

# Stability and mean-field limits of age dependent Hawkes processes

Mads Bonde Raad<sup>a</sup>, Susanne Ditlevsen<sup>a</sup> and Eva Löcherbach<sup>b</sup>

<sup>a</sup>*Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark.  
E-mail: [madsraad@gmail.com](mailto:madsraad@gmail.com)*

<sup>b</sup>*Université de Paris 1 Panthéon-Sorbonne, SAMM, F-75013 Paris*

Received 25 June 2018; revised 5 August 2019; accepted 7 September 2019

---

**Abstract.** In the last decade, Hawkes processes have received a lot of attention as good models for functional connectivity in neural spiking networks. In this paper we consider a variant of this process; the age dependent Hawkes process, which incorporates individual post-jump behavior into the framework of the usual Hawkes model. This allows to model recovery properties such as refractory periods, where the effects of the network are momentarily being suppressed or altered. We show how classical stability results for Hawkes processes can be improved by introducing age into the system. In particular, we neither need to a priori bound the intensities nor to impose any conditions on the Lipschitz constants. When the interactions between neurons are of mean-field type, we study large network limits and establish the propagation of chaos property of the system.

**Résumé.** Depuis la dernière décennie il s'est avéré que la classe des processus de Hawkes fournit un bon modèle pour décrire la connectivité fonctionnelle dans un réseau de neurones. Dans cet article nous étudions une variante de ce processus, le processus de Hawkes structuré en âge. Cette structure en âge rajoute un comportement individuel après les sauts à la dynamique de chaque composante, ce qui permet en particulier de décrire une période refractaire durant laquelle l'influence du réseau est supprimée ou au moins modifiée. Nous améliorons les résultats de stabilité classiques pour les processus de Hawkes dans ce cadre. En particulier, nous n'avons ni besoin de supposer que les intensités sont bornées, ni d'imposer une condition aux normes Lipschitz des fonctions taux de saut. Lorsque les interactions entre les neurones sont du type champ moyen, nous étudions les limites en grande population et nous démontrons la propriété de propagation du chaos du système.

MSC: 60G55; 60G57; 60K05

*Keywords:* Multivariate nonlinear Hawkes processes; Multivariate point processes; Mean-field approximations; Age dependency; Stability; Coupling; Piecewise deterministic Markov processes

---

## 1. Introduction

In nature, macroscopic events caused by many microscopic events in an interacting network of units often exhibit a cascading structure, so that they come in waves, for example caused by some events inhibiting or exciting the occurrence of other events. Technological development permitting high-frequency data sampling in recent decades has made it possible to perform detailed analysis of the dependencies between units interacting in a cascade structure. It is therefore relevant to develop models which can quantify temporal interactions between many units on a microscopic level. There are many examples of phenomena of interest which occur in cascades and have been analyzed by Hawkes processes. Examples include bankruptcies in finance that propagate through a market, giving rise to volatility clustering observations [1], interactions on social media [22], and pattern dependencies in DNA [21].

Hawkes processes are point processes where the intensity function is stochastic and allowed to depend on the past history, introducing memory in the temporal evolution of the stochastic process. Initially introduced in [14] and [15], they have commonly been used to model neurophysiological processes [3,4,8,11,20], and the application we have in mind is to model functional connectivity between neurons in a network. When neurons send an electric signal, the so-called *action potential* or *spike*, they excite or inhibit recipient neurons in the network (the *post-synaptic* neurons). Jumps of the  $i$ th unit of the Hawkes process are then identified with the spike times of the  $i$ th neuron. Moreover, the biological process imposes a strong self-inhibition on a neuron that has just emitted a spike. This period of about  $2ms$  is called the *absolute refractory*

period, and in this phase it is virtually impossible for a neuron to spike again. The neuron then gradually regains its ability to spike in the longer *relative refractory period*. It was proposed in [3] to model absolute and relative refractory periods in neuronal spike trains by *age dependent Hawkes processes*, where the age of a unit is defined as the time passed since the last time it jumped, and thus, it resets to zero at each jump time. The present article is devoted to a thorough study of its stability properties and associated mean-field limits.

It turns out that the classical stability results for Hawkes processes can be improved by introducing age into the system. In particular, we neither need a priori bound the intensities nor to impose any conditions on the Lipschitz constants. This is interesting not only from a mathematical but also from a biological point of view. If a network of neurons is transmitting some information over time, it is not operational nor realistic that intensities should be bounded, a given signal should be transmitted without any necessary delay. However, if the activity explodes, the entire system breaks down. Introducing a refractory period on the individual neuron stabilizes the system, while at the same time the information is still transferred effectively by the network.

In the present paper we consider multivariate counting processes  $(Z_t^i)_{t \geq 0}$ ,  $i = 1, \dots, N$ , where  $N$  is the number of units in the network.

The counting processes  $(Z_t^i)_{t \geq 0}$ ,  $i = 1, \dots, N$ , are characterized by their *conditional intensities* which can informally be described as the instantaneous jump rate, given the past, that is,

$$\lambda_t^i dt \approx P(Z^i \text{ has a jump in } (t, t + dt] | \mathcal{F}_t),$$

where  $\mathcal{F}_t$  is the history of the entire network of neurons. The Hawkes process is defined by imposing a specific structure on the conditional intensity.

Before giving a precise definition of the age dependent Hawkes process in Definition 1.1 below, let us start by describing and discussing its related components in a less formal way. We consider an  $N$ -dimensional point process  $(Z^i)_{i \leq N}$ , where each coordinate  $Z^i$  counts the jump events of the  $i$ th unit. The intensity of this process is a predictable process depending on the history  $\mathcal{F}_t$  before and up to time  $t$ . It is assumed to have the form

$$\lambda_t^i = \psi^i(X_t^i, A_t^i), \tag{1}$$

where  $X_t^i$  is the *memory process*, a predictable process depending on the history  $\mathcal{F}_t$  of the process,  $\psi^i$  is the *rate function*, and  $A_t^i$  is the *age process* of  $Z^i$ . Here follows a brief introduction of the involved objects.

The *Rate Function*  $(x, a) \mapsto \psi^i(x, a)$  describes how the memory and the age influence the intensity of the  $i$ th unit. The existence of a non-exploding Hawkes process is generally ensured by assuming that  $\psi^i$  is sub-linear in  $x$ . Often, stronger assumptions such as a uniform bound on  $\psi^i$  is also imposed to prove basic properties. In this article we will work under standard Lipschitz- and linear-growth-conditions, but we shall not need to bound the rate function nor its Lipschitz constant.

The *Age Process*  $A_t^i$  associated to the  $i$ th process  $Z^i$  is the time elapsed since the last jump time of  $Z^i$  before time  $t$ , that is,

$$A_t^i = \begin{cases} A_0^i + t, & \text{if } Z^i \text{ has not jumped between time 0 and time } t, \\ t - \sup\{s < t : \Delta Z_s^i > 0\}, & \text{otherwise,} \end{cases}$$

where  $\Delta Z_s^i = Z_s^i - Z_{s-}^i = Z_s^i - \lim_{\varepsilon \rightarrow 0+} Z_{s-\varepsilon}^i$  are the jumps.

The *Memory Process*  $X_t^i$  integrates the effects of previous jumps in the network, where the influence from the past is a weighted average of all previous jumps of all units that directly affect unit  $i$  (the *pre-synaptic* neurons). Each unit has its own memory process, even if they all depend on the same common history of all units, but they are affected in individual ways. More precisely, the  $i$ th memory process is assumed to have the structure

$$X_t^i = \sum_{j=1}^N \sum_{\tau < t: \Delta Z_\tau^j = 1} h_{ij}(t - \tau) + R_t^i = \sum_{j=1}^N \int_0^{t-} h_{ij}(t - s) Z^j(ds) + R_t^i.$$

In the definition of the memory process we have introduced two new objects.

The *Weight Function*  $h_{ij}(t)$  determines how much a jump of unit  $j$  that occurred  $t$  time units ago contributes to the present memory of unit  $i$ . Positive  $h_{ij}(t)$  means excitation of unit  $i$  when a jump of unit  $j$  occurred  $t$  time units ago, while negative  $h_{ij}(t)$  means inhibition.

The *Initial Signal*  $R_t^i$  is a process assumed to be known at time  $t = 0$ . It should be thought of as a memory process which the process inherits from past time.

When  $R^i \equiv 0$  and  $\psi^i(x, a) = f^i(x)$  for a suitable function  $f^i$ , we obtain the usual non-linear Hawkes process which has been studied in detail, e.g., in [2].

In Section 2, we discuss stability of the age dependent Hawkes process. The main assumption is a post-jump bound on the intensity, corresponding to a strong self-inhibition for a short time interval after a spike. This models the refractory period. We do not impose any a priori bounds on the intensities. Within this sub-model, we are able to prove stability properties for the  $N$ -dimensional Hawkes process. The results we obtain are similar to what has been shown for ordinary nonlinear Hawkes processes in [2] and recently in [5]. This last paper is however entirely devoted to the study of weight functions which are of compact support giving rise to explicit regeneration points when the process comes back to the all zero measure. Compared to these studies, it turns out that the natural self-inhibition by the age processes eliminates the need of controlling the Lipschitz constant of  $\psi$ , and we do not need any restriction on the support of the weight functions. We also discuss which starting conditions (that is, which form of an initial process) will ensure that the Hawkes process jumps in synchrony with the invariant process eventually. When this holds, the Hawkes process is said to *couple* with the invariant process. These results are collected within our first main theorem, Theorem 2.2.

During the proof of the stability properties, other interesting properties of the model are discussed, such as a nice-behaving domination of the intensities (Lemma 2.4).

In Section 3, we study a mean-field setup and associated mean-field limits. More precisely, we consider  $N$  interacting units which are organized within  $\mathcal{K}$  classes of populations. Each unit belongs to one of these classes, and any two units within the same class  $k$  are assumed to be similar,  $k = 1, \dots, \mathcal{K}$ . This means that they have the same rate function  $\psi^k$ , memory process  $X^k$  and initial signal  $R^k$ , and the weight function describing the influence of any unit belonging to another class  $l$  is given by  $N^{-1}h_{kl}$ . However, each unit still has its own age process.

In this setup we establish a limiting distribution for a large scale network,  $N \rightarrow \infty$ . Let  $N_k$  be the number of units in class  $k$  with proportion  $N_k/N$ , and assume  $\lim_{N \rightarrow \infty} N_k/N = p_k > 0$ . We index the  $j$ th unit within class  $k$  by  $Z^{kj}$ ,  $k = 1, \dots, \mathcal{K}$ ,  $j = 1, \dots, N_k$ . It is sensible to assume that small contributions from unit  $Z^{lj}$  to the memory process  $X^k$  of the  $k$ th class disappear in the large-scale dynamics, meaning that  $N^{-1} \sum_{j=1}^{N_l} \int_0^{t-s} h_{kl}(t-s) Z^{lj}(ds) \approx p_l \int_0^t h_{kl}(t-s) d\mathbb{E}Z^{l1}(s)$  for large  $N$ , for any  $1 \leq l \leq \mathcal{K}$ . Therefore, if a limiting point process  $(Z^{kj})_{j \in \mathbb{N}}$  exists, we expect that any  $Z^{kj}$ ,  $k \leq \mathcal{K}$ ,  $j \geq 1$ , should have intensity

$$\lambda_t^{kj} = \psi^k(x_t^k, A_t^{kj}),$$

where  $A_t^{kj}$  is the age process of  $Z^{kj}$ , and the process  $x_t^k$  is deterministic, given by

$$x_t^k = \sum_{l=1}^{\mathcal{K}} p_l \int_0^t h_{kl}(t-s) d\mathbb{E}Z^{l1}(s) + r_t^k,$$

where  $r_t^k$  is a suitable limit of the initial processes in population  $k$ . In Theorem 3.2, we discuss criteria under which such a system exists. Our second main theorem, Theorem 3.3, shows that this system will indeed be a limit process for the age dependent Hawkes processes for  $N \rightarrow \infty$ . We also discuss in Lemma 3.4 how robust the system is to adjusting the weight functions. Not only is this robustness a good model feature in itself, but it also allows approximation of an arbitrary age dependent Hawkes process, using weight functions with better features. Examples are weight functions given by Erlang densities or exponential polynomials which induce Markovian systems, see [8].

We close our article with the [Appendix](#) where we collect some proofs and useful results about counting processes.

### Notation, definitions and core assumptions

Throughout this article, we will be working on a background probability space  $(\Omega, \mathcal{F}, P)$ , and all random variables are assumed to be defined on this space. If  $v = (v_1, \dots, v_d)$  is a  $d$ -dimensional Euclidian vector, then  $|v| = \sum_{i=1}^d |v_i|$  denotes the 1-distance. Moreover, for a  $d$ -dimensional process  $X$  we define the running-supremum of the 1-distance as  $\|X\|_t = \sup_{s \leq t} |X_s|$ .

We recall the basic Stieltjes integration notation. Let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a càdlàg function. The variation of  $g$  on a bounded interval  $I$  is given by

$$V_g(I) = \sup_{(x_i) \in \mathcal{I}} |g(x_i) - g(x_{i-1})| < \infty.$$

where  $\mathcal{I}$  denotes the system of all finite sets of increasing indices  $(x_i) \subset I$ .  $g$  is said to be of finite variation, if the variation is finite on all bounded intervals. For such  $g$  there exist two singular  $\sigma$ -finite measures  $\mu_g^+, \mu_g^-$  s.t.  $\mu_g := \mu_g^+ - \mu_g^-$  which

satisfies  $\mu_g(a, b) = g(b) - g(a)$ . The variation measure  $|dg| := \mu_g^+ + \mu_g^-$  satisfies  $|dg|(a, b) = V_g(a, b)$ . If  $f : \mathbb{R} \rightarrow \mathbb{R}$  is a measurable function such that  $\int |f| |dg| < \infty$  then we define the *Lebesgue-Stieltjes integral* as

$$\int f(x) dg(x) = \int f(x) d\mu_g(x),$$

see e.g. [13]. If  $\nu$  is a measure on  $\mathbb{R}^2$  we shall also use the following notation for the integral over semi-closed boxes

$$\int_a^b \int_c^d f(x, y) \nu(dx, dy) = \int \mathbb{1}\{(a, b]\}(x) \mathbb{1}\{(c, d]\}(y) f(x, y) \nu(dx, dy).$$

In the following we introduce the core mathematical objects and assumptions needed to discuss the age dependent Hawkes process.

$\pi$  |  $\pi$  and  $\pi^i, i \in \mathbb{N}$ , are i.i.d. Poisson Random Measures (PRMs) on  $\mathbb{R} \times \mathbb{R}_+$  with Lebesgue intensity measure. For any  $t \in \mathbb{R}$ , we define the  $\sigma$ -algebra  $\tilde{\mathcal{F}}_t$  induced by the projections

$$\pi(A \cap ((-\infty, t] \times \mathbb{R}_+)), \pi^i(A \cap ((-\infty, t] \times \mathbb{R}_+)) \quad \text{for } A \in \mathcal{B}(\mathbb{R} \times \mathbb{R}_+), i \in \mathbb{N}.$$

We equip the space  $(\Omega, \mathcal{F}, P)$  with the filtration  $(\mathcal{F}_t)_{t \in \mathbb{R}}$  which is the completion of  $(\tilde{\mathcal{F}}_t)_{t \in \mathbb{R}}$ .

$h$  | weight functions: For all  $1 \leq i, j \leq N, h_{ij} : \mathbb{R}_+ \rightarrow \mathbb{R}$  is a locally integrable function.

$R_t$  | initial signals: For all  $1 \leq i \leq N, (R_t^i)_{t \geq 0}$  is an  $\mathcal{F}_0 \otimes \mathcal{B}$  measurable process on  $t \in \mathbb{R}_+$  such that  $t \mapsto \mathbb{E}R_t^i$  is locally bounded.

$\psi$  | rate functions: For all  $1 \leq i \leq N, \psi^i : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a measurable function which is  $L$ -Lipschitz in  $x$  when the age variables agree, and otherwise sub-linear in  $x$ , i.e.,

$$|\psi^i(x, a) - \psi^i(x', a')| \leq L|x - x'| \mathbb{1}\{a = a'\} + L(\max(|x'|, |x|) + 1) \mathbb{1}\{a \neq a'\}, \tag{2}$$

for some  $L \geq 1$ . By taking a possibly larger  $L$  we may, and will, also assume that

$$\psi^i(x, a) \leq L(1 + |x|) \quad \forall (x, a) \in \mathbb{R} \times \mathbb{R}_+. \tag{3}$$

$A_0$  | initial ages  $(A_0^i)_{i \in \mathbb{N}}$  are  $\mathcal{F}_0$ -measurable random variables with support in  $\mathbb{R}_+$ .

With these definitions, we introduce the age dependent Hawkes process.

**Definition 1.1 (The age dependent Hawkes process).** Let  $N \in \mathbb{N}$  and let  $(Z, X, A) = ((Z^i)_{1 \leq i \leq N}, (X^i)_{1 \leq i \leq N}, (A^i)_{1 \leq i \leq N})$  be a triple consisting of an  $N$ -dimensional counting process  $Z$ , an  $N$ -dimensional predictable process  $X$ , and an  $N$ -dimensional adapted càglàd process  $A$ . The triple is an  $N$ -dimensional age dependent Hawkes process with weight functions  $(h_{ij})_{i, j \leq N}$ , spiking rates  $(\psi^i)_{i \leq N}$ , initial ages  $(A_0^i)_{i \leq N}$ , and initial signals  $(R^i)_{i \leq N}$  if almost surely all sample paths solve the system

$$\begin{aligned} Z_t^i &= \int_0^t \int_0^\infty \mathbb{1}\{z \leq \psi^i(X_s^i, A_s^i)\} \pi^i(ds, dz), \\ X_t^i &= \sum_{j=1}^N \int_0^{t-} h_{ij}(t-s) Z^j(ds) + R_t^i, \\ A_t^i - A_0^i &= t - \int_0^{t-} A_s^i Z^i(ds), \end{aligned} \tag{4}$$

for all  $t \geq 0$ .

**Remark 1.** Notice that we choose the càglàd version of  $A$ , that is,  $A$  is left continuous and has right limits. All age processes presented in the article will be càglàd as well. This is notationally convenient as the age process will appear in the intensities for most point processes treated in this article.

