Microscopic approach of a time elapsed neural model

Julien Chevallier
Laboratoire J. A. Dieudonné, UMR CNRS 6621, Université de Nice Sophia-Antipolis, Parc Valrose
06108 Nice Cedex 2, France
julien.chevallier@unice.fr

María José Cáceres
Departamento de Matemática Aplicada , Universidad de Granada, Campus de Fuentenueva E-18071 Granada, Spain
caceresg@ugr.es

Marie Doumic
UPMC University of Paris 6, JL Lions Lab., 4 place Jussieu
75005 Paris, France

Patricia Reynaud-Bouret
Laboratoire J. A. Dieudonné, UMR CNRS 6621, Université de Nice Sophia-Antipolis, Parc Valrose
06108 Nice Cedex 2, France
patricia.reynaud-bouret@unice.fr

The spike trains are the main components of the information processing in the brain. To model spike trains several point processes have been investigated in the literature. And more macroscopic approaches have also been studied, using partial differential equation models. The main aim of the present article is to build a bridge between several point processes models (Poisson, Wold, Hawkes) that have been proved to statistically fit real spike trains data and age-structured partial differential equations as introduced by Pakdaman, Perthame and Salort.

Keywords: Hawkes process; Wold process; renewal equation; neural network

AMS Subject Classification:35F15, 35B10, 92B20, 60G57, 60K15

Introduction

In Neuroscience, the action potentials (spikes) are the main components of the real-time information processing in the brain. Indeed, thanks to the synaptic integration, the membrane voltage of a neuron depends on the action potentials emitted by some others, whereas if this membrane potential is sufficiently high, there is production of action potentials.

To access those phenomena, schematically, one can proceed in two ways: extracellularly record in vivo several neurons, at a same time, and have access to simultaneous spike trains (only the list of events corresponding to action potentials) or intracellularly record the whole membrane voltage of only one neuron at a time, being blind to the nearby neurons.

Many people focus on spike trains. Those data are fundamentally random and can be modelled easily by time point processes, i.e. random countable sets of points on $\mathbb{R}_+$. Several point processes models have been investigated in the literature, each of them reproducing different features of the neuronal reality. The easiest model is the homogeneous Poisson process, which can only reproduce a constant firing rate for the neuron, but which, in particular, fails to reproduce refractory periods[a]. It is commonly admitted that this model is too poor to be realistic. Indeed, in such a model, two points or spikes can be arbitrary close as soon as their overall frequency is respected in average. Another more realistic model is the renewal process [37], where the occurrence of a point or spike depends on the previous occurrence. More precisely, the distribution of delays between spikes (also called inter-spike intervals, ISI) is given and a distribution, which provides small weights to small delays, is able to mimic refractory periods. A deeper statistical analysis has shown that Wold processes is showing good results, with respect to goodness-of-fit test on real data sets [38]. Wold processes are point processes for which the next occurrence of a spike depends on the previous occurrence but also on the previous ISI. From another point of view, the fact that spike trains are usually non stationary can be easily modelled by inhomogeneous Poisson processes [43]. All those models do not reflect one of the main features of spike trains, which is the synaptic integration and there has been various attempts to catch such phenomenon. One of the main model is the Hawkes model, which has been introduced in [13] and which has been recently shown to fit several stationary data [40]. Several studies have been done in similar directions (see for instance [5]). More recently a vast interest has been shown to generalized linear models [36], with which one can infer functional connectivity and which are just an exponential variant of Hawkes models.

There has also been several models of the full membrane voltage such as Hodgkin-Huxley models. It is possible to fit some of those probabilistic stochastic differential equations (SDE) on real voltage data [22] and to use them to estimate meaningful physiological parameters [18]. However, the lack of simultaneous data (voltages of different neurons at the same time) prevent these models to be used as statistical models that can be fitted on network data, to estimate network parameters. A simple SDE model taking synaptic integration into account is the well-known Integrate-and-Fire (IF) model. Several variations have been proposed to describe several features of real neural networks such as oscillations [7,8]. In particular, there exists hybrid IF models including inhomogeneous voltage driven Poisson process [21] that are able to mimic real membrane potential data. However up to our knowledge

[a]Biologically, a neuron cannot produce two spikes too closely in time.

and unlike point processes models, no statistical test have been applied to show that any of the previous variations of the IF model fit real network data.

Both, SDE and point processes, approaches are microscopic descriptions, where random noise explains the intrinsic variability. Many authors have argued that there must be some more macroscopic approach describing huge neural networks as a whole, using PDE formalism [15,42]. Some authors have already been able to perform link between PDE approaches as the macroscopic system and SDE approach (in particular IF models) as the microscopic model [39,30,26]. Another macroscopic point of view on spike trains is proposed by Pakdaman, Perthame and Salort in a series of articles [31,32,33]. It uses a nonlinear age-structured equation to describe the spikes density. Adopting a population view, they aim at studying relaxation to equilibrium or spontaneous periodic oscillations. Their model is justified by a qualitative, heuristic approach. As many other models, their model shows several qualitative features such as oscillations that make it quite plausible for real networks, but once again there is no statistical proof of it, up to our knowledge.

In this context, the main purpose of the present article is to build a bridge between several point processes models that have been proved to statistically fit real spike trains data and age structured PDE of the type of Pakdaman, Perthame and Salort. The point processes are the microscopic models, the PDE being their meso-macroscopic counterpart. In this sense, it extends PDE approaches for IF models to models that statistically fit true spike trains data. In the first section, we introduce Pakdaman, Perthame and Salort PDE (PPS) via its heuristic informal and microscopic description, which is based on IF models. Then, in Section 2, we develop the different point process models, quite informally, to draw the main heuristic correspondences between both approaches. In particular, we introduce the conditional intensity of a point process and a fundamental construction, called Ogata's thinning [29], which allows a microscopic understanding of the dynamics of a point process. Thanks to Ogata's thinning, in Section 3, we have been able to rigorously derive a microscopic random weak version of (PPS) and to propose its expectation deterministic counterpart. An independent and identically distributed (i.i.d) population version is also available. Several examples of applications are discussed in Section 4. To facilitate reading, technical results and proofs are included in two appendices. The present work is clearly just a first to link point processes and PDE: there are much more open questions than answered ones and this is discussed in the final conclusion. However, we think that this can be fundamental to acquire a deeper understanding of spike train models, their advantages as well as their limitations.

1. Synaptic integration and (PPS) equation

Based on the intuition that every neuron in the network should behave in the same way, Pakdaman, Perthame and Salort proposed in [31] a deterministic PDE denoted (PPS) in the sequel. The origin of this PDE is the classical (IF) model. In this section we describe the link between the (IF) microscopic model and the mesoscopic (PPS)

model, the main aim being to show thereafter the relation between (PPS) model and other natural microscopic models for spike trains: point processes.

1.1. *Integrate-and-fire*

Integrate-and-fire models describe the time evolution of the membrane potential, $V(t)$, by means of ordinary differential equations as follows

$$C_m \frac{\mathrm{d}tV}{\mathrm{d}t} = -g_L(V - V_L) + I(t), \tag{1.1}$$

where C_m is the capacitance of the membrane, g_L is the leak conductance and V_L is the leak reversal potential. If $V(t)$ exceeds a certain threshold θ, the neuron fires / emits an action potential (spike) and $V(t)$ is reset to V_L. The *synaptic current* $I(t)$ takes into account the fact that other presynaptic neurons fire and excite the neuron of interest, whose potential is given by $V(t)$.

As stated in [31], the origin of (PPS) equation comes from [35], where the explicit solution of a classical IF model as (1.1) has been discussed. To be more precise the membrane voltage of one neuron at time t is described by:

$$V(t) = V_r + (V_L - V_r)e^{-(t-T)/\tau_m} + \int_T^t h(t-u)N_{input}(du), \tag{1.2}$$

where V_r is the resting potential satisfying $V_L < V_r < \theta$, T is the last spike emitted by the considered neuron, τ_m is the time constant of the system (normally $\tau_m = g_L/C_m$), h is the excitatory post synaptic potential (EPSP) and N_{input} is the sum of Dirac masses at each spike of the presynaptic neurons. Since after firing, $V(t)$ is reset to $V_L < V_r$, there is a refractory period when the neuron is less excitable than at rest. The constant time τ_m indicates whether the next spike can occur more or less rapidly. The other main quantity, $\int_T^t h(t-u)N_{input}(du)$, is the *synaptic integration term.*

In [35], they consider a whole random network of such IF neurons and look at the behavior of this model, where the only randomness is in the network. In many other studies [7,8,9,11,26,42,30] IF models as (1.1) are considered to finally obtain other systems of partial differential equations (different to (PPS)) describing neural networks behavior. In these studies, each presynaptic neuron is assumed to fire as an independent Poisson process and via a diffusion approximation, the synaptic current is then approximated by a continuous in time stochastic process of Ornstein-Uhlenbeck.

1.2. *The (PPS) equation*

The deterministic PDE proposed by Pakdaman, Perthame and Salort, whose origin is also the microscopic IF model (1.2), is the following:

$$\text{(PPS)} \qquad \begin{cases} \frac{\partial n(s,t)}{\partial t} + \frac{\partial n(s,t)}{\partial s} + p\left(s, X\left(t\right)\right) n\left(s,t\right) = 0 \\ m\left(t\right) := n\left(0,t\right) = \int_0^{+\infty} p\left(s, X\left(t\right)\right) n\left(s,t\right) \mathrm{d}s. \end{cases}$$

In this equation, $n(s,t)$ represents a probability density of neurons at time t having discharged at time $t-s$. Therefore, s represents the time elapsed since the last

discharge. The fact that the equation is an elapsed time structured equation is natural, because the IF model (1.2) clearly only depends on the time since the last spike. More informally, the variable s represents the "age" of the neuron.

The first equation of the system (PPS) represents a pure transport process and means that as time goes by, neurons of age s and past given by $X(t)$ are either aging linearly or reset to age 0 with rate $p\left(s, X\left(t\right)\right)$.

The second equation of (PPS) describes the fact that when neurons spike, the age (the elapsed time) returns to 0. Therefore, $n(0,t)$ depicts the density of neurons undergoing a discharge at time t and it is denoted by $m(t)$. As a consequence of this boundary condition, for n at $s=0$, the following conservation law is obtained:

$$\int_0^\infty n\left(s,t\right)\mathrm{d}s = \int_0^\infty n\left(s,0\right)\mathrm{d}s$$

This means that if $n\left(\cdot,0\right)$ is a probabilistic density then $n\left(\cdot,t\right)$ can be interpreted as a density at each time t. Denoting by $\mathrm{d}t$ the Lebesgue measure and since $m(t)$ is the density of firing neurons at time t in (PPS), $m(t)\mathrm{d}t$ can also be interpreted as the limit of $N_{input}(\mathrm{d}t)$ in (1.2) when the population of neurons becomes continuous.

The system (PPS) is nonlinear since the rate $p\left(s, X(t)\right)$ depends on $n(0,t)$ by means of the quantity $X(t)$:

$$X(t) = \int_0^t h(u)m(t-u)\mathrm{d}u = \int_0^t h(u)n(0,t-u)\mathrm{d}u. \tag{1.3}$$

The quantity $X(t)$ represents the interactions between neurons. It "takes into account the averaged propagation time for the ionic pulse in this network" [31]. More precisely with respect to the IF models (1.2), this is the synaptic integration term, once the population becomes continuous. The only difference is that in (1.2) the memory is cancelled once the last spike has occurred and this is not the case here. However informally, both quantities have the same interpretation. Note nevertheless, that in [31], the function h can be much more general than the h of the IF models which clearly corresponds to EPSP. From now on and in the rest of the paper, h is just a general non negative function without forcing the connection with EPSP.

The larger $p\left(s, X(t)\right)$ the more likely neurons of age s and past $X(t)$ fire. Most of the time (but it is not a requisite), p is assumed to be less than 1 and is interpreted as the probability that neurons of age s fire. However, as shown in Section 3 and as interpreted in many population structured equation [14,19,34], $p(s, X(t))$ is closer to a hazard rate, i.e. a positive quantity such that $p\left(s, X(t)\right)\mathrm{d}t$ is informally the probability to fire given that the neuron has not fired yet. In particular, it could be not bounded by 1 and does not need to integrate to 1. A toy example is obtained if $p\left(s, X(t)\right) = \lambda > 0$, where a steady state solution is $n(s,t) = \lambda e^{-\lambda s}\mathbf{1}_{s\geq 0}$: this is the density of an exponential variable with parameter λ.

However, based on the interpretation of $p\left(s, X(t)\right)$ as a probability bounded by 1, one of the main model that Pakdaman, Perthame and Salort consider is $p\left(s, X(t)\right) = \mathbf{1}_{s\geq\sigma(X(t))}$. This again can be easily interpreted by looking at (1.2).

Indeed, since in the IF models the spike happens when the threshold θ is reached, one can consider that $p(s, X(t))$ should be equal to 1 whenever

$$V(t) = V_r + (V_L - V_r)e^{-(t-T)/\tau_m} + X(t) \geq \theta,$$

and 0 otherwise. Since $V_L - V_r < 0$, $p(s, X(t)) = 1$ is indeed equivalent to $s = t - T$ larger than some decreasing function of $X(t)$. This has the double advantage to give a formula for the refractory period ($\sigma(X(t))$) and to model excitatory systems: the refractory period decreases when the whole firing rate increases via $X(t)$ and this makes the neurons fire even more. This is for this particular case that Pakdaman, Perthame and Salort have shown existence of oscillatory behavior [32].

Another important parameter in the (PPS) model and introduced in [31] is J, which can be seen with our formalism as $\int h$ and which describes the network connectivity or the strength of the interaction. In [31] it has been proved that, for highly or weakly connected networks, (PPS) model exhibits relaxation to steady state and periodic solutions have also been numerically observed for moderately connected networks. The authors in [32] have quantified the regime where relaxation to a stationary solution occurs in terms of J and described periodic solution for intermediate values of J.

