# Mean-field limit of generalized Hawkes processes

Julien Chevallier*

Univ. Nice Sophia Antipolis, CNRS, LJAD, UMR 7351, 06100 Nice, France.

## Abstract

We generalize multivariate Hawkes processes mainly by including a dependence with respect to the age of the process, i.e. the delay since the last point.

Within this class, we investigate the limit behaviour, when $n$ goes to infinity, of a system of $n$ mean-field interacting age-dependent Hawkes processes. We prove that such a system can be approximated by independent and identically distributed age dependent point processes interacting with their own mean intensity. This result generalizes the study performed by Delattre, Fournier and Hoffmann (2015).

In continuity with the work of Chevallier et al. (2015), the second goal of this paper is to give a proper link between these generalized Hawkes processes as microscopic models of individual neurons and the age-structured system of partial differential equations introduced by Pakdaman, Perthame and Salort (2010) as macroscopic model of neurons.

## I  Introduction

In the recent years, the self-exciting point process known as the Hawkes process [22] has been used in very diverse areas. First introduced to model earthquake replicas [25] or [35] (ETAS model), it has been used in criminology to model burglary [32], in genomic data analysis to model occurrences of genes [20, 44], in social networks analysis to model viewing or popularity [3, 10], as well as in finance [1, 2]. We refer to [28] or [47] for more extensive reviews on applications of Hawkes processes. A univariate (nonlinear) Hawkes process is a point process $N$ admitting a stochastic intensity of the form

$$\lambda_t = \Phi\left(\int_0^{t-} h(t-z)N(dz)\right),\tag{1}$$

where $\Phi : \mathbb{R} \to \mathbb{R}_+$ is called the intensity function, $h : \mathbb{R}_+ \to \mathbb{R}$ is called the self-interaction function and $N(dz)$ denotes the point measure associated with $N$.

Such a form of the intensity is motivated by practical cases where all the previous points of the process may impact the rate of appearance of a new point. The influence of the past points is formulated in terms of the delay between those past occurrences and the present time, through the weight function $h$. In the natural framework where $h$ is non-negative and $\Phi$ increasing, this choice of interaction models an excitatory phenomenon: each time the process has a jump, it excites itself in the sense that it increases its intensity and thus the probability of finding a new point. A classical case is the *linear* Hawkes process for which $h$ is non-negative and $\Phi(x) = \mu + x$ where $\mu$ is a positive constant called the spontaneous rate. Note however that Hawkes processes can also describe inhibitory phenomena. For example, the function $h$ may take negative values, $\Phi$ being the positive part modulo the spontaneous rate $\mu$, i.e. $\Phi(x) = \max(0, \mu + x)$.

---

*Corresponding author: e-mail: julien.chevallier@unice.fr

Hawkes processes are also really suitable to model interacting particles such as bids and asks orders in limit order books [33]. Multivariate Hawkes processes consist of multivariate point processes $(N^1, \ldots, N^n)$ whose intensities are respectively given for $i = 1, \ldots, n$ by

$$\lambda_t^i = \Phi_i \left( \sum_{j=1}^n \int_0^{t-} h_{j\to i}(t-z)N^j(dz) \right), \tag{2}$$

where $\Phi_i : \mathbb{R} \to \mathbb{R}_+$ is the intensity function associated with the particle $i$ and $h_{j\to i}$ is the *interaction function* describing the influence of each point of $N^j$ on the appearance of a new point onto $N^i$, via its intensity $\lambda^i$.

When the number of interacting particles is huge (as, for instance, financial or social networks agents), one may be willing to let the number of particles goes to infinity. This is especially true for multivariate Hawkes processes subject to mean-field interactions. In such a case, we may indeed expect propagation of chaos, namely the particles are expected to become asymptotically independent, provided that they start from independent and identically distributed (i.i.d.) initial conditions and submitted to i.i.d. sources of noise. Mean-field type interactions involve some homogeneity and some symmetry through coefficients that depend upon the empirical measure of the processes: In the limit regime, the coefficients depend upon the common asymptotic distribution of the particles, which satisfies nonlinear dynamics, sometimes called of McKean-Vlasov type.

The study of mean-field situations for Hawkes processes was initiated by Delattre et al. [13] by considering the following form of intensity

$$\lambda_t^i = \Phi \left( \frac{1}{n} \sum_{j=1}^n \int_0^{t-} h(t-z)N^j(dz) \right), \tag{3}$$

where, in comparison with (2), all the $\Phi_i$'s and the $h_{j\to i}$'s are the same. In particular, it is shown in [13] that mean-field interacting Hawkes processes are well approximated, when the size of the network $n$ goes to infinity, by i.i.d. Poisson processes of the McKean-Vlasov type in the sense that their intensity is given by the following implicit formula $\overline{\lambda}(t) = \Phi(\int_0^t h(t-z)\overline{\lambda}(z)dz)$.

In the present article, a generalized version of Hawkes processes with mean-field interactions, namely Age Dependent Random Hawkes Processes (ADRHP for short), is studied. For any point process $N$, we call *age process* associated with $N$ the predictable process $(S_{t-})_{t\geq 0}$ given by

$$S_{t-} = t - \sup\{T \in N, T < t\}, \quad \text{for all } t > 0, \tag{4}$$

and extended by continuity in $t = 0$. In particular, its value in $t = 0$ is entirely determined by $N \cap \mathbb{R}_-$ and is well-defined as soon as there is a point therein. In comparison with the standard mean-field type Hawkes processes studied in [13] we assume here that the intensity function $\Phi$ in (3) (which is denoted by $\Psi$ to avoid confusion) may also depend on the age process $(S_{t-}^i)_{t\geq 0}$ associated with the point process $N^i$, like for instance

$$\lambda_t^i = \Psi \left( S_{t-}^i, \frac{1}{n} \sum_{j=1}^n \int_0^{t-} h(t-z)N^j(dz) \right). \tag{5}$$

This more general choice for the intensity makes the main difference with [13], where the intensity is assumed to be of the simpler form (3) only. We then show that, instead of Poisson processes of the McKean-Vlasov type, the limit processes associated with mean-field interacting age-dependent Hawkes processes are point processes of the McKean-Vlasov type whose stochastic intensity not only depends on the time but also on the age. More precisely, for the toy example (5), the intensity of the limit process $\overline{N}$ would be given by the following implicit formula $\overline{\lambda}_t = \Psi(\overline{S}_{t-}, \int_0^t h(t-z)\mathbb{E}\left[\overline{\lambda}_z\right] dz)$ where $(\overline{S}_{t-})_{t\geq 0}$ is the age process associated with $\overline{N}$.

Part of our analysis finds its motivation in the use of Hawkes processes for the modelling in neuroscience. First of all, at a microscopic scale, Hawkes processes are commonly used in theoretical studies [9, 21, 40, 43] to describe the time occurrences of the action potentials of different neurons. These action potentials are associated with brutal changes of the membrane potential, called *spikes* in the rest of the article. The motivation for using Hawkes process is well-understood and linked

with the *synaptic integration* phenomenon: the interaction functions $h_{j\to i}$ describe the fact that, whenever a neuron spikes, the membrane potential of the other neurons in the network (and thus their firing rate as well) may change. In that sense, the $L^1$ norm of the interaction function $h_{j\to i}$, for $j \neq i$, is the analogue of the synaptic weight of neuron $j$ over neuron $i$, that is the strength of the influence of neuron $j$ over neuron $i$ through their synaptic connection. For example, if one considers $h_{j\to i} = \alpha_{j\to i} h$ for a fixed function $h$ then $\alpha_{j\to i}$ represents the (relative) synaptic weight of neuron $j$ over neuron $i$. Notice that in the present paper we allow the functions $h_{j\to i}$ to be random and thus the synaptic weights to be random as well (as in [15] for instance).

To model a transition in the behaviour of the network at the shifting time $t = 0$, the distribution of $N \cap \mathbb{R}_-$ is considered as an initial condition of the dynamics of the point process and may be different from the distribution of a Hawkes process. Therefore, to specify the dependence of the dynamics (on $\mathbb{R}_+$) upon the initial condition, the following form of intensity can be considered:

$$\lambda_t = \Phi\left(\int_{-\infty}^{t-} h(t-z)N(dz)\right) = \Phi\left(\int_0^{t-} h(t-z)N(dz) + F(t)\right), \qquad (6)$$

where $F(t) := \int_{-\infty}^0 h(t-z)N(dz)$ models, in a Hawkes manner, the influence of the initial condition. This choice of $F$ is taken from [8]. However, other choices are conceivable. For example, more general functions $F$ may describe a stimulus at a given time $t_0 < 0$ which is more convenient for peristimulus analyses like [41].

However, standard Hawkes processes fail to model, in a convenient way, the neurophysiological constraint known as *refractory period*, that is the fact that a neuron cannot spike twice in a too short delay. This is the main reason why we allow the intensity of the Hawkes process to depend upon the age in the present study. In comparison with (1), one may represent strict refractory period by considering, for instance, the following form of intensity:

$$\lambda_t = \Phi\left(\int_0^{t-} h(t-z)N(dz)\right) \mathbb{1}_{S_{t-} \geq \delta}, \qquad (7)$$

where $(S_{t-})_{t\geq 0}$ is the age process associated with $N$ and $\delta$ is a parameter corresponding to the time length of the strict refractory period of a neuron. This sounds as an alternative to the strategy used in [9]. Therein, refractory periods are described by choosing, in the standard formulation of Hawkes processes, strongly negative self-interaction functions at a very short range. The strategy used in the present article is more flexible: synaptic integration and refractory period involve different aspects of the physiology of a neuron and so we prefer to describe each of them by different elements in the modelling.

Mean-field approaches have been already used to pass from a microscopic to a macroscopic description of neural networks. Taking for granted that the network is symmetric enough, the mean-field modelling sounds quite fair. Indeed neural networks admit a large number of vertices and are highly connected (see [15] for a review). One may distinguish three types of models: intrinsically spike generating models (like the FitzHugh–Nagumo model [29]), threshold spike generating models (like the integrate-and-fire model [6, 12, 11]) and point processes models ([17] or [18, 24]).

As usual with McKean-Vlasov dynamics, the asymptotic evolution (when $n$ goes to infinity) of the distribution of the population at hand can be described as the solution of a nonlinear partial differential equation (PDE). In the present article, the candidate to describe the dynamics at a macroscopic level is the following age structured system of nonlinear PDEs studied by Pakdaman, Perthame and Salort in a series of articles [36, 37, 38].

$$\begin{cases} \dfrac{\partial n\,(s,t)}{\partial t} + \dfrac{\partial n\,(s,t)}{\partial s} + p\,(s, X\,(t))\,n\,(s,t) = 0, \\ m\,(t) := n\,(0,t) = \displaystyle\int_0^{+\infty} p\,(s, X\,(t))\,n\,(s,t)\,ds. \end{cases} \qquad \text{(PPS)}$$

Here, $n(s,t)$ represents the probability density of the age $s$ of a neuron at time $t$ where the age of a neuron is the delay since its last spike. Of course, the definition of the age of a neuron fits with the definition of the age associated with a point process as soon as the spike train is modelled by a point process. The function $p$ represents the firing rate which may depend on the age $s$. As already explained, this dependence describes for instance the phenomenon of refractory period

(e.g. $p(s,x) = \mathbb{1}_{s \geq \delta}$ for some $\delta > 0$). The function $p$ may also depend on the global activity of the network which is denoted by $X(t) := \int_0^t d(z)n(0, t-z)dz$ where $d$ is some delay function. This global (deterministic) variable $X(t)$ corresponds to the mean of the integral that appears in (1). This correspondence forms the basis of the previous work [8] where a bridge is made between a modified version of (PPS) and the distribution of the age of a single neuron (modelled by a point process). From a neural network point of view, this distribution can of course be recovered as the limit of the empirical distribution associated with a network of i.i.d. neurons.

The study of the link between the (PPS) system and a mean-field interacting neural network (modelled by point processes) was left as an open question in [8]. The heuristic of this mean-field interpretation comes from the specific structure of the variable $X(t)$ which brings out a non-linearity of the McKean-Vlasov type. One of the main purpose of the present paper is to answer that left open question. To be precise, this kind of study is performed in a preliminary work [42] for a firing rate $p$ that is continuous and non-decreasing in both variables and under Markovian assumptions. Transposed to the Hawkes framework, this last point corresponds to interaction functions of the form $h_{j \to i}(t) = e^{-\beta(t-\tau_j)}\mathbb{1}_{[\tau_j, +\infty)}(t)$ where $\beta$ is a constant and the $\tau_j$'s are i.i.d. random variables describing the propagation time of the signal from the neuron to the network. The convergence of the empirical measure is discussed in [42] when $p$ is continuous only but without any rate of convergence. In the present study, rates of convergence are given for non Markovian Hawkes processes (that is non necessary exponential interaction functions) as well as for firing rates that are discontinuous with respect to the age, like (7) for instance. However, we make the crucial assumption that the firing rate $p$ is Lipschitz continuous with respect to the second variable.

To sum up, we call Age Dependent Random Hawkes Process (ADRHP) a multivariate age dependent Hawkes process (like (7) for instance) with some general dependence with respect to the initial condition (6) and with some randomness regarding the interaction functions $h_{j \to i}$. This article has two main purposes: extend the mean-field approximation obtained in [13] to this generalization of Hawkes processes and establish a proper link between the microscopic modelling of individual neurons given by a $n$-particle system of mean-field interacting age-dependent Hawkes processes (like (5) for instance) and the macroscopic modelling given by the (PPS) system.

The paper is organized as follows. In section II, we introduce ADRHPs and we show how to represent them as solutions of an SDE driven by a Poisson noise. As a by-product of this representation, we get, on the one hand, the existence of such processes, and on the other hand, an efficient way to get a coupling between our $n$-particle system and $n$ i.i.d. limit processes. As a first step towards the mean-field approximation, the limit dynamics is studied in Section III. Existence and uniqueness of a solution of the (PPS) system, which is our candidate to drive the limit dynamics, are proved in Theorem III.5. As a consequence, we get the existence of point processes of the McKean-Vlasov type whose intensity depends on both the time and the age. In Section IV, these processes are proved to be the mean-field approximation of age dependent random Hawkes processes (Theorem IV.1 and Corollary IV.4) using coupling arguments under either of the two following main assumptions: the intensity is bounded or the intensity does not depend on the age. Notice that even when the intensity does not depend on the age, the results presented here extend the ones given in [13] since random interaction functions $h_{j \to i}$ as well as dependences with respect to the dynamics before time 0 cannot be taken into account in [13]. Finally, the link between age dependent random Hawkes processes and the (PPS) system is given by Corollary IV.4. For sake of readability, most of the computations and technical lemmas are given in two appendices.

**General notations**

- The space of continuous function from $E$ to $\mathbb{R}$ is denoted by $\mathcal{C}(E)$.

- The space of Radon (resp. probability) measures on $E$ is denoted by $\mathcal{M}(E)$ (resp. $\mathcal{P}(E)$).

- For $\nu$ in $\mathcal{P}(E)$, $X \sim \nu$ means that $X$ is a random variable distributed according to $\nu$.

- For $f : \mathbb{R} \to \mathbb{R}$, $||f||_1$, $||f||_2$ and $||f||_\infty$ respectively denote the $L^1$, $L^2$ and $L^\infty$ norms of $f$.

# II   Age dependent random Hawkes processes

In all the sequel, we focus on locally finite point processes, $N$, on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that are random countable sets of points of $\mathbb{R}$ such that for any bounded measurable set $K \subset \mathbb{R}$, the number of points in $N \cap K$ is finite almost surely (a.s.). The associated points define an ordered sequence of points $(T_n)_{n \in \mathbb{Z}}$. For a measurable set $A$, $N(A)$ denotes the number of points of $N$ in $A$. We are interested in the behaviour of $N$ on $(0, +\infty)$ and we denote $t \in \mathbb{R}_+ \mapsto N_t := N((0, t])$ the associated counting process. Furthermore, the point measure associated with $N$ is denoted by $N(dt)$. In particular, for any non-negative measurable function $f$, $\int_{\mathbb{R}} f(t) N(dt) = \sum_{i \in \mathbb{Z}} f(T_i)$.

We work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and suppose that the canonical filtration associated with $N$, namely $(\mathcal{F}_t^N)_{t \geq 0}$ defined by $\mathcal{F}_t^N := \sigma(N \cap (-\infty, t))$, is such that for all $t \geq 0$, $\mathcal{F}_t^N \subset \mathcal{F}_t$. Let us denote $\mathbb{F} := (\mathcal{F}_t)_{t \geq 0}$. We call $\mathbb{F}$-(predictable) intensity of $N$ any non-negative $\mathbb{F}$-predictable process $(\lambda_t)_{t \geq 0}$ such that $(N_t - \int_0^t \lambda_s ds)_{t \geq 0}$ is an $\mathbb{F}$-local martingale. Informally, $\lambda_t dt$ represents the probability that the process $N$ has a new point in $[t, t + dt]$ given $\mathcal{F}_{t-}$. Under some assumptions that are supposed here, this intensity process exists, is essentially unique and characterizes the point process (see [4] for more insight).

In analogy with the study of the dynamics of a variable over time, we use a dichotomy between the behaviour of the point process before time 0 (which is treated as an initial condition) and its behaviour after time 0 (which is supposed to admit a "Hawkes type" intensity). For every point process $N$, we denote $N_- = N \cap \mathbb{R}_-$ and $N_+ = N \cap (0, +\infty)$. In the rest of the paper, a point process on $\mathbb{R}$ is characterized by:

1. the distribution of $N_-$, namely $\zeta_{N_-}$, which gives the dynamics of $N$ on $\mathbb{R}_-$;

2. the $\mathbb{F}$-predictable intensity $\lambda_t$, which gives the dynamics of $N$ on $(0, \infty)$.

In particular, $\zeta_{N_-}$ characterizes the distribution of $T_0$ that is the last point (spike) before time 0. Notice that the $\sigma$-algebra $\mathcal{F}_0$ is such that $N_-$ is $\mathcal{F}_0$-measurable.

## II.1   Parameters of the model

The definition of an *age dependent random Hawkes process* (ADRHP) is given bellow, but let us first introduce the parameters of the model:

- a positive integer $n$ which is the number of particles (e.g. neurons) in the network (for $i = 1, \ldots, n$, $N^i$ represents the occurrences of the events (e.g. spikes) associated with the particle $i$);

- a distribution $\zeta_{N_-}$ determining the initial conditions $(N_-^i)_{i=1,..,n}$ which are i.i.d. point processes on $\mathbb{R}_-$ distributed according to $\zeta_{N_-}$;

- a distribution $\mu_H$ determining the matrix of interaction functions $\mathbf{H} = (H_{ij})_{1 \leq i,j \leq n}$ where $H_{ij} : \mathbb{R}_+ \to \mathbb{R}$ are $\mathcal{F}_0$-measurable random functions distributed according to $\mu_H$ such that

$$\begin{cases} \text{for any fixed } i = 1, \ldots, n, \text{ the variables } H_{i1}, \ldots, H_{in} \text{ are independent,} \\ \text{the vectors } (H_{i1}, \ldots, H_{in}) \text{ are exchangeable (with respect to } i), \\ \text{the matrix } \mathbf{H} \text{ is independent from the initial conditions } (N_-^i)_{i=1,..,n}; \end{cases} \tag{8}$$

- a distribution $\nu_F$ determining the matrix of functions $\mathbf{F} = (F_{ij})_{1 \leq i,j \leq n}$ where $F_{ij} : \mathbb{R}_+ \to \mathbb{R}$ are $\mathcal{F}_0$-measurable random functions distributed according to $\nu_F$ such that

$$\begin{cases} \text{for any fixed } i = 1, \ldots, n, \text{ the variables } F_{i1}, \ldots, F_{in} \text{ are independent,} \\ \text{the vectors } (F_{i1}, \ldots, F_{in}) \text{ are exchangeable (with respect to } i); \end{cases} \tag{9}$$

- an intensity function $\Psi : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$.

