ESAIM: PS 23 (2019) 770–796
https://doi.org/10.1051/ps/2019005

ESAIM: Probability and Statistics
www.esaim-ps.org

STABILITY, CONVERGENCE TO EQUILIBRIUM AND SIMULATION OF NON-LINEAR HAWKES PROCESSES WITH MEMORY KERNELS GIVEN BY THE SUM OF ERLANG KERNELS

Aline Duarte[1], Eva Löcherbach[2] and Guilherme Ost[3,*]

Abstract. Non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels are considered. It is shown that their stability properties can be studied in terms of an associated class of piecewise deterministic Markov processes, called *Markovian cascades of successive memory terms*. Explicit conditions implying the positive Harris recurrence of these processes are presented. The proof is based on integration by parts with respect to the jump times. A crucial property is the non-degeneracy of the transition semigroup which is obtained thanks to the invertibility of an associated Vandermonde matrix. For Lipschitz continuous rate functions we also show that these Markovian cascades converge to equilibrium exponentially fast with respect to the Wasserstein distance. Finally, an extension of the classical thinning algorithm is proposed to simulate such Markovian cascades.

Mathematics Subject Classification. 60G55, 60J75, 60K10.

Received July 12, 2017. Accepted March 28, 2019.

1. Introduction

Hawkes processes have regained a lot of interest in the recent years, in particular in econometrics, as good models to account for contagion risk and clustering arrival of events. They have shown to be very useful in neuroscience also due to their capacity of reproducing the typical time dependencies observed in spike trains of neurons as well as the interaction structure of neural nets. Originally introduced by [13] and [14] as a model for the appearances of earthquakes, their key feature is the fact that any event is able to trigger future events – for this reason, Hawkes processes are sometimes called “self-exciting point processes”. In their by now classical paper, Bremaud and Massoulie [4] develop the stability theory of general non-linear Hawkes processes, also in a multivariate frame. Hansen, Reynaud-Bouret and Rivoirard [12] have put the foundations for the use of Hawkes processes as models of spike trains in neuroscience, see [6] also, and recently some effort has been spent to study Hawkes processes in high dimensions, especially focusing on properties such as the propagation of chaos, see [8] and [5], see [9] also in a multi-class frame. Finally, we refer to [23] for a study of the large deviation properties of non-linear Hawkes processes having Markovian intensity function.

Keywords and phrases: Hawkes processes, Erlang kernels, piecewise deterministic Markov process (PDMP), longtime behaviour, coupling.

[1] Universidade de São Paulo, São Paulo, Brazil.
[2] Université de Paris 1 Panthéon Sorbonne, Paris, France.
[3] Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

[*] Corresponding author: guilhermeost@im.ufrj.br

In the present paper Hawkes processes with memory kernels given by the sum of Erlang kernels are considered. It is shown that the longtime behaviour and stability properties of these processes can be studied in terms of an associated class of piecewise deterministic Markov processes (PDMPs). More precisely, let N be a counting process on $\mathbb{R}_+$ characterised by its intensity process $(\lambda_t)_{t\geq 0}$ defined, for each $t \geq 0$, through the relation

$$\mathbb{P}(N \text{ has a jump in }]t, t+\mathrm{d}t] | \mathcal{F}_t) = \lambda_t \mathrm{d}t,$$

where $\mathcal{F}_t = \sigma(N(]u, s]), 0 \leq u < s \leq t)$ and

$$\lambda_t = f\left(\delta + \int_{]0,t[} h(t-s)\mathrm{d}N_s\right). \tag{1.1}$$

Here, $f : \mathbb{R} \to \mathbb{R}_+$ is the *jump rate function* and $h : \mathbb{R}_+ \to \mathbb{R}$ is the *memory kernel.* The parameter $\delta \in \mathbb{R}$ is interpreted as an initial input to the jump rate function.

The memory kernel h is assumed to be given by the sum of Erlang kernels, that is, $h = \sum_{i=1}^L h_i$, where each function h_i is of the form

$$h_i(t) = c_i e^{-\alpha_i t} \frac{t^{n_i}}{n_i!}, t \geq 0, \tag{1.2}$$

where $c_i \in \mathbb{R}, \alpha_i > 0$ and $n_i \in \mathbb{N}$.

Erlang kernels are widely used in the modelling literature, for example to model delays in the hemodynamics in nephrons, see [10, 22] or to prove the existence of oscillations in large-scale limits of interacting neurons in a mean-field frame, see [9]. This is the main motivation for the particular choice taken for the memory kernel h. Moreover, it is well known that the class of memory kernels having this form is dense in $L^1(\mathbb{R}_+)$, see *e.g.*, [16]. Therefore, any Hawkes process having integrable memory kernel can be well approximated by a Hawkes process having an Erlang memory kernel, see (2.9) below, at least over compact time intervals. Finally, the specific memory structure induced by Erlang kernels allow a completely new approach to simulation of non-linear Hawkes processes.

Erlang kernels depend on three parameters, c_i, n_i and α_i. Here, $n_i + 1$ is the *order* of the delay of the influence of a past event on a future event. It takes its maximum absolute value at $(n_i + 1)/\alpha_i$ time units back in time. The mean is $(n_i + 1)/\alpha_i$ (if normalising to a probability density). The higher the order of the delay, the more concentrated is the delay around its mean value, and in the limit of $n_i \to \infty$ while keeping $(n_i + 1)/\alpha_i$ fixed, the delay converges to a discrete delay. The sign of c_i indicates if the influence of past events on future events is inhibitory or excitatory.

The use of Erlang kernels allows to relate the study of the longtime behaviour of a Hawkes process having intensity (1.1) to the study of an associated system of PDMPs. More specifically, it is easily shown that the system of stochastic processes $X_t^{(i,0)} = \int_{]0,t]} h_i(t-s)\mathrm{d}N_s, t \geq 0$, for each $1 \leq i \leq L$, can be completed, by introducing $\sum_{i=1}^L n_i$ auxiliary processes, to a piecewise deterministic Markov process in dimension $L + \sum_{i=1}^L n_i$, see (2.4) below. Between successive jumps of N, the evolution of each $X_t^{(i,0)}$, together with its auxiliary processes, is explicitly given by a deterministic flow. Jumps do only occur in the auxiliary variables $X^{(i,n_i)}, 1 \leq i \leq L$. We shall call this class of PDMPs *Markovian cascades of successive memory terms.*

We prove that these Markovian cascades are recurrent in the sense of Harris under the usual sub-criticality condition

$$\|f\|_{Lip} \int_0^\infty |h(t)|\mathrm{d}t < 1.^{1} \tag{1.3}$$

[1] if all α_i are equal and all c_i are of the same sign. In the case of bounded rate functions, we do not need to impose this condition.

Under (1.3), we are able to construct a Lyapunov function implying that these processes return to a compact set infinitely often, almost surely. Under the additional condition of some minimal *ellipticity*, that is, some minimal jump activity, we establish, in Theorem 3.11, a Doeblin like lower bound based on integration by parts with respect to the jump times. A crucial property is the non-degeneracy of the transition semigroup which is obtained, thanks to the invertibility of an associated Vandermonde matrix and structure of the flow of the Markovian cascades (see (2.8) below). In the case of Lipschitz continuous rate functions we also show that the Markovian cascades converge to equilibrium exponentially fast with respect to the Wasserstein distance.

The fact that the flow governing the evolution of the Markovian cascades in between successive jump times is explicitly given and enables us to introduce an efficient simulation algorithm which allows to sample from N on $[0,T]$ for any finite time horizon $T>0$ and any fixed parameter $\delta\in\mathbb{R}$. This method is straightforward to implement, and can be easily extended to multi-dimensional versions.

The paper is organised as follows. In Section 2, we present the model and provide some preliminary remarks. In Section 3, the long-time behaviour of the Markovian cascades is investigated. The statement and proof for the case $L=1$ of Theorem 3.11, establishing the Doeblin lower bound for the Markovian cascades, is also included in this section. In Section 4, a simulation algorithm to simulate a Hawkes process simultaneously with memory kernel given by the sum of Erlang kernels and its Markovian cascade is proposed. In Section 5, numerical examples are presented. Finally, in the Appendix A, we prove Theorem 3.11 in the general case.

2. Model definition and preliminary remarks

Throughout the article the set $\mathbb{N}$ denotes the set of non-negative integers, $\mathbb{N}^*$ the set of positive integers $\{1,2,\ldots\}$ and $\mathcal{B}((0,\infty))$ (resp. $\mathcal{B}((a,b])$, for real numbers $0\le a\le b<\infty$) the Borel sigma-algebra on $(0,\infty)$ (resp. on $(a,b]$).

We work on the following filtered space $(\Omega,\mathcal{F},\mathbb{F})$. Let Ω be the canonical path space of simple point processes, *i.e.*,

$$\Omega=\{w=(t_n)_{n\in\mathbb{N}^*}\in]0,\infty]^{\mathbb{N}^*}:t_n\le t_{n+1},t_n<t_{n+1}\text{ if }t_n<+\infty,\ \lim_{n\to+\infty}t_n=+\infty\}.$$

For each $w\in\Omega$ and $n\in\mathbb{N}^*$, we define $T_n(w)=t_n$. For each $w\in\Omega$, we associate the canonical point measure $\mathcal{B}((0,\infty))\ni A\mapsto N(w)(A)=\sum_{n\in\mathbb{N}^*}\delta_{T_n(w)}(A)$. We shall write for short $N(A)$ rather than $N(w)(A)$; when $A=(0,t]$ for some $t\ge 0$, we simply write N_t to denote $N((0,t])$. Finally, we define $\mathcal{F}_t=\sigma\{N(A):A\in\mathcal{B}((0,t])\}$ for each $t\ge0$, $\mathcal{F}=\sigma\{N(A):A\in\mathcal{B}((0,\infty))\}$ and $\mathbb{F}=(\mathcal{F}_t)_{t\ge0}$.

Let $f:\mathbb{R}\to\mathbb{R}_+$ and $h:\mathbb{R}_+\to\mathbb{R}$ be measurable functions and let n be a deterministic point process on $]-\infty,0]$ such that $\int_{]-\infty,0]}h(t-s)n(\mathrm{d}s)$ is finite for all $t\ge0$.

Definition 2.1. A Hawkes process with parameters (f,h) and with initial condition n is a probability measure P on the filtered space $(\Omega,\mathcal{F},\mathbb{F})$ such that the compensator of $(N_t)_{t\ge0}$ is given by $(\int_0^t\lambda_s\mathrm{d}s)_{t\ge0}$, where $(\lambda_t)_{t\ge0}$ is the non-negative $\mathbb{F}-$predictable process defined for $t\ge0$ by

$$\lambda_t=f\left(\int_{]-\infty,0]}h(t-s)n(\mathrm{d}s)+\int_{]0,t[}h(t-s)\mathrm{d}N_s\right).\tag{2.1}$$

The stochastic process $(\lambda_t)_{t\ge0}$ is called *intensity process*. The functions $f:\mathbb{R}\to\mathbb{R}_+$ and $h:\mathbb{R}_+\to\mathbb{R}$ are called *jump rate function* and *memory kernel* respectively. We shall work under the following assumptions.

Assumption 2.2. The rate function $f:\mathbb{R}\to\mathbb{R}_+$ is either bounded or Lipschitz continuous with Lipschitz constant $\|f\|_{Lip}$.

Assumption 2.3. The memory kernel $h : \mathbb{R}_+ \to \mathbb{R}$ can be written as sum of Erlang kernels, *i.e*, for each $t \geq 0$,

$$h(t) = \sum_{i=1}^{L} c_i e^{-\alpha_i t} \frac{t^{n_i}}{n_i!}, \tag{2.2}$$

where for each $1 \leq i \leq L$, $c_i \in \mathbb{R}$, $\alpha_i > 0$ and $n_i \in \mathbb{N}$.

Under Assumption 2.3, the intensity process (2.1) can be described by an associated PDMP. Indeed, for each $1 \leq i \leq L$ and $0 \leq k \leq n_i$, writing for each $t \geq 0$,

$$X_t^{(i,k)} = \int_{]-\infty,0]} c_i e^{-\alpha_i (t-s)} \frac{(t-s)^{(n_i-k)}}{(n_i-k)!} n(\mathrm{d}s) + \int_{]0,t]} c_i e^{-\alpha_i (t-s)} \frac{(t-s)^{(n_i-k)}}{(n_i-k)!} \mathrm{d}N_s, \tag{2.3}$$

we have

$$\lambda_t = f\left(X_{t-}^{(1,0)} + \ldots + X_{t-}^{(L,0)}\right),$$

and one easily deduces that for $t \geq 0$ and $1 \leq i \leq L$,

$$\begin{aligned} \mathrm{d}X_t^{(i,0)} &= X_t^{(i,1)}\mathrm{d}t - \alpha_i X_t^{(i,0)}\mathrm{d}t \\ &\;\vdots \\ \mathrm{d}X_t^{(i,n_i-1)} &= X_t^{(i,n_i)}\mathrm{d}t - \alpha_i X_t^{(i,n_i-1)}\mathrm{d}t \\ \mathrm{d}X_t^{(i,n_i)} &= -\alpha_i X_t^{(i,n_i)}\mathrm{d}t + c_i \mathrm{d}N_t, \end{aligned} \tag{2.4}$$

with initial condition $X_0^{(i,k)} = x_0^{(i,k)} = \int_{]-\infty,0]} c_i e^{\alpha_i s} \frac{(-s)^{(n_i-k)}}{(n_i-k)!} n(\mathrm{d}s)$.

