Exponential ergodicity for diffusions with jumps driven by a Hawkes process
Charlotte Dion, Sarah Lemler, Eva Löcherbach

To cite this version:

HAL Id: hal-02094514
https://hal.science/hal-02094514v3
Submitted on 20 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Exponential ergodicity for diffusions with jumps driven by a Hawkes process

December 19, 2019

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 driven by a multivariate nonlinear Hawkes process. Our goal is to 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 $(X,Y)$, where $X$ denotes the diffusion process and $Y$ the piecewise deterministic Markov process (PDMP) defining 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.

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

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

1 Introduction

1.1 Motivation

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

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

where $W$ is the standard one-dimensional Brownian motion and $N$ is a multivariate Hawkes process having $M$ components with intensity process $\lambda$. $N$ 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 $\lambda$ 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 the entire 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.

Our main motivation is to be able to describe systems of interacting neurons, where we are in particular interested in modeling the membrane potential of a single neuron together with the sequence of spike times of its presynaptic neurons. Indeed, neurons are communicating through the emission of electrical signals, forming a network. When a (presynaptic) neuron sends a signal to another (postsynaptic) neuron, the membrane potential of the neuron 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 of one of its presynaptic neurons 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 presynaptic neurons which are modeled by the Hawkes process. Hawkes processes have been studied recently in the context of neurosciences [see 34, 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 aim of the present article is to study the longtime behavior of the process \( X \). Since \( X \) is not Markov on its own, we propose a study of the couple of processes \((X, \lambda)\) in the case when the intensity \( \lambda \) of \( N \) can be described in terms of a piecewise deterministic Markov process (PDMP) \( Y \), see (2.4) below. Such a description is possible when the memory kernels are exponential. We prove the positive Harris recurrence of \((X, Y)\) together with ergodicity, and we establish the speed of \( \beta\)-mixing. These are important probabilistic properties which are interesting in its own right but which are also the key ingredients for the theoretical study of estimators of the model parameters. The estimation of the coefficients of the process is the issue of the companion paper [11].

### 1.2 State of the art

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 of diffusion processes or the by now classical references [21, 33] in which all properties of diffusions are summarized. Ergodicity and strong mixing have been established in the pioneering papers [37, 23]. 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 ergodicity and mixing properties have been studied from a probabilistic point of view by [29].

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 (or to inhibit) 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, 8, 5]. Ergodicity for general non-linear Hawkes processes with general memory kernels has been recently established by [5]. For linear Hawkes processes, [15] proves exponential ergodicity, giving very precise estimates on the regeneration times. Finally [8] establishes exponential ergodicity in the particular case of linear Hawkes processes with exponential memory kernels.

Relying on one of these recent results on ergodicity of Hawkes processes, it is therefore reasonable to argue that the solution of (1.1) should be ergodic itself, provided we are able to ensure that the diffusion part of (1.1), that is the dynamic (1.1) with \( a(\cdot) \equiv 0 \), driving the motion in between successive
jumps, is ergodic.

However, no such general criterion, ensuring that the ergodicity of the underlying jump process implies the ergodicity of the associated diffusion with jumps, exists, at least not to our knowledge.

The present paper proposes therefore a short study of the ergodicity of diffusions with jumps driven by Hawkes processes with exponential memory kernels. In this case, the associated stochastic intensity processes of the jump part can be expressed in terms of particular Markov processes, PDMP’s, see e.g. [13]. For these PDMP’s, ergodicity has been established by [8] in the case of linear Hawkes processes, and by [13] in the case of non-linear Hawkes processes with Erlang kernels. Here, we extend these results to the general non-linear case, including also inhibitions \(^1\) and more importantly, we prove the joint ergodicity as well as the \(\beta\)-mixing property for the couple \((X, Y)\) in the case of exponential kernel functions. Finally, we give criteria ensuring that the mixing is exponential.

1.3 Main contributions and plan of the paper

We propose a general study of the longtime behavior of a jump diffusion process, where the jumps are driven by a multivariate non-linear Hawkes process in dimension \(M\), having exponential memory kernels. Our proof of the Harris recurrence of \((X, Y)\) follows a well-known scheme, see e.g. [30, 31].

Firstly, we establish a local Döblin lower bound for the transition semigroup of the process \((X, Y)\), based on a lower bound of the joint transition density of \((X, Y)\). This is the content of Theorem 3.1. The ideas that we use to obtain this lower bound are very natural, and they have appeared independently of our work already in [8]. They have also been exploited in [13] and in [26]: more precisely, we use \(M^2\) consecutive jumps which are carefully chosen to produce a part of Lebesgue density for the transition of \(Y\). 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. This part of the proof relies on fine support properties of \((X, Y)\) and provides, as a by-product, a simulation algorithm for the process. In particular we need to localize the possible positions of the diffusion part, conditionally on the jump times of the underlying Hawkes process that have already been chosen.

In a second step, a Foster-Lyapunov condition (see Proposition 3.4) 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 couple \((X, Y)\), as well as the ergodic theorem. This is our first main result, stated in Theorem 3.6.

Finally, following the steps of [29] where the ergodicity and exponential \(\beta\)-mixing bounds are established for diffusions with jumps driven by a Lévy process and the general framework of [24], our second main theorem 3.8 states the mixing property of the process together with a control on the rate of convergence. To conclude, our Theorem 3.10 provides sufficient criteria implying the exponential control of the \(\beta\)-mixing coefficient of \((X, Y)\).

Finally, let us mention that it is possible to extend our results to multivariate diffusions driven by Hawkes processes. However, the primary aim of our article is to provide a theoretical probabilistic foundation for a statistical application where we want to deal with data describing the membrane potential of one single neuron together with the spike trains (the point process data associated to the subsequent times of action potentials) of its presynaptic neurons. This is why we concentrate on a one-dimensional diffusion model, driven by a multivariate Hawkes process in this paper. Indeed, in this modeling context, the joint evolution of the membrane potentials of, say, \(d\) neurons, is independent, conditional on the evolution of the spike trains of their presynaptic neurons. This is due to the fact that the membrane potential processes of different neurons within a network do only interact via the

\(^1\) Notice that [7] has obtained a control of the \(\tau\)-mixing coefficient in the inhibitory case – but this does not imply the \(\beta\)-mixing property of the process.
incoming stimuli, that is, the incoming spike trains of the presynaptic neurons. In other words, their driving Brownian motions are independent - no interaction is present in their drift nor in their diffusion coefficient.

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 prove the ergodicity and finally the exponential $\beta$–mixing of the coupled process $(X,Y)$.

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 jump rate functions $f_i : \mathbb{R} \to \mathbb{R}_+, 1 \leq i \leq M$, and in terms of interaction functions (also called memory kernels) $h_{ij} : \mathbb{R}_+ \to \mathbb{R}, 1 \leq i, j \leq M$, 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.

**Assumption 2.1.** We suppose that each $f_i, 1 \leq i \leq M$, is Lipschitz continuous with Lipschitz constant $\gamma_i$.