**Example 1.2 (Examples of rate functions).** A possible choice of rate function is  $\psi^i(x, a) = f^i(x)g^i(a)$ , where  $f^i$  is Lipschitz and  $g^i$  is bounded. In particular, for the neuroscience application, it is possible to model an absolute refractory

period of length  $\delta$  by putting  $\psi^i(x, a) = f^i(x)\mathbb{1}\{a > \delta\}$ . Another example are the rate functions considered in the time elapsed neuron network model of [19] given by  $\psi^i(x, a) = f^i(x)g^i(a - s^i(x))$ , where  $g^i \equiv 0$  on  $\mathbb{R}_-$ , and where  $g^i, s^i$  are Lipschitz and bounded.

**Example 1.3 (Initial process).** Let  $z^i(ds), 1 \leq i \leq N$ , be  $\mathcal{F}_0$ -measurable point measures on  $[-\infty, 0]$  which we interpret as the initial condition of the age dependent Hawkes process. We then typically think of initial processes  $R_t^i$  of the form

$$R_t^i = \sum_{j=1}^N \int_{-\infty}^0 h_{ij}(t-s)z^j(ds),$$

provided the above expression is well-defined.

Well-posedness of the system (4) follows from

**Proposition 1.4.** *Almost surely, there is a unique sample path  $(Z, X, A)$  solving (4). Moreover,  $Z$  is non-exploding.*

**Proof.** Let  $h = \sum_{i,j=1}^N |h_{ij}|$  and  $R = \sum_{i=1}^N |R^i|$ . Assume first that  $R$  is bounded by some constant  $M > 0$ . Consider the linear Hawkes processes

$$\begin{aligned} \tilde{Z}_t^i &= \int_0^t \int_0^\infty \mathbb{1}\{z \leq L(1 + Y_s)\} \pi^i(ds, dz), \quad 1 \leq i \leq N, \\ Y_t &= \sum_{j=1}^N \int_0^{t-} h(t-s)\tilde{Z}^j(ds) + M. \end{aligned} \tag{5}$$

Notice that  $\tilde{Z}$  is driven by the same PRMs as  $Z$ . It is well known that the system (5) almost surely has a path-wise unique solution, which is defined for all  $t \in [0, \infty)$  (see for example [7], Theorem 6). By induction over jump times of  $\sum_{j=1}^N \tilde{Z}^j$  it follows that (4) has a unique solution satisfying  $Z_t^i \leq \tilde{Z}_t^i$  for all  $t \in \mathbb{R}_+$ , implying that  $Z$  does not explode.

In the general case define for each  $m \in \mathbb{N}$  the  $N$ -dimensional age dependent Hawkes processes  $(Z^m, X^m, A^m)$  with the same starting conditions and parameters as  $(Z, X, A)$  except for the initial processes, which are instead defined as  $R^{mi} = -m \vee R^i \wedge m$ . Write also  $\lambda_t^{m,i} = \psi^i(X_t^{m,i}, A_t^{m,i})$  for the associated intensity and  $\lambda_t^{m \wedge m+1, i} := \lambda_t^{m,i} \wedge \lambda_t^{m+1, i}$ . We have that

$$\begin{aligned} &P(Z_t^m \neq Z_t^{m+1}, t \in [0, T]) \\ &\leq P(\exists i : \int_0^\infty \int_0^T \mathbb{1}\{z \in (\lambda_s^{m \wedge m+1, i}, \lambda_s^{m \wedge m+1, i} + L|R_s^{m,i} - R_s^{m+1, i}|\})\pi^i(ds, dz) \geq 1) \\ &\leq L \sum_{i=1}^N \mathbb{E} \int_0^T |R_s^{m,i} - R_s^{m+1, i}| ds. \end{aligned}$$

Since  $\mathbb{E} \int_0^T |R_s^i| ds < \infty$ , we conclude that

$$\sum_m P(Z_t^m \neq Z_t^{m+1}, t \in [0, T]) < \infty,$$

implying that almost surely, the limit  $Z = \lim_{m \rightarrow \infty} Z^m$  exists. It is straightforward to show that  $Z$  solves (5). □

## 2. Stability

We start this section by discussing the stability of the age dependent Hawkes process within a sub-model where age acts as an inhibitor. For nonlinear Hawkes processes with no age dependence, a thorough investigation of invariant distributions and couplings was done in [2]. However, for results where boundedness is not forced upon the system (Theorem 1 of [2]), stability depends on the Lipschitz constant  $L$  of  $\psi$  in (2). As we show in Theorem 2.2 below, such restrictions are not necessary when age is incorporated as an inhibitor. The fact that the model has desirable stability properties is indicated by the result of Lemma 2.4, where we state a strong control of the intensity.

Throughout this section, the processes are defined on the entire real line  $\mathbb{R}$  unless otherwise mentioned. We do this for the following reason. When studying stability and thus the existence of stationary versions of infinite memory processes such as (age dependent) Hawkes processes, a widely used approach is to construct the process starting from  $t = -\infty$ . If such a construction is feasible, this implicitly implies that the state of the process at time  $t = 0$  must be in a stationary regime. Therefore, throughout this section we will work with random measures  $Z$  defined on the entire real line, with the usual identification of processes and random measures given by  $Z_t = Z((0, t])$ , for all  $t \geq 0$ , and  $Z_t = -Z((t, 0])$ , for all  $t < 0$ . We shall also use the shift operator  $\theta^r$  which is defined for any  $r \in \mathbb{R}$  by

$$\theta^r Z(C) := Z(r + C) := Z(\{r + x : x \in C\}), \tag{6}$$

for any  $C \in \mathcal{B}(\mathbb{R})$ .

*Setup in this section*

We consider a system with a fixed number of units  $N$ . Introduce the functions

$$\bar{h}_{ij}(t) = \sup_{s \geq t} |h_{ij}|(s), \quad h(t) = \sum_{i,j=1}^N |h_{ij}|.$$

In addition to the fundamental assumptions we add the following set of assumptions.

**Assumption 1.** 1. There exist  $K$  and  $\delta > 0$  such that

$$\psi^i(x, a) \leq K \quad \text{for all } 1 \leq i \leq N, a \in [0, \delta], x \in \mathbb{R}. \tag{7}$$

2. There exist  $x^*, a^*, c > 0$  such that for all  $|x| \leq x^*, a \geq a^*$  and for all  $1 \leq i \leq N$ ,

$$\psi^i(x, a) \geq c > 0. \tag{8}$$

3. We suppose that

$$[0, \infty[\ni t \mapsto \bar{h}_{ij}(t) \in \mathcal{L}^1 \cap \mathcal{L}^2 \quad \text{and} \quad [0, \infty[\ni t \mapsto th_{ij}(t) \in \mathcal{L}^1. \tag{9}$$

Notice that (9) implies that

$$\bar{h} := \sum_{i,j=1}^N \bar{h}_{ij} \in \mathcal{L}^1 \cap \mathcal{L}^2 \tag{10}$$

is a decreasing function that dominates  $h_{ij}$  for all  $i, j \leq N$ .

**Remark 2.** The existence of  $K, \delta$  in (7) excludes instantaneous bursting by imposing a bound on the immediate post-jump intensity. Moreover the existence of  $x^*, a^*, c$  in (8) ensures that no unit will eventually stop spiking. A main example of rate functions that satisfy this assumption are those inducing absolute refractory periods as given in Example 1.2.

The assumption  $\bar{h} \in \mathcal{L}^1$  is natural, at least in the context of modeling interacting neurons. To obtain stability, it is usually assumed that the weight functions are integrable. Here we impose the slightly stronger assumption that  $\bar{h}_{ij} \in \mathcal{L}^1$ ; that is, there exists a decreasing integrable function dominating  $h_{ij}$ .

Throughout this section we use the following notation. For  $K > 0$  as in (7) above, we denote the PRMs

$$\pi_K(ds) := \pi(ds, [0, K]), \quad \pi_K^i(ds) := \pi^i(ds, [0, K]) \quad \text{and} \quad \pi_{NK} := \sum_{i=1}^N \pi_K^i. \tag{11}$$

**Example 2.1 (Hawkes processes with Erlang weight functions).** Weight functions given by Erlang kernels are widely used in the modeling literature to describe delay in the information transmission. They are given by

$$h_{ij}(t) = c_{ij} t^{n_{ij}} e^{-v_{ij} t}, \quad t \geq 0,$$

where  $c_{ij} \in \mathbb{R}$ ,  $v_{ij} > 0$  and  $n_{ij} \in \mathbb{N} \cup \{0\}$  are fixed constants. The order of the delay is given by  $n_{ij}$ . The delay of the influence of particle  $j$  on particle  $i$  is distributed and taking its maximum absolute value at  $n_{ij}/v_{ij}$  time units back in time. The sign of  $c_{ij}$  indicates if the influence is inhibitory or excitatory, and the absolute value of  $c_{ij}$  scales how strong the influence is. All  $h_{ij}$  clearly satisfy (9).

The main result of this chapter shows existence of a unique stationary  $N$ -dimensional age dependent Hawkes process following the dynamics of (4). In order to state the result, we first introduce the notion of *compatibility* (see e.g. [2]). Let  $M_{\mathbb{R}_-}$  be the set of all bounded measures defined on  $\mathbb{R}_-$  equipped with the weak-hat metric and the associated Borel  $\sigma$ -algebra  $\mathcal{M}_{\mathbb{R}_-}$  (see the Appendix for details). We shall say that  $Z$  is *compatible* (to  $\pi^1, \dots, \pi^N$ ) if there is a measurable map  $H : M_{\mathbb{R}_-}^N \rightarrow M_{\mathbb{R}_-}$  such that for all  $t \in \mathbb{R}$ ,

$$(\theta^t Z)_{|\mathbb{R}_-} = H((\theta^t \pi^1, \dots, \theta^t \pi^N)_{|\mathbb{R}_-}) \tag{12}$$

Likewise, we say that a stochastic process  $X$  is compatible, if  $X_t = H((\theta^t \pi^1, \dots, \theta^t \pi^N)_{|\mathbb{R}_-})$  for an appropriate measurable mapping  $H$ .

**Remark 3.** Note that if  $Z^1, \dots, Z^n$  are compatible random measures, then  $(Z^1, \dots, Z^n)$  is a stationary and ergodic  $n$ -tuple of random measures.

Let  $Z = (Z^i)$ ,  $1 \leq i \leq N$ , be compatible random measures on  $\mathbb{R}$ . Let  $X = (X^i)_{i \leq N}$ ,  $A = (A^i)_{i \leq N}$  be compatible processes defined on  $t \in \mathbb{R}$  such that  $A_t^i$  is adapted and càglàd and  $X_t^i$  is predictable for all  $1 \leq i \leq N$ . We say that  $Z$  is an  $N$ -dimensional age dependent Hawkes process on  $t \in \mathbb{R}$ , if almost surely

$$\begin{aligned} Z^i(t_1, t_2] &= \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}\{z \leq \psi^i(X_s^i, A_s^i)\} \pi^i(ds, dz), \\ X_t^i &= \sum_{j=1}^N \int_{-\infty}^{t^-} h_{ij}(t-s) Z^j(ds) \quad t \in \mathbb{R}, \\ A_{t_2}^i - A_{t_1}^i &= t_2 - t_1 - \int_{t_1}^{t_2^-} A_t^i Z^i(dt), \end{aligned} \tag{13}$$

for all  $-\infty < t_1 \leq t_2$ .

**Theorem 2.2.** *Grant Assumption 1.*

1. *There exists an  $N$ -dimensional age dependent Hawkes process  $Z$  on  $\mathbb{R}$ , compatible to  $(\pi^1, \dots, \pi^N)$ .*
2. *Let  $\check{Z}$  be another  $N$ -dimensional age dependent Hawkes process with the same weight functions  $(h_{ij})_{i,j \leq N}$  and driven by the same PRMs  $(\pi^1, \dots, \pi^N)$ , following the dynamics (4), that is, starting at time 0 with arbitrary initial ages  $(A_0^i)_{i \leq N}$  and initial signals  $(R^i)_{i \leq N}$  such that*

$$\mathbb{E} \int_0^\infty |R_s^i| ds < \infty \tag{14}$$

*for all  $1 \leq i \leq N$ . Then almost surely,  $\check{Z}$  and  $Z$  couple eventually, i.e.,*

$$\exists t_0 \in \mathbb{R}_+ : \check{Z}_{|[t_0, \infty)} = Z_{|[t_0, \infty)}.$$

3. *If  $Z'$  is another  $N$ -dimensional age dependent Hawkes process on  $\mathbb{R}$ , compatible to  $(\pi^1, \dots, \pi^N)$ , then  $Z = Z'$  almost surely.*

The proof of the above theorem will be given in the next subsection. An immediate corollary of it is an ergodic theorem for additive functionals of age dependent Hawkes processes depending only on a finite time horizon. More precisely, let  $T > 0$  be a fixed time horizon and let  $M_T$  be the set of all bounded measures defined on  $] - T, 0]$ , equipped with its Borel  $\sigma$ -algebra  $\mathcal{M}_T$  (see the Appendix).

**Corollary 2.3.** *Grant Assumption 1. Let  $(Z, X, A)$  be the stationary age dependent Hawkes process and let  $\check{Z}$  be as in Item 2. of Theorem 2.2. Let  $f : M_T \rightarrow \mathbb{R}$  be any measurable function such that*

$$\mu(f) := \mathbb{E}f((Z_{|]-T,0]}) < \infty. \tag{15}$$

Then

$$\frac{1}{t} \int_0^t f((\theta^s \check{Z}_{|]-T,0]}) ds = \frac{1}{t} \int_0^t f((\check{Z}_{|]s-T,s]}) ds \rightarrow \mu(f) \tag{16}$$

almost surely, as  $t \rightarrow \infty$ .

**Proof.** By ergodicity of  $(Z, X, A)$ , it holds that

$$\frac{1}{t} \int_0^t f((\theta^s Z_{|]-T,0]}) ds = \frac{1}{t} \int_0^t f((Z_{|]s-T,s]}) ds \rightarrow \mu(f).$$

Since  $\check{Z}_{|[t_0,\infty)} = Z_{|[t_0,\infty)}$ , we have that

$$f((\theta^s \check{Z}_{|]-T,0]}) = f((\theta^s Z_{|]-T,0]}) \quad \text{for all } s \geq t_0 + T,$$

which implies (16). □

2.1. Proof of Theorem 2.2

This section is devoted to the proof of Theorem 2.2, and we will thus work under Assumption 1.

The proof of the existence part relies on the Picard iteration

$$X_t^{n,i} = \sum_{j=1}^N \int_{-\infty}^{t-} h_{ij}(t-s) Z^{n-1,j}(ds),$$

$$Z^{n,i}(t_1, t_2] = \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}\{z \leq \psi^i(X_s^{n,i}, A_s^{n,i})\} \pi^i(ds, dz), \quad t_1 < t_2 \in \mathbb{R}, \tag{17}$$

where  $A^{n,j}$  is the age process of  $Z^{n,j}$ . The Picard iteration is similar to the one found in [2], but since the system has an age variable and the intensity is not necessarily bounded, the following issues must be addressed before proving convergence.

- We need to produce an integrable intensity  $\hat{\lambda}$  that a priori dominates the intensities  $\psi^i(X_t^{n,i}, A_t^{n,i})$ . This is done in Proposition 2.6.
- Using the Lipschitz part of (2), we will construct events  $E_t \in \mathcal{F}_t$  for all  $t \in \mathbb{R}$  such that  $A_t^{n,i} = A_t^{n+1,i}$  on  $E_t$ , for all  $n \in \mathbb{N}, i \leq N$ . This is done in Lemma 2.7.
- We need to ensure that the  $i$ th iteration is well defined, i.e. for a given  $X^{n,i}$  there exist  $Z^{n,i}, A^{n,i}$  such that (17) is satisfied. This is done in Lemma 2.8.

Finally we combine these results to complete the proof of Theorem 2.2. We start with the following useful result which provides bounds on the intensities.

**Lemma 2.4.** *Let  $K, \delta$  be the constants from (7). Let  $X$  be a predictable stochastic process and assume that  $(Z, A)$  solves the system*

$$Z(t_1, t_2] = \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}\{z \leq \psi(X_s, A_s)\} \pi(ds, dz), \quad t_1 \leq t_2 \in \mathbb{R},$$

where  $A$  is the age process of  $Z$  and where  $\psi$  satisfies (2) and (7). Suppose moreover that (9) is satisfied. Then almost surely, for any  $1 \leq i, j \leq N, t_1 \leq t_2$ ,

$$Y_{ij}(t_1, t_2) = \int_{-\infty}^{t_1-} h_{ij}(t_2-s) Z(ds)$$



is well-defined and

$$|Y_{ij}(t_1, t_2)| \leq \sum_{k=0}^{\infty} \bar{h}_{ij}(t_2 - t_1 + A_{t_1} + k\delta) + \int_{-\infty}^{t_1 - A_{t_1}} \bar{h}_{ij}(t_2 - s)\pi_K(ds). \tag{18}$$

Moreover,

$$\mathbb{E} \int_{-\infty}^t \bar{h}_{ij}(t - s)\pi_K(ds) < \infty$$

for all  $t$ .

**Corollary 2.5.** *If we suppose in addition that  $\bar{h}(0) < \infty$ , then*

$$\mathbb{E}Y_{ij}(t, t) \leq K \int_0^{\infty} \bar{h}(u) du + \sum_{k \geq 0} \bar{h}(k\delta) < \infty. \tag{19}$$

**Proof of Lemma 2.4.** For any  $i, j \leq N, t \leq t_2$  we have

$$\begin{aligned} |Y_{ij}(t, t_2)| &\leq \int_{-\infty}^{t^-} \bar{h}_{ij}(t_2 - s)\mathbb{1}\{A_s \geq \delta\}Z(ds) + \int_{-\infty}^{t^-} \bar{h}_{ij}(t_2 - s)\mathbb{1}\{A_s < \delta\}Z(ds) \\ &= \int_{-\infty}^{t - A_t} \bar{h}_{ij}(t_2 - s)\mathbb{1}\{A_s \geq \delta\}Z(ds) + \int_{-\infty}^{t - A_t} \bar{h}_{ij}(t_2 - s)\mathbb{1}\{A_s < \delta\}Z(ds) \\ &\leq \int_{-\infty}^{t - A_t} \bar{h}_{ij}(t_2 - s)G(ds) + \int_{-\infty}^{t - A_t} \bar{h}_{ij}(t_2 - s)\pi_K(ds) \\ &:= \hat{Y}_{ij}(t, t_2) + \tilde{Y}_{ij}(t, t_2), \end{aligned}$$

where  $G(dt) = \mathbb{1}\{A_t \geq \delta\}Z(dt)$ . Define now for fixed  $t \in \mathbb{R}$  and for all  $l \in \mathbb{N}$ ,  $\tau_l(t) := \sup\{s < \tau_{l-1}(t) : \Delta G_s = 1\}$ , where we have put  $\tau_0(t) := t - A_t$ . Thus,  $\tau_l(t)$  is the  $l$ th jump-time of  $G$  before  $t - A_t$  – which is itself the last jump-time of  $Z$  strictly before time  $t$ .