Note that the functions $H_{ij}$'s can in particular be equal to a given deterministic function $h$ which corresponds to more standard Hawkes processes.

## II.2   Definition via the intensity

The definition of an age dependent random Hawkes process is given by providing the form of its intensity.

**Definition II.1.** *An age dependent random Hawkes process (ADRHP) with parameters $(n, \mu_H, \nu_F, \Psi, \zeta_{N_-})$ is a family $(N^i)_{i=1,\dots,n}$ of point processes on $\mathbb{R}$ such that $(N^i_-)_{i=1,\dots,n}$ is a family of i.i.d. point processes on $\mathbb{R}_-$ distributed according to $\zeta_{N_-}$ and $(N^i_+)_{i=1,\dots,n}$ is a family of point processes on $\mathbb{R}_+$ with $\mathbb{F}$-intensity given for all $i = 1, \dots, n$ by*

$$\lambda^i_t = \Psi\left(S^i_{t-}, \frac{1}{n}\sum_{j=1}^{n}\left(\int_0^{t-} H_{ij}(t-z)N^j_+(dz) + F_{ij}(t)\right)\right), \tag{10}$$

*where $(S^i_{t-})_{t\geq 0}$ is the predictable age process associated with $N^i$ defined in (4) and $(H_{ij})_{1\leq i,j\leq n}$ (respectively $(F_{ij})_{1\leq i,j\leq n}$) is a random matrix with entries distributed according to $\mu_H$ (resp. $\nu_F$) and satisfying (8) (resp. (9)).*

An age dependent random Hawkes process admits two different behaviours:

1. before time 0, the processes $(N^i_-)_{i=1,\dots,n}$ are independent and identically distributed;

2. after time 0, the processes $(N^i_+)_{i=1,\dots,n}$ are dependent (in general) and driven by their respective intensities which can be different from one process to another.

**Remark II.2.**   *1. The dichotomy of behaviours can model a change of regime at time $t = 0$. It should be interesting to see whether the results could be extended to initial conditions given by a mean-field dynamics and not necessarily i.i.d. ones. However, it is not in the scope of this article.*

*2. Assumptions (8) and (9) mean that a single particle receives i.i.d. interactions from its neighbours and that the particles are exchangeable: one can permute the particles without modifying their joint distribution.*

*3. Given $\mathcal{F}_0$, the randomness of $\lambda^i_t$ in Equation (10) only lies in the point measures $N^j_+(dz)$ and the age process $(S^i_{t-})_{t\geq 0}$. These intensities, and so the point processes, are not exchangeable given $\mathcal{F}_0$. However, they are exchangeable when they are considered with respect to all the randomness (including the $N^i_-$'s, $H_{ij}$'s and $F_{ij}$'s).*

*4. The link between synaptic weights (that is the strength with which one given neuron influences an other one) and interaction functions can be well emphasized by the following choice of interaction functions. Consider a fixed function $h : \mathbb{R}_+ \to \mathbb{R}$ and, independently of everything else, a sequence $(\alpha_j)_{j=1,\dots,n}$ of i.i.d. random variables with values in $[0,1]$. Then, $(H_{ij})_{1\leq i,j\leq n}$ defined by $H_{ij} = \alpha_j h$ satisfies (8). The $\alpha_j$'s represent the (relative) synaptic weight of neuron $j$ over all the other ones.*

*The interaction functions, even if they are random, are fixed at time 0. The dynamics of synaptic weights is not taken into account here.*

*5. As presented in the introduction (see Equation (7)), a particular case we have in mind in this study is when there exists a function $\Phi : \mathbb{R} \to \mathbb{R}_+$ and a non-negative real number $\delta$ such that*

$$\Psi(s,x) = \Phi(x)\mathbb{1}_{s\geq\delta}. \tag{11}$$

*This particular choice of $\Psi$ provides an interesting modelling of the strict refractory period of a neuron. Furthermore, when $\delta = 0$, there is no refractory period and one recovers more standard Hawkes processes. In particular, if $\mu_H$ is the Dirac mass located at some fixed function $h$ and $\nu_F$ is the Dirac mass located at the null function, then one recovers the Hawkes processes studied in [13]. Remark that the exchangeability of the Hawkes processes studied in [13] is obvious since they have the same intensity at each time $t$.*

*6. Since the auto-interaction given by $H_{ii}$ is scaled by $1/n$, it vanishes when $n$ goes to infinity and so the asymptotic behaviour proved in this article (Corollary IV.4) remains the same if one assumes that $H_{ii} = 0$.*

7. *Unlike the matrix* $\mathbf{H}$, *the matrix* $\mathbf{F}$ *can depend on the initial conditions as it can be seen in the following particular case which is derived from* (6) *for instance. For the same matrix* $\mathbf{H} = (H_{ij})_{1 \leq i,j \leq n}$ *as in* (8), *we may choose for all* $1 \leq i, j \leq n$, *the function* $F_{ij} : \mathbb{R}_+ \to \mathbb{R}$ *defined for all* $t \geq 0$ *by*

$$F_{ij}(t) = \int_{-\infty}^{0} H_{ij}(t - z) N_{-}^{j}(dz). \tag{12}$$

*These random functions are* $\mathcal{F}_0$-*measurable and they satisfy the first two lines of* (9) *thanks to the independence of the* $H_{ij}$'s *and the* $N_{-}^{j}$'s. *Hence, one can consider the intensity given by* (10) *with such a choice of* $F$ *to represent the contribution of the processes* $(N_{-}^{i})_{i=1,\ldots,n}$ *to the dynamics after time* 0. *In this example, the* $F_{ij}$'s *are obviously dependent from the* $N_{-}^{j}$'s.

8. *In the case of neurons modelling, one can model external inputs via the functions* $F_{ij}$. *For example, one could take* $F_{ij} = H_{ij}(t - \tau)$ *where* $\tau$ *is some non-positive real number that may be random (independent of anything else) modelling that all the neurons have spiked at the same time* $\tau < 0$ *thanks to a common stimulus.*

## II.3 List of assumptions

In the present article, several assumptions on the parameters of the model are used depending on the context. For sake of simplicity, all these assumptions are gathered here.

**Main assumptions**

$\left(\mathcal{A}_{u^{\text{in}}}^{\zeta_{N_-}}\right)$: If $N_-$ is distributed according to $\zeta_{N_-}$ ($N_- \sim \zeta_{N_-}$) and $T_0$ denotes the closest point of $N_-$ to 0, then $-T_0$ admits a density with respect to the Lebesgue measure denoted by $u^{\text{in}}$ ("in" stands for "initial"). Furthermore, $u^{\text{in}}$ is uniformly bounded.

$\left(\mathcal{A}_{\infty}^{\mu_H}\right)$: If $H \sim \mu_H$, then there exists a deterministic function $G : \mathbb{R}_+ \to \mathbb{R}$ such that a.s., for all $t \geq 0$, $|H(t)| \leq G(t)$. The smallest possible deterministic function $G$, denoted by $M_{\mu_H}$, is moreover supposed to be locally integrable. In particular, $\mathbb{E}[H(t)]$ is well-defined and we let $m_{\mu_H}(t) := \mathbb{E}[H(t)]$.

$\left(\mathcal{A}_{1}^{\nu_F}\right)$: If $F \sim \nu_F$, then $t \in \mathbb{R}_+ \mapsto \mathbb{E}[|F(t)|]$ is locally bounded. In particular, for all $t \geq 0$, $\mathbb{E}[F(t)]$ is well-defined and we let $m_{\nu_F}(t) := \mathbb{E}[F(t)]$.

$\left(\mathcal{A}_{\text{Lip}}^{\Psi}\right)$: The function $\Psi : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$ is uniformly Lipschitz continuous with respect to the second coordinate: there exists a constant $C > 0$ such that, for all $s \geq 0$, the function $x \mapsto \Psi(s, x)$ is Lipschitz with constant $C$. The smallest constant $C$ is denoted by $\text{Lip}(\Psi)$. Furthermore, $s \in \mathbb{R}_+ \mapsto \Psi(s, 0)$ is uniformly bounded.

$\left(\mathcal{A}_{\infty}^{\Psi}\right)$: The function $\Psi$ is uniformly bounded, that is $||\Psi||_\infty < +\infty$.

$\left(\mathcal{A}_{\Psi = \Psi_0}\right)$: There exists a function $\Psi_0 : \mathbb{R} \to \mathbb{R}_+$ such that, for all $s \geq 0$, $\Psi(s, \cdot) = \Psi_0(\cdot)$. In this case, if $\left(\mathcal{A}_{\text{Lip}}^{\Psi}\right)$ is satisfied then $\text{Lip}(\Psi)$ is rather denoted by $\text{Lip}(\Psi_0)$.

**Additional assumptions**

$\left(\mathcal{A}_{\infty}^{\zeta_{N_-}}\right)$: If $N_- \sim \zeta_{N_-}$ and $T_0$ denotes the closest point of $N_-$ to 0, then $-T_0$ is upper bounded a.s. that is there exists a constant $C > 0$ such that $-T_0 \leq C$ a.s. The smallest possible constant $C$ is denoted by $M_{T_0}$.

$\left(\mathcal{A}_{\infty,2}^{\mu_H}\right)$: $\left(\mathcal{A}_{\infty}^{\mu_H}\right)$ is satisfied and $M_{\mu_H}$ is furthermore locally square integrable.

$\left(\mathcal{A}_{2}^{\nu_F}\right)$: $\left(\mathcal{A}_{1}^{\nu_F}\right)$ is satisfied and if $F \sim \nu_F$, then for all $t \geq 0$, $F(t)$ admits a variance denoted by $V_{\nu_F}(t)$ satisfying that for all $t \geq 0$, $\int_0^t V_{\nu_F}(t')^{1/2} dt' < +\infty$. Furthermore, $m_{\nu_F}$ is a continuous function.

Remark that:

- Assumptions $\left(\mathcal{A}_{\infty}^{\mu_H}\right)$, $\left(\mathcal{A}_{1}^{\nu_F}\right)$ and $\left(\mathcal{A}_{\text{Lip}}^{\Psi}\right)$ are used to prove the existence of the $n$-particle system.

- Assumptions $(\mathcal{A}_{\infty,2}^{\mu_H})$, $(\mathcal{A}_2^{\nu_F})$ and $(\mathcal{A}_{\text{Lip}}^{\Psi})$ are used to prove the mean-field approximation under the addition of either $(\mathcal{A}_{\infty}^{\Psi})$ and $(\mathcal{A}_{u^{\text{in}}}^{\zeta_{N_-}})$ or $(\mathcal{A}_{\Psi=\Psi_0})$.

## II.4   Representation via a stochastic differential equation

Definition II.1 describes ADRHPs as weak solutions. It characterizes their distribution but not their path-wise dynamics. As it is well emphasized in [30], point processes can be either represented as weak solutions thanks to their stochastic intensity or represented as strong solutions of a stochastic differential equation (SDE) driven by Poisson noise. The idea to represent point processes as *strong* solutions of SDEs driven by Poisson measures was first introduced by Lewis and Shedler [27], for inhomogeneous Poisson process, and extended by Ogata [34] (thinning procedure) under some weak assumptions on the intensity. It says that, if $N$ admits $(\lambda_t)_{t \geq 0}$ as a $\mathbb{F}$-predictable intensity, then the point measure associated with $N$ can be represented by $N(dt) = \Pi(dt \times [0, \lambda_t])$ where $\Pi$ is a Poisson measure with intensity 1 on $\mathbb{R}_+^2$. This has been used to show existence or stability results for some classes of point processes by Brémaud and Massoulié in [5] or [30] and more recently to exhibit some suitable coupling between interacting Hawkes processes and their mean field approximation in [13]. We introduce here the representation of ADRHPs based on such a thinning procedure.

**Representation II.3.** *Let $(N_-^i)_{i \geq 1}$ be some i.i.d. point processes on $\mathbb{R}_-$ distributed according to $\zeta_{N_-}$. Let $(H_{ij})_{1 \leq i,j \leq n}$ (respectively $(F_{ij})_{1 \leq i,j \leq n}$) be a random matrix with entries distributed according to $\mu_H$ (resp. $\nu_F$) and satisfying (8) (resp. (9)). Let $(\Pi^i(dt, dx))_{i \geq 1}$ be some i.i.d. $\mathbb{F}$-Poisson measures with intensity 1 on $\mathbb{R}_+^2$.*

*Let $(N_t^i)_{t \geq 0}^{i=1,..,n}$ be a family of counting processes such that, for $i = 1,..,n$ and all $t \geq 0$,*

$$
N_t^i = \int_0^t \int_0^\infty \mathbb{1}_{\left\{ x \leq \Psi\left( S_{t'-}^i, \frac{1}{n} \sum_{j=1}^n \left( \int_0^{t'-} H_{ij}(t'-z)N_+^j(dz) + F_{ij}(t') \right) \right) \right\}} \Pi^i(dt', dx), \quad (13)
$$

*where $(S_{t-}^i)_{t \geq 0}$ is the predictable age process associated with $N^i = N_-^i \cup N_+^i$ and $N_+^i$ is the point process associated with the counting process $(N_t^i)_{t \geq 0}$. Then, $(N^i)_{i=1,..,n}$ is an age dependent random Hawkes process with parameters $(n, \mu_H, \nu_F, \Psi, \zeta_{N_-})$.*

This representation is mainly used in this paper in order to provide a suitable coupling between ADRHPs and i.i.d. point processes describing the mean-field dynamics.

Going back and forth between the weak solution of Definition II.1 and the strong solution of Representation II.3 is classic: the thinning Theorem (see [5, Lemma 2] or [8, Theorem B.11] for a complete proof) states that a strong solution is also a weak solution; and the Poisson inversion [5, Lemma 4] states that, from a weak solution $(N^i)_{i=1,...,n}$, one can construct Poisson measures on an enlarged probability space such that (13) is satisfied.

At this stage, one has two equivalent concepts of ADRHPs but no result on the existence of such processes. Indeed, if there is too much self-excitation, then there may be an infinite number of points in finite time. In the present paper, point processes that do not explode in finite time are considered and, thanks to Representation II.3, one can prove existence of these non-explosive processes.

**Proposition II.4.** *Under $(\mathcal{A}_{\infty}^{\mu_H})$, $(\mathcal{A}_1^{\nu_F})$ and $(\mathcal{A}_{\text{Lip}}^{\Psi})$, there exists an ADRHP $(N^i)_{i=1,..,n}$ with parameters $(n, \mu_H, \nu_F, \Psi, \zeta_{N_-})$ such that $t \mapsto \mathbb{E}\left[N_t^1\right]$ is locally bounded.*

This result can be challenging in an infinite dimensional framework like in [13]. However, in our finite dimensional framework, it is quite clear since the ADRHP can be stochastically dominated by some multivariate *linear* Hawkes process (thanks to the Lipschitz assumption $(\mathcal{A}_{\text{Lip}}^{\Psi})$). Yet the well-posedness of linear Hawkes processes is standard thanks to their branching structure [23]. Nevertheless, a proof of Proposition II.4 is given in Appendix A.1.

# III   Study of the limit dynamics

The interactions between the point processes involved in the definition of an ADRHP are of mean-field type. Therefore, the limit version of Equation (13) is proposed below in Equation (14)

(informally, the empirical means involved in (13) are replaced by their expected values). The *limit equation* with parameters $(h, f_0, \Psi, \zeta_{N_-})$ is given by

$$\forall t > 0, \; \overline{N}_t = \int_0^t \int_0^\infty \mathbb{1}_{\left\{ x \leq \Psi \left( \overline{S}_{t'-}, \int_0^{t'-} h(t'-z) \mathbb{E}\left[ \overline{N}_+(dz) \right] + f_0(t') \right) \right\}} \Pi(dt', dx), \quad (14)$$

where $h$ and $f_0$ are some functions from $\mathbb{R}_+$ to $\mathbb{R}$, $\Pi(dt', dx)$ is an $\mathbb{F}$-Poisson measure on $\mathbb{R}_+^2$ with intensity 1 and $(\overline{S}_{t-})_{t \geq 0}$ is the predictable age process associated with $\overline{N} = \overline{N}_- \cup \overline{N}_+$ where $\overline{N}_-$ is a point process distributed according to $\zeta_{N_-}$ and $\overline{N}_+$ is the point process associated with the counting process $(\overline{N}_t)_{t \geq 0}$.

Looking simultaneously at Equations (13) and (14) shows that the empirical mean of the random interaction functions $H_{ij}$ (*respectively the random functions* $F_{ij}$) are replaced by $h$ (*resp.* $f_0$) which should be the mean interaction function $m_{\mu_H}$ (*resp.* $m_{\nu_F}$). Moreover, the empirical mean of the point measures $N_+^j(dz)$ in (13) is replaced by the expectation of the point measure $\overline{N}_+(dz)$.

Finally, let us note that the dependence with respect to the age process is still present in the limit equation. This matches with experimental data in neuroscience where refractory periods are highlighted. By comparison, there is no such dependence in the limit process given in [13] which is an inhomogeneous Poisson process.

This limit equation is used in the next section to provide suitable couplings to prove the mean-field approximation. Hence, the main point of this section is to prove the well-posedness of the limit equation (14). However, to study the probabilistic formulation of the mean-field dynamics described in Equation (14), one first needs to find a representation of the distribution of a possible solution of (14). As a first step, we prove existence/uniqueness results for a linearisation of the (PPS) system (Proposition III.1) as well as we give a representation of the solution given by the method of characteristics (Proposition III.2). The second step is to deduce existence/uniqueness results for the (PPS) system (Theorem III.5) from the linearised system via a fixed point argument. Then, the well-posedness of the limit equation (14) is proved thanks to the results obtained for the (PPS) system. Finally, the link between the (PPS) system and the processes defined by the limit equation is fully investigated.

Note that the analytical study of the fixed point equation satisfied by the expectation of the solution of (14) (as it is done in [13]) can be extended to the case when the intensity does depend on the age. However, the results for the (PPS) system are valid in a more general framework so they are favoured here.

## III.1 Study of the linear system

In comparison with the (PPS) system, the linear system studied below corresponds to the case where the firing rate $p$ in (PPS) is a function of the time $t$ and the age $s$ only. More precisely, we consider the system

$$\begin{cases} \dfrac{\partial u(t, s)}{\partial t} + \dfrac{\partial u(t, s)}{\partial s} + f(t, s) u(t, s) = 0, \\ u(t, 0) = \displaystyle\int_{s \in \mathbb{R}} f(t, s) u(t, s) \, ds, \end{cases} \quad (15)$$

where $f$ is a bounded function.

We state uniqueness of the solution of this system in a measure space as a consequence of the uniqueness result stated in [7]. More precisely, the result is stated in $\mathcal{BC}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}))$ that is the space of bounded continuous curves on $\mathcal{M}(\mathbb{R})$ (the space of Radon measures on $\mathbb{R}$) endowed with the *bounded Lipschitz norm* as considered in [7]. As we are interested in probability measures, let us remark that the bounded Lipschitz norm on $\mathcal{P}(\mathbb{R}_+)$ is equivalent, thanks to the duality of Kantorovich-Rubinstein, to the modified 1-Wasserstein distance defined by

$$\tilde{W}_1(\mu, \nu) := \inf \mathbb{E}\left[ \min(|X - Y|, 1) \right], \quad (16)$$

where the infimum is taken over all joint distributions of the random variables $X$ and $Y$ with marginals $\mu$ and $\nu$ respectively.

Since measure solutions are considered, a weak form of the system is given. The following set of test functions is used:

$\mathcal{C}_{c,b}^{\infty}(\mathbb{R}_+^2)$ 

The function $\varphi$ belongs to $\mathcal{C}_{c,b}^{\infty}(\mathbb{R}_+^2)$ if
- $\varphi$ is continuous, uniformly bounded,
- $\varphi$ has uniformly bounded derivatives of every order,
- there exists $T > 0$ such that $\varphi(t, s) = 0$ for all $t > T$ and $s \geq 0$.