**Definition 2.2.** A non-linear Hawkes process with parameters $(f_i, h_{ij})_{1 \leq i,j \leq M}$, and with initial condition $n^{(i)}, 1 \leq i \leq M$, is a multivariate counting process $N_t = (N_{1t}, \ldots, N_{Mt}), t \geq 0$, 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 $N_{it}^{(i)}$ is given by $\int_0^t \lambda_{it}^{(i)} ds$, where

$$\lambda_{it}^{(i)} = f_i \left( \sum_{j=1}^M \int_0^{t-} h_{ij}(t-u)n^{(j)}_u du + \sum_{j=1}^M \int_{-\infty}^0 h_{ij}(t-u)dn^{(j)}_u \right). \quad (2.1)$$

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. [6] or [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

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

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

**Proposition 2.3.** Suppose that Assumption 2.1 holds, that the coefficients $b, \sigma$ are globally Lipschitz continuous, that $a$ is measurable and that the memory kernel functions $h_{ij}$ are locally integrable. Then equation $(2.2)$ admits a unique strong solution.
Proof. Theorem 6 of [10] implies that for any \( T > 0 \), almost surely, \( N_t \) has only a finite number of jumps on \( [0, T] \). We can therefore work conditionally with respect to the choice of these jumps and construct the solution of (2.2) by pasting together trajectories of strong solutions to the diffusion equation (that is, the solution of (2.2) with \( a(\cdot) \equiv 0 \)) at the successive jump times (see [20] for related ideas).

In the sequel we shall also need estimates on the transition densities of the diffusion process underlying (2.2). We want to be able to deal with a general drift term \( b \) that is not bounded in order to include Ornstein-Uhlenbeck type processes to the class of models that we consider. Unbounded drift coefficients add however a substantial difficulty when one wishes to apply techniques from Malliavin calculus to obtain bounds on the transition densities. Therefore, we impose the following additional assumptions which are inspired by [14].

\[ \text{Assumption 2.4.} \]

1. \( a \) is continuous.
2. \( b \) and \( \sigma \) are of class \( C^2 \).
3. There exist positive constants \( c, q \) such that for all \( x \in \mathbb{R} \), \( |b'(x)| + |\sigma'(x)| \leq c \) and \( |b''(x)| + |\sigma''(x)| \leq c(1 + |x|^q) \).
4. There exist strictly positive constants \( \sigma_0 \) and \( \sigma_1 \) such that \( \sigma_0 \leq \sigma^2(x) \leq \sigma_1 \) for all \( x \in \mathbb{R} \).

\[ \text{Remark 2.5.} \] We do only need the above assumption to obtain good lower bounds on the transition densities of the underlying non-jumping diffusion process, see (3.3) below.

To study the longtime behavior of \( X \) and to ensure its ergodicity, we introduce two additional conditions which are classical in the study of diffusion processes and which are derived from [16], see also [37].

In the following, we assume we are in one of the following frames.

\[ \text{Assumption 2.6. Exponential frame.} \] There exist \( d \geq 0 \), \( r > 0 \) such that for all \( x \) satisfying \( |x| > r \), we have \( xb(x) \leq -dx^2 \). Moreover, one of the two following conditions holds true.

1. For all \( |x| > r \), \( 2xa(x) + a^2(x) \leq 0 \).
2. For all \( |x| > r \), \( |a(x)| \leq C|x|^\eta \) for some \( \eta < 1 \).

\[ \text{Assumption 2.7. Polynomial rate.} \] There exist \( \gamma > \sigma_1/2, r > 0 \) and \( m \in ]2, 1 + \frac{2\gamma}{\sigma_1}[, \) such that for all \( x \) satisfying \( |x| > r \), we have \( xb(x) \leq -\gamma \) and \( |x + a(x)|^m - |x|^m \leq 0 \).

\[ \text{2.2 Markovian framework} \]

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

\[ \text{Assumption 2.8.} \] For all \( 1 \leq i, j \leq M \),

\[ h_{ij}(t) = c_{ij} e^{-\alpha_{ij}t}, c_{ij} \in \mathbb{R}, \alpha_{ij} > 0. \]

(2.3)

In this case we may introduce the auxiliary Markov process \( Y = Y^{(ij)} \),

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

(2.4)
The intensities can be expressed in terms of sums of these Markov processes, that is, for all $1 \leq i \leq M$, 
\[ \lambda_t^{(i)} = f_i \left( \sum_{j=1}^{M} Y_t^{(ij)} \right). \]

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

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 $\sigma$. Note also that $N_t$ is an autonomous process, that is, $N_t$ does not depend on $X$.

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 2.9. Grant Assumptions 2.1 and 2.4. Then the process $Z$ is a Feller process, that is, $P_t$ belongs to $C_b(\mathbb{R} \times \mathbb{R}^{M \times M})$ whenever $\Phi \in C_b(\mathbb{R} \times \mathbb{R}^{M \times M})$.

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 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, see [20]. We start by giving the principal elements needed to construct this flow.

The associated stochastic Brownian flow. Thanks to our assumptions on $b$ and $\sigma$, by Theorem 4.2.5 of [25], there exists a unique stochastic flow of homeomorphisms $\Phi_t(x)$ for $0 \leq t < \infty, x \in \mathbb{R}$, such that 
\[ \Phi_t(x) = x + \int_{0}^{t} b(\Phi_u(x))du + \int_{0}^{t} \sigma(\Phi_u(x))dW_u. \quad (2.6) \]

In particular, we have that 
\[ (t, x) \mapsto \Phi_t(x) \text{ is continuous } (0 \leq t < \infty, \ x \in \mathbb{R}). \]

The stochastic flow $\Phi$ 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 
\[ \varphi_t(x) = (\varphi_t^{(ij)}(x))_{1 \leq i, j \leq M}, \quad \text{where} \quad \varphi_t^{(ij)}(x) = e^{-\alpha_{ij}t}y^{(ij)}, 1 \leq i, j \leq M. \quad (2.7) \]

We are now ready to give the simulation algorithm for $Z$. 

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 $\tilde{N}_1$ on the number of jumps of $Z$, that is, of $N$, during some finite time interval $[0, t]$. To do so, observe that by Lipschitz continuity of $f_i$ with Lipschitz constant $\gamma_i$, we have for $1 \leq i \leq M$,

$$f_i(x) \leq f_i(0) + \gamma_i|x|.$$  

Let

$$\tilde{\gamma} := \max_i \gamma_i, \quad \tilde{c} := \sup_{1 \leq i, j \leq M} |c_{ij}|.$$  

Consider now the one-dimensional auxiliary linear Hawkes process $N^*_t$ having intensity $\lambda^*_t, t \geq 0$, where

$$\lambda^*_0 := \sup_{y \in K_2} \sup_{t \geq 0} \sum_{i=1}^M f_i \left( \sum_{j=1}^M e^{-\alpha_{ij} t} g^{(ij)} \right)$$

and

$$\lambda^*_t = \lambda^*_0 + M\tilde{\gamma}\tilde{c} \int_0^t dN^*_u = \lambda^*_0 + M\tilde{\gamma}\tilde{c} N^*_{t-}, \quad t \geq 0.$$ 