Write $\kappa = L + \sum_{i=1}^{L} n_i$. The associated PDMP is the Markov process $X = (X_t)_{t\geq 0}$ having càdlàg paths and taking values in $\mathbb{R}^\kappa$, defined, for each $t \geq 0$, by

$$X_t = \left(X_t^{(1)}, \ldots, X_t^{(L)}\right) \text{ with } X_t^{(i)} = \left(X_t^{(i,0)}, \ldots, X_t^{(i,n_i)}\right),\ 1 \leq i \leq L. \tag{2.5}$$

If $L = 1$, that is, h is a pure Erlang kernel, we write for short $X_t = (X_t^{(k)}, 0 \leq k \leq n_1)$. We call the process X *Markovian cascade of successive memory terms*. Its infinitesimal generator $\mathcal{L}$ is given for any smooth test function $g : \mathbb{R}^\kappa \mapsto \mathbb{R}$ by

$$\mathcal{L}g(x) = \langle F(x), \nabla g(x)\rangle + f\Big(\sum_{i=1}^{L} x^{(i,0)}\Big)\Big(g\big(x + \sum_{i=1}^{L} c_i e_{(i,n_i)}\big) - g(x)\Big), \tag{2.6}$$

where $x = \big(x^{(1)}, \ldots, x^{(L)}\big)$ with $x^{(i)} = (x^{(i,0)}, \ldots, x^{(i,n_i)})$ and $e_{(i,n_i)} \in \mathbb{R}^\kappa$ is the unit vector having entry 1 in the coordinate (i, n_i), and 0 elsewhere. Finally, $F : \mathbb{R}^\kappa \mapsto \mathbb{R}^\kappa$ is the vector field associated to the system of first-order ODE's

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} x_t^{(i,0)} = x_t^{(i,1)} - \alpha_i x_t^{(i,0)} \\ \vdots \\ \frac{\mathrm{d}}{\mathrm{d}t} x_t^{(i,n_i-1)} = x_t^{(i,n_i)} - \alpha_i x_t^{(i,n_i-1)} \\ \frac{\mathrm{d}}{\mathrm{d}t} x_t^{(i,n_i)} = -\alpha x_t^{(i,n_i)}, \ 1 \le i \le L, \end{cases} \tag{2.7}$$

given by $F(x) = ((F^{(1)}(x), \ldots, F^{(L)}(x)),$ where $F^{(i)}(x) = (F^{(i,0)}(x), \ldots, F^{(i,n_i)}(x))$ with

$$F^{(i,k)}(x) = -\alpha_i x^{(i,k)} + x^{(i,k+1)} \text{ for } 0 \le k < n_i, \text{ and } F^{(i,n_i)}(x) = -\alpha_i x^{(i,n_i)}.$$

Notice that jumps introduce discontinuities only in the coordinates $X_t^{(i,n_i)}$ of X_t. Figure 1 depicts a realisation of the joint processes $(N_t, X_t)_{t\ge 0}$ in the case $L = 1$.

Hereafter, we write $\varphi_t(x) = (\varphi_t^{(1)}(x), \ldots, \varphi_t^{(L)}(x))$ for the unique solution, starting from $x \in \mathbb{R}^\kappa$, of the system (2.7). It is immediate to check that for each $1 \le i \le L$ and $t \ge 0$,

$$\varphi_t^{(i)}(x) = \begin{pmatrix} e^{-\alpha_i t}(x^{(i,0)} + t x^{(i,1)} + \ldots + \frac{t^{n_i}}{n_i!} x^{(i,n_i)}) \\ \vdots \\ e^{-\alpha_i t}(x^{(i,n_i-1)} + t x^{(i,n_i)}) \\ e^{-\alpha_i t} x^{(i,n_i)} \end{pmatrix}. \tag{2.8}$$

Notice that $\varphi_t^{(i)}(x)$ depends only on the variable $x^{(i)}$. Given the Markovian cascade of successive memory terms (2.4)–(2.6), one recovers immediately the non-linear Hawkes processes with intensity (2.1) as shows the following proposition. In what follows, for any $x_0 \in \mathbb{R}^\kappa$, we write P_{x_0} for the probability measure on $\mathbb{R}^\kappa$ under which $X_0 = x_0$, and denote by E_{x_0} the expectation taken with respect to P_{x_0}.

Proposition 2.4. *Suppose Assumptions 2.2 and 2.3. Fix any initial condition n on $]-\infty, 0]$ such that $\int_{]-\infty,0]} h(t-s) n(\mathrm{d}s) < \infty$ for all $t \ge 0$ and put, for each $1 \le i \le L$ and $0 \le k \le n_i$, $x_0^{(i,k)} = \int_{]-\infty,0]} c_i e^{\alpha_i s} \frac{(-s)^{(n_i-k)}}{(n_i-k)!} n(\mathrm{d}s)$. Let $X = (X_t)_{t\ge 0}$ be the Markov process having generator (2.6), starting from $X_0 = x_0$. Then X is non-explosive, i.e., X has P_{x_0}-almost surely a finite number of jumps on each interval $[s,t], 0 \le s < t < \infty$. Finally, introduce $N_t = \sum_{s\le t} 1_{\{\Delta X_s \ne 0\}}$ the counting process associated to the jumps of X. Then $(N_t)_{t\ge 0}$ is a non-linear Hawkes process with intensity (2.1).*

Proof. Let us define $\mu(\mathrm{d}t) = \sum_{n\ge 1} \delta_{T_n}(\mathrm{d}t)$ and $\nu(\mathrm{d}t) = f\big(\sum_{i=1}^L X_t^{(i,0)}\big)\mathrm{d}t$. Then Proposition 3.1 in [15] implies that ν is the predictable compensator of μ. In particular, the compensator of N_t is given by $\int_0^t \lambda_s \mathrm{d}s$ with $\lambda_s = f\big(\sum_{i=1}^L X_{s-}^{(i,0)}\big)$.

It remains to prove the non-explosiveness of the process X. In the case of bounded f, nothing has to be proved. Suppose therefore that f is Lipschitz continuous and define $g(x) = f(0)(\sum_{i=1}^L |c_i|) + \sum_{i=1}^L \sum_{k=0}^{n_i} |x^{(i,k)}|$ for $x \in \mathbb{R}^\kappa$. Let $A = \max_{1\le i\le L} \alpha_i$ and $c = \sum_{i=1}^L |c_i|$. By plugging g in (2.6),

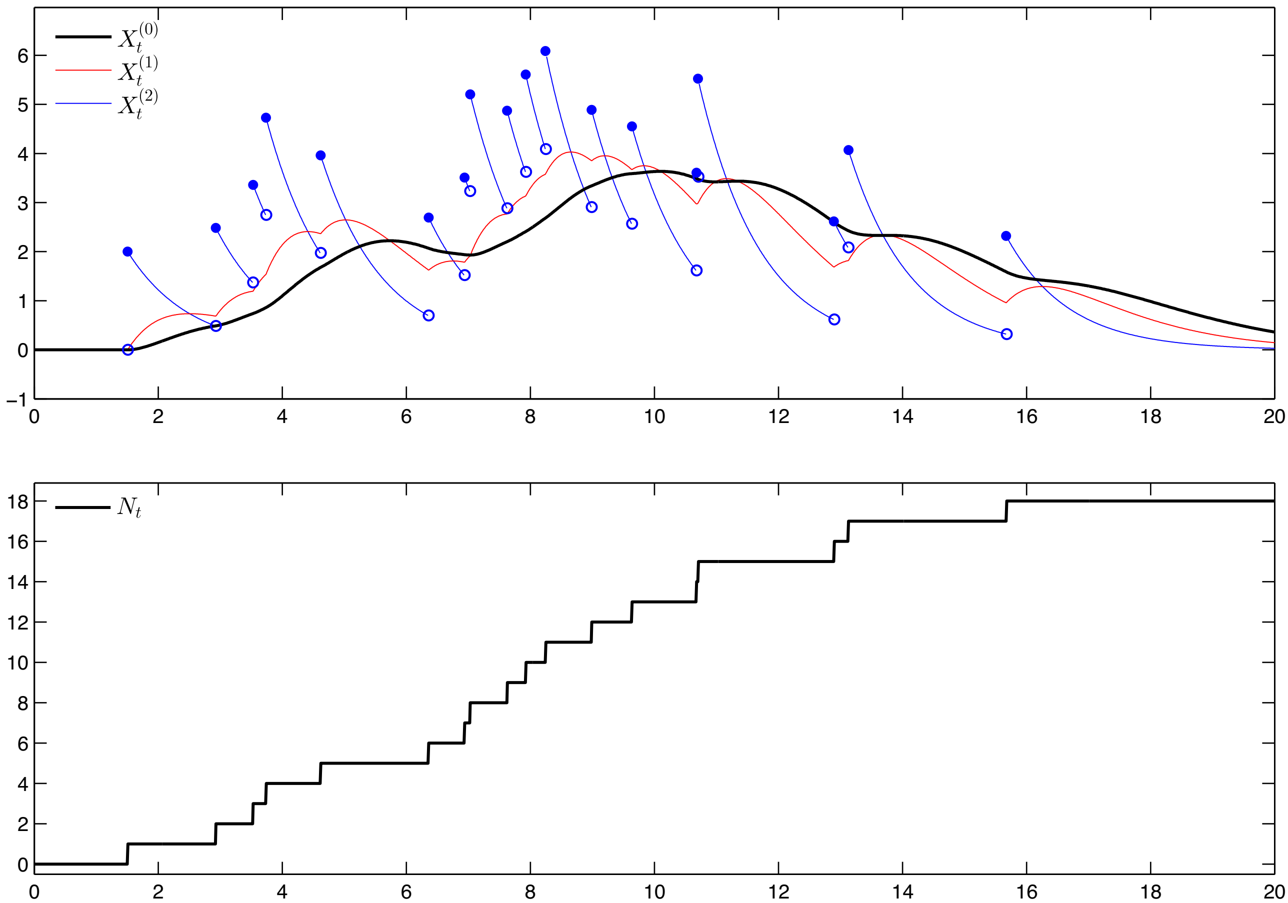

FIGURE 1. A finite joint realisation of the Markovian cascade $X = (X_t)_{0\le t\le T}$ (upper panel) and its associated counting process $N = (N_t)_{0\le t\le T}$ (lower panel) for the choices $L = 1$, $n_1 = 2$, $c_1 = 2$, $\alpha_1 = 1$, $T = 20$ and $f(x) = x/5 + 1$ with initial input $x_0 = (x_0^{(0)}, x_0^{(1)}, x_0^{(2)}) = (0, 0, 0)$. The blue (resp. red and black) trajectory corresponds to the realisation of $(X_t^{(2)})_{0\le t\le T}$ (resp. $(X_t^{(1)})_{0\le t\le T}$ and $(X_t^{(0)})_{0\le t\le T}$). Notice that the smaller the index $0 \le k \le n_1$ the smoother the correspondent process $(X^{(k)}_{0\le t\le T})$.

we have

$$\begin{aligned}
Lg(x) &\le \sum_{i=1}^{L}\Big(\sum_{k=0}^{n_i-1} sg(x^{(i,k)})\big(x^{(i,k+1)} - \alpha_i x^{(i,k)}\big) - \alpha_i|x^{(i,n_i)}| + |c_i| f\big(\sum_{j=1}^{L} x^{(j,0)}\big)\Big)\\
&\le cf(0) + (c\,\|f\|_{Lip} + \alpha)\sum_{i=1}^{L}|x^{(i,0)}| + (1+A)\sum_{i=1}^{L}\sum_{k=1}^{n}|x^{(i,k)}|\\
&\le Cg(x),
\end{aligned}$$

where $C = C(A, c, \|f\|_{Lip})$ and $sg(y)$ is the sign of $y \in \mathbb{R}$. In the second inequality above we have used that $f(y) \le \|f\|_{Lip}|y| + f(0)$ for any $y \in \mathbb{R}$. Thus, by applying Dynkin's formula and then using $Lg(x) \le Cg(x)$, one concludes that

$$E_{x_0}[g(X_t)] = g(x_0) + \int_0^t E_{x_0}[Lg(X_s)]\mathrm{d}s \leq g(x_0) + C\int_0^t E_{x_0}[g(X_s)]\mathrm{d}s.$$

Then by Gronwall's inequality, $E_{x_0}[g(X_t)] \leq g(x_0)e^{Ct}$. From this last estimate we conclude the proof noticing that

$$E_{x_0}(N([s,t])) = E_{x_0}\int_s^t f\big(\sum_{i=1}^L X_u^{(i,0)}\big)\mathrm{d}u \leq \|f\|_{Lip}\int_s^t \sum_{i=1}^L E_{x_0}|X_u^{(i,0)}|\mathrm{d}u + f(0)(t-s).$$

Since $\sum_{i=1}^L |X_u^{(i,0)}| \leq g(X_u)$, it follows immediately from the inequality above that

$$\begin{aligned} E_{x_0}(N([s,t])) &\leq \|f\|_{Lip}\int_s^t E_{x_0}[g(X_u)]\mathrm{d}u + f(0)(t-s) \\ &\leq g(x_0)\frac{\|f\|_{Lip}}{C}(e^{Ct} - e^{Cs}) + f(0)(t-s) < \infty. \end{aligned}$$

□

Remark 2.5. The converse statement of the above proposition does also hold true. More precisely, let $(N_t)_{t\geq 0}$ be a non-linear Hawkes process with intensity (2.1) where h is given by (2.2). Suppose moreover that N starts from the $n(\mathrm{d}s)$ on $\mathbb{R}_-$, where n is some discrete point measure on $\mathbb{R}_-$ such that $x_0^{(i,k)} := \int_{-\infty}^0 c_i e^{-\alpha_i(-s)}\frac{(-s)^{(n_i-k)}}{(n_i-k)!}n(\mathrm{d}s)$ are well-defined. Introduce the associated processes $X_t^{(i,k)}, 1 \leq i \leq L, 0 \leq k \leq n_i$ as in (2.3). Then $X = (X_t)_{t\geq 0}$ is Markov with generator (2.6).

2.1. Some comments on the use of Erlang kernels

Erlang kernels are widely used in the modelling literature. They have been introduced by Erlang in the 1920's to provide an efficient approach for analyzing telephone networks. Nowadays, they are widely used in the theoretical and mathematical biology literature, see *e.g.*, [10, 22] where they serve as a good model to describe delays in the hemodynamics in nephrons. They are also the building block to prove the existence of oscillations in large-scale limits of interacting neurons in a mean-field frame, see [9].

Notice also that the class of Erlang memory kernels is dense in $L^1(\mathbb{R}_+)$, see *e.g.*, [16]. Therefore, any Hawkes process N having general integrable memory kernel h can be approximated by a sequence of Hawkes processes $N^{(n)}$ having Erlang memory kernel $h^{(n)}$ such that $\|h^{(n)} - h\|_{L^1(\mathbb{R}_+)} \to 0$ as $n \to \infty$ and

$$E|||N - N^{(n)}|||_t \leq C_T\int_0^t |h^{(n)} - h|(s)\mathrm{d}s, \tag{2.9}$$

for all $t \leq T$ (see Lem. (3.4) of [21]), where $|||N - N^{(n)}|||_t$ denotes the total variation distance between N and $N^{(n)}$ on $[0,t]$.

Finally, the Markovian representation of Hawkes processes having memory kernels as in Assumption 2.3 in terms of the PDMP (2.5)–(2.6) has two advantages. The first advantage is that stability properties and the longtime behaviour of such Hawkes processes can be studied *via* the well-established theory of PDMPs. Since it is straightforward to simulate the PDMP (2.5)–(2.6) (see Sect. 4), one can also simulate Hawkes processes with memory kernels given by sum of Erlang kernels by using this representation. This is the second advantage.

In the next section we discuss stability properties of the associated PDMP (2.5)–(2.6) with random jump heights. A simulation algorithm for this PDMP will be presented in Section 4.