The result stated below is a consequence of [7, Theorem 2.4.] in the same essence than the one presented in [7, Section 3.3.]. Its proof is given in Appendix A.2.

**Proposition III.1.** *Assume that $f : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ is bounded and continuous (uniformly in the first variable) with respect to the second variable. Assume that $u^{\mathrm{in}}$ belongs to $\mathcal{M}(\mathbb{R}_+)$.*

*Then, there exists a unique solution in the weak sense $u$ such that $t \mapsto u(t, \cdot)$ belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}_+))$ of the system* (15) *with initial condition $u(0, \cdot) = u^{\mathrm{in}}$. The weak sense means here that for every $\varphi$ in $\mathcal{C}_{c,b}^{\infty}(\mathbb{R}_+^2)$,*

$$\int_{\mathbb{R}_+^2} \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right) \varphi(t, s) \, u(t, ds) \, dt + \int_{\mathbb{R}_+} \varphi(0, s) u^{\mathrm{in}}(ds)$$
$$+ \int_{\mathbb{R}_+^2} [\varphi(t, 0) - \varphi(t, s)] f(t, s) u(t, ds) dt = 0. \quad (17)$$

Remark that the system is mass-conservative (e.g. take a sequence of functions converging to $t \mapsto \mathbb{1}_{[0,T]}(t)$ as test functions in the weak equation (17)). As we are interested in probability measures as solutions, let us remark that the mass-conservation alone cannot ensure that the solution is a probability even if the initial condition is a probability. However, when the initial condition is a probability which admits a density, the method of characteristics shows that the solution of (15) is probability density function for all time $t \geq 0$.

**Proposition III.2.** *Under the assumptions of Proposition III.1, assume that $u^{\mathrm{in}}$ is a probability which admits a density (denoted by $u^{\mathrm{in}}$ as well) with respect to the Lebesgue measure. Then, there exists a unique locally bounded function $u_0 : \mathbb{R}_+ \to \mathbb{R}$ (which is furthermore non-negative) such that $u$ defined by*

$$\begin{cases} u(t, s) = u^{\mathrm{in}}(s - t) \exp\left( - \int_0^t f(t', s - t + t') dt' \right), & \text{for } s \geq t \quad (18) \\[2mm] u(t, s) = u_0(t - s) \exp\left( - \int_0^s f(t - s + s', s') ds' \right), & \text{for } t \geq s \quad (19) \end{cases}$$

*is the unique solution of* (15). *In particular,*

- *$u$ satisfies the second equation of* (15) *in a strong sense,*

- *since $u_0$ is non-negative and the system is mass-conservative, the function $u(t, \cdot)$ is a density for all time $t \geq 0$.*

A detailed proof of this result is given in Appendix A.3. Here are listed some properties of the solution $u$ of the linear system.

**Proposition III.3.** *Under the assumptions of Proposition III.2, assume furthermore that there exists $M > 0$ such that for all $s \geq 0$, $0 \leq u^{\mathrm{in}}(s) \leq M$.*

*Then, the solution $u$ of* (15) *is such that the function $t \mapsto u(t, \cdot)$ belongs to $\mathcal{C}(\mathbb{R}_+, L^1(\mathbb{R}_+))$, the function $t \mapsto u(t, 0)$ is continuous and*

$$0 \leq u(t, s) \leq \max(M, ||f||_\infty), \quad \text{for all } t, s \geq 0. \quad (20)$$

*Proof.* The first continuity property is rather classic thanks to a fixed point argument in the space $\mathcal{C}([0, T], L^1(\mathbb{R}_+))$ for a good choice of $T > 0$ (see [39, Section 3.3.] for instance). The second one is given by the second equation of (15) (which is satisfied in a strong sense by the solution given by the characteristics). Indeed,

$$\begin{aligned} |u(t + t', 0) - u(t, 0)| &\leq \int_0^{+\infty} f(t + t', s) |u(t + t', s) - u(t, s)| ds \\ &\quad + \int_0^{+\infty} |f(t + t', s) - f(t, s)| u(t, s) ds \\ &\leq ||f||_\infty ||u(t + t', \cdot) - u(t, \cdot)||_1 + \sup_{s \geq 0} |f(t + t', s) - f(t, s)|. \end{aligned}$$

10

Yet, the continuity properties of both functions f and $t \mapsto u(t, \cdot)$ and give that $|u(t + t', 0) - u(t, 0)|$ goes to 0 as $t'$ goes to 0, hence the continuity of $t \mapsto u(t, 0)$.

Finally, one can prove that $u$ satisfies (20) thanks to the representation given by the characteristics. On the one hand, the function $u_0$ given in Proposition III.2 is non-negative and so is $u$. On the other hand, it follows from (18) that for $s \geq t$, $u(t, s) \leq M$ and it follows from the second equation of (15) that for all $t \geq 0$, $u(t, 0) \leq ||f||_\infty$ and so (19) implies that for $t \geq s$, $u(t, s) \leq ||f||_\infty$. □

## III.2   Study of the (PPS) system

Here, a global existence result for the nonlinear system (PPS) is stated under suitable assumptions. Since this result is one of the cornerstone of this work, its proof is given even if its sketch is pretty similar to the proof of [36, Theorem 5.1].

In comparison with the uniqueness for the linear system which takes place in $\mathcal{BC}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}_+))$, the uniqueness result stated in this section takes place in $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$. However, this last result is sufficient for our purpose since it is applied to measures that are probabilities a priori.

First of all, a technical lemma is needed to fully understand the non linearity involved in the system (PPS).

**Lemma III.4.** *Under ($\mathcal{A}^{\Psi}_{\text{Lip}}$) and ($\mathcal{A}^{\Psi}_{\infty}$), assume that $h : \mathbb{R}_+ \to \mathbb{R}$ is locally integrable and that $f_0 : \mathbb{R}_+ \to \mathbb{R}$ is continuous. Then, for all $u$ in $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ there exists a unique function $X_u : \mathbb{R}_+ \mapsto \mathbb{R}$ such that*

$$X_u(t) = \int_{z=0}^{t} \int_{s=0}^{+\infty} h(t - z) \Psi\left(s, X_u(z) + f_0(z)\right) u\left(z, ds\right) dz. \tag{21}$$

*Furthermore, the function $X_u$ is continuous.*

*Proof.* The proof is divided in three steps:

-1. Establish a priori estimates on $X_u$ to show that it is locally bounded. Indeed, using the boundedness of $\Psi$ and the fact that $u$ belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$, one deduces that

$$X_u(t) \leq ||\Psi||_\infty \int_{z=0}^{t} \int_{s=0}^{+\infty} |h(t - z)| u\left(z, ds\right) dz \leq ||\Psi||_\infty \int_{z=0}^{t} |h(t - z)| dz.$$

Hence, the local integrability of $h$ implies the local boundedness of $X_u$.

-2. Show that $X_u$ exists and is unique as a fixed point. For any $T > 0$, consider the following application

$$
\begin{aligned}
G_T : \quad L^\infty([0, T]) &\longrightarrow && L^\infty([0, T]) \\
X &\longmapsto && \left(t \mapsto \int_{z=0}^{t} \int_{s=0}^{+\infty} h(t - z) \Psi\left(s, X(z) + f_0(z)\right) u\left(dz, ds\right)\right).
\end{aligned}
$$

The Lipschitz continuity of $\Psi$ leads, for any $X_1, X_2$ in $L^1([0, T])$ and $t$ in $[0, T]$, to

$$
\begin{aligned}
|G_T(X_1)(t) - G_T(X_2)(t)| &\leq \text{Lip}(\Psi) \int_0^t |h(t - z)| \, |X_1(z) - X_2(z)| \int_{s=0}^{+\infty} u(z, ds) dz \\
&\leq \text{Lip}(\Psi) \int_0^t |h(t - z)| \, |X_1(z) - X_2(z)| \, dz \\
&\leq \text{Lip}(\Psi) ||X_1 - X_2||_{L^\infty([0,T])} \int_0^T |h(z)| dz.
\end{aligned}
$$

Fix $T > 0$ such that $\text{Lip}(\Psi) \int_0^T h(z) dz \leq 1/2$ so that $G_T$ is a contraction and admits a unique fixed point. Iterating this fixed point gives the existence and uniqueness of $X_u$ in the space of locally bounded functions.

For instance, we give the idea for the first iteration. Denoting $W$ the fixed point of $G_T$, one can consider the following application

$$
\begin{aligned}
G_{2T}^W : \quad L^\infty([T, 2T]) &\longrightarrow && L^\infty([T, 2T]) \\
X &\longmapsto && \left(t \mapsto \int_{z=0}^{t} \int_{s=0}^{+\infty} h(t - z) \Psi\left(s, \tilde{X}(z) + f_0(z)\right) u\left(z, ds\right) dz\right),
\end{aligned}
$$

where $\tilde{X}(t) = W(t)$ if $0 \le t < T$, $\tilde{X}(t) = X(t)$ if $T \le t \le 2T$, and $\tilde{X}(t) = 0$ otherwise. Applying the same argument as for the fixed point of $G_T$ leads to existence and uniqueness of the trace of $X_u$ on $[0, 2T]$.

-3. Finally, let us show that $X_u$ is continuous thanks to a generalized Gronwall lemma. Using the Lipschitz continuity and the boundedness of $\Psi$, one deduces from (21) that

$$|X_u(t+t') - X_u(t)| \le \mathrm{Lip}(\Psi) \int_0^t |h(y)| |(X_u + f_0)(t - y) - (X_u + f_0)(t + t' - y)| dy$$

$$+ ||\Psi||_\infty \int_t^{t+t'} |h(y)| dy.$$

This means that the function $Y_u^{(t')} := |X_u(\cdot + t') - X_u(\cdot)|$ satisfies

$$Y_u^{(t')} \le g^{(t')}(t) + \mathrm{Lip}(\Psi) \int_0^t |h(t-z)| Y_u^{(t')}(z) dz$$

where $g^{(t')}(t) := \mathrm{Lip}(\Psi) \int_0^t |h(t-z)| |f_0(z+t') - f_0(z)| dz + ||\Psi||_\infty \int_t^{t+t'} |h(y)| dy$. Applying Lemma B.4-(i) gives, for any $T > 0$, $\sup_{t \in [0,T]} Y_u^{(t')}(t) \le C_T \sup_{t \in [0,T]} g^{(t')}(t)$. Yet the continuity (hence uniform continuity on compact time intervals) of $f_0$ and the local integrability of $h$ gives that $\sup_{t \in [0,T]} g^{(t')}(t) \to 0$ as $t' \to 0$. □

Now, we have all the ingredients to state the existence/uniqueness result for the (PPS) system in a measure space of possible solutions. Notice that the existence/uniqueness result used for the linear system would not directly apply to the non-linear system. In that sense, we extend the result stated in [7].

**Theorem III.5.** *Under $(\mathcal{A}_{\mathrm{Lip}}^\Psi)$ and $(\mathcal{A}_\infty^\Psi)$, assume that $h : \mathbb{R}_+ \to \mathbb{R}$ is locally integrable and that $f_0 : \mathbb{R}_+ \to \mathbb{R}$ is continuous. Assume that $u^{\mathrm{in}}$ is a non-negative function such that both $\int_0^{+\infty} u^{\mathrm{in}}(s) ds = 1$ and there exists $M > 0$ such that for all $s \ge 0$, $0 \le u^{\mathrm{in}}(s) \le M$.*

*Then, there exists a unique solution in the weak sense $u$ such that $t \mapsto u(t, \cdot)$ belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ of the following (PPS) system*

$$\begin{cases} \dfrac{\partial u(t,s)}{\partial t} + \dfrac{\partial u(t,s)}{\partial s} + \Psi(s, X(t) + f_0(t)) u(t,s) = 0, \\ u(t,0) = \displaystyle\int_{s \in \mathbb{R}} \Psi(s, X(t) + f_0(t)) u(t,s) ds, \end{cases} \tag{22}$$

*with initial condition that $u(0, \cdot) = u^{\mathrm{in}}$, where for all $t \ge 0$, $X(t) = \int_0^t h(t-z) u(z, 0) dz$. The weak sense means here that for every $\varphi$ in $\mathcal{C}_{c,b}^\infty(\mathbb{R}_+^2)$,*

$$\int_{\mathbb{R}_+^2} \left( \frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right) \varphi(t,s) u(t,ds) dt + \int_{\mathbb{R}_+} \varphi(0,s) u^{\mathrm{in}}(s) ds$$

$$+ \int_{\mathbb{R}_+^2} [\varphi(t,0) - \varphi(t,s)] \Psi(s, X(t) + f_0(t)) u(t,ds) dt = 0 \tag{23}$$

*where $X$ is the continuous function given by Lemma III.4 and satisfying*

$$X(t) = \int_{z=0}^t \int_{s=0}^{+\infty} h(t-z) \Psi(s, X(z) + f_0(z)) u(z, ds) dz. \tag{24}$$

*Moreover, the solution $u$ is such that, for all $t \ge 0$, the measure $u(t, \cdot)$ is a probability and admits a density which is identified to the solution itself. Furthermore, the function $t \mapsto u(t, \cdot)$ belongs to $\mathcal{C}(\mathbb{R}_+, L^1(\mathbb{R}_+))$, the function $t \mapsto u(t, 0)$ is continuous and*

$$0 \le u(t,s) \le \max(M, ||\Psi||_\infty), \quad \text{for all } t, s \ge 0. \tag{25}$$

As for the linear case, the system is mass-conservative (e.g. take a sequence of functions converging to $t \mapsto \mathbb{1}_{[0,T]}(t)$ as test functions in the weak equation (23)).

12

*Proof.* The proof is divided in two steps. First, we apply the results of Section III.1 to a linearised version of the non-linear system (22) and then we find the auxiliary function $X$ corresponding to the solution $u$ as a fixed point in a space of continuous functions in order to deal with the non linearity of the system (22).

-1. The linearised version of the system takes the form

$$\begin{cases} \dfrac{\partial u\,(t,s)}{\partial t} + \dfrac{\partial u\,(t,s)}{\partial s} + \Psi\,(s,Y(t)+f_0(t))\,u\,(t,s) = 0, \\ u\,(t,0) = \displaystyle\int_{s\in\mathbb{R}} \Psi\,(s,Y(t)+f_0(t))\,u\,(t,s)\,ds, \end{cases} \tag{26}$$

for a fixed continuous function $Y$. Note that the funciton $f:(t,s)\mapsto \Psi\,(s,Y(t)+f_0(t))$ is bounded and continuous (uniformly in $s$) with respect to $t$. So the assumptions of Propositions III.1, III.2 and III.3 are satisfied. In particular, for any continuous function $Y$, there exists a unique solution $u_Y$ (with initial condition $u^{\mathrm{in}}$) in $\mathcal{BC}(\mathbb{R}_+,\mathcal{P}(\mathbb{R}_+)) \subset \mathcal{BC}(\mathbb{R}_+,\mathcal{P}(\mathbb{M}_+))$ of the system (26) which furthermore satisfies the properties listed in Proposition III.3.

-2. Let us notice that for all $T>0$, if $Y$ belongs to the Banach space $(\mathcal{C}([0,T]),||.||_{\infty,[0,T]})$ then $t\mapsto \int_0^t h(t-z)u_Y(z,0)dz$ belongs to $\mathcal{C}([0,T])$ too. Indeed, remind that $t\mapsto u_Y(t,0)$ is continuous thanks to Proposition III.3 and that $\int_0^t h(t-z)u_Y(z,0)dz = \int_0^t h(z)u_Y(t-z,0)dz$. Hence one can consider the application

$$\begin{array}{rcl} F_T: \quad \mathcal{C}([0,T]) & \longrightarrow & \mathcal{C}([0,T]) \\ Y & \longmapsto & \left(t\mapsto \displaystyle\int_0^t h(t-z)u_Y(z,0)dz\right), \end{array} \tag{27}$$

and show that it admits a fixed point for a good choice of $T$. Computations given in Appendix A.4 provide the following statement

$$\exists T>0, \forall Y_1, Y_2 \in L^\infty([0,T]), \quad ||F_T(Y_1)-F_T(Y_2)||_{L^\infty([0,T])} \leq \frac{1}{2}||Y_1-Y_2||_{L^\infty([0,T])}, \tag{28}$$

where $T$ depends neither on the initial condition nor on $f_0$.

Until the end of the proof, let fix such a $T$. Then, there exists a unique $W$ in $\mathcal{C}([0,T])$ such that $F_T(W)=W$. In particular, $u_W$ is a solution of (22) on $[0,T]$ so we have existence on $[0,T]$.

For the uniqueness, let us consider the trace of a solution $u \in \mathcal{BC}(\mathbb{R}_+,\mathcal{P}(\mathbb{R}_+))$ to (22) on $[0,T]$. Then, the auxiliary function $X_u$ associated with $u$ defined in Lemma III.4 is continuous. Since $u$ is a solution of (22), the trace of $X_u$ on $[0,T]$ is a fixed point of $F_T$ and so $X_u=W$ and $u=u_W$. This gives the uniqueness of the solution of (22) on $[0,T]$ in $\mathcal{BC}(\mathbb{R}_+,\mathcal{P}(\mathbb{R}_+))$.

Taking $u_W(T,\cdot)$ instead of $u^{\mathrm{in}}$ as initial condition, the function $t\mapsto f_0(t+T)+\int_0^T h(t-z)u_W(z,0)dz$ instead of $f_0$ and applying the same find of fixed point argument gives the trace of the solution on $[T,2T]$. Iterating this fixed point argument, one deduces that there exists a unique solution $u$ of (22) on $\mathbb{R}_+$ (remind that $T$ depends neither on the initial condition nor on $f_0$). In particular, the iteration is possible since the boundedness of the initial condition is carried on by the equation (see Equation (20)).

The regularity and boundedness of the solution $u$, i.e. the continuity properties and Equation (25) listed at the end of the statement, come from the regularity and boundedness of the solutions $u_Y$ since $u$ is one of the $u_Y$'s. $\qquad\square$

## III.3  Limit process

Let us remind the limit equation (14), that is

$$\forall t>0,\ \overline{N}_t = \int_0^t \int_0^\infty \mathbb{1}_{\left\{x\leq \Psi\left(\overline{S}_{t'-},\, \int_0^{t'-} h(t'-z)\mathbb{E}\left[\overline{N}_+(dz)\right]+f_0(t')\right)\right\}} \Pi(dt',dx).$$

This limit equation describes an age dependent point process interacting with its own mean intensity in an Hawkes manner. More precisely, a solution $(\overline{N}_t)_{t\geq 0}$ of (14), if it exists, admits

13

an intensity $\overline{\lambda}_t$ which depends on the time $t$ and the age $\overline{S}_{t-}$ in a McKean-Vlasov manner in the sense that it satisfies the following implicit equation $\overline{\lambda}_t = \Psi(\overline{S}_{t-}, \int_0^{t'-} h(t'-z)\mathbb{E}\left[\overline{\lambda}_z\right]dz + f_0(t'))$. Equation (14) is in particular a non trivial fixed point problem of the McKean-Vlasov type. Notice that the dependence of $\overline{N}$ with respect to the past $\overline{N}_-$ is reduced to the age at time 0, i.e. $\overline{S}_{0-} = -T_0$. Throughout this article, a solution of the limit equation is called a *point process of the McKean-Vlasov type whose intensity depends on time and on the age.*

The fixed point problem of the limit equation (14) is proved to be well-posed in two cases as given in the next two statements. In either case, the idea of the proof is first to compute the mean intensity (denoted by $\overline{\lambda}(t)$) of a possible solution of (14). The next two propositions state the same result under different sets of assumptions and can be summarized as follows:

- in the first case, the intensity is bounded, i.e. $\Psi$ satisfies $(\mathcal{A}_\infty^\Psi)$, and the mean intensity is given by the system (22). More precisely, $\overline{\lambda}(t) = u(t,0)$ with $u$ given by Theorem III.5.