Observe that $N^*$ is a one dimensional linear Hawkes process with memory kernel $\hat{h}(t) = M\tilde{\gamma}\tilde{c} 1_{[0, t]}(t)$. Obviously, for all $t \geq 0$,

$$\sum_{i=1}^M N^{(i)}_t \leq N^*_t.$$  

Moreover, for all $t \geq 0, N^*_t < \infty$ 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, also known as acceptance-rejection 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 $z_s = z_s(z, t^n_1, U^n_1)$, depending on these choices, for $0 \leq s \leq t$. Here, $t^n_1 = (t_1, \ldots, t_n)$ denotes the choices of the successive jump times and $U^n_1 = (U_1, \ldots, U_n)$ the sequence of indices of jumping particles. This process is defined recursively as follows. Firstly, we put

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

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

$$q(y, t, \{i\}) = \frac{f_i(\sum_j y^{(ij)})}{\lambda^*_i}, \quad 1 \leq i \leq M, \quad q(y, t, \{0\}) = 1 - \sum_{i=1}^M q(y, t, \{i\}).$$  

(2.9)

$U_1$ 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 $\{U_1 = 1\}$, we put

$$x_{t_1} := x_1 + a(x_1), \quad y^{(ij)}_{t_1} := y^{(ij)}_1 + c_{ij} 1_{\{U_1 = j\}}, \text{ for all } 1 \leq i, j \leq M,$$

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

Then we put

$$x_s = \Phi_{s-t_1}(x_{t_1}), \quad y_s = \varphi_{s-t_1}(y_{t_1})\text{ for all } t_1 \leq s < t_2,$$
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 
\[ \mathcal{L}(z_t(z, t^n_1, U^n_1)) = \mathcal{L}(Z_t|N^*_i = n, T^*_i = t_1, \ldots, T^n_i = t_n). \]
The important point is that the above construction ensures the continuity of the application 
\[ z \mapsto z_t(z, t^n_1, U^n_1). \]

Alternative proof of Proposition \[ 2.9 \]. Noticing that the law of \( N^* \) does not depend on the starting point \( z \) but only on an upper bound of the associated intensities, the above construction implies that for any \( \Phi \in C_b(\mathbb{R} \times \mathbb{R}^{M \times M}) \), the mapping 
\[ K \ni z \mapsto P.t \Phi(z) \in \mathbb{R} \]
is continuous, implying the assertion of Proposition \[ 2.9 \]. Indeed, we can write

\[
P_t \Phi(z) = \sum_{n \geq 0} \mathbb{P}(N^*_i = n) \int_{[0,t]^n} \mathcal{L}(T^*_1, \ldots, T^n_i|N^*_i = n)(dt_1, \ldots, dt_n) 
\sum_{i=1}^{n} \sum_{u_i=0}^{M} \mathbb{P}(U_1 = u_1, \ldots, U_n = u_n|T_1 = t_1, \ldots, T_n = t_n, Z_0 = z) \mathbb{E}^W \Phi(z_t(z, u^n_1)),
\]
where the last expectation \( \mathbb{E}^W \) is only taken with respect to the underlying Brownian motion \( W \).
Suppose now that \( (z^k)_k \subset K_1 \times K_2 \) is a sequence of initial configurations converging to \( z \) as \( k \to \infty \). Then, by the structure of \[ (2.9) \],
\[
\mathbb{P}(U_1 = u_1, \ldots, U_n = u_n|T_1 = t_1, \ldots, T_n = t_n, Z_0 = z^k) \to 
\mathbb{P}(U_1 = u_1, \ldots, U_n = u_n|T_1 = t_1, \ldots, T_n = t_n, Z_0 = z),
\]
as \( k \to \infty \), by continuity of \( f_i, 1 \leq i \leq M \), and of \( z \mapsto z_s(z, t^n_1, U^n_1) \), for all \( s \leq t \). Moreover, using dominated convergence, \( \mathbb{E}^W \Phi(z_t(z, u^n_1)) \to \mathbb{E}^W \Phi(z_t(z, u^n_1)) \), as \( k \to \infty \), implying that \( P_t \Phi(z^k) \to P_t \Phi(z) \), as \( k \to \infty \), where we have used dominated convergence once more.

3  Ergodicity of the process \((X, Y)\)

In this section, we start proving the positive Harris recurrence of \( Z_t = (X_t, Y_t) \) using the regeneration technique.

3.1 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 \((Z_t)\).

**Theorem 3.1.** Grant Assumptions \[ 2.4 \] and \[ 2.8 \] and suppose moreover that \( f_i(x) > 0 \) for all \( 1 \leq i \leq M \), for all \( x \in \mathbb{R} \). Then there exists \( T > 0 \) such that for all \( z^* = (x^*, y^*) \in \mathbb{R} \times \mathbb{R}^{M \times M} \) and for all \( x^{**} \in \mathbb{R} \) the following holds. There exist \( R > 0 \), open sets \( I_1 \subset \mathbb{R} \) and \( I_2 \subset \mathbb{R}^{M \times M} \) with strictly positive Lebesgue measure such that

\[ x^{**} \in I_1, \quad (3.1) \]
and a constant $\beta \in (0, 1)$, depending on $I_1, I_2, R$ and the coefficients of the system with
\[ P_T(z, dz') \geq \beta \mathbb{I}_{C}(z)\nu(dz'), \tag{3.2} \]
where $C = B_R(z^*)$ is the (open) ball of radius $R$ centred at $z^*$, and where $\nu$ is the uniform probability measure on $I_1 \times I_2$.

**Remark 3.2.**

1. Together with a Lyapunov argument implying that the process $Z$ comes back to the set $C$ infinitely often, the above result will imply the $\nu-$irreducibility of the sampled chain $(Z_{kT})_{k \geq 0}$.

We want to stress that the above construction implies that for any given $x^{**} \in \mathbb{R}$ and some $\varepsilon > 0$ we can construct $\nu$ such that $B_{\varepsilon}(x^{**})$ lies in the support of the projection on the $x-$variable of $\nu$. Of course, this property is related to the support properties of the underlying diffusion process, granted by Assumption 2.4. It will imply that the invariant density of $X$ is bounded away from 0 on each compact set, see Proposition 3.7 below.

2. A lower bound of the type (3.2) is trivial if we are only interested in the transitions of the process $X$. Indeed, in this case it is sufficient to consider the event where no jump has appeared up to time $T$ and to use known lower bounds on the transition densities of the diffusion part. However, since $X$ is not Markov on its own, we do need to work with the couple of processes $(X, Y)$, and therefore have to establish such lower bounds for the joint transition of $X$ and $Y$.

The main idea of the proof is to use the jump noise of $M^2$ successive jumps of the underlying Hawkes process to create firstly a Lebesgue density for the process $X$. Such ideas have already been exploited in [8] and in [13]. The important second step is then to use density estimates of the underlying diffusion to prove that a joint Lebesgue density of $(X_T, Y_T)$ exists which is strictly lower-bounded on $C$.

