]. Therefore we impose the following additional assumption.

Assumption 2.2. *1. a, b, σ are globally Lipschitz, and b and σ are of class $\mathcal{C}^2$.*

2. There exist positive constants c, q such that for all $x \in \mathbb{R}$, $|b'(x)| + |\sigma'(x)| \leq c$ and $|b''(x)| + |\sigma''(x)| \leq c(1 + |x|^q)$.

3. There exist positive constants a_1 and σ_1 such that $a(x) < a_1$ and $\sigma(x) < \sigma_1$ for all $x \in \mathbb{R}$.

Under these assumptions, assuming only that the memory kernel functions h_{ij} are locally integrable, Equation (2.2) admits a unique strong solution.

To study the longtime behavior of X and to ensure its ergodicity, we introduce two additional conditions which are classical in the study of diffusions and which are derived from [15], see also [37].

Assumption 2.3. *1. We have that $\sigma(x) \geq \sigma_0 > 0$ for all $x \in \mathbb{R}$.*

2. There exist $d \geq 0$, $r > 0$ such that for all x satisfying $|x| > r$, we have $xb(x) \leq -dx^2$.

2.2 Markovian framework

Our study relies on the general theory of Markov processes. To be able to work within a Markovian framework, we impose a special structure on the interaction functions and suppose that

Assumption 2.4. *For all $1 \leq i, j \leq M$,*

$$h_{ij}(t) = c_{ij} e^{-\alpha_{ij} t}, c_{ij} \geq 0, \alpha_{ij} > 0. \tag{2.3}$$

In this case we may introduce the auxiliary Markov processes

$$Y_t^{(ij)} = c_{ij} \int_0^t e^{-\alpha_{ij}(t-s)} dN_s^{(j)} + c_{ij} \int_{-\infty}^0 e^{-\alpha_{ij}(t-s)} dn_s^{(j)}, 1 \leq i, j \leq M, \tag{2.4}$$

and the intensities can be expressed in terms of sums of these Markov processes, that is, $\lambda_t^{(i)} = \xi_i + \sum_{j=1}^M Y_{t-}^{(ij)}$. Notice that

$$Y_0^{(ij)} = y_0^{(ij)} = c_{ij} \int_{-\infty}^0 e^{\alpha_{ij} s} dn_s^{(j)}, \quad \text{and} \quad dY_t^{(ij)} = -\alpha_{ij} Y_t^{(ij)} dt + c_{ij} dN_t^{(j)}. \tag{2.5}$$

In particular, the process

$$(X_t, Y_t^{(ij)}, 1 \leq i, j \leq M)$$

is a Markov process. Its longtime behavior is determined on the one hand by the longtime behavior of the underlying Hawkes process N_t and on the other hand on the one of the continuous diffusion process with drift b and diffusion coefficient σ. Notice that N_t is an autonomous process, that is, N_t does not depend on X.

[1] that is, the process X in between its successive jump times

2.3 First properties of the process (X, Y) and an associated stochastic flow.

In the sequel, we shall denote the whole process by $Z_t := (X_t, Y_t)$. It takes values in $\mathbb{R} \times \mathbb{R}^{M\times M}$. We write $(P_t)_{t\geq 0}$ for its associated transition semigroup and $z = (x, y), y = (y^{(ij)})_{1\leq i,j\leq M}$ for generic elements of its state space.

Proposition 2.5. *Grant Assumption 2.2. Then the process Z is a Feller process, that is, $P_t f$ belongs to $C_b(\mathbb{R} \times \mathbb{R}^{M\times M})$ whenever $f \in C_b(\mathbb{R} \times \mathbb{R}^{M\times M})$.*

The proof of this result follows from classical arguments, see e.g. the proof of Proposition 4.8 in [19], or [20]. In what follows we shall give an alternative proof relying on an explicit construction of the process Z as a *stochastic flow*. This construction is interesting in its own right and based on the fact that Z is a sort of *piecewise continuous Markov process*, that is, a generalization of the piecewise deterministic Markov processes to those traveling in between successive jumps according to *stochastic flows* instead of deterministic ones. We start by giving the principal elements needed to construct this flow.

The associated stochastic Brownian flow. Thanks to our assumptions on b and σ, by Theorem 4.2.5 of [25], there exists a unique stochastic flow of homeomorphisms

$$\Phi_t(x),\ 0 \leq t < \infty, x \in \mathbb{R}$$

such that for all $t \geq 0,\ x \in \mathbb{R}$,

$$\Phi_t(x) = x + \int_0^t b(\Phi_u(x))du + \int_0^t \sigma(\Phi_u(x))dW_u.$$

In particular, we have that

$$(t, x) \to \Phi_t(x) \text{ is continuous } (0\leq t<\infty,\ x \in \mathbb{R}).$$

The stochastic flow Φ describes the evolution of X in between successive jumps of X (or, equivalently, of N).

The associated deterministic flow. In between successive jump events of N, the process Y evolves according to the deterministic flow

$$\varphi_t(y) = (\varphi_t^{(ij)}(y))_{1\leq i,j\leq M}, \quad \text{where } \varphi_t^{(ij)}(y) = e^{-\alpha_{ij}t}y^{(ij)}, 1 \leq i, j \leq M. \tag{2.6}$$

Simulation algorithm for Z. We propose a simulation algorithm for Z for any family of starting configurations $z \in K_1 \times K_2$, where $K_1 \subset \mathbb{R}$, $K_2 \subset \mathbb{R}^{M\times M}$ are compact sets. The first step is to construct an upper bound $\bar{N}_t$ on the number of jumps of Z, that is, of N, during some finite time interval $[0, t]$. To do so, let

$$\bar{c} := \sup_{1\leq i,j\leq M} c_{ij} \tag{2.7}$$

be the maximal coefficient (weight) and consider the one-dimensional auxiliary Hawkes process N_t^* having intensity λ_t^*

$$\lambda_0^* := \sum_{j=1}^M \xi_j + \sup_{y\in K_2} \|y\|_1,\ \lambda_s^* = \bar{\lambda}_0^* + M\bar{c}\int_0^{s-} N^*(du) = \lambda_0^* + M\bar{c}N_{s-}^*,\ s \geq 0.$$

Here, $\|y\|_1 = \sum_{1\le i,j\le M} |y^{(ij)}|$. Observe that N^* is a linear Hawkes process with memory kernel $\bar{h}(t) = M\bar{c}1_{\mathbb{R}_+}(t)$. Obviously, for all $t \ge 0$,

$$\sum_{i=1}^{M} N_t^{(i)} \le N_t^*.$$

Moreover, for all $t \ge 0$, $N_t^* < \infty$ almost surely. In what follows, we shall write $T_1^* < T_2^* < \ldots < T_n^* < \ldots$ for the successive jump times of N^*.

We work conditionally on the realization of N^* on $[0,t]$, that is, on the event $\{N_t^* = n\}$, and on a realization of the associated jump times $0 < t_1 < t_2 \ldots < t_n < t$. Our goal is to construct a version of Z, conditionally on these choices, which is continuous in the starting point z. This construction relies on the classical thinning method.