- in the second case, the intensity does not depend on the age process, i.e. $\Psi$ satisfies $(\mathcal{A}_{\Psi=\Psi_0})$, and the mean intensity is given by a generalization of [13, Lemma 24].

**Proposition III.6.** *Under $(\mathcal{A}_{u^{\text{in}}}^{\zeta_{N_-}})$, $(\mathcal{A}_{\text{Lip}}^\Psi)$ and $(\mathcal{A}_\infty^\Psi)$, assume that $h : \mathbb{R}_+ \to \mathbb{R}$ is locally integrable and that $f_0 : \mathbb{R}_+ \to \mathbb{R}$ is continuous. Denote by $u$ the unique solution of (22) with initial condition $u^{\text{in}}$ as given by Theorem III.5 and let, for all $t \geq 0$, $\overline{\lambda}(t) := u(t,0)$. Then, the following statements hold*

*(i) if $(\overline{N}_t)_{t\geq 0}$ is a solution of (14) then $\mathbb{E}\left[\overline{N}_+(dt)\right] = \overline{\lambda}(t)dt$,*

*(ii) there exists a unique (once $\Pi$ and $\overline{N}_-$ are fixed) solution $(\overline{N}_t)_{t\geq 0}$ of the following system*

$$
\begin{cases}
\overline{N}_t = \int_0^t \int_0^\infty \mathbb{1}_{\left\{x \leq \Psi\left(\overline{S}_{t'-}, \int_0^{t'} h(t'-z)\overline{\lambda}(z)dz + f_0(t')\right)\right\}} \Pi(dt', dx), \\
\mathbb{E}\left[\overline{N}_t\right] = \int_0^t \overline{\lambda}(t')dt',
\end{cases}
\tag{29}
$$

*where $(\overline{S}_{t-})_{t\geq 0}$ is the predictable age process associated with $\overline{N} = \overline{N}_- \cup \overline{N}_+$ where $\overline{N}_-$ is a point process distributed according to $\zeta_{N_-}$ and $\overline{N}_+$ is the point process associated with the counting process $(\overline{N}_t)_{t\geq 0}$*

*In particular, $\overline{\lambda}$ is a continuous function satisfying $\overline{\lambda}(t) = \mathbb{E}\left[\Psi\left(\overline{S}_{t-}, \int_0^{t-} h(t-z)\overline{\lambda}(z)dz + f_0(t)\right)\right]$ and the solution of (29) is the unique (once $\Pi$ and $\overline{N}_-$ are fixed) solution of (14).*

*Proof.* -(i) Suppose that $(\overline{N}_t)_{t\geq 0}$ is a solution of (14). The thinning procedure implies that $(\overline{N}_t)_{t\geq 0}$ admits an intensity which only depends on the time $t$ and the age $\overline{S}_{t-}$. This allows us to denote by $f$ the bivariate function such that the intensity of $\overline{N}$ at time $t$ is given by $\overline{\lambda}_t = f(t, \overline{S}_{t-})$. It satisfies for all $t, s \geq 0$,

$$
f(t,s) = \Psi\left(s, \int_0^{t-} h(t-z)\mathbb{E}\left[f(z, \overline{S}_{z-})\right]dz + f_0(t)\right).
\tag{30}
$$

In particular, the intensity is bounded since $\Psi$ is bounded. So, if we denote $w_t = w(t, \cdot)$ the distribution of the age $\overline{S}_{t-}$, Lemma B.2 gives that $w$ belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ and [8, Section 4.1] implies that $w$ satisfies the system

$$
\begin{cases}
\dfrac{\partial w(t,s)}{\partial t} + \dfrac{\partial w(t,s)}{\partial s} + f(t,s)w(t,s) = 0, \\
w(t,0) = \int_0^{+\infty} f(t,s)w(t,s)\,ds.
\end{cases}
$$

Yet, by definition of $w$, $\mathbb{E}\left[f(z, \overline{S}_{z-})\right] = \int_{s=0}^{+\infty} f(z,s)w(z,ds)$, so (30) rewrites as

$$
f(t,s) = \Psi\left(s, X(t) + f_0(t)\right),
\tag{31}
$$

where $X$ satisfies

$$
X(t) = \int_0^t h(t-z) \int_0^{+\infty} \Psi(s, X(z) + f_0(z))w(z, ds)dz.
$$

14

Hence, $w$ is a solution in $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$ of the system (22). Yet, the solution of (22) is unique (Theorem III.5) so we have $w = u$ (defined in Proposition III.6) and in particular $\mathbb{E}\left[\overline{N}_+(dt)\right] = \mathbb{E}\left[f(t, \overline{S}_{t-})\right] dt = u(t, 0)dt = \overline{\lambda}(t)dt$.

-$(ii)$ The first equation of (29) is a standard thinning equation with $\overline{\lambda}$ given by the first step so its solution $(\overline{N}_t)_{t \geq 0}$ is a measurable function of $\Pi$ and $\overline{N}_-$ hence it is unique (once $\Pi$ and $\overline{N}_-$ are fixed).

To conclude this step, it suffices to check that $(\overline{N}_t)_{t \geq 0}$ satisfies the second equation of (29). Identifying $\overline{\lambda}(t)$ with $\mathbb{E}\left[f(t, \overline{S}_{t-})\right]$ in (30), the intensity of $\overline{N}$ is given by

$$f(t, \overline{S}_{t-}) = \Psi\left(\overline{S}_{t-}, \int_0^{t-} h(t-z)\overline{\lambda}(z)dz + f_0(t)\right) \tag{32}$$

which is bounded. Hence, [8, Section 4.1] implies that the distribution of the age $\overline{S}_{t-}$ denoted by $v(t, \cdot)$ is the unique solution of

$$\begin{cases} \dfrac{\partial v(t,s)}{\partial t} + \dfrac{\partial v(t,s)}{\partial s} + \Psi\left(s, \int_0^t h(t-z)\overline{\lambda}(z)dz + f_0(t)\right) v(t,s) = 0, \\ v(t,0) = \displaystyle\int_0^\infty \Psi\left(s, \int_0^t h(t-z)\overline{\lambda}(z)dz + f_0(t)\right) v(t,s)\, ds. \end{cases} \tag{33}$$

Since $\overline{\lambda}(t) = u(t, 0)$ and $u$ is a solution of (22), it is clear that $u$ satisfies this system, so $u(t, \cdot)$ is the density of $\overline{S}_{t-}$. Finally, using Fubini's Theorem we have

$$\begin{aligned} \mathbb{E}\left[\overline{N}_t\right] = \int_0^t \mathbb{E}\left[f(t', \overline{S}_{t'-})\right] dt' &= \int_0^t \mathbb{E}\left[\Psi\left(\overline{S}_{t'-}, \int_0^{t'-} h(t'-z)\overline{\lambda}(z)dz + f_0(t')\right)\right] dt' \\ &= \int_0^t \int_0^\infty \Psi\left(s, \int_0^{t'} h(t'-z)\overline{\lambda}(z)dz + f_0(t')\right) u(t', s)ds\, dt' \\ &= \int_0^{t'} \overline{\lambda}(t')dt', \end{aligned}$$

since $u$ satisfies the second equation of (33).

Finally, the three remaining points are rather simple. Firstly, the continuity of $\overline{\lambda}$ comes from Theorem III.5. Secondly, using (32) and $(i)$ one has

$$\overline{\lambda}(t) = \mathbb{E}\left[f(t, \overline{S}_{t-})\right] = \mathbb{E}\left[\Psi\left(\overline{S}_{t-}, \int_0^{t-} h(t-z)\overline{\lambda}(z)dz + f_0(t)\right)\right].$$

Lastly, the solution of (29) is clearly a solution of (14) and $(i)$ tells that a solution of (14) is necessarily a solution of (29) which gives uniqueness. $\square$

**Proposition III.7.** *Under $(\mathcal{A}_{\text{Lip}}^\Psi)$ and $(\mathcal{A}_{\Psi=\Psi_0})$ assume that $h : \mathbb{R}_+ \to \mathbb{R}$ is locally integrable and that $f_0 : \mathbb{R}_+ \to \mathbb{R}$ is continuous.*

*Then, there exists a unique function $\overline{\lambda}$ (which is furthermore continuous on $\mathbb{R}_+$) depending only on $\Psi_0$, $h$ and $f_0$ such that the following statements hold*

*(i) if $(\overline{N}_t)_{t \geq 0}$ is a solution of (14) then $\mathbb{E}\left[\overline{N}_+(dt)\right] = \overline{\lambda}(t)dt$,*

*(ii) there exists a unique (once $\Pi$ is fixed) solution $(\overline{N}_t)_{t \geq 0}$ to the following system*

$$\begin{cases} \overline{N}_t = \displaystyle\int_0^t \int_0^\infty \mathbb{1}_{\left\{x \leq \Psi_0\left(\int_0^{t'} h(t'-z)\overline{\lambda}(z)dz + f_0(t')\right)\right\}} \Pi(dt', dx), \\ \mathbb{E}\left[\overline{N}_t\right] = \displaystyle\int_0^t \overline{\lambda}(t')dt'. \end{cases} \tag{34}$$

*In particular, $\overline{\lambda}(t) = \Psi_0\left(\int_0^{t-} h(t-z)\overline{\lambda}(z)dz + f_0(t)\right)$ and the solution of (34) is the unique (once $\Pi$ is fixed) solution of (14).*

The proof follows [13, Theorem 8-$(i)$] and is given in Appendix A.5 for sake of exhaustiveness.

## III.4 Link via the age process

The link between the limit equation (14) and system (22) is even deeper than what is stated in Proposition III.6. Indeed, the distribution of the age process associated with a solution of the limit equation is a solution of (22) as described in the next statement.

**Proposition III.8.** *Under the assumptions of Proposition III.6, the unique solution $u$ to the system* (22) *with initial condition that $u(0, \cdot) = u^{\mathrm{in}}$ is such that $u(t, \cdot)$ is the density of the age $\overline{S}_{t-}$ associated with the solution of the limit equation* (14) *given in Proposition III.6.*

This result is in fact given in the proof of Proposition III.6 below Equation (33).

# IV Mean-field dynamics

The convergence to the limit dynamics is proved by using a path-wise coupling (like in [13, 31]) between the processes given by Representation II.3 on the one hand and by the limit equation on the other hand. Then, this coupling is studied in two different cases: when the intensity is bounded, i.e. $\Psi$ satisfies $(\mathcal{A}_\infty^\Psi)$, or when the intensity does not depend on the age process, i.e. $\Psi$ satisfies $(\mathcal{A}_{\Psi=\Psi_0})$.

The precise statement regarding the convergence of the $n$-particle system towards point processes of the McKean-Vlasov type whose intensity depends on time and on the age is given in Corollary IV.4.

## IV.1 Coupling

Once the limit equation is well-posed, following the ideas of Sznitman in [46], it is easy to construct a suitable coupling between ADRHPs and i.i.d. solutions of the limit equation (14). More precisely, consider

- a sequence $(N_-^i)_{i \geq 1}$ of i.i.d. point processes distributed according to $\zeta_{N_-}$;

- an infinite matrix $(H_{ij})_{i,j \geq 1}$ (independent of $(N_-^i)_{i \geq 1}$) with entries distributed according to $\mu_H$ such that

$$\begin{cases} \text{for any fixed } i \geq 1, \text{ the variables } H_{i1}, \ldots, H_{in}, \ldots \text{ are independent,} \\ \text{the sequences } H_{i1}, \ldots, H_{in}, \ldots \text{ are exchangeable (with respect to } i), \\ \text{the matrix } (H_{ij})_{i,j \geq 1} \text{ is independent from the initial conditions } (N_-^i)_{i=1,..,n}; \end{cases} \quad (35)$$

- an infinite matrix $(F_{ij})_{i,j \geq 1}$ with entries distributed according to $\nu_F$ such that

$$\begin{cases} \text{for any fixed } i \geq 1, \text{ the variables } F_{i1}, \ldots, F_{in}, \ldots \text{ are independent,} \\ \text{the sequences } F_{i1}, \ldots, F_{in}, \ldots \text{ are exchangeable (with respect to } i), \end{cases} \quad (36)$$

- a sequence $(\Pi^i(dt', dx))_{i \geq 1}$ of i.i.d. $\mathbb{F}$-Poisson measures with intensity 1 on $\mathbb{R}_+^2$.

Notice that (35) (resp. (36)) is the equivalent of (8) (resp. (9)) for infinite matrices. Under $(\mathcal{A}_{\infty,2}^{\mu_H})$, $(\mathcal{A}_2^{\nu_F})$ and $(\mathcal{A}_{\mathrm{Lip}}^\Psi)$, the assumptions of Proposition II.4 are satisfied. Furthermore, if we assume either:

- $H_1$: "$(\mathcal{A}_\infty^\Psi)$ and $(\mathcal{A}_{u^{\mathrm{in}}}^{\zeta_{N_-}})$ are satisfied",

- $H_2$: "$(\mathcal{A}_{\Psi=\Psi_0})$ is satisfied",

then $m_{\mu_H}$ and $m_{\nu_F}$ (defined in $(\mathcal{A}_\infty^{\mu_H})$ and $(\mathcal{A}_1^{\nu_F})$) satisfy the assumptions of either Proposition III.6 (under $H_1$) or III.7 (under $H_2$) so one can build simultaneously:

- a sequence (indexed by $n \geq 1$) $(N^{n,i})_{i=1,\ldots,n}$ of ADRHPs with parameters $(n, \mu_H, \nu_F, \Psi, \zeta_{N_-})$ according to Representation II.3 that is

$$N_t^{n,i} = \int_0^t \int_0^\infty \mathbb{1}_{\left\{ x \leq \Psi \left( S_{t'-}^{n,i}, \frac{1}{n} \sum_{j=1}^n \left[ \int_0^{t'^-} H_{ij}(t'-z) N_+^{n,j}(dz) + F_{ij}(t') \right] \right) \right\}} \Pi^i(dt', dx) \quad (37)$$

with age processes $(S_{t-}^{n,i})_{t\geq 0}^{i=1,...,n}$ and past given by $N_-^i$,

- and a sequence $(\overline{N}_t^i)_{t\geq 0}^{i\geq 1}$ of i.i.d. solutions of the limit equation with parameters $(m_{\mu_H}, m_{\nu_F}, \Psi, \zeta_{N_-})$ that is

$$\overline{N}_t^i = \int_0^t \int_0^\infty \mathbb{1}_{\left\{x \leq \Psi\left(\overline{S}_{t'-}^i, \int_0^{t'} m_{\mu_H}(t'-z)\overline{\lambda}(z)dz + m_{\nu_F}(t')\right)\right\}} \Pi^i(dt', dx), \qquad (38)$$

where $\overline{\lambda}$ is defined either in Proposition III.6 (under $H_1$) or III.7 (under $H_2$) and $(\overline{S}_{t-}^i)_{t\geq 0}$ is the predictable age process associated with $\overline{N}^i = N_-^i \cup \overline{N}_+^i$.

Notice that this coupling is based on the sharing of a common past $(N_-^i)_{i\geq 1}$ and a common underlying randomness, that are the $\mathbb{F}$-Poisson measures $(\Pi^i(dt', dx))_{i\geq 1}$. Note that the sequence of ADRHPs is indexed by the size of the network $n$ whereas the solutions of the limit equation that represent the behaviour under the mean field approximation are not.

The following result states the control of the mean-field approximation.

**Theorem IV.1.** *Under* $(\mathcal{A}_{\infty,2}^{\mu_H})$, $(\mathcal{A}_2^{\nu_F})$ *and* $(\mathcal{A}_{\text{Lip}}^{\Psi})$, *assume either* $H_1$ *or* $H_2$.

*Then, the sequence of ADRHP* $(N^{n,i})_{i=1,...,n}$ *(with* $\mathbb{F}$*-intensities on* $\mathbb{R}_+$ *denoted by* $(\lambda^{n,i})_{i=1,...,n}$*) defined by* (37) *and the i.i.d. copies* $(\overline{N}_t^i)_{t\geq 0}^{i\geq 1}$ *of the solution of the limit equation (with* $\mathbb{F}$*-intensities on* $\mathbb{R}_+$ *denoted by* $(\overline{\lambda}^i)_{i=1,...,n}$*) defined by* (38) *are such that for all* $i = 1,\ldots,n$ *and* $\theta > 0$,

$$\mathbb{E}\left[\sup_{t\in[0,\theta]} |N_t^{n,i} - \overline{N}_t^i|\right] \leq \int_0^\theta \mathbb{E}\left[|\lambda_t^{n,i} - \overline{\lambda}_t^i|\right] dt \leq C(\theta, \Psi, \mu_H, \nu_F) n^{-1/2}, \qquad (39)$$

*where the constant* $C(\theta, \Psi, \mu_H, \nu_F)$ *does not depend on* $n$.

*Proof.* First, let us note that in each cases, the coupling is well-defined thanks to either Proposition III.6 or III.7. Let us fix some $n \geq 1$. Let us denote $\Delta_n^i(t) = \int_0^t |N_+^{n,i}(dt') - \overline{N}_+^i(dt')|$ and $\delta_n^i(t) = \mathbb{E}[\Delta_n^i(t)]$ its expectation. Denoting $A\triangle B$ the symmetric difference of the sets $A$ and $B$ and $\text{Card}(A)$ the cardinal of the set $A$ we have

$$\Delta_n^i(t) = \int_0^t |N_+^{n,i}(dt') - \overline{N}_+^i(dt')| = \text{Card}\left((N^{n,i}\triangle \overline{N}^i) \cap [0,t]\right), \qquad (40)$$

that is the number of points that are not common to $N^{n,i}$ and $\overline{N}^i$ between 0 and $t$. Then, it is clear that, for all $i = 1,\ldots,n$ and $\theta > 0$, $\sup_{t\in[0,\theta]} |N_t^i - \overline{N}_t^i| \leq \Delta_n^i(\theta)$ and so

$$\mathbb{E}\left[\sup_{t\in[0,\theta]} |N_t^{n,i} - \overline{N}_t^i|\right] \leq \mathbb{E}\left[\Delta_n^i(\theta)\right] = \delta_n^i(\theta). \qquad (41)$$

On the one hand, the $\overline{N}^i$'s are i.i.d. hence exchangeable. On the other hand, thanks to the form of the intensity and the assumptions on the matrices $(H_{ij})_{i,j\geq 1}$ and $(F_{ij})_{i,j\geq 1}$ - (35) and (36) - the family $(N^{n,i})_{i=1,...,n}$ is exchangeable too. Hence $\delta_n^i$ does not depend on $i$ and is simply denoted by $\delta_n$ in the sequel. Let us focus on the case $i = 1$. First, let us remind that $\overline{\lambda}_t^1$ is the intensity of $\overline{N}^1$, so

$$\Delta_n^1(\theta) = \int_0^\theta \int_0^\infty \left|\mathbb{1}_{\left\{x\leq\lambda_t^{n,1}\right\}} - \mathbb{1}_{\left\{x\leq\overline{\lambda}_t^1\right\}}\right| \Pi^1(dt, dx).$$

Taking expectation we find

$$\begin{aligned}
\delta_n(\theta) &= \mathbb{E}\left[\int_0^\theta \mathbb{E}\left[\int_0^\infty \left|\mathbb{1}_{\left\{x\leq\lambda_t^{n,1}\right\}} - \mathbb{1}_{\left\{x\leq\overline{\lambda}_t^1\right\}}\right| \Pi^1(dt, dx)\Big| \mathcal{F}_{t-}\right]\right] \\
&= \mathbb{E}\left[\int_0^\theta \int_0^\infty \left|\mathbb{1}_{\left\{x\leq\lambda_t^{n,1}\right\}} - \mathbb{1}_{\left\{x\leq\overline{\lambda}_t^1\right\}}\right| dx dt\right] \\
&= \mathbb{E}\left[\int_0^\theta |\lambda_t^{n,1} - \overline{\lambda}_t^1| dt\right] = \int_0^\theta \mathbb{E}\left[|\lambda_t^{n,1} - \overline{\lambda}_t^1|\right] dt, \qquad (42)
\end{aligned}$$

where the last equality comes from Fubini's Theorem. It remains to show that the rate of convergence is $n^{-1/2}$.