Recently, in [33], the (PPS) model has been extended including a fragmentation term, which describes the adaptation and fatigue of the neurons. In this sense, this new term incorporates the past activity of the neurons. For this new model, in the linear case there is exponential convergence to the steady states, while in the weakly nonlinear case a total desynchronization in the network is proved. Moreover, for greater nonlinearities, synchronization can again been numerically observed.

2. Point processes and conditional intensities as models for spike trains

We first start by quickly reviewing the main basic concepts and notations of point processes, in particular, conditional intensities and Ogata's thinning [29]. We refer the interested reader to [3] for exhaustiveness and to [6] for a much more condensed version, with the main useful notions.

2.1. *Counting processes and conditional intensities*

We focus on locally finite point processes on $\mathbb{R}$, equipped with the borelians $\mathcal{B}(\mathbb{R})$.

Definition 2.1 (Locally finite point process). A locally finite point process N on $\mathbb{R}$ is a random set of points such that it has almost surely (a.s.) a finite number of points in finite intervals. Therefore, associated to N there is an ordered sequence of extended real valued random times $(T_z)_{z\in\mathbb{Z}}$: $\cdots \leq T_{-1} \leq T_0 \leq 0 < T_1 \leq \cdots$.

For a measurable set A, N_A denotes the number of points of N in A. This is a random variable with values in $\mathbb{N} \cup \{\infty\}$.

Definition 2.2 (Counting process associated to a point process). The process on $\mathbb{R}_+$ defined by $t \mapsto N_t := N_{(0,t]}$ is called the counting process associated to the point process N.

The natural and the predictable filtrations are fundamental for the present work.

Definition 2.3 (Natural filtration of a point process). The natural filtration of N is the family $\left(\mathcal{F}_t^N\right)_{t\in\mathbb{R}}$ of σ-algebras defined by $\mathcal{F}_t^N = \sigma\left(N \cap (-\infty, t]\right)$.

Definition 2.4 (Predictable filtration of a point process). The predictable filtration of N is the family of σ-algebra $\left(\mathcal{F}_{t-}^N\right)_{t\in\mathbb{R}}$ defined by $\mathcal{F}_{t-}^N = \sigma\left(N \cap (-\infty, t)\right)$.

The intuition behind this concept is that $\mathcal{F}_t^N$ contains all the information given by the point process at time t. In particular, it contains the information whether t is a point of the process or not while $\mathcal{F}_{t-}^N$ only contains the information given by the point process strictly before t. Therefore, it does not contain (in general) the information whether t is a point or not. In this sense, $\mathcal{F}_{t-}^N$ represents (the information contained in) the past.

Under some rather classical conditions [3], which are always assumed to be satisfied here, one can associate to $(N_t)_{t\geq 0}$ a stochastic intensity $\lambda(t, \mathcal{F}_{t-}^N)$, which is a non negative random quantity. The notation $\lambda(t, \mathcal{F}_{t-}^N)$ for the intensity refers to the predictable version of the intensity associated to the natural filtration and $(N_t - \int_0^t \lambda(u, \mathcal{F}_{u-}^N)\mathrm{d}u)_{t\geq 0}$ forms a local martingale [3]. Informally, $\lambda(t, \mathcal{F}_{t-}^N)\mathrm{d}t$ represents the probability to have a new point in interval $[t, t + \mathrm{d}t)$ given the past. Note that $\lambda(t, \mathcal{F}_{t-}^N)$ should not be understood as a function, in the same way as density is for random variables. It is a "recipe" explaining how the probability to find a new point at time t depends on the past configuration: since the past configuration depends on its own past, this is closer to a recursive formula. In this respect, the intensity should obviously depend on $N\cap(-\infty, t)$ and not on $N\cap(-\infty, t]$ to predict the occurrence at time t, since we cannot know whether t is already a point or not.

The distribution of the point process N on $\mathbb{R}$ is completely characterized by the knowledge of the intensity $\lambda(t, \mathcal{F}_{t-}^N)$ on $\mathbb{R}_+$ and the distribution of $N_- = N \cap \mathbb{R}_-$, which is denoted by $\mathbb{P}_0$ in the sequel. The information about $\mathbb{P}_0$ is necessary since each point of N may depend on the occurrence of all the previous points: if for all $t > 0$, one knows the "recipe" $\lambda(t, \mathcal{F}_{t-}^N)$ that gives the probability of a new point at time t given the past configuration, one still needs to know the distribution of N_- to obtain the whole process.

Two main assumptions are used depending on the type of results we seek:

$\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,a.s.}\right)$ | for any $T \geq 0$, $\int_0^T \lambda(t, \mathcal{F}_{t-}^N)\mathrm{d}t$ is finite a.s.

$\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp}\right)$ | for any $T \geq 0$, $\mathbb{E}\left[\int_0^T \lambda(t, \mathcal{F}_{t-}^N)\mathrm{d}t\right]$ is finite.

Clearly $\left(\mathcal{A}_{loc}^{\mathbb{L}^1,exp}\right)$ implies $\left(\mathcal{A}_{loc}^{\mathbb{L}^1,a.s.}\right)$. Note that $\left(\mathcal{A}_{loc}^{\mathbb{L}^1,a.s.}\right)$ implies non-explosion in finite time for the counting processes (N_t).

Definition 2.5 (Point measure associated to a point process). The point measure associated to N is denoted by $N(\mathrm{d}t)$ and defined by $N(\mathrm{d}t) = \sum_{i\in\mathbb{Z}} \delta_{T_i}(\mathrm{d}t)$, where δ_u is the Dirac mass in u.

By analogy with (PPS), and since points of point processes correspond to spikes

(or times of discharge) for the considered neuron in spike train analysis, $N(\mathrm{d}t)$ is the microscopic equivalent of the distribution of discharging neurons $m(t)\mathrm{d}t$. Following this analogy, and since T_{N_t} is the last point less or equal to t for every $t \geq 0$, the age S_t at time t is defined by $S_t = t - T_{N_t}$. In particular, if t is a point of N, then $S_t = 0$. Note that S_t is $\mathcal{F}_t^N$ measurable for every $t \geq 0$ and therefore, $S_0 = -T_0$ is $\mathcal{F}_0^N$ measurable. To define an age at time $t = 0$, one assumes that

$(\mathcal{A}_{T_0})$ | There exists a first point before 0 for the process N_-, i.e. $-\infty < T_0$.

As we have remarked before, conditional intensity should depend on $N \cap (-\infty, t)$. Therefore, it cannot be function of S_t, since S_t informs us if t is a point or not. That is the main reason for considering this $\mathcal{F}_{t-}^N$ measurable variable

$$S_{t-} = t - T_{N_{t-}}, \tag{2.1}$$

where $T_{N_{t-}}$ is the last point strictly before t (see Figure 1). Note also that knowing $(S_{t-})_{t\geq 0}$ or $(N_t)_{t\geq 0}$ is completely equivalent given $\mathcal{F}_0^N$.

The last and most crucial equivalence between (PPS) and the present point process set-up, consists in noting that the quantities $p(s, X(t))$ and $\lambda(t, \mathcal{F}_{t-}^N)$ have informally the same meaning: they both represent a firing rate, i.e. both give the rate of discharge as a function of the past. This dependence is made more explicit in $p(s, X(t))$ than in $\lambda(t, \mathcal{F}_{t-}^N)$.

2.2. *Examples*

Let us review the basic point processes models of spike trains and see what kind of analogy is likely to exist between both models ((PPS) equation and point processes). These informal analogies are possibly exact mathematical results (see Section 4).

Homogeneous Poisson process This is the simplest case where $\lambda(t, \mathcal{F}_{t-}^N) = \lambda$, with λ a fixed positive constant representing the firing rate. There is no dependence in time t (it is homogeneous) and no dependence with respect to the past. This case should be equivalent to $p(s, X(t)) = \lambda$ in (PPS). This can be made even more explicit. Indeed in the case where the Poisson process exists on the whole real line (stationary case), it is easy to see that

$$\mathbb{P}\left(S_{t-} > s\right) = \mathbb{P}\left(N_{[t-s,t)} = 0\right) = \exp(-\lambda s),$$

meaning that the age S_{t-} obeys an exponential distribution with parameter λ, i.e. the steady state of the toy example developed for (PPS) when $p(s, X(t)) = \lambda$.

Inhomogeneous Poisson process To model non stationarity, one can use $\lambda(t, \mathcal{F}_{t-}^N) = \lambda(t)$, which only depends on time. This case should be equivalent to the replacement of $p(s, X(t))$ in (PPS) by $\lambda(t)$.

Renewal process This model is very useful to take refractory period into account. It corresponds to the case where the ISIs (delays between spikes) are independent and identically distributed (i.i.d.) with a certain given density ν on $\mathbb{R}_+$. The asso-

ciated hazard rate is

$$f(s) = \frac{\nu(s)}{\int_s^{+\infty} \nu(x)\mathrm{d}x},$$

when $\int_s^{+\infty} \nu(x)\mathrm{d}x > 0$. Roughly speaking, $f(s)\mathrm{d}s$ is the probability that a neuron spikes with age s given that its age is larger than s. In this case, considering the set of spikes as the point process N, it is easy to show (see the Appendix B.1) that its corresponding intensity is $\lambda(t, \mathcal{F}_{t-}^N) = f(S_{t-})$ which only depends on the age. One can also show quite easily that the process $(S_{t-})_{t>0}$, which is equal to $(S_t)_{t>0}$ almost everywhere (a.e.), is a Markovian process in time. This renewal setting should be equivalent in the (PPS) framework to $p(s, X(t)) = f(s)$.

Note that many people consider IF models (1.2) with Poissonian inputs with or without additive white noise. In both cases, the system erases all memory after each spike and therefore the ISIs are i.i.d. Therefore as long as we are only interested by the spike trains and their point process models, those IF models are just a particular case of renewal process [8,10,17,35].

Wold process and more general structures Let A_t^1 be the delay (ISI) between the last point and the occurrence just before (see also Figure 1), $A_t^1 = T_{N_{t-}} - T_{N_{t-}-1}$. A Wold process [24,16] is then characterized by $\lambda(t, \mathcal{F}_{t-}^N) = f(S_{t-}, A_t^1)$. This model has been matched to several real data thanks to goodness-of-fit tests [38] and is therefore one of our main example with the next discussed Hawkes process case. One can show in this case that the successive ISI's form a Markov chain of order 1 and that the continuous time process (S_{t-}, A_t^1) is also Markovian.

This case should be equivalent to the replacement of $p(s, X(t))$ in (PPS) by $f(s, a^1)$, with a^1 denoting the delay between the two previous spikes. Naturally in this case, one should expect a PDE of higher dimension with third variable a^1.

More generally, one could define

$$A_t^k = T_{N_{t-}-(k-1)} - T_{N_{t-}-k}, \tag{2.2}$$

and point processes with intensity $\lambda(t, \mathcal{F}_{t-}^N) = f(S_{t-}, A_t^1, ..., A_t^k)$. Those processes satisfy more generally that their ISI's form a Markov chain of order k and that the continuous time process $(S_{t-}, A_t^1, ..., A_t^k)$ is also Markovian (see the Appendix B.2).

Remark 2.1. The dynamics of the successive ages is pretty simple. On the one hand, the dynamics of the vector of the successive ages $(S_{t-}, A_t^1, ..., A_t^k)_{t>0}$ is deterministic between two jumping times. The first coordinate increases with rate 1. On the other hand, the dynamics at any jumping time T is given by the following shift:

$$\begin{cases} \text{the age process goes to 0, i.e. } S_T = 0, \\ \text{the first delay becomes the age, i.e. } A_{T+}^1 = S_{T-}, \\ \text{the other delays are shifted, i.e. } A_{T+}^i = A_T^{i-1} \text{ for all } i \leq k. \end{cases} \tag{2.3}$$

Hawkes processes The most classical setting is the linear (univariate) Hawkes process, which corresponds to

$$\lambda(t, \mathcal{F}_{t-}^N) = \mu + \int_{-\infty}^{t-} h(t-u)N(du),$$

where the positive parameter μ is called the spontaneous rate and the non negative function h, with support in $\mathbb{R}_+$, is called the interaction function, which is generally assumed to satisfy $\int_{\mathbb{R}_+} h < 1$ to guarantee the existence of a stationary version [16]. This model has also been matched to several real neuronal data thanks to goodness-of-fit tests [40]. Since it can mimic synaptic integration, as explained below, this represents the main example of the present work.

In the case where T_0 tends to $-\infty$, this is equivalent to say that there is no point on the negative half-line and in this case, one can rewrite

$$\lambda(t, \mathcal{F}_{t-}^N) = \mu + \int_{0}^{t-} h(t-u)N(du).$$

By analogy between $N(\mathrm{d}t)$ and $m(t)\mathrm{d}t$, one sees that $\int_0^{t-} h(t-u)N(du)$ is indeed the analogous of $X(t)$ the synaptic integration in (1.3). So one could expect that the PDE analogue is given by $p(s, X(t)) = \mu + X(t)$. In Section 4, we show that this does not hold stricto sensu, whereas the other analogues work well.

Note that this model shares also some link with IF models. Indeed, the formula for the intensity is close to the formula for the voltage (1.2), with the same flavor for the synaptic integration term. The main difference comes from the fact that when the voltage reaches a certain threshold, it fires deterministically for the IF model, whereas the higher the intensity, the more likely is the spike for the Hawkes model, but without certainty. In this sense Hawkes models seem closer to (PPS) since as we discussed before, the term $p(s, X(t))$ is closer to a hazard rate and never imposes deterministically the presence of a spike.

To model inhibition (see [41] for instance), one can use functions h that may take negative values and in this case $\lambda(t, \mathcal{F}_{t-}^N) = \left(\mu + \int_{-\infty}^{t-} h(t-u)N(du)\right)_+$, which should correspond to $p(s, X(t)) = (\mu + X(t))_+$. Another possibility is $\lambda(t, \mathcal{F}_{t-}^N) = \exp\left(\mu + \int_{-\infty}^{t-} h(t-u)N(du)\right)$, which is inspired by the generalized linear model as used by [36] and which should correspond to $p(s, X(t)) = \exp(\mu + X(t))$.