During this construction, we choose successively random variables $U_1, \ldots, U_n$ taking values in $\{0, 1, \ldots, M\}$ and define a process $z_s(z, U_1^n)$, depending on these choices, for $0 \le s \le t$. Here, $U_1^n = (U_1, \ldots, U_n)$. This process is defined recursively as follows. Firstly, we put

$$z_s = (x_s, y_s) = (\Phi_s(x), \varphi_s(y)) \text{ for all } 0 \le s < t_1.$$

Then, conditionally on $z_{t_1-} = z_1 = (x_1, y_1)$, we choose a random variable $U_1 \in \{0, 1, \ldots, M\}$ with law $q(z_1, t_1, \cdot)$, where

$$q(z,t,\{i\}) = \frac{\xi_i + \sum_j y_1^{(ij)}}{\bar{\lambda}_t},\ 1 \le i \le M,\ q(z,t,\{0\}) = 1 - \sum_{i=1}^{M} q(z,t,\{i\}).$$

U_1 gives the label of the particle (or component) that will jump at time t_1, if $U_1 = 0$, no jump happens at this time. More precisely, on $\{U_1 \ge 1\}$, we put

$$x_{t_1} := x_1 + a(x_1),\ y_{t_1}^{(ij)} := y_1^{(ij)} + c_{ij}1_{\{U_1=j\}},\ \text{for all } 1 \le i,j \le M,$$

and $z_{t_1} := z_1$ on $\{U_1 = 0\}$.

Then we put

$$x_s = \Phi_{s-t_1}(x_{t_1}), y_s = \varphi_{s-t_1}(y_{t_1}) \text{ for all } t_1 \le s < t_2,$$

and we proceed iteratively by choosing, conditionally on $y_{t_2-} = y_2$, a random variable

$$U_2 \sim q(y_2, t_2, \cdot),$$

and so on. Finally, we obtain a terminal value $x_t = \Phi_{t-t_n}(x_{t_n})$ and $y_t = \varphi_{t-t_n}(y_{t_n})$. It is easy to check that

$$\mathcal{L}(z_t(z, U_1^n)) = \mathcal{L}(Z_t | N_t^* = n, T_1^* = t_1, \ldots, T_n^* = t_n).$$

The important point is that the above construction ensures the continuity of the application

$$z \mapsto z_1(z, U_1^n).$$

Alternative proof of Proposition 2.5. Noticing that the law of N^* does not depend on the starting point z but only on an upper bound of the associated intensities, this implies that for any $f \in C_b(\mathbb{R}\times\mathbb{R}^{M\times M})$, the mapping

$$K \ni z \mapsto P_t f(z) \in \mathbb{R}$$

is continuous, implying the assertion of Proposition 2.5. □

3 Wasserstein coupling

This section concerns the case of a general Hawkes process with intensity given by Equation 2.1. We start by discussing the longtime behavior of the Hawkes process N. To do so, it is well known that linear Hawkes processes can be represented as a spatio-temporal branching process with immigrations which can be either recurrent or transient, depending on whether the underlying branching mechanism is sub-critical, critical or super-critical. Therefore we assume that

Assumption 3.1. *Let H be the matrix with entries $H_{ij} = \int_0^\infty h_{ij}(t)dt$. Then its spectral radius satisfies*

$$\rho := \rho(H) < 1.$$

Assumption 3.1 ensures that we are in the sub-critical framework and its implies that a stationary version of N_t exists [see 6]. To be able to implement statistical methods of parameter estimation, we need however more qualitative properties of $(X_t, Y_t)_t$; in particular, we need ergodicity and exponential β-mixing.

We start by proving the exponential convergence of the process $Y_t := (Y_t^{(ij)})_{1\leq i,j\leq M}$ to equilibrium in Wasserstein distance, under the condition (3.1).

Recall that Y_t takes values in $\mathbb{R}^{M\times M}$ and that for any $y \in \mathbb{R}^{M\times M}$, $\|y\|_1 = \sum_{i,j=1}^M |y^{(ij)}|$ denotes the associated L^1-norm. Let μ and ν be two probability measures on $\mathbb{R}^{M\times M}$. We call coupling of μ and ν any probability measure on $\mathbb{R}^{M\times M} \times \mathbb{R}^{M\times M}$ whose marginals are μ and ν, and we denote by $\Gamma(\mu,\nu)$ the set of all such couplings. The Wasserstein distance between μ and ν is defined by

$$W_1(\mu,\nu) = \inf\left\{\int_{\mathbb{R}^{M\times M}}\int_{\mathbb{R}^{M\times M}} \|x-y\|_1\gamma(dx,dy), \gamma \in \Gamma(\mu,\nu)\right\}.$$

Write $(P_t^Y)_{t\geq 0}$ for the transition semigroup of the process Y and introduce

$$\alpha := \min_{1\leq i,j\leq M} \alpha_{ij} > 0. \tag{3.1}$$

The following theorem states the exponential rate of convergence to equilibrium of the process N with respect to the Wasserstein distance.

Theorem 3.2. *Grant Assumption* (3.1).

1. *Then there exist $C > 0, \kappa > 0$ such that for any choice of probability measures μ and ν on $\mathcal{B}(\mathbb{R}^{M\times M})$,*
$$W_1(\mu P_t^Y, \nu P_t^Y) \leq Ce^{-\kappa t} W_1(\mu,\nu). \tag{3.2}$$

2. *In particular, there exists a unique invariant probability measure π^Y of the process Y such that for any probability measure ν on $\mathcal{B}(\mathbb{R}^{M\times M})$,*
$$W_1(\pi^Y, \nu P_t^Y) \leq Ce^{-\kappa t} W_1(\pi^Y,\nu).$$

Proof. The assertion of point 1. follows from a standard contraction argument. Let Y_t and $\tilde{Y}_t$ be two copies of the process Y, starting from initial configurations y_0 and $\tilde{y}_0 \in \mathbb{R}^{M\times M}$, and let $N, \tilde{N}$ be the associated counting processes. Then

$$Y_t^{(ij)} - \tilde{Y}_t^{(ij)} = e^{-\alpha_{ij}t}(y_0^{(ij)} - \tilde{y}_0^{(ij)}) + \int_0^t h_{ij}(t-s)(dN_s^{(j)} - d\tilde{N}_s^{(j)}).$$

Taking expectation, this implies that $\delta_t^{(ij)} := \mathbb{E}|Y_t^{(ij)} - \tilde{Y}_t^{(ij)}|$ satisfies

$$\delta_t^{(ij)} \leq e^{-\alpha t}\delta_0^{(ij)} + \sum_{k=1}^{M}\int_0^t h_{ij}(t-s)\delta_s^{(jk)}ds.$$

Let $\delta_t = (\delta_t^{(ij)})_{1\leq i,j\leq M}$ and $h(t) = (h_{ij}(t))_{1\leq i,j\leq M}$. Hence we have just shown that

$$\delta_t \leq e^{-\alpha t}\delta_0 + h * \delta(t),$$

where $h * \delta(t)$ denotes matrix convolution. Put $\Gamma(t) = \sum_{n\geq 1} h^{*n}(t)$. Then by Corollary 3.1 of [8],

$$\delta_t \leq e^{-\alpha t}\delta_0 + \left(\int_0^t \Gamma(t-s)e^{-\alpha s}ds\right)\delta_0$$

and there exist $C > 0$, $\kappa > 0$ such that

$$\delta_t^{(ij)} \leq Ce^{-\kappa t}$$

for all $1 \leq i, j \leq t, t \geq 0$. This concludes the proof of item 1.