Computations given in Section A.6 show that in each cases there exists some locally bounded function $g$ depending on $\Psi$, $\mu_H$ and $\nu_F$ such that $\delta_n$ satisfies,

$$
\begin{cases}
\delta_n(\theta) \leq n^{-1/2}g(\theta) + \displaystyle\int_0^\theta \left[ ||\Psi||_\infty + \mathrm{Lip}(\Psi)M_{\mu_H}(\theta - z) \right] \delta_n(z)dz & \text{(under } H_1\text{)}, \\[4mm]
\delta_n(\theta) \leq n^{-1/2}g(\theta) + \displaystyle\int_0^\theta \mathrm{Lip}(\Psi)M_{\mu_H}(\theta - z)\delta_n(z)dz & \text{(under } H_2\text{)}.
\end{cases}
\tag{43}
$$

Remark that the only dependence with respect to $n$ lies in $\delta_n$. Since $g$ is locally bounded and $M_{\mu_H}$ is locally integrable, using Lemma B.4-(i), we end up with $\delta_n(\theta) \leq C(\theta, \Psi, \mu_H, \nu_F) \, n^{-1/2}$ where $C$ does not depend on $n$. $\qquad\square$

There are mainly two reasons for the dichotomy of the assumptions $H_1$ and $H_2$. Firstly, up to our knowledge, existence/uniqueness results on the macroscopic system (22) are only valid if the function $\Psi$ is bounded. Secondly, as it appears in (43), when the intensity of the $n$-particle system depends on the age, the control of $\delta_n(\theta)$ involves the $L^\infty$ norm of the function $\Psi$. This boundedness condition is used in order to control the coupling as soon as the ages of the $n$-particle system on the one hand and the i.i.d. copies of the solution of the limit equation on the other hand are different. Notice that even under $H_2$, this coupling result extends [13, Theorem 8-(ii)] since there are two novelties in the present article: random interaction functions $H_{ij}$ as well as dependences with respect to the past $F_{ij}$.

Under more restrictive assumptions (corresponding informally to uniform controls instead of local ones), the rate (with respect to $\theta$) of $C(\theta, \Psi, \mu_H, \nu_F)$ given in Theorem IV.1 is linear in comparison with a rate which is at least exponential in general. The main assumption corresponds to the stability criterion of Hawkes processes [5].

**Proposition IV.2.** *Under $(\mathcal{A}_{\infty,2}^{\mu_H})$, $(\mathcal{A}_2^{\nu_F})$ and $(\mathcal{A}_{\mathrm{Lip}}^{\Psi})$, assume either $H_1$ or $H_2$. Furthermore, assume that both:*

- *the functions $M_{\mu_H}$ and $\Psi$ are such that $\alpha := \mathrm{Lip}(\Psi)||M_{\mu_H}||_1 < 1$ and $||M_{\mu_H}||_2 < \infty$;*

- *the functions $m_{\nu_F}$ and $V_{\nu_F}$ are uniformly bounded.*

*Then, the constant $C(\theta, \Psi, \mu_H, \nu_F)$ given in Theorem IV.1 can be bounded by $\beta(\Psi, \mu_H, \nu_F)\theta$ where the constant $\beta(\Psi, \mu_H, \nu_F)$ depends neither on $\theta$ nor on $n$.*
*This bound holds for all $\theta \geq 0$ under $H_2$ whereas it holds for $\theta < (1 - \alpha)/||\Psi||_\infty$ under $H_1$.*

A proof is given in Appendix A.7 where an explicit expression of $\beta$ can be found in Equation (81) or (83) depending on the context.

As said in the introduction, Hawkes processes seem to be the microscopic point processes underpinning the (PPS) system introduced in [36]. There is a striking similarity, modulo a change of time, between :

- on the one hand, the mean intensity, denoted by $m(t)$, of a Hawkes process which is a function of $\int_0^t h(t - z)m(z)dz$

- and on the other hand, the firing rate $p$ in (PPS) which is a function of $\int_0^t d(z)n(0, t - z)dz$.

A first step in this direction has been made in [8] in the framework of a network of i.i.d. Hawkes processes. In that case, there is no direct bridge between Hawkes processes and the (PPS) system as it is shown in [8]. Indeed, when the size of the network goes to infinity, one recovers conditional expectation of the intensity with respect to the age (instead of the mean intensity). By comparison, the limit system of the mean-field interacting age-dependent Hawkes processes system considered in the present article involves the mean intensity as it can be seen in Equation (14). Furthermore, this mean intensity is, in some cases, given by the macroscopic system.

System (37)-(38) provides an efficient coupling between the spikes attached with the $n$-particle system and the spikes associated with the limiting process. In order to go one step further, a natural question is to wonder about a possible coupling between the ages associated with the two dynamics. This question is not addressed in [13] in which the propagation of chaos is discussed

at the level of the counting processes only. In comparison, we are here willing to investigate this question carefully. The underlying motivation is not of a mathematical essence only: exhibiting a suitable coupling between the ages of the ADRHP and the ages of the point processes of the McKean-Vlasov type whose intensity depends on time and on the age is the right and proper way to make the connection between the microscopic description of the neural dynamics and the macroscopic equation (PPS).

In the sequel, Assumption $(\mathcal{A}_\infty^{\zeta_N-})$ is used. It appears that the dependence of the age at time 0 with respect to the initial condition generates additional difficulties for investigating the mean-field approximation. To limit the complexity of the analysis, it is quite convenient to assume that the age at time 0 is bounded, which is precisely what Assumption $(\mathcal{A}_\infty^{\zeta_N-})$ says.

**Corollary IV.3.** *With the notations and assumptions of Theorem IV.1, assume that $(\mathcal{A}_\infty^{\zeta_N-})$ is satisfied.*

*Then, the age processes $(S_{t-}^{n,i})_{t\geq 0}^{i=1,\ldots,n}$ associated with the sequence of ADRHP $(N^{n,i})_{i=1,\ldots,n}$ and the age processes $(\overline{S}_{t-}^i)_{t\geq 0}^{i\geq 1}$ associated with the i.i.d. solutions $(\overline{N}_t^i)_{t\geq 0}^{i\geq 1}$ of the limit equation satisfy for all $i=1,\ldots,n$ and $\theta>0$,*

$$\mathbb{E}\left[\sup_{t\in[0,\theta]}|S_{t-}^{n,i}-\overline{S}_{t-}^i|\right] \leq (M_{T_0}+\theta)\,C(\theta,\Psi,\mu_H,\nu_F)\,n^{-1/2}, \tag{44}$$

*where $C(\theta,\Psi,\mu_H,\nu_F)$ is given in Theorem IV.1 and $M_{T_0}$ is defined in $(\mathcal{A}_\infty^{\zeta_N-})$.*

*Proof.* Let us note that, for all $n\geq 1$ and $i=1,\ldots,n$, $N^{n,i}$ and $\overline{N}^i$ coincide on the non-positive part, i.e. $N_-^{n,i}=\overline{N}_-^i$. Therefore, $S_0^{n,i}=\overline{S}_0^i$ and $\sup_{t\in[0,\theta]}|S_{t-}^{n,i}-\overline{S}_{t-}^i|$ is a.s. upper bounded by $M_{T_0}+\theta$ when the trajectories $(S_{t-}^{n,i})_{t\in[0,\theta]}$ and $(\overline{S}_{t-}^i)_{t\in[0,\theta]}$ are different and is equal to 0 otherwise. Therefore, we have the following bound

$$\mathbb{E}\left[\sup_{t\in[0,\theta]}|S_{t-}^{n,i}-\overline{S}_{t-}^i|\right] \leq (M_{T_0}+\theta)\mathbb{P}\left(\left(S_{t-}^{n,i}\right)_{t\in[0,\theta]}\neq\left(\overline{S}_{t-}^i\right)_{t\in[0,\theta]}\right).$$

Yet if the trajectories are different, there is at least one point between 0 and $\theta$ which is not common to both $N_+^{n,i}$ and $\overline{N}_+^i$, that is $\sup_{t\in[0,\theta]}|N_t^{n,i}-\overline{N}_t^i|\neq 0$, hence

$$\mathbb{P}\left(\left(S_{t-}^{n,i}\right)_{t\in[0,\theta]}\neq\left(\overline{S}_{t-}^i\right)_{t\in[0,\theta]}\right) = \mathbb{P}\left(\sup_{t\in[0,\theta]}|N_t^{n,i}-\overline{N}_t^i|\neq 0\right). \tag{45}$$

Moreover, since counting processes are piecewise constant with jumps of height 1 a.s., it is clear that

$$\mathbb{P}\left(\sup_{t\in[0,\theta]}|N_t^{n,1}-\overline{N}_t^1|\neq 0\right) = \mathbb{P}\left(\sup_{t\in[0,\theta]}|N_t^{n,1}-\overline{N}_t^1|\geq 1\right) \leq \mathbb{E}\left[\sup_{t\in[0,\theta]}|N_t^{n,1}-\overline{N}_t^1|\right], \tag{46}$$

where we used Markov's inequality. Finally, inequality (44) clearly follows from Theorem IV.1. $\square$

## IV.2 Mean-field approximations

Inspired by the seminal work of Sznitman [46], we now obtain, from the results of the previous section, the convergence of the $n$-particle system towards the limit equation:

- the empirical distribution of the point processes associated with the $n$-particle system converges to the distribution of the point process solution of the limit equation,

- the empirical distribution of the age processes associated with the $n$-particle system converges to the distribution of the age process associated with the solution of the limit equation,

This result, together with the ones from the previous paragraphs, are typical of what is known as the *propagation of chaos theory of interacting particle system*. In particular, it says that $k$ fixed neurons behave independently and identically when the size of the network goes to infinity. Their spiking dynamics being described by the limit equation (14).

**Corollary IV.4.** *Let $\mathcal{L}(X)$ denote the distribution of some random variable $X$ and $\mathcal{D}(\mathbb{R}_+)$ denote the space of càdlàg functions from $\mathbb{R}_+$ to $\mathbb{R}$ endowed with the Skorohod topology.*

*With notations and assumptions of Theorem IV.1, we have the following mean-field approximations:*

- *the weak convergence in $\mathcal{P}(\mathcal{D}(\mathbb{R}_+))$ of the empirical measure of counting processes,*

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{(N_t^{n,i})_{t\geq 0}} \xrightarrow[n\to\infty]{} \mathcal{L}\big((\overline{N}_t^1)_{t\geq 0}\big); \tag{47}$$

- *if furthermore $(\mathcal{A}_\infty^{\zeta_{N-}})$ holds, the weak convergence in $\mathcal{P}(\mathcal{D}(\mathbb{R}_+))$ of the empirical measure of age processes,*

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{(S_{t-}^{n,i})_{t\geq 0}} \xrightarrow[n\to\infty]{} \mathcal{L}\big((\overline{S}_{t-}^1)_{t\geq 0}\big). \tag{48}$$

*Both convergences also hold in probability since the limits are constant in $\mathcal{P}(\mathcal{D}(\mathbb{R}_+))$.*

*Finally, if furthermore $(\mathcal{A}_\infty^\Psi)$ and $(\mathcal{A}_{u^{\mathrm{in}}}^{\zeta_{N-}})$ hold, then the unique solution $u$ of the system (22) with initial condition that $u(0,\cdot) = u^{\mathrm{in}}$ is such that $u_t := u(t,\cdot)$ is the density of the age $\overline{S}_{t-}^1$ and for all $\theta > 0$,*

$$\sup_{t\in[0,\theta]} \mathbb{E}\left[W_1\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{S_{t-}^{n,i}}, u_t\right)\right] \leq D(\theta,\Psi,\mu_H,\nu_F,M_{T_0})\,n^{-1/2}, \tag{49}$$

*where $W_1$ denotes the standard 1-Wasserstein distance and the constant $D(\theta,\Psi,\mu_H,\nu_F,M_{T_0})$ does not depend on $n$.*

*Proof.* The space of càdlàg functions $\mathcal{D}(\mathbb{R}_+)$ endowed with the Skorohod topology is a Polish space. So, according to [46, Proposition 2.2] or [31, Proposition 4.2], to show the first limit (47), it suffices to check that $((N_t^{n,1})_{t\geq 0}, (N_t^{n,2})_{t\geq 0})$ converges in distribution, as $n \to +\infty$, to two independent copies of $(\overline{N}_t^1)_{t\geq 0}$. Since the convergence with respect to Lipschitz continuous test functions is sufficient in order to prove the convergence in distribution (Portemanteau Theorem [26]), the first limit clearly follows from both (39) and the fact that the uniform convergence topology on compact time intervals is finer than the Skorohod topology. The proof of the second limit is similar with the difference that it follows from (44) instead of (39).

The link between the solution of (22) and the age process associated with the solution of the limit equation (14) is given by Proposition III.8.

The rate of convergence for the 1-Wasserstein distance stated in (49) is a consequence of the rate of convergence for i.i.d. real valued random variables. Fix $\theta > 0$ and let $t$ be in $[0,\theta]$. First, using the exchangeability of the particles, it follows from Corollary IV.3 that there exists a constant $C(\theta,\Psi,\mu_H,\nu_F,M_{T_0})$ such that

$$\mathbb{E}\left[W_1\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{S_{t-}^{n,i}}, \frac{1}{n}\sum_{i=1}^{n}\delta_{\overline{S}_{t-}^i}\right)\right] \leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\sup_{t\in[0,\theta]}|S_{t-}^{n,i} - \overline{S}_{t-}^i|\right] \leq C(\theta,\Psi,\mu_H,\nu_F,M_{T_0})\,n^{-1/2}.$$

Then, applying [16, Theorem 1] to the i.i.d. random variables $\overline{S}_{t-}^i$, that are bounded by $M_{T_0}+\theta$, we deduce that there exists a constant $\tilde{C}(\theta,M_{T_0})$ such that $\mathbb{E}[W_1(\frac{1}{n}\sum_{i=1}^{n}\delta_{\overline{S}_{t-}^i}, u_t)] \leq \tilde{C}(M_{T_0},\theta)n^{-1/2}$. Finally, the triangular inequality for the 1-Wasserstein distance gives (49). □

The first mean-field approximation (47) is a generalization of the one given in [13, Theorem 8-*(iii)*] where mean-field interacting Hawkes processes are approximated by Poisson processes of the McKean-Vlasov type. Here, the limit processes are point processes of the McKean-Vlasov type whose intensity depends on time (like Poisson processes) and on the age.

Moreover, Equation (49) extends the result on the rate of convergence for the age processes given in [42, Section 5].

# V    Conclusion

We present a generalization of mean-field interacting Hawkes processes, namely age dependent random Hawkes processes (ADRHPs), which are well-adapted to neuroscience modelling. From a biological point of view, they encompass some interesting features such as refractory period, synaptic integration or random synaptic weights. These processes are studied in a mean-field situation and we show in Theorem IV.1 and Corollary IV.4 that, as the number of particles goes to infinity, they can be well approximated by point processes of the McKean-Vlasov type whose intensity depends on time and on the age. These limit point processes are closely related to the age structured PDE system introduced by Pakdaman, Perthame and Salort, namely (PPS), as shown in Proposition III.8.

Hence, using the theory of mean-field approximations, the present article makes a bridge between the microscopic modelling given by Hawkes processes, or more generally age dependent random Hawkes processes, and the macroscopic modelling given by the (PPS) system. This bridge is presented under the main assumption that the intensity of the microscopic point processes is bounded. In this sense, the present article offers an answer to the question left open in [8]. This legitimises the convolution term $X(t)$ in the (PPS) system as well as opens the way to the study of new assumptions on the spiking rate $p$ appearing in the (PPS) system from a more analytical point of view. Up to our knowledge, this has not been done yet.

The present article gives somehow the law of large numbers for a generalization of Hawkes processes. It could be interesting to investigate how these processes fluctuates around their mean limit or in other words find some kind of functional central limit theorem for Hawkes processes in a mean-field framework.

As noted, random synaptic weights can be considered in this study. However, they are supposed to be, in some sense, independent and identically distributed which can be considered as an unrealistic assumption. Inspired by [14], it could be interesting to see how correlated synaptic weights could be handled in the Hawkes processes framework.

On a different path, it could be interesting to see how locally stationary Hawkes processes, as introduced in [45], behave in a mean-field situation. Indeed, these processes may take into account the dynamics of the synaptic weights occurring in the neural network.

# A  Proofs

## A.1  Proof of Proposition II.4

Let us denote $G_i : t \mapsto \sup_{s \geq 0} \Psi(s, 0) + \text{Lip}(\Psi)n^{-1} \sum_{j=1}^n |F_{ij}(t)|$. Thanks to $(\mathcal{A}_1^{\nu_F})$ and Fubini's Theorem we have, for all $T > 0$, $\mathbb{E}[\int_0^T |F_{11}(t)| dt] = \int_0^T \mathbb{E}[|F_{11}(t)|] dt < +\infty$. In particular, for all $1 \leq i, j \leq n$, $t \mapsto |F_{ij}(t)|$ is locally integrable almost surely. Hence, there exists a subset $\Omega$ of probability 1 such that, on $\Omega$, $G_i$ is locally integrable for all $i$. Fixing the $G_i$'s, one can apply Lemma B.1 (with $a_i = \text{Lip}(\Psi)$ and $g_i = G_i$) to deduce that the processes $(N_t^i)_{t \geq 0}^{i=1,..,n}$ are dominated by the processes $(\tilde{N}_t^i)_{t \geq 0}^{i=1,..,n}$ (defined by (85)) and so are well-defined.

It remains to show that the function $t \mapsto \mathbb{E}[N_t^1]$ is locally bounded. First, let us study the dominating processes. We have

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \tilde{N}_t^i\right] &\leq \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n \int_0^t G_i(t') dt'\right] + \text{Lip}(\Psi)\frac{1}{n} \sum_{i=1}^n \int_0^t \int_0^{t'} M_{\mu_H}(t' - z) \mathbb{E}\left[\frac{1}{n} \sum_{j=1}^n \tilde{N}^j(dz)\right] dt' \\
&\leq \frac{1}{n} \sum_{i=1}^n \int_0^t \mathbb{E}[G_i(t')] dt' + \text{Lip}(\Psi) \int_0^t M_{\mu_H}(t - t') \mathbb{E}\left[\frac{1}{n} \sum_{j=1}^n \tilde{N}_{t'}^j\right] dt',
\end{aligned}
$$

where we used Lemma B.3. Next, $t \mapsto \mathbb{E}[G_i(t)] = \sup_{s \geq 0} \Psi(s, 0) + \text{Lip}(\Psi)\mathbb{E}[|F_{11}(t)|]$ is locally bounded and $M_{\mu_H}$ is locally integrable so Lemma B.4-*(i)* gives that $t \mapsto \mathbb{E}[n^{-1} \sum_{i=1}^n \tilde{N}_t^i] = \mathbb{E}[\tilde{N}_t^1]$ is locally bounded. Finally, the stochastic domination (in particular, $\mathbb{E}[N_t^1] \leq \mathbb{E}[\tilde{N}_t^1]$) gives the result.