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

### 3.2 Some useful properties of the underlying stochastic flow and proof of Theorem

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 $\sigma$, the stochastic flow $\Phi_t(x)$ given by Equation (2.6), possesses a transition density $x \mapsto p_t(x, y)$ with explicit lower bounds. More precisely, Proposition 1.2 of [14] implies that, for some suitable constants $c, C$,
\[ e^{-1}t^{-1/2}e^{-C(x-y)^2/t}e^{-Ct}\leq p_t(x, y) \leq ct^{-1/2}e^{-\frac{1}{2}(x-y)^2/t}e^{Ct}. \tag{3.3} \]

Here, the constants $c$ and $C$ do only depend on the coefficients $b$ and $\sigma$.\footnote{It is possible to replace our Condition 2.4 by any other condition ensuring that $p_t(x, y)$ is strictly lower bounded on compact sets for any $t > 0$.}

**Support properties.** We will use the control theorem which goes back to Stroock and Varadhan (1972) [35], see also Millet and Sanz-Sole (1994) [32], Theorem 3.5. For some time horizon $T_1 < \infty$ which is arbitrary but fixed, write $H$ for the Cameron-Martín space of measurable functions $h : [0, T_1] \rightarrow \mathbb{R}$ having absolutely continuous functions $h(t) = \int_0^t h(s)ds$ with $\int_0^{T_1} (h)^2(s)ds < \infty$. For $x \in \mathbb{R}$ and $h \in H$, consider the deterministic flow
\[ (\varphi_t^{[b]}(x))_{t \geq 0} \text{ solution to } d\varphi_t^{[b]}(x) = \tilde{b}(\varphi_t^{[b]}(x))dt + \sigma(\varphi_t^{[b]}(x))h(t)dt, \text{ with } \varphi_0^{[b]}(x) = x, \tag{3.4} \]
on $[0, T_1]$. In the above formula (3.4), $\tilde{b}$ is Stratonovich drift given by
\[ \tilde{b}(x) = b(x) - \frac{1}{2} \sigma(x)\sigma'(x). \]
Denote by $Q^{T_i}_x$ the law of the solution $(\Phi_i(x))_{0 \leq t \leq T_i}$. Then for any $x \in \mathbb{R}$ and $h \in H$,

$$
(\varphi^{(b)}_t(x))_{t \in [0,T_i]} \in \text{supp} (Q^{T_i}_x).
$$

(3.5)

We are now able to give the proof of Theorem 3.1.

**Proof.** We suppose without loss of generality that for all $1 \leq j \leq M$

$$c_{ij} \neq 0 \text{ for all } i, \alpha_{ij} \neq \alpha_{kj}, \text{ for all } 1 \leq i, k \leq M, i \neq k. \quad \text{[3]}
$$

In what follows, we impose first $M$ consecutive jumps of $N^{(1)}$, followed by $M$ jumps of $N^{(2)}$ etc up to $M$ consecutive jumps of $N^{(M)}$, and we suppose that they all happen before time $T$. Then we can lower-bound, for all $A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{B}(\mathbb{R}^{M \times M})$, $z \in \mathbb{R} \times \mathbb{R}^{M \times M}$,

$$P_T(z \times A) \geq \mathbb{E}_z \left( \mathbb{1}_A(X_T) \mathbb{1}_B(Y_T), N^{(i)}_T = M, \forall 1 \leq i \leq M, T^{(1)}_1 < T^{(1)}_2 < \ldots < T^{(1)}_M < T^{(2)}_1 < \ldots < T^{(2)}_M < \ldots < T^{(M)}_1 < \ldots < T^{(M)}_M < T \right),$$

where $T^{(i)}_k, k \geq 1$, are the successive jump times of $N^{(i)}, 1 \leq i \leq M$. In what follows we shall write

$$A_T := \{N^{(i)}_T = M, \forall 1 \leq i \leq M, T^{(1)}_1 < T^{(2)}_1 < \ldots < T^{(1)}_M < T^{(2)}_1 < \ldots < T^{(2)}_M < \ldots < T^{(M)}_1 < \ldots < T^{(M)}_M < T \}.$$

**Step 1.** We first deal with the process $Y_t$. This part of the proof follows well-known arguments that have been developed independently from our work in [8], and that we have also already used in [13] and [20]. We reproduce the main arguments here since the second step will be a control of the diffusion part conditionally on the results of this first step.

Recall that in between successive jump events of $N$, the process $Y$ evolves according to the deterministic flow $\varphi_t(y)$ introduced in (2.7) above. Thus, on the event $A_T$, starting from $Y_0 = y \in \mathbb{R}^{M \times M}$, we first let the flow $\varphi$ evolve starting from $y$ up to some first jump time $t_1$. At that jump time each particle having index $(i)$ gains an additional value $c_{ij}$. We then successively choose the following inter-jump waiting times $t_2, \ldots, t_{M^2}$ under the constraint $t_1 + \ldots + t_{M^2} < T$. We write

$$s_1 = T - t_1, s_2 = T - (t_1 + t_2), \ldots, s_{M^2} = T - (t_1 + \ldots + t_{M^2}).$$

Conditionally on $Y_0 = y$, the successive choices of $s = (s_1, \ldots, s_{M^2})$ as above, the position of $Y^{(i)}_T$ is given by

$$\gamma^{(ij)}(y,s) = e^{-\alpha_{ij}T}y^{(ij)} + c_{ij}[e^{-\alpha_{ij}s_{M-M+1}} + \ldots + e^{-\alpha_{ij}s_{M}}].$$

In what follows we work at fixed $y$ and we write

$$\gamma_y : s \mapsto \gamma(y,s) = (\gamma^{(11)}, \ldots, \gamma^{(M1)}, \gamma^{(12)}, \ldots, \gamma^{(M2)}, \ldots, \gamma^{(1M)}, \ldots, \gamma^{(MM)})(y,s).$$

We will use the jump noise which is created by the $M^2$ jumps, i.e., we will use a change of variables on the account of $s_1, \ldots, s_{M^2}$. Therefore, let

$$\frac{\partial \gamma_y(s)}{\partial s} = \left[ \frac{\partial \gamma_y(s)}{\partial s_1}, \ldots, \frac{\partial \gamma_y(s)}{\partial s_{M^2}} \right].$$

[3] Otherwise, if $\alpha_{ij} = \alpha_{kj}$, we introduce the new process $Y^{(i+k)}_{t} := Y^{ij}_t + Y^{kj}_t$ which is Markov again.
be the Jacobian matrix of the map $\mathbf{s} \mapsto \gamma_y(\mathbf{s})$. This matrix does not depend on the initial position $y$. Indeed, one easily finds that

$$
\frac{\partial \gamma_y(\mathbf{s})}{\partial \mathbf{s}} = C(\mathbf{s}) = \begin{pmatrix}
C^{(1)} & 0 & \ldots & 0 & 0 \\
0 & C^{(2)} & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & 0 & C^{(M)}
\end{pmatrix}(\mathbf{s}),
$$

where for each $1 \leq j \leq M$, $C^{(j)}(\mathbf{s})$ is the $M \times M$ matrix given by

$$
C^{(j)}(\mathbf{s}) = \begin{pmatrix}
c_{1j}(\alpha_{1j})^{-1}e^{-\alpha_{1j}s_{j,M-M+1}} & \cdots & c_{1j}(\alpha_{1j})^{-1}e^{-\alpha_{1j}s_{j,M}} \\
c_{2j}(\alpha_{2j})^{-1}e^{-\alpha_{2j}s_{j,M-M+1}} & \cdots & c_{2j}(\alpha_{2j})^{-1}e^{-\alpha_{2j}s_{j,M}} \\
\vdots & \vdots & \vdots \\
c_{Mj}(\alpha_{Mj})^{-1}e^{-\alpha_{Mj}s_{j,M-M+1}} & \cdots & c_{Mj}(\alpha_{Mj})^{-1}e^{-\alpha_{Mj}s_{j,M}}
\end{pmatrix}.
$$