3. Long-time behaviour of the associated Markovian cascade with random jump heights

In this section we consider the Markov process $X = (X_t)_{t\geq 0}$ taking values in $\mathbb{R}^\kappa$ with (possibly) random jump heights. Its generator is given for any smooth and bounded function $g : \mathbb{R}^\kappa \mapsto \mathbb{R}$ by

$$\mathcal{L}g(x) = \langle F(x), \nabla g(x)\rangle + f\big(\sum_{i=1}^{L} x^{(i,0)}\big) \int \big(g(x + \sum_{i=1}^{L} c_i e_{(i,n_i)}) - g(x)\big) G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L), \tag{3.1}$$

where $F : \mathbb{R}^\kappa \mapsto \mathbb{R}^\kappa$ is the vector field associated to the system (2.7) and $G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L)$ is a probability measure on $\mathbb{R}^L$.

Assumption 3.1. The probability measure G on $\mathbb{R}^L$ has finite first moments, *i.e.*,

$$\int \sum_{i=1}^{L} |c_i| G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L) < \infty. \tag{3.2}$$

The above process is well-defined under Assumptions 2.2 and 3.1, as shows the following proposition.

Proposition 3.2. *Assume Assumptions 2.2 and 3.1. Let $N = (N_t)_{t\geq 0}$ be the counting process associated to the jumps of the Markov Process $X = (X_t)_{t\geq 0}$ having generator given by (3.1), starting from $x_0 \in \mathbb{R}^\kappa$. Then N has P_{x_0}-almost surely a finite number of jumps on each interval $[s, t], 0 \leq s < t < \infty$.*

The proof of this proposition is analogous to the proof of Proposition 2.4.

3.1. A Foster–Lyapunov condition

We start showing that there exists a compact set K of $\mathbb{R}^\kappa$ such that the process $X = (X_t)_{t\geq 0}$ possessing the generator defined in (3.1) visits K infinitely often almost surely. Let $n = \max_{1\leq i\leq L} n_i$ and $\alpha = \min_{1\leq i\leq L} \alpha_i$. In what follows, we write 0^κ to denote the vector in $\mathbb{R}^\kappa$ having all coordinates equal to 0.

Proposition 3.3. *Suppose Assumptions 2.2 and 3.1. Let $\mathcal{L}$ be the generator defined in (3.1) and consider the function $V : \mathbb{R}^\kappa \mapsto \mathbb{R}_+$ defined by*

$$V(x) = 1 + \sum_{i=1}^{L} \sum_{k=0}^{n_i} \frac{b(k+1)}{\alpha_i^k} |x^{(i,k)}|, \tag{3.3}$$

where $b : \{0, 1, \ldots n+1\} \to \mathbb{R}_+$ is a strictly increasing function. If f is not bounded but only Lipschitz continuous, we suppose moreover that

$$\|f\|_{Lip} \left(\int \sum_{i=1}^{L} \frac{1}{\alpha^{n_i}} |c_i| G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L) \right) < \alpha, \tag{3.4}$$

and choose the function b so that

$$\frac{b(n+1)}{b(1)} \|f\|_{Lip} \left(\int \sum_{i=1}^{L} \frac{1}{\alpha^{n_i}} |c_i| G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L) \right) < \alpha. \tag{3.5}$$

Then there exist positive constants λ, β and R such that the following Foster–Lyapunov type drift condition holds

$$\mathcal{L}V(x) \leq -\lambda V(x) + \beta 1_K(x), \tag{3.6}$$

where $K = \bar{B}_R(0^\kappa)$ is the (closed) ball of center 0^κ and radius R.

Remark 3.4. If $\alpha_i = \alpha$ for all $1 \leq i \leq L$ and $sg(c_i) = sg(c_j)$ (where $sg(u)$ is the sign of $u \in \mathbb{R}$) for all $i \neq j$, then condition (3.4) is equivalent to the sub-criticality condition (1.3) required in Theorem 1 of [4]. For values of $\alpha = \min_{1\leq i\leq L} \alpha_i \geq 1$, we could have taken the simpler Lyapunov function $V(x) = 1 + \sum_{i=1}^{L}\sum_{k=0}^{n} |x^{(i,k)}|$.

Proof. Indeed, one immediately verifies that

$$\mathcal{L}V(x) = A(x) + B(x),$$

where

$$A(x) = \sum_{i=1}^{L}\left(\sum_{k=0}^{n_i-1} \frac{b(k+1)}{\alpha_i^k}\big(x^{(i,k+1)} - \alpha_i x^{(i,k)}\big) sg\big(x^{(i,k)}\big) - \frac{b(n_i+1)}{\alpha_i^{n_i-1}}|x^{(i,n)}|\right),$$

and

$$B(x) = f\big(\sum_{i=1}^{L} x^{(i,0)}\big) \int \sum_{i=1}^{L} \frac{b(n_i+1)}{\alpha_i^{n_i}}\big(|x^{(i,n_i)} + c_i| - |x^{(i,n_i)}|\big) G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L).$$

Defining $b_* = \min\{b(k+1) - b(k), 0 \leq k \leq n\}$ and $r = \alpha b_*(b(n+1))^{-1}$, it is also straightforward to check that

$$\begin{aligned} A(x) &\leq -b(1)\sum_{i=1}^{L}\alpha_i|x^{(i,0)}| - \alpha\sum_{i=1}^{L}\sum_{k=1}^{n_i}\frac{1}{\alpha_i^k}\big(b(k+1) - b(k)\big)|x^{(i,k)}| \\ &\leq -\alpha\sum_{i=1}^{L} b(1)|x^{(i,0)}| - r\sum_{i=1}^{L}\sum_{k=1}^{n_i}\frac{b(k+1)}{\alpha_i^k}|x^{(i,k)}|. \end{aligned} \tag{3.7}$$

Suppose first that is f is bounded by f^*. In this case, one can easily verify that

$$B(x) \leq f^* b(n+1) \int \sum_{i=1}^{L} \alpha_i^{-n_i}|c_i| G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L).$$

Since $r < \alpha$, it follows from the above estimates that

$$\mathcal{L}V(x) \leq -rV(x) + p,$$

where $p = r + f^* b(n+1) \int \sum_{i=1}^{L} \alpha_i^{-n_i}|c_i| G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L)$.

Let $q = 1 + b(1)(1 \wedge \alpha^{-n})R$ and observe that $V(x) \geq q$, for $x \notin K = \bar{B}_R(0^\kappa)$. Thus, taking any R sufficiently large such that $p/q < r$, we deduce that

$$\mathcal{L}V(x) \leq -\Big(r - \frac{p}{q}\Big)V(x) + p 1_K(x),$$

which proves (3.6) for bounded jump rates f with $\lambda = r - \frac{p}{q} > 0$ and $\beta = p > 0$.

Assuming now that f is unbounded and Lipschitz continuous, we have that

$$B(x) \leq \Big(\|f\|_{Lip} \sum_{i=1}^{L} |x^{(i,0)}| + f(0)\Big) b(n+1) \int \sum_{i=1}^{L} \alpha_i^{-n_i} |c_i| G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L),$$

which, together with the first inequality in (3.7), implies that

$$\mathcal{L}V(x) \leq -\mathrm{d}V(x) + d + f(0) b(n+1) \int \sum_{i=1}^{L} \alpha_i^{-n_i} |c_i| G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L),$$

where $d = \Big(\alpha - \|f\|_{Lip} \frac{b(n+1)}{b(1)} \int \sum_i \alpha_i^{-n_i} |c_i| G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L)\Big) \wedge r$ is positive thanks to (3.4). Using the inequality above and proceeding as before, we establish also the drift condition (3.6) for Lipschitz jump rates. □

As a corollary of Proposition 3.3, we obtain exponential moments for the return times to the compact set K appearing in (3.6).

Corollary 3.5. *Let $K = \bar{B}_R(0^\kappa)$ and V be as in Proposition 3.3. Write $T_K = \inf\{t > 0 : X_t \in K\}$. Then for all $\eta \leq \lambda$ and $x_0 \in \mathbb{R}^\kappa$,*

$$E_{x_0}[e^{\eta T_K}] \leq V(x_0). \tag{3.8}$$

The proof of this corollary is classical, see for instance Theorem 6.1 of [11].

3.2. Wasserstein contraction for Lipschitz jump rates

Throughout this section we suppose that the jump rate f is Lipschitz continuous. In this case, we are able to prove the exponential convergence to equilibrium in Wasserstein distance, under the sub-criticality condition (3.4).

More precisely, in the sequel, for any $x \in \mathbb{R}^\kappa$, we will write $\|x\|_1 = \sum_{i=1}^{L} \sum_{k=0}^{n_i} |x^{(i,k)}|$. Let μ and ν be two probability measures on $\mathbb{R}^\kappa$. We call coupling of μ and ν any probability measure on $\mathbb{R}^\kappa \times \mathbb{R}^\kappa$ 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

$$W_1(\mu, \nu) = \inf \left\{ \int_{\mathbb{R}^\kappa} \int_{\mathbb{R}^\kappa} \|x - y\|_1 \gamma(\mathrm{d}x, \mathrm{d}y), \gamma \in \Gamma(\mu, \nu) \right\}. \tag{3.9}$$

In the following, we write $(P_t)_{t \geq 0}$ for the transition semigroup of the process X with generator (3.1). Recall that $A = \max_{1 \leq i \leq L} \alpha_i$, $\alpha = \min_{1 \leq i \leq L} \alpha_i$ and $n = \max_{1 \leq i \leq L} n_i$. The following theorem states the exponential rate of convergence to equilibrium of the process with respect to the Wasserstein distance.

Theorem 3.6. *Suppose f is Lipschitz continuous, assume condition (3.4) and choose the function b as in (3.5).*

1. Then, for any choice of probability measures μ and ν on $\mathcal{B}(\mathbb{R}^\kappa)$,

$$W_1(\mu P_t, \nu P_t) \leq C e^{-\mathrm{d}t} W_1(\mu, \nu), \tag{3.10}$$

where

$$C = \frac{A^n \vee 1}{1 \wedge \alpha^{n+1}} \frac{b(n+1)}{b(1)},$$

and $d = \left(\alpha - \|f\|_{Lip} \frac{b(n+1)}{b(1)} \int \sum_{i=1}^{L} \alpha_i^{-n_i} |c_i| G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L)\right) \wedge \left(\alpha b_* (b(n+1))^{-1}\right)$, *with* $b_* = \min\{b(k+1) - b(k), 0 \le k \le n\}$.

2. In particular, there exists a unique invariant probability measure π *of the process* X *such that for any probability measure* ν *on* $\mathcal{B}(\mathbb{R}^\kappa)$,

$$W_1(\pi, \nu P_t) \le C e^{-\mathrm{d}t} W_1(\pi, \nu).$$

Proof. The assertion of point 1. follows from a standard Wasserstein coupling. More precisely, denote by $(X, \tilde{X})$ the Markov processes taking values in $\mathbb{R}^\kappa \times \mathbb{R}^\kappa$ having the infinitesimal generator defined for any smooth test function $\varphi(x, y) : \mathbb{R}^\kappa \times \mathbb{R}^\kappa \to \mathbb{R}$ by

$$\begin{aligned}
\mathcal{L}_2 \varphi(x, y) =& \langle F(x), \nabla_x \varphi(x, y) \rangle + \langle F(y), \nabla_y \varphi(x, y) \rangle \\
&+ f\Big(\sum_i x^{(i,0)}\Big) \wedge f\Big(\sum_i y^{(i,0)}\Big) \int G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L) \left[\varphi(x + \sum_i c_i e_{(i,n_i)}, y + \sum_i c_i e_{(i,n_i)}) - \varphi(x, y)\right] \\
&+ \left(f\Big(\sum_i x^{(i,0)}\Big) - f\Big(\sum_i y^{(i,0)}\Big)\right)_+ \int G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L) \left[\varphi(x + \sum_i c_i e_{(i,n_i)}, y) - \varphi(x, y)\right] \\
&+ \left(f\Big(\sum_i y^{(i,0)}\Big) - f\Big(\sum_i x^{(i,0)}\Big)\right)_+ \int G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L) \left[\varphi(x, y + \sum_i c_i e_{(i,n_i)}) - \varphi(x, y)\right], \quad (3.11)
\end{aligned}$$

where $F : \mathbb{R}^\kappa \mapsto \mathbb{R}^\kappa$ is the vector field associated to the system (2.7).

This is the usual coupling which consists of making the two processes jump together as much as possible. Define

$$H(x, y) = \sum_{i=1}^{L} \sum_{k=0}^{n_i} \frac{b(k+1)}{\alpha_i^k} |x^{(i,k)} - y^{(i,k)}|.$$

Then an analogous calculus as the one used in the proof of Proposition 3.3 yields

$$\mathcal{L}_2 H(x, y) \le -\mathrm{d}H(x, y),$$

implying that

$$E_{x,y} H(X_t, \tilde{X}_t) \le H(x, y) e^{-\mathrm{d}t}.$$

Observing that

$$\|x - y\|_1 \le \frac{A^n \vee 1}{b(1)} H(x, y); \; H(x, y) \le \frac{b(n+1)}{1 \wedge \alpha^{n+1}} \|x - y\|_1,$$

we conclude the proof of item 1.

To prove item 2., let $\mu_n := \mu P_n$ for any probability measure μ on $\mathcal{B}(\mathbb{R}^\kappa)$. Observe that $W_1(\mu_{n+m}, \mu_m) \le C e^{-\mathrm{d}m} W_1(\mu_n, \mu)$, implying that $(\mu_n)_n$ is Cauchy and thus, by the completeness of the space of all probability measures on $(\mathbb{R}^\kappa, \mathcal{B}(\mathbb{R}^\kappa))$, endowed with the metric induced by W_1 (see *e.g.*, [20] or [3]), convergent to some limit measure μ_∞. This limit measure must be invariant. Indeed we have $W_1(\mu_m P_t, \mu_m) \le C e^{-\mathrm{d}m} W_1(\mu P_t, \mu)$.

But

$$|W_1(\mu_m P_t, \mu_m) - W_1(\mu_\infty P_t, \mu_\infty)| \le W_1(\mu_m P_t, \mu_\infty P_t) + W_1(\mu_\infty, \mu_m)$$
$$\le [Ce^{-\mathrm{d}t} + 1] W_1(\mu_\infty, \mu_m) \to 0,$$

as $m \to \infty$, where we have used (3.10) once again to obtain the second inequality. As a consequence, $W_1(\mu_\infty P_t, \mu_\infty) = 0$, implying that μ_∞ is the (necessarily unique) invariant measure. This concludes our proof. □

Remark 3.7. Under the conditions of Theorem 3.6, write $\bar{N}_t$ for the stationary version of the non-linear Hawkes process having intensity (2.1); that is, $\bar{N}_t$ has intensity $\bar{\lambda}_t = f\big(\sum_{i=1}^L \bar{X}_t^{(i,0)}\big)$, where $\bar{X}$ is the stationary process evolving according to (2.6). Let moreover N_t be the non-linear Hawkes process with intensity $\lambda_t = f\big(\sum_{i=1}^L X_t^{(i,0)}\big)$ starting from some fixed initial condition $X_0 = x_0 \in \mathbb{R}^\kappa$. Then the Lipschitz-continuity of f together with (3.10) imply that $E\int_0^\infty |\lambda_t - \bar{\lambda}_t| \mathrm{d}t < \infty$. It is then straightforward to deduce from this by standard coupling arguments, as explained *e.g.*, the proof of Theorem 1 in [4], that N and $\bar{N}$ couple almost surely in finite time; that is, there exists $T > 0$ such that for all $t \ge 0$, $N_{T+t} - N_T = \bar{N}_{T+t} - \bar{N}_t$, meaning that N and $\bar{N}$ have the same jump times after time T. This is what is called *stability in variation* in [4] (see their Def. 1). Therefore, our Theorem 3.6 implies Theorem 1 of [4].