## A.2  Proof of Proposition III.1

First, in order to be consistent with the formalism used in [7] we must rewrite the system (15) in a single equation in the following way

$$
\frac{\partial u(t, s)}{\partial t} + \frac{\partial u(t, s)}{\partial s} = N_1(t, s, u) + \delta_{s=0} n_2(t, u), \tag{50}
$$

with initial condition $u^{\text{in}}$ where

$$
\begin{cases}
N_1(t, s, u) := -f(t, s)u(t, s) \\
n_2(t, u) := \int_0^{+\infty} f(t, s')u(t, s') ds'.
\end{cases}
$$

The use of the Dirac mass localized in age equal to 0 represents the boundary condition that is the second equation of (15).

Note that the general result [7, Theorem 2.4.] gives existence and uniqueness of solution in $BC(\mathbb{R}_+, \mathcal{M}(\mathbb{R}))$ and not $BC(\mathbb{R}_+, \mathcal{M}(\mathbb{R}_+))$ even if the initial condition has support contained in $\mathbb{R}_+$. However, as explained in [7, Section 3.3.], it suffices to extend the equation for $s$ in $\mathbb{R}$, to apply [7, Theorem 2.4.] and then to check that the support of the solution is conserved in the sense that: if $u^{\text{in}}$ has support on $\mathbb{R}_+$ then the unique solution given by the Theorem has also support contained in $\mathbb{R}_+$ for all time $t \geq 0$.

Hence, consider Equation (50) but with $s$ being in $\mathbb{R}$ by mirror symmetry for definiteness (that is $N_1(t, s, u) = N_1(t, -s, u)$ and so $f(t, s) = f(t, -s)$). Let us check that the assumptions of [7, Theorem 2.4.] are satisfied.

-(H1) and (H2) are clearly satisfied.

-(H3). We need to verify that $N_1$ and $n_2$ are continuous in $t$ with respect to the usual topology and in $u$ with respect to the topology induced by the bounded Lipschitz norm (denoted by $||.||_{BL}$). On the one hand, using the boundedness of $f$, we have, with $u_t$ denoting the measure $u(t, \cdot)$,

$$
||N_1(t + t', s, u + \tilde{u}) - N_1(t, s, u)||_{BL} \leq \sup_{s \in \mathbb{R}} |f(t + t', s) - f(t, s)|||u_t||_{BL}
$$
$$
+ ||f||_\infty ||\tilde{u}_{t+t'}||_{BL} + ||f||_\infty ||u_{t+t'} - u_t||_{BL}.
$$

As $t'$ and $\tilde{u}$ converge to $0$, the first term converges to $0$ since $f$ is (uniformly in $s$) continuous with respect to $t$, the second one clearly converges to $0$ and the third one converges to $0$ since $u$ belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{M}(\mathbb{R}))$.

On the other hand, using once again the boundedness of $f$, we have,

$$|n_2(t + t', u + \tilde{u}) - n_2(t, u)| \leq \sup_{s \in \mathbb{R}} |f(t + t', s) - f(t, s)| ||u_t||_{BL} + ||f||_\infty ||u_{t+t'} - u_t||_{BL},$$

which converges to $0$ as $t'$ tends to $0$.

-(H4). It suffices to show that $N_1$ and $n_2$ are Lipschitz continuous with respect to the variable $u$. On the one hand, we have

$$||N_1(t, s, u) - N_1(t, s, v)||_{BL} \leq ||f||_\infty ||u_t - v_t||_{BL}.$$

On the other hand, we have

$$|n_2(t, u) - n_2(t, v)| \leq ||f||_\infty ||u_t - v_t||_{BL}.$$

-(H5). Here, it suffices to check that $N_1(t, s, u) + \delta_{s=0} n_2(t, u)$ carries bounded sets in total variation norm to bounded sets in total variation norm. Denoting $||.||_{TV}$ the total variation norm, we have $||N_1(t, s, u) + \delta_{s=0} n_2(t, u)||_{TV} \leq 2||f||_\infty ||u||_{TV}$.

Finally, the argument to prove conservation of the support for solutions being the same as the one elaborated in [7, Section 3.3.], it is not reproduced here.

## A.3  Proof of Proposition III.2

The method of characteristics applied to the first equation of (15) suggests to consider:

- for all $z \geq 0$, $u^z : t \mapsto u(t, t + z)$ satisfying

$$\frac{d}{dt} u^z(t) = -f(t, z + t) u^z(t)$$

hence for all $t \geq 0$,

$$u^z(t) = u^z(0) \exp\left(-\int_0^t f(t', z + t') dt'\right)$$

and so, using that $u^z(0) = u^{\text{in}}(z)$ and letting $s = z + t$, one has (18).

- for all $z \geq 0$, $u^z : s \mapsto u(s + z, s)$ satisfying

$$\frac{d}{ds} u^z(s) = -f(s + z, s) u^z(s)$$

hence for all $s \geq 0$,

$$u^z(s) = u^z(0) \exp\left(-\int_0^s f(s' + z, s') ds'\right)$$

and so, using that $u^z(0) = u(z, 0)$ and letting $t = z + s$, one has

$$u(t, s) = u(t - s, 0) \exp\left(-\int_0^s f(t - s + s', s') ds'\right), \quad \text{for } t \geq s,$$

which is not exactly (19). Here, $u(t - s, 0)$ is just a parameter which is not constrained by the first equation of (15). However, it is characterized by the second equation of (15) as explained below.

For any $T > 0$, consider the application

$$G_T : \quad \begin{array}{ccc} L^\infty([0, T]) & \longrightarrow & L^\infty([0, T]) \\ u_0 & \longmapsto & (t \mapsto G(u_0)(t)), \end{array}$$

where

$$G(u_0)(t) := \int_0^t f(t, s) u_0(t - s) \exp\left(-\int_0^s f(t - s + s', s') ds'\right) ds$$

$$+ \int_t^{+\infty} f(t, s) u^{\text{in}}(s - t) \exp\left(-\int_0^t f(t', s - t + t') dt'\right) ds.$$

23

Note that the characteristics and the second equation of (15) suggest that, denoting $u$ the solution of (15), $u(\cdot, 0)$ is bounded and that its trace on $[0, T]$ is a fixed point of $G_T$. Using the boundedness of $f$ and the fact that the argument in the exponential is non-positive, we have for any $u_0, v_0$ in $L^\infty([0, T])$,

$$|G_T(u_0)(t) - G_T(v_0)(t)| \leq ||f||_\infty \int_0^t |u_0(t-s) - v_0(t-s)| ds \leq T||f||_\infty ||u_0 - v_0||_{L^\infty([0,T])}.$$

Now, fix $T > 0$ such that $T||f||_\infty \leq 1/2$ so that $G_T$ is a contraction and admits a unique fixed point. Note that $G_T$ maps non-negative functions to non-negative functions, so that the fixed point of $G_T$ is a non-negative function. Iterating this fixed point gives the existence and uniqueness of a locally bounded function $u_0$ (which is non-negative) such that for all $t \geq 0$, $u_0(t) = G(u_0)(t)$ (see the end of the proof of Lemma III.4 for the same kind of argument in a more detailed form). Until the end of the proof, $u_0$ will denote this fixed point.

It only remains to check that $u$ is a solution of (15) in the weak sense. Let $\varphi$ be in $\mathcal{C}_{c,b}^\infty(\mathbb{R}_+^2)$, let us compute $\int_{\mathbb{R}_+^2} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \varphi(t, s) u(t, s) dt ds$. Sticking with the decomposition of the representation given by (18)-(19) and using integration by parts to go backward in the heuristic given by the method of characteristics, one has

$$\int_{s \geq t} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \varphi(t, s) u(t, s) dt ds = \int_{s \geq t} \varphi(t, s) f(t, s) u(t, s) dt ds - \int_{s=0}^{+\infty} \varphi(0, s) u^{\text{in}}(s) ds, \quad (51)$$

and

$$\int_{t \geq s} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \varphi(t, s) u(t, s) dt ds = \int_{t \geq s} \varphi(t, s) f(t, s) u(t, s) dt ds - \int_{t=0}^{+\infty} \varphi(t, 0) u_0(t) dt. \quad (52)$$

Remarking that the definition of $u_0$ as the fixed point of $G$ gives that $u$ defined by (18)-(19) satisfies the second equation of (15) in a strong sense, one deduces that (17) is satisfied by gathering (51) and (52).

## A.4   Proof of (28)

Let $Y_1$ and $Y_2$ be two fixed locally bounded functions and denote $u_1 := u_{Y_1}$ and $u_2 := u_{Y_2}$ the two solutions associated with $Y_1$ and $Y_2$ (with same initial condition $u^{\text{in}}$) and $\tilde{u}(t) := u_1(t, 0) - u_2(t, 0)$ for all $t \geq 0$. Using the characteristics, i.e. (18) and (19), one deduces from the second equation in (26) that for $i = 1$ or $2$, $u_i(t, 0) = v_i^{0,t} + v_i^{t,\infty}$ where

$$\begin{cases} v_i^{0,t} := \int_0^t \Psi(s, Y_i(t) + f_0(t)) u_i(t-s, 0) \exp\left(-\int_0^s \Psi(s', Y_i(t-s+s') + f_0(t-s+s')) ds'\right) ds \\ v_i^{t,\infty} := \int_t^{+\infty} \Psi(s, Y_i(t) + f_0(t)) u^{\text{in}}(s-t) \exp\left(-\int_0^t \Psi(s-t+t', Y_i(t') + f_0(t')) dt'\right) ds. \end{cases}$$

and so $\tilde{u}(t) = A(t) + B(t)$ where $A(t) := v_1^{0,t} - v_2^{0,t}$ and $B(t) := v_1^{t,\infty} - v_2^{t,\infty}$. Before studying the functions $A$ and $B$, let us remark that in order to prove (28), it suffices to prove that there exists a non-decreasing function $C$ (which depends neither on the initial condition nor on $f_0$) such that for all $T \geq 0$,

$$||\tilde{u}||_{L^\infty([0,T])} \leq C(T)||Y_1 - Y_2||_{L^\infty([0,T])}. \quad (53)$$

Indeed, using the definition of $F_T$ given by (27), one then deduces that for all $T \geq 0$,

$$||F_T(Y_1) - F_T(Y_2)||_{L^\infty([0,T])} \leq C(T)||Y_1 - Y_2||_{L^\infty([0,T])} \int_0^T |h(z)| dz$$

and since $h$ is locally integrable and $C$ is non-decreasing, there exists $T > 0$ small enough such that $C(T) \int_0^T |h(z)| dz \leq 1/2$ which ends the proof. To prove (53), let $t$ be a positive real number.

24

**Study of $A$.** We have $A = A_1 + A_2 + A_3$ with

$$
\begin{cases}
A_1(t) := \displaystyle\int_0^t [\Psi(s, Y_1(t) + f_0(t)) - \Psi(s, Y_2(t) + f_0(t))]u_1(t - s, 0)e^{-\int_0^s \Psi(s', Y_1 + f_0)ds'}ds \\[2mm]
A_2(t) := \displaystyle\int_0^t \Psi(s, Y_2(t) + f_0(t))[u_1(t - s, 0) - u_2(t - s, 0)]e^{-\int_0^s \Psi(s', Y_1 + f_0)ds'}ds \\[2mm]
A_3(t) := \displaystyle\int_0^t \Psi(s, Y_2(t) + f_0(t))u_2(t - s, 0)[e^{-\int_0^s \Psi(s', Y_1 + f_0)ds'} - e^{-\int_0^s \Psi(s', Y_2 + f_0)ds'}]ds.
\end{cases}
$$

where the arguments "$t - s + s'$" in the exponentials are not written for simplicity.

- Study of $A_1$. Using the Lipschitz continuity of $\Psi$, the fact that the argument in the exponential is non-positive and the a priori bound on $u_1$, we have

$$
|A_1(t)| \leq \text{Lip}(\Psi) \int_0^t u_1(t - s, 0)|Y_1(t) - Y_2(t)|ds \leq \text{Lip}(\Psi)||u_1||_{L^\infty(\mathbb{R}_+^2)}\, t\, |Y_1(t) - Y_2(t)|
$$

$$
\leq \text{Lip}(\Psi)\max(M, ||\Psi||_\infty)\, t\, ||Y_1 - Y_2||_{L^\infty([0,t])}. \quad (54)
$$

- Study of $A_2$. Using the boundedness of $\Psi$ and the fact that the argument in the exponential is non-positive, we have

$$
|A_2(t)| \leq ||\Psi||_\infty \int_0^t |u_1(z, 0) - u_2(z, 0)|dz = ||\Psi||_\infty \int_0^t |\tilde{u}(z, 0)|dz. \quad (55)
$$

- Study of $A_3$. The arguments of the exponentials are non-positive and the exponential function is Lipschitz with constant 1 on $\mathbb{R}_-$ so, using the a priori bound on $u_2$, we have

$$
\begin{aligned}
|A_3(t)| &\leq ||\Psi||_\infty ||u_2||_{L^\infty(\mathbb{R}_+^2)} \int_0^t \left| \int_0^s \Psi(s', Y_1 + f_0)ds' - \int_0^s \Psi(s', Y_2 + f_0)ds' \right| ds \\
&\leq ||\Psi||_\infty \max(M, ||\Psi||_\infty)\text{Lip}(\Psi) \int_0^t \int_0^s |Y_1(t - s + s') - Y_2(t - s + s')|\, ds'ds \\
&\leq ||\Psi||_\infty \max(M, ||\Psi||_\infty)\text{Lip}(\Psi)\, t^2\, ||Y_1 - Y_2||_{L^\infty([0,t])}. \quad (56)
\end{aligned}
$$

where the arguments "$t - s + s'$" are not written in the first equation for simplicity.

**Study of $B$.** We have $B = B_1 + B_2$ with

$$
\begin{cases}
B_1(t) := \displaystyle\int_t^{+\infty} [\Psi(s, Y_1(t) + f_0(t)) - \Psi(s, Y_2(t) + f_0(t))] \\[2mm]
\hspace{4cm} u^{\text{in}}(s - t)\exp\left(-\int_0^t \Psi(s - t + t', Y_1(t') + f_0(t'))dt'\right)ds \\[2mm]
B_2(t) := \displaystyle\int_t^{+\infty} \Psi(s, Y_2(t) + f_0(t))u^{\text{in}}(s - t) \\[2mm]
\hspace{3cm} [e^{-\int_0^t \Psi(s-t+t', Y_1(t')+f_0(t'))dt'} - e^{-\int_0^t \Psi(s-t+t', Y_2(t')+f_0(t'))dt'}]ds.
\end{cases}
$$

- Study of $B_1$. Using the Lipschitz continuity of $\Psi$ and the fact that the argument in the exponential is non-positive, we have

$$
|B_1(t)| \leq \text{Lip}(\Psi) \int_t^{+\infty} |Y_1(t) - Y_2(t)|u^{\text{in}}(s - t)ds \leq \text{Lip}(\Psi)||Y_1 - Y_2||_{L^\infty([0,t])}, \quad (57)
$$

where we used that $\int_0^{+\infty} u^{\text{in}}(s)ds = 1$.

- Study of $B_2$. As for $A_3$, we have

$$
\begin{aligned}
|B_2(t)| &\leq ||\Psi||_\infty \int_t^{+\infty} u^{\text{in}}(s - t) \\
&\qquad \left| \int_0^t \Psi(s - t + t', Y_1(t') + f_0(t')) - \Psi(s - t + t', Y_2(t') + f_0(t'))dt' \right| ds \\
&\leq ||\Psi||_\infty \text{Lip}(\Psi) \int_t^{+\infty} u^{\text{in}}(s - t) \int_0^t |Y_1(t') - Y_2(t')|\, dt'ds \\
&\leq ||\Psi||_\infty \text{Lip}(\Psi)\, t\, ||Y_1 - Y_2||_{L^\infty([0,t])}, \quad (58)
\end{aligned}
$$

where we used once again that $\int_0^{+\infty} u^{\mathrm{in}}(s)ds = 1$.

Gathering (54), (55), (56), (57) and (58), we get that there exists a non-decreasing function $g$ such that for all $t > 0$,

$$|\tilde{u}(t)| \leq g(t)||Y_1 - Y_2||_{L^\infty([0,t])} + ||\Psi||_\infty \int_0^t |\tilde{u}(z)|dz.$$

Then, Lemma B.4-(i) gives that for all $T \geq 0$,

$$||\tilde{u}||_{L^\infty([0,T])} \leq C_T \sup_{t \in [0,T]} \left\{ g(t)||Y_1 - Y_2||_{L^\infty([0,t])} \right\},$$

with $C_T$ being a non-decreasing function of $T$. Then, (53) follows since $g$ is non-decreasing.

## A.5   Proof of Proposition III.7

-(i) Suppose that $(\overline{N}_t)_{t \geq 0}$ is a solution of (14). Then, the mean cumulative intensity $\overline{\Lambda}(t) = \mathbb{E}[\overline{N}_t]$ is a non-decreasing locally bounded function satisfying

$$\overline{\Lambda}(t) = \int_0^t \Psi_0 \left( \int_0^{t'} h(t' - z)d\overline{\Lambda}_z + f_0(t') \right) dt' \quad \text{for every } t \geq 0. \tag{59}$$

By Lemma B.5, we know that (59) admits a unique solution which is furthermore of class $\mathcal{C}^1$ and we denote $\overline{\lambda}$ its derivative. Thus, we have $\mathbb{E}\left[ \overline{N}_+(dt) \right] = \overline{\Lambda}'(t)dt = \overline{\lambda}(t)dt$.

-(ii) The first equation of (34) is a classical thinning equation so its solution $(\overline{N}_t)_{t \geq 0}$ is a measurable function of $\Pi$ hence it is unique (once $\Pi$ is fixed).

To conclude this step, it suffices to check that $(\overline{N}_t)_{t \geq 0}$ satisfies the second equation of (34) where we remind that $\overline{\lambda}$ is the derivative of $\overline{\Lambda}$ which is the unique solution of (59). But $\mathbb{E}[\overline{N}_t] = \int_0^t \Psi_0(\int_0^{t'} h(t'-z)\overline{\lambda}(z)dz + f_0(t'))dt'$, which is equal to $\overline{\Lambda}(t) = \int_0^t \overline{\lambda}(t')dt'$ since $\overline{\Lambda}$ is a solution of (59).

Finally, the two remaining points are rather simple. First, taking the derivative of (59) gives that $\overline{\lambda}(t) = \Psi_0 \left( \int_0^{t-} h(t - z)\overline{\lambda}(z)dz + f_0(t) \right)$. Secondly, the solution of (34) is clearly a solution of (14) and (i) tells that a solution of (14) is necessarily a solution of (34) which gives uniqueness.

## A.6   Proof of (43)

For simplicity of notations in (37) and (38), let us denote for all $t \geq 0$,

$$\begin{cases} \gamma_t^{n,i} := \dfrac{1}{n} \displaystyle\sum_{j=1}^n \left( \int_0^{t-} H_{ij}(t - z)N_+^{n,j}(dz) + F_{ij}(t) \right), \\[2mm] \overline{\gamma}(t) := \displaystyle\int_0^t m_{\mu_H}(t - z)\overline{\lambda}(z)dz + m_{\nu_F}(t), \end{cases} \tag{60}$$

where $\overline{\lambda}$ is defined either in Proposition III.6 (under $H_1$) or III.7 (under $H_2$). We have $\lambda_t^{n,i} = \Psi(S_{t-}^{n,i}, \gamma_t^{n,i})$ and $\overline{\lambda}_t^i = \Psi(\overline{S}_{t-}^i, \overline{\gamma}(t))$. Notice that $\overline{\gamma}$ is a deterministic function (which does not depend on $i$) whereas $\gamma^{n,i}$ is random.