Note finally that Hawkes models in Neuroscience (and their variant) are usually multivariate meaning that they model interaction between spike trains thanks to interaction functions between point processes, each process representing a neuron. To keep the present analogy as simple as possible, we do not deal with those multivariate models in the present article. Some open questions in this direction are presented in conclusion.

2.3. *Ogata's thinning algorithm*

To turn the analogy between $p(s, X(t))$ and $\lambda(t, \mathcal{F}_{t-}^N)$ into a rigorous result on the PDE level, we need to understand the intrinsic dynamics of the point process. This dynamics is often not explicitly described in the literature (see e.g. the reference book by Brémaud [3]) because martingale theory provides a nice mathematical setting in which one can perform all the computations. However, when one wants to simulate point processes based on the knowledge of their intensity, there is indeed a dynamics that is required to obtain a practical algorithm. This method has been described at first by Lewis in the Poisson setting [25] and generalized by Ogata in [29]. If there is a sketch of proof in [29], we have been unable to find any complete mathematical proof of this construction in the literature and we propose a full and mathematically complete version of this proof with minimal assumptions in the Appendix B.4. Let us just informally describe here, how this construction works.

The principle consists in assuming that is given an external homogeneous Poisson process Π of intensity 1 in $\mathbb{R}_+^2$ and with associated point measure $\Pi(\mathrm{d}t, \mathrm{d}x) = \sum_{(T,V)\in\Pi} \delta_{(T,V)}(\mathrm{d}t, \mathrm{d}x)$. This means in particular that

$$\mathbb{E}\left[\Pi(\mathrm{d}t, \mathrm{d}x)\right] = \mathrm{d}t\ \mathrm{d}x. \tag{2.4}$$

Once a realisation of N_- fixed, which implies that $\mathcal{F}_0^N$ is known and which can be seen as an initial condition for the dynamics, the construction of the process N on $\mathbb{R}_+$ only depends on Π.

More precisely, if we know the intensity $\lambda(t, \mathcal{F}_{t-}^N)$ in the sense of the "recipe" that explicitly depends on t and $N \cap (-\infty, t)$, then once a realisation of Π and of N_- is fixed, the dynamics to build a point process N with intensity $\lambda(t, \mathcal{F}_{t-}^N)$ for $t \in \mathbb{R}_+$ is purely deterministic. It consists (see also Figure 1) in successively projecting on the abscissa axis the points that are below the graph of $\lambda(t, \mathcal{F}_{t-}^N)$. Note that a point projection may change the shape of $\lambda(t, \mathcal{F}_{t-}^N)$, just after the projection. Therefore the graph of $\lambda(t, \mathcal{F}_{t-}^N)$ evolves thanks to the realization of Π. For a more mathematical description, see Theorem B.11 in the Appendix B.4. Note in particular that the construction ends on any finite interval $[0, T]$ a.s. if $\left(\mathcal{A}_{\lambda,loc}^{1,a.s}\right)$ holds.

Then the point process N, result of Ogata's thinning, is given by the union of N_- on $\mathbb{R}_-$ and the projected points on $\mathbb{R}_+$. It admits the desired intensity $\lambda(t, \mathcal{F}_{t-}^N)$ on $\mathbb{R}_+$. Moreover, the point measure can be represented by

$$\mathbf{1}_{t>0}\ N(\mathrm{d}t) = \sum_{\substack{(T,X)\in\Pi\ /\\ X\le\lambda\left(T,\mathcal{F}_{T-}^N\right)}} \delta_T(\mathrm{d}t) = \left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)} \Pi(\mathrm{d}t, \mathrm{d}x)\right). \tag{2.5}$$

NB: The last equality comes from the following convention. If $\delta_{(c,d)}$ is a Dirac mass in $(c, d) \in \mathbb{R}_+^2$, then $\int_{x=a}^b \delta_{(c,d)}(\mathrm{d}t, \mathrm{d}x)$, as a distribution in t, is $\delta_c(\mathrm{d}t)$ if $d \in [a, b]$ and 0 otherwise.

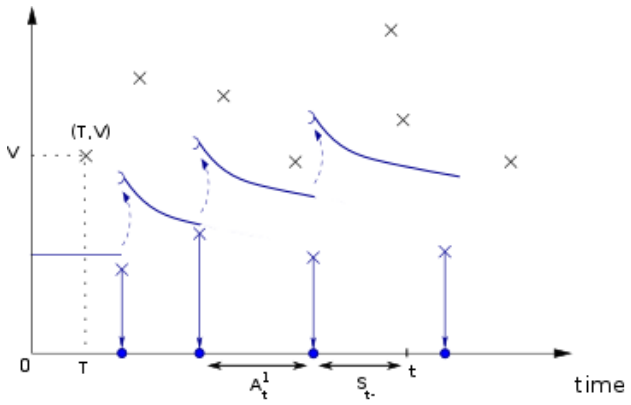


Fig. 1. Example of Ogata's thinning algorithm on a linear Hawkes process with interaction function $h(u) = e^{-u}$ and no point before 0 (i.e. $N_- = \emptyset$). The crosses represent a realization of Π, Poisson process of intensity 1 on $\mathbb{R}_+^2$. The blue piecewise continuous line represents the intensity $\lambda(t, \mathcal{F}_{t-}^N)$, which starts in 0 with value μ and then jumps each time a point of Π is present underneath it. The resulting Hawkes process (with intensity $\lambda(t, \mathcal{F}_{t-}^N)$) is given by the blue circles. Age S_{t-} at time t and the quantity A_t^1 are also represented.

3. From point processes to PDE

Let us now present our main results. Informally, we want to describe the evolution of the distribution in s of the age S_t according to the time t. Note that at fixed time t, $S_{t-} = S_t$ a.s. and therefore it is the same as the distribution of S_{t-}. We prefer to study S_{t-} since its predictability, i.e. its dependence in $N \cap (-\infty, t)$, makes all definitions proper from a microscopic/random point of view. Microscopically, the interest lies in the evolution of $\delta_{S_{t-}}(\mathrm{d}s)$ as a random measure. But it should also be seen as a distribution in time, for equations like (PPS) to make sense. Therefore, we need to go from a distribution only in s to a distribution in both s and t. Then one can either focus on the microscopic level, where the realisation of Π in Ogata's thinning construction is fixed or focus on the expectation of such a distribution.

3.1. *A clean setting for bivariate distributions in age and time*

In order to obtain, from a point process, (PPS) system we need to define bivariate distributions in s and t and marginals (at least in s), in such a way that weak solutions of (PPS) are correctly defined. Since we want to possibly consider more than two variables for generalized Wold processes, we consider the following definitions.

In the following, $< \varphi, \nu >$ denotes the integral of the integrable function φ with respect to the measure ν.

Let $k \in \mathbb{N}$. For every bounded measurable function φ of $(t, s, a_1, ..., a_k) \in \mathbb{R}_+^{k+2}$, one can define

$$\varphi_t^{(1)}(s, a_1, ..., a_k) = \varphi(t, s, a_1, ..., a_k) \quad \text{and} \quad \varphi_s^{(2)}(t, a_1, ..., a_k) = \varphi(t, s, a_1, ..., a_k).$$

Let us now define two sets of regularities for φ.

$\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$ | The function φ belongs to $\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$ if and only if
- φ is a measurable bounded function,
- there exists $T>0$ such that for all $t>T$, $\varphi_t^{(1)}=0$.

$C_{c,b}^{\infty}(\mathbb{R}_+^{k+2})$ | The function φ belongs to $C_{c,b}^{\infty}(\mathbb{R}_+^{k+2})$ if and only if
- φ is continuous, uniformly bounded,
- φ has uniformly bounded derivatives of every order,
- there exists $T>0$ such that for all $t>T$, $\varphi_t^{(1)}=0$.

Let $(\nu_1^t)_{t\geq 0}$ be a (measurable w.r.t. t) family of positive measures on $\mathbb{R}_+^{k+1}$, and $(\nu_2^s)_{s\geq 0}$ be a (measurable w.r.t. s) family of positive measures $\mathbb{R}_+^{k+1}$. Those families satisfy the Fubini property if

$(\mathcal{P}_{Fubini})$ | for any $\varphi\in\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$, $\quad\int\langle\varphi_t^{(1)},\nu_1^t\rangle\mathrm{d}t=\int\langle\varphi_s^{(2)},\nu_2^s\rangle\mathrm{d}s$.

In this case, one can define ν, measure on $\mathbb{R}_+^{k+2}$, by the unique measure on $\mathbb{R}_+^{k+2}$ such that for any test function φ in $\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$,

$$<\varphi,\nu>=\int\langle\varphi_t^{(1)},\nu_1^t\rangle\mathrm{d}t=\int\langle\varphi_s^{(2)},\nu_2^s\rangle\mathrm{d}s.$$

To simplify notations, for any such measure $\nu(t,\mathrm{d}s,\mathrm{d}a_1,...,\mathrm{d}a_k)$, we define

$$\nu(t,\mathrm{d}s,\mathrm{d}a_1,...,\mathrm{d}a_k)=\nu_1^t(\mathrm{d}s,\mathrm{d}a_1,...,\mathrm{d}a_k),\quad \nu(\mathrm{d}t,s,\mathrm{d}a_1,...,\mathrm{d}a_k)=\nu_2^s(\mathrm{d}t,\mathrm{d}a_1,...,\mathrm{d}a_k).$$

In the sequel, we need in particular a measure on $\mathbb{R}_+^2$, η_x, defined for any real x by its marginals that satisfy $(\mathcal{P}_{Fubini})$ as follows

$$\forall\, t,s\geq 0,\qquad \eta_x(t,\mathrm{d}s)=\delta_{t-x}(\mathrm{d}s)\mathbf{1}_{t-x>0}\quad\text{and}\quad\eta_x(\mathrm{d}t,s)=\delta_{s+x}(\mathrm{d}t)\mathbf{1}_{s\geq 0}. \tag{3.1}$$

It represents a Dirac mass "travelling" on the positive diagonal originated in $(x,0)$.

3.2. *The microscopic construction of a random PDE*

For a fixed realization of Π, we therefore want to define a random distribution $U(\mathrm{d}t,\mathrm{d}s)$ in terms of its marginals, thanks to $(\mathcal{P}_{Fubini})$, such that, $U(t,\mathrm{d}s)$ represents the distribution at time $t>0$ of the age S_{t-}, i.e.

$$\forall\, t>0,\quad U(t,\mathrm{d}s)=\delta_{S_{t-}}(\mathrm{d}s) \tag{3.2}$$

and satisfies similar equations as (PPS). This is done in the following proposition.

Proposition 3.1. *Let* Π, $\mathcal{F}_0^N$ *and an intensity* $\left(\lambda(t,\mathcal{F}_{t-}^N)\right)_{t>0}$ *be given as in Section 2.3 and satisfying* $(\mathcal{A}_{T_0})$ *and* $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,a.s.}\right)$. *On the event* Ω *of probability 1, where Ogata's thinning is well defined, let* N *be the point process on* $\mathbb{R}$ *that is constructed thanks to Ogata's thinning with associated predictable age process* $(S_{t-})_{t>0}$ *and whose points are denoted* $(T_i)_{i\in\mathbb{Z}}$. *Let the (random) measure* U *and its corresponding marginals be defined by*

$$U\left(\mathrm{d}t,\mathrm{d}s\right)=\sum_{i=0}^{+\infty}\eta_{T_i}(\mathrm{d}t,\mathrm{d}s)\ \mathbb{1}_{0\leq t\leq T_{i+1}}. \tag{3.3}$$

Then, on Ω, U satisfies $(\mathcal{P}_{Fubini})$ and $U(t,\mathrm{d}s) = \delta_{S_{t-}}(\mathrm{d}s)$. Moreover, on Ω, U is a solution in the weak sense of the following system

$$\frac{\partial}{\partial t}U(\mathrm{d}t,\mathrm{d}s) + \frac{\partial}{\partial s}U(\mathrm{d}t,\mathrm{d}s) + \left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)} \Pi(\mathrm{d}t,\mathrm{d}x)\right) U(t,\mathrm{d}s) = 0, \tag{3.4}$$

$$U(\mathrm{d}t,0) = \int_{s\in\mathbb{R}_+}\left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)} \Pi(\mathrm{d}t,\mathrm{d}x)\right) U(t,\mathrm{d}s) + \delta_0(\mathrm{d}t)\mathbb{1}_{T_0=0}, \tag{3.5}$$

$$U(0,\mathrm{d}s) = \delta_{-T_0}(\mathrm{d}s)\mathbb{1}_{T_0<0} = U^{in}(\mathrm{d}s)\mathbb{1}_{s>0}, \tag{3.6}$$

where $U^{in}(ds) = \delta_{-T_0}(ds)$. The weak sense means that for any $\varphi \in C_{c,b}^{\infty}(\mathbb{R}_+^2)$,

$$\int_{\mathbb{R}_+\times\mathbb{R}_+}\left(\frac{\partial}{\partial t}\varphi(t,s) + \frac{\partial}{\partial s}\varphi(t,s)\right)U(\mathrm{d}t,\mathrm{d}s) + \int_{\mathbb{R}_+\times\mathbb{R}_+}[\varphi(t,0)-\varphi(t,s)]\left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)}\Pi(\mathrm{d}t,\mathrm{d}x)\right)U(t,\mathrm{d}s) + \varphi(0,-T_0) = 0. \tag{3.7}$$

The proof of Proposition 3.1 is included in Appendix A.1. Note also that thanks to the Fubini property, the boundary condition (3.5) is satisfied also in a strong sense.