We have to prove that each $C^{(j)}(\mathbf{s})$ is invertible. This is difficult for general choices of $\mathbf{s}$. But if we choose $t_0$ such that $M^2t_0 < T$ and then $s^*_1 = M^2t_0$, $s^*_2 = (M^2 - 1)t_0$, $\ldots$, $s^*_M = t_0$ such that $s_k - s_{k+1} = t_0$ for all $1 \leq k < M^2$, then

$$
\det C^{(j)}(\mathbf{s}^*) = \left( \prod_{i=1}^{M} c_{ij}(\alpha_{ij})^{-1}e^{-\alpha_{ij}s_{j,M}} \right) \det V^{(j)}(t_0),
$$

where

$$
V^{(j)}(t_0) = \begin{pmatrix}
e^{-((M-1)\alpha_{1j}t_0)} & \cdots & e^{-2\alpha_{1j}t_0} & e^{-\alpha_{1j}t_0} & 1 \\
e^{-((M-1)\alpha_{2j}t_0)} & \cdots & e^{-2\alpha_{2j}t_0} & e^{-\alpha_{2j}t_0} & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
e^{-((M-1)\alpha_{Mj}t_0)} & \cdots & e^{-2\alpha_{Mj}t_0} & e^{-\alpha_{Mj}t_0} & 1
\end{pmatrix}.
$$

Therefore, $\det V^{(j)}(t_0)$ is a Vandermonde determinant which is different from 0 since by assumption all $e^{-\alpha_{ij}t_0}$ and $e^{-\alpha_{ij}t_0}$ are different, for all $i \neq k$.

By continuity we therefore have that for any choice of $t_0 < T/M^2$ there exists $\varepsilon > 0$ such that for all $\mathbf{s} \in B_\varepsilon(\mathbf{s}^*)$, $\frac{\partial \gamma_y(\mathbf{s})}{\partial \mathbf{s}}$ is invertible.

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 (3.2). For that sake, we shall also need the following notation. For each couple $(y, \mathbf{s})$, we write $y_0 = y$, and then define recursively for all $jM - M + 1 \leq k \leq jM$, where $j$ varies between $1 \leq j \leq M$,

$$
y_k^{(i\ell)} = \varphi_{s_k - s_{k-1}}(y_{k-1}) + c_{ij}I\{t = j\}, \quad (3.6)
$$

for all $1 \leq i, \ell \leq M$, where we put $s_0 = T$.

The sequence $y_1, \ldots, y_{M^2}$ corresponds to the positions of the process $Y$ right after the successive jumps, starting from the initial location $y \in \mathbb{R}^{M \times M}$, induced by the inter-jump waiting times $T - s_1, s_1 - s_2, \ldots, s_{M^2-1} - s_{M^2}$ which are determined by $\mathbf{s}$.

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

$$
\bar{f}(y) := \sum_{i=1}^{M} f_i \left( \sum_j y^{(ij)} \right)
$$

and for each $t \geq 0$, the survival rate

$$
e(y, t) = \exp \left\{ - \int_0^t \bar{f}(\varphi_s(y)) ds \right\}. \quad (3.7)
$$
We define for each couple \((y,s)\) (recall that \(s_0 = T\)),
\[
q_y(s) = \left( \prod_{j=1}^M \prod_{k=1}^{M-1} f_j(\varphi_{s_k - s_{k+1}}(y_k)) e(y_k, s_k - s_{k+1}) \right) e(y_{M^2}, s_{M^2}). \tag{3.8}
\]

Since \(f(y) > 0\) for all \(y \in \mathbb{R}^{M \times M}\) and from the definition of \(e(\cdot, \cdot)\), we deduce that for any couple \((y^*, s^*)\) there are neighborhoods \(W_{s^*}\) and \(U_{y^*}\) of \(s^*\) and \(y^*\) respectively such that
\[
\inf_{(y,s) \in U_{y^*} \times W_{s^*}} q_y(s) > 0. \tag{3.9}
\]

Let us now fix \((y^*, s^*)\) such that the matrix \(\frac{\partial \gamma_y(s^*)}{\partial s}\) is invertible. By Lemma 6.2 of [4], there exist an open neighborhood \(B_R(y^*)\) of \(y^*\), an open set \(I_2 \subset \mathbb{R}^{M \times M}\), and for any \(y \in B_R(y^*)\), an open set \(W_y\) such that
\[
\tilde{\gamma}_y(s) : \begin{cases} W_y \to I_2 \\ s \mapsto \gamma_y(s) \end{cases}
\]

is a diffeomorphism, with \(W_y \subset W_{s^*}\), and also
\[
\inf_{y \in B_R(y^*)} \inf_{s \in W_y} \left| \det \left( \frac{\partial \gamma_y(s)}{\partial s} \right)^{-1} \right| > 0. \tag{3.10}
\]

Reducing (if necessary) \(R\), we may assume also that \(B_R(y^*) \subset U_{y^*}\). Thus we have that by \(3.9\) and \(3.10\),
\[
\beta_1 := \inf_{y \in B_R(y^*)} \inf_{s \in W_y} q_y(s) \det \left( \frac{\partial \gamma_y(s)}{\partial s} \right)^{-1} > 0. \tag{3.11}
\]

Once we have done all these steps we can conclude with the following preliminary result. For any measurable \(B \in \mathcal{B}(\mathbb{R}^{M \times M})\) and for any \(z = (x,y)\) such that \(y \in B_R(y^*)\), using the change of variables \(y = \tilde{\gamma}_y(s)\), we obtain the lower bound
\[
\begin{aligned}
\mathbb{E}_y \left( \mathbf{1}_B(Y_T), N_T^{(i)} = M, \forall 1 \leq i \leq M, T_1^{(1)} < \ldots < T_{M}^{(1)} < T_1^{(2)} < \ldots < T_{M}^{(M)} < T \right) \\
\geq \int_{W_y} q_y(s) \mathbf{1}_B(\tilde{\gamma}_y(s)) ds_1 \ldots ds_{M^2} \geq \beta_1 \int_{I_2 \cap B} dy_1 \ldots dy_{M^2}. \tag{3.12}
\end{aligned}
\]

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 \(s_1, \ldots, s_{M^2}\) of the first step. Analogously to \(3.6\) we therefore introduce the successive jump positions of the process \(X\) which are given by \(x_0 = x\) and then for all \(1 \leq k \leq M^2\),
\[
x_k = \Phi_{s_{k-1} - s_k}(x_{k-1}) + a(\Phi_{s_{k-1} - s_k}(x_{k-1})), \tag{3.13}
\]

where \(s_0 = T\) as before.