In the next section we prove a stronger result, showing that the process is even recurrent in the sense of Harris.

3.3. Harris recurrence

In this section, we use the regeneration method based on Nummelin splitting to show that X is recurrent in the sense of Harris having a unique invariant probability measure π. We recall (*e.g.*, from [1]) that

Definition 3.8. The process $(X_t)_{t\ge 0}$ is said to be *recurrent in the sense of Harris* if there exists a sigma-finite measure m on $\mathcal{B}(\mathbb{R}^\kappa)$ such that $m(A) > 0$ implies that for all $x \in \mathbb{R}^\kappa$, P_x−almost surely,

$$\limsup_{t\to\infty} 1_A(X_t) = 1.$$

By [1], Harris recurrence of X implies in particular the existence of a unique invariant measure (which is sigma-finite but does not need to be finite) π such that the above property holds with π in place of m. X is called positive Harris recurrent if $\pi(\mathbb{R}^\kappa) < \infty$. We have the following

Theorem 3.9. *Suppose that f is bounded or Lipschitz continuous satisfying* (3.4). *Suppose moreover that Assumption 3.1 holds and that $G(\mathrm{d}c_1, \ldots, \mathrm{d}c_L) = \prod_{i=1}^L G_i(\mathrm{d}c_i)$ for probability measures G_i on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfying supp $(G_i) \cap \{0\}^c \neq \emptyset$, for all $1 \le i \le L$. Finally, suppose that f is lower bounded.*

1. Then $(X_t)_{t\ge 0}$ is positive Harris recurrent with unique invariant measure $\pi(\mathrm{d}x)$.

2. Let $\bar{X}_t$ be a stationary version of the process and suppose that $(X_t)_{t\ge 0}$ starts from $X_0 = x_0 \in \mathbb{R}^\kappa$, both evolving according to (2.6). *Then $\bar{X}$ and X couple almost surely in finite time; that is, it is possible to construct them on the same probability space such that there exists $\tau_c < \infty$ almost surely satisfying*

$$t \ge \tau_c \text{ implies that } X_t = \bar{X}_t \text{ and } P(\tau_c > t) \le C(p)V(x)t^{-p}, \tag{3.12}$$

for every $p \ge 1$, where $C(p)$ is a constant depending on p.

Remark 3.10. In particular, using the notation of Remark 3.7 above, (3.12) implies that $\lambda_t = \bar{\lambda}_t$ for all $t \ge \tau_c$ meaning that $\bar{N}$ and N couple as well. As a consequence, our Theorem 3.9 is a refinement of the results of Theorems 1 and 2 in [4] – however at the prize of imposing a lower bound on f.

The proof of this theorem uses the regeneration technique based on Nummelin splitting. It is well known that it is easier to implement this method in the frame of discrete time Markov chains rather then Markov processes in continuous time – although some effort has been spent to introduce regeneration times in a continuous time framework, see *e.g.*, [18]. Therefore, we start by showing that the sampled chain $(Y_k)_{k\geq 0} = (X_{kT})_{k\geq 0}$, for some fixed $T > 0$, is positive Harris recurrent.

We recall that the chain $(Y_k)_{k\geq 0}$ is said to be *recurrent in the sense of Harris* with invariant measure π on $\mathcal{B}(\mathbb{R}^\kappa)$ if whenever $\pi(A) > 0$, we have, for all $x \in \mathbb{R}^\kappa$, P_x–almost surely, $\limsup_{k\to\infty} 1_A(Y_k) = 1$. Obviously, Harris recurrence of the chain $(Y_k)_{k\geq 0}$ implies the Harris recurrence of the process X, and the invariant probability measures of both processes coincide (if they exist).

The rest of this section is devoted to prove that the sampled chain $(Y_k)_{k\geq 0}$ is Harris which follows from the following Doeblin type lower bound. Recall that $(P_t)_{t\geq 0}$ denotes the transition semigroup of the process X, therefore, P_T is the transition operator of the sampled chain $(Y_k)_{k\geq 0}$.

Theorem 3.11. *Assume the assumptions of Theorem 3.9. For all $x^* \in \mathbb{R}^\kappa$, there exist $R > 0$, an open set $I \subset \mathbb{R}^\kappa$ and a constant $\beta \in (0, 1)$, depending on $I, R, L, n_1, \ldots, n_L, \alpha_1, \ldots, \alpha_L$ and f, such that*

$$P_{LT}(x, \mathrm{d}y) \geq \beta 1_C(x)\nu(\mathrm{d}y), \tag{3.13}$$

where $C = B_R(x^*)$ is the (open) ball of radius R centered at x^*, and where ν is the uniform probability measure on I.

Proof. **Part I.** $L = 1$.
We start by proving the result in the case $L = 1, c_1 = c, \alpha_1 = \alpha$ and $n_1 = n$, that is, $h(t) = ce^{-\alpha t}\frac{t^n}{n!}$. The corresponding Markov process is then given by $X_t = (X_t^{(0)}, \ldots, X_t^{(n)})$ taking values in $\mathbb{R}^{n+1}$. Clearly, for all $A \in \mathcal{B}(\mathbb{R}^{n+1})$,

$$P_T(x, A) \geq E_x(1_A(X_T), N_T = n + 1).$$

Recall the definition of the flow in (2.8). On the event $\{N_T = n + 1\}$, starting from $X_0 = x$, we first let the flow evolve starting from x up to some first jump time t_1. At that jump time we choose an associated jump height c_1. We then successively choose the following inter-jump waiting times $t_2, \ldots, t_{n+1}$ under the constraint $t_1 + \ldots + t_{n+1} < T$ and the associated jump heights $c_2, \ldots, c_{n+1}$. Write $s_1 = T - t_1, s_2 = T - (t_1 + t_2), \ldots, s_{n+1} = T - (t_1 + \ldots + t_{n+1})$.

Conditionally on $X_0 = x$, the successive choices of $\underline{\mathbf{c}} = (c_1, \ldots, c_{n+1})$ and $\underline{\mathbf{s}} = (s_1, \ldots, s_{n+1})$ as above, the position of X_T is given by

$$\gamma(x, \underline{\mathbf{c}}, \underline{\mathbf{s}}) = \varphi_T(x) + c_1 e^{-\alpha s_1} v(s_1) + \ldots + c_{n+1} e^{-\alpha s_{n+1}} v(s_{n+1}), \tag{3.14}$$

where for each $1 \leq k \leq n + 1$,

$$v(s_k) = \begin{pmatrix} \frac{s_k^n}{n!} \\ \frac{s_k^{n-1}}{(n-1)!} \\ \vdots \\ s_k \\ 1 \end{pmatrix}. \tag{3.15}$$

We omitted the dependence on T of the map $\gamma(x, \underline{\mathbf{c}}, \underline{\mathbf{s}})$ since we keep the value $T > 0$ fixed once for all and work with sequences $\underline{\mathbf{s}}$ satisfying the constraints $0 < s_{n+1} < s_n < \ldots < s_1 < T$. Finally, in what follows we shall write, for any fixed pair $(x, \underline{\mathbf{c}})$,

$$\gamma_{(x,\underline{\mathbf{c}})} : \underline{\mathbf{s}} \mapsto \gamma(x, \underline{\mathbf{c}}, \underline{\mathbf{s}}).$$

We will use the jump noise which is created by the $n+1$ jumps, *i.e.*, we will use a change of variables on the account of $s_1, \ldots, s_{n+1}$. Therefore, in what follows we write

$$\frac{\partial \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})}{\partial \underline{\mathbf{s}}} = \Big[\frac{\partial \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})}{\partial s_1}, \ldots, \frac{\partial \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})}{\partial s_{n+1}}\Big]$$

to denote the Jacobian matrix of the the map $\underline{\mathbf{s}} \mapsto \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})$. This matrix does not depend on the initial position x nor on the first jump height c_0. Indeed, one easily finds that

$$\frac{\partial \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})}{\partial \underline{\mathbf{s}}} = \Big[C^{(1)}, \ldots, C^{(n+1)}\Big],$$

where for each $1 \le k \le n+1$, $C^{(k)}$ is a column vector given by

$$C^{(k)} = c_k e^{-\alpha s_k} \begin{pmatrix} \frac{s_k^{n-1}}{(n-1)!} - \alpha \frac{s_k^n}{n!} \\ \frac{s_k^{n-2}}{(n-2)!} - \alpha \frac{s_k^{n-1}}{(n-1)!} \\ \vdots \\ 1 - \alpha s_k \\ -\alpha \end{pmatrix}.$$

As a consequence the determinant of $\frac{\partial \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})}{\partial \underline{\mathbf{s}}}$ is given by

$$\det\Big(C^{(1)}, \ldots, C^{(n+1)}\Big) = (-1)^{n+1} \alpha \prod_{k=1}^{n+1} c_k e^{-\alpha s_k} \det\Big(\tilde{C}^{(1)}, \ldots, \tilde{C}^{(n+1)}\Big), \tag{3.16}$$

where for each $1 \le k \le n+1$, $\tilde{C}^{(k)}$ is a column vector given by

$$p\tilde{C}^{(k)} = \begin{pmatrix} \alpha \frac{s_k^n}{n!} - \frac{s_k^{n-1}}{(n-1)!} \\ \alpha \frac{s_k^{n-1}}{(n-1)!} - \frac{s_k^{n-2}}{(n-2)!} \\ \vdots \\ \alpha s_k - 1 \\ 1 \end{pmatrix}.$$

Thus the invertibility of the matrix $\frac{\partial \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})}{\partial \underline{\mathbf{s}}}$ follows from the invertibility of the matrix $J = [\tilde{C}^{(1)}, \ldots, \tilde{C}^{(n+1)}]$. In the sequel, for each $1 \le k \le n+1$, let r_k denote the k th row of J. By replacing successively (bottom-up) r_i

by $\alpha^{-1}(r_i + r_{i-1})(n+1-i)!$, we deduce that J is equivalent to the Vandermonde matrix

$$\begin{pmatrix} s_1^n & s_2^n & \dots & s_{n+1}^n \\ s_1^{n-1} & s_2^{n-1} & \dots & s_{n+1}^{n-1} \\ \vdots & & & \\ s_1 & s_2 & \dots & s_{n+1} \\ 1 & 1 & \dots & 1 \end{pmatrix},$$

which is know to be invertible if and only if $0 < s_{n+1} < s_n < \ldots < s_1$. In conclusion, we have just shown that for any $x \in \mathbb{R}^{n+1}$, any choice of $\underline{\mathbf{c}}$ having non null coordinates, the Jacobian of the map $\underline{s} \mapsto \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})$ is invertible at any $\underline{\mathbf{s}}$ such that $0 < s_{n+1} < s_n < \ldots < s_1$.

It will be proved now that this uniform invertibility of the Jacobian matrix of the map $\underline{\mathbf{s}} \mapsto \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})$ implies inequality (3.13). For that sake, we shall also need the following notation. For each triple $(x, \underline{\mathbf{c}}, \underline{\mathbf{s}})$, we write $x_0 = x$, $x_1 = \varphi_{T-s_1}(x) + c_1 e_{n+1}$[2] and then recursively $x_k = \varphi_{s_{k-1}-s_k}(x_{k-1}) + c_k e_{n+1}$ for $2 \le k \le n+1$. The sequence $x_1, \ldots x_{n+1}$ corresponds to the positions right after successive jumps, starting from the initial location $x \in \mathbb{R}^{n+1}$, induced by the heights $\underline{\mathbf{c}}$ and the inter-jump waiting times $T - s_1, s_1 - s_2, \ldots s_n - s_{n+1}$ which are determined by $\underline{\mathbf{s}}$.

Introduce now for each $x \in \mathbb{R}^{n+1}$ and $t \ge 0$,

$$e(x,t) = \exp\Big\{ - \int_0^t f\big(\varphi_s^{(0)}(x)\big)\mathrm{d}s \Big\} \tag{3.17}$$

and define for each triple $(x, \underline{\mathbf{c}}, \underline{\mathbf{s}})$ (here we set $s_0 = T$),

$$q_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}}) = \left(\prod_{k=0}^{n} f(\varphi^{(0)}_{s_k - s_{k+1}}(x_k)) e(x_k, s_k - s_{k+1}) \right) e(x_{n+1}, s_{n+1}). \tag{3.18}$$

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

$$\inf_{(x,\underline{\mathbf{c}},\underline{\mathbf{s}}) \in U_{x^*} \times V_{\underline{\mathbf{c}}^*} \times W^*_{\underline{\mathbf{s}}}} q_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}}) > 0. \tag{3.19}$$

Let us now fix a triple $(x^*, \underline{\mathbf{c}}^*, \underline{\mathbf{s}}^*)$ such that the matrix $\frac{\partial \gamma_{x^*,\underline{\mathbf{c}}^*}(\underline{\mathbf{s}}^*)}{\partial \underline{\mathbf{s}}}$ is invertible. Recall that by (3.16), the vector $\underline{\mathbf{c}}^*$ must have all coordinates non-null. By Lemma 6.2 of [2], there exist an open neighborhood $J_{x^*,\underline{\mathbf{c}}^*} = B_R(x^*) \times B_R(\underline{\mathbf{c}}^*)$ of the pair $(x^*, \underline{\mathbf{c}}^*)$, an open set $I \subset \mathbb{R}^{n+1}$, and for any pair $(x, \underline{\mathbf{c}}) \in J_{x^*,\underline{\mathbf{c}}^*}$, an open set $W_{x,\underline{\mathbf{c}}}$ such that

$$\tilde{\gamma}_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}}) : \begin{cases} W_{x,\underline{\mathbf{c}}} \to I \\ \underline{\mathbf{s}} \mapsto \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}}), \end{cases}$$

is a diffeomorphism, with $W_{x,\underline{\mathbf{c}}} \subset W^*_{\underline{\mathbf{s}}}$ and also

$$\inf_{x,\underline{\mathbf{c}} \in J_{x^*,\underline{\mathbf{c}}^*}} \inf_{\underline{s} \in W_{x,\underline{\mathbf{c}}}} \left| \det\Big(\frac{\partial \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})}{\partial \underline{\mathbf{s}}} \Big)^{-1} \right| > 0. \tag{3.20}$$