System (3.4)–(3.6) is a random microscopic version of (PPS) if $T_0 < 0$, where $n(s,t)$ the density of the age at time t is replaced by $U(t,\cdot) = \delta_{S_{t-}}$, the Dirac mass in the age at time t. The assumption $T_0 < 0$ is satisfied *a.s.* if T_0 has a density, but this may not be the case for instance if the experimental device gives an impulse at time zero (e.g. [38] studied Peristimulus time histograms (PSTH), where the spike trains are locked on a stimulus given at time 0).

This result may seem rather poor from a PDE point of view. However, since this equation is satisfied at a microscopic level, we are able to define correctly all the important quantities at a macroscopic level. Indeed, the analogy between $p(s,X(t))$ and $\lambda(t,\mathcal{F}_{t-}^N)$ is actually on the random microscopic scale a replacement of $p(s,X(t))$ by $\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)}\Pi(\mathrm{d}t,\mathrm{d}x)$, whose expectancy given the past is, heuristically speaking, equal to $\lambda\left(t,\mathcal{F}_{t-}^N\right)$ because the mean behaviour of Π is given by the Lebesgue measure (see (2.4)). Thus, the main question at this stage is : can we make this argument valid by taking the expectation of U? This is addressed in the next section.

The property $(\mathcal{P}_{Fubini})$ and the quantities η_{T_i} mainly allows to define $U(\mathrm{d}t,0)$ as well as $U(t,\mathrm{d}s)$. As expected, with this definition, (3.2) holds as well as

$$U(\mathrm{d}t,0) = \mathbb{1}_{t\geq 0}\, N(\mathrm{d}t), \tag{3.8}$$

i.e. the spiking measure (the measure in time with age 0) is the point measure.

Note also that the initial condition is given by $\mathcal{F}_0^N$, since $\mathcal{F}_0^N$ fixes in particular the value of T_0 and $(\mathcal{A}_{T_0})$ is required to give sense to the age at time 0. To understand the initial condition, remark that if $T_0 = 0$, then $U(0,\cdot) = 0 \neq \lim_{t\to 0^+} U(t,\cdot) = \delta_0$ by definitions of η_{T_i}but that if $T_0 < 0$, $U(0,\cdot) = \lim_{t\to 0^+} U(t,\cdot) = \delta_{-T_0}$.

The conservativeness (i.e. for all $t \geq 0$, $\int_0^\infty U(t, \mathrm{d}s) = 1$) is obtained by using (a sequence of test functions converging to) $\varphi = \mathbb{1}_{t \leq T}$.

Proposition 3.1 shows that the (random) measure U, defined by (3.3), in terms of a given point process N, is a weak solution of System (3.4)-(3.6). The study of the well-posedness of this system could be addressed following, for instance, the ideas given in [12]. In this case U should be the unique solution of system (3.4)–(3.6).

As last comment about Proposition 3.1, we analyse the particular case of the linear Hawkes process, in the following remark.

Remark 3.1. In the linear Hawkes process, $\lambda(t, \mathcal{F}_{t-}^N) = \mu + \int_{-\infty}^{t-} h(t-z)N(\mathrm{d}z)$. Thanks to (3.8) one decomposes the intensity into a term given by the initial condition plus a term given by the measure U, $\lambda(t, \mathcal{F}_{t-}^N) = \mu + F_0(t) + \int_0^{t-} h(t-z)U(\mathrm{d}z, 0)$, where $F_0(t) = \int_{-\infty}^0 h(t-z)N_-(\mathrm{d}z)$ is $(\mathcal{F}_0^N)$-measurable and considered as an initial condition. Hence, (3.4)–(3.6) becomes a closed system in the sense that $\lambda(t, \mathcal{F}_{t-}^N)$ is now an explicit function of the solution of the system. This is not true in general.

3.3. *The PDE satisfied in expectation*

In this section, we want to find the system satisfied by the expectation of the random measure U. First, we need to give a proper definition of such an object. The construction is based on the construction of U and is summarized in the following proposition. (The proofs of all the results of this subsection are in Appendix A.1).

Proposition 3.2. *Let* Π, $\mathcal{F}_0^N$ *and an intensity* $\left(\lambda(t, \mathcal{F}_{t-}^N)\right)_{t>0}$ *be given as in Section 2.3 and satisfying* $(\mathcal{A}_{T_0})$ *and* $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp}\right)$. *Let* N *be the process resulting of Ogata's thinning and let* U *be the random measure defined by* (3.3). *Let* $\mathbb{E}$ *denote the expectation with respect to* Π *and* $\mathcal{F}_0^N$.

Then for any test function φ *in* $\mathcal{M}_{c,b}(\mathbb{R}_+^2)$, $\mathbb{E}\left[\int \varphi(t,s)U(t,\mathrm{d}s)\right]$ *and* $\mathbb{E}\left[\int \varphi(t,s)U(\mathrm{d}t,s)\right]$ *are finite and one can define* $u(t,\mathrm{d}s)$ *and* $u(\mathrm{d}t,s)$ *by*

$$\begin{cases} \forall\, t \geq 0, \quad \displaystyle\int \varphi(t,s)u(t,\mathrm{d}s) = \mathbb{E}\left[\int \varphi(t,s)U(t,\mathrm{d}s)\right], \\ \forall\, s \geq 0, \quad \displaystyle\int \varphi(t,s)u(\mathrm{d}t,s) = \mathbb{E}\left[\int \varphi(t,s)U(\mathrm{d}t,s)\right]. \end{cases}$$

Moreover, $u(t,\mathrm{d}s)$ *and* $u(\mathrm{d}t,s)$ *satisfy* $(\mathcal{P}_{Fubini})$ *and one can define* $u(\mathrm{d}t,\mathrm{d}s) = u(t,\mathrm{d}s)\mathrm{d}t = u(\mathrm{d}t,s)\mathrm{d}s$ *on* $\mathbb{R}_+^2$, *such that for any test function* φ *in* $\mathcal{M}_{c,b}(\mathbb{R}_+^2)$,

$$\int \varphi(t,s)u(\mathrm{d}t,\mathrm{d}s) = \mathbb{E}\left[\int \varphi(t,s)U(\mathrm{d}t,\mathrm{d}s)\right],$$

quantity which is finite.

In particular, since $\int \varphi(t,s)u(t,\mathrm{d}s) = \mathbb{E}\left[\int \varphi(t,s)U(t,\mathrm{d}s)\right] = \mathbb{E}\left[\varphi(t,S_{t-})\right]$, $u(t,\cdot)$ is therefore the distribution of S_{t-}, the (predictable version of the) age at time t. Now let us show that as expected, u satisfies a system similar to (PPS).

Theorem 3.3. *Let* Π, $\mathcal{F}_0^N$ *and an intensity* $\left(\lambda(t, \mathcal{F}_{t-}^N)\right)_{t>0}$ *be given as in Section*

2.3 and satisfying $(\mathcal{A}_{T_0})$ *and* $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp}\right)$. *If* N *is the process resulting of Ogata's thinning,* $(S_{t-})_{t>0}$ *its associated predictable age process,* U *its associated random measure, defined by* (3.3)*, and* u *its associated mean measure, defined in Proposition 3.2, then, there exists a bivariate measurable function* $\rho_{\lambda,\mathbb{P}_0}$ *satisfying*

$$\begin{cases} \forall\, T \geq 0,\ \displaystyle\int_0^T \int_s \rho_{\lambda,\mathbb{P}_0}(t,s) u(\mathrm{d}t,\mathrm{d}s) < \infty, \\ \rho_{\lambda,\mathbb{P}_0}(t,s) = \mathbb{E}\left[\lambda\left(t,\mathcal{F}_{t-}^N\right) \middle| S_{t-} = s\right] \quad u(\mathrm{d}t,\mathrm{d}s)\text{- a.e} \end{cases} \tag{3.9}$$

and such that u *is solution in the weak sense of the following system*

$$\frac{\partial}{\partial t} u\left(\mathrm{d}t,\mathrm{d}s\right) + \frac{\partial}{\partial s} u\left(\mathrm{d}t,\mathrm{d}s\right) + \rho_{\lambda,\mathbb{P}_0}(t,s) u\left(\mathrm{d}t,\mathrm{d}s\right) = 0, \tag{3.10}$$

$$u\left(\mathrm{d}t,0\right) = \int_{s\in\mathbb{R}_+} \rho_{\lambda,\mathbb{P}_0}(t,s) u\left(t,\mathrm{d}s\right)\, \mathrm{d}t + \delta_0(\mathrm{d}t) u^{in}(\{0\}), \tag{3.11}$$

$$u\left(0,\mathrm{d}s\right) = u^{in}(\mathrm{d}s)\mathbb{1}_{s>0}, \tag{3.12}$$

where u^{in} *is the law of* $-T_0$. *The weak sense means here that for any* $\varphi \in C_{c,b}^{\infty}(\mathbb{R}_+^2)$,

$$\int_{\mathbb{R}_+\times\mathbb{R}_+} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \varphi\left(t,s\right) u\left(\mathrm{d}t,\mathrm{d}s\right) + \int_{\mathbb{R}_+\times\mathbb{R}_+} [\varphi(t,0) - \varphi(t,s)] \rho_{\lambda,\mathbb{P}_0}(t,s) u(\mathrm{d}t,\mathrm{d}s) + \int_{\mathbb{R}_+} \varphi(0,s) u^{in}(\mathrm{d}s) = 0, \tag{3.13}$$

Comparing this system to (PPS), one first sees that $n(\cdot,t)$, the density of the age at time t, is replaced by the mean measure $u(t,\cdot)$. If $u^{in} \in L^1(\mathbb{R}_+)$ we have $u^{in}(\{0\}) = 0$ so we get an equation which is exactly of renewal type, as (PPS). In the general case where u^{in} is only a probability measure, the difference with (PPS) lies in the term $\delta_0(\mathrm{d}t)u^{in}(\{0\})$ in the boundary condition for $s = 0$ and in the term $\mathbb{1}_{s>0}$ in the initial condition for $t = 0$. Both these extra terms are linked to the possibility for the initial measure u^{in} to charge zero. This possibility is not considered in [31] - else, a similar extra term would be needed in the setting of [31] as well. As said above in the comment of Proposition 3.1, we want to keep this term here since it models the case where there is a specific stimulus at time zero [38].

In general and without more assumptions on λ, it is not clear that u is not only a measure satisfying $(\mathcal{P}_{Fubini})$ but also absolutely continuous wrt to $\mathrm{d}t\ \mathrm{d}s$ and that the equations can be satisfied in a strong sense.

Concerning $p(s,X(t))$, which has always been thought of as the equivalent of $\lambda(t,\mathcal{F}_{t-}^N)$, it is not replaced by $\lambda(t,\mathcal{F}_{t-}^N)$, which would have no meaning in general since this is a random quantity, nor by $\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\right]$ which would have been a first possible guess; it is replaced by $\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)|S_{t-}=s\right]$. Indeed intuitively, since

$$\mathbb{E}\left[\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)} \Pi\left(\mathrm{d}t,\mathrm{d}x\right) \middle| \mathcal{F}_{t-}^N\right] = \lambda\left(t,\mathcal{F}_{t-}^N\right)\mathrm{d}t,$$

the corresponding weak term can be interpreted as, for any test function φ,

$$\mathbb{E}\left[\int \varphi(t,s)\left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^{N}\right)} \Pi(\mathrm{d}t,\mathrm{d}x)\right) U(t,\mathrm{d}s)\right] = \mathbb{E}\left[\int \varphi(t,s)\,\lambda\left(t,\mathcal{F}_{t-}^{N}\right)\delta_{S_{t-}}(\mathrm{d}s)\mathrm{d}t\right]$$
$$= \int_t \mathbb{E}\left[\varphi(t,S_{t-})\,\lambda\left(t,\mathcal{F}_{t-}^{N}\right)\right]\mathrm{d}t$$
$$= \int_t \mathbb{E}\left[\varphi(t,S_{t-})\,\mathbb{E}\left[\lambda\left(t,\mathcal{F}_{t-}^{N}\right)|S_{t-}\right]\right]\mathrm{d}t,$$

which is exactly $\int \varphi(t,s)\rho_{\lambda,\mathbb{P}_0}(t,s)u(\mathrm{d}t,\mathrm{d}s)$.

This conditional expectation makes dependencies particularly complex, but this also enables to derive equations even in non-Markovian setting (as Hawkes processes for instance, see Section 4). More explicitly, $\rho_{\lambda,\mathbb{P}_0}(t,s)$ is a function of the time t, of the age s, but it also depends on λ, the shape of the intensity of the underlying process and on the distribution of the initial condition N_-, that is $\mathbb{P}_0$. As explained in Section 2, it is both the knowledge of $\mathbb{P}_0$ and λ that characterizes the distribution of the process and in general the conditional expectation cannot be reduced to something depending on less than that. In Section 4, we discuss several examples of point processes where one can (or cannot) reduce the dependence.

Note that here again, we can prove that the equation is conservative by taking (a sequence of functions converging to) $\varphi = \mathbb{1}_{t\leq T}$ as a test function.

A direct corollary of Theorem 3.3 can be deduced thanks to the law of large numbers. This can be seen as the interpretation of (PPS) equation at a macroscopic level, when the population of neurons is i.i.d..