We may upper bound  $\hat{Y}_{ij}$  by

$$\hat{Y}_{ij}(t, t_2) \leq \sum_{l=0}^{G(-\infty, t)} \bar{h}_{ij}(t_2 - \tau_l(t)).$$

Since  $\tau_l(t) - \tau_{l-1}(t) \geq \delta$  by construction of  $G$  and since  $\bar{h}_{ij}$  is decreasing, we get the bound

$$\hat{Y}_{ij}(t, t_2) \leq \sum_{l=0}^{\infty} \bar{h}_{ij}(t_2 - t + A_t + l\delta).$$

Note that almost surely,  $A_t$  never attains the value 0 for any  $t \in \mathbb{R}$ , and in that event, each term in the above sum is finite for all  $t \leq t_2 \in \mathbb{R}$ . Moreover, since  $\bar{h}_{ij}$  is  $\mathcal{L}^1$  and decreasing the sum is finite as well. The expectation  $t \mapsto \mathbb{E}\tilde{Y}(t, t)$  is given by

$$\begin{aligned} \mathbb{E}\tilde{Y}_{ij}(t, t) &= \lim_{T \rightarrow \infty} \mathbb{E} \int_{-T}^{t - A_t} \bar{h}_{ij}(t - s)\pi_K(ds) \\ &\leq \lim_{T \rightarrow \infty} \mathbb{E} \int_{-T}^t \bar{h}_{ij}(t - s)\pi_K(ds) = K \int_0^{\infty} \bar{h}_{ij}(u) du < \infty. \end{aligned} \quad \square$$

We now construct the dominating intensity, as mentioned in the start of the section. Recall that  $L \geq 1$  is the Lipschitz constant appearing in (2) and let  $K, \delta$  be the constants from (7), we suppose w.l.o.g. that  $K \geq c$ , where  $c$  is the lower bound from (8).

**Proposition 2.6.** *Let  $C \geq \max\{1 + \sum_{k \geq 1} \bar{h}(k\delta), K\}$ . There exists a compatible process  $(\hat{Z}, \hat{A}, \hat{\lambda})$  which is defined for any  $t \in \mathbb{R}$  by*

$$\hat{\lambda}_t = L \left( C + \int_{-\infty}^{t-} \bar{h}(t-s) \pi_{NK}(ds) + \bar{h}(\hat{A}_t) \right), \tag{20}$$

where

$$\hat{Z}(t_1, t_2] = \sum_{i=1}^N \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}_{\{z \leq \hat{\lambda}_s\}} \pi^i(ds, dz) \tag{21}$$

for all  $t_1 \leq t_2$ , together with its age process  $\hat{A}_t$ . Moreover, we have that

$$\mathbb{E}(\hat{\lambda}_t) < \infty. \tag{22}$$

**Proof.** By construction,  $\hat{\lambda}_t \geq K$  for all  $t$ , and therefore, any jump time  $\tau$  of  $\pi_K^i$  is also a jump of  $\hat{Z}$ . Hence, at  $\tau$ , the age process  $\hat{A}_t$  is reset to 0. It is therefore possible to construct a unique solution to (20) on  $t \in (\tau, \infty)$ . This solution is non-exploding since the process is stochastically dominated by a classical linear Hawkes process  $Z'$  having intensity  $L(C + \int_{-\infty}^{t-} \bar{h}(t-s) \pi_{NK}(ds) + 2 \int_{\tau}^{t-} \bar{h}(t-s) Z'(ds))$  which is non-exploding by Proposition 1.4 since  $\bar{h} \in \mathcal{L}^1$ . A solution on the entire real line may be constructed by pasting together the solutions constructed in between the successive jump times of  $\pi_{NK}$ . It is unique and compatible by construction.

It remains to prove that  $\mathbb{E}(\hat{\lambda}_t) < \infty$ . Due to stationarity, it is sufficient to prove that  $\mathbb{E}\bar{h}(\hat{A}_0) < \infty$ . Also from stationarity, writing  $T_1$  for the first jump time of  $\hat{Z}$  after time 0, it follows that  $\mathcal{L}(T_1) = \mathcal{L}(\hat{A}_0)$ . It follows from Lemma A.3 in the Appendix that

$$P(T_1 > t) = P(Z[0, t] = 0) = \mathbb{E} \left( \exp \left( - \int_0^t (LC + L\xi_s + L\bar{h}(\hat{A}_0 + s)) ds \right) \right),$$

where  $\xi_s := \int_{-\infty}^{0-} \bar{h}(s-u) \pi_{NK}(du)$ , implying that, since  $\hat{A}_0 \geq 0$  and  $\bar{h}$  is decreasing,

$$\begin{aligned} \mathbb{E}(\bar{h}(\hat{A}_0)) &= \int_0^\infty \mathbb{E}(\bar{h}(t) e^{-\int_0^t (LC + L\xi_s + L\bar{h}(\hat{A}_0 + s)) ds} (LC + L\xi_t + L\bar{h}(\hat{A}_0 + t))) dt \\ &\leq L \int_0^\infty (\bar{h}(t)^2 + \bar{h}(t) \mathbb{E}(\xi_0) + \bar{h}(t) C) dt < \infty, \end{aligned}$$

since  $\bar{h} \in \mathcal{L}^1 \cap \mathcal{L}^2$ . □

We now proceed to the construction of events  $E_t$  which a priori will serve by coupling the age processes in the Picard iteration. Indeed, Assumption (8) will enable us to construct common jumps for any two point processes  $Z^1, Z^2$  having intensity  $\psi(\tilde{X}_t^1, A_t^1)$  and  $\psi(\tilde{X}_t^2, A_t^2)$ , where  $A_t^i$  is the age process of  $Z^i$ , and  $\tilde{X}_t^i$  is a predictable process such that

$$\psi(X_t^i, A_t^i) \leq \hat{\lambda}_t,$$

for  $i = 1, 2$ .

Fix some  $p > a^*$  such that

$$\sum_{k \geq 1} L\bar{h}(p + k\delta) < \frac{x^*}{3N}, \tag{23}$$

where  $a^*$  and  $x^*$  are given in (8), and fix some  $M > LC$  where  $L$  and  $C$  are as in (20). Then necessarily  $M \geq K \geq c$ . Introduce for all  $t \in \mathbb{R}$  the events

$$\begin{aligned}
 E_t^1 &:= \{ \pi^1(ds, [0, c]) \text{ has a unique jump } \tau^1 \text{ in } (t - 2Np + p, t - 2Np + 2p) \} \\
 &\cap \bigcap_{j=1}^N \left\{ \int_{t-2Np}^{\tau^1-} \int_{\mathbb{R}_+} \mathbb{1}\{z \leq M\} \pi^j(ds, dz) = 0 \right\} \\
 &\cap \bigcap_{j=1}^N \left\{ \int_{\tau^1}^{t-2Np+2p} \int_{\mathbb{R}_+} \mathbb{1}\{z \leq M + 2L\bar{h}(s - \tau^1)\} \pi^j(ds, dz) = 0 \right\}, \tag{24}
 \end{aligned}$$

and for all  $i = 2, \dots, N$ ,

$$\begin{aligned}
 E_t^i &:= \{ \pi^i(ds, [0, c]) \text{ has a unique jump } \tau^i \text{ in } (t - 2Np + 2(i - 1)p + p, t - 2Np + 2ip) \} \\
 &\cap \bigcap_{j=1}^N \left\{ \int_{t-2Np+2(i-1)p}^{\tau^i-} \int_{\mathbb{R}_+} \mathbb{1}\{z \leq M + 2L\bar{h}(s - (t - 2Np + 2(i - 1)p))\} \pi^j(ds, dz) = 0 \right\} \\
 &\cap \bigcap_{j=1}^N \left\{ \int_{\tau^i}^{t-2Np+2ip} \int_{\mathbb{R}_+} \mathbb{1}\{z \leq M + 2L\bar{h}(s - \tau^i)\} \pi^j(ds, dz) = 0 \right\}, \tag{25}
 \end{aligned}$$

where the constant  $c$  is given in (8). This event splits the interval  $(t - 2Np, t)$  up in intervals of length  $2p$ , where the  $i$ th truncated PRM has exactly one jump in the second part, and no other events (of truncated PRMs) occur.

To control the past up to time  $t - 2Np$ , we also introduce the event

$$E_t^0 := \{ \hat{\lambda}_{t-2Np} + x^* \leq M \} \cap \left\{ \int_{-\infty}^{t-2Np} \bar{h}(t - 2Np - s) \pi_{NK}(ds) \leq \frac{x^*}{3N} \right\}$$

and put

$$E_t := \bigcap_{i=0}^N E_t^i. \tag{26}$$

The event  $E_{2Np}$  is illustrated in Figure 1, for  $N = 2$  and  $\bar{h}(t) \approx t^{-0.4}$ . The grey area is the relevant part for the truncated PRMs.

The main feature of the event  $E_{2Np}$  with  $N = 2$  is the fact that the process is forced to have some regeneration events during the intervals  $[p, 2p]$  and  $[3p, 4p]$ . Indeed, on these intervals, the corresponding age processes will have values larger than  $p > a^*$ , and the associated memory processes will be bounded by  $x^*$ , such that we can use (8).

Let us return to the general definition of the events  $E_t$ . Using induction and the strong Markov property, it follows from integrability of  $\bar{h}$  that  $P(\bigcap_{i=0}^j E_t^i) > 0$  for all  $t \in \mathbb{R}$ ,  $j \leq N$ . In particular  $P(E_t) > 0$ . Let us define

$$Y_t := \int_{-\infty}^{t-\hat{A}_t} \bar{h}(t - s) \pi_{NK}(ds) + \sum_{k=0}^{\infty} \bar{h}(\hat{A}_t + k\delta). \tag{27}$$

We summarize the most important features of the event  $E_t$  in the next lemma.

**Lemma 2.7.** *On  $E_t$ , for all  $1 \leq i \leq N$ , each measure  $\pi^i(ds, [0, c])$  has a jump at time  $\tau^i \in (t - 2Np, t)$  such that*

$$\int_{t-2Np}^{\tau^i} \int_{\mathbb{R}_+} \mathbb{1}\{s \neq \tau^i, z \leq \hat{\lambda}_s\} \pi^i(ds, dz) = 0, \tag{28}$$

$$\hat{\lambda}_{\tau^i} \leq \hat{\lambda}_{t-2Np} + \frac{2i}{3N} x^*, \tag{29}$$

$$Y_{\tau^i} \leq \frac{2i}{3N} x^*. \tag{30}$$

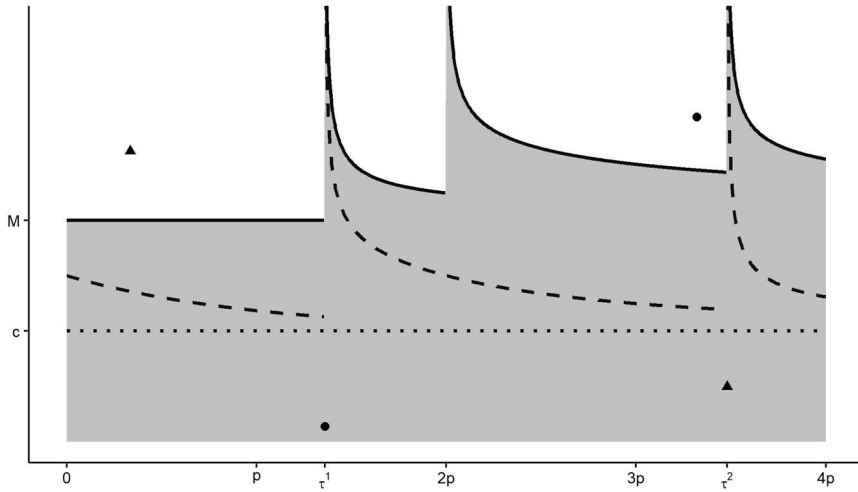


Fig. 1. An illustration of the event  $E_{2Np}$  with  $N = 2$  and  $\bar{h}(t) \approx t^{-0.4}$ . The figure shows a superposition of  $\pi^1(\bullet)$ ,  $\pi^2(\blacktriangle)$ , including the jump times  $\tau^1$  and  $\tau^2$ , and three curves: The dotted curve is the constant  $c$ . The dashed curve is the intensity process  $\hat{\lambda}$ . The solid curve is enclosing the area (in grey) of the plane that is relevant to the event  $E_{2Np}$ , and it is given by  $M$  for  $0 < s \leq \tau^1$ ,  $M + 2L\bar{h}(s - \tau^1)$  for  $\tau^1 < s \leq 2p$ ,  $M + 2L\bar{h}(s - 2p)$  for  $2p < s \leq \tau^2$  and  $M + 2L\bar{h}(s - \tau^2)$  for  $\tau^2 < s \leq 4p$ .

Moreover,  $|\tau^i - \tau^{i-1}| \geq a^*$ , where we put  $\tau^0 = t - 2Np$ . In particular, take any two point processes  $Z^1, Z^2$  having intensity  $\psi(\tilde{X}_t^1, A_t^1)$  and  $\psi(\tilde{X}_t^2, A_t^2)$ , where  $A_t^i$  is the age process of  $Z^i$ , and  $\tilde{X}_t^i$  is a predictable process such that

$$\psi(X_t^i, A_t^i) \leq \hat{\lambda}_t,$$

for  $i = 1, 2$ . It holds that  $A_t^1 = A_t^2$  under the event  $E_t$ .

**Proof.** Let  $(\tau^i)_{i \leq N}$  be the jump times as given in the definition of  $E_t$ . By construction, the inter-distances are at least equal to  $p$  and thus strictly larger than  $a^*$ , since we chose  $p > a^*$ . We shall prove by induction over  $j$  that

$$\int_{t-2Np}^{t-2Np+2jp} \int_{\mathbb{R}_+} \mathbb{1}\{s \neq \tau^i, z \leq \hat{\lambda}_s\} \pi^i(ds, dz) = 0 \quad \forall i \leq N$$

as well as (29) and (30) hold for  $i \in \{0, \dots, j\}$  in the event  $E_t$ . The induction start is trivial, so assume that the assertion is true up to  $j - 1$ . Notice that by the induction assumption

$$\hat{\lambda}_s \leq \hat{\lambda}_{t-2Np} + \frac{2(j-1)}{3N} x^* + 2L\bar{h}(s - \tau^{j-1}),$$

for  $s \geq \tau^{j-1}$  and until the next jump of  $\hat{Z}$ . It follows from the construction of  $E^j \cap E^{j-1}$  that  $\hat{Z}(\tau^{j-1}, \tau^j) = 0$ . This proves the first claim. It also shows that  $\hat{A}_{\tau^j} > p$  so the properties of  $p$  gives  $2L\bar{h}(\tau^j - \tau^{j-1}) \leq \frac{2x^*}{3N}$  which implies the remaining claims.  $\square$

The next result ensures that for a well-behaving process  $X^i$  there exist couples  $(Z^i, A^i)$  such that  $Z^i$  has intensity  $\psi^i(X^i, A^i)$  and  $A^i$  is the age of  $Z^i$ . The proof relies on a Picard iteration of (13) that alternately updates  $(X^i)_{i \leq N}$  and  $(Z^i, A^i)_{i \leq N}$ .

**Lemma 2.8.** Let  $(\hat{Z}, \hat{A}, \hat{\lambda})$  be as in Proposition 2.6 and let  $(X_t^i)_{t \in \mathbb{R}}, 1 \leq i \leq N$ , be compatible and predictable stochastic processes satisfying that almost surely,

$$|X_t^i| \leq Y_t, \tag{31}$$

for all  $1 \leq i \leq N, t \in \mathbb{R}$ .

Then there exist random counting measures  $Z^i, 1 \leq i \leq N$ , on  $\mathbb{R}$  which are compatible, and compatible càglàd processes  $A^i, 1 \leq i \leq N$ , which almost surely satisfy

$$Z^i(B) = \int_B \int_0^\infty \mathbb{1}\{z \leq \psi^i(X_s^i, A_s^i)\} \pi^i(ds, dz), \quad \forall B \in \mathcal{B}(\mathbb{R}), \tag{32}$$

for all  $1 \leq i \leq N$ , where  $A^i$  is the age process of  $Z^i$ .

**Proof.** The proof relies on Picard iteration. For that sake, define recursively for all  $n \geq 1$ , for all  $1 \leq i \leq N$ ,

$$Z^{n,i}(t_1, t_2] = \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}\{z \leq \psi^i(X_s^i, A_s^{n-1,i})\} \pi^i(ds, dz), \quad t_1 < t_2 \in \mathbb{R}, \tag{33}$$

where  $A^{n-1,i}$  is the age process corresponding to  $Z^{n-1,i}$ . We initialize the iteration with  $Z^{0,i} \equiv \pi_K^i$ .

We start by proving inductively over  $n$  that the Picard iteration is well-posed, and  $Z^n$  is non-exploding and compatible.

The induction start is trivial. We assume that the hypothesis holds for  $n - 1$ . Clearly  $Z^n$  is compatible. Moreover,  $Z^{n,i}$  has intensity  $\psi^i(X_t^i, A_t^{n-1,i}) \leq L(1 + Y_t)$ , and  $\mathbb{E}Y_t < \infty$  implying that  $Z^{n,i}$  does not explode.