Item 2. follows then from standard arguments, see e.g. the proof of Theorem 1 in [13]. □

Remark 3.3. *Under the conditions of Theorem 3.2, write $\tilde{N}_t$ for the stationary version of the Hawkes process having intensity (2.1); that is, $\tilde{N}_t^{(i)}$ has intensity $\tilde{\lambda}_t^{(i)} = \xi_i + \sum_{j=1}^M \tilde{Y}_t^{(ij)}$, where $\tilde{Y}$ is the stationary process evolving according to (2.5). Let moreover N_t be the Hawkes process such that each $N_t^{(i)}$ has intensity $\lambda_t^{(i)} = \xi_i + \sum_{j=1}^M Y_t^{(ij)}$ starting from some fixed initial condition $Y_0 = y_0 \in \mathbb{R}^{M\times M}$. Then (3.2) implies that for any $1 \leq i \leq M$,*

$$\int_0^\infty \mathbb{E}|\lambda_t^{(i)} - \tilde{\lambda}_t^{(i)}|dt < \infty.$$

It is then straightforward to deduce from this by standard coupling arguments, as explained e.g. the proof of Theorem 1 in [6], that N and $\tilde{N}$ couple almost surely in finite time; that is, there exists $T_{coupling} > 0$ such that for all $t \geq 0$, $N_{T_{coupling}+t} - N_{T_{coupling}} = \tilde{N}_{T_{coupling}+t} - \tilde{N}_{T_{coupling}}$, meaning that N and $\tilde{N}$ have the same jump times after time $T_{coupling}$. Moreover, we have the control

$$\mathbb{P}(T_{coupling} > t) \leq Ce^{-\kappa t},$$

which follows from

$$\mathbb{P}(T_{coupling} > t) \leq \mathbb{E}\left(|\sum_i N^{(i)}([t,\infty[) - \tilde{N}^{(i)}([t,\infty[)|\right) \leq \int_t^\infty \sum_{i,j}\delta_s^{(ij)}ds \leq Ce^{-\kappa t}.$$

In the next section we prove a stronger result which is the positive Harris recurrence of the process $(Y_t)_t$, together with $(X_t)_t$.

4 Exponential ergodicity of the process (X, Y)

We start proving the positive Harris recurrence of $(X_t, Y_t)_t$ using the regeneration technique. For simplicity we shall assume that

Assumption 4.1. *For all i, j we have that $\alpha_{ij} = \alpha_i > 0$.*

This assumption is reasonable at least in the perspective of modeling systems of interacting neurons as described in the introduction. Indeed, it means that the leakage rate of the $i-$th neuron does only depend on the neuron i itself and not on the index of the presynaptic neuron j.

Under Assumption 4.1, the analysis of the $M \times M-$dimensional Markov process $Y_t = (Y_t^{(ij)})$ boils down to the one of the $M-$dimensional process

$$Y_t^{(i)} = \sum_{j=1}^{M} Y_t^{(ij)} = \sum_{j=1}^{M} c_{ij} \int_0^t e^{-\alpha_i(t-s)} dN_s^{(j)}, 1 \leq i \leq M,$$

having generator

$$A^Y f(y) = -\sum_{i=1}^{M} \alpha_i \, y^{(i)} \partial_{y^{(i)}} f(y) + \sum_{i=1}^{M} (\xi_i + y^{(i)})[f(y + \sum_j c_{ji} e_j) - f(y)], \tag{4.1}$$

where e_j is the $j-$the unit vector in $\mathbb{R}^M$. In the sequel, without changing notation, we shall therefore work with the $1 + M-$dimensional process

$$Z_t = (X_t, Y_t).$$

As before, we write for short $z = (x, y) \in \mathbb{R} \times \mathbb{R}^M$ for the elements of the state space of Z; we write P_t^Y for the transition semigroup of Y, now acting on smooth functions $f : \mathbb{R}^M \to \mathbb{R}$, and P_t for the transition semigroup of the whole process Z.

Finally, we work under the following identifiability assumption.

Assumption 4.2. *The offspring matrix H is invertible. Moreover we suppose that $\sum_{i=1}^{M} \xi_i > 0$.*

Remark 4.3. *Notice that the $j-$th column of H gives the total offspring vector that is produced by a jump of the $j-$th component. If H would not be invertible, then this total offspring could be represented by a linear combination of the total offspring vectors of the other components. In particular, in this case the model would not be irreducible in the sense that M is the minimal number of independent component (neurons for example) that is needed to describe to process.*

4.1 A Döblin type lower bound

The following theorem is the main result of this section. It states a local Döblin lower bound for the transition semigroup of the process, which is the main ingredient towards ergodicity of the process $Z = (X, Y)$.

Theorem 4.4. *Grant Assumptions 2.2, 2.3, 3.1 and 4.2. Then there exists $T > 0$ such that for all $z^* = (x^*, y^*) \in \mathbb{R} \times \mathbb{R}^M$ and for all $x^{**} \in \mathbb{R}$ the following holds. There exist $R > 0$, open sets $I_1 \subset \mathbb{R}$, such that $x^{**} \in I_1$, and $I_2 \subset \mathbb{R}^M$ with strictly positive Lebesgue measure and a constant $\beta \in (0, 1)$, depending on I_1, I_2, R and the coefficients of the system with*

$$P_T(z, dz') \geq \beta 1_C(z) \nu(dz'), \tag{4.2}$$

where $C = B_R(z^*)$ is the (open) ball of radius R centred at z^*, and where ν is the uniform probability measure on $I_1 \times I_2$.

The main idea of the proof is to use the jump noise of M successive jumps of the underlying Hawkes process to create a Lebesgue density for the process N, and then to use density estimates of the underlying diffusion to deal with X.

The proof is done in several steps which are the subject of the next subsection.

4.2 Some useful properties of the underlying stochastic flow and proof of Theorem 4.4

We start by collecting useful properties of the stochastic flow governing the evolution of X in between successive jumps.

Transition densities. Due to our assumptions on b and σ, $\Phi_t(x,\cdot)$ possesses a transition density $p_t(x,y)$ with explicit lower bounds. More precisely, Proposition 1.2 of [14] implies that, for some suitable constants c, C,

$$c^{-1}t^{-1/2}e^{-C(x-y)^2/t}e^{-Ctx^2} \leq p_t(x,y) \leq ct^{-1/2}e^{-\frac{1}{C}(x-y)^2/t}e^{Ctx^2}. \tag{4.3}$$

Here, the constants c and C do only depend on the coefficients b and σ and on σ_0 and σ_1.