[2] e_{n+1} denotes the $n+1$−st unit vector in $\mathbb{R}^{n+1}$

Reducing (if necessary) R, we may assume also that $J_{x^*,\underline{\mathbf{c}}^*} \subset U_{x^*} \times V_{\underline{\mathbf{c}}^*}$. Thus we have that by (3.19) and (3.20),

$$\inf_{x,\underline{\mathbf{c}} \in J_{x^*,\underline{\mathbf{c}}^*}} \inf_{\underline{s} \in W_{x,\underline{\mathbf{c}}}} q_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}}) \left| \det\left(\frac{\partial \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})}{\partial \underline{\mathbf{s}}}\right)^{-1} \right| > 0. \tag{3.21}$$

Since $supp(G) \cap \{0\}^c \neq \emptyset$ there exists an interval (a,b) such that $0 \notin (a,b)$ and $G((a,b)) > 0$. Thus, by taking $\underline{\mathbf{c}}^* = ((a+b)/2, \ldots, (a+b)/2)$, we have (reducing R again if necessary) that for $1 \leq k \leq n$, $G((\underline{c}^{*(k)} - R, \underline{c}^{*(k)} + R)) > 0$ which together with (3.21) implies

$$\tilde{\beta} = \left(\prod_{k=1}^{n} G((\underline{c}^{*(k)} - R, \underline{c}^{*(k)} + R)) \right) \inf_{x,\underline{\mathbf{c}} \in J_{x^*,\underline{\mathbf{c}}^*}} \inf_{\underline{s} \in W_{x,\underline{\mathbf{c}}}} q_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}}) \left| \det\left(\frac{\partial \gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})}{\partial \underline{\mathbf{s}}}\right)^{-1} \right| > 0. \tag{3.22}$$

Finally, we have for any measurable $A \in \mathbb{B}(\mathbb{R}^{n+1})$ and $x \in B_R(x^*)$, using the change of variables $y = \tilde{\gamma}_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})$,

$$\begin{aligned} E_x(1_A(X_T, N_T = n+1) &\geq \int_{B_R(\underline{\mathbf{c}}^*)} \underline{G}(\mathrm{d}\underline{\mathbf{c}}) \int_{W_{x,\underline{\mathbf{c}}}} q_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}}) 1_A(\gamma_{x,\underline{\mathbf{c}}}(\underline{\mathbf{s}})) \mathrm{d}s_1 \ldots \mathrm{d}s_{n+1} \\ &\geq \tilde{\beta} \int_{I \cap A} \mathrm{d}y_1 \ldots \mathrm{d}y_{n+1} = \beta \nu(A), \end{aligned} \tag{3.23}$$

where $\underline{G}(\mathrm{d}\underline{\mathbf{c}}) = G(\mathrm{d}c_1) \ldots G(\mathrm{d}c_{n+1})$ and $\beta = \tilde{\beta}\nu(I)$, establishing the desired result in case $L = 1$.

The proof of the general case $L > 1$ follows the same strategy and is given in the Appendix. $\square$

We are now able to conclude the proof of Theorem 3.9.

Proof of Theorem 3.9.
(1) By Corollary 3.5, we know that X comes back to K infinitely often almost surely. Moreover, $\sup_{x \in K} |\varphi_t(x)| \to 0$ as $t \to \infty$, by the explicit form of the flow in (2.8). Therefore, for any $\varepsilon > 0$ there exists t^* such that $\varphi_t(x) \in B_\varepsilon(0)$ for all $t \geq t^*$, for all $x \in K$. Since f is bounded on $\tilde{K} := \{\varphi_t(x) : t \geq 0, x \in K\}$, we have $\inf_{x \in K} P_x(T_1 > t_* + 2T) > 0$. This implies that

$$\inf_{x \in K} P(\text{ the trajectory of } X \in B_\varepsilon(0) \text{ during a time interval of length} > T \,|X_0 = x) > 0,$$

and therefore, using a conditional version of the Borel–Cantelli lemma, $(Y_k)_{k \in \mathbb{N}}$ visits $B_\varepsilon(0)$ infinitely often almost surely.
(2) Applying the result of Theorem 3.11 with $x^* = 0$ and $\varepsilon = R$ and using the standard regeneration technique allows to conclude that $(Y_{Lk})_k$ and therefore $(X_t)_t$ are Harris recurrent. This implies item 1. of the theorem.
(3) To prove item 2., it is straightforward to show that Proposition 3.3 implies the existence a coupling of X_t and $\tilde{X}_t$, both evolving according to (2.6), such that for $T_{K \times K} := \inf\{t \geq 0 : (X_t, \tilde{X}) \in K \times K\}$, we have

$$E_{x,y}(e^{\eta T_{K \times K}}) \leq V(x) + V(y).$$

Indeed, if suffices to define the $2\kappa-$dimensional Lyapunov function $\bar{V}(x,y) := V(x) + V(y)$ and to check that (3.6) holds for $\bar{\mathcal{L}}$ where $\bar{\mathcal{L}}$ denotes the generator of the process $(X_t, \tilde{X}_t)$. Moreover, (3.13) can be immediately extended to a lower bound for the joint transition kernel of $(X_t, \tilde{X}_t)$, whenever both of them start within the set $C = B_R(0)$. Thus X and $\tilde{X}$ couple at least with probability β, each time they are within C at the same time. The proof that this coupling time has polynomial moments of any order follows then the same arguments as those given in the proof of Proposition 2.15 in [17], implying that

$$E_{x,y}[\tau_c^p] \leq C(p)[V(x) + V(y)]. \tag{3.24}$$

Finally, Theorem 4.3 of [19] implies that $\int V \mathrm{d}\pi < \infty$ such that we are able to integrate (3.24) against $\pi(\mathrm{d}x)$ in order to replace X_t by the invariant process $\bar{X}_t$ starting from $\bar{X}_0 \sim \pi$. This concludes the proof. $\square$

4. Simulation Algorithm

As a consequence of Proposition 2.4 it follows that any Hawkes process possessing memory kernels given by the sum of Erlang kernels can be represented as the counting process associated to the jumps of its Markovian cascade. Based on this Markovian representation we propose an algorithm (hereafter Algorithm 1) for simulating such Hawkes processes.

In what follows, for any $x \in \mathbb{R}^\kappa$, we shall write $\|x\|_\infty = \max\{|x^{(i,k)}|, 1 \le i \le L, 0 \le k \le n_i\}$. For a practical implementation of our algorithm the remark below will be important. Recall that $n = \max_{1\le i\le L} n_i$ and $\alpha = \min_{1\le i \le L} \alpha_i$.

Lemma 4.1. *For each $x \in \mathbb{R}^\kappa$, let $M(x) = \max\{|\varphi_t^{(i,0)}(x)| : 1 \le i \le L, t \ge 0\}$ where $\varphi_t^{(i,0)}(x)$ is defined in (2.8). Then*

$$M(x) \le e\|x\|_\infty \Big(1 \vee \Big(\frac{n}{\alpha e}\Big)^n\Big). \tag{4.1}$$

Proof. Indeed, it follows from (2.8) that for each $x \in \mathbb{R}^{n+1}$ and $t \ge 0$,

$$|\varphi_t^{(0)}(x)| \le \|x\|_\infty e^{-\alpha t}(1 + t + \ldots + t^n/n!) \le e\|x\|_\infty e^{-\alpha t}(1 \vee t^n) \le e\|x\|_\infty e^{-(\alpha-1)t},$$

so that if $\alpha > 1$, then clearly (4.1) holds. Now, assume $0 < \alpha \le 1$. Under this assumption, from standard calculus arguments we deduce that $\arg\max\{e^{-\alpha t}(1 \vee t^n) : t \ge 1\} = n/\alpha$. This fact and the second inequality above imply the bound in (4.1) as stated. $\square$

In the sequel, for any rate function f satisfying Assumption 2.2 we define the function $\mathbb{R}^\kappa \ni x \mapsto f^*(x)$ by

$$f^*(x) = \left\{ \begin{array}{ll} \max\{f(y) : y \in [0, LM(x)]\}, & \text{if } x \in \mathbb{R}^\kappa_+ \\ \max\{f(y) : y \in [-LM(x), 0]\}, & \text{if } x \in \mathbb{R}^\kappa_- \\ \max\{f(y) : y \in [-LM(x), LM(x)]\}, & \text{else} \end{array} \right\}.$$

Here, L is the number of terms in the sum defining the memory kernel h (recall Assumption 2.3). It follows immediately from Lemma 4.1 that the function f^* is well-defined, that is $f^*(x)$ is finite for all $x \in \mathbb{R}^\kappa$. Let $T_0 = 0$ and $(T_k)_{k\ge1}$ denote the sequence of jump times of the Markovian cascade X having generator (2.6). Observe that the non-explosiveness of X (thanks to Proposition 2.4) ensures that the sequence $(T_k)_{k\ge1}$ is well-defined. Suppose that $X_{T_k} = x$ is given for some $k \in \mathbb{N}$. Algorithm 1 works as follows. Draw an exponential random variable τ with parameter $f^*(x)$ and a uniform random variable U on $[0,1]$. If $U \le f(\sum_{i=1}^L \varphi_{T_k+\tau}^{(i,0)}(x))/f^*(x)$, then define the next jump time $T_{k+1} = T_k + \tau$. If not, repeat this procedure starting from $X_{T_k+\tau} = \varphi_\tau(x)$. Notice that Algorithm 1 is an extension (to our framework) of the classical thinning algorithm for simulating non-homogeneous Poisson processes. Moreover, it provides an exact simulation of the Markovian cascade X (and consequently of the associated Hawkes process) in the sense that no approximation procedure is required. Its formal definition is given below as a pseudo-code.

Proposition 2.4 and Lemma 4.1 ensure that Algorithm 1 is well-defined and works properly. More precisely, we have the following result.

Proposition 4.2. *Assume Assumption 2.2. For any choice of $T > 0$, $x_0 \in \mathbb{R}^\kappa$, $L \ge 1$ and $n_i \in \mathbb{N}$, $c_i \in \mathbb{R}$, $\alpha_i > 0$ for $1 \le i \le L$, Algorithm 1 terminates almost surely within finite time. If additionally x_0 is given as in Theorem 2.4, the output of Algorithm 1 follows the distribution of a Hawkes process with intensity (2.1).*

Algorithm 1. Simulation algorithm for the Markovian cascade X.

Input: bounded or Lipschitz continuous f, constants $\alpha_i > 0$, $c_i \in \mathbb{R}$ and $T > 0$; and a vector of initial conditions $\big(X_0^{(i,k)}, 1 \le i \le L, 0 \le k \le n_i\big) = \big(x_0^{(i,k)}, 1 \le i \le L, 0 \le k \le n_i\big) \in \mathbb{R}^\kappa$.
Output: The counting process $(N_t)_{t\in[0,T]}$.
Initial values: $x \leftarrow \big(x_0^{(i,k)}, 1 \le i \le L, 0 \le k \le n_i\big)$, $D \leftarrow 0$ and $N_0 \leftarrow 0$.
while $D < T$ **do**
  $f^* \leftarrow f^*(x)$
  draw $\tau \sim \mathcal{E}(f^*)$
  if $\tau \le T - D$ **then**
    draw $U \sim U[0,1]$
    if $U \le f(\sum_i \varphi_{D+\tau}^{(i,0)}(x))/f^*$ **then**
      $x \leftarrow \varphi_\tau(x) + \sum_{i=1}^L c_i e_{(i,n_i)}$
      $N_t \leftarrow N_D$, for $D \le t < D + \tau$
      $N_{D+\tau} \leftarrow N_D + 1$
    else
      $x \leftarrow \varphi_\tau(x)$
      $N_t \leftarrow N_D$, for $D \le t \le D + \tau$
    end if
  else
    $N_t \leftarrow N_D$, for $D \le t \le T$
  end if
  $D \leftarrow D + \tau$
end while
Return $(N_t)_{t\in[0,T]}$.

It is worth noting that Algorithm 1 does not require the sub-criticality condition (1.3) for non-linear Hawkes processes. Indeed, Algorithm 1 applies for instance for the choice $L = 1$, $n_1 = 0$, $c_1 = \alpha_1$ and $f(x) = (\mu + x)1_{[0,\infty)}(x)$ with $\mu \ge 0$ for which (1.3) does not hold. The only restriction we have to impose is to work with memory kernels which are sum of Erlang kernels, which is a generalization of the approach proposed in [7]. In the next section some numerical examples are presented both for bounded and unbounded Lipschitz jump rates f.

5. Numerical examples

In this section four numerical examples are given. Specifically, we generate first a sample of the Markovian cascade with $L = 1$, for a time window $T = 100$, order delay $n_1 = 3$, jump height $c_1 = 1$, decay rate $\alpha_1 = 1$ and jump rate $f(x) = (\mu + x)1_{[0,\infty)}(x)$ with $\mu = 1$. We also simulate a Markovian cascade with random jump heights c_1 following a Normal distribution $\mathcal{N}(0, 100)$ and $f(x) = (1 + (x/2)^{3/2}) \wedge 30$, keeping all others parameters as in the preceding example. The extension of Algorithm 1 for random jump heights is straightforward. Next, we simulate jointly three Markovian cascades with $L = 1$ possessing rates of decay $\alpha_1 = 0.8, \alpha_1 = 1$ and $\alpha_1 = 1.4$ respectively; in this example $T = 500, n_1 = 3$, the jump heights follow a Normal distribution $\mathcal{N}(0, 100)$ and $f(x) = 1 + \sigma/(1 + e^{-\beta(x-\rho)})$ where $\sigma = 20, \beta = 1/3$ and $\rho = 10$. Finally, we simulate a Markovian cascade for the choices $L = 3$, $n_1 = 1$, $n_2 = 3$, $n_3 = 2$, $\alpha_1 = 1.3$, $\alpha_2 = 0.8$, $\alpha_3 = 1$, $T = 30$, $f(x) = (2 + \exp(x/10)) \wedge 20$, random jump heights $c_1 = c_2 = c_3$ following a Normal distribution $\mathcal{N}(0, 25)$.

The results are presented in Figures 2, 3, 4 and 6 respectively. In order to test if Algorithm 1 works properly, we use the following result.