We will now prove the convergence of the above scheme. To do so, define measures  $\underline{Z}^i$  and  $\overline{Z}^i$  by

$$\underline{Z}^i[t] = \liminf_{n \rightarrow \infty} Z^{ni}[t], \quad \overline{Z}^i[t] = \limsup_{n \rightarrow \infty} Z^{ni}[t]$$

for any  $t \in \mathbb{R}$ , and

$$\tilde{Z} = \sum_{i=1}^N (\overline{Z}^i - \underline{Z}^i).$$

That is,  $\tilde{Z}$  counts the sum of the differences of the superior and inferior limit processes. We claim that  $\tilde{Z}$  is almost surely the trivial measure. It will follow that  $Z^{n,i}$ , and thus also  $A^{n,i}$  converge.

To prove this claim, consider the event

$$G_t = \{\tilde{Z}(t, \infty) = 0\}.$$

Notice that  $\{\tilde{Z}(t, \infty) = 0\} = \{\theta^t(\pi^i)_{i=1}^N \in V\}$  for some  $V \in \mathcal{M}$  and thus  $\{\tilde{Z}(t, \infty) = 0$  infinitely often $\}$  is an invariant set, and thus also a 0/1 event. It follows by standard arguments that  $P(\tilde{Z}(\mathbb{R}) = 0) = 1$  if  $P(G_0) > 0$ .

We now prove that  $P(G_0) > 0$  by showing that  $E_0 \subset G_0$ , where  $E_0$  was defined in (26) above (that is, we choose  $t = 0$ ).

The assumption  $|X_t^i| \leq Y_t$  implies that  $\lambda_t^{n,i} \leq \hat{\lambda}_t$  for all  $i, n$  and  $t$ . Lemma 2.7 implies that on  $E_t$  we have  $\hat{A}_{\tau^i} \geq a^*$ , and therefore also  $A_{\tau^i}^{n,i} \geq a^*$ . Moreover, (30) implies that  $|X_{\tau^i}^i| \leq x^*$ . Therefore, (8) implies

$$\lambda_{\tau^i}^{n,i} \geq c \tag{34}$$

for all  $n, i$ . As a consequence, at time  $\tau^i$ , all  $Z^{ni}$  have a common jump. From (28) it follows that  $Z^{ni}(\tau^i, 0) = 0$ , and therefore,  $A_0^{n,i} = -\tau^i$ . In particular, they are all equal. We may now conclude that on  $E_0$ ,  $Z_{\mathbb{R}_+}^{n,i}$  is a constant sequence over  $n$ , for all  $i$ . In particular, we have  $\tilde{Z}(0, \infty) = 0$ . To conclude the proof, we have proven that  $E_0 \subset G_0$ , and thus

$$P(G_0 \cap E_0) = P(E_0) > 0,$$

implying the result. □

We are now ready to prove Theorem 2.2.

**Proof of Theorem 2.2.** First we construct a stationary solution to (13). For this sake we consider the Picard iteration

$$X_t^{n,i} = \sum_{j=1}^N \int_{-\infty}^{t-} h_{ij}(t-s) Z^{n-1,j}(ds),$$

$$Z^{n,i}(t_1, t_2) = \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}\{z \leq \psi^i(X_s^{n,i}, A_s^{n,i})\} \pi^i(ds, dz), \quad t_1 < t_2 \in \mathbb{R},$$

where  $A^{n,j}$  is the age process of  $Z^{n,j}$ . We initialize the iteration with  $Z^{0,i} \equiv \pi_K^i$ ,  $X^0 \equiv 0$ .

We start by proving inductively over  $n$  that the Picard iteration is well-posed,  $Z^{n,i}$  is non-exploding and compatible, and almost surely

$$L(1 + |X_t^{n,i}|) \leq \hat{\lambda}_t \quad \forall t \in \mathbb{R}, \tag{35}$$

for all  $n, i$ , where  $\hat{\lambda}$  is defined in (20) above. The induction start is trivial. Suppose now that the assertion holds for  $n - 1$ . We apply Lemma 2.8 with  $X^i = X^{n,i}$ , and show that the conditions of this Lemma are met, then well-posedness, ergodicity and stationarity of  $Z^{n,i}, A^{n,i}$  follow.

Next, we prove the upper bound on  $X^{n,i}$ . By construction,

$$X_t^{n,i} = \sum_{j=1}^N \int_{-\infty}^{t-} h_{ij}(t-s) Z^{n-1,j}(ds) = \sum_{j=1}^N \int_{-\infty}^{t-A_t^{n-1,j}} h_{ij}(t-s) Z^{n-1,j}(ds).$$

We apply Lemma 2.4 to each of the  $N$  terms within the above sum and obtain

$$\int_{-\infty}^{t-A_t^{n-1,j}} h_{ij}(t-s) Z^{n-1,j}(ds) \leq \sum_{k \geq 0} \bar{h}_{ij}(A_t^{n-1,j} + k\delta) + \int_{-\infty}^{t-A_t^{n-1,j}} \bar{h}_{ij}(t-s) \pi_K^j(ds). \tag{36}$$

Since  $\hat{\lambda}_t \geq \psi^i(X_t^{n-1,i}, A_t^{n-1,i})$  for all  $i$ , it follows that  $\hat{A}_t \leq A_t^{n-1,i}$  for all  $i$ , implying that

$$|X_t^{n,i}| \leq \int_{-\infty}^{t-\hat{A}_t} \bar{h}(t-s) \pi_{NK}(ds) + \sum_{k \geq 0} \bar{h}(\hat{A}_t + k\delta),$$

which is (31). Finally, since  $A^{n-1}, Z^{n-1}$  are compatible, it is straight-forward to show that  $Z^n$  is compatible as well.

Define now

$$\underline{\lambda}_t^i = \liminf_{n \rightarrow \infty} \psi^i(X_t^{n,i}, A_t^{n,i}), \quad \bar{\lambda}_t^i = \limsup_{n \rightarrow \infty} \psi^i(X_t^{n,i}, A_t^{n,i}), \quad 1 \leq i \leq N.$$

Note that by (3) and (35),  $\underline{\lambda}_t^i \leq \bar{\lambda}_t^i \leq \limsup_{n \rightarrow \infty} L(1 + |X_t^{n,i}|) \leq \hat{\lambda}_t$ . So almost surely,  $\underline{\lambda}^i, \bar{\lambda}^i$  have finite sample paths. Note also that they are limits of predictable processes (see Lemma A.1 in the Appendix), and thus they are predictable as well. Define also

$$\tilde{Z}^i[t] = \limsup_{n \rightarrow \infty} Z^{n,i}[t] - \liminf_{n \rightarrow \infty} Z^{n,i}[t] = \pi^i(\{t\} \times (\underline{\lambda}_t^i, \bar{\lambda}_t^i]),$$

for  $i \leq N, t \in \mathbb{R}$ . That is,  $\tilde{Z}^i$  counts the difference of the superior and inferior limit process. We claim that

$$\tilde{Z} = \sum_{j=1}^N \tilde{Z}^j \tag{37}$$

is almost surely the zero-measure. It will follow that  $Z^{n,j}$ , and thus also  $X^{n,i}, A^{n,i}$  converge. Moreover, it is straight forward to check that the limit variables solve (13).

To prove this claim, note that we may also find measurable  $H^i : M_{\mathbb{R} \times \mathbb{R}^+} \rightarrow \mathbb{R}^2$  such that almost surely

$$H^i(\theta^t(\pi^i)_{i=1}^N) = (\underline{\lambda}_t^i, \bar{\lambda}_t^i), \quad \forall t \in \mathbb{R}.$$

Consider the events  $E_t$  defined in (26) above as well as

$$G_t = (\tilde{Z}(t, \infty) = 0).$$

Using the functionals obtained previously, it follows that  $\{\tilde{Z}(t, \infty) = 0\} = \{\theta^t(\pi^i)_{i=1}^N \in V\}$  for some  $V \in \mathcal{M}$  and thus  $\{\tilde{Z}(t, \infty) = 0 \text{ infinitely often}\}$  is an invariant set, and thus also a 0/1 event. As before this implies that  $P(\tilde{Z}(\mathbb{R}) = 0) = 1$  if  $P(G_0 \cap E_0) > 0$ .

To prove that  $P(G_0 \cap E_0) > 0$ , note that we have  $\lambda_t^{n,i} = \psi(X_t^{n,i}, A_t^{n,i}) \leq \hat{\lambda}_t$  and  $|X_t^{n,i}| \leq Y_t$ . Therefore, the same arguments as those exposed in the proof of Lemma 2.8, show that on  $E_0$ , we have  $A_0^{n,i} = A_0^{m,i}$ ; for all  $n, m$  and  $i$ , that is, the age variables are all equal at time 0.

Moreover, on  $G_0$ , either no jumps happen any more, or they happen conjointly, and so the Lipschitz criterion (2) ensures the bound

$$|\bar{\lambda}_t^i - \underline{\lambda}_t^i| \leq L \lim_{n \rightarrow \infty} \sup_{m, k \geq n} |X_t^m - X_t^k| \leq L \int_{-\infty}^{0-} h(t-s) \tilde{Z}(ds) := \bar{X}_t,$$

for all  $t \geq 0$ , which holds on  $G_0 \cap E_0$ . Therefore we may write

$$P(G_0 \cap E_0) \geq P\left(E_0 \cap \left\{ \sum_{j=1}^N \int_0^\infty \int_0^\infty \mathbb{1}\{z \in (\underline{\lambda}_s^j, \underline{\lambda}_s^j + \bar{X}_s]\} \pi^j(ds, dz) = 0 \right\}\right).$$

Note that Lemma A.3 in the Appendix reveals that the compensator of the integral-sum above is

$$t \mapsto N \int_0^t \bar{X}_s ds = NL \int_0^t \int_{-\infty}^{0-} h(s-u) \tilde{Z}(du) ds.$$

The same lemma gives an expression for  $P(\tilde{Z}(0, \infty) = 0 | \mathcal{F}_0)$ , and so implies the lower bound

$$P(G_0 \cap E_0) \geq \mathbb{E} \mathbb{1}\{E_0\} \exp\left(-NL \int_0^\infty \int_{-\infty}^{0-} h(s-u) \tilde{Z}(du) ds\right).$$

To prove that the right hand side is positive, it suffices to show that the double integral inside the exponential is almost surely finite. Notice that by construction,  $\tilde{Z}_t \leq \hat{Z}_t$ , and recall from Proposition 2.6 that  $\hat{\lambda}$  is stationary with  $\mathbb{E} \hat{\lambda}_0 < \infty$ . After taking expectation,

$$\begin{aligned} \mathbb{E} \int_0^\infty \int_{-\infty}^{0-} h(s-u) \tilde{Z}(du) ds &\leq \mathbb{E} \int_0^\infty \int_{-\infty}^{0-} h(s-u) \hat{Z}(du) ds \\ &= \mathbb{E} \int_0^\infty \int_{-\infty}^{0-} h(s-u) \hat{\lambda}_u du ds \\ &= \mathbb{E}(\hat{\lambda}_0) \int_0^\infty \int_{-\infty}^{0-} h(s-u) du ds \\ &= \mathbb{E}(\hat{\lambda}_0) \int_0^\infty th(t) dt < \infty. \end{aligned}$$

This proves the desired result.

We now prove the coupling part. This will be done in two steps. First suppose that  $|R_t|$  is bounded by a constant  $C_R$ . We suppose w.l.o.g. that  $\hat{\lambda}$  defined in (20) is such that also  $C \geq C_R$ .

Let  $((\check{Z}^i)_{i \leq N}, (\check{X}^i), (\check{A}^i)_{i \leq N})$  be the  $N$ -dimensional age dependent Hawkes process with initial conditions  $(\check{A}_0^i), (R^i)$  and denote  $\check{\lambda}_t^i := \psi(\check{X}_t^i, \check{A}_t^i)$ . Then we clearly have that

$$\check{\lambda}_0^i \leq \hat{\lambda}_0$$

for all  $i$ , and it can be shown inductively over the successive jumps of  $\hat{Z}$ , using Lemma 2.4, that this inequality is preserved over time, that is,

$$\check{\lambda}_t^i \leq \hat{\lambda}_t \tag{38}$$

for all  $t \geq 0$ . Now introduce

$$\check{E}_t^0 := \left\{ \sum_{i=1}^N \int_{t-2Np}^t \mathbb{1}\left\{z \leq \frac{3Nc}{x^*} |R_s|\right\} \pi^i(ds, dz) = 0 \right\}$$

and put

$$E'_t := E_t \cap \check{E}_t^0. \tag{39}$$

Let  $(\tau^i)_{i \leq N}$  be the jump times from Lemma 2.7. We necessarily have that on  $E'_t$ ,

$$|R_{\tau^i}| \leq \frac{x^*}{3N},$$

for all  $1 \leq i \leq N$ . As a consequence, (30) implies that

$$|\check{X}_{\tau^i}| \leq Y_{\tau^i} + |R_{\tau^i}| \leq x^*.$$

Due to (38), we have  $\check{A}_{\tau^i}^i \geq \hat{A}_{\tau^i} \geq a^*$ . Therefore, using (8), we conclude that

$$\check{\lambda}_{\tau^i}^i \geq c,$$

implying that  $\tau^i$  is also a jump of  $\check{Z}^i$ . Thus, on  $E'_t$ , at time  $t$ , all  $(\check{A}^i, A^i)$ ,  $1 \leq i \leq N$ , are coupled. Therefore, we have a Lipschitz bound under the event  $E'_t$ ; so with  $Z = \sum_{j=1}^N Z^j$ ,  $\check{Z} = \sum_{j=1}^N \check{Z}^j$ , we may write

$$\sum_{i=1}^N |\psi^i(X_t^i, A_t^i) - \psi^i(\check{X}_t^i, \check{A}_t^i)| \leq L \sum_{i=1}^N |X_t^i - \check{X}_t^i| \leq L \left( \int_{-\infty}^{0^-} h(t-s)Z(ds) + |R_t| + \int_0^{t^-} h(t-s)\check{Z}(ds) \right),$$

which holds on  $E'_t$ .

As before  $\check{Z} := |Z - \check{Z}|$  and  $G_t := \{\check{Z}(t, \infty) = 0\}$ . Equivalent considerations as in the first part of the proof yield

$$\begin{aligned} &P(G_t \cap E'_t | \mathcal{F}_t) \\ &\geq \mathbb{1}\{E'_t\} \exp\left(-L \int_t^\infty \left[ \int_{-\infty}^0 h(s-u)Z(du) + |R_s| + \int_0^{t^-} h(s-u)\check{Z}(du) \right] ds\right) \\ &\geq \mathbb{1}\{E'_t\} \exp\left(-L \int_t^\infty \left[ \int_{-\infty}^{t^-} h(s-u)Z(du) + |R_s| + \int_0^{t^-} h(s-u)\check{Z}(du) \right] ds\right). \end{aligned} \tag{40}$$

Since  $\check{\lambda}_t \leq \hat{\lambda}_t$  for all  $t \geq 0$ ,

$$\int_0^{t^-} h(s-u)\check{Z}(du) \leq \int_{-\infty}^{t^-} h(s-u)\hat{Z}(du).$$

As a consequence,

$$\begin{aligned} &\int_t^\infty \left[ \int_{-\infty}^{t^-} h(s-u)Z(du) + |R_s| + \int_0^{t^-} h(s-u)\check{Z}(du) \right] ds \\ &\leq 2 \int_t^\infty \left[ \int_{-\infty}^{t^-} h(s-u)\hat{Z}(du) \right] ds + \int_t^\infty |R_s| ds = C_t + D_t, \end{aligned}$$

where

$$C_t := 2 \int_t^\infty \left[ \int_{-\infty}^{t^-} h(s-u)\hat{Z}(du) \right] ds$$

is stationary and ergodic, and where

$$D_t := \int_t^\infty |R_s| ds.$$

Clearly,  $D_t \rightarrow 0$  as  $t \rightarrow \infty$  almost surely.



Now apply Lemma A.4 in the Appendix with  $U_t := \mathbb{1}_{E_t} e^{-LC_t}$ ,  $r_t := \mathbb{1}_{E_t} e^{-LC_t} - \mathbb{1}_{E'_t} e^{-LC_t - LD_t}$ . Clearly,  $U_t$  is ergodic and satisfies  $P(U_t > 0) > 0$ . To see that  $r_t \rightarrow 0$  almost surely as  $t \rightarrow \infty$ , it suffices to prove that  $\mathbb{1}_{E'_t} - \mathbb{1}_{E_t} \rightarrow 0$  almost surely, as  $t \rightarrow \infty$ , which is equivalent to proving that  $\mathbb{1}_{\check{E}_t^0} \rightarrow 1$  almost surely. But this follows from

$$\sum_{i=1}^N \mathbb{E} \int_0^\infty \int_{\mathbb{R}_+} \mathbb{1} \left\{ z \leq \frac{3N}{x^*} |R_s| \right\} \pi^i(ds, dz) < \infty,$$

which follows from (14).