Support properties. We will use the control theorem which goes back to Strook and Varadhan (1972) [35], see also Millet and Sanz-Sole (1994) [30], Theorem 3.5. For some time horizon $T_1 < \infty$ which is arbitrary but fixed, write $\mathtt{H}$ for the Cameron-Martin space of measurable functions $\mathtt{h} : [0,T_1] \to \mathbb{R}$ having absolutely continuous components $\mathtt{h}(t) = \int_0^t \dot{\mathtt{h}}(s)ds$ with $\int_0^{T_1}(\dot{\mathtt{h}})^2(s)ds < \infty$. For $x \in \mathbb{R}$ and $\mathtt{h} \in \mathtt{H}$, consider the deterministic flow

$$(\varphi_t^{(\mathtt{h})}(x))_{t\geq 0} \text{ solution to } d\varphi_t^{(\mathtt{h})}(x) = \tilde{b}(\varphi_t^{(\mathtt{h})}(x))dt + \sigma(\varphi^{(\mathtt{h})_t}(x))\dot{\mathtt{h}}(t)dt, \text{ with } \varphi_0^{(\mathtt{h})}(x) = x, \tag{4.4}$$

on $[0,T_1]$. In the above formula (4.4), $\tilde{b}$ is Stratonovich drift given by

$$\tilde{b}(x) = b(x) - \frac{1}{2}\sigma(x)\sigma'(x).$$

Denote by $Q_x^{T_1}$ the law of the solution $(\Phi_t(x))_{0\leq t\leq T_1}$. Then for any $x \in \mathbb{R}$ and $\mathtt{h} \in \mathtt{H}$,

$$\left(\varphi_t^{(\mathtt{h})}(x)\right)_{|t\in[0,T_1]} \in \overline{\text{supp}\left(Q_x^{T_1}\right)}. \tag{4.5}$$

We are now able to give the proof of Theorem 4.4.

Proof. Clearly, for all $A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{B}(\mathbb{R}^M)$, $z \in \mathbb{R}\times\mathbb{R}^M$,

$$P_T(z, A\times B) \geq E_z(1_A(X_T)1_B(Y_T), N_T^{(i)} = 1, \forall 1 \leq i \leq M, T_1^{(1)} < T_1^{(2)} < \ldots < T_1^{(M)} < T),$$

where $T_1^{(i)} = \inf\{t \geq 0 : N_t^{(i)} = 1\}$ is the first jump time of $N^{(i)}, 1 \leq i \leq M$.

Step 1. We first deal with the process Y_t. In between successive jump events of N, the process Y evolves according to the deterministic flow

$$\varphi_t(y) = (\varphi_t^{(1)}(y), \ldots, \varphi_t^{(M)}(y)), \text{ where } \varphi_t^{(i)}(y) = e^{-\alpha_i t}y^{(i)}, 1 \leq i \leq M. \tag{4.6}$$

Thus, on the event $\{N_T^{(i)} = 1,\ \forall 1 \leq i \leq M,\ T_1^{(1)} < T_1^{(2)} < \ldots < T_1^{(M)} < T\}$, starting from $Y_0 = y \in \mathbb{R}^M$, we first let the flow φ evolve starting from y up to some first jump time t_1. At that jump time each particle j gains an additional value c_{j1}. We then successively choose the following inter-jump waiting times $t_2, \ldots, t_M$ under the constraint $t_1 + \ldots + t_M < T$. We write

$$s_1 = T - t_1, s_2 = T - (t_1 + t_2), \ldots,\ s_M = T - (t_1 + \ldots + t_M).$$

Conditionally on $Y_0 = y$, the successive choices of $\mathbf{s} = (s_1, \ldots, s_M)$ as above, the position of $Y_T^{(i)}$ is given by

$$\gamma^{(i)}(y, \mathbf{s}) = e^{-\alpha_i T}y^{(i)} + c_{i1}e^{-\alpha_i s_1} + \ldots + c_{iM}e^{-\alpha_i s_M}.$$

In what follows we work at fixed y and we write

$$\gamma_y : \mathbf{s} \mapsto \gamma(y, \mathbf{s}).$$

We will use the jump noise which is created by the M jumps, i.e., we will use a change of variables on the account of $s_1, \ldots, s_M$. Therefore, let

$$\frac{\partial \gamma_y(\mathbf{s})}{\partial \mathbf{s}} = \Big[\frac{\partial \gamma_y(\mathbf{s})}{\partial s_1}, \ldots, \frac{\partial \gamma_y(\mathbf{s})}{\partial s_M}\Big]$$

be the Jacobian matrix of the the map $\mathbf{s} \mapsto \gamma_y(\mathbf{s})$. This matrix does not depend on the initial position y. Indeed, one easily finds that

$$\frac{\partial \gamma_y(\mathbf{s})}{\partial \mathbf{s}} = C(\mathbf{s}) = \Big[C^{(1)}, \ldots, C^{(M)}\Big](\mathbf{s}),$$

where for each $1 \leq j \leq M$, $C^{(j)}(\mathbf{s})$ is a column vector given by

$$C^{(j)}(\mathbf{s}) = -\begin{pmatrix} c_{1j}(\alpha_1)^{-1} e^{-\alpha_1 s_j} \\ c_{2j}(\alpha_2)^{-1} e^{-\alpha_2 s_j} \\ \vdots \\ c_{Mj}(\alpha_M)^{-1} e^{-\alpha_M s_j} \end{pmatrix}.$$

By Assumption 4.2, $C(\mathbf{0})$ is invertible. Therefore, we may choose T sufficiently small such that $\frac{\partial \gamma_y(\mathbf{s})}{\partial \mathbf{s}}$ is invertible. In conclusion, we have just shown that for any $y \in \mathbb{R}^M$, the Jacobian of the map $\mathbf{s} \mapsto \gamma_y(\mathbf{s})$ is invertible at any $\mathbf{s}$ such that $0 < s_M < \ldots < s_1 < T$, for sufficiently small T.

It will be proved now that this uniform invertibility of the Jacobian matrix of the map $\mathbf{s} \mapsto \gamma_y(\mathbf{s})$ implies the first part of inequality (4.2). For that sake, we shall also need the following notation. For each couple $(y, \mathbf{s})$, we write $y_0 = x$, and then define recursively for all $1 \leq k \leq M$,

$$y_k = \varphi_{s_{k-1} - s_k}(y_{k-1}) + c_{\cdot k}, \quad \text{where } c_{\cdot k} := \sum_i c_{ik} e_i, \tag{4.7}$$

with ϕ given by Equation (4.6) and $s_0 = T$.

The sequence $y_1, \ldots, y_M$ corresponds to the positions of the process Y right after successive jumps, starting from the initial location $y \in \mathbb{R}^M$, induced by the inter-jump waiting times $T - s_1, s_1 - s_2, \ldots, s_{M-1} - s_M$ which are determined by $\mathbf{s}$.