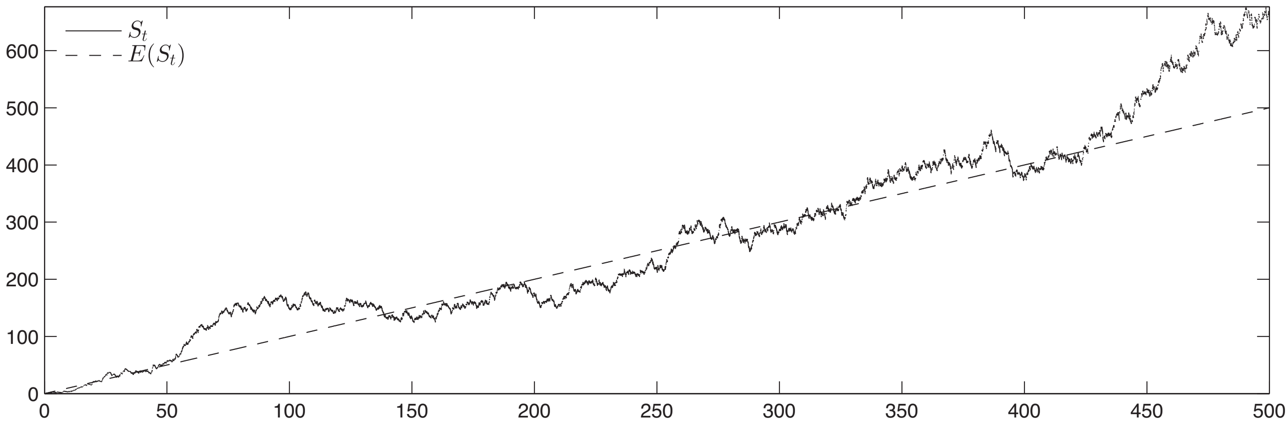

FIGURE 2. A realization of the process $S = (S_t)_{0\le t\le T}$ where $S_t = \sum_{k=0}^{n_1} X_t^{(k)}$ for the choices $L = 1$, $n_1 = 3$, $c_1 = \alpha_1 = 1$, $T = 100$ and $f(x) = (1 + x/5)1_{[0,\infty)}(x)$ with initial configuration $S_0 = 0$. The dashed line corresponds to the theoretical mean $E[S_t] = t$ of S_t, conditionally on $S_0 = 0$, obtained by Proposition 5.1.

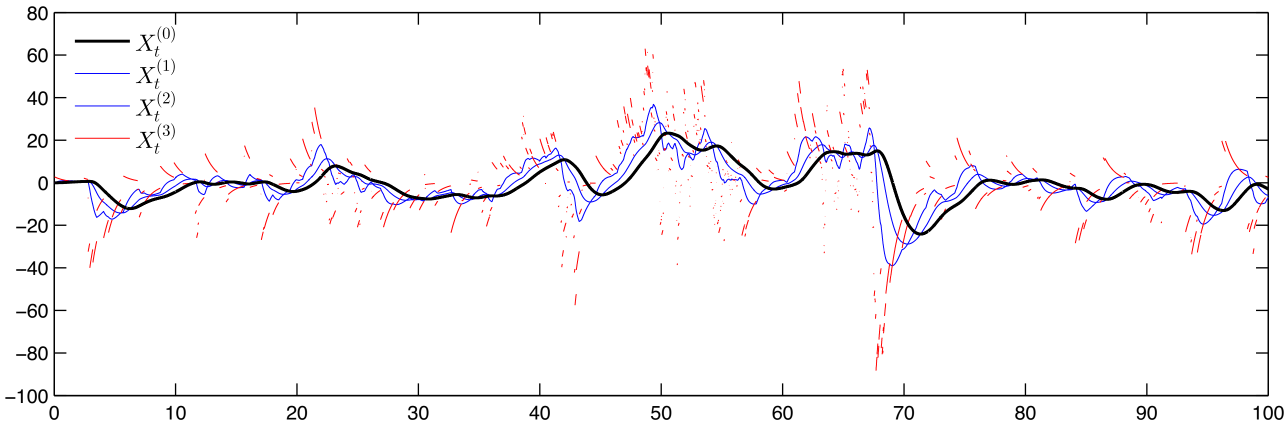

FIGURE 3. A realization of the process $X = (X_t)_{0\le t\le T}$ with random jump heights following a Normal distribution $\mathcal{N}(0, 100)$ for the choices $L = 1, n_1 = 3$, $\alpha_1 = 1$, $T = 100$, $f(x) = (1 + (x/2)^{3/2}) \wedge 30$ and $X_0 = (0, 0, 0, 0)$. The black trajectory (resp. red) corresponds to the realization of the process $(X_t^{(0)})_{0\le t\le T}$ (resp. $(X_t^{(3)})_{0\le t\le T}$). The blue trajectories correspond to the realization of the processes $(X_t^{(1)})_{0\le t\le T}$ and $(X_t^{(2)})_{0\le t\le T}$.

Proposition 5.1. *Let X be the Markov process whose generator is given by* (2.6) *with $L = 1, c_1 = 1, \alpha_1 > 0, n_1 \in \mathbb{N}$ and $f(x) = (\mu + x)1_{[0,\infty)}(x)$ with $\mu \ge 0$. For any $t \ge 0$, we write $S_t = \sum_{k=0}^{n_1} X_t^{(k)}$. Then for any $x = (x^{(0)}, \ldots, x^{(n_1)}) \in \mathbb{R}_+^{n_1+1}$,*

$$E[S_t] = \sum_{k=0}^{n_1} x^{(k)} + \begin{cases} \dfrac{\mu}{1-\alpha}\left(e^{t(1-\alpha)} - 1\right), & \textit{if } \alpha \ne 1 \\ \mu t & \textit{if } \alpha = 1. \end{cases} \tag{5.1}$$

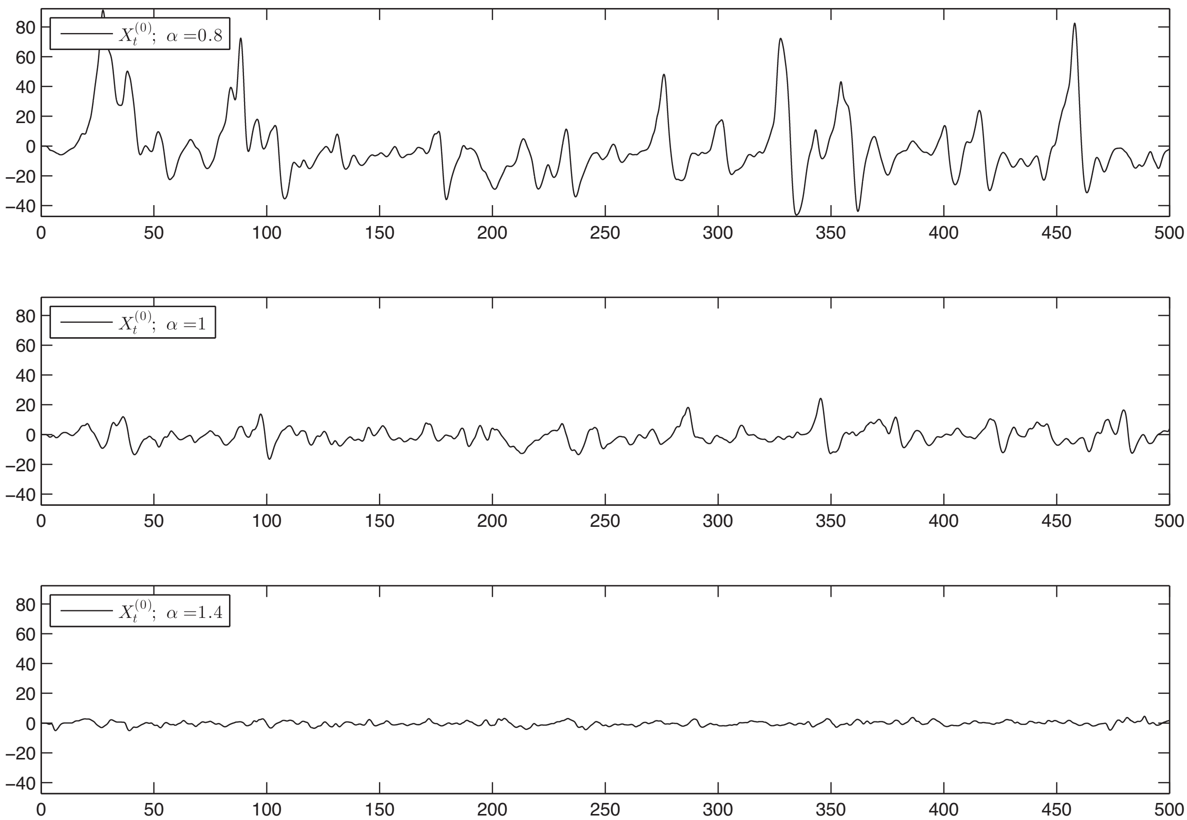

FIGURE 4. A joint realization of three Markovian cascades with $\alpha_1 = 0.8$ (upper panel), $\alpha_1 = 1$ (middle panel) and $\alpha_1 = 1.2$ (lower panel) respectively. Here, $L = 1$, $T = 500, n = 3$, the jump heights follow a Normal distribution $\mathcal{N}(0, 100)$ and $f(x) = 1 + \sigma/(1 + e^{-\beta(x-\rho)})$ where $\sigma = 20, \beta = 1/3$ and $\rho = 10$. Notice that the smaller the rate of decay α the larger the oscillations of the process $(X_t^{(0)})_{0\le t\ge T}$ are.

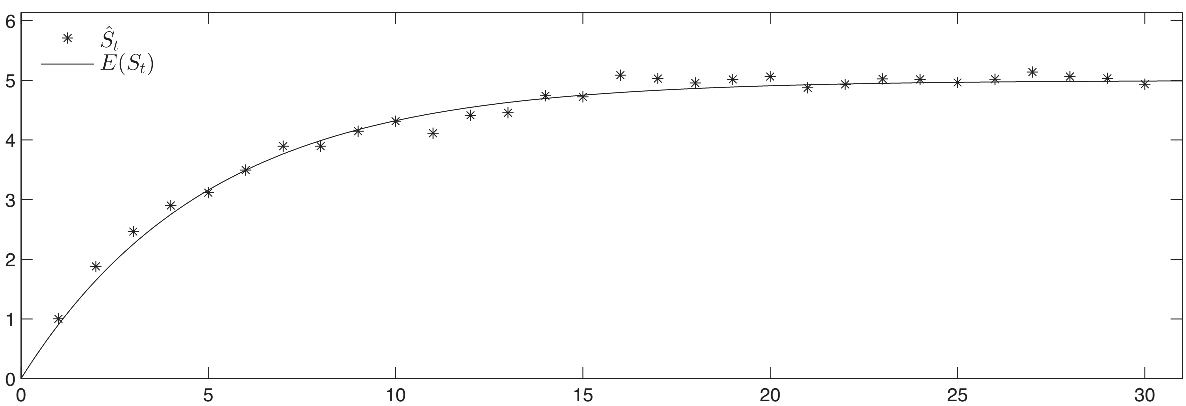

FIGURE 5. The graph of $(E[S_t])_{0\le t\le T}$, conditionally on $S_0 = 0$, for the choices $L = 1, n_1 = 3, c_1 = 1, \alpha_1 = 1.2, T = 30$ and $f(x) = (1 + x/5)1_{[0,\infty)}(x)$. The marks $*$ corresponds to the empirical expected value $\hat{S}_t$ of S_t computed at times $t \in \{0, 1, \ldots, 30\}$ based on 100 simulated samples of $(S_t)_{0\le t\le T}$.

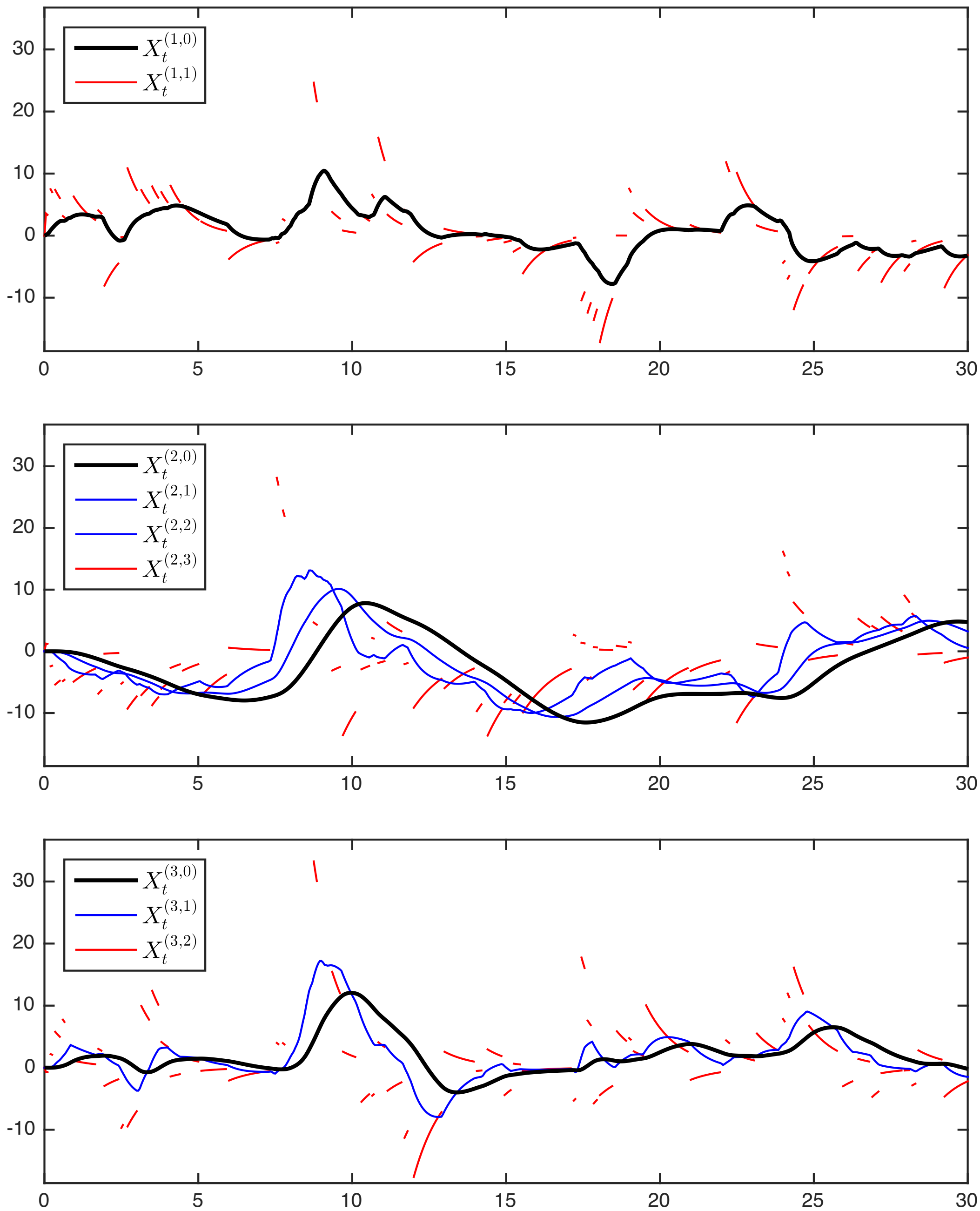

FIGURE 6. A realization of the process $X = (X_t)_{0\le t\le T}$ for $L = 3$, $n_1 = 1$, $n_2 = 3$, $n_3 = 2$, $\alpha_1 = 1.3$, $\alpha_2 = 0.8$, $\alpha_3 = 1$, $T = 30$, $f(x) = (2 + \exp(x/10)) \wedge 20$, random jump heights $c_1 = c_2 = c_3$ following a Normal distribution $\mathcal{N}(0, 25)$ and initial configuration $X_0 = 0^9$. In the upper panel it is shown separately the realization of the process $(X_t^{(1)})_{0\le t\le T}$, in the middle panel (resp. lower panel) that of $(X_t^{(2)})_{0\le t\le T}$ (resp. $(X_t^{(3)})_{0\le t\le T}$).