This finishes the first part of the proof of the coupling result. Finally, suppose that the initial process only satisfies (14). Take then a sequence  $(\check{Z}^m, \check{X}^m, \check{A}^m)$  of  $N$ -dimensional age dependent Hawkes processes with starting condition  $(\check{A}_0^i)_{i \leq n}$  and initial processes  $(-m \vee R^i \wedge m)_{i \leq N}$ . Write  $\check{\lambda}_t^{m,i} = \psi^i(\check{X}_t^{m,i}, \check{A}_t^{m,i})$  for the associated intensity and  $\check{\lambda}_t^{m \wedge m+1,i} := \check{\lambda}_t^{m,i} \wedge \check{\lambda}_t^{m+1,i}$ . Denote  $R_t^{m,i} := -m \vee R_t^i \wedge m$ . As in the proof of Proposition 1.4, we have that

$$P(\check{Z}^m \neq \check{Z}^{m+1}) \leq L \sum_{i=1}^N \mathbb{E} \int_0^\infty |R_t^{m,i} - R_t^{m+1,i}| dt.$$

Since  $\mathbb{E} \int_0^\infty |R_t^i| dt < \infty$ , we conclude that

$$\sum_m P(\check{Z}^m \neq \check{Z}^{m+1}) < \infty,$$

implying that almost surely,  $\check{Z}^m = \check{Z}$  on  $\mathbb{R}_+$  for sufficiently large  $m$ . Since  $\check{Z}^m$  and  $Z$  couple eventually almost surely, this proves the coupling part. It remains to prove uniqueness of the stationary solution. Let  $(Z', X', A')$  be another age dependent Hawkes process on  $t \in \mathbb{R}$  compatible to  $\pi^1, \dots, \pi^N$ . Lemma 2.4 gives the inequality

$$|X_t^i| \leq \sum_{j=1}^N \left( \sum_{k \geq 0} \bar{h}_{ij}(A_t^j + k\delta) + \int_{-\infty}^{t-A_t^j} \bar{h}_{ij}(t-s) \pi_K^j(ds) \right).$$

Let  $\tau$  be a jump of  $\hat{Z}$ . Using the above inequality, it is shown inductively over future jumps of  $\hat{Z}$  that  $\hat{\lambda}_t \geq |\lambda^i|$  for all  $t \in \tau, \infty[$ . Thus it follows that  $\hat{\lambda}_t \geq |\lambda^i|$  for all  $t \in \mathbb{R}$ . Note that the  $(Z', X', A')$  system may be written in terms of Definition 1.1 with initial signals  $R_t^i := \sum_{j=1}^N \int_{-\infty}^0 h_{ij}(t-s) dZ_s^j$ . The same arguments as in (38) give

$$\mathbb{E} \int_0^\infty |R_s^i| ds \leq \mathbb{E} \int_0^\infty \int_{-\infty}^{0-} h(s-u) \hat{Z}(du) ds < \infty.$$

Therefore it follows from the 2nd point of this theorem that

$$P(\exists t_0 \in \mathbb{R} : Z'_{|t_0, \infty) = Z_{|t_0, \infty}) = P\left( \bigcup_{n=-\infty}^\infty (Z'_{|n, \infty) = Z_{|n, \infty}) \right) = 1. \tag{41}$$

Since  $Z$  and  $Z'$  are both compatible, it follows that  $(Z, Z')$  is compatible and therefore also stationary. Thus, the events  $\{Z'_{|n, \infty) = Z_{|n, \infty})\}$  have the same probability for all  $n \in \mathbb{Z}$ , and from (41) it follows that the probability is equal to 1. This proves that  $Z = Z'$  almost surely. □

### 2.2. Age dependent Hawkes processes with Erlang weight functions

Here we show how Theorem 2.2 can be applied for weight functions given by Erlang kernels as in Example 2.1, and consider a one-dimensional ( $N = 1$ ) age dependent Hawkes process  $(Z, X, A)$ , solution of

$$\begin{aligned} Z_t &= \int_0^t \int_0^\infty \mathbb{1} \{ z \leq \psi(X_s, A_s) \} \pi(ds, dz), \\ X_t &= \int_0^{t-} h(t-s) Z(ds) + R_t, \\ A_t &= A_0 + t - \int_0^{t-} A_s Z(ds), \end{aligned} \tag{42}$$

where

$$h(t) = b \frac{t^n}{n!} e^{-\nu t}, \tag{43}$$

for some  $b \in \mathbb{R}$ ,  $\nu > 0$  and  $n \geq 0$ , and where the initial signal is given by

$$R_t = \int_{-\infty}^0 h(t-s)z(ds),$$

for some fixed discrete point measure  $z$  defined on  $(-\infty, 0)$  such that  $\int_{-\infty}^0 h(t-s)z(ds)$  is well defined.

The process  $(X_t, A_t)$  is not Markov, but it is well-known (see e.g. [8]) that it can be completed to a Markovian system  $(X_t^{(0)} := X_t, X_t^{(1)}, \dots, X_t^{(n)}, A_t)$ , by introducing the auxiliary processes

$$X_t^{(k)} := \int_0^{t^-} b \frac{(t-s)^{n-k}}{(n-k)!} e^{-\nu(t-s)} Z(ds) + \int_{-\infty}^0 b \frac{(t-s)^{n-k}}{(n-k)!} e^{-\nu(t-s)} z(ds), \quad \text{for all } 0 \leq k \leq n.$$

By [8], these satisfy the system of coupled differential equations, driven by the PRM  $\pi$ , given by

$$dX_{t+}^{(k)} = -\nu X_{t+}^{(k)} dt + X_{t+}^{(k+1)} dt, \quad 0 \leq k < n, \tag{44}$$

$$dX_{t+}^{(n)} = -\nu X_{t+}^{(n)} dt + b \int_0^\infty \mathbb{1}\{z \leq \psi(X_t^{(0)}, A_t)\} \pi(dt, dz), \tag{45}$$

and

$$A_t - A_0 = t - \int_0^{t^-} \int_0^\infty A_s \mathbb{1}\{z \leq \psi(X_s, A_s)\} \pi(ds, dz) = t - \int_0^{t^-} A_s Z(ds),$$

for  $t \geq 0$ . Evidently,  $h$  satisfies (9). We suppose that  $\psi(x, a)$  satisfies (2) and we strengthen (8) to the following assumption.

**Assumption 2.**  $\psi(x, a)$  is continuous in  $x$  and  $a$ ; and  $\psi(x, a) \geq c > 0$  for all  $x, a$  with  $a \geq a^*$ .

Then the following result strengthens Theorem 2.2 in this Markovian setting.

**Theorem 2.9.** *Grant Assumptions 1 and 2. Then the process  $(X_t^{(0)}, X_t^{(1)}, \dots, X_t^{(n)}, A_t)$  is positively recurrent in the sense of Harris having unique invariant probability measure  $\mu$ .*

**Proof.** *Step 1.* By Lemma 2.4 and Corollary 2.5,  $t \mapsto \mathbb{E}(\lambda_t) = \mathbb{E}(\psi(X_t, A_t))$  and  $t \mapsto \mathbb{E}(|X_t|) = \mathbb{E}(|X_t^{(0)}|)$  are bounded on  $\mathbb{R}$ . By the same argument, also  $t \mapsto \mathbb{E}(|X_t^{(k)}|)$  is bounded for  $1 \leq k \leq n$ . Therefore,  $(X_t^{(0)}, X_t^{(1)}, \dots, X_t^{(n)})$  is a 1-ultimately bounded Feller process (the Feller property follows from the continuity of  $\psi$ ), see e.g. [17].

We write  $x = (x^0, \dots, x^n) \in \mathbb{R}^{n+1}$  for the elements of  $\mathbb{R}^{n+1}$  and denote by  $P_t((x, a), \cdot)$  the transition semigroup of  $(X_t^{(0)}, \dots, X_t^{(n)}, A_t)$ . Let  $B_k = \{(x, a) : |x| + |a| \leq k\}$ . Then for any  $x_0 \in \mathbb{R}^{n+1}$ ,  $a_0 \geq 0$ ,

$$P_t((x_0, a_0), B_k^c) \leq P_{(x_0, a_0)}\left(|X_t| \geq \frac{k}{2}\right) + P_{(x_0, a_0)}\left(A_t \geq \frac{k}{2}\right),$$

where

$$P_{(x_0, a_0)}\left(|X_t| \geq \frac{k}{2}\right) \leq \frac{2 \sup_t \mathbb{E}_{(x_0, a_0)}(|X_t|)}{k},$$

and

$$P_{(x_0, a_0)}\left(A_t \geq \frac{k}{2}\right) \leq e^{-c(\frac{k}{2} - a^*)_+},$$

implying inequality (6) of [17]. Thus, by Theorem 1 of [17],  $(X_t^{(0)}, \dots, X_t^{(n)}, A_t)_{t \geq 0}$  possesses invariant probability measures  $\mu$  (not necessarily unique ones).<sup>1</sup>

*Step 2.* We shall now use the coupling property proved in Theorem 2.2 to prove uniqueness of the invariant measure  $\mu$ . In what follows we write  $(Z, X, A)$  for the stationary version of (42), which exists according to Theorem 2.2. Moreover, we write  $(\tilde{Z}, \tilde{X}, \tilde{A})$  for a version of (42) starting at time  $t = 0$  from an arbitrary initial age  $a_0$  and an initial configuration  $x_0 = (x_0^{(0)}, \dots, x_0^{(n)})$  with  $x_0^{(k)} = b \int_{-\infty}^0 \frac{(-s)^{n-k}}{(n-k)!} e^{-\nu s} z(ds)$ . Write

$$\tau_c := \inf\{t > 0 : Z \text{ and } \tilde{Z} \text{ couple at time } t\} \vee 1.$$

Note that

$$|h(s + u)| \leq C |h(s)| |h(u)|$$

for all  $s, u \geq 1$ , where  $C$  is an appropriate constant. It follows that almost surely, for all  $t \geq \tau_c + 1$

$$|X_t - \tilde{X}_t| \leq \bar{h}(t - \tau_c)(Z([0, \tau_c]) + \tilde{Z}([0, \tau_c]))\tilde{Z}([0, \tau_c]) + C |h(t - \tau_c)| (|X_{\tau_c}| + |\tilde{X}_{\tau_c}|),$$

showing that

$$\lim_{t \rightarrow \infty} |X_t - \tilde{X}_t| = 0 \tag{46}$$

almost surely, since  $|h(t - \tau_c)| \rightarrow 0$  as  $t \rightarrow \infty$ . In the same way one proves that also

$$\lim_{t \rightarrow \infty} |X_t^{(k)} - \tilde{X}_t^{(k)}| = 0 \tag{47}$$

almost surely, for all  $1 \leq k \leq n$ . Moreover, we obviously have that  $\tilde{A} = A$  on  $[T_1 \circ \theta_{\tau_c}, \infty[$ , where  $T_1 \circ \theta_{\tau_c} = \inf\{t > \tau_c : Z([t]) = \tilde{Z}([t]) = 1\}$ . Since  $\psi(x, a) \geq c > 0$  for all  $a \geq a^*$ ,  $T_1 \circ \theta_{\tau_c} < \infty$  almost surely. This implies the uniqueness of the invariant measure.

*Step 3.* Finally, to prove the Harris recurrence of the process  $(X_t^{(0)}, \dots, X_t^{(n)}, A_t)$ , we rely on the following local Doebler lower bound. It states that for all  $(x^{**}, a^{**}) \in \mathbb{R}^{n+1} \times \mathbb{R}_+$ , there exist  $R > 0$ , an open set  $I \subset \mathbb{R}^{n+1} \times \mathbb{R}_+$  and a constant  $\beta \in ]0, 1[$ , such that for any  $T > (n + 2)a^*$ ,

$$P_T((x_0, a_0), \cdot) \geq \beta \mathbb{1}_C(x_0, a_0) U(\cdot), \tag{48}$$

where  $C = B_R(x^{**}, a^{**})$  is the (open) ball of radius  $R$  centered at  $(x^{**}, a^{**})$ , and where  $U$  is the uniform measure on  $I$ . This lower bound follows easily adapting the proof of Theorem 3 in [9] to our framework.

We may apply the above result with  $(x^{**}, a^{**}) \in \text{supp}(\mu)$  where  $\mu$  is the (unique) invariant measure of the process. Then for the stationary version of the process,  $(X_t^{(0)}, \dots, X_t^{(n)}, A_t) \in B_{R/2}(x^{**}, a^{**})$  infinitely often. Then (46) and (47) imply that also  $(\tilde{X}_t^{(0)}, \dots, \tilde{X}_t^{(n)}, \tilde{A}_t) \in B_R(x^{**}, a^{**}) = C$  infinitely often, almost surely. The classical regeneration technique, see e.g. [18], allows to conclude that indeed the process is positively recurrent in the sense of Harris.  $\square$

### 3. Mean-field limit and propagation of chaos

In this section we focus on a multi-class mean-field setup of the age dependent Hawkes process on  $\mathbb{R}_+$ , starting from initial signals and ages. We propose a limit system, and show how the high dimensional system couples with the limit system. This can be seen as a generalization of the work of Chevallier [3] where a single class is considered under the assumption that the spiking rate function  $\psi$  is uniformly bounded. The multi-class setup is similar to the one in [8] for ordinary Hawkes Processes, and to the data transmission model in [12]. We also discuss how to approximate a Hawkes process induced by one weight function, by another Hawkes process, induced by different weight functions.

<sup>1</sup>As a matter of fact, this provides a different approach to prove the existence of a stationary version of the age dependent Hawkes process.

Setup in this section

In addition to the fundamental assumptions, we introduce the following specifications for the mean-field setup, which will be used throughout this section. We partition the indices of individual units into  $\mathcal{K} \in \mathbb{N}$  different populations, where  $\mathcal{K} > 0$ . More precisely, for each fixed total population size  $N \in \mathbb{N}$ ,

$$N_k := N_k(N) := \#\{i \leq N : i \text{ in population } k\}$$

will denote the number of units belonging to population  $k$ ,  $1 \leq k \leq \mathcal{K}$ , and

$$N = N_1 + \dots + N_{\mathcal{K}}.$$

We assume that each population represents an asymptotic part of all units, i.e., there exists  $p_k > 0$  such that

$$\frac{N_k}{N} \xrightarrow{N \rightarrow \infty} p_k.$$

For a fixed  $N \in \mathbb{N}$ , we re-index the  $N$ -dimensional age dependent Hawkes process of (4) as

$$(Z^{kj})_{k \leq \mathcal{K}, j \leq N_k},$$

where the superscript  $kj$  denotes the  $j$ th unit in population  $k$ . The weight function from the  $i$ th unit of population  $l$  to the  $j$ th unit of population  $k$  is given by  $N^{-1}h_{kl}$ . Moreover, all units within the same population have the same spiking rate  $\psi^k$ . Finally, we assume that all units in population  $k$  have the same initial signal  $R^k$ , and that the initial ages are interchangeable in groups and mutually independent in and between groups. With this set of parameters, the age dependent Hawkes process  $(Z, X, A)$  from (4) defined on  $t \in \mathbb{R}_+$  becomes

$$\begin{aligned} Z_t^{kj} &= \int_0^t \int_0^\infty \mathbb{1}\{z \leq \psi^k(X_s^k, A_s^{kj})\} d\pi^{kj}(ds, dz), \quad j \leq N_k, k \leq \mathcal{K}, \\ X_t^k &= \frac{1}{N} \sum_{l=1}^{\mathcal{K}} \sum_{j=1}^{N_l} \int_0^{t-} h_{kl}(t-s) Z^{lj}(ds) + R_t^k, \quad k \leq \mathcal{K}, \end{aligned} \tag{49}$$

where  $A^{kj}$  is the age process of  $Z^{kj}$ , starting from  $A_0^{kj}$  at time  $t = 0$ . Sometimes, to explicitly indicate the dependency on  $N$ , we add  $N$  to the superscript and write  $Z_t^{Nki}$ ,  $X^{Nk}$  and  $A^{Nki}$ .

Model observations

1. Suppose that the initial ages  $(A_0^i)_{i \in \mathbb{N}}$  are exchangeable. Then the symmetry of the system gives interchangeability between units within the same population, i.e.,  $Z^{kj} \stackrel{\mathcal{L}}{=} Z^{ki}$  for  $i, j \leq N_k, k \leq \mathcal{K}$ .
2. In the mean-field setup, all units within a population  $k$  share the same memory process  $X^k$ .

3.1. The limit system

We propose a limit system for  $N \rightarrow \infty$ . To pursue this goal, take finite variation functions  $t \mapsto \alpha_t^k$ , locally bounded functions  $t \mapsto \beta_t^k$ , and PRMs  $\pi^k$  for  $k \leq \mathcal{K}$ , and consider the stochastic convolution equation

$$\begin{aligned} \phi_t^k &= \int_0^t \mathbb{E} \psi^k(x_s^k, A_s^k) ds, \\ x_t^k &= \sum_{l=1}^{\mathcal{K}} p_l \int_0^t h_{kl}(t-s) d\alpha_s^l + \beta_t^k, \\ A_t^k - A_0^k &= t - \int_0^{t-} \int_0^\infty A_s^k \mathbb{1}\{z \leq \psi^k(x_s^k, A_s^k)\} \pi^k(ds, dz), \end{aligned} \tag{50}$$

with unknown  $(\phi, x, A) = (\phi^k, x^k, A^k)_{k \leq \mathcal{K}}$ . Notice that only  $A$  is stochastic. Introducing

$$Z_t^k = \int_0^t \int_0^\infty \mathbb{1}\{z \leq \psi^k(x_s^k, A_s^k)\} \pi^k(ds, dz),$$

we can interpret  $A^k$  as age process of  $Z^k$ . Hence, in the limit, the network activity can be resumed via the deterministic quantities  $x^k$ ,  $1 \leq k \leq \mathcal{K}$ , the only remaining randomness is in the individual age processes. Finally, notice that  $\phi$  depends on the law of  $A$ .

We are motivated by what for the moment is a heuristic.

$$N^{-1} \sum_{j=1}^{N_k} Z^{Nkj} \approx p_k \mathbb{E} Z^k$$

for large  $N$ , where  $(Z^1, \dots, Z^{\mathcal{K}})$  denotes the limit process such that each  $Z^k$  describes the jump activity of a typical unit belonging to population  $k$ . This relation invites the idea that the memory process for  $N \rightarrow \infty$ ,  $t \mapsto x_t$  should satisfy the integral system (50) with  $\phi_t^k = \alpha_t^k = \mathbb{E} Z_t^k$  and  $\beta_t^k = \mathbb{E} R_t^{k1}$ . This motivates the following result.

**Lemma 3.1.** *Let  $\beta_t = (\beta_t^k)_{k \leq \mathcal{K}}$  be measurable and locally bounded. There is a unique function  $\alpha$  such that  $\alpha = \phi$ , where  $(\phi, x, A)$  is the solution to (50). Moreover,  $\phi$  is continuous and  $x$  is bounded on  $[0, T]$  by a constant  $C$  which depends on  $h := \sum_{k,j} |h_{kj}|$ ,  $\|\beta\|_T$ ,  $T$  and  $L$ .*

The proof is given in the [Appendix](#). Once this lemma is established, we can ensure existence of the limit process.

**Theorem 3.2.**

Let  $\beta_t = (\beta_t^k)_{k \leq \mathcal{K}}$  be measurable and locally bounded. There is a unique solution  $(Z, A)$  to the integral equation

$$Z_t^k = \int_0^t \int_0^\infty \mathbb{1} \left\{ z \leq \psi^k \left( \sum_{l=1}^{\mathcal{K}} p_l \int_0^s h_{kl}(s-u) d\mathbb{E} Z_u^l + \beta_s^k, A_s^k \right) \right\} \pi^k(ds, dz), \quad 1 \leq k \leq \mathcal{K},$$

where  $A^k$  is the age process corresponding to  $Z^k$ , initialized at  $A_0^k$ .