**First point of (43).**   Assume that $H_1$ is satisfied. Then one can use the decomposition,

$$1 = \mathbb{1}_{\{S_{t-}^{n,1} = \overline{S}_{t-}^1\}} + \mathbb{1}_{\{S_{t-}^{n,1} \neq \overline{S}_{t-}^1\}} \tag{61}$$

and the fact that $|\lambda_t^{n,1} - \overline{\lambda}_t^1| \leq ||\Psi||_\infty$ to deduce from (42) that

$$\delta_n(\theta) \leq \int_0^\theta \mathbb{E}\left[ \left| \lambda_t^{n,1} - \overline{\lambda}_t^1 \right| \mathbb{1}_{\{S_{t-}^{n,1} = \overline{S}_{t-}^1\}} \right] dt + ||\Psi||_\infty \int_0^\theta \mathbb{P}\left( S_{t-}^{n,1} \neq \overline{S}_{t-}^1 \right) dt. \tag{62}$$

Let us denote

$$\begin{cases} A^n(\theta) := \displaystyle\int_0^\theta \mathbb{E}\left[ \left| \lambda_t^{n,1} - \overline{\lambda}_t^1 \right| \mathbb{1}_{\{S_{t-}^{n,1} = \overline{S}_{t-}^1\}} \right] dt, \\[2mm] D^n(\theta) := \displaystyle\int_0^\theta \mathbb{P}\left( S_{t-}^{n,1} \neq \overline{S}_{t-}^1 \right) dt. \end{cases}$$

**Study of $A^n(\theta)$.** Using the Lipschitz continuity of $\Psi$, it is clear that for all $t$ in $[0,\theta]$, if $S_{t-}^{n,1} = \overline{S}_{t-}^1$, then $|\lambda_t^{n,1} - \overline{\lambda}_t^1| \leq \mathrm{Lip}(\Psi)|\gamma_t^{n,1} - \overline{\gamma}(t)|$. So, one deduces that

$$A^n(\theta) \leq \mathrm{Lip}(\Psi)(B_1^n(\theta) + B_2^n(\theta) + B_3^n(\theta) + C^n(\theta)), \tag{63}$$

where

$$
\begin{cases}
B_1^n(\theta) := \int_0^\theta \mathbb{E}\Big[\Big|\int_0^t \frac{1}{n}\sum_{j=1}^n H_{1j}(t-z)[N_+^{n,j}(dz) - \overline{N}_+^j(dz)]\Big|\Big]dt, \\[3mm]
B_2^n(\theta) := \int_0^\theta \mathbb{E}\Big[\Big|\int_0^t \frac{1}{n}\sum_{j=1}^n H_{1j}(t-z)[\overline{N}_+^j(dz) - \overline{\lambda}_z^j dz]\Big|\Big]dt, \\[3mm]
B_3^n(\theta) := \int_0^\theta \mathbb{E}\Big[\Big|\int_0^t \frac{1}{n}\sum_{j=1}^n \Big(H_{1j}(t-z)\overline{\lambda}_z^j - m_{\mu_H}(t-z)\overline{\lambda}(z)\Big)dz\Big|\Big]dt, \\[3mm]
C^n(\theta) := \int_0^\theta \mathbb{E}\Big[\Big|\frac{1}{n}\sum_{j=1}^n F_{1j}(t) - m_{\nu_F}(t)\Big|\Big]dt.
\end{cases}
\tag{64}
$$

- <u>Study of $B_1^n$</u>. Firstly, using $(\mathcal{A}_\infty^{\mu_H})$ and then Lemma B.3, we have

$$
\begin{aligned}
B_1^n(\theta) &\leq \int_0^\theta \mathbb{E}\Big[\int_0^t \frac{1}{n}\sum_{j=1}^n \Big|H_{1j}(t-z)\Big|\Big|N_+^{n,j}(dz) - \overline{N}_+^j(dz)\Big|\Big]dt \\[2mm]
&\leq \int_0^\theta \mathbb{E}\Big[\int_0^t \frac{1}{n}\sum_{j=1}^n M_{\mu_H}(t-z)\Big|N_+^{n,j}(dz) - \overline{N}_+^j(dz)\Big|\Big]dt \\[2mm]
&= \int_0^\theta \frac{1}{n}\sum_{j=1}^n \mathbb{E}\Big[\int_0^t M_{\mu_H}(t-z)d\Delta_n^j(z)\Big]dt = \int_0^\theta M_{\mu_H}(\theta-z)\delta_n(z)dz, \tag{65}
\end{aligned}
$$

where the $\Delta_n^j$'s are given in (40).

- <u>Study of $B_2^n$</u>. Secondly, using Cauchy-Schwartz inequality, we have

$$
\begin{aligned}
B_2^n(\theta) &\leq \frac{1}{n}\int_0^\theta \mathbb{E}\Big[\Big|\sum_{j=1}^n \int_0^t H_{1j}(t-z)[\overline{N}_+^j(dz) - \overline{\lambda}_z^j dz]\Big|^2\Big]^{1/2}dt \\[2mm]
&= \frac{1}{n}\int_0^\theta \mathbb{E}\Big[\sum_{j=1}^n \int_0^t H_{1j}(t-z)^2\overline{\lambda}_z^j dz\Big]^{1/2}dt \\[2mm]
&\leq \frac{1}{n}\int_0^\theta \Big(n\int_0^t M_{\mu_H}(t-z)^2||\Psi||_\infty dz\Big)^{1/2}dt \\[2mm]
&= \frac{1}{\sqrt{n}}||\Psi||_\infty^{1/2}\int_0^\theta \Big(\int_0^t M_{\mu_H}(t-z)^2 dz\Big)^{1/2}dt := \frac{1}{\sqrt{n}}\tilde{B}_2(\theta), \tag{66}
\end{aligned}
$$

by computing the bracket of a sum over a compensated point process (see [19, Proposition II.4.1.]), using $(\mathcal{A}_\infty^{\mu_H})$ and the fact that $\overline{\lambda}^j$ is bounded by $||\Psi||_\infty$.

- <u>Study of $B_3^n$</u>. Let us fix some $t$ in $[0,\theta]$ and $z$ in $[0,t]$ and denote $Y_j := H_{1j}(t-z)\overline{\lambda}_z^j$. Since $\overline{\lambda}_z^j$ is the intensity of a solution of the limit equation, it is independent of $H_{1j}$. Moreover the $H_{1j}$'s are independent (see (8)) and the $\overline{\lambda}^j$'s are independent so the $Y_j$'s are independent. Hence,

$$\mathbb{E}[Y_j] = \mathbb{E}[H_{1j}(t-z)]\,\mathbb{E}\Big[\overline{\lambda}_z^j\Big] = m_{\mu_H}(t-z)\overline{\lambda}(z),$$

and

$$\mathrm{Var}(Y_j) = \mathrm{Var}(H_{1j}(t-z))\,\mathrm{Var}\Big(\overline{\lambda}_z^j\Big) + \mathrm{Var}(H_{1j}(t-z))\,\overline{\lambda}(z)^2 + m_{\mu_H}(t-z)^2\mathrm{Var}\Big(\overline{\lambda}_z^j\Big). \tag{67}$$

On the one hand, it follows from $(\mathcal{A}_\infty^{\mu_H})$ that $\mathrm{Var}(H_{1j}(s)) \leq M_{\mu_H}(s)^2$ for all $s \geq 0$. On the other hand, it follows from $(\mathcal{A}_\infty^\Psi)$ that $\mathrm{Var}(\overline{\lambda}_z^j) \leq ||\Psi||_\infty^2$ for all $z \geq 0$. Finally (67) leads to

$\mathbb{V}\mathrm{ar}\,(Y_j) \leq 3M_{\mu_H}(t-z)^2||\Psi||_\infty^2$ since $m_{\mu_H}$ is dominated by $M_{\mu_H}$ and $\overline{\lambda}$ is bounded by $||\Psi||_\infty$. Using the fact that the $Y_j$'s are independent, one has

$$\mathbb{V}\mathrm{ar}\left(\frac{1}{n}\sum_{j=1}^n Y_j\right) = \mathbb{E}\left[\left|\left|\frac{1}{n}\sum_{j=1}^n H_{1j}(t-z)\overline{\lambda}_z^j - m_{\mu_H}(t-z)\overline{\lambda}(z)\right|\right|^2\right] \leq \frac{3}{n}M_{\mu_H}(t-z)^2||\Psi||_\infty^2.$$

So, starting from the definition of $B_3^n$, thanks to Cauchy-Schwartz inequality, one has

$$\begin{aligned} B_3^n(\theta) &\leq & \int_0^\theta \int_0^t \left(\frac{3}{n}M_{\mu_H}(t-z)^2||\Psi||_\infty^2\right)^{1/2} dzdt \\ &\leq & \frac{\sqrt{3}}{\sqrt{n}}||\Psi||_\infty \int_0^\theta \int_0^t M_{\mu_H}(t-z)dzdt := \frac{1}{\sqrt{n}}\tilde{B}_3(\theta) \end{aligned} \tag{68}$$

- Study of $C^n$. Since for all $t \geq 0$, $F_{11}(t),\ldots,F_{1n}(t)$ are i.i.d. random variables (see (9)) with expectation $m_{\nu_F}(t)$ and variance $V_{\nu_F}(t)$, one deduces from Cauchy-Schwartz inequality that

$$C^n(\theta) \leq \frac{1}{\sqrt{n}}\int_0^\theta V_{\nu_F}(t)^{1/2}dt := \frac{1}{\sqrt{n}}\tilde{C}(\theta). \tag{69}$$

Finally, one deduces from (63), (65), (66), (68) and (69) that

$$A^n(\theta) \leq \frac{\mathrm{Lip}(\Psi)}{\sqrt{n}}\left(\tilde{B}_2(\theta) + \tilde{B}_3(\theta) + \tilde{C}(\theta)\right) + \mathrm{Lip}(\Psi)\int_0^\theta M_{\mu_H}(\theta - z)\delta_n(z)dz. \tag{70}$$

**Study of $D^n(\theta)$.** Since the initial conditions of $N^{n,1}$ and $\overline{N}^1$ are the same (equal to $N_-^1$), it holds that $S_0^{n,1} = \overline{S}_0^1$ a.s. If, at a fixed time $t \geq 0$, $S_{t-}^{n,1} \neq \overline{S}_{t-}^1$, then there is at least one point between 0 and $t$ which is not common to both $N_+^{n,1}$ and $\overline{N}_+^1$, that is $\sup_{t'\in[0,t]}|N_{t'}^{n,i} - \overline{N}_{t'}^i| \neq 0$, hence $\mathbb{P}(S_{t-}^{n,1} \neq \overline{S}_{t-}^1) \leq \mathbb{P}(\sup_{t'\in[0,t]}|N_{t'}^{n,i} - \overline{N}_{t'}^i| \neq 0)$. Moreover, since counting processes are piecewise constant with jumps of height 1 a.s., it is clear that

$$\mathbb{P}\left(\sup_{t'\in[0,t]}|N_{t'}^{n,1} - \overline{N}_{t'}^1| \neq 0\right) = \mathbb{P}\left(\sup_{t'\in[0,t]}|N_{t'}^{n,1} - \overline{N}_{t'}^1| \geq 1\right) \leq \mathbb{E}\left[\sup_{t'\in[0,t]}|N_{t'}^{n,1} - \overline{N}_{t'}^1|\right],$$

where we used Markov's inequality. Using (41), one has $\mathbb{P}\left(S_{t-}^{n,1} \neq \overline{S}_{t-}^1\right) \leq \delta_n(t)$ and so

$$D^n(\theta) \leq \int_0^\theta \delta_n(t)dt. \tag{71}$$

Rewriting (62) under the form $\delta_n(\theta) \leq A^n(\theta) + ||\Psi||_\infty D^n(\theta)$, one deduces from (70) and (71) that

$$\delta_n(\theta) \leq \frac{\mathrm{Lip}(\Psi)}{\sqrt{n}}\left(\tilde{B}_2(\theta) + \tilde{B}_3(\theta) + \tilde{C}(\theta)\right) + \int_0^\theta \left[||\Psi||_\infty + \mathrm{Lip}(\Psi)M_{\mu_H}(\theta - z)\right]\delta_n(z)dz \tag{72}$$

where $\tilde{B}_2$, $\tilde{B}_3$, $\tilde{C}$ are respectively defined in (66), (68) and (69). Since $M_{\mu_H}$ is locally square integrable, $\theta \mapsto \tilde{B}_2(\theta)$ is locally bounded; since $M_{\mu_H}$ is locally integrable, $\theta \mapsto \tilde{B}_3(\theta)$ is locally bounded; since $V_{\nu_F}$ is locally square root integrable, $\theta \mapsto \tilde{C}(\theta)$ is locally bounded. Hence we proved the first point of (43).

**Second point of (43).** Assume that $H_2$ is satisfied. Then the decomposition (61) is not helpful anymore. Whatever the age processes are, one always has $\lambda_t^{n,i} = \Psi_0(\gamma_t^{n,i})$ and $\overline{\lambda}_t^i = \Psi_0(\overline{\gamma}(t))$ for all $i = 1,\ldots,n$ and $t > 0$. Remark that in this case, the intensities $\overline{\lambda}^i$ of the limit processes defined by (38) are deterministic and equal to $\overline{\lambda}$ defined in Proposition III.7. Instead of (62) one should start from $\delta_n(\theta) = \int_0^\theta \mathbb{E}[|\lambda_t^{n,1} - \overline{\lambda}_t^1|]dt$. One can prove in the same way as above that

$$\delta_n(\theta) \leq \mathrm{Lip}(\Psi_0)(B_1^n(\theta) + B_2^n(\theta) + B_3^n(\theta) + C^n(\theta)), \tag{73}$$

where $B_1^n$, $B_2^n$, $B_3^n$ and $C^n$ are defined by (64). Since the uniform boundedness of $\Psi$ was not used in the study of $B_1^n$ and $C^n$ then (65) and (69) still hold. It remains to control $B_2^n$ and $B_3^n$ under 1ssumption $(\mathcal{A}_{\Psi=\Psi_0})$.

- Study of $B_2^n$. Firstly, remind that for all $j$, $\overline{\lambda}_t^j = \overline{\lambda}(t)$ so, using Cauchy-Schwartz inequality, we have

$$
\begin{aligned}
B_2^n(\theta) &\leq \frac{1}{n} \int_0^\theta \mathbb{E}\Big[ \Big| \sum_{j=1}^n \int_0^t H_{1j}(t-z)[\overline{N}_+^j(dz) - \overline{\lambda}(z)dz] \Big|^2 \Big]^{1/2} dt \\
&= \frac{1}{n} \int_0^\theta \mathbb{E}\Big[ \sum_{j=1}^n \int_0^t H_{1j}(t-z)^2 \overline{\lambda}(z)dz \Big]^{1/2} dt \\
&\leq \frac{1}{\sqrt{n}} \int_0^\theta \left( \int_0^t M_{\mu_H}(t-z)^2 \overline{\lambda}(z)dz \right)^{1/2} dt := \frac{1}{\sqrt{n}}\overline{B}_2(\theta),
\end{aligned}
\tag{74}
$$

by computing the bracket of a sum over a compensated point process (see [19, Proposition II.4.1.]) and then using $(\mathcal{A}_\infty^{\mu_H})$.

- Study of $B_3^n$. Secondly, since $\overline{\lambda}_t^j = \overline{\lambda}(t)$ for all $j$, then $B_3^n$ rewrites as

$$
B_3^n(\theta) = \int_0^\theta \mathbb{E}\Big[ \Big| \int_0^t \frac{1}{n} \sum_{j=1}^n (H_{1j}(t-z) - m_{\mu_H}(t-z)) \overline{\lambda}(z)dz \Big| \Big] dt
$$

$$
\leq \int_0^\theta \int_0^t \mathbb{E}\Big[ \Big| \frac{1}{n} \sum_{j=1}^n (H_{1j}(t-z) - m_{\mu_H}(t-z)) \Big| \Big] \overline{\lambda}(z)dzdt.
$$

Using Cauchy-Schwartz inequality, the fact that the $H_{1j}$'s are i.i.d. with mean function $m_{\mu_H}$ and that for all $s \geq 0$, $\mathbb{V}\text{ar}(H_{1j}(s)) \leq M_{\mu_H}(s)^2$ (which follows from $(\mathcal{A}_\infty^{\mu_H})$), we have

$$
B_3^n(\theta) \leq \int_0^\theta \int_0^t \frac{1}{\sqrt{n}} \mathbb{V}\text{ar}\left(H_{1j}(t-z)\right)^{1/2} \overline{\lambda}(z)dzdt \leq \frac{1}{\sqrt{n}} \int_0^\theta \int_0^t M_{\mu_H}(t-z)\overline{\lambda}(z)dzdt := \frac{1}{\sqrt{n}}\overline{B}_3(\theta).
\tag{75}
$$

Finally, one deduces from (73), (65), (74), (75) and (69) that

$$
\delta_n(\theta) \leq \frac{\text{Lip}(\Psi_0)}{\sqrt{n}} \left( \overline{B}_2(\theta) + \overline{B}_3(\theta) + \tilde{C}(\theta) \right) + \int_0^\theta \text{Lip}(\Psi_0) M_{\mu_H}(\theta-z)\delta_n(z)dz,
\tag{76}
$$

where $\overline{B}_2$, $\overline{B}_3$, $\tilde{C}$ are respectively defined in (74), (75) and (69).

Since $M_{\mu_H}$ is locally square integrable and $\overline{\lambda}$ is continuous, $\theta \mapsto \overline{B}_2(\theta)$ is locally bounded; since $M_{\mu_H}$ is locally integrable and $\overline{\lambda}$ is continuous, $\theta \mapsto \overline{B}_3(\theta)$ is locally bounded; since $V_{\nu_F}$ is locally square root integrable, $\theta \mapsto \tilde{C}(\theta)$ is locally bounded. Hence we proved the second point of (43).

## A.7   Proof of Proposition IV.2

-1. Assume that $H_1$ is satisfied. The mean intensity $\overline{\lambda}$ defined in Proposition III.6 is clearly uniformly bounded by $||\Psi||_\infty$.

Looking at (72), one wants to find some uniform bounds on $\tilde{B}_2$, $\tilde{B}_3$, $\tilde{C}$ respectively defined in (66), (68) and (69). Firstly, since $M_{\mu_H}$ is square integrable and $\overline{\lambda}$ is uniformly bounded, it is clear from (66) that

$$
\tilde{B}_2(\theta) \leq ||M_{\mu_H}||_2 ||\Psi||_\infty^{1/2} \theta.
\tag{77}
$$

Secondly, using the integrability of $M_{\mu_H}$ and the boundedness of $\overline{\lambda}$, one deduces from (68) that

$$
\tilde{B}_3(\theta) \leq \sqrt{3} ||M_{\mu_H}||_1 ||\Psi||_\infty \theta.
\tag{78}
$$

Finally, since $V_{\nu_F}$ is uniformly bounded, one deduces from (69) that

$$
\tilde{C}(\theta) \leq ||V_{\nu_F}||_\infty^{1/2} \theta.
\tag{79}
$$

Moreover, using the fact that $\delta_n$ is a non-decreasing function, we find (with $\alpha = \mathrm{Lip}(\Psi)||M_{\mu_H}||_1$)

$$\begin{cases} \int_0^\theta \mathrm{Lip}(\Psi) M_{\mu_H}(\theta - z)\delta_n(z)dz \le \alpha\delta_n(\theta) \\ \int_0^\theta ||\Psi||_\infty \delta_n(z)dz \le ||\Psi||_\infty \theta \delta_n(\theta). \end{cases} \tag{80}$$

Gathering (77), (78), (79) and (80) one deduces from (72) that

$$\delta_n(\theta) \le \frac{\mathrm{Lip}(\Psi)}{\sqrt{n}} \left( ||M_{\mu_H}||_2 ||\Psi||_\infty^{1/2} + \sqrt{3}||M_{\mu_H}||_1 ||\Psi||_\infty + ||V_{\nu_F}||_\infty^{1/2} \right)\theta + (\alpha + ||\Psi||_\infty\theta)\delta_n(\theta),$$

which leads to $\delta_n(\theta) \le \beta(\Psi, \mu_H, \nu_F)\theta n^{-1/2}$ with

$$\beta(\Psi, \mu_H, \nu_F) := \frac{\mathrm{Lip}(\Psi)}{1 - (\alpha + ||\Psi||_\infty\theta)} \left( ||M_{\mu_H}||_2 ||\Psi||_\infty^{1/2} + \sqrt{3}||M_{\mu_H}||_1 ||\Psi||_\infty + ||V_{\nu_F}||_\infty^{1/2} \right), \tag{81}$$

as soon as $\alpha + ||\Psi||_\infty\theta < 1$, i.e. $\theta < (1 - \alpha)/||\Psi||_\infty$.