Proof. For $\mathbb{R}_+^{n_1+1} \ni (y^{(0)}, \ldots, y^{(n_1)}) \mapsto g(y^{(0)}, \ldots, y^{(n_1)}) = y^{(0)} + \ldots + y^{(n_1)}$ one checks that

$$Lg(y^{(0)}, \ldots, y^{(n_1)}) = \mu + (1-\alpha) g(y^{(0)}, \ldots, y^{(n_1)}).$$

By Dynkin's formula it follows that for each $t \geq 0$,

$$\begin{aligned} E[S_t] = E[g(X_t)] &= E[g(X_0)] + \int_0^t E[Lg(X_s)] \mathrm{d}s \\ &= \sum_{k=0}^{n_1} x^{(k)} + \mu t + (1-\alpha) \int_0^t E[S_s] \mathrm{d}s, \end{aligned}$$

from which it is easy to deduce the result by applying Gronwall's inequality. □

A comparison between the formula for $E[S_t]$ and and its estimated counterpart denoted by $\hat{S}_t$ is presented in Figure 5.

Appendix A. Proof of Theorem 3 for general L

Proof. We now prove Theorem 3.11 in the case $L > 1$. We write $X_t = (X_t^{(1)}, \ldots, X_t^{(L)})$ and $\varphi_t(x) = (\varphi_t^{(1)}(x), \ldots, \varphi_t^{(L)}(x))$ for the flow given in (2.8). Recall that elements of $\mathbb{R}^\kappa$ are denoted by $x = (x^{(1)}, \ldots, x^{(L)})$, where $x^{(i)} = (x^{(i,0)}, \ldots, x^{(i,n_i)})$ for each $1 \leq i \leq L$.

We prove by induction that for all $1 \leq k \leq L$, for all $x^* \in \mathbb{R}^\kappa$ there exists a neighbourhood C of x^* such that for all $x \in C$,

$$\begin{aligned} &P_{kT}(x, \mathrm{d}y) \geq \beta_k \nu_k(y^{(1)}, \ldots, y^{(k)}) \\ &Q_k((x, y^{(1)}, \ldots, y^{(k)}), \mathrm{d}y^{(k+1)} \ldots \mathrm{d}y^{(L)}) \mathrm{d}y^{(1)} \ldots \mathrm{d}y^{(k)}, \end{aligned} \tag{A.1}$$

where for open sets $I_l \subset \mathbb{R}^{n_l+1}, 1 \leq l \leq k$, ν_k is the uniform density on $I_1 \times \ldots \times I_k$, and where Q_k is a transition kernel from $\mathbb{R}^\kappa \times \mathbb{R}^{n_1+\ldots+n_k+k} \to \mathbb{R}^{n_{k+1}+\ldots+n_L+L-k}$.

Proof of (A.1) for $\boldsymbol{k = 1}$. We proceed as in part I and use the jump noise produced by $n_1 + 1$ jumps occurring during $[0, T]$ to produce a density for $X^{(1)}$: We impose inter-jump waiting times $t_1, \ldots, t_{n_1+1}$ under the constraint that $t_1 + \ldots + t_{n_1+1} < T$. To each jump time we associate jump heights $c_1, \ldots, c_{n_1+1}$, where each c_l is an element of $\mathbb{R}^L$, that is, $c_l = (c_l^{(i)}, 1 \leq i \leq L), 1 \leq l \leq n_1 + 1$.

In what follows we shall write $\underline{\mathbf{c}} = (c_1, \ldots, c_{n_1+1})$ and $\underline{\mathbf{c}}^{(i)} = (c_1^{(i)}, \ldots, c_{n_1+1}^{(i)})$. Moreover, we define $\underline{\mathbf{s}} = (s_1, \ldots, s_{n_1+1})$, for $s_k = T - t_1 + \ldots + t_k$. We call $\underline{\mathbf{s}}$ *admissible* if $T > s_1 > \ldots > s_{n_1+1} > 0$.

Then, conditionally on $X_0 = x$ and on the above choices, the position of X_T is given by

$$\gamma(x, \underline{\mathbf{c}}, \underline{\mathbf{s}}) := \varphi_T(x) + \sum_{i=1}^{L} \Big(c_1^{(i)} e^{-\alpha_i s_1} v_i(s_1) + \ldots + c_{n_1+1}^{(i)} e^{-\alpha_i s_{n_1+1}} v_i(s_{n_1+1}) \Big). \tag{A.2}$$

Here, $v_i(s) \in \mathbb{R}^\kappa, 1 \leq i \leq L$, is the vector given by $(v_i(s))^{(i,k)} = \frac{s^{n_i-k}}{(n_i-k)!}$, for $0 \leq k \leq n_i$, and with zero entries else. We shall write shortly $\gamma^{(i)}(x, \underline{\mathbf{c}}, \underline{\mathbf{s}}) \in \mathbb{R}^{n_i+1}$ for the $i-$the coordinate of $\gamma(x, \underline{\mathbf{c}}, \underline{\mathbf{s}})$, that is,

$$\gamma(x, \underline{\mathbf{c}}, \underline{\mathbf{s}}) = (\gamma^{(i)}(x, \underline{\mathbf{c}}, \underline{\mathbf{s}}), 1 \leq i \leq L).$$

In what follows, we will use the product form of the flow (2.8). By this we mean the fact that by the explicit form of the flow in (2.8), for each $1 \le i \le L$, we have that

$$\varphi_t^{(i,k)}(x) = e^{-\alpha_i t} \sum_{m=0}^{n_i - k} \frac{t^m}{m!} x^{(i,k+m)} =: \varphi_t^{(i,k)}(x^{(i)});$$

that is $\varphi_t^{(i)}(x) = \varphi_t^{(i)}(x^{(i)})$ does only depend on $x^{(i)}$, for any $1 \le i \le L$. As a consequence,

$$\gamma^{(i)}(x, \underline{\mathbf{c}}, \underline{\mathbf{s}}) = \gamma^{(i)}(x^{(i)}, \underline{\mathbf{c}}^{(i)}, \underline{\mathbf{s}})$$

does also depend only on $x^{(i)}$ and on $\underline{\mathbf{c}}^{(i)}$. As usual, we shall write, for any fixed pair $(x^{(i)}, \underline{\mathbf{c}}^{(i)})$,

$$\gamma^{(i)}_{(x^{(i)}, \underline{\mathbf{c}}^{(i)})} : \underline{\mathbf{s}} \mapsto \gamma^{(i)}(x^{(i)}, \underline{\mathbf{c}}^{(i)}, \underline{\mathbf{s}})$$

and similarly,

$$\gamma^{(i)}_{(\underline{\mathbf{c}}^{(\mathbf{i})}, \underline{\mathbf{s}})} : x^{(i)} \mapsto \gamma^{(i)}(x^{(i)}, \underline{\mathbf{c}}^{(i)}, \underline{\mathbf{s}}).$$

Fix any $x^* \in \mathbb{R}^\kappa$ and fix $\underline{\mathbf{c}}^*$ such that $\underline{\mathbf{c}}_l^{(*,i)} = (a_i + b_i)/2$, for all $1 \le l \le n_1 + 1$, where $(a_i, b_i) \subset \mathbb{R}$ such that $0 \notin (a_i, b_i)$ and $G_i((a_i, b_i)) > 0$, for all $1 \le i \le L$. Then there exists an open neighbourhood $J_1 = B_R(x^{(*,1)}) \times B_R(\underline{\mathbf{c}}^{(*,1)})$ of the pair $(x^{(*,1)}, \underline{\mathbf{c}}^{(*,1)})$ with $R < \min_i(b_i - a_i)$, and an open set $I \subset \mathbb{R}^{n_1+1}$ and for any pair $(x^{(1)}, \underline{\mathbf{c}}^{(1)}) \in J_1$ an open set $W_{x^{(1)}, \underline{\mathbf{c}}^{(1)}} \subset \mathbb{R}_+^{n_1+1}$ such that

$$\tilde{\gamma}^{(1)}_{x^{(1)}, \underline{\mathbf{c}}^{(1)}}(\underline{\mathbf{s}}) : \begin{cases} W_{x^{(1)}, \underline{\mathbf{c}}^{(1)}} \to I \\ \underline{\mathbf{s}} \mapsto \gamma^{(1)}(x^{(1)}, \underline{\mathbf{c}}^{(1)}, \underline{\mathbf{s}}), \end{cases}$$

is a diffeomorphism. Moreover,

$$\tilde{\beta}_1 = \inf_{x \in B_R(x^*), \underline{\mathbf{c}} \in B_R(\underline{\mathbf{c}}^*)} \inf_{\underline{\mathbf{s}} \in W_{x^{(1)}, \underline{\mathbf{c}}^{(1)}}} q_{x, \underline{\mathbf{c}}}(\underline{\mathbf{s}}) \left| \det \left(\frac{\partial \tilde{\gamma}^{(1)}_{x^{(1)}, \underline{\mathbf{c}}^{(1)}}(\underline{\mathbf{s}})}{\partial \underline{\mathbf{s}}} \right)^{-1} \right| > 0.$$

Let now $A_i \in \mathcal{B}(\mathbb{R}^{n_i+1})$, for all $1 \le i \le L$. Then for all $x \in B_R(x^*)$, using the change of variables $y^{(1)} = \tilde{\gamma}^{(1)}_{x^{(1)}, \underline{\mathbf{c}}^{(1)}}(\underline{\mathbf{s}})$,

$$\begin{aligned} P_T(x, A_1 \times \ldots \times A_L) \ge \tilde{\beta}_1 \int_I 1_{A_1}(y^{(1)}) \mathrm{d}y^{(1)} \Big[\int_{B_R(\underline{\mathbf{c}}^*)} \prod_{i=1}^L \underline{G}_i(\mathrm{d}\underline{\mathbf{c}}^{(i)}) \\ 1_{A_2}(\gamma^{(2)}_{x^{(2)}, \underline{\mathbf{c}}^{(2)}} \circ (\tilde{\gamma}^{(1)}_{x^{(1)}, \underline{\mathbf{c}}^{(1)}})^{-1}(y^{(1)})) \ldots 1_{A_L} \ (\gamma^{(L)}_{x^{(L)}, \underline{\mathbf{c}}^{(L)}} \circ (\tilde{\gamma}^{(1)}_{x^{(1)}, \underline{\mathbf{c}}^{(1)}})^{-1}(y^{(1)})) \Big] \\ =: \beta_1 \int_{\mathbb{R}^{n_1+1}} \nu_1(y^{(1)}) 1_{A_1}(y^{(1)}) \mathrm{d}y^{(1)} \int_{A_2 \times \ldots \times A_L} Q_1((x, y^{(1)}), \mathrm{d}y^{(2)} \ldots \mathrm{d}y^{(L)}), \end{aligned}$$

where $\nu^{(1)}$ is the uniform density on I, $\beta_1 = \tilde{\beta} \nu_1(I)$ and $\underline{G}_i(\mathrm{d}\underline{\mathbf{c}}^{(i)}) = \prod_{l=1}^{n_1+1} G_i(\mathrm{d}c_l^{(i)})$. This proves (A.1) for $k = 1$.

Induction step: $k-1$ implies k. Suppose that we have already established the result for $k-1$. Let $A_i \in \mathcal{B}(\mathbb{R}^{n_i+1})$, for all $1 \le i \le L$. We have

$$\begin{aligned} P_{kT}(x, A_1 \times \ldots \times A_L) &= \int P_T(y, A_1 \times \ldots \times A_L) P_{(k-1)T}(x, \mathrm{d}y) \\ &\ge \beta_{k-1} 1_C(x) \int P_T(y, A_1 \times \ldots \times A_L) \nu_{k-1}(y^{(1)}, \ldots, y^{(k-1)}) \\ &\quad Q_{k-1}((x, y^{(1)}, \ldots, y^{(k-1)}), \mathrm{d}y^{(k)} \ldots \mathrm{d}y^{(L)}) \mathrm{d}y^{(1)} \ldots \mathrm{d}y^{(k-1)}. \end{aligned} \tag{A.3}$$

We work conditionally on the choice of $y = (y^{(1)}, \ldots, y^{(L)})$ and proceed as in the first part, using the jump noise (of a sufficient number of jumps) to create a density for the variable $y^{(k)}$ and proving that the already produced density $\nu_{k-1}(y^{(1)}, \ldots, y^{(k-1)})$ of the first $k-1$ variables is well preserved.

As in the first step, we start with

$$P_T(y, A_1 \times \ldots \times A_L) \ge E_y \left(\prod_{i=1}^{L} 1_{A_i}(X_T^{(i)}), N_T = n_k + 1 \right). \tag{A.4}$$

We impose inter-jump waiting times $t_1, \ldots, t_{n_k+1}$ under the constraint that $t_1 + \ldots + t_{n_k+1} < T$ and associated jump heights $c_1, \ldots, c_{n_k+1}$, where each c_l is an element of $\mathbb{R}^L$, that is, $c_l = (c_l^{(i)}, 1 \le i \le L)$. These inter-jump waiting times will produce a density for the $k-$th variable $X_T^{(k)}$ as in the preceding steps.