**Proof.** Let  $(\phi, x, A)_{k \leq \mathcal{K}}$  be the tuple given in Lemma 3.1. Define the counting process

$$Z_t^k := \int_0^t \int_0^\infty \mathbb{1} \{ z \leq \psi^k(x_s^k, A_s^k) \} \pi^k(ds, dz).$$

It is clear that  $A^k$  is the age process of  $Z^k$ , and since  $d\mathbb{E} Z_t^k = \mathbb{E} \psi^k(x_t^k, A_t^k) dt$ ,  $Z^k$  will satisfy the desired identity. For uniqueness, consider another solution  $(\tilde{Z}^k, \tilde{A}^k)_{k \leq \mathcal{K}}$ , which satisfies the same identity:

$$\tilde{Z}_t^k = \int_0^t \int_0^\infty \mathbb{1} \left\{ z \leq \psi^k \left( \sum_{l=1}^{\mathcal{K}} p_l \int_0^s h_{kl}(s-u) d\mathbb{E} \tilde{Z}_u^l + \beta_s^k, \tilde{A}_s^k \right) \right\} \pi^k(ds, dz).$$

Defining  $\tilde{x}_t^k = \sum_{l=1}^{\mathcal{K}} p_l \int_0^t h_{kl}(t-s) d\mathbb{E} \tilde{Z}_u^l + \beta_t^k$  and  $\tilde{\phi}_t^k = \int_0^t \mathbb{E} \psi^k(\tilde{x}_s^k, \tilde{A}_s^k) ds$ , we note that  $\mathbb{E} \tilde{Z}_t^k = \tilde{\phi}_t^k$ . Thus, if we insert  $\alpha = \tilde{\phi}$  in (50),  $(\tilde{\phi}, \tilde{x}, \tilde{A})$  is a solution, and hence the uniqueness part of Lemma 3.1 gives that  $(\tilde{\phi}, \tilde{x}, \tilde{A}) = (\phi, x, A)$  and thus also  $Z = \tilde{Z}$ .  $\square$

### 3.2. Large network asymptotics and weight approximations

In this section we couple the  $N$ -dimensional Hawkes process with the limit system proposed in the previous section. This coupling implies that the finite-dimensional system converges to the limit system. The result is traditionally named *Propagation of Chaos*, a typical result within mean-field theory. Specifically for Hawkes processes, there are several variants of this result. Some of the recent results may be found in [3,7] and in [8].

#### Framework for propagation of chaos

We will first introduce a set of assumptions.

**Assumption 3.** We are given, for each  $1 \leq k \leq \mathcal{K}$ , a sequence  $(R^{Nk})_{N \in \mathbb{N}}$  of initial signals with  $\sup_{k \leq \mathcal{K}, N \in \mathbb{N}} \|\mathbb{E}R^{Nk}\|_t < \infty$ , such that there is a locally bounded function  $t \mapsto r_t^k$  with

$$\int_0^t \mathbb{E}|R_s^{Nk} - r_s^k| ds \rightarrow 0 \quad \text{as } N \rightarrow \infty, \tag{51}$$

for all  $t \geq 0$ .

**Assumption 4.** The initial ages  $A_0^{ki}, 1 \leq k \leq \mathcal{K}, 1 \leq i < \infty$  are i.i.d.

**Assumption 5.** The weight functions  $h_{kl}^N : \mathbb{R}_+ \rightarrow \mathbb{R}$  satisfy  $h_{kl}^N \rightarrow h_{kl}$  as  $N \rightarrow \infty$  locally in  $\mathcal{L}^1$ , where  $h_{kl} \in \mathcal{L}_{\text{loc}}^2$  for all  $1 \leq k, l \leq \mathcal{K}$ .

Consider an i.i.d. sequence of driving PRMs  $\pi^{kj}, 1 \leq k \leq \mathcal{K}, j \geq 1$ . Define for each  $N \in \mathbb{N}$ , the  $N$ -dimensional Hawkes process

$$(Z^N, X^N, A^N) = (Z^{Nki}, X^{Nk}, A^{Nki})_{k \leq \mathcal{K}, i \leq Nk},$$

given by (49), driven by  $(\pi^{kj})$ , with weight functions  $(N^{-1}h_{kl}^N)$ , spiking rate  $(\psi^k)$  and initial processes  $(R^{Nk})$ .

Applying Theorem 3.2 with weight functions  $(h_{kl})$  and initial functions  $\beta^k = r^k$ , we obtain, for any  $1 \leq k \leq \mathcal{K}$  and for all  $i \in \mathbb{N}$ , a solution  $(Z^{ki}, X^k, A^{ki})$  to the equation

$$Z_t^{ki} = \int_0^t \int_0^\infty \mathbb{1} \left\{ z \leq \psi^k \left( \sum_{l=1}^{\mathcal{K}} p_l \int_0^s h_{kl}(s-u) d\mathbb{E}Z_u^{lj} + r_s^k, A_s^{ki} \right) \right\} \pi^{ki}(ds, dz),$$

$1 \leq k \leq \mathcal{K}, i \in \mathbb{N}$ , driven by the same sequence of PRMs.

**Theorem 3.3 (Propagation of Chaos).** Consider the framework described above and grant Assumptions 3–5. Then for all  $t \geq 0$ ,

$$\mathbb{E}|d(Z_t^{Nki} - Z_t^{ki})| \rightarrow 0, \quad \text{for } N \rightarrow \infty, \tag{52}$$

for all  $k \leq \mathcal{K}, i \in \mathbb{N}$ . In particular, for any finite set of indices  $(k_1, i_1, \dots, k_n, i_n)$ , we have weak convergence

$$(Z^{Nk_1 i_1}, \dots, Z^{Nk_n i_n})_{t \geq 0} \xrightarrow{wk} (Z^{k_1 i_1}, \dots, Z^{k_n i_n})_{t \geq 0}$$

as  $N \rightarrow \infty$  (in  $D(\mathbb{R}_+, \mathbb{R}_+^n)$ , endowed with the topology of locally-uniform convergence).

To prove this theorem we shall need the following lemma.

**Lemma 3.4.** Let  $(h_{kl})_{1 \leq k, l \leq \mathcal{K}}, (\tilde{h}_{kl})_{1 \leq k, l \leq \mathcal{K}}$  be sets included in a family  $\mathcal{E}$  of real-valued functions defined on  $\mathbb{R}_+$  which is uniformly integrable on  $[0, T]$ . Define  $(Z, X, A), (\tilde{Z}, \tilde{X}, \tilde{A})$  as the  $N$ -dimensional age dependent Hawkes process with weight functions  $(N^{-1}h_{kl})_{1 \leq k, l \leq \mathcal{K}}, (N^{-1}\tilde{h}_{kl})_{1 \leq k, l \leq \mathcal{K}}$ , rate functions  $(\psi^k)_{k \leq \mathcal{K}}$ , and with initial conditions  $A_0, R^k$ . There exists  $C > 0$  depending on the family  $\mathcal{E}$ , on  $T, L, \mathcal{K}$  and on  $\sup_{k \leq \mathcal{K}} \|\mathbb{E}R^k\|_T$  (but not on  $N$ ) such that

$$\sum_{k=1}^{\mathcal{K}} \mathbb{E}|d(Z_t^{k1} - \tilde{Z}_t^{k1})| \leq C_T \sum_{k, l=1}^{\mathcal{K}} \int_0^t |h_{kl} - \tilde{h}_{kl}|(s) ds,$$

for all  $t \leq T$ .

The proofs of Theorem 3.3 and Lemma 3.4 may be found in the Appendix.

**Remark 4.** The result shows that finitely many units will be asymptotically independent for  $N \rightarrow \infty$ .

3.3. *The mean-field limit in the case of a hard refractory period*

In this section we consider the mean-field limit of age dependent Hawkes processes with one single population ( $\mathcal{K} = 1$ ) and a weight function given by an Erlang kernel as in Example 2.1, that is,

$$h(t) = be^{-\nu t} \frac{t^n}{n!},$$

for some fixed constants  $b \in \mathbb{R}$ ,  $\nu > 0$ ,  $n \in \mathbb{N}$ . Throughout this section we suppose a hard refractory period of length  $\delta$  after a jump where no new jumps can occur as given in the following assumption.

**Assumption 6.**

$$\psi(x, a) = f(x)\mathbb{1}\{a \geq \delta\}.$$

We start by rewriting the limit system in this frame. Recall that

$$\phi_t = \int_0^t E(\psi(x_s, A_s)) ds = \int_0^t \bar{\lambda}_s ds, \tag{53}$$

where

$$\bar{\lambda}_t = E(\psi(x_t^{(0)}, A_t))$$

denotes the expected number of jumps up to time  $t$  of a typical unit in the limit system. As in Section 2.2 above, we write  $x^{(0)} := x$ , and we add auxiliary variables  $x^{(i)}$ ,  $1 \leq i \leq n$  to obtain the system

$$A_t = A_0 + t - \int_0^t \int_{\mathbb{R}_+} A_s \mathbb{1}\{z \leq \psi(x_s^{(0)}, A_s)\} \pi(ds, dz)$$

together with

$$\begin{aligned} dx_t^{(0)} &= -\nu x_t^{(0)} dt + x_t^{(1)} dt, \\ &\vdots \\ dx_t^{(n-1)} &= -\nu x_t^{(n-1)} dt + x_t^{(n)} dt, \\ dx_t^{(n)} &= -\nu x_t^{(n)} dt + bd\phi_t = -\nu x_t^{(n)} dt + b\bar{\lambda}_t dt. \end{aligned} \tag{54}$$

Let us now study the age process of this limit system. Write  $\tau_t = \sup\{0 \leq s \leq t : \Delta A_s \neq 0\}$  for the last jump time of the process before time  $t$ , where by convention,  $\sup \emptyset := 0$ . Then obviously,

$$A_{t+} = (t - \tau_t)\mathbb{1}\{\tau_t > 0\} + (A_0 + t)\mathbb{1}\{\tau_t = 0\}.$$

Due to Assumption 6, we have the following

**Proposition 3.5.**

$$\mathcal{L}(\tau_t)(dz) = \mathbb{E}(e^{-\int_0^t f(x_s^{(0)})\mathbb{1}\{A_0+s \geq \delta\}} ds) \delta_0(dz) + f(x_z^{(0)}) p_z e^{-\int_z^t f(x_s^{(0)}) ds} \mathbb{1}\{0 < z < t\} dz,$$

where  $p_t = P(A_t \geq \delta)$  is given by

$$\begin{aligned} p_t &= \mathbb{E}(\mathbb{1}\{A_0 \geq \delta - t\} e^{-\int_{(\delta-A_0) \vee 0}^t f(x_s^{(0)}) ds}) + \int_0^{t-\delta} f(x_s^{(0)}) p_s e^{-\int_{s+\delta}^t f(x_u^{(0)}) du} ds \\ &= \int_{(\delta-t) \vee 0}^\infty \mu_0(da) e^{-\int_{(\delta-a) \vee 0}^t f(x_s^{(0)}) ds} + \int_0^{t-\delta} f(x_s^{(0)}) p_s e^{-\int_{s+\delta}^t f(x_u^{(0)}) du} ds, \end{aligned}$$

where  $A_0 \sim \mu_0(da)$ .

In particular, the above representation shows that, even starting from a non-smooth initial trajectory,  $p_t$  is eventually smooth.

**Corollary 3.6.** For any starting law  $\mu_0(da)$ ,  $t \mapsto p_t$  is continuous on  $]\delta, \infty]$ , and thus, taking into account (54),  $C^1(]2\delta, \infty[, \mathbb{R})$ , solving

$$dp_t = -f(x_t^{(0)})p_t dt + f(x_{t-\delta}^{(0)})p_{t-\delta} dt, \quad \text{for all } t > 2\delta.$$

If the starting law is smooth, we can say more.

**Corollary 3.7.** If  $\mu_0(da) = \mu_0(a) da$ , with  $\mu_0 \in C(\mathbb{R}, \mathbb{R}_+)$ , then for all  $t < \delta$ ,

$$p_t = \int_{\delta-t}^{\infty} \mu_0(a) e^{-\int_{(\delta-a) \vee 0}^t f(x_s^{(0)}) ds} da$$

is continuous and thus, taking into account (54),  $C^1([0, \delta[, \mathbb{R})$ . In particular, on  $[0, \delta[, t \mapsto p_t$  solves

$$dp_t = \mu_0(\delta - t) - f(x_t^{(0)})p_t dt.$$

By induction, this implies that  $t \mapsto p_t$  is continuous on  $\mathbb{R}_+$  and  $C^1$  on  $]\delta, \infty[,$  with

$$dp_t = -f(x_t^{(0)})p_t dt + f(x_{t-\delta}^{(0)})p_{t-\delta} dt, \quad \text{for all } t > \delta.$$

Moreover, at  $t = \delta$ , writing  $\dot{p} := \frac{d}{dt} p_t$ ,

$$\dot{p}_{\delta-} = \mu_0(0) - f(x_{\delta}^{(0)})p_{\delta} \quad \text{and} \quad \dot{p}_{\delta+} = -f(x_{\delta}^{(0)})p_{\delta} + f(x_0^{(0)})p_0.$$

Let us now look for possible stationary solutions of (54). At equilibrium, we necessarily have that

$$x^{(0)} \equiv x^{**}$$

for a given value  $x^{**} \in \mathbb{R}$ . It follows that  $A_t$  is a renewal process with dynamics

$$dA_{t+} = dt - A_t \int_{\mathbb{R}_+} 1_{\{z \leq \psi(x^{**}, A_t)\}} \pi(dt, dz). \tag{55}$$

$A$  is recurrent in the sense of Harris if it comes back to 0 infinitely often almost surely. This happens if  $\int_0^\infty \psi(x^{**}, A_t) dt = \infty$  almost surely, which is granted by the following condition.

**Assumption 7.** For all  $x$ , there exists  $r(x) \geq 0$  such that  $\psi(x, a)$  is lower bounded for all  $a \geq r(x)$ .

The stationary distribution of (55) is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}_+$ , having the density (see Proposition 21 of [10])

$$g_{x^{**}}(a) = \kappa e^{-\int_0^a \psi(x^{**}, z) dz}$$

on  $\mathbb{R}_+$ , where  $\kappa$  is chosen such that  $\int_0^\infty g_{x^{**}}(a) da = 1$ . Recall that  $\bar{\lambda}_t = \frac{d\phi_t}{dt}$  denotes the (expected) jump rate of the limit system at time  $t$ . Then at equilibrium, the total jump rate is constant and given by  $\bar{\lambda}_t = \bar{\lambda}$ . From (53) we get that

$$\bar{\lambda} = \kappa \int_0^\infty \psi(x^{**}, a) e^{-\int_0^a \psi(x^{**}, z) dz} da = \kappa,$$

where we have used the change of variables  $y = \int_0^a \psi(x^{**}, z) dz$ ,  $dy = \psi(x^{**}, a) da$ .

As a consequence,

$$\bar{\lambda} = \kappa$$



implying that at equilibrium, the jump rate of the system is solution of

$$\bar{\lambda}^{-1} = \int_0^\infty \exp\left(-\int_0^a \psi\left(\frac{b}{\nu^{n+1}}\bar{\lambda}, z\right) dz\right) da. \quad (56)$$

Here we have used that at equilibrium

$$x^{**} = x^{(0)} = \frac{1}{\nu}x^{(1)} = \dots = \frac{b}{\nu^{n+1}}\bar{\lambda},$$

which follows from (54).

**Proposition 3.8.** *Suppose that  $x \mapsto \psi(x, a)$  is strictly increasing for any fixed  $a \geq 0$  and that  $b < 0$ . There exists a unique solution  $\lambda^*$  to (56).*

Recall that we suppose that  $\psi(x, a) = f(x)\mathbb{1}\{a \geq \delta\}$ , for some  $\delta > 0$ . We calculate the right hand side of (56) and obtain the fixed point equation

$$\int_0^\infty \exp\left(-\int_0^a \psi\left(\frac{b}{\nu^{n+1}}\bar{\lambda}, z\right) dz\right) da = \delta + \left(f\left(\frac{b}{\nu^{n+1}}\bar{\lambda}\right)\right)^{-1} = \bar{\lambda}^{-1}. \quad (57)$$

More generally, for any Hawkes process with mean-field interactions, rate function  $\psi(x, a)$  given by  $\psi(x, a) = f(x)\mathbb{1}\{a \geq \delta\}$  and general weight function  $h \in \mathcal{L}^1(\mathbb{R}_+)$ , we obtain the fixed point equation

$$\bar{\lambda}^{-1} = \delta + \left(f\left(\bar{\lambda} \int_0^\infty h(t) dt\right)\right)^{-1} \quad (58)$$

for the limit intensity. This limit intensity depends on the length of the refractory period, we write  $\bar{\lambda} = \bar{\lambda}(\delta)$  to indicate this dependence.

It is then natural to study the influence of the length of the refractory period  $\delta$  on the limit intensity. If  $f$  is increasing and the system inhibitory, that is,  $\int_0^\infty h(t) dt < 0$ , then clearly

$$\delta \mapsto \bar{\lambda}(\delta)$$

is decreasing: increasing the length of the refractory period ‘‘calms down the system’’.

In the excitatory case when  $\int_0^\infty h(t) dt > 0$ , to ensure that the fixed point equation (56) has a solution, suppose that  $f$  is strictly increasing and bounded from above and below, away from zero. Then the function

$$\bar{\lambda} \mapsto \left(f\left(\bar{\lambda} \int_0^\infty h(t) dt\right)\right)^{-1}$$

is a strictly decreasing function mapping  $[0, \infty]$  onto  $[\frac{1}{f(0)}, \frac{1}{f(\infty)}]$ . Therefore, there is exactly one fixed point solution of (56), and  $\delta \mapsto \bar{\lambda}(\delta)$  is again decreasing.

## Appendix

Here we prove Lemma 3.1, Theorem 3.3 and Lemma 3.4. Then we collect some useful results about counting processes.