Corollary 3.4. *Let $\left(N^i\right)_{i=1}^{\infty}$ be some i.i.d. point processes with intensity given by $\lambda(t,\mathcal{F}_{t-}^{N^i})$ on $(0,+\infty)$ satisfying $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp}\right)$ and associated predictable age processes $(S_{t-}^i)_{t>0}$. Suppose furthermore that the distribution of N^1 on $(-\infty,0]$ is given by $\mathbb{P}_0$ which is such that $\mathbb{P}_0(N_-^1=\emptyset)=0$.*

Then there exists a measure u satisfying $(\mathcal{P}_{Fubini})$, weak solution of Equations (3.10) *and* (3.11)*, with $\rho_{\lambda,\mathbb{P}_0}$ defined by*

$$\rho_{\lambda,\mathbb{P}_0}(t,s) = \mathbb{E}\left[\lambda\left(t,\mathcal{F}_{t-}^{N^1}\right)|S_{t-}^1=s\right], \quad u(\mathrm{d}t,\mathrm{d}s)-a.e.$$

and with u^{in} distribution of the age at time 0, such that for any $\varphi\in C_{c,b}^{\infty}(\mathbb{R}_+^2)$

$$\forall\, t>0, \quad \int \varphi(t,s)\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{S_t^i}(\mathrm{d}s)\right) \xrightarrow[n\to\infty]{a.s.} \int \varphi(t,s)u(t,\mathrm{d}s), \tag{3.14}$$

In particular, informally, the fraction of neurons at time t with age in $[s,s+\mathrm{d}s)$ in this i.i.d. population of neurons indeed tends to $u(t,\mathrm{d}s)$.

4. Application to the various examples

Let us now apply these results to the examples presented in Section 2.2.

4.1. *When the intensity only depends on time and age*

If $\lambda\left(t, \mathcal{F}_{t-}^N\right) = f\left(t, S_{t-}\right)$ (homogeneous and inhomogeneous Poisson processes and renewal processes are particular examples) then the intuition giving that $p(s, X(t))$ is analogous to $\lambda\left(t, \mathcal{F}_{t-}^N\right)$ works. Let us assume that $f(t,s) \in L^\infty(\mathbb{R}_+^2)$. We have $\mathbb{E}\left[\lambda\left(t, \mathcal{F}_{t-}^N\right) | S_{t-} = s\right] = f(t,s)$. Under this assumption, we may apply Theorem 3.3, so that we know that the mean measure u associated to the random process is a solution of System (3.10)–(3.12). Therefore the mean measure u satisfies a completely explicit PDE of the type (PPS) with $\rho_{\lambda,\mathbb{P}_0}(t,s) = f(t,s)$ replacing $p(s, X(t))$. In particular, in this case $\rho_{\lambda,\mathbb{P}_0}(t,s)$ does not depend on the initial condition. As already underlined, in general, the distribution of the process is characterized by $\lambda\left(t, \mathcal{F}_{t-}^N\right) = f\left(t, S_{t-}\right)$ and by the distribution of N_-. Therefore, in this special case, this dependence is actually reduced to the function f and the distribution of $-T_0$. Since $f(\cdot,\cdot) \in L^\infty\big([0,T] \times \mathbb{R}_+\big)$, assuming also $u^{in} \in L^1(\mathbb{R}_+)$, it is well-known that there exists a unique solution u such that $(t \mapsto u(t,\cdot)) \in \mathcal{C}\big([0,T], L^1(\mathbb{R}_+)\big)$, see for instance [34] Section 3.3. p.60. Note that following [12] uniqueness for measure solutions may also be established, hence the mean measure u associated to the random process is the unique solution of System (3.10)–(3.12), and it is in $\mathcal{C}\big([0,T], L^1(\mathbb{R}_+)\big)$: the PDE formulation, together with existence and uniqueness, has provided a regularity result on u which is obtained under weaker assumptions than through Fokker-Planck / Kolmogorov equations. This is another possible application field of our results: using the PDE formulation to gain regularity. Let us now develop the Fokker-Planck / Kolmogorov approach for renewal processes.

Renewal processes The renewal process, i.e. when $\lambda\left(t, \mathcal{F}_{t-}^N\right) = f\left(S_{t-}\right)$, with f a continuous function on $\mathbb{R}_+$, has particular properties. As noted in Section 2.2, the renewal age process $(S_{t-})_{t>0}$ is an homogeneous Markovian process. It is known for a long time that it is easy to derive PDE on the corresponding density through Fokker-Planck / Kolmogorov equations, once the variable of interest (here the age) is Markovian (see for instance [1]). Here we briefly follow this line to see what kind of PDE can be derived through the Markovian properties and to compare the equation with the (PPS) type system derived in Theorem 3.3.

Since f is continuous, the infinitesimal generator[b] of $(S_t)_{t>0}$ is given by

$$(\mathcal{G}\phi)(x) = \phi'(x) + f(x)\left(\phi(0) - \phi(x)\right), \tag{4.1}$$

for all $\phi \in C^1(\mathbb{R}_+)$ (see [2]). Note that, since for every $t > 0$ $S_{t-} = S_t$ a.s., the process $(S_{t-})_{t>0}$ is also Markovian with the same infinitesimal generator.

[b]The infinitesimal generator of an homogeneous Markov process $(Z_t)_{t\geq 0}$ is the operator $\mathcal{G}$ which is defined to act on every function $\phi : \mathbb{R}^n \to \mathbb{R}$ in a suitable space $\mathcal{D}$ by

$$\mathcal{G}\phi(x) = \lim_{t\to 0^+} \frac{\mathbb{E}\left[\phi(Z_t) | Z_0 = x\right] - \phi(x)}{t}.$$

Let us now define for all $t > 0$ and all $\phi \in C^1(\mathbb{R}_+)$,

$$P_t\phi(x) = \mathbb{E}\left[\phi(S_{t-})|S_0 = x\right] = \int \phi(s)u_x(t, \mathrm{d}s),$$

where $x \in \mathbb{R}_+$ and $u_x(t, \cdot)$ is the distribution of S_{t-} given that $S_0 = x$. Note that $u_x(t, \mathrm{d}s)$ corresponds to the marginal in the sense of $(\mathcal{P}_{Fubini})$ of u_x given by Theorem 3.3 with $\rho_{\lambda,\mathbb{P}_0}(t, s) = f(s)$ and initial condition δ_x, i.e. $T_0 = -x$ a.s.

In this homogeneous Markovian case, the forward Kolmogorov equation gives

$$\frac{\partial}{\partial t}P_t = P_t\mathcal{G}.$$

Let $\varphi \in C^\infty_{c,b}(\mathbb{R}^2_+)$ and let $t > 0$. This implies that

$$\begin{aligned}\frac{\partial}{\partial t}\left(P_t\varphi(t, s)\right) &= P_t\mathcal{G}\varphi(t, s) + P_t\frac{\partial}{\partial t}\varphi(t, s)\\ &= P_t\left[\frac{\partial}{\partial s}\varphi(t, s) + f(s)\left(\varphi(t, 0) - \varphi(t, s)\right) + \frac{\partial}{\partial t}\varphi(t, s)\right].\end{aligned}$$

Since φ is compactly supported in time, an integration with respect to t yields

$$-P_0\varphi(0, s) = \int P_t\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\varphi(t, s)\mathrm{d}t + \int P_t f(s)\left(\varphi(t, 0) - \varphi(t, s)\right)\mathrm{d}t,$$

or equivalently

$$-\varphi(0, x) = \int\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\varphi\left(t, s\right)u_x\left(t, \mathrm{d}s\right)\mathrm{d}t - \int(\varphi(t, s) - \varphi(t, 0))f(s)u_x(t, \mathrm{d}s)\mathrm{d}t, \tag{4.2}$$

in terms of u_x. This is exactly Equation (3.13) with $u^{in} = \delta_x$.

The result of Theorem 3.3 is stronger than the application of the forward Kolmogorov equation on homogeneous Markovian systems since the result of Theorem 3.3 never used the Markov assumption and can be applied to non Markovian processes (see Section 4.3). So the present work is a general set-up where one can deduce PDE even from non Markovian microscopic random dynamics. Note also that only boundedness assumptions and not continuity ones are necessary to directly obtain (4.2) via Theorem 3.3: to obtain the classical Kolmogorov theorem, one would have assumed $f \in \mathcal{C}^0(\mathbb{R}^2_+)$ rather than $f \in L^\infty(\mathbb{R}^2_+)$.

4.2. *Generalized Wold process*

In the case where $\lambda\left(t, \mathcal{F}^N_{t-}\right) = f(S_{t-}, A^1_t, ..., A^k_t)$, with f being a non-negative function, one can define in a similar way $u_k\left(t, s, a_1, \ldots, a_k\right)$ which is informally the distribution at time t of the processes with age s and past given by $a_1, ...a_k$ for the last k ISI's. We want to investigate this case not for its Markovian properties, which are nevertheless presented in Proposition B.2 in the appendix for sake of completeness, but because this is the first basic example where the initial condition is indeed impacting $\rho_{\lambda,\mathbb{P}_0}$ in Theorem 3.3.

To do so, the whole machinery applied on $u(\mathrm{d}t, \mathrm{d}s)$ is first extended in the next result to $u_k\left(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a^1, \ldots, \mathrm{d}a^k\right)$ which represents the dynamics of the age and the last k ISI's. This could have been done in a very general way by an easy generalisation of Theorem 3.3. However to avoid too cumbersome equations, we express it only for generalized Wold processes to provide a clean setting to illustrate the impact of the initial conditions on $\rho_{\lambda,\mathbb{P}_0}$. Hence, we similarly define a random distribution $U_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, \ldots, \mathrm{d}a_k)$ such that its evaluation at any given time t exists and is

$$U_k(t, \mathrm{d}s, \mathrm{d}a_1, \ldots, \mathrm{d}a_k) = \delta_{(S_{t-}, A^1_t, \ldots, A^k_t)}(\mathrm{d}s, \mathrm{d}a_1, \ldots, \mathrm{d}a_k). \tag{4.3}$$

The following result states the PDE satisfied by $u_k = \mathbb{E}[U_k]$.

Proposition 4.1. *Let k be a positive integer and f be some non negative function on $\mathbb{R}_+^{k+1}$. Let N be a generalized Wold process with predictable age process $(S_{t-})_{t>0}$, associated points $(T_i)_{i\in\mathbb{Z}}$ and intensity $\lambda(t, \mathcal{F}^N_{t-}) = f(S_{t-}, A^1_t, ..., A^k_t)$ satisfying $\left(\mathcal{A}^{\mathbb{L}^1, exp}_{\lambda, loc}\right)$, where $A^1_t, \ldots, A^k_t$ are the successive ages defined by (2.2). Suppose that $\mathbb{P}_0$ is such that $\mathbb{P}_0(T_{-k} > -\infty) = 1$. Let U_k be defined by*

$$U_k\left(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, \ldots, \mathrm{d}a_k\right) = \sum_{i=0}^{+\infty} \eta_{T_i}(\mathrm{d}t, \mathrm{d}s) \prod_{j=1}^{k} \delta_{A^j_{T_i}}(\mathrm{d}a_j)\ \mathbb{1}_{0\le t\le T_{i+1}}, \tag{4.4}$$

If N is the result of Ogata's thinning on the Poisson process Π, then U_k satisfies (4.3) and $(\mathcal{P}_{Fubini})$ a.s. in Π and $\mathcal{F}^N_0$. Assume that the initial condition u^{in}_k, defined as the distribution of $(-T_0, A^1_0, \ldots, A^k_0)$ which is a random vector in $\mathbb{R}^{k+1}$, is such that $u^{in}_k(\{0\} \times \mathbb{R}^k_+) = 0$. Then U_k admits a mean measure u_k which also satisfies $(\mathcal{P}_{Fubini})$ and the following system in the weak sense: on $\mathbb{R}_+ \times \mathbb{R}^{k+1}_+$,

$$\left\{\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right\} u_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_k) + f(s, a_1, ..., a_k) u_k(\mathrm{d}t, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_k) = 0, \tag{4.5}$$

$$u_k\left(\mathrm{d}t, 0, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_{k-1}\right) = \int_{a_k=0}^{\infty} f(s, a_1, ..., a_k)\, u_k(t, \mathrm{d}s, \mathrm{d}a_1, ..., \mathrm{d}a_k)\ \mathrm{d}t, \tag{4.6}$$

$$u_k\left(0, \mathrm{d}s, \mathrm{d}a_1, \ldots, \mathrm{d}a_k\right) = u^{in}_k\left(\mathrm{d}s, \mathrm{d}a_1, \ldots, \mathrm{d}a_k\right). \tag{4.7}$$

We have assumed $u^{in}_k(\{0\}\times\mathbb{R}^k_+) = 0$ (i.e. $T_0 \neq 0$ a.s.) for the sake of simplicity, but this assumption may of course be relaxed and Dirac masses at 0 should then be added in a similar way as in Theorem 3.3.

If $f \in L^\infty(\mathbb{R}^{k+1}_+)$, we may apply Proposition 4.1, so that the mean measure u_k satisfy System (4.5)–(4.7). Assuming an initial condition $u^{in}_k \in L^1(\mathbb{R}^{k+1}_+)$, we can prove exactly as for the renewal equation (with a Banach fixed point argument for instance) that there exists a unique solution u_k such that $(t \mapsto u_k(t, \cdot)) \in \mathcal{C}\big(\mathbb{R}_+, L^1(\mathbb{R}^{k+1}_+)\big)$ [34] to the generalized Wold case, the boundary assumption on the kth penultimate point before time 0 being necessary to give sense to the successive ages at time 0. By uniqueness, this proves that the mean measure u_k is this solution, so that it belongs to $\mathcal{C}\big(\mathbb{R}_+, L^1(\mathbb{R}^{k+1}_+)\big)$: Proposition 4.1 leads to a regularity result on the mean measure.