To start with, let us introduce the following notation. For all $1 \le l < m \le L$, we write $y^{l:m} := (y^{(l)}, \ldots, y^{(m)})$, $\mathrm{d}y^{l:m} := \mathrm{d}y^{(l)} \ldots \mathrm{d}y^{(m)}$,$\underline{\mathbf{c}}^{l:m} := (\underline{\mathbf{c}}^{(l)}, \ldots, \underline{\mathbf{c}}^{(m)})$. For $x \in C$ we have

$$\begin{aligned} &\int P_T(y, A_1 \times \ldots \times A_L) \nu_{k-1}(y^{1:k-1}) Q_{k-1}((x, y^{1:k-1}), \mathrm{d}y^{k:L}) \mathrm{d}y^{1:k-1} \\ &\quad \ge \int \nu_{k-1}(y^{1:k-1}) Q_{k-1}((x, y^{1:k-1}), \mathrm{d}y^{k:L}) \mathrm{d}y^{1:k-1} \int \prod_{i=1}^{L} \underline{G}_i(\mathrm{d}\underline{\mathbf{c}}^{(i)}) \\ &\int \mathrm{d}s_1 \ldots \mathrm{d}s_{n_k+1} 1_{A_1 \times \ldots \times A_{k-1}}(\gamma^{1:k-1}(y^{1:k-1}, \underline{\mathbf{c}}^{1:k-1}, \underline{\mathbf{s}})) 1_{A_k}(\gamma^{(k)}(y^{(k)}, \underline{\mathbf{c}}^{(k)}, \underline{\mathbf{s}})) \\ &\quad 1_{A_{k+1} \times \ldots \times A_L}(\gamma^{k+1:L}(y^{k+1:L}, \underline{\mathbf{c}}^{k+1:L}, \underline{\mathbf{s}})) q_{y, \underline{\mathbf{c}}}(\underline{\mathbf{s}}). \end{aligned} \tag{A.5}$$

Let $N_1 := \sum_{l=1}^{k-1} (n_l + 1)$ be the dimension of $y^{1:k-1}$. We introduce now for any fixed $\underline{\mathbf{c}}^{\mathbf{1:k}}$ having all entries non zero and $y^{(k)} \in \mathbb{R}^{n_k+1}$,

$$\Phi_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})} : \left\{ \begin{array}{lll} \mathbb{R}^{N_1} \times \mathbb{R}^{n_k+1} & \to & \mathbb{R}^{N_1+n_k+1} \\ (y^{1:k-1}, \underline{\mathbf{s}}) & \mapsto & (\gamma^{1:k-1}(y^{1:k-1}, \underline{\mathbf{c}}^{1:k-1}, \underline{\mathbf{s}}), \gamma^{(k)}(y^{(k)}, \underline{\mathbf{c}}^{(k)}, \underline{\mathbf{s}})). \end{array} \right.$$

We write

$$\frac{\partial \Phi_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})}(y^{1:k-1}, \underline{\mathbf{s}})}{\partial y^{1:k-1} \partial \underline{\mathbf{s}}} = \Big[\frac{\partial \Phi_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})}(y^{1:k-1}, \underline{\mathbf{s}})}{\partial y^{(1)}}, \ldots, \frac{\partial \Phi_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})}(y^{1:k-1}, \underline{\mathbf{s}})}{\partial s_{n_k+1}} \Big]$$

to denote the Jacobian matrix of the the map $(y^{1:k-1}, \underline{\mathbf{s}}) \mapsto \Phi_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})}(y^{1:k-1}, \underline{\mathbf{s}})$. By the properties of the flow (2.8), it follows that

$$\frac{\partial \Phi_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})}(y^{1:k-1}, \underline{\mathbf{s}})}{\partial y^{(1)}} = \begin{pmatrix} \frac{\partial \gamma^{(1)}(y^{(1)}, \underline{\mathbf{c}}^{(\mathbf{1})}, \underline{\mathbf{s}})}{\partial y^{(1)}} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\partial \varphi_T(y^{(1)})}{\partial y^{(1)}} \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where by the "cascade structure" of the flow (2.8),

$$\frac{\partial \varphi_T(y^{(1)})}{\partial y^{(1)}} = \begin{pmatrix} e^{-\alpha_1 T} & * & * & * \\ 0 & e^{-\alpha_1 T} & * & * \\ \vdots & & \ddots & * \\ 0 & \cdots & 0 & e^{-\alpha_1 T} \end{pmatrix}.$$

Therefore,

$$\frac{\partial \Phi_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})}(y^{1:k-1}, \underline{\mathbf{s}})}{\partial y^{1:k-1} \partial \underline{\mathbf{s}}} = \begin{pmatrix} A & * \\ 0 & B \end{pmatrix},$$

where A is an $N_1 \times N_1$ upper diagonal matrix having entries of the type $e^{-\alpha_i T}, 1 \leq i \leq k-1$, on the diagonal, where 0 is the $(n_k+1) \times N_1$ matrix having all zero entries, and where $B = \left(\frac{\partial \gamma^{(k)}(y^{(k)}, \underline{\mathbf{c}}^{(k)}, \underline{\mathbf{s}})}{\partial \underline{\mathbf{s}}} \right)$. According to Part I of the proof, $\det B \neq 0$ whenever $\underline{\mathbf{c}}^{(k)}$ has only non zero coordinates, for any fixed $y^{(k)}$, for admissible $\underline{\mathbf{s}}$. It is immediate to see that

$$\det \left(\frac{\partial \Phi_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})}(y^{1:k-1}, \underline{\mathbf{s}})}{\partial y^{1:k-1} \partial \underline{\mathbf{s}}} \right) = e^{-(\alpha_1(n_1+1) + \ldots \alpha_{k-1}(n_{k-1}+1))T} \cdot \det B \neq 0$$

in this case. Therefore, arguing as in Part I, for any $y^{(*,k)} \in \mathbb{R}^{n_k+1}$ and $\underline{\mathbf{c}}^{(*,\mathbf{1:k})}$ having non zero entries, for all $y^{(*,1:k-1)}$ and an admissible $\underline{\mathbf{s}}^*$ there exist open neighbourhoods $J_k = B_R(y^{(*,k)}) \times B_R(\underline{\mathbf{c}}^{(*,1:k)})$ and an open set $I \subset \mathbb{R}^{N_1+n_k+1}$ containing $(y^{(*,1:k-1)}, \underline{\mathbf{s}}^*)$ and for any pair $(y^{(k)}, \underline{\mathbf{c}}^{(1:k)}) \in J_k$ an open set $W_{y^{(k)}, \underline{\mathbf{c}}^{(1:k)}}$ such that

$$\tilde{\Phi}_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})}(y^{1:k-1}, \underline{\mathbf{s}}) : \begin{cases} W_{y^{(k)}, \underline{\mathbf{c}}^{(\mathbf{1:k})}} & \to I \\ (y^{1:k-1}, \underline{\mathbf{s}}) & \mapsto \Phi_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})}(y^{1:k-1}, \underline{\mathbf{s}}), \end{cases}$$

is a diffeomorphism.

In what follows, in order to ease notation, we shall shortly write

$$\tilde{\Phi}(y^{1:k-1}, \underline{\mathbf{s}}) := \tilde{\Phi}_{(\underline{\mathbf{c}}^{\mathbf{1:k}}, y^{(k)})}(y^{1:k-1}, \underline{\mathbf{s}})$$

and

$$\Psi = \tilde{\Phi}^{-1} \tag{A.6}$$

for the associated inverse function. Ψ taking values in some (subset of) $\mathbb{R}^{N_1+n_k+1}$, we shall write as before $\Psi^{(i)}$ for its coordinates and $\Psi^{1:N_1}$ for the first N_1 of its coordinates, corresponding to $y^{1:k-1}$, and $\Psi^{N_1+1:N_1+n_k+1}$ for the last coordinates, corresponding to $\underline{\mathbf{s}}$.

We choose any $x \in C$ and $y^* \in$ supp $(\nu_{k-1}(y^{1:k-1})Q_{k-1}((x, y^{1:k-1}), \mathrm{d}y^{k:L})\mathrm{d}y^{1:k-1})$ and obtain (recall (A.4) and (A.5)),

$$\begin{aligned}
&\int P_T(y, A_1 \times \ldots \times A_L)\nu_{k-1}(y^{1:k-1})Q_{k-1}((x, y^{1:k-1})), \mathrm{d}y^{k:L})\mathrm{d}y^{1:k-1} \\
&\quad \geq \int_{B_R(y^*)} \nu_{k-1}(y^{1:k-1})Q_{k-1}((x, y^{1:k-1}), \mathrm{d}y^{k:L})\mathrm{d}y^{1:k-1} \int_{B_R(\underline{\mathbf{c}}^*)} \prod_{i=1}^{L} \underline{G}_i(\mathrm{d}\underline{\mathbf{c}}^{(i)}) \\
&\int_{\mathbb{R}_+^{n_k+1}} 1_{W_{y^{(k)},\underline{\mathbf{c}}^{(1:\mathbf{k})}}}(y^{1:k-1}, \underline{\mathbf{s}})\; q_{y,\underline{\mathbf{c}}}(\underline{\mathbf{s}})\; 1_{A_1\times\ldots\times A_k}(\tilde\Phi(y^{1:k-1}, \underline{\mathbf{s}})) \\
&\quad 1_{A_{k+1}\times\ldots A_L}(\gamma^{k+1:L}_{y^{k+1:L},\underline{\mathbf{c}}}(\underline{\mathbf{s}}))\mathrm{d}s_1 \ldots \mathrm{d}s_{n_k+1}.
\end{aligned}$$

In the above formula, R is chosen sufficiently small such that $B_R(\underline{\mathbf{c}}^*)$ contains only jump heights with non zero entries. We then use the change of variables

$$z^{1:k} := \tilde\Phi(y^{1:k-1}, \underline{\mathbf{s}}).$$

Choose now R sufficiently small such that

$$\begin{aligned}
\tilde\beta_k := \inf_{y \in B_R(y^*)} \inf_{\underline{\mathbf{c}} \in B_R(\underline{\mathbf{c}}^*)} \inf_{\underline{\mathbf{s}}\,:\,(y^{1:k-1},\underline{\mathbf{s}})\in W_{y^{(k)},\underline{\mathbf{c}}^{\mathbf{1:k}}}} q_{y,\underline{\mathbf{c}}}(\underline{\mathbf{s}}) \\
\times\left|\det\Big(\frac{\partial\tilde\Phi(y^{1:k-1}, \underline{\mathbf{s}})}{\partial y^{1:k-1}\partial\underline{\mathbf{s}}}\Big)^{-1}\right|\nu_{k-1}(y^{1:k-1}) > 0.
\end{aligned}$$

Let $C_2^* = B_R(y^{(*,k)}, \ldots, y^{(*,L)})$. Then

$$\begin{aligned}
&\int P_T(y, A_1 \times \ldots \times A_L)\nu_{k-1}(y^{1:k-1})Q_{k-1}((x, y^{1:k-1})), \mathrm{d}y^{k:L})\mathrm{d}y^{1:k-1} \\
&\geq \tilde\beta_k \int_I (\prod_{j=1}^{k} 1_{A_j}(z^{(j)})\mathrm{d}z^{1:k}\Big[\int_{B_R(\underline{\mathbf{c}}^*)} \prod_{i=1}^{L} \underline{G}_i(\mathrm{d}\underline{\mathbf{c}}^{(i)}) \\
&\int_{C_2^*} Q_{k-1}((x, \Psi^{1:N_1}(z^{1:k}), \mathrm{d}y^{k:L})1_{A_{k+1}\times\ldots A_L}(\gamma^{k+1:L}_{y^{k+1:L},\underline{\mathbf{c}}} \circ \Psi^{N-1+1:N_1+n_k+1})(z^{1:k})\Big] \\
&=: \tilde\beta_k \int_I (\prod_{j=1}^{k} 1_{A_j}(z^{(j)})\mathrm{d}z^{1:k}\left[\int Q_k((x, z^{1:k}), \mathrm{d}z^{k+1:L})1_{A_{k+1}}(z^{(k+1)}) \cdot \ldots \cdot 1_{A_L}(z^{(L)})\right].
\end{aligned}$$

Together with (A.3), this shows that (A.1) holds also for k, and this finishes the induction step. By taking finally $k = L$ in (A.1), this implies the assertion of the Theorem.

$\square$

Acknowledgements. The authors thank an anonymous referee for critical reading and useful remarks. E.L. thanks B. Cloez for fruitful discussions concerning Wasserstein coupling at an early stage of this work. This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01); it is part of USP project *Mathematics, computation, language and the brain*, FAPESP project *Research, Innovation and Dissemination Center for Neuromathematics* (grant 2013/07699-0), CNPq projects *Stochastic modelling of the brain activity* (grant 480108/2012-9) and *Plasticity in the brain after a brachial plexus lesion* (grant 478537/2012-3). A.D. and G.O. were fully supported by a FAPESP fellowship (grants 2016/17791-9 and 2016/17789-4 respectively).

References

[1] J. Azéma, M. Duflo and D. Revuz, Mesures invariantes des processus de Markov récurrents. Sém. Proba. III, Vol. 88 of *Lecture Notes in Mathematics.* Springer, Berlin (1969) 24–33.

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

[3] F. Bolley, Separability and completeness for the wasserstein distance. Sém. Proba. XLI, Vol. 1934 of *Lecture Notes in Mathematics. Springer, Berlin* (2008) 371–377.

[4] P. Brémaud and L. Massoulié, Stability of nonlinear Hawkes processes. *Ann. Probab.* **24** (1996) 1563–1588.

[5] J. Chevallier, Mean-field limit of generalized Hawkes processes. *Stoch. Proc. Appl.* **127** (2017) 3870–3912.

[6] J. Chevallier, M.J. Caceres, M. Doumic and P. R. Bouret, Microscopic approach of a time elapsed neural model. *Math. Models Methods Appl. Sci.* **25** (2015) 2669–2719.

[7] A. Dassios and H. Zhao, Exact simulation of Hawkes process with exponentially decaying intensity. *Electron. Commun. Probab.* **18** (2013) 1–13.

[8] S. Delattre, N. Fournier and M. Hoffmann, Hawkes processes on large networks. *Ann. App. Probab.* **26** (2016) 216–261.

[9] S. Ditlevsen and E. Löcherbach, Multi-class oscillating systems of interacting neurons. *Stoch. Proc. Appl.* **127** (2017) 1840–1869.

[10] S. Ditlevsen, K.P. Yip and N.H Holstein-Rathlou, Parameter estimation in a stochastic model of the tubuloglomerular feedback mechanism in a rat nephron. *Math. Biosci.* **194** (2005) 49–69.

[11] D. Down, S.P. Meyn and R.L. Tweedie, Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **23** (1995) 1671–1691.

[12] N. Hansen, P.R Bouret and V. Rivoirard, Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli* **21** (2015) 83–143.

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

[14] A.G. Hawkes and D. Oakes, A cluster process representation of a self-exciting process. *J. Appl. Probab.* **11** (1974) 93–503.

[15] J. Jacod, Multivariate Point Processes: Predictable Projection, Radon-Nikodym Derivatives, Representation of Martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **31** (1975) 235–253.

[16] D.W. Kammler, Approximation with sums of exponentials in $l_p[0,\infty)$. *J. Approx. Theory* **16** (1976) 384–408.

[17] E. Löcherbach, Convergence to equilibrium for time inhomogeneous jump diffusions with state dependent jump intensity. Preprint arXiv:1712.03507 (2017).

[18] 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. Process. Appl.* **118** (2008) 1301–1321.

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

[20] S.T. Rachev, Probability Metrics and the Stability of Stochastic Models. John Wiley and Sons, Chichester, USA (1991).

[21] M.B. Raad, S. Ditlevsen and E. Löcherbach, Age dependent Hawkes process. Preprint arXiv:1806.06370 (2018).

[22] A.C. Skeldon and I. Purvey, The effect of different forms for the delay in a model of the nephron. *Math. Biosci. Eng.* **2** (2005) 97–109.

[23] L. Zhu, Large deviations for Markovian nonlinear Hawkes processes. *Ann. Appl. Probab.* **25** (2015) 548–581.