Introduce now for each $y \in \mathbb{R}^M$ the total jump rate

$$f(y) := \sum_{i=1}^{M} \xi_i + y^{(i)}$$

(notice that $f(y) \geq \underline{f} := \sum_{i=1}^{M} \xi_i > 0$,) and for each $t \geq 0$, the survival rate

$$e(y, t) = \exp\Big\{ - \int_0^t f\big(\varphi_s(y)\big) ds \Big\}. \tag{4.8}$$

We define for each couple $(y, \mathbf{s})$ (here we set $s_0 = T$),

$$q_y(\mathbf{s}) = \left(\prod_{k=0}^{M-1} \left(\xi_{k+1} + \varphi^{(k+1)}_{s_k - s_{k+1}}(y_k)\right) e(y_k, s_k - s_{k+1}) \right) e(y_M, s_M). \tag{4.9}$$

Since f is bounded away from 0 and from the definition of $e(\cdot,\cdot)$, we deduce that for any couple $(y^*, \mathbf{s}^*)$ there are neighborhoods $W_{\mathbf{s}^*}$ and U_{y^*} of $\mathbf{s}^*$ and y^* respectively such that

$$\inf_{(y,\mathbf{s})\in U_{y^*}\times W_{\mathbf{s}^*}} q_y(\mathbf{s}) > 0. \tag{4.10}$$

Let us now fix $(y^*, \mathbf{s}^*)$ such that the matrix $\frac{\partial \gamma_{y^*}(\mathbf{s}^*)}{\partial \mathbf{s}}$ is invertible. By Lemma 6.2 of [5], there exist an open neighborhood $B_R(y^*)$ of y^*, an open set $I_2 \subset \mathbb{R}^M$, and for any $y \in B_R(y^*)$, an open set W_y such that

$$\tilde{\gamma}_y(\mathbf{s}) : \left\{ \begin{array}{l} W_y \to I_2 \\ \mathbf{s} \mapsto \gamma_y(\mathbf{s}) \end{array} \right.$$

is a diffeomorphism, with $W_y \subset W_{\mathbf{s}^*}$, and also

$$\inf_{y\in B_R(y^*)} \inf_{\mathbf{s}\in \mathbf{W_y}} \Big|\det\Big(\frac{\partial \gamma_y(\mathbf{s})}{\partial \mathbf{s}}\Big)^{-1}\Big| > 0. \tag{4.11}$$

Reducing (if necessary) R, we may assume also that $B_R(y^*) \subset U_{y^*}$. Thus we have that by (4.10) and (4.11),

$$\beta_1 := \inf_{y\in B_R(y^*)} \inf_{s\in W_y} q_y(\mathbf{s})\Big|\det\Big(\frac{\partial \gamma_y(\mathbf{s})}{\partial \mathbf{s}}\Big)^{-1}\Big| > 0. \tag{4.12}$$

Once we have done all these steps we can conclude with the following preliminary result. For any measurable $B \in \mathbb{B}(\mathbb{R}^M)$ and for any $z = (x, y)$ such that $y \in B_R(y^*)$, using the change of variables $y = \tilde{\gamma}_y(\mathbf{s})$, we obtain the lower bound

$$\begin{aligned} E_y(1_B(Y_T), N_T^{(i)} = 1, \forall 1 \le i \le M, T_1^{(1)} < T_1^{(2)} < \ldots < T_1^{(M)} < T) \\ \ge \int_{W_y} q_y(\mathbf{s}) 1_B(\tilde{\gamma}_y(\mathbf{s})) ds_1 \ldots ds_M \ge \beta_1 \int_{I_2\cap B} dy_1 \ldots dy_M. \end{aligned} \tag{4.13}$$

This would establish the desired result if we were only interested in the autonomous process Y.

Step 2. We now deal with the process X. Of course, we still work conditionally on the choice of $s_1, \ldots, s_M$ of the first step. Analogously to (4.7) we therefore introduce the successive jump positions of the process X which are given by $x_0 = x$ and then for all $1 \le k \le M$,

$$x_k = \Phi_{s_{k-1}-s_k}(x_{k-1}) + a(\Phi_{s_{k-1}-s_k}(x_{k-1})), \tag{4.14}$$

where $s_0 = T$ as before.

Conditionally on $X_0 = x, Y_0 = y$, the successive choices of $\mathbf{s} = (s_1, \ldots, s_M)$ as above, the position of X_T is then given by

$$\gamma(z, \mathbf{s}) = \Phi_{s_M}(x_M),$$

where $z = (x, y)$. Notice that x_M depends on the choices of $s_1, \ldots, s_M$ and of course on the evolution of the stochastic flow between the successive jump times.

Step 3. Therefore, conditioning with respect to $X_{T-s_M} = x_M$, we obtain

$$\begin{aligned} P_T(z, A \times B) \ge E_z(1_A(X_T)1_B(Y_T), N_T^{(i)} = 1, \forall 1 \le i \le M, T_1^{(1)} < T_1^{(2)} < \ldots < T_1^{(M)} < T) \\ \ge \int_{W_y} q_y(\mathbf{s}) 1_B(\tilde{\gamma}_y(\mathbf{s}))\, \mathbb{E}\left(\int_A p_{s_M}(x_M, u) du\right) ds_1 \ldots ds_M, \end{aligned} \tag{4.15}$$

where $p_{s_M}(x_M, u)$ is the transition density of (4.3).

Notice that in the last line, expectation $\mathbb{E}$ is taken with respect to Brownian motion only, that is, with respect to the law of X_{T-s_M} under the Wiener measure, conditionally on the choices $s_1, \dots, s_M$.

The lower bound estimate given in (4.3) implies that for fixed s_M and x_M, $p_{s_M}(x_M, u)$ is lower bounded in small neighborhoods of x_M. Therefore, in what follows we need to localize the possible positions of x_M, conditionally on the choices of the jumps times $s_1, \dots, s_M$.

Step 4. Localization of x_M.