Now that we have clarified the dynamics of the successive ages, one can look at this system from the point of view of Theorem 3.3, that is when only two variables s and t are considered. In this respect, let us note that U defined by (3.3) is such that $U(\mathrm{d}t,\mathrm{d}s)=\int_{a_1,\dots,a_k}U_k(\mathrm{d}t,\mathrm{d}s,\mathrm{d}a_1,\dots,\mathrm{d}a_k)$. Since the integrals and the expectations are exchangeable in the weak sense, the mean measure u defined in Proposition 3.2 is such that $u(\mathrm{d}t,\mathrm{d}s)=\int_{a_1,\dots,a_k}u_k(\mathrm{d}t,\mathrm{d}s,\mathrm{d}a_1,\dots,\mathrm{d}a_k)$. But (4.5) in the weak sense means, for all $\varphi\in C^\infty_{c,b}(\mathbb{R}^{k+2})$,

$$\int\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)\varphi(t,s,a_1,...,a_k)u_k\left(\mathrm{d}t,\mathrm{d}s,\mathrm{d}a_1,\dots,\mathrm{d}a_k\right)$$
$$+\int\left[\varphi\left(t,0,a_1,\dots,a_k\right)-\varphi(t,s,a_1,\dots,a_k)\right]f\left(s,a_1,\dots,a_k\right)u_k\left(\mathrm{d}t,\mathrm{d}s,\mathrm{d}a_1,\dots,\mathrm{d}a_k\right)$$
$$+\int\varphi\left(0,s,a_1,\dots,a_k\right)u_k^{in}\left(\mathrm{d}s,\mathrm{d}a_1,\dots,\mathrm{d}a_k\right)=0. \quad (4.8)$$

Letting $\psi\in C^\infty_{c,b}(\mathbb{R}^2)$ and $\varphi\in C^\infty_{c,b}(\mathbb{R}^{k+2})$ being such that

$$\forall\, t,s,a_1,\dots,a_k,\quad \varphi(t,s,a_1,\dots,a_k)=\psi(t,s),$$

we end up proving that the function $\rho_{\lambda,\mathbb{P}_0}$ defined in Theorem 3.3 satisfies

$$\rho_{\lambda,\mathbb{P}_0}(t,s)u\left(\mathrm{d}t,\mathrm{d}s\right)=\int_{a_1,\dots,a_k}f\left(s,a_1,\dots,a_k\right)u_k\left(\mathrm{d}t,\mathrm{d}s,\mathrm{d}a_1,\dots,\mathrm{d}a_k\right), \quad (4.9)$$

$u(\mathrm{d}t,\mathrm{d}s)-$almost everywhere ($a.e.$). Equation (4.9) means exactly from a probabilistic point of view that

$$\rho_{\lambda,\mathbb{P}_0}(t,s)=\mathbb{E}\left[f(S_{t-},A^1_t,...,A^k_t)|S_{t-}=s\right],\quad u(\mathrm{d}t,\mathrm{d}s)-a.e.$$

Therefore, in the particular case of generalized Wold process, the quantity $\rho_{\lambda,\mathbb{P}_0}$ depends on the shape of the intensity (here the function f) and also on u_k. But, by Proposition 4.1, u_k depends on its initial condition given by the distribution of $(-T_0,A^1_0,\dots,A^k_0)$, and not only $-T_0$ as in the initial condition for u. That is, as announced in the remarks following Theorem 3.3, $\rho_{\lambda,\mathbb{P}_0}$ depends in particular on the whole distribution of the underlying process before time 0, namely $\mathbb{P}_0$ and not only on the initial condition for u. Here, for generalized Wold processes, it only depends on the last k points before time 0. For more general non Markovian settings, the integration cannot be simply described by a measure u_k in dimension $(k+2)$ being integrated with respect to $\mathrm{d}a^1...\mathrm{d}a^k$. In general, the integration has to be done on all the "randomness" hidden behind the dependence of $\lambda(t,\mathcal{F}^N_{t-})$ with respect to the past once S_{t-} is fixed and in this sense it depends on the whole distribution $\mathbb{P}_0$ of N_-. This is made even clearer on the following non Markovian example: the Hawkes process.

4.3. *Hawkes process*

As we have seen in Section 2.2, there are many different examples of Hawkes processes that can all be expressed as $\lambda\left(t,\mathcal{F}^N_{t-}\right)=\phi\left(\int_{-\infty}^{t-}h\left(t-x\right)N(\mathrm{d}x)\right)$, where the main case is $\phi(\theta)=\mu+\theta$, for μ some positive constant, which is the linear case.

When there is no point before 0, $\lambda\left(t, \mathcal{F}_{t-}^N\right) = \phi\left(\int_0^{t-} h\left(t-x\right) N(\mathrm{d}x)\right)$. In this case, the interpretation is so close to (PPS) that the first guess, which is wrong, would be that the analogous in (PPS) is

$$p(s, X(t)) = \phi(X(t)), \tag{4.10}$$

where $X(t) = \mathbb{E}\left[\int_0^{t-} h\left(t-x\right) N(\mathrm{d}x)\right] = \int_0^t h\left(t-x\right) u(\mathrm{d}x, 0)$. This is wrong, even in the linear case since $\lambda\left(t, \mathcal{F}_{t-}^N\right)$ depends on all the previous points. Therefore $\rho_{\lambda,\mathbb{P}_0}$ defined by (3.9) corresponds to a conditioning given only the last point.

By looking at this problem through the generalized Wold approach, one can hope that for h decreasing fast enough:

$$\lambda\left(t, \mathcal{F}_{t-}^N\right) \simeq \phi\left(h(S_{t-}) + h(S_{t-} + A_t^1) + ... + h(S_{t-} + A_t^1 + ... + A_t^k)\right).$$

In this sense and with respect to generalized Wold processes described in the previous section, we are informally integrating on "all the previous points" except the last one and not integrating over all the previous points. This is informally why (4.10) is wrong even in the linear case. Actually, $\rho_{\lambda,\mathbb{P}_0}$ is computable for linear Hawkes processes : we show in the next section that $\rho_{\lambda,\mathbb{P}_0}(t,s) \neq \phi(\int_{-\infty}^t h(t-x)u(\mathrm{d}x,0)) = \mu + \int_0^\infty h(t-x)u(\mathrm{d}x,0)$ and that $\rho_{\lambda,\mathbb{P}_0}$ explicitly depends on $\mathbb{P}_0$.

4.3.1. *Linear Hawkes process*

We are interested in Hawkes processes with a past before time 0 given by $\mathcal{F}_0^N$, which is not necessarily the past given by a stationary Hawkes process. To illustrate the fact that the past is impacting the value of $\rho_{\lambda,\mathbb{P}_0}$, we focus on two particular cases:
$\left(\mathcal{A}_{N_-}^1\right) \mid N_- = \{T_0\}$ a.s. and T_0 admits a bounded density f_0 on $\mathbb{R}_-$
$\left(\mathcal{A}_{N_-}^2\right) \mid N_-$ is an homogeneous Poisson process with intensity α on $\mathbb{R}_-$

Before stating the main result, we need some technical definitions. Indeed the proof is based on the underlying branching structure of the linear Hawkes process described in Section B.3.1 of the appendix and the following functions (L_s, G_s) are naturally linked to this branching decomposition (see Lemma B.7).

Lemma 4.2. *Let $h \in L^1(\mathbb{R}_+)$ such that $\|h\|_{L^1} < 1$. For all $s \geq 0$, there exist a unique solution $(L_s, G_s) \in L^1(\mathbb{R}_+) \times L^\infty(\mathbb{R}_+)$ of the following system*

$$\log(G_s(x)) = \int_0^{(x-s)\vee 0} G_s(x-w)h(w)\mathrm{d}w - \int_0^x h(w)\mathrm{d}w, \tag{4.11}$$

$$L_s(x) = \int_{s\wedge x}^x \left(h\left(w\right) + L_s(w)\right) G_s(w)h(x-w)\,\mathrm{d}w, \tag{4.12}$$

where $a \vee b$ (resp. $a \wedge b$) denotes the maximum (resp. minimum) between a and b. Moreover, $L_s(x \leq s) \equiv 0$, $G_s : \mathbb{R}_+ \to [0,1]$, and L_s is uniformly bounded in L^1.

This result allows to define two other important quantities, K_s and q, by, for all $s, t \geq 0, z \in \mathbb{R}$,

$$K_s(t,z) := \int_0^{(t-s)\vee 0} \left[h(t-x) + L_s(t-x)\right] G_s(t-x) h(x-z) \mathrm{d}x,$$

$$\log(q(t,s,z)) := -\int_{(t-s)\vee 0}^{t} h(x-z)\mathrm{d}x - \int_0^{(t-s)\vee 0} \left[1 - G_s(t-x)\right] h(x-z)\mathrm{d}x. \quad (4.13)$$

Finally, the following result is just an obvious remark that helps to understand the resulting system.

Remark 4.1. For a non negative $\Phi \in L^\infty(\mathbb{R}_+)$ and $v^{in} \in L^\infty(\mathbb{R}_+)$, there exists a unique solution $v \in L^\infty(\mathbb{R}_+^2)$ in the weak sense to the following system,

$$\frac{\partial}{\partial t} v(t,s) + \frac{\partial}{\partial s} v(t,s) + \Phi(t,s) v(t,s) = 0, \quad (4.14)$$

$$v(t,0) = 1 \qquad v(t=0,s) = v^{in}(s) \quad (4.15)$$

Moreover $t \mapsto v(t,.)$ is in $C(R_+, L^1_{loc}(\mathbb{R}_+))$.

If v^{in} is a survival function (i.e. non increasing from 0 to 1), then $v(t,.)$ is a survival function and $-\partial_s v$ is a probability measure for all $t > 0$.

Proposition 4.3. *Using the notations of Theorem 3.3, let N be a Hawkes process with past before 0 given by N_- satisfying either $\left(\mathcal{A}^1_{N_-}\right)$ or $\left(\mathcal{A}^2_{N_-}\right)$ and with intensity on $\mathbb{R}_+$ given by*

$$\lambda(t, \mathcal{F}^N_{t-}) = \mu + \int_{-\infty}^{t-} h(t-x) N(\mathrm{d}x),$$

where μ is a positive real number and $h \in L^\infty(\mathbb{R}_+)$ is a non-negative function with support in $\mathbb{R}_+$ such that $\int h < 1$.

Then, the mean measure u defined in Proposition 3.2 satisfies Theorem 3.3 and moreover its integral $v(t,s) := \int_s^\infty u(t, d\sigma)$ is the unique solution of the system (4.14)–(4.15) *where v^{in} is the survival function of $-T_0$, and where $\Phi = \Phi^{\mu,h}_{\mathbb{P}_0} \in L^\infty(\mathbb{R}_+)$ is defined by*

$$\Phi^{\mu,h}_{\mathbb{P}_0} = \Phi^{\mu,h}_+ + \Phi^h_{-,\mathbb{P}_0}, \quad (4.16)$$

where for all non negative s, t

$$\Phi^{\mu,h}_+(t,s) = \mu \left(1 + \int_{s\wedge t}^t (h(x) + L_s(x)) G_s(x) \mathrm{d}x\right), \quad (4.17)$$

and where under Assumption $\left(\mathcal{A}^1_{N_-}\right)$,

$$\Phi^h_{-,\mathbb{P}_0}(t,s) = \frac{\int_{-\infty}^{0\wedge(t-s)} \left(h(t-t_0) + K_s(t,t_0)\right) q(t,s,t_0) f_0(t_0) \mathrm{d}t_0}{\int_{-\infty}^{0\wedge(t-s)} q(t,s,t_0) f_0(t_0) \mathrm{d}t_0}, \quad (4.18)$$

or, under Assumption $\left(\mathcal{A}^2_{N_-}\right)$,

$$\Phi^h_{-,\mathbb{P}_0}(t,s) = \alpha \int_{-\infty}^{0\wedge(t-s)} \left(h(t-z) + K_s(t,z)\right) q(t,s,z)\mathrm{d}z. \tag{4.19}$$

In these formulae, L_s, G_s, K_s *and* q *are given by Lemma 4.2 and* (4.13). *Moreover*

$$\forall\, s \geq 0, \quad \int_s^{+\infty} \rho_{\lambda,\mathbb{P}_0}(t,x)u(t,\mathrm{d}x) = \Phi^{\mu,h}_{\mathbb{P}_0}(t,s) \int_s^{+\infty} u(t,\mathrm{d}x). \tag{4.20}$$

The proof is included in Appendix B.3. Proposition 4.3 gives a purely analytical definition for v, and thus for u, in two specific cases, namely $\left(\mathcal{A}^1_{N_-}\right)$ or $\left(\mathcal{A}^2_{N_-}\right)$. In the general case, treated in Appendix B (Proposition B.5), there remains a dependence with respect to the initial condition $\mathbb{P}_0$, *via* the function $\Phi^h_{-,\mathbb{P}_0}$.

Remark 4.2. Contrarily to the general result in Theorem 3.3, Proposition 4.3 focuses on the equation satisfied by $v(\mathrm{d}t,s) = \int_s^{+\infty} u(\mathrm{d}t,\mathrm{d}x)$ because in Equation (4.14) the function parameter $\Phi = \Phi^{\mu,h}_{\mathbb{P}_0}$ may be defined independently of the definitions of v or u, which is *not* the case for the rate $\rho_{\lambda,\mathbb{P}_0}$ appearing in Equation (3.10). Thus, it is possible to depart from the system of equations defining v, study it, prove existence, uniqueness and regularity for v under some assumptions on the initial distribution u^{in} as well as on the birth function h, and then deduce regularity or asymptotic properties for u without any previous knowledge on the underlying process.
In Sections 4.1 and 4.2, we were able to use the PDE formulation to prove that the distribution of the ages u has a density. Here, since we only obtain a closed formula for v and not for u, we would need to derive Equation (4.14) in s to obtain a similar result, so that we need to prove more regularity on $\Phi^{\mu,h}_{\mathbb{P}_0}$. Such regularity for $\Phi^{\mu,h}_{\mathbb{P}_0}$ is not obvious since it depends strongly on the assumptions on N_-. This paves the way for future research, where the PDE formulation would provide regularity on the distribution of the ages, as done above for renewal and Wold processes.