### A.1. Proofs

**Proof of Lemma 3.1.** It suffices to show that a unique solution exists on  $[0, T]$ , for arbitrary  $T \geq 0$ . In the following proof,  $C := C_T$  will denote a dynamic constant depending on the parameters described in the lemma. It need not represent the same constant from line to line, nor from equation to equation.

First we prove existence of a solution to (50) with  $\phi_t = \alpha_t$  using Picard-iteration. For  $n \in \mathbb{N}$  define  $(\phi^n, x^n, A^n) = (\phi^{n,k}, x^{n,k}, A^{n,k})_{k \leq \mathcal{K}}$  as follows. Initialize the system for  $n = 0$  by putting  $(\phi^{0,k}, x^{0,k}, A^{0,k}) \equiv (0, 0, A_0)$ . For general

$n \in \mathbb{N}$ ,  $n \geq 1$ , the triple  $(\phi^n, x^n, A^n)$  is defined as the solution to (50) with  $\alpha = \phi^{n-1}$ . Inductively it is seen that these processes are well-defined. Recall that  $h = \sum_{k,l=1}^{\mathcal{K}} |h_{kl}|$ . Using (3) we bound  $x^n$  by

$$|x_t^n| \leq \sum_{l=1}^{\mathcal{K}} \int_0^t |h(t-s)| d\phi_s^{n-1,l} + |\beta_t| \leq C \int_0^t |h(s)| ds + C \int_0^t |h(t-s)| |x_s^{n-1}| ds + |\beta_t|.$$

It follows from Lemma A.2 in the Appendix that there exists a constant  $C > 0$  which bounds all  $\|x^n\|_T$ ,  $n \in \mathbb{N}$ . Using this upper bound on  $x^n$ , we also bound the difference of two consecutive solutions. Define

$$\delta_t^n = \sum_{k=1}^{\mathcal{K}} \int_0^t \mathbb{E} |\psi^k(x_s^{n,k}, A_s^{n,k}) - \psi^k(x_s^{n-1,k}, A_s^{n-1,k})| ds.$$

The Lipschitz property of  $\psi$  and the bound on  $x^n$  yield

$$\delta_t^{n+1} \leq C \int_0^t \left( |x_s^{n+1} - x_s^n| + \sum_{k=1}^{\mathcal{K}} P(\|A^{n+1,k} - A^{n,k}\|_s > 0) \right) ds.$$

For the probability term, we note that a necessity for the age processes to differ, is that one of their corresponding intensities catches a  $\pi$ -singularity which the other one does not catch. This leads to the inequality

$$\sum_{k=1}^{\mathcal{K}} P(\|A^{n+1,k} - A^{n,k}\|_t > 0) \tag{59}$$

$$\begin{aligned} &\leq \sum_{k=1}^{\mathcal{K}} P\left(\int_0^t \int_0^\infty |\mathbb{1}\{z \leq \psi^k(x_s^{n+1,k}, A_s^{n+1,k})\} - \mathbb{1}\{z \leq \psi^k(x_s^{n,k}, A_s^{n,k})\}| \pi^k(ds, dz) \geq 1\right) \\ &\leq \delta_t^{n+1}, \end{aligned} \tag{60}$$

where the latter inequality follows by the Markov inequality. By Gronwall’s inequality we obtain

$$\delta_t^{n+1} \leq C \int_0^t |x_s^{n+1} - x_s^n| ds. \tag{61}$$

Moreover, Lemma 22 of [7] gives

$$\int_0^t |x_s^{n+1} - x_s^n| ds \leq \sum_{l=1}^{\mathcal{K}} \int_0^t \int_0^s h(s-u) |d(\phi_u^{n,l} - \phi_u^{n-1,l})| ds \leq \int_0^t h(t-s) \delta_s^n ds. \tag{62}$$

It therefore follows from Lemma A.2 in the Appendix that for all  $1 \leq k \leq \mathcal{K}$ ,

$$\sum_{n=1}^\infty \sup_{t \leq T} |\phi_t^{n+1,k} - \phi_t^{n,k}| \leq \sum_{n=1}^\infty \delta_T^n < \infty.$$

Thus,  $\phi^n$  and therefore also  $x^n$  converge locally-uniformly to some  $\phi, x$ , respectively. Moreover,

$$P(A_{s \leq T}^n \neq A_{s \leq T}^{n+1} i.o.) = P\left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \{\|A^n - A^{n+1}\|_T > 0\}\right) \leq \lim_{m \rightarrow \infty} \sum_{n \geq m} \delta_T^n = 0.$$

It follows that almost surely,  $A^n$  converges to some limit  $A$  after finitely many iterations.

We need to show that the limit triple  $(\phi, x, A)$  satisfies (50) with  $\phi_t = \alpha_t$ . Recall that  $x \mapsto \psi^k(x, a)$  is continuous for fixed  $a \in \mathbb{R}_+$ . Since  $A^n$  reaches its limit in finitely many iterations, and  $\psi$  is continuous in  $x$  for fixed  $a$ , it follows that  $\lim_{n \rightarrow \infty} \psi^k(x_s^{n,k}, A_s^{n,k})$  exists for all  $s \leq T$ , almost surely. By dominated convergence and (3), it follows that  $\mathbb{E} \psi^k(x_s^{n,k}, A_s^{n,k})$  converges as well. Therefore, once again by dominated convergence,

$$\phi_t^k = \lim_{n \rightarrow \infty} \phi_t^{n,k} = \lim_{n \rightarrow \infty} \int_0^t \mathbb{E} \psi^k(x_s^{n,k}, A_s^{n,k}) ds = \int_0^t \mathbb{E} \psi^k(x_s^k, A_s^k) ds,$$

that is,  $\phi$  satisfies (50). One shows similarly that

$$\begin{aligned} x_t^k &= \sum_{l=1}^{\mathcal{K}} \lim_{n \rightarrow \infty} \int_0^t h_{kl}(t-s) d\phi_s^{n,l} + \beta_t^k \\ &= \sum_{l=1}^{\mathcal{K}} \lim_{n \rightarrow \infty} \int_0^t h_{kl}(t-s) \mathbb{E} \psi^l(x_s^{n,l}, A_s^{n,l}) ds + \beta_t^k \\ &= \sum_{l=1}^{\mathcal{K}} \int_0^t h_{kl}(t-s) \mathbb{E} \psi^l(x_s^l, A_s^l) ds + \beta_t^k \\ &= \sum_{l=1}^{\mathcal{K}} \int_0^t h_{kl}(t-s) d\phi_s^l + \beta_t^k, \end{aligned}$$

and  $x$  satisfies (50) as well. For the age process, notice that the càglàd process

$$\varepsilon(t) = \sum_{k=1}^{\mathcal{K}} \int_0^{t-} \int_0^\infty \mathbb{1}\{z = \psi^k(x_s^k, A_s^k)\} \pi^k(ds, dz)$$

has a compensator which is equal to zero for all  $t \geq 0$ , almost surely, by Lemma A.3. Therefore,  $\varepsilon_t = 0$  for all  $t \geq 0$  almost surely. This implies that with probability 1,  $\mathbb{1}\{z \leq \psi^k(x_s^{nk}, A_s^k)\}$  converges  $\pi^k$ -a.e. to  $\mathbb{1}\{z \leq \psi^k(x_s^k, A_s^k)\}$  for all  $k \leq \mathcal{K}$ . As a consequence,

$$\begin{aligned} A_t^k - A_0^k &= t - \lim_{n \rightarrow \infty} \int_0^{t-} A_s^{nk} \mathbb{1}\{z \leq \psi^k(x_s^{n,k}, A_s^{nk})\} \pi^k(ds, dz) \\ &= t - \lim_{n \rightarrow \infty} \int_0^{t-} A_s^k \mathbb{1}\{z \leq \psi^k(x_s^{n,k}, A_s^k)\} \pi^k(ds, dz) \\ &= t - \int_0^{t-} \int_0^\infty A_s^k \mathbb{1}\{z \leq \psi^k(x_s^k, A_s^k)\} \pi^k(ds, dz), \end{aligned}$$

where we have used dominated convergence. Since  $x$  is locally bounded, it follows that  $\phi$  is  $C^0$ .

To prove uniqueness, we assume that  $(\tilde{\phi}, \tilde{x}, \tilde{A})$  also solves (50) with

$$\tilde{x}_t^k = \sum_{l=1}^{\mathcal{K}} p_l \int_0^t h_{kl}(t-s) d\tilde{\phi}_s^l + \beta_t^k.$$

Define

$$\delta_t = \sum_{k=1}^{\mathcal{K}} \int_0^t \mathbb{E} |\psi^k(x_s^k, A_s^k) - \psi^k(\tilde{x}_s^k, \tilde{A}_s^k)| ds.$$

Considerations analogous to the ones given in the proof of existence, gives that

$$|x_t - \tilde{x}_t| \leq \delta_t \leq C \int_0^t h(t-s) \delta_s ds.$$

From Gronwall's inequality it follows that  $\delta \equiv 0$ , and therefore also that  $x = \tilde{x}$  on  $[0, T]$ . From (50) it follows immediately  $\phi = \tilde{\phi}$  and  $A = \tilde{A}$  almost surely. □

**Proof of Lemma 3.4.** Throughout this proof,  $C$  is a dynamic constant with dependencies as declared in the theorem. Define the functions  $h = \sum_{k,l} |h_{kl}|$ ,  $\tilde{h} = \sum_{k,l} |\tilde{h}_{kl}|$ . First we prove that the memory processes  $\mathbb{E}|X_t|$ ,  $\mathbb{E}|\tilde{X}_t|$  are bounded

on  $[0, T]$  by a suitable constant  $C$ . Note that

$$\begin{aligned} \mathbb{E}|X_t| &\leq \sum_{l=1}^{\mathcal{K}} \left( \int_0^t h(t-s) \mathbb{E} \psi^l(X_s^l, A_s^l) ds + \mathbb{E}|R_t^l| \right) \\ &\leq C \int_0^t h(t-s) \mathbb{E}|X_s| ds + C \int_0^t h(s) ds + \mathbb{E}|R_t|. \end{aligned}$$

Since  $\mathcal{E}$  is uniformly integrable, the direct sum  $\{\sum_{k,l=1}^{\mathcal{K}} |f_{kl}|, f_{kl} \in \mathcal{E}\}$  is uniformly integrable as well. Thus, there exists  $b > 0$  satisfying

$$\int_0^T \sum_{k,l=1}^{\mathcal{K}} |f_{kl}|(s) \mathbb{1} \left\{ \sum_{k,l=1}^{\mathcal{K}} |f_{kl}|(s) > b \right\} ds < 2^{-1} \tag{63}$$

for all choices of  $(f_{kl}) \subset \mathcal{E}$ . It follows from Lemma A.2 that  $\mathbb{E}\|X\|_T \leq C$  for a suitable  $C$ . The same argument shows that also  $\mathbb{E}\|\tilde{X}\|_T \leq C$ . Define the total variation measure  $\delta_t = \sum_{k=1}^{\mathcal{K}} \mathbb{E}|d(Z_t^{k1} - \tilde{Z}_t^{k1})|$ . We may write

$$\begin{aligned} \delta_t &\leq \mathbb{E} \sum_{k=1}^{\mathcal{K}} \int_0^t |\psi^k(\tilde{X}_s^k, \tilde{A}_s^{k1}) - \psi^k(X_s^k, A_s^{k1})| ds \\ &\leq C \sum_{k=1}^{\mathcal{K}} \int_0^t \mathbb{E}|\tilde{X}_s^k - X_s^k| + P(\|\tilde{A}^{k1} - A^{k1}\|_s > 0) ds. \end{aligned}$$

As in the proof of Lemma 3.1 we apply Markov's inequality to achieve

$$\sum_{k=1}^{\mathcal{K}} P(\|\tilde{A}^{k1} - A^{k1}\|_t > 0) \leq \delta_t^n.$$

We insert this inequality into (64) to get

$$\delta_t \leq C \left( \int_0^t \mathbb{E}|\tilde{X}_s - X_s| ds + \int_0^t \delta_s ds \right). \tag{64}$$

We now wish to bound the difference of the memory processes. First, define  $\gamma = \sum_{k,l=1}^{\mathcal{K}} |h_{kl} - \tilde{h}_{kl}|$ , and note that for any fixed  $k, l, j$  we have for any  $s \geq 0$

$$\begin{aligned} &\left| \int_0^{s^-} h_{kl}(s-u) dZ_u^{lj} - \int_0^{s^-} \tilde{h}_{kl}(s-u) d\tilde{Z}_u^{lj} \right| \\ &\leq \int_0^{s^-} |h_{kl} - \tilde{h}_{kl}|(s-u) dZ_u^{lj} + \int_0^{s^-} |\tilde{h}_{kl}|(s-u) |d(Z_u^{lj} - \tilde{Z}_u^{lj})| \\ &\leq \int_0^{s^-} \gamma(s-u) d \sum_{l=1}^{\mathcal{K}} Z_u^{lj} + \int_0^{s^-} \tilde{h}(s-u) \left| d \sum_{l=1}^{\mathcal{K}} (Z_u^{lj} - \tilde{Z}_u^{lj}) \right|. \end{aligned} \tag{65}$$

We take expectation and apply Lemma 22 of [7] to obtain

$$\mathbb{E} \int_0^t \left| \int_0^{s^-} h_{kl}(s-u) dZ_u^{lj} - \int_0^{s^-} \tilde{h}_{kl}(s-u) d\tilde{Z}_u^{lj} \right| ds \leq \int_0^t \gamma(t-s) L(1 + \mathbb{E}\|X_T\|) ds + \int_0^t \tilde{h}(t-s) \delta_s ds.$$

Note that this expression does not depend on  $k, l$  nor  $j$ . Thus we get

$$\begin{aligned} & \sum_{k=1}^{\mathcal{K}} \int_0^t \mathbb{E} |\tilde{X}_s^k - X_s^k| ds \\ & \leq \sum_{k=1}^{\mathcal{K}} \int_0^t N^{-1} \sum_{l=1}^{\mathcal{K}} \sum_{j=1}^{N_l} \mathbb{E} \left| \int_0^{s-} h_{kl}(s-u) dZ_u^{lj} - \int_0^{s-} \tilde{h}_{kl}(s-u) d\tilde{Z}_u^{lj} \right| ds \\ & \leq C \left( \int_0^t \gamma(s) ds + \int_0^t |\tilde{h}(t-s)| \delta_s ds \right). \end{aligned} \tag{66}$$

Inserting inequality (66) into (64), we obtain

$$\delta_t \leq C \left( \int_0^t \gamma(s) ds + \int_0^t (\tilde{h}(t-s) + 1) \delta_s ds \right).$$

The proof will be complete, after repeating the argument for bounded  $\mathbb{E}|X|$ , but with  $\delta$  in place of  $\mathbb{E}|X|$ . □

**Proof of Theorem 3.3.** Let  $(\tilde{Z}^N, \tilde{X}^N, \tilde{A}^N)$  be the  $N$ -dimensional age dependent Hawkes process induced by the same parameters as  $(Z^N, X^N, A^N)$ , except the weight functions  $h_{kl}$  instead of  $h_{kl}^N$ .

Fix  $T > 0$  and consider  $t \in [0, T]$ . We have

$$\sum_{k=1}^{\mathcal{K}} |d(Z_t^{Nk1} - Z_t^{k1})| \leq \sum_{k=1}^{\mathcal{K}} \mathbb{E} |d(Z_t^{Nk1} - \tilde{Z}_t^{Nk1})| + \sum_{k=1}^{\mathcal{K}} \mathbb{E} |d(\tilde{Z}_t^{Nk1} - Z_t^{k1})| := \tilde{\delta}_t^{Nk} + \delta_t^{Nk}.$$

The first term converges by Lemma 3.4, and so it remains to prove convergence of  $\delta_t^N$ . This part of the proof follows closely the proof given by Chevallier in [3], but we include it here for completeness. Let  $C$  be a dynamic constant depending on  $p_k, L, T, \mathcal{K}, \|r\|_T$  and  $(h_{kl})$ . We use the symbol  $\varepsilon(N)$  for any function depending on the same parameters as  $C$ , and  $N$  such that  $\varepsilon(N) \xrightarrow{N \rightarrow \infty} 0$ . Recall that  $\|x\|_T$  is bounded by  $C$  sufficiently large by Lemma 3.1.