We apply the support theorem, that is, (4.5), to our process X_t in between the successive jump times $0, T_1 := t_1, T_2 := t_1 + t_2, \dots, T_M := t_1 + \dots t_M$, by choosing on each time interval $[T_n, T_{n+1}[, n = 0, \dots, M$, the control function $\mathtt{h} \equiv \mathtt{0}$. Consequently, introducing

$$\tilde{x}_1 := \varphi^{(\mathtt{0})}_{t_1}(x) + a(\varphi^{(\mathtt{0})}_{t_1}(x)), \dots, \tilde{x}_M := \varphi^{(\mathtt{0})}_{s_{M-1}-s_M}(x) + a(\varphi^{(\mathtt{0})}_{s_{M-1}-s_M}(x)),$$

there exists an open neighborhood $U_{\tilde{x}_M}$ of $\tilde{x}_M$, such that

$$\mathbb{P}(X_{T-s_M} \in U_{\tilde{x}_M}) > 0.$$

Notice that $\tilde{x}_M$ is a continuous function of the starting point x and of $\mathbf{s}$; that is, $\tilde{x}_M = F(x, \mathbf{s})$ for some continuous function F. This implies that for any starting point x^* and any $R > 0$, there exists a compact $K = K(x^*, \mathbf{s})$ such that $F(x, \mathbf{s}) \in K$ for all $x \in B_R(x^*)$ and for all $\mathbf{s} \in W_{\mathbf{s}^*}$, whence

$$\inf_{x \in B_R(x^*)} \inf_{\mathbf{s} \in W_{\mathbf{s}^*}} \mathbb{P}(X_{T-s_M} \in J_1) > 0,$$

where

$$J_1 = \bigcup_{\tilde{x}_M \in K} U_{\tilde{x}_M}.$$

Let now $x^{**} \in \mathbb{R}$ be arbitrarily chosen. The lower bound of (4.3) implies that there exists an open interval $I_1 \ni x^{**}$, such that

$$\inf_{x \in J_1, y \in I_1} P_{s_M}(x, y) > 0.$$

Therefore,

$$\beta_2 := \inf_{x \in B_R(x^*)} \inf_{y \in B_R(y^*)} \inf_{s \in W_y} \inf_{u \in J_1, v \in I_1} p_{s_M}(u, v) \mathbb{P}_x(X_{T-s_M} \in J_1) > 0. \tag{4.16}$$

Therefore, coming back to (4.15), for all $z \in B_R(x^*, y^*)$,

$$\begin{aligned}
P_T(z, A \times B) &\geq E_z(1_A(X_T) 1_B(Y_T), N_T^{(i)} = 1, \forall 1 \leq i \leq M, T_1^{(1)} < T_1^{(2)} < \dots < T_1^{(M)} < T) \\
&\geq \int_{W_y} q_y(\mathbf{s}) 1_B(\tilde{\gamma}_y(\mathbf{s}))\, \mathbb{E}\left(1_{\{X_{T-s_M} \in J_1\}} \int_A p_{s_M}(X_{T-s_M}, u) du\right) ds_1 \dots ds_M \\
&\geq \int_{W_y} q_y(\mathbf{s}) 1_B(\tilde{\gamma}_y(\mathbf{s}))\, \mathbb{E}\left(1_{\{X_{T-s_M} \in J_1\}} \int_{A \cap I_1} \inf_{x \in J_1} p_{s_M}(x, u) du\right) ds_1 \dots ds_M \\
&\geq \beta_2 \lambda(A \cap I_1) \int_{W_y} q_y(\mathbf{s}) 1_B(\tilde{\gamma}_y(\mathbf{s})) ds_1 \dots ds_M \\
&\geq \beta_2 \lambda(A \cap I_1) \beta_1 \lambda^M(B \cap I_2),
\end{aligned}$$

where we have finally applied (4.13), and where λ and λ^M denote the Lebesgue measure on $\mathbb{R}$, $\mathbb{R}^M$, respectively. This implies the desired result putting

$$\beta := \beta_1 \beta_2 \lambda(I_1) \lambda^M(I_2) \quad \text{and} \quad \nu := \mathcal{U}_{I_1 \times I_2},$$

the uniform probability law on $I_1 \times I_2$. □

4.3 A Foster-Lyapunov type condition

In order to prove the positive Harris recurrence of the process Z, we need of course a stability condition which is a Lyapunov type condition. Notice that the process $Z_t = (X_t, Y_t)$, has the following extended generator, defined for sufficiently smooth test functions f by

$$\begin{aligned}\mathcal{A}^Z f(x,y) &= -\sum_{i=1}^{M} \alpha_i y^{(i)} \partial_{y^{(i)}} f(x,y) + \partial_x f(x,y) b(x) + \frac{1}{2}\sigma^2(x)\partial_x^2 f(x,y) \\ &\quad + \sum_{i=1}^{M} y^{(i)} \left[f\left(x + a(x), y + c_{i\cdot}\right) - f(x,y) \right],\end{aligned}$$

with $c_{i\cdot} = \sum_{j=1}^{M} c_{i,j} e_j$.

Proposition 4.5. *Consider the function $V : \mathbb{R}^{M+1} \to \mathbb{R}^+,\ (x,y) \mapsto x^4 + \sum_{i=1}^{M} (y^{(i)})^2$. Then, there exist positive constants d_1, d_2 such that the following Foster-Lyapunov type drift condition holds*

$$\mathcal{A}^Z V \leq d_1 - d_2 V. \tag{4.17}$$

Proof. If $|x| \geq r$, under Assumption 2.2 and Assumption 2.3.2, with α given in (3.1) and $\bar{c}$ as in (2.7), we obtain

$$\begin{aligned}\mathcal{A}^Z V(x,y) &= -2\alpha \sum_{i=1}^{M} (y^{(i)})^2 + 4x^2 (xb(x) + \frac{3}{2}\sigma^2(x)) + \sum_{i=1}^{M} y^{(i)} \left[2\sum_{j=1}^{M} y^{(j)} c_{i,j} + \sum_{j=1}^{M} c_{i,j}^2 + (x + a(x))^4 - x^4 \right] \\ &\leq d_1 - d_2 \left(x^4 + \sum_{j=1}^{M} (y^{(j)})^2 \right),\end{aligned}$$

with two non negative constants d_1, d_2 depending on $\sigma_1, a_1, M, \alpha, \bar{c}, d$.
If $|x| \leq r$, we obtain

$$\mathcal{A}^Z V(x,y) = -\sum_{i=1}^{M} 2\alpha_i (y^{(i)})^2 + 4r^3 \sup_{|x| \leq r} |b(x)| + \frac{3}{2}\sigma_1^2 r^2 + \sum_{i=1}^{M} y^{(i)} \left[2\sum_{j=1}^{M} y^{(j)} c_{i,j} + \sum_{j=1}^{M} c_{i,j}^2 + 7r^4 + a_1^4 \right].$$

Finally for $x \in \mathbb{R}, y \in \mathbb{R}^+$, $\mathcal{A}^Z V(x,y) \leq d_1{}' - d_2{}' V(x,y)$ with $d_1{}', d_2{}' > 0$ depending on $\sigma_1, a_1, M, \alpha, \bar{c}$. □

4.4 Harris recurrence of Z

We do now possess all ingredients to obtain our main results.

Theorem 4.6. *Grant Assumptions 2.2, 2.3, 3.1 and 4.2. Then $(Z_t)_{t \geq 0}$ is positive Harris recurrent with unique invariant measure π. In particular, for any starting point z and any positive measurable function $g : \mathbb{R} \times \mathbb{R}^M \to \mathbb{R}_+$, as $T \to \infty$, $\mathbb{P}_z-$almost surely,*

$$\frac{1}{T}\int_0^T g(Z_s) ds \to \pi(g).$$

Proof of Theorem 4.6. 1) We fix any $x^* \in \mathbb{R}$ and we wish to apply Theorem 4.4 with $y^* = 0$ and $x^* = x^{**}$. Let R be the associated radius.