Remark 4.3. These two cases $\left(\mathcal{A}^1_{N_-}\right)$ and $\left(\mathcal{A}^2_{N_-}\right)$ highlight the dependence with respect to all the past before time 0 (i.e. $\mathbb{P}_0$) and not only the initial condition (i.e. the age at time 0). In fact, they can give the same initial condition u^{in} : for instance, $\left(\mathcal{A}^1_{N_-}\right)$ with $-T_0$ exponentially distributed with parameter $\alpha > 0$ gives the same law for $-T_0$ as $\left(\mathcal{A}^2_{N_-}\right)$ with parameter α. However, if we fix some non-negative real number s, one can show that $\Phi^h_{-,\mathbb{P}_0}(0,s)$ is different in those two cases. It is clear from the definitions that for every real number z, $q(0,s,z) = 1$ and $K_s(0,z) = 0$. Thus, in the first case,

$$\Phi^h_{-,\mathbb{P}_0}(0,s) = \frac{\int_{-\infty}^{-s} h(-t_0)\alpha e^{\alpha t_0}\mathrm{d}t_0}{\int_{-\infty}^{-s} \alpha e^{\alpha t_0}\mathrm{d}t_0} = \frac{\int_s^{\infty} h(z)\alpha e^{-\alpha z}\mathrm{d}z}{\int_s^{\infty} \alpha e^{-\alpha z}\mathrm{d}z},$$

while in the second case, $\Phi^h_{-,\mathbb{P}_0}(0,s) = \alpha \int_{-\infty}^{-s} h(-z)\mathrm{d}z = \alpha \int_s^{\infty} h(w)\mathrm{d}w$. Therefore $\Phi^h_{-,\mathbb{P}_0}$ clearly depends on $\mathbb{P}_0$ and not just on the distribution of the last point before 0, and so is $\rho_{\lambda,\mathbb{P}_0}$.

Remark 4.4. If we follow our first guest, $\rho_{\lambda,\mathbb{P}_0}$ would be either $\mu + \int_0^t h(t-x)u(\mathrm{d}x,0)$ or $\mu + \int_{-\infty}^t h(t-x)u(\mathrm{d}x,0)$. In particular, it would not depend on the age s. Therefore by (4.20), so would $\Phi^{\mu,h}_{\mathbb{P}_0}$. But for instance at time $t=0$, when N_- is an homogeneous Poisson process of parameter α, $\Phi^{\mu,h}_{\mathbb{P}_0}(0,s) = \mu + \alpha \int_s^{+\infty} h(w)\mathrm{d}w$, which obviously depends on s. Therefore the intuition linking Hawkes processes and (PPS) does not apply.

4.3.2. *Linear Hawkes process with no past before time 0*

A classical framework in point processes theory is the case in $\left(\mathcal{A}^1_{N_-}\right)$ where $T_0 \to -\infty$, or equivalently, when N has intensity $\lambda(t, \mathcal{F}^N_{t-}) = \mu + \int_0^{t-} h(t-x)N(\mathrm{d}x)$. The problem in this case is that the age at time 0 is not finite. The age is only finite for times greater than the first spiking time T_1.

Here again, the quantity $v(t,s)$ reveals more informative and easier to use: having the distribution of T_0 going to $-\infty$ means that $\mathrm{Supp}(u^{in})$ goes to $+\infty$, so that the initial condition for v tends to value uniformly 1 for any $0 \le s < +\infty$. If we can prove that the contribution of $\Phi^h_{-,\mathbb{P}_0}$ vanishes, the following system is a good candidate to be the limit system:

$$\frac{\partial}{\partial t}v^\infty(t,s) + \frac{\partial}{\partial s}v^\infty(t,s) + \Phi^{\mu,h}_+(t,s)\, v^\infty(t,s) = 0, \tag{4.21}$$

$$v^\infty(t,0) = 1, \qquad v^\infty(0,s) = 1, \tag{4.22}$$

where $\Phi^{\mu,h}_+$ is defined in Proposition 4.3. This leads us to the following proposition.

Proposition 4.4. *Under the assumptions and notations of Proposition 4.3, consider for all $M \ge 0$, v_M the unique solution of system* (4.14)-(4.15) *with Φ given by Proposition 4.3, case $\left(\mathcal{A}^1_{N_-}\right)$, with T_0 uniformly distributed in $[-M-1,-M]$. Then, as M goes to infinity, v_M converges uniformly on any set of the type $(0,T)\times(0,S)$ towards the unique solution v^∞ of System* (4.21)-(4.22).

Conclusion

We present in this article a bridge between univariate point processes, that can model the behavior of one neuron through its spike train, and a deterministic age structured PDE introduced by Pakdaman, Perthame and Salort, named (PPS). More precisely Theorem 3.3 present a PDE that is satisfied by the distribution u of the age s at time t, where the age represents the delay between time t and the last spike before t. This is done in a very weak sense and some technical structure, namely $(\mathcal{P}_{Fubini})$, is required.

The main point is that the "firing rate" which is a deterministic quantity written as $p(s, X(t))$ in (PPS) becomes the conditional expectation of the intensity given the age at time t in Theorem 3.3. This first makes clear that $p(s, X(t))$ should be interpreted as a hazard rate, which gives the probability that a neuron fires given that it has not fired yet. Next, it makes clearly rigorous several "easy guess" bridges between both set-ups when the intensity only depends on the age. But it also explained why when the intensity has a more complex shape (Wold, Hawkes), this term can keep in particular the memory of all that has happened before time 0.

One of the main point of the present study is the Hawkes process, for which what was clearly expected was a legitimation of the term $X(t)$ in the firing rate $p(s, X(t))$ of (PPS), which models the synaptic integration. This is not the case, and the interlinked equations that have been found for the cumulative distribution function $v(t, \cdot)$ do not have a simple nor direct deterministic interpretation. However one should keep in mind that the present bridge, in particular in the population wide approach, has been done for independent neurons. This has been done to keep the complexity of the present work reasonable as a first step. But it is also quite obvious that interacting neurons cannot be independent. So one of the main question is: can we recover (PPS) as a limit with precisely a term of the form $X(t)$ if we consider multivariate Hawkes processes that really model interacting neurons ?

Acknowledgment

This research was partly supported by the European Research Council (ERC Starting Grant SKIPPERAD number 306321), by the french Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration) and by the interdisciplanary axis MTC-NSC of the University of Nice Sophia-Antipolis. MJC acknowledges support from the projects MTM2014-52056-P (Spain) and P08-FQM-04267 from Junta de Andalucía (Spain). We warmly thank François Delarue for very fruitful discussions and ideas.

A. Proofs linked with the PDE

A.1. *Proof of Proposition 3.1*

First, let us verify that U satisfies Equation (3.2). For any $t > 0$,

$$U(t, \mathrm{d}s) = \sum_{i \geq 0} \eta_{T_i}(t, \mathrm{d}s) \mathbb{1}_{0 \leq t \leq T_{i+1}},$$

by definition of U. Yet, $\eta_{T_i}(t, \mathrm{d}s) = \delta_{t-T_i}(\mathrm{d}s)\mathbb{1}_{t>T_i}$, and the only $i \in \mathbb{N}$ such that $T_i < t \leq T_{i+1}$ is $i = N_{t-}$. So, for all $t > 0$, $U(t, \mathrm{d}s) = \delta_{t-T_{N_{t-}}}(\mathrm{d}s) = \delta_{S_{t-}}(\mathrm{d}s)$.

Secondly, let us verify that U satisfies $(\mathcal{P}_{Fubini})$. Let $\varphi \in \mathcal{M}_{c,b}(\mathbb{R}_+^2)$, and let T be

such that for all $t > T$, $\varphi_t^{(1)} = 0$. Then since $U(t, \mathrm{d}s) = \sum_{i=0}^{+\infty} \eta_{T_i}(t, \mathrm{d}s)\mathbb{1}_{0\leq t\leq T_{i+1}}$,

$$\left|\int_{\mathbb{R}_+}\left(\int_{\mathbb{R}_+}\varphi(t,s)U(t,\mathrm{d}s)\right)\mathrm{d}t\right| \leq \int_{\mathbb{R}_+}\left(\int_{\mathbb{R}_+}|\varphi(t,s)|\sum_{i\geq 0}\eta_{T_i}(t,\mathrm{d}s)\mathbb{1}_{0\leq t\leq T_{i+1}}\right)\mathrm{d}t$$
$$= \sum_{i\geq 0}\int_{\mathbb{R}_+}|\varphi(t,t-T_i)|\mathbb{1}_{t>T_i}\mathbb{1}_{0\leq t\leq T_{i+1}}\mathrm{d}t = \sum_{i\geq 0}\int_{\max(0,T_i)}^{T_{i+1}}|\varphi(t,t-T_i)|\mathrm{d}t$$
$$= \int_0^{T_1}|\varphi(t,t-T_0)| + \sum_{i/0<T_i<T}\int_{T_i}^{T_{i+1}}|\varphi(t,t-T_i)|\mathrm{d}t.$$

Since there is a finite number of points of N between 0 and T, on Ω, this quantity is finite and one can exchange $\sum_{i\geq 0}$ and $\int_{t=0}^{+\infty}\int_{s=0}^{+\infty}$. Therefore, since all the η_{T_i} satisfy $(\mathcal{P}_{Fubini})$ and $\varphi(t,s)\mathbb{1}_{0\leq t\leq T_{i+1}}$ is in $\mathcal{M}_{c,b}(\mathbb{R}_+^2)$, so does U.

For the dynamics of U, similar computations lead for every $\varphi \in C_{c,b}^{\infty}({\mathbb{R}_+}^2)$ to

$$\int \varphi(t,s)\, U(\mathrm{d}t, \mathrm{d}s) = \sum_{i\geq 0}\int_{\max(0,-T_i)}^{T_{i+1}-T_i}\varphi(s+T_i, s)\,\mathrm{d}s.$$

We also have

$$\int\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)\varphi(t,s)\,U(\mathrm{d}t,\mathrm{d}s) = \sum_{i\geq 0}\int_{\max(0,-T_i)}^{T_{i+1}-T_i}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)\varphi(s+T_i,s)\,\mathrm{d}s$$
$$= \sum_{i\geq 1}\left[\varphi(T_{i+1}, T_{i+1}-T_i) - \varphi(T_i, 0)\right] + \varphi(T_1, T_1-T_0) - \varphi(0, -T_0). \quad (A.1)$$

It remains to express the term with $\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)}\Pi(\mathrm{d}t,\mathrm{d}x) = \sum_{i\geq 0}\delta_{T_{i+1}}(\mathrm{d}t)$, that is

$$\int\varphi(t,s)\,U(t,\mathrm{d}s)\sum_{i\geq 0}\delta_{T_{i+1}}(\mathrm{d}t) = \int\left(\int\varphi(t,s)\,U(t,\mathrm{d}s)\right)\sum_{i\geq 0}\delta_{T_{i+1}}(\mathrm{d}t)$$
$$= \int\varphi(t,S_{t-})\sum_{i\geq 0}\delta_{T_{i+1}}(\mathrm{d}t) = \sum_{i\geq 0}\varphi(T_{i+1}, T_{i+1}-T_i), \quad (A.2)$$

and, since $\int U(t,\mathrm{d}s) = 1$ for all $t > 0$,

$$\int\int\varphi(t,0)\,U(t,\mathrm{d}s)\sum_{i\geq 0}\delta_{T_{i+1}}(\mathrm{d}t) = \sum_{i\geq 0}\varphi(T_{i+1}.0), \quad (A.3)$$

Identifying all the terms in the right-hand side of Equation (A.1), this lead to Equation (3.7), which is the weak formulation of System (3.4)–(3.6).

A.2. *Proof of Proposition 3.2*

Let $\varphi \in \mathcal{M}_{c,b}(\mathbb{R}_+^2)$, and let T be such that for all $t > T$, $\varphi_t^{(1)} = 0$. Then,

$$\int|\varphi(t,s)|U(t,\mathrm{d}s) \leq ||\varphi||_{L^\infty}\mathbb{1}_{0\leq t\leq T}, \quad (A.4)$$

since at any fixed time $t > 0$, $\int U(t, \mathrm{d}s) = 1$. Therefore, the expectation $\mathbb{E}\left[\int \varphi(t,s)U(t,\mathrm{d}s)\right]$ is well-defined and finite and so $u(t,.)$ is well-defined.

On the other hand, at any fixed age s,

$$\begin{aligned}\int |\varphi(t,s)|U(\mathrm{d}t,s) &= \sum_{i=0}^{\infty} |\varphi(s+T_i,s)|\mathbb{1}_{0\le s\le T_{i+1}-T_i} \\ &= \sum_{i\ge 0} |\varphi(s+T_i,s)|\mathbb{1}_{0\le s+T_i\le T}\mathbb{1}_{0\le s\le T_{i+1}-T_i},\end{aligned}$$

because for all $t > T$, $\varphi_t^{(1)} = 0$. Then, one can deduce the following bound

$$\begin{aligned}&\int |\varphi(t,s)|U(\mathrm{d}t,s) \\ &\quad\le |\varphi(s+T_0,s)|\mathbb{1}_{-T_0\le s\le T-T_0}\mathbb{1}_{0\le s\le T_1-T_0} + \sum_{i\ge 1}|\varphi(s+T_i,s)|\mathbb{1}_{0\le s\le T}\mathbb{1}_{T_i\le T} \\ &\quad\le ||\varphi||_{L^\infty}\left(\mathbb{1}_{-T_0\le s\le T-T_0} + N_T\mathbb{1}_{0\le s\le T}\right).\end{aligned}$$

Since the intensity is L^1_{loc} in expectation, $\mathbb{E}[N_T] = \mathbb{E}\left[\int_0^T \lambda(t,\mathcal{F}^N_{t-})\mathrm{d}t\right] < \infty$ and

$$\mathbb{E}\left[\int |\varphi(t,s)|U(\mathrm{d}t,s)\right] \le ||\varphi||_{L^\infty}\left(\mathbb{E}\left[\mathbb{1}_{-T_0\le s\le T-T_0}\right] + \mathbb{E}\left[N_T\right]\mathbb{1}_{0\le s\le T}\right), \tag{A.5}$$

so the expectation is well-defined and finite and so $u(\cdot,s)$ is well-defined.