Conditionally on \(X_0 = x, Y_0 = y\), the successive choices of \(s = (s_1, \ldots, s_{M^2})\) as above, the position of \(X_T\) is then given by
\[
\Gamma(z, s) = \Phi_{s_{M^2}}(x_{M^2}),
\]

where \(z = (x,y)\). Notice that \(x_{M^2}\) depends on the choices of \(s_1, \ldots, s_{M^2}\) and of course on the evolution of the stochastic flow between the successive jump times.
Step 3. Therefore, conditioning with respect to \( X_{T-s_{M2}} = x_{M2} \), we obtain

\[
P_T(z, A \times B) \geq \mathbb{E}_z \left( \mathbb{I}_A(X_T) \mathbb{I}_B(Y_T), A_T \right) x
\]

\[
\geq \int_{W_\gamma} q_\gamma(s) \mathbb{I}_B(\tilde{\gamma}_\gamma(s)) \mathbb{E} \left( \int_A p_{s_{M2}}(x_{M2}, u) du \right) ds_1 \ldots ds_{M2}, \quad (3.14)
\]

where \( p_{s_{M2}}(x_{M2}, u) \) is the transition density of \((3.3)\).

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 \( X_{T-s_{M2}} \) under the Wiener measure, conditionally on the choices \( s_1, \ldots, s_{M2} \).

The lower bound estimate given in \((3.3)\) implies that for fixed \( s_{M2} \) and \( x_{M2} \), \( u \mapsto p_{s_{M2}}(x_{M2}, u) \) is lower bounded in small neighborhoods of \( x_{M2} \). Therefore, in what follows we need to localize the possible positions of \( x_{M2} \), conditionally on the choices of the jumps times \( s_1, \ldots, s_{M2} \).

Step 4. Localization of \( x_{M2} \).

We apply the support theorem, that is, \((3.5)\), to our process \( X_t \) in between the successive jump times \( 0, T_1 := t_1, T_2 := t_1 + t_2, \ldots, T_{M2} := t_1 + \ldots + t_{M2} \), by choosing on each time interval \([T_n, T_{n+1}], n = 0, \ldots, M^2 - 1\), the control function \( h \equiv 0 \). Consequently, introducing

\[
\tilde{x}_1 := \varphi^{(0)}_t(x) + a(\varphi^{(0)}_t(x)), \ldots, \tilde{x}_{M2} := \varphi^{(0)}_{s_{M2-1}-s_{M2}}(\tilde{x}_{M2-1}) + a(\varphi^{(0)}_{s_{M2-1}-s_{M2}}(\tilde{x}_{M2-1})),
\]

there exists an open neighborhood \( U_{\tilde{x}_{M2}} \) of \( \tilde{x}_{M2} \), such that

\[
\mathbb{P}(X_{T-s_{M2}} \in U_{\tilde{x}_{M2}}) > 0.
\]

Notice that \( \tilde{x}_{M2} \) is a continuous function of the starting point \( x \) and of \( s \); that is, \( \tilde{x}_{M2} = F(x, s) \) for some continuous function \( F \). This implies that for any starting point \( x^* \) and for \( R > 0 \) sufficiently small, reducing \( W_{s^*} \) if necessary, there exists a compact \( K = K(x^*, s^*) \) such that \( F(x, s) \in K \) for all \( x \in B_R(x^*) \) and for all \( s \in W_{s^*} \), whence

\[
\inf_{x \in B_R(x^*)} \inf_{s \in W_{s^*}} \mathbb{P}(X_{T-s_{M2}} \in J_1) > 0,
\]

where

\[
J_1 = \bigcup_{\tilde{x}_{M2} \in K} U_{\tilde{x}_{M2}}.
\]

Notice that \( J_1 \) has compact closure.

Let now \( x^{**} \in \mathbb{R} \) be arbitrarily chosen. The lower bound of \((3.3)\) implies that there exists an open interval \( I_1 \ni x^{**} \), such that

\[
\inf_{v \in I_1} \inf_{y \in [x^{**}, s_{M2}]} p_{s_{M2}}(x, y) > 0.
\]

Therefore,

\[
\beta_2 := \inf_{x \in B_R(x^*)} \inf_{y \in [x^{**}, s_{M2}]} \inf_{u \in I_1, v \in I_1} p_{s_{M2}}(u, v) \mathbb{P}_x(X_{T-s_{M2}} \in J_1) > 0. \quad (3.15)
\]

Therefore, coming back to \((3.14)\), for all \( z \in B_R(x^*, y^*) \),

\[
P_T(z, A \times B) \geq \mathbb{E}_z \left( \mathbb{I}_A(X_T) \mathbb{I}_B(Y_T), A_T \right) x
\]

\[
\geq \int_{W_\gamma} q_\gamma(s) \mathbb{I}_B(\tilde{\gamma}_\gamma(s)) \mathbb{E} \left( \int_A p_{s_{M2}}(x_{M2}, u) du \right) ds_1 \ldots ds_{M2}
\]

\[
\geq \int_{W_\gamma} q_\gamma(s) \mathbb{I}_B(\tilde{\gamma}_\gamma(s)) \mathbb{E} \left( \int_{A \cap I_1} p_{s_{M2}}(x, u) du \right) ds_1 \ldots ds_{M2}
\]

\[
\geq \beta_2 \lambda(A \cap I_1) \int_{W_\gamma} q_\gamma(s) \mathbb{I}_B(\tilde{\gamma}_\gamma(s)) ds_1 \ldots ds_{M2}
\]

\[
\geq \beta_2 \lambda(A \cap I_1) \beta_1 \lambda_{M \times M}(B \cap I_2),
\]

13
where we have finally applied (3.12), and where λ and \( \lambda^{M \times M} \) denote the Lebesgue measure on \( \mathbb{R} \), \( \mathbb{R}^{M \times M} \), respectively. This implies the desired result putting

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

the uniform probability law on \( I_1 \times I_2 \).

3.3 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 \( g \) by

\[
\mathcal{A}^Z g(x, y) = - \sum_{i,j=1}^{M} \alpha_{ij} y^{(ij)} \partial_{y^{(ij)}} g(x, y) + \partial_x g(x, y) b(x) + \frac{1}{2} \sigma^2(x) \partial_x^2 g(x, y)
+ \sum_{j=1}^{M} f_j \left( \sum_{k=1}^{M} y^{(jk)} \right) [g(x + a(x), y + \Delta_j) - g(x, y)],
\]

with \((\Delta_j)^{(il)} = c_{ij} \mathbb{I}_{\{j=i\}}\) for all \( 1 \leq i, l \leq M \).

To obtain stability we first introduce the following classical stability assumption for Hawkes processes. Recall that \( \gamma_i \) denotes the Lipschitz constant of the rate function \( f_i \).

**Assumption 3.3.** Let \( H \) be the \( M \times M \)-matrix with entries \( H_{ij} = \gamma_j \frac{|c_{ij}|}{\alpha_{ij}} \), \( 1 \leq i, j \leq M \). Then its spectral radius satisfies

\[
\rho := \rho(H) < 1.
\]

Under the above stability condition, let \( \kappa \in \mathbb{R}^M_+ \) be a left eigenvector of \( H \), associated to the eigenvalue \( \rho \) and having non-negative components, that is, for all \( j \), \( \sum_i \kappa_i H_{ij} = \rho \kappa_j \geq 0 \). Such a vector \( \kappa \) exists thanks to the theorem of Perron-Frobenius. Following [8], we introduce a Lyapunov function by defining \( m_{ij} = \frac{\kappa_i}{\alpha_{ij}} \) for all \( 1 \leq i, j \leq M \) and \( V : \mathbb{R} \times \mathbb{R}^{M \times M} \to \mathbb{R}_+ \) by

\[
V(x, y) = V_1(x) + e^{\sum_{i,j} m_{ij} |y^{(ij)}|},
\]

where \( V_1 : \mathbb{R} \to \mathbb{R}_+ \) will be chosen in the sequel. Notice that \( V(x, y) \geq 1 \) for all \( x, y \).