-2. Assume $H_2$ is satisfied. Under the assumptions of Proposition IV.2, the mean intensity $\overline{\lambda}$ defined in Proposition III.7 is uniformly bounded thanks to Lemma B.6.

Looking at (76), one wants to find some uniform bounds on $\overline{B}_2$, $\overline{B}_3$, $\tilde{C}$ respectively defined in (74), (75) and (69). In the same way as above, one can deduce from (74), (75) and (69) that

$$\begin{cases} \overline{B}_2(\theta) \le ||M_{\mu_H}||_2 ||\overline{\lambda}||_\infty^{1/2}\theta \\ \overline{B}_3(\theta) \le ||M_{\mu_H}||_1 ||\overline{\lambda}||_\infty\theta \\ \tilde{C}(\theta) \le ||V_{\nu_F}||_\infty^{1/2}\theta. \end{cases} \tag{82}$$

Using (82) and the first equation of (80) which is still valid is this case, one deduces from (76) that

$$\delta_n(\theta) \le \frac{\mathrm{Lip}(\Psi)}{\sqrt{n}} \left( ||M_{\mu_H}||_2 ||\overline{\lambda}||_\infty^{1/2} + ||M_{\mu_H}||_1 ||\overline{\lambda}||_\infty + ||V_{\nu_F}||_\infty^{1/2} \right)\theta + \alpha\delta_n(\theta),$$

which leads to $\delta_n(\theta) \le \beta(\Psi, \mu_H, \nu_F)\theta n^{-1/2}$ with

$$\beta(\Psi, \mu_H, \nu_F) := \frac{\mathrm{Lip}(\Psi)}{1 - \alpha} \left( ||M_{\mu_H}||_2 ||\overline{\lambda}||_\infty^{1/2} + ||M_{\mu_H}||_1 ||\overline{\lambda}||_\infty + ||V_{\nu_F}||_\infty^{1/2} \right), \tag{83}$$

for every $\theta \ge 0$. Notice that an explicit expression of $||\overline{\lambda}||_\infty$ with respect to $M_{\mu_H}$, $m_{\nu_F}$ and $\Psi$ can be obtained thanks to Lemma B.6.

# B  Lemmas

## B.1  Point processes

Here we collect some technical lemmas about point processes.

The following lemma is used to show the well-posedness of the studied point processes.

**Lemma B.1.** *Let $n \ge 1$ be an integer, let $(g_i)_{i=1,\ldots,n}$ be a family of locally integrable functions, $(a_i)_{i=1,\ldots,n}$ be a family of non-negative real numbers and $h : \mathbb{R}_+ \to \mathbb{R}$ be a locally integrable function. Let $(\Pi^i(dt, dx))_{i \ge 1}$ be some i.i.d. $\mathbb{F}$-Poisson measures with intensity 1 on $\mathbb{R}_+^2$.*

*Let $(N_t^i)_{t \ge 0}^{i=1,\ldots,n}$ be a family of counting processes such that for $i = 1,..,n$ and all $t \ge 0$ ,*

$$N_t^i = \int_0^t \int_0^\infty \mathbb{1}_{\left\{ x \le \lambda_{t'}^i \right\}} \Pi^i(dt', dx), \tag{84}$$

*where the $\lambda^i$'s are $\mathbb{F}$-predictable processes such that $\lambda_t^i \le g_i(t) + a_i\frac{1}{n}\sum_{j=1}^n \int_0^{t-} |h(t - z)|N^j(dz)$. Then, the linear multivariate Hawkes process $(\tilde{N}_t^i)_{t \ge 0}^{i=1,\ldots,n}$ defined by*

$$\tilde{N}_t^i = \int_0^t \int_0^\infty \mathbb{1}_{\left\{ x \le g_i(t') + a_i\frac{1}{n}\sum_{j=1}^n \int_0^{t'-} |h(t' - z)|\tilde{N}^j(dz) \right\}} \Pi^i(dt', dx), \tag{85}$$

*is such that for all $i = 1, \ldots, n$, $\tilde{N}^i$ stochastically dominates $N^i$ in the sense that $N^i \subset \tilde{N}^i$ where $N^i$ (resp. $\tilde{N}^i$) denotes the point process associated with the counting process $(N_t^i)_{t \ge 0}$ (resp. $(\tilde{N}_t^i)_{t \ge 0}$). In particular, the processes $(N_t^i)_{t \ge 0}^{i=1,\ldots,n}$ are well-defined.*

*Proof.* First, let us note that the processes $(\tilde{N}_t^i)_{t\geq 0}^{i=1,..,n}$ are well-defined by the Galton-Watson representation of the linear Hawkes process introduced in [23] when the $g_i$'s are constant in time (see [8, Proposition B.3] when the $g_i$'s are more generally locally integrable functions).

We are going to show by induction that

$$\forall t \geq 0, \ \lambda_t^i \leq \tilde{\lambda}_t^i := g_i(t) + a_i \frac{1}{n} \sum_{j=1}^{n} \int_0^{t-} |h(t-u)| \tilde{N}^j(du).$$

Indeed, for all time $t$ less than the first point of either $N := \cup_{i=1}^n N^i$ or $\tilde{N} := \cup_{i=1}^n \tilde{N}^i$, the respective intensities are such that $\lambda_t^i \leq g_i(t) = \tilde{\lambda}_t^i$. Hence, the first point of $N \cup \tilde{N}$ (denoted by $T_1$) is a point of $\tilde{N}$ (and possibly a point of $N$). Let us denote $(T_k)_{k\geq 1}$ the ordered sequence of the points of $N \cup \tilde{N}$.

Let us fix some $k_0 \geq 1$. Suppose that for every $i = 1, \ldots, n$, $\lambda_t^i \leq \tilde{\lambda}_t^i$, for all $t \leq T_{k_0}$. Then, it is clear that for every $k = 1, \ldots, k_0$, $T_k \in \tilde{N}$, hence for every $i = 1, \ldots, n$, $\tilde{N}^i$ stochastically dominates $N^i$ up to time $T_{k_0+1-}$. Moreover, it implies that for every $i = 1, \ldots, n$ and for all $t$ in $(T_{k_0}, T_{k_0+1}]$,

$$
\begin{aligned}
\lambda_t^i &\leq & g_i(t) + a_i \frac{1}{n} \sum_{j=1}^{n} \int_0^{t-} |h(t-z)| N^j(dz) \\
&\leq & g_i(t) + a_i \frac{1}{n} \sum_{j=1}^{n} \int_0^{t-} |h(t-z)| \tilde{N}^j(dz) = \tilde{\lambda}_t^i,
\end{aligned}
$$

since $|h|$ is a non negative function. Therefore, by induction on $k$, the desired stochastic domination holds true for all time. In particular, the dominated processes are well-defined. $\square$

**Lemma B.2.** *If $N$ admits the bounded $\mathbb{F}$-intensity $\lambda_t$ and $(S_{t-})_{t\geq 0}$ denote its associated age process, then the distribution of $S_{t-}$ denoted by $w_t$ is such that $t \mapsto w_t$ belongs to $\mathcal{BC}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}_+))$.*

*Proof.* This continuity result comes from the fact that the probability that $N$ has a point in an interval goes to 0 as the size of the interval goes to 0. Indeed, let $t, t'$ be positive real numbers, $\mathbb{P}\left(\overline{N}([t, t+t']) \neq 0\right) \leq \mathbb{E}\left[\overline{N}([t, t+t'])\right] = \mathbb{E}[\int_t^{t+t'} \lambda_z dz]$ goes to 0 as $t'$ goes to 0. Moreover, $S_{(t+t')-} = S_{t-} + t'$ as soon as there is no point of $N$ in the interval $[t, t+t')$ and so one has

$$\tilde{W}_1(w_{t+t'}, w_t) \leq \mathbb{E}\left[\min\left(\left|S_{(t+t')-} - S_{t-}\right|, 1\right)\right] \leq t' + \mathbb{P}\left(\overline{N}([t, t+t']) \neq 0\right) \xrightarrow[t' \to 0]{} 0,$$

reminding that $\tilde{W}_1$ denotes the modified Wasserstein distance defined in (16). The same argument for $t' < 0$ gives continuity. $\square$

## B.2 Analytic lemmas

Here, we collect some analytic lemmas regarding the convolution equations used throughout the present paper. First, here are two lemmas introduced in [13].

The first one is an easy application of Fubini's Theorem [13, Lemma 22].

**Lemma B.3.** *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}$ be locally integrable and let $\alpha : \mathbb{R}_+ \to \mathbb{R}$ have bounded variations on compact intervals, satisfying $\alpha(0) = 0$. Then, for all $t \geq 0$,*

$$\int_0^t \int_0^s \Phi(s-u) \, d\alpha(u) \, ds = \int_0^t \Phi(t-s) \alpha(s) \, ds,$$

*where the integral has to be understood in the Stieltjes' sense.*

The second one is a rather classical generalization of Grönwall Lemma [13, Lemma 23].

**Lemma B.4.** *Let $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ be locally integrable and $g : \mathbb{R}_+ \to \mathbb{R}_+$ be locally bounded.*

*(i) Consider a locally bounded non-negative function $u$ such that for all $t \geq 0$, $u_t \leq g_t + \int_0^t \Phi(t-s) u_s ds$. Then, $\sup_{t\in[0,T]} u_t \leq C_T \sup_{t\in[0,T]} g_t$, for some constant $C_T$ depending only on $T > 0$ and $\Phi$. Moreover, $C_T$ can be taken as a non-decreasing function of $T$.*

(ii) *Consider a sequence of locally bounded non-negative functions $u^n$ such that for all $t \geq 0$, all $n \geq 0$, $u_t^{n+1} \leq \int_0^t \Phi(t-s) u_s^n ds$. Then, $\sup_{t \in [0,T]} \sum_{n \geq 0} u_t^n \leq C_T$, for some constant $C_T$ depending only on $T > 0$, $u^0$ and $\Phi$.*

(iii) *Consider a sequence of locally bounded non-negative functions $u^n$ such that for all $t \geq 0$, all $n \geq 0$, $u_t^{n+1} \leq g_t + \int_0^t \Phi(t-s) u_s^n ds$. Then, $\sup_{t \in [0,T]} \sup_{n \geq 0} u_t^n \leq C_T$, for some constant $C_T$ depending only on $T > 0$, $u^0$, $g$ and $\Phi$.*

Note that we added to the first statement that $C_T$ can be taken as a non-decreasing function of $T$. It is not given in [13, Lemma 23] but it is a direct consequence of the proof.

Then, here is a well-posedness result which is a generalization of [13, Lemma 24].

**Lemma B.5.** *Let $\Phi : \mathbb{R} \to \mathbb{R}_+$ be Lipschitz-continuous, $h : \mathbb{R}_+ \to \mathbb{R}$ be locally integrable and $f_0 : \mathbb{R}_+ \to \mathbb{R}$ be continuous. The equation*

$$m_t = \int_0^t \Phi \left( \int_0^{t'} h(t'-z) dm_z + f_0(t') \right) dt' \tag{86}$$

*has a unique locally bounded solution. Furthermore, $m$ is of class $\mathcal{C}^1$ on $\mathbb{R}_+$.*

*Proof.* The proof is similar to [13, Lemma 24]. We refrain from reproducing it here; instead, we only indicate the minor changes that are required to make it fit the current framework, i.e. the addition of the function $f_0$. The "uniqueness" part is exactly the same. The "existence" part requires $f_0$ to be locally integrable in order to have locally boundedness in the Picard iteration. Finally, we need $f_0$ to be continuous to show by induction that at each step of the Picard iteration the function is $\mathcal{C}^1$ on $\mathbb{R}_+$ and so it is for the limit, that is the solution of (86). $\square$

Here is given an analytic result which is used to give a uniform bound on the mean intensity of a Hawkes process under stationary conditions.

**Lemma B.6.** *Let $\Phi : \mathbb{R} \to \mathbb{R}_+$ be Lipschitz-continuous and $h : \mathbb{R}_+ \to \mathbb{R}$ be integrable such that $\mathrm{Lip}(\Phi)||h||_1 < 1$. Moreover, let $f_0 : \mathbb{R}_+ \to \mathbb{R}$ be uniformly bounded.*

*If $g : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous function satisfying*

$$g(t) \leq \Phi \left( \int_0^t h(t-u) g(u) du + f_0(t) \right) \tag{87}$$

*for every $t > 0$, then $g$ is uniformly upper bounded by*

$$M := \frac{\Phi(0) + \mathrm{Lip}(\Phi)||f_0||_\infty}{1 - \mathrm{Lip}(\Phi)||h||_1}.$$

*Proof.* For any $t > 0$,

$$g(t) \leq \Phi(0) + \mathrm{Lip}(\Phi) \left( \int_0^t |h(t-u)| g(u) du + ||f_0||_\infty \right)$$

hence, thanks to the continuity of $g$, for every $T \geq 0$,

$$\sup_{t \in [0,T]} g(t) \leq \Phi(0) + \mathrm{Lip}(\Phi) \left( ||h||_1 \sup_{t \in [0,T]} g(t) + ||f_0||_\infty \right)$$

and

$$\sup_{t \in [0,T]} g(t) \leq \frac{1}{1 - \mathrm{Lip}(\Phi)||h||_1} [\Phi(0) + \mathrm{Lip}(\Phi)||f_0||_\infty] = M.$$

$\square$

# References

[1] E. Bacry, K. Dayri, and J.-F. Muzy. Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *The European Physical Journal B-Condensed Matter and Complex Systems*, 85(5):1–12, 2012.

[2] E. Bacry, S. Delattre, M. Hoffmann, and J. F. Muzy. Scaling limits for Hawkes processes and application to financial statistics, Feb. 2012.

[3] P. Bao, H.-W. Shen, X. Jin, and X.-Q. Cheng. Modeling and predicting popularity dynamics of microblogs using self-excited hawkes processes. *arXiv preprint arXiv:1503.02754*, 2015.

[4] P. Brémaud. *Point processes and queues.* Springer-Verlag, New York, 1981. Martingale dynamics, Springer Series in Statistics.

[5] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *Ann. Probab.*, 24(3):1563–1588, 1996.

[6] N. Brunel and V. Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural computation*, 11(7):1621–1671, 1999.

[7] J. A. Cañizo, J. A. Carrillo, and S. Cuadrado. Measure solutions for some models in population dynamics. *Acta applicandae mathematicae*, 123(1):141–156, 2013.

[8] J. Chevallier, M. J. Caceres, M. Doumic, and P. Reynaud-Bouret. Microscopic approach of a time elapsed neural model. arXiv:1506.02361, to appear in M3AS, 2015.

[9] E. Chornoboy, L. Schramm, and A. Karr. Maximum likelihood identification of neural point process systems. *Biological Cybernetics*, 59(4-5):265–275, 1988.

[10] R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, 105(41):15649–15653, 2008.

[11] F. Delarue, J. Inglis, S. Rubenthaler, and E. Tanré. Particle systems with a singular mean-field self-excitation. application to neuronal networks. *Stochastic Processes and their Applications*, 125(6):2451–2492, 2015.

[12] F. Delarue, J. Inglis, S. Rubenthaler, E. Tanré, et al. Global solvability of a networked integrate-and-fire model of mckean–vlasov type. *The Annals of Applied Probability*, 25(4):2096–2133, 2015.

[13] S. Delattre, N. Fournier, and M. Hoffmann. High dimensional hawkes processes. *arXiv preprint arXiv:1403.5764*, 2014.

[14] O. Faugeras and J. Maclaurin. Asymptotic description of stochastic neural networks. i. existence of a large deviation principle. *Comptes Rendus Mathematique*, 352(10):841–846, 2014.

[15] O. Faugeras, J. Touboul, and B. Cessac. A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs. *Frontiers in computational neuroscience*, 3, 2009.

[16] N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, pages 1–32, 2014.

[17] N. Fournier and E. Löcherbach. On a toy model of interacting neurons. *arXiv preprint arXiv:1410.3263*, 2014.

[18] A. Galves and E. Löcherbach. Modeling networks of spiking neurons as interacting processes with memory of variable length. *arXiv preprint arXiv:1502.06446*, 2015.

[19] R. D. Gill, N. Keiding, and P. K. Andersen. *Statistical models based on counting processes.* Springer, 1997.

[20] G. Gusto and S. Schbath. Fado: A statistical method to detect favored or avoided distances between occurrences of motifs using the hawkes' model. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 2005.

[21] N. R. Hansen, P. Reynaud-Bouret, V. Rivoirard, et al. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143, 2015.

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

[23] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, pages 493–503, 1974.

[24] P. Hodara and E. Löcherbach. Hawkes processes with variable length memory and an infinite number of components. *arXiv preprint arXiv:1410.5235*, 2014.

[25] Y. Y. Kagan. Statistical distributions of earthquake numbers: consequence of branching process. *Geophysical Journal International*, 180(3):1313–1328, 2010.

[26] A. Klenke. *Probability theory: a comprehensive course.* Springer, 2007.

[27] P. A. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

[28] T. J. Liniger. *Multivariate hawkes processes.* PhD thesis, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 18403, 2009, 2009.

[29] E. Luçon and W. Stannat. Mean field limit for disordered diffusions with singular interactions. *The Annals of Applied Probability*, 24(5):1946–1993, 2014.

[30] L. Massoulié. Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic processes and their applications*, 75(1):1–30, 1998.

[31] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995)*, volume 1627 of *Lecture Notes in Math.*, pages 42–95. Springer, Berlin, 1996.

[32] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493), 2011.

[33] I. Muni Toke and F. Pomponio. Modelling trades-through in a limited order book using hawkes processes. *Economics discussion paper*, (2011-32), 2011.

[34] Y. Ogata. On Lewis simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–30, 1981.

[35] Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

[36] K. Pakdaman, B. Perthame, and D. Salort. Dynamics of a structured neuron population. *Nonlinearity*, 23(1):55, 2010.

[37] K. Pakdaman, B. Perthame, and D. Salort. Relaxation and self-sustained oscillations in the time elapsed neuron network model. *SIAM Journal on Applied Mathematics*, 73(3):1260–1279, 2013.

[38] K. Pakdaman, B. Perthame, and D. Salort. Adaptation and fatigue model for neuron networks and large time asymptotics in a nonlinear fragmentation equation. *The Journal of Mathematical Neuroscience (JMN)*, 4(1):1–26, 2014.

[39] B. Perthame. *Transport equations in biology.* Springer Science & Business Media, 2006.

[40] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

[41] C. Pouzat and A. Chaffiol. Automatic spike train analysis and report generation. an implementation with r, {R2HTML} and {STAR}. *Journal of Neuroscience Methods*, 181(1):119 – 144, 2009.

[42] C. Quininao. A microscopic spiking neuronal network for the age-structured model. *arXiv preprint arXiv:1503.00115*, 2015.

[43] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *The Journal of Mathematical Neuroscience (JMN)*, 4(1):1–41, 2014.

[44] P. Reynaud-Bouret, S. Schbath, et al. Adaptive estimation for hawkes processes; application to genome analysis. *The Annals of Statistics*, 38(5):2781–2822, 2010.

[45] F. Roueff, R. Von Sachs, and L. Sansonnet. Time-frequency analysis of locally stationary hawkes processes. prepublication on HAL, 2015.

[46] A.-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

[47] L. Zhu. *Nonlinear Hawkes Processes*. PhD thesis, New York University, 2013.