By Proposition 4.5, we know that for a suitable compact set $K = K_1 \times K_2$, with $K_1 \subset \mathbb{R}, K_2 \subset \mathbb{R}^M$, Z comes back to K infinitely often almost surely. Moreover,

$$\sup_{y \in K_2, t \geq 0} \|\varphi_t(y)\|_1 := F < \infty \quad \text{and} \quad \sup_{y \in K_2} \|\varphi_t(y)\|_1 \to 0$$

as $t \to \infty$, by the explicit form of the flow in (4.6). Therefore there exists t^* such that $\varphi_t(y) \in B_R(0)$ for all $t \geq t^*$, for all $y \in K_2$.

Applying once more the support theorem for diffusions and observing that σ is strictly positive, Equation (4.5) implies that

$$\inf_{x \in K_1} \mathbb{P}(\Phi_{t^*+s}(x) \in B_R(x^*), 0 \leq s \leq 2T) > 0$$

and thus

$$\inf_{z \in K} \mathbb{P}_z(X_{t^*+s} \in B_R(x^*), Y_{t^*+s} \in B_R(0), 0 \leq s \leq 2T) > 0.$$

Consequently, using a conditional version of the Borel-Cantelli lemma, the sampled Markov chain $(Z_{kT})_{k \in \mathbb{N}}$ visits $B_R(x^*, 0)$ infinitely often almost surely.
2) The standard regeneration technique (see e.g. [26]) allows to conclude that $(Z_{kT})_{k \in \mathbb{N}}$ and therefore $(Z_t)_t$ are Harris recurrent. This concludes the proof. □

The following by-product of the above result will prove to be useful when dealing with statistical inference within this new model class.

Proposition 4.7. *Grant Assumptions 2.2, 2.3, 3.1 and 4.2 and write π^X for the projection of the invariant measure π onto the $X-$coordinate, that is, $\pi^X(dx) = \int_{\mathbb{R}^M} \pi(dx, dy)$. Then π^X possesses a Lebesgue density which is bounded away from zero on each compact of $\mathbb{R}$.*

Proof. Let $A \in \mathcal{B}(\mathbb{R})$. Then for any $t > 0$,

$$\pi^X(A) = \int_{\mathbb{R} \times \mathbb{R}^M} \pi(dz) \mathbb{E}_z[1_A(X_t)]. \tag{4.18}$$

Let $L_t := \sup\{s \leq t : \exists j : \Delta N_s^{(j)} = 1\}$ be the last jump time of the process before time t. Then by Fubini,

$$\mathbb{E}_z[1_A(X_t)] = \int_A \mathbb{E}_z[p_{t-L_t}(X_{L_T}, y)] dy,$$

where $p_t(x, y)$ is the transition density of (4.3). This implies the existence of $\pi^X(x)$ given by

$$\pi^X(x) = \int_{\mathbb{R} \times \mathbb{R}^M} \pi(dz) \mathbb{E}_z[p_{t-L_t}(X_{L_t}, x)]$$

for any $t > 0$.

Notice that we do not dispose of any regularity results of $\pi^X(x)$ with respect to x. Indeed, the upper bound in (4.3) does not allow to conclude that the almost sure continuity in x of $p_{t-L_t}(X_{L_t}, x)$ survives the integration $\pi(dz)\mathbb{E}_z(\ldots)$.

We are however able to prove that π^X is lower bounded on compacts $K \subset \mathbb{R}$. For that sake, fix any $x^{**} \in K$ and apply (4.2) to $(x^*, y^*) \in supp(\pi)$ such that $\pi(B_R(z^*)) = \pi(C) > 0$ and to x^{**}. Then for any measurable $A \subset I_1 = I_1(x^{**})$, applying the lower bound of (4.2) to (4.18),

$$\int_A \pi^X(x) dx \geq \beta \pi(C) \int_A dx$$

implying that

$$\inf_{x\in I_1} \pi^X(x) \geq \beta\pi(C) > 0.$$

Therefore we have just shown that for all $x^{**} \in K$, there exists an open interval $x^{**} \subset I_1 = I_1(x^{**})$, such that π^X is strictly lower bounded on I_1. Since we can cover the compact K by a finite collection of such open intervals $I_1(x^{**})$, this implies the desired lower bound of π^X on compacts.

□

In the sequel, following [29], we introduce

$$\bar{V}(z) := V(z) + 1 \text{ and } \|\mu\|_{\bar{V}} := \sup_{g:|g|\leq\bar{V}} |\mu(g)|.$$

It is now straightforward to obtain our second main result.

Theorem 4.8 (Exponential ergodicity)**.** *Grant Assumptions 2.2, 2.3, 3.1 and 4.2. Then there exist $c_1, c_2 > 0$ such that for all $z \in \mathbb{R} \times \mathbb{R}^M$,*

$$\|P_t(z,\cdot) - \pi\|_{\bar{V}} \leq c_1 \bar{V}(z) e^{-c_2 t}. \tag{4.19}$$

Proof. The sampled chain $(Z_{kT})_{k\geq 0}$ is Feller according to Proposition 2.5. Moreover it is $\nu-$irreducible, where ν is the measure introduced in Theorem 4.4, associated with the point $(x^*, 0)$ and x^{**}, for any choice of $x^*, x^{**} \in \mathbb{R}$, used in the proof of Theorem 4.6. Since ν is the uniform measure on some open set of strictly positive Lebesgue measure, the support of ν has non-empty interior. Theorem 3.4 of [28] implies that all compact sets are 'petite' sets of the sampled chain. The Lyapunov condition established in Proposition 4.5 allows to apply Theorem 6.1 of [29] which implies the assertion. □

4.5 Exponential β-mixing for $Z = (X, Y)$.

It is now easy to deduce from the above results the exponentially β-mixing property of the process. Recall that the $\beta-$mixing coefficient of Z is given by

$$\beta_Z(t) = \sup_{s\geq 0} \int \|P_t(z,\cdot) - \mu P_{s+t}(\cdot)\|_{TV}\ \mu P_s(dz),$$

where $\mu = \mathcal{L}(Z_0)$ is the law of the initial configuration and where

$$\|\mu\|_{TV} := \sup_{g:|g|\leq 1} \mu(g)$$

denotes the total variation distance. Notice that if $\mu = \pi$, then the process is in its stationary regime, and

$$\beta_Z(t) = \int \|P_t(z,\cdot) - \pi\|_{TV} \pi(dz).$$

Theorem 4.9. *Grant Assumptions 2.2, 2.3, 3.1 and 4.2. Then Z is exponentially $\beta-$mixing, that is, there exist constants $K, \theta > 0$ such that for any initial law μ with $\mu(V) < \infty$,*

$$\beta_Z(t) \leq K e^{-\theta t}.$$

Proof. Suppose firstly that $\mu = \pi$. Then Theorem 4.3 of [29] implies that $\int V d\pi < \infty$ such that we are able to integrate (4.19) against $\pi(dx)$ to obtain

$$\beta_Z(t) \leq c_1 \pi(\bar{V}) e^{-c_2 t}.$$

Putting $K := c_1$ and $\theta = c_2$, this implies the result in this case.