**Proposition 3.4.** Grant Assumptions 2.4, 2.6 (Exponential frame) and 3.3. Let \( V_1(x) = x^2 \). Then there exist positive constants \( d_1, d_2 \) such that the following Foster-Lyapunov type drift condition holds

\[
\mathcal{A}^Z V \leq d_1 - d_2 V. \tag{3.17}
\]

**Proof.** We write \( V(x, y) = V_1(x) + V_2(y) \), where \( V_1(x) = x^2 \), \( V_2(y) = e^{\sum_{i,j} m_{ij} |y^{(ij)}|} \), and \( \mathcal{A}^Z V = \mathcal{A}_1^Z V + \mathcal{A}_2^Z V \), where

\[
\mathcal{A}_1^Z V(x, y) = 2x b(x) + \sigma^2(x) + \sum_{j=1}^{M} f_j \left( \sum_{k=1}^{M} y^{(jk)} \right) [(x + a(x))^2 - x^2]
\]

is the diffusion part and \( \mathcal{A}_2^Z V = \mathcal{A}_2^Z V - \mathcal{A}_1^Z V \) is the jump part of the generator. The arguments of the proof of Proposition 4.5 of [8] imply that

\[
\mathcal{A}_2^Z V(x, y) = \mathcal{A}_2^Z V_2(y) \leq -c_1 V_2(y) + c_2 \mathbb{I}_{K_1}(y), \tag{3.18}
\]

where

\[
K_1(y) = \{ (x, y) : |y^{(ij)}| \leq \beta \},
\]

and \( \beta \) is the maximum over \( \mathbb{R}^M \) of \( \frac{|c_{ij}|}{\alpha_{ij}} \).
where \(c_1, c_2 > 0\) and where \(K_1 \subset \mathbb{R}^{M \times M}\) is some compact. Moreover, writing for short \(\tilde{f}(y) = \sum_i f_i(\sum_j y^{(ij)})\) for the total jump rate,

\[
\mathcal{A}_y^2 V(x, y) = 2xb(x) + \sigma^2(x) + (2xa(x) + a(x)^2)\tilde{f}(y).
\] (3.19)

Thanks to Assumption 2.7 and bound of \(\sigma^2\) from Assumption 2.4,

\[
2xb(x) + \sigma^2(x) \leq -c_3x^2 + c_41_{K_2}(x),
\] (3.20)

where \(K_2 \subset \mathbb{R}\) is a compact set. Now, suppose that Assumption 2.6 is satisfied. Then for all \(|x| > r\), \((2xa(x) + a(x)^2)\tilde{f}(y) \leq 0\) (recall that \(\tilde{f}\) is non-negative) implying that

\[
\mathcal{A}_y^2 V(x, y) \leq -c_3x^2 + c_41_{K_2}(x) + c_5\tilde{f}(y).
\]

This implies (3.17) since

\[
\tilde{f}(y) \leq c_6 + c_7 \sum_{i,j} |y^{(ij)}| \leq c_6 + c_7 \log(V_2(y))
\]

which is thus negligible with respect to the negative term \(-c_3V_2(y)\) of (3.18).

If only Assumption 2.7 holds, then we upper bound the jump part of (3.19) by

\[
(2xa(x) + a(x)^2)\tilde{f}(y) \leq C\tilde{f}(y) + C|x|^{1+\eta}\tilde{f}(y).
\]

Choose \(1 < p < 2\) and \(q > 2\) such that \((1+\eta)p < 2\) and \(\frac{1}{p} + \frac{1}{q} = 1\). Then

\[
|x|^{1+\eta}\tilde{f}(y) \leq \frac{1}{p}|x|^{p(1+\eta)} + \frac{1}{q}\tilde{f}(y).
\]

Since \((1+\eta)p < 2\), the first term

\[
C\frac{1}{p}|x|^{(1+\eta)p}
\]

will again be negligible with respect to the negative term \(-c_3x^2\) of (3.20). Finally, the second term \(C\tilde{f}(y) + C\frac{1}{q}\tilde{f}(y)\) is treated as previously. This concludes our proof. \(\square\)

**Proposition 3.5.** Grant Assumptions 2.4, 2.7 (Polynomial frame) and 3.3. Let \(V_1(x) = 1 + |x|^\alpha\) and \(\alpha = 2/m \in [0, 1]\). Then there exist positive constants \(d_1, d_2\) such that the following Foster-Lyapunov type drift condition holds

\[
\mathcal{A}_y^2 V \leq d_1 - d_2V^{1-\alpha}.
\] (3.21)

**Proof.** Using the same arguments as in the proof of Proposition 3.4, we have that

\[
\mathcal{A}_y^2 V(x, y) = \mathcal{A}_y^2 V_2(y) \leq -c_1V_2(y) + c_21_{K_1}(y).
\]

Moreover, for all \(|x| \geq r\), since \(|x + a(x)|^m - |x|^m| \leq 0,

\[
\mathcal{A}_y^2 V(x, y) = b(x)V_1'(x) + \frac{1}{2}\sigma^2(x)V_1''(x) + \sum_{j=1}^{M} f_j \left( \sum_{k=1}^{M} y^{(jk)} \right) |x + a(x)|^m - |x|^m
\]

\[
\leq b(x)V_1'(x) + \frac{1}{2}\sigma^2(x)V_1''(x).
\]

A standard calculus, see e.g. 37 or 28, shows that, for suitable constants \(e_1, e_2 > 0\),

\[
b(x)V_1'(x) + \frac{1}{2}\sigma^2(x)V_1''(x) \leq e_1 - e_2V_1(x)^{1-\alpha}.
\]

Using that \(V_1(x) \geq 1\) such that \((V_1(x) + V_2(y))^{1-\alpha} \leq V_1(x)^{1-\alpha} + (1-\alpha)V_2(y) \leq V_1(x)^{1-\alpha} + V_2(y)\), we deduce from this the drift condition (3.21) as in the proof of Proposition 3.4. \(\square\)
3.4 Harris recurrence of $Z$

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

**Theorem 3.6.** Grant Assumptions 2.4, 2.6, 2.8 and 3.3 and suppose that for all $1 \leq i \leq M$, $f_i(x) > 0$ for all $x \in \mathbb{R}$. Then $(Z_t)_{t \geq 0}$ is positive Harris recurrent with unique invariant measure $\pi$. In particular, for any starting point $x$ and any positive measurable function $g : \mathbb{R} \times \mathbb{R}^{M \times M} \to \mathbb{R}_+$, as $T \to \infty$, $\mathbb{P}_x$–almost surely,

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

**Proof of Theorem 3.6.** 1) We fix any $x^* \in \mathbb{R}$ and we wish to apply Theorem 3.1 with $y^* = 0$ and $x^* = x^*$. Let $R$ be the associated radius.