As in the proof of Lemma 3.1 we obtain

$$\delta_t^N \leq C \left( \int_0^t \mathbb{E} |\tilde{X}_s^N - x_s| ds + \int_0^t \delta_s^N ds \right). \tag{67}$$

This inequality prepares for an application of Gronwall's inequality, but first we bound  $\int_0^t \mathbb{E} |\tilde{X}_s - x_s| ds$  using  $\delta_t^N$  as well. Indeed, set  $\Lambda_t^{kj} := \int_0^{t-} \psi(x_s^k, A_s^{kj}) ds$ , which is the compensator of  $Z^{kj}$ . We write  $p_k = N_k/N + \varepsilon(N)$  and obtain

$$x_t^k = N^{-1} \sum_{l=1}^{\mathcal{K}} \sum_{j=1}^{N_l} \int_0^{t-} h_{kl}(t-s) d\phi_s^l + r_t^k + \varepsilon(N) \sum_{l=1}^{\mathcal{K}} \int_0^t h_{kl}(t-s) d\phi_s^l.$$

Since  $d\phi_s^l = \mathbb{E} \psi(x_s^l, A_s^{lj}) ds$  and  $\mathbb{E} \psi(x_s^l, A_s^{lj})$  are locally bounded, the entire right term may be replaced by an  $\varepsilon$ -function. For fixed  $k \leq \mathcal{K}$ , we apply the triangle inequality

$$\begin{aligned} & \int_0^t (\mathbb{E} |\tilde{X}_s^{Nk} - x_s^k| - \mathbb{E} |R_s^{Nk} - r_s^k|) ds \\ & \leq \varepsilon(N) \\ & + \int_0^t N^{-1} \sum_{l=1}^{\mathcal{K}} \mathbb{E} \left| \sum_{j=1}^{N_l} \int_0^{s-} h_{kl}(s-u) d(\phi_u^l - \Lambda_u^{lj}) \right| ds \end{aligned} \tag{68}$$

$$+ \int_0^t N^{-1} \sum_{l=1}^{\mathcal{K}} \mathbb{E} \left| \sum_{j=1}^{N_l} \int_0^{s-} h_{kl}(s-u) d(\Lambda_u^{lj} - Z_u^{lj}) \right| ds \tag{69}$$

$$\begin{aligned}
 & + \int_0^t N^{-1} \sum_{l=1}^{\mathcal{K}} \mathbb{E} \left| \sum_{j=1}^{N_l} \int_0^{s-} h_{kl}(s-u) d(Z_u^{lj} - \tilde{Z}_u^{N_l j}) \right| ds \\
 & := \varepsilon(N) + B_t^{1k} + B_t^{2k} + B_t^{3k}.
 \end{aligned} \tag{70}$$

We now proceed to bound  $B^i := \sum_{k=1}^{\mathcal{K}} B^{ik}$ ,  $i \leq 3$ . Define  $h = \sum_{k,l=1}^{\mathcal{K}} |h_{kl}|$ . Rewrite  $\phi$  and  $\Lambda$  in terms of their densities, and thereby obtain a bound for the inner-most sum in (68) for  $s \in [0, t]$ ,  $l \leq \mathcal{K}$ , which is given by

$$\mathbb{E} \int_0^s \sum_{j=1}^{N_l} h(s-u) |d(\phi_u^l - \Lambda_u^{lj})| \leq \int_0^s h(s-u) \mathbb{E} \sum_{j=1}^{N_l} |\psi^l(x_u^l, A_u^{lj}) - \mathbb{E} \psi^l(x_u^l, A_u^{lj})| du.$$

Notice that the sum consists of i.i.d. terms, so we may apply Cauchy–Schwarz to bound it by  $\sqrt{N_l \text{Var}(\psi(x_u^l, A_u^{lj}))}$ , which is bounded for  $u \in [0, T]$  by  $\sqrt{N_l} C(1 + \|r^l\|_T)$  using (3). Insert this into (68) to see that

$$B_t^1 = \sum_{k=1}^{\mathcal{K}} B_t^{1k} \leq \varepsilon(N).$$

For  $B_t^2$ , recall that  $(Z^{lj} - \Lambda^{lj})_j$  are i.i.d. for fixed  $l$ . By Cauchy-Schwarz, we obtain a bound for the inner-most sum of (69)

$$N_l^{1/2} \sqrt{\text{Var} \int_0^s h_{kl}(s-u) d(Z_u^{l1} - \Lambda_u^{l1})}. \tag{71}$$

To treat the process inside the root, fix  $s \geq 0$ ,  $l \leq \mathcal{K}$  and consider the process

$$I : r \mapsto \int_0^{r \wedge s} h_{kl}(s-u) d(Z_u^{l1} - \Lambda_u^{l1}).$$

Then  $I$  is a martingale, and

$$\text{Var } I_s = \mathbb{E}[I]_s = \mathbb{E} \int_0^s h_{kl}^2(s-u) d\Lambda_u^{l1} = \int_0^s h_{kl}^2(s-u) \mathbb{E} \psi^l(x_u^l, A_u^{l1}) du.$$

Since  $h_{kl} \in \mathcal{L}_{\text{loc}}^2$  it follows that  $\text{Var } I_s$  is bounded on  $s \in [0, T]$ , and so

$$B_t^2 \leq \varepsilon(N)$$

for all  $t \leq T$ . For  $B^3$  the triangle inequality, and Lemma 22 [7] gives

$$B_t^3 \leq C \int_0^t h(t-s) \delta_s^N ds.$$

We plug the bounds for  $B_1$ ,  $B_2$  and  $B_3$  into (67) to obtain

$$\delta_t^N \leq C \left( \int_0^t h(t-s) \delta_s^N + \varepsilon(N) + \sum_{k=1}^{\mathcal{K}} \mathbb{E} |R_s^{Nk} - r_s^k| ds \right).$$

Applying Lemma A.2 in the Appendix yields

$$\delta_t^N \leq \varepsilon(N) + C \int_0^T \mathbb{E} |R_s^N - r_s| ds = \varepsilon(N) \tag{72}$$

for all  $t \leq T$ , which implies the desired result. □

A.2. Results about counting processes

A.2.1. Measure theory on  $(M_E, \mathcal{M}_E)$

This section provides a brief overview of measure theory on the measurable space  $(M_E, \mathcal{M}_E)$  of bounded measures defined on a Polish space  $E$ . We refer to A.2.6 in [6] for more details. Let  $d$  be a distance so that  $(E, d)$  is complete and separable. A measure  $\nu$  on  $E$  is said to be *boundedly finite* if  $\text{diam}(A) < \infty$  implies that  $\nu(A) < \infty$  for  $A \in \mathcal{B}(E)$ . Let  $M_E$  be the space of all boundedly-finite measures on  $(E, \mathcal{B}(E))$ . This space may now be equipped with the *weak-hat* metric  $\hat{d}$ , making  $(M_E, \hat{d})$  a complete and separable space itself. The Borel-algebra  $\mathcal{M}_E$  is easily characterized by the projections  $\Pi_A : M_E \ni \nu \mapsto \nu(A)$  in the sense that

$$\mathcal{M}_E = \sigma(\Pi_A, A \in \mathbb{D}), \tag{73}$$

where  $\mathbb{D} \subset \mathcal{B}(D)$  is a semi-ring of bounded sets. A random variable taking values in the space of measures is called a *random measure*. A particular interesting example is when  $E = \mathbb{R} \times \mathbb{R}_+$  as considered in this paper. A Poisson random measure  $\pi : \Omega \rightarrow M_{\mathbb{R} \times \mathbb{R}_+}$  on  $\mathbb{R} \times \mathbb{R}_+$  with Lebesgue intensity measure is a random measure such that for any disjoint  $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R}_+)$ ,

- $\pi(A_i) \sim \text{Pois}(\int_{A_i} ds dz)$
- $\pi(A_1) \perp \dots \perp \pi(A_n)$ .

Since the underlying space is on the form  $\mathbb{R} \times \mathbb{R}_+$ , the first coordinate of  $\pi$  may be thought of as the time coordinate; and concepts like stationarity and ergodicity transfer naturally to random measures. Define the shift operator as the automorphism on  $M_{\mathbb{R} \times \mathbb{R}_+}$  given by

$$(\theta^r \nu)(C) = \nu(\{(t, x) \in \mathbb{R} \times \mathbb{R}_+ : (t - r, x) \in C\}).$$

Then a random measure  $\sigma$  on  $\mathbb{R} \times \mathbb{R}_+$  is said to be *stationary* if the distribution is invariant under shift

$$\mathcal{L}(\sigma) = \mathcal{L}(\theta^r \sigma)$$

for all  $r \in \mathbb{R}$ . Stationarity is equivalent to have invariance of the finite dimensional distributions (fidi's) (Proposition 6.2.III of [6])

$$P\left(\bigcap_{i=1}^n \{\sigma(A_i) \in B_i\}\right) = P\left(\bigcap_{i=1}^n \{\theta^r \sigma(A_i) \in B_i\}\right)$$

for all  $r \in \mathbb{R}, n > 0, A_1, \dots, A_n \in \mathcal{B}(\mathbb{R} \times \mathbb{R}_+), B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}_+)$ . A stationary random measure  $\sigma$  is *mixing* if

$$P(\sigma \in V, \theta^r \sigma \in W) \xrightarrow{|r| \rightarrow \infty} P(\sigma \in V)P(\sigma \in W) \tag{74}$$

for all  $V, W \in \mathcal{M}_{\mathbb{R} \times \mathbb{R}_+}$ . We refer to Chapter 10.2–10.3 of [6] for a thorough introduction to ergodic theory for random measures. As is the case for processes, mixing implies that  $\mathcal{L}(\sigma)$  is ergodic w.r.t. the shift operator  $\theta^r$  for all  $r > 0$ , meaning that all invariant events have probability 0/1.

Finally, we present a core measurability result, which can be applied to show measurability of all processes treated in this article.

**Lemma A.1.** *Let  $D, E$  be complete and separable metric spaces, and let  $H : D \times E \rightarrow \mathbb{R}_+$  be measurable. The section integral  $F : M_E \times D \rightarrow \overline{\mathbb{R}}$*

$$F(\nu, d) = \int_E H(d, s) d\nu(s)$$

is  $\mathcal{M}_E \times \mathcal{B}(D) \rightarrow \mathcal{B}(\overline{\mathbb{R}})$  measurable.

**Proof.** We start by defining  $G : M_{D \times E} \rightarrow \overline{\mathbb{R}_+}$  as

$$G : \rho \mapsto \int H(x, y)\rho(dx, dy).$$

It is easily seen that  $G$  is measurable.

Consider the map  $m : M_E \times D \rightarrow M_{D \times E}$  given by  $(\nu, d) \mapsto \delta_d \otimes \nu$ . To prove measurability of  $m$  it is sufficient to treat projections into bounded boxes  $A \times B$ ,  $A \in \mathcal{B}(D)$ ,  $B \in \mathcal{B}(E)$ . Such projections are simply given as  $\Pi_{A \times B} m : (\nu, d) \mapsto \mathbb{1}\{B\}(d)\Pi_A(\nu)$  and are therefore measurable. We conclude that

$$G \circ m(\nu, d) = \int \int H(u, s) d\delta_d(u) d\nu(s) = \int H(d, s) d\nu(s) = F(\nu, d),$$

proving that  $F$  is measurable. □

We shall need the following version of Gronwall’s lemma which has been proven in [7]. Recall that for any function  $g : \mathbb{R}_+ \rightarrow \mathbb{R}$  and any  $T > 0$ , we have introduced  $\|g\|_T = \sup_{t \leq T} |g(t)|$ .

**Lemma A.2 (Lemma 23 of [7]).** *Let  $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be locally integrable and  $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be locally bounded. Let  $T \geq 0$ .*

1. *Let  $u$  be a locally bounded nonnegative function satisfying  $u_t \leq g_t + \int_0^t h(t-s)u_s ds$  for all  $t \in [0, T]$ . If  $b > 0$  satisfies that*

$$\int_0^T h(s)\mathbb{1}\{h(s) \geq b\} ds < \frac{1}{2}, \tag{75}$$

*then  $\|u\|_T \leq 2e^{2bT} \|g\|_T =: C_T \|g\|_T$ .*

2. *Let  $(u^n)$  be a sequence of locally bounded nonnegative functions such that  $u_t^{n+1} \leq g_t + \int_0^t h(t-s)u_s^n ds$  for all  $t \in [0, T]$ . Then  $\sup_n \|u^n\|_T \leq C_T (\|g\|_T + \|u^0\|_T)$ . Moreover, if the inequality is satisfied with  $g \equiv 0$ , then  $\sum_n u^n$  converges uniformly on  $[0, T]$ .*

A.2.2. Point process results

We collect some useful results on point processes known in the literature.

**Lemma A.3.**

*Let  $H : (\Omega \times \mathbb{R} \times \mathbb{R}) \rightarrow \mathbb{R}$  be  $\mathcal{P} \otimes \mathcal{B} \rightarrow \mathcal{B}$  measurable and assume that almost surely*

$$Z : t \mapsto \int_0^t \int_0^\infty H(s, z)\pi(ds, dz)$$

*does not explode; that is, for all  $t > 0$ ,*

$$\int_0^t \int_0^\infty |H(s, z)|\pi(ds, dz) < \infty$$

*almost surely.*

1. *If  $H$  is bounded, then the compensator  $\Lambda$  of  $Z$  is given by*

$$\Lambda : t \mapsto \int_0^t \int_0^\infty H(s, z) dz ds,$$

*i.e.  $Z - \Lambda$  is a local  $(\mathcal{F}_t)$ -martingale.*

2. *If moreover  $s \mapsto \mathbb{E} \int |H(s, z)| dz$  is locally integrable, then  $Z - \Lambda$  is a martingale.*

3. *Fix  $T \geq 0$  and assume that  $\Lambda$  can be written as*

$$\Lambda_t = \int_0^t \lambda_s ds.$$

*Assume also that  $\lambda(s) = F(Z_{|(-\infty, T]}, s) + \varepsilon(s)$  where  $F$  is  $\mathcal{M}_{\mathbb{R}} \times \mathcal{B}(\mathbb{R}) \mapsto \mathcal{B}(\mathbb{R})$  measurable and  $t \mapsto \varepsilon(t)$  is  $(\mathcal{F}_{t \wedge T})$ -predictable. It holds that*

$$P(Z(T, \infty) = 0 | \mathcal{F}_T) \stackrel{a.s.}{=} \exp\left(-\int_T^\infty F(Z_{|(-\infty, T]}, s) + \varepsilon(s) ds\right).$$



**Proof.** The first point follows from [16], Theorem 1.8 of Chapter II, by using the localizing sequence  $T_n = \inf\{t : \int_0^t \int_0^\infty |H(s, z)| \pi(ds, dz) \geq n\}$ ,  $n \geq 1$ , since

$$\int_0^{T_n} \int_0^\infty |H(s, z)| \pi(ds, dz) \leq n + \|H\|_\infty.$$

For the second point, let  $M_t := Z_t - \Lambda_t$ . It suffices to show that  $\mathbb{E}(\sup_{s \leq t} |M_s|) < \infty$  which follows from

$$\mathbb{E}\left(\sup_{s \leq t} |M_{s \wedge T_n}|\right) \leq 2\mathbb{E} \int_0^t \int_0^\infty |H(s, z)| ds dz < \infty$$

by monotone convergence. The third point is Lemma 1 in [2]. □

**Lemma A.4 (Lemma 5 of [2]).** Assume that  $X, Y$  are  $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ -progressive and that for all  $s \geq 0$

$$P(X_t = Y_t \forall t > s | \mathcal{F}_s) \geq U_s - r(s),$$

where  $U$  is ergodic,  $P(U_s > 0) > 0$  and  $r(t) \xrightarrow{a.s.} 0$  for  $t \rightarrow \infty$ . Then almost surely,  $X$  and  $Y$  couple in finite time.

## References

- [1] E. Bacry and J. F. Muzy. Hawkes model for price and trades high-frequency dynamics. *Quant. Finance* **14** (2014) 1147–1166. MR3219705 <https://doi.org/10.1080/14697688.2014.897000>
- [2] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Ann. Probab.* **24** (3) (1996) 1563–1588. MR1411506 <https://doi.org/10.1214/aop/1065725193>
- [3] J. Chevallier. Mean-field limit of generalized Hawkes processes. *Stochastic Process. Appl.* **127** (2017) 3870–3912. MR3718099 <https://doi.org/10.1016/j.spa.2017.02.012>
- [4] J. Chevallier, M. J. Caceres, M. Doumic and P. Reynaud-Bouret. Microscopic approach of a time elapsed neural model. *Math. Models Methods Appl. Sci.* **25** (14) (2015) 2669–2719. MR3411353 <https://doi.org/10.1142/S021820251550058X>
- [5] M. Costa, C. Graham, L. Marsalle and V. C. Tran. Renewal in Hawkes processes with self-excitation and inhibition, 2018. Available at [arXiv.org/abs/1801.04645](https://arxiv.org/abs/1801.04645).
- [6] D. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer-Verlag, New York, 1988. MR0950166
- [7] S. Delattre, N. Fournier and M. Hoffmann. Hawkes processes on large networks. *Ann. Appl. Probab.* **26** (2016) 216–261. MR3449317 <https://doi.org/10.1214/14-AAP1089>
- [8] S. Ditlevsen and E. Löcherbach. Multi-class oscillating systems of interacting neurons. *Stochastic Process. Appl.* **127** (2017) 1840–1869. MR3646433 <https://doi.org/10.1016/j.spa.2016.09.013>
- [9] A. Duarte, E. Löcherbach and G. Ost. Stability, convergence to equilibrium and simulation of non-linear Hawkes Processes with memory kernels given by the sum of Erlang kernels. Available at [arXiv:1610.03300](https://arxiv.org/abs/1610.03300), 2017. Submitted.
- [10] N. Fournier and E. Löcherbach. A toy model of interacting neurons. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** (2016) 1844–1876. MR3573298 <https://doi.org/10.1214/15-AIHP701>
- [11] F. Gerhard, M. Deger and W. Truccolo. On the stability and dynamics of stochastic spiking neuron models: Nonlinear Hawkes process and point process GLMs. *PLoS Comput. Biol.* **13** (2) (2017), e1005390. <https://doi.org/10.1371/journal.pcbi.1005390>
- [12] C. Graham and P. Robert. Interacting multi-class transmissions in large stochastic systems. *Ann. Appl. Probab.* **19** (6) (2009) 2334–2361. MR2588247 <https://doi.org/10.1214/09-AAP614>
- [13] P. R. Halmos. *Measure Theory*. Springer-Verlag, New York, 1974.
- [14] A. G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** (1971) 83–90. MR0278410 <https://doi.org/10.1093/biomet/58.1.83>
- [15] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *J. Appl. Probab.* **11** (1974) 493–503. MR0378093 <https://doi.org/10.2307/3212693>
- [16] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*, 2nd edition. Springer-Verlag, Berlin, 2003. MR1943877 <https://doi.org/10.1007/978-3-662-05265-5>
- [17] Y. Miyahara. Invariant measures of ultimately bounded stochastic processes. *Nagoya Math. J.* **49** (1973) 149–153. MR0321196
- [18] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete* **43** (1978) 309–318.
- [19] K. Pakdaman, B. Perthame and D. Salort. Relaxation and self-sustained oscillations in the time elapsed neuron network model. *SIAM J. Appl. Math* **73** (2013) 1260–1279. MR3071416
- [20] P. Reynaud-Bouret, V. Rivoirard and C. Tuleau-Malot In *Inference of Functional Connectivity in Neurosciences via Hawkes Processes*. 1st IEEE Global Conference on Signal and Information Processing, Austin, Texas, 2013.
- [21] P. Reynaud-Bouret and S. Schbath. Adaptive estimation for Hawkes processes; application to genome analysis. *Ann. Statist.* **5** (2010) 2781–2822. MR2722456 <https://doi.org/10.1214/10-AOS806>
- [22] A. H. Zadeh and R. Sharda. *Hawkes Point Processes for Social Media Analytics. Reshaping Society Through Analytics, Collaboration, and Decision Support: Role of Business Intelligence and Social Media*. Springer, Berlin, 2015.