In order to deal with the general non-stationary process, we apply Lemma 3.9 of [27] with $h = \bar{V}$, $\delta(t) = c_1 e^{-c_2 t}$ and

$$\kappa = \sup_{s \geq 0} \mathbb{E}(\bar{V}(Z_s)),$$

to deduce that

$$\beta_Z(t) \leq 2c_1 \kappa e^{-c_2 t}.$$

Putting $K := 2c_1\kappa$ and $\theta = c_2$, this implies the result, if we have already shown that κ is finite. This last fact follows immediately from (4.17), following the first lines of the proof of Theorem 6.1 of [29]. Indeed, we have by Dynkin's formula that

$$e^{\alpha t}\mathbb{E}_z(V(Z_t)) \leq V(z) + \frac{\beta}{\alpha}e^{\alpha t},$$

implying that

$$\mathbb{E}_z(V(Z_t)) \leq e^{-\alpha t}V(z) + \frac{\beta}{\alpha}.$$

Integrating this last inequality with respect to $\mu(dz)$ implies the result. □

5 Discussion

In this paper, we have established unique ergodicity and exponential β-mixing for a new class of Markov processes (X, λ) where X is a jump-diffusion and λ is the intensity process of a M-dimensional multivariate Hawkes process. These properties will be used in a following paper in a statistical context to study nonparametric estimators of the parameters of the process.

We can generalise these results to the non-linear case with eventually negative exponential kernels, that is, where negative weights $c_{ij} < 0$ are allowed. The same kind of probabilistic results should also hold if we replace the exponential kernels by Erlang kernels in the definition of the Hawkes intensity (see [12]).

Lastly, even if this will add some complexity to the model, it will be interesting to consider also in our model the case where we add jumps of the underlying diffusion process, which are driven by an underlying independent Lévy process. If we take the example of neurons, this means that we will add to the model the jumps of the fixed neurons (whose dynamic is described by the evolution X in between jumps triggered by the Hawkes process) besides the jumps of the M neurons around. It will be challenging to study this complete model.

References

[1] D. Applebaum. *Lévy processes and stochastic calculus.* Cambridge university press, 2009.

[2] E. Bacry, S. Delattre, M. Hoffmann, and J-F Muzy. Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Processes and their Applications*, 123(7):2475–2499, 2013.

[3] E. Bacry, I. Mastromatteo, and J-F Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.

[4] E. Bacry and J-F Muzy. Second order statistics characterization of Hawkes processes and non-parametric estimation. *arXiv preprint arXiv:1401.0903*, 2014.

[5] M. Benaïm, S. Le Borgne, F. Malrieu, and P-A Zitt. Qualitative properties of certain piecewise deterministic Markov processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 51(3):1040–1075, 08 2015.

[6] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, pages 1563–1588, 1996.

[7] S. Chen, A. Shojaie, E. Shea-Brown, and D. Witten. The multivariate Hawkes process in high dimensions: Beyond mutual excitation. *arXiv preprint arXiv:1707.04928*, 2017.

[8] K.S Crump. On systems of renewal equations. *Journal of Mathematical Analysis and Applications*, 30(2):425 – 434, 1970.

[9] Daryl J. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

[10] S. Delattre, N. Fournier, and M Hoffmann. Hawkes processes on large networks. *The Annals of Applied Probability*, 26(1):216–261, 2016.

[11] C. Dion and S. Lemler. Nonparametric drift estimation for diffusions with jumps driven by a Hawkes process. *HAL*, 2019.

[12] S. Ditlevsen and E. Löcherbach. Multi-class oscillating systems of interacting neurons. *Stoch. Proc. Appl.*, 127:840–1869, 2017.

[13] A. Duarte, E. Löcherbach, and G. Ost. Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels. *arXiv preprint arXiv:1610.03300*, 2016.

[14] E. Gobet. Lan property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(5):711–737, 2002.

[15] R.Z. Has'minskii. *Stochastic stability of differential equations*. Sijthoff & Noordhoff, 1980.

[16] A. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[17] A.G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[18] P. Hodara and E. Löcherbach. Hawkes processes with variable length memory and an infinite number of components. *Advances in Applied Probability*, 49(1):84–107, 2017.

[19] R. Höpfner and E. Löcherbach. Statistical models for Birth and Death on a Flow: Local absolute continuity and likelihood ratio processes. *Scandinavian Journal of Statistics*, 26(1):107–128, 1999.

[20] N. Ikeda, M. Nagasawa, and S. Watanabe. A construction of Markov processes by piecing out. *Proc. Japan Acad.*, 42(4):370–375, 1966.

[21] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland Mathematical Library. Elsevier Science, 2014.

[22] K. Ito and H.P McKean. Diffusion processes and their sample paths, 1965.

[23] T. Jaisson and M Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability*, 25(2):600–631, 2015.

[24] SA. Klokov and A Yu Veretennikov. On subexponential mixing rate for Markov processes. *Theory of Probability & Its Applications*, 49(1):110–122, 2005.

[25] H. Kunita. *Stochastic flows and stochastic differential equations*. Cambridge University Press, Cambridge, 1992.

[26] E. Löcherbach and D. Loukianova. On Nummelin splitting for continuous time Harris recurrent Markov processes and application to kernel estimation for multi-dimensional diffusions. *Stoch. Proc. Appl.*, 118:1301–1321, 2008.

[27] H Masuda. Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Processes and their Applications*, 117(1):35–56, 2007.

[28] S.P. Meyn and R.L. Tweedie. Stability of Markovian processes I : Criteria for discrete-time chains. *Adv. Appl. Probab.*, 24:542–574, 1992.

[29] S.P. Meyn and R.L. Tweedie. Stability of Markovian processes III : Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.*, 25:487–548, 1993.

[30] A. Millet and M Sanz-Solé. A simple proof of the support theorem for diffusion processes. *Séminaire de probabilités de Strasbourg*, 28:36–48, 1994.

[31] D Oakes. The Markovian self-exciting process. *Journal of Applied Probability*, 12(1):69–77, 1975.

[32] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media, 2013.

[33] P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Inference of functional connectivity in neurosciences via Hawkes processes. In *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, pages 317–320. IEEE, 2013.

[34] P. Reynaud-Bouret and E Roy. Some non asymptotic tail estimates for Hawkes processes. *Bulletin of the Belgian Mathematical Society-Simon Stevin*, 13(5):883–896, 2007.

[35] D.W. Stroock and S. R. S. Varadhan. On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Probability Theory*, pages 333–359, Berkeley, Calif., 1972. University of California Press.

[36] P. Tankov and E. Voltchkova. Jump-diffusion models: a practitioner's guide. *Banque et Marchés*, 99(1):24, 2009.

[37] A.Yu. Veretennikov. On polynomial mixing bounds for stochastic differential equations. *Stochastic Processes and their Applications*, 70(1):115 – 127, 1997.