By Proposition 3.4 or Proposition 3.5 we know that for a suitable compact set $K = K_1 \times K_2$, with $K_1 \subset \mathbb{R}, K_2 \subset \mathbb{R}^{M \times M}$, $Z$ comes back to $K$ infinitely often almost surely. Moreover,

$$\sup_{y \in K_1, t \geq 0} \|\varphi_t(y)\|_1 := F < \infty \quad \text{and} \quad \sup_{y \in K_2} \|\varphi_t(y)\|_1 \to 0$$

as $t \to \infty$, by the explicit form of the flow in (2.7). 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 $\sigma$ is strictly positive, Equation (3.5) implies that

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

and thus

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

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. [27]) 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 3.7.** Grant all assumptions of Theorem 3.6 and write $\pi^X$ for the projection of the invariant measure $\pi$ onto the $X$–coordinate, that is, $\pi^X(dx) = \int_{\mathbb{R}^{M \times M}} \pi(dx, dy)$. Then $\pi^X$ 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 $t > 0$,

$$\pi^X(A) = \int_{\mathbb{R}^{M \times M}} \pi(dx) \mathbb{E}_z[\mathbb{1}_A(X_t)]. \quad (3.22)$$

Let $L_t := \sup\{s \leq t : \exists j : \Delta N^{(j)}_s = 1\}$ be the last jump time of the process before time $t$, $L_t = 0$ if there is no such jump. Then by Fubini,

$$\mathbb{E}_z[\mathbb{1}_A(X_t)] = \int_A \mathbb{E}_z[p_{t-L_t}(X_{L_t}, y)]dy,$$
where \( p_t(x, y) \) is the transition density of (3.3). This implies the existence of the density \( \pi^X(x) \) which is given by

\[
\pi^X(x) = \int_{\mathbb{R}^2} \pi(dz)E_z[p_{t-L_t}(X_{L_t}, x)]
\]

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 (3.3) does not allow to conclude that the almost sure continuity in \( x \) of \( p_{t-L_t}(X_{L_t}, x) \) survives the integration \( \pi(dz)E_z(\ldots) \).

We are however able to prove that \( \pi^X \) is lower bounded on compacts \( K \subset \mathbb{R} \). For that sake, fix any \( x^{**} \in K \) and apply (3.2) to \((x^*, y^*) \in \text{supp}(\pi)\) 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 (3.2) to (3.22) and taking \( t = T \),

\[
\int_A \pi^X(x)dx \geq \beta \pi(C) \frac{1}{\lambda(I_1)} \int_A dx
\]

implying that

\[
\inf_{x \in I_1} \pi^X(x) \geq \beta \pi(C) \frac{1}{\lambda(I_1)} > 0.
\]

Therefore we have just shown that for all \( x^{**} \in K \), there exists an open interval \( I_1 = I_1(x^{**}) \) containing \( x^{**} \in I_1 \) such that \( \pi^X \) 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 \( \pi^X \) on compacts.

\[ \square \]

In the sequel, following [31], in any of the two frames (Assumption 2.6 or 2.7) and for the choice of \( V \) as in Proposition 3.4 or Proposition 3.5 we introduce

\[
\|\mu\|_V := \sup_{g:|g|\leq V} |\mu(g)|, \quad \|\mu\|_{TV} := \sup_{g:|g|\leq 1} \mu(g).
\]

It is now straightforward to obtain our second main result.

**Theorem 3.8 (Ergodicity).** Grant all assumptions of Theorem 3.6. Then there exist \( c_1, c_2 > 0 \) such that for all \( z \in \mathbb{R} \times \mathbb{R}^{2M \times 2M} \), under Assumption 2.6

\[
\|P_t(z, \cdot) - \pi\|_V \leq c_1 V(z) e^{-c_2 t}, \tag{3.23}
\]

and under Assumption 2.7

\[
\|P_t(z, \cdot) - \pi\|_{TV} \leq c_1 V(z) t^{\frac{1}{2} - 1}, \tag{3.24}
\]

where \( \alpha \) is as in Proposition 3.5.

**Proof.** The sampled chain \((Z_{kT})_{k\geq 0}\) is Feller according to Proposition 2.9. Moreover it is \( \nu \)-irreducible, where \( \nu \) is the measure introduced in Theorem 3.1 associated with the point \((x^*, 0)\) and \( x^{**} \), for any choice of \( x^*, x^{**} \in \mathbb{R} \), used in the proof of Theorem 3.6. Since \( \nu \) is the uniform measure on some open set of strictly positive Lebesgue measure, the support of \( \nu \) has non-empty interior. Theorem 3.4 of [30] implies that all compact sets are ‘petite’ sets of the sampled chain. Under Condition 2.6 the Lyapunov condition established in Proposition 3.4 allows to apply Theorem 6.1 of [31] which implies the assertion in the exponential frame. In the polynomial frame, the assertion follows from Theorem 3.2 of [12] or from Theorems 1.19 and 1.23 in [24]. \( \square \)
Remark 3.9. In the above Theorem 3.8, we do only treat two cases for the rate of convergence to equilibrium: the exponential and the polynomial one. Using slightly different Lyapunov functions, we could also deal with more general sub-exponential rates of convergence, as they have for example been considered in Theorem 1.20 in [24]. We do not detail these calculations here since we are mostly interested in exponential rates of convergence. Finally, notice that it is also possible to interpolate between the total variation norm considered in (3.24) and the weighted total variation norm $\| \cdot \|_V$ considered in (3.23), however at the cost of slower rates of convergence. Details can be found in the very instructive paper [12].

3.5 Exponential $\beta$-mixing for $Z = (X,Y)$.

It is now easy to deduce from the above results the exponentially $\beta$-mixing property of the process, under Condition 2.6. Recall that the $\beta$-mixing coefficient of $Z$ is given by

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

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 3.10. Grant all assumptions of Theorem 3.6 and suppose that Condition 2.6 holds. Then $Z$ is exponentially $\beta$-mixing, that is, there exist constants $K, \theta > 0$ such that for any initial law $\mu$ with $\mu(V) < \infty$,

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

Proof. Suppose firstly that $\mu = \pi$. Then Theorem 4.3 of [31] implies that $\int Vd\pi < \infty$ such that we are able to integrate (3.23) against $\pi(dx)$ to obtain

$$
\beta_Z(t) \leq c_1 \pi(V)e^{-c_2t}.
$$

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 [29] with $h = \bar{V}$, $\delta(t) = c_1e^{-c_2t}$ and

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

to deduce that

$$
\beta_Z(t) \leq 2c_1\kappa e^{-c_2t}.
$$

Putting $K := 2c_1\kappa$ and $\theta = c_2$, this implies the result, if we have already shown that $\kappa$ is finite. This last fact follows immediately from (3.17), following the first lines of the proof of Theorem 6.1 of [31]. Indeed, we have by Dynkin’s formula that

$$
e^{-at}E_z(V(Z_t)) \leq V(z) + \frac{\beta}{\alpha}e^{at},$$

implying that

$$
E_z(V(Z_t)) \leq e^{-at}V(z) + \frac{\beta}{\alpha}.
$$

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