Now, let us show $(\mathcal{P}_{Fubini})$. First Equation (A.4) implies

$$\int \mathbb{E}\left[\int |\varphi(t,s)|U(t,\mathrm{d}s)\right]\mathrm{d}t \le T||\varphi||_{L^\infty},$$

and Fubini's theorem implies that the following integrals are well-defined and that the following equality holds,

$$\int \mathbb{E}\left[\int \varphi(t,s)U(t,\mathrm{d}s)\right]\mathrm{d}t = \mathbb{E}\left[\int\int \varphi(t,s)U(t,\mathrm{d}s)\mathrm{d}t\right]. \tag{A.6}$$

Secondly, Equation (A.5) implies

$$\int \mathbb{E}\left[\int |\varphi(t,s)|U(\mathrm{d}t,s)\right]\mathrm{d}s \le ||\varphi||_{L^\infty}\left(T + T\mathbb{E}\left[N_T\right]\right),$$

by exchanging the integral with the expectation and Fubini's theorem implies that the following integrals are well-defined and that the following equality holds,

$$\int \mathbb{E}\left[\int \varphi(t,s)U(\mathrm{d}t,s)\right]\mathrm{d}s = \mathbb{E}\left[\int\int \varphi(t,s)U(\mathrm{d}t,s)\mathrm{d}s\right]. \tag{A.7}$$

Now, it only remains to use $(\mathcal{P}_{Fubini})$ for U to deduce that the right members of Equations (A.6) and (A.7) are equal. Moreover, $(\mathcal{P}_{Fubini})$ for U tells that these two quantities are equal to $\mathbb{E}\left[\int\int \varphi(t,s)U(\mathrm{d}t,\mathrm{d}s)\right]$. This concludes the proof.

A.3. *Proof of Theorem 3.3*

Let $\rho_{\lambda,\mathbb{P}_0}(t,s) := \liminf_{\varepsilon\downarrow 0} \frac{\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\mathbb{1}_{|S_{t-}-s|\leq\varepsilon}\right]}{\mathbb{P}(|S_{t-}-s|\leq\varepsilon)}$, for every $t>0$ and $s\geq 0$. Since $(\lambda(t,\mathcal{F}_{t-}^N))_{t>0}$ and $(S_{t-})_{t>0}$ are predictable processes, and a fortiori progressive processes (see page 9 in [3]), $\rho_{\lambda,\mathbb{P}_0}$ is a measurable function of (t,s).

For every $t>0$, let μ_t be the measure defined by $\mu_t(A)=\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\mathbb{1}_A(S_{t-})\right]$ for all measurable set A. Since Assumption $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp}\right)$ implies that $\mathrm{d}t$-a.e. $\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\right]<\infty$ and since $u(t,\mathrm{d}s)$ is the distribution of S_{t-}, μ_t is absolutely continuous with respect to $u(t,\mathrm{d}s)$ for $\mathrm{d}t$-almost every t.

Let f_t denote the Radon Nikodym derivative of μ_t with respect to $u(t,\mathrm{d}s)$. For $u(t,\mathrm{d}s)$-a.e. s, $f_t(s)=\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\middle|\, S_{t-}=s\right]$ by definition of the conditional expectation. Moreover, a Theorem of Besicovitch [27] claims that for $u(t,\mathrm{d}s)$-a.e. s, $f_t(s)=\rho_{\lambda,\mathbb{P}_0}(t,s)$. Hence, the equality$\rho_{\lambda,\mathbb{P}_0}(t,s)=\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\middle|\, S_{t-}=s\right]$ holds $u(t,\mathrm{d}s)\mathrm{d}t=u(\mathrm{d}t,\mathrm{d}s)$-almost everywhere.

Next, in order to use $(\mathcal{P}_{Fubini})$, let us note that for any $T,K>0$,

$$\rho_{\lambda,\mathbb{P}_0}^{K,T}:(t,s)\mapsto\left(\rho_{\lambda,\mathbb{P}_0}(t,s)\wedge K\right)\mathbf{1}_{0\leq t\leq T}\in\mathcal{M}_{c,b}(\mathbb{R}_+^2) \tag{A.8}$$

Hence, $\int\int\rho_{\lambda,\mathbb{P}_0}^{K,T}(t,s)u(\mathrm{d}t,\mathrm{d}s)=\int\left(\int\rho_{\lambda,\mathbb{P}_0}^{K,T}(t,s)u(t,\mathrm{d}s)\right)\mathrm{d}t$ which is always upper bounded by $\int_0^T\left(\int\rho_{\lambda,\mathbb{P}_0}(t,s)u(t,\mathrm{d}s)\right)\mathrm{d}t=\int_0^T\mu_t(\mathbb{R}_+)\mathrm{d}t=\int_0^T\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\right]\mathrm{d}t<\infty$.

Letting $K\to\infty$, one has that $\int_0^T\int\rho_{\lambda,\mathbb{P}_0}(t,s)u(\mathrm{d}t,\mathrm{d}s)$ is finite for all $T>0$.

Once $\rho_{\lambda,\mathbb{P}_0}$ correctly defined, the proof of Theorem 3.3 is a direct consequence of Proposition 3.1.

More precisely, let us show that (3.7) implies (3.13). Taking the expectation of (3.7) gives that for all $\varphi\in C_{c,b}^\infty(\mathbb{R}_+^2)$,

$$\mathbb{E}\left[\int\left[\varphi\left(t,s\right)-\varphi(t,0)\right]\left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)}\Pi\left(\mathrm{d}t,\mathrm{d}x\right)\right)U\left(t,\mathrm{d}s\right)\right]-\int\varphi\left(0,s\right)u^{in}\left(\mathrm{d}s\right)$$
$$-\int\left(\partial_t+\partial_s\right)\varphi\left(t,s\right)u\left(\mathrm{d}t,\mathrm{d}s\right)=0. \tag{A.9}$$

Let us denote $\psi(t,s):=\varphi(t,s)-\varphi(t,0)$. Due to Ogata's thinning construction, $\left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)}\Pi\left(\mathrm{d}t,\mathrm{d}x\right)\right)=N(\mathrm{d}t)\mathbb{1}_{t>0}$ where N is the point process constructed by thinning, and so,

$$\mathbb{E}\left[\int\psi\left(t,s\right)\left(\int_{x=0}^{\lambda\left(t,\mathcal{F}_{t-}^N\right)}\Pi\left(\mathrm{d}t,\mathrm{d}x\right)\right)U\left(t,\mathrm{d}s\right)\right]=\mathbb{E}\left[\int_{t>0}\psi\left(t,S_{t-}\right)N(\mathrm{d}t)\right]. \tag{A.10}$$

But $\psi(t,S_{t-})$ is a $(\mathcal{F}_t^N)$-predictable process and

$$\mathbb{E}\left[\int_{t>0}|\psi(t,S_{t-})|\lambda(t,\mathcal{F}_{t-}^N)\mathrm{d}t\right]\leq\|\psi\|_{L^\infty}\,\mathbb{E}\left[\int_0^T\lambda(t,\mathcal{F}_{t-}^N)\mathrm{d}t\right]<\infty,$$

hence, using the martingale property of the predictable intensity,

$$\mathbb{E}\left[\int_{t>0}\psi\left(t,S_{t-}\right)N(\mathrm{d}t)\right]=\mathbb{E}\left[\int_{t>0}\psi\left(t,S_{t-}\right)\lambda\left(t,\mathcal{F}_{t-}^{N}\right)\mathrm{d}t\right]. \tag{A.11}$$

Moreover, thanks to Fubini's Theorem, the right-hand term is finite and equal to $\int\mathbb{E}[\psi\left(t,S_{t-}\right)\lambda(t,\mathcal{F}_{t-}^{N})]\mathrm{d}t$, which can also be seen as

$$\int\mathbb{E}\left[\psi\left(t,S_{t-}\right)\rho_{\lambda,\mathbb{P}_0}(t,S_{t-})\right]\mathrm{d}t=\int\psi(t,s)\rho_{\lambda,\mathbb{P}_0}(t,s)u(t,\mathrm{d}s)\mathrm{d}t. \tag{A.12}$$

For all $K>0$, $((t,s)\mapsto\psi(t,s)\left(\rho_{\lambda,\mathbb{P}_0}(t,s)\wedge K\right))\in\mathcal{M}_{c,b}(\mathbb{R}_+^2)$ and, from $(\mathcal{P}_{Fubini})$, it is clear that $\int\psi(t,s)\left(\rho_{\lambda,\mathbb{P}_0}(t,s)\wedge K\right)u(t,\mathrm{d}s)\mathrm{d}t=\int\psi(t,s)\left(\rho_{\lambda,\mathbb{P}_0}(t,s)\wedge K\right)u(\mathrm{d}t,\mathrm{d}s)$. Since one can always upper-bound this quantity in absolute value by $\|\psi\|_{L^\infty}\int_0^T\int_s\rho_{\lambda,\mathbb{P}_0}(t,s)u(\mathrm{d}t,\mathrm{d}s)$, this is finite. Letting $K\to\infty$ one can show that

$$\int\psi(t,s)\rho_{\lambda,\mathbb{P}_0}(t,s)u(t,\mathrm{d}s)\mathrm{d}t=\int\psi(t,s)\rho_{\lambda,\mathbb{P}_0}(t,s)u(\mathrm{d}t,\mathrm{d}s). \tag{A.13}$$

Gathering (A.10)-(A.13) with (A.9) gives (3.13).

A.4. *Proof of Corollary 3.4*

For all $i\in\mathbb{N}^*$, let us denote $N_+^i=N^i\cap(0,+\infty)$ and $N_-^i=N^i\cap\mathbb{R}_-$. Thanks to Proposition B.12, the processes N_+^i can be seen as constructed via thinning of independent Poisson processes on $\mathbb{R}_+^2$. Let $(\Pi^i)_{i\in\mathbb{N}}$ be the sequence of point measures associated to independent Poisson processes of intensity 1 on $\mathbb{R}_+^2$ given by Proposition B.12. Let T_0^i denote the closest point to 0 in N_-^i. In particular, $(T_0^i)_{i\in\mathbb{N}^*}$ is a sequence of i.i.d. random variables.

For each i, let U^i denote the solution of the microscopic equation corresponding to Π^i and T_0^i as defined in Proposition 3.1 by (3.3). Using (3.2), it is clear that $\sum_{i=1}^n\delta_{S_{t-}^i}(\mathrm{d}s)=\sum_{i=1}^n U^i(t,\mathrm{d}s)$ for all $t>0$. Then, for every $\varphi\in C_{c,b}^\infty(\mathbb{R}_+^2)$,

$$\int\varphi(t,s)\left(\frac{1}{n}\sum_{i=1}^n\delta_{S_t^i}(\mathrm{d}s)\right)=\frac{1}{n}\sum_{i=1}^n\int\varphi(t,s)U^i(t,\mathrm{d}s).$$

The right-hand side is a sum n i.i.d. random variables with mean $\int\varphi(t,s)u(t,\mathrm{d}s)$, so (3.14) clearly follows from the law of large numbers.

B. Proofs linked with the various examples

B.1. *Renewal process*

Proposition B.1. *With the notations of Section 2, let N be a point process on $\mathbb{R}$, with predictable age process $(S_{t-})_{t>0}$, such that $T_0=0$ a.s. The following statements are equivalent:*

(i) $N_+=N\cap(0,+\infty)$ is a renewal process with ISI's distribution given by some density $\nu:\mathbb{R}_+\to\mathbb{R}_+$.

(ii) N admits $\lambda(t,\mathcal{F}_{t-}^N)=f(S_{t-})$ as an intensity on $(0,+\infty)$ and $\left(\lambda(t,\mathcal{F}_{t-}^N)\right)_{t>0}$ satisfies $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,a.s.}\right)$, for some $f:\mathbb{R}_+\to\mathbb{R}_+$.

In such a case, for all $x\geq 0$, f and ν satisfy

- $\nu(x)=f(x)\exp(-\int_0^x f(y)dy)$ *with the convention* $\exp(-\infty)=0$, (B.1)
- $f(x)=\dfrac{\nu(x)}{\int_x^\infty \nu(y)dy}$ *if* $\int_x^\infty \nu(y)dy\neq 0$, *else* $f(x)=0$. (B.2)

Proof. For (ii) $\Rightarrow$ (i). Since $T_0=0$ a.s., Point (2) of Proposition B.2 given later on for the general Wold case implies that the ISI's of N forms a Markov chain of order 0 i.e. they are i.i.d. with density given by (B.1).

For (i) $\Rightarrow$ (ii). Let $x_0=\inf\{x\geq 0,\int_x^\infty \nu(y)dy=0\}$. It may be infinite. Let us define f by (B.2) for every $0\leq x<x_0$ and let $\tilde{N}$ be a point process on $\mathbb{R}$ such that $\tilde{N}_-=N_-$ and $\tilde{N}$ admits $\lambda(t,\mathcal{F}_{t-}^{\tilde{N}})=f(S_{t-}^{\tilde{N}})$ as an intensity on $(0,+\infty)$ where $(S_{t-}^{\tilde{N}})_{t>0}$ is the predictable age process associated to $\tilde{N}$. Applying (ii) $\Rightarrow$ (i) to $\tilde{N}$ gives that the ISI's of $\tilde{N}$ are i.i.d. with density given by

$$\tilde{\nu}(x)=\frac{\nu(x)}{\int_x^\infty \nu(y)dy}\exp\left(-\int_0^x \frac{\nu(y)}{\int_y^\infty \nu(z)\mathrm{d}z}dy\right),$$

for every $0\leq x<x_0$ and $\tilde{\nu}(x)=0$ for $x\geq x_0$. It is clear that $\nu=\tilde{\nu}$ since the function $x\mapsto \frac{1}{\int_x^\infty \nu(y)dy}\exp\left(-\int_0^x \frac{\nu(y)}{\int_y^\infty \nu(z)\mathrm{d}z}dy\right)$ is differentiable with derivative equal to 0. Since N and $\tilde{N}$ are renewal processes with same density ν and same first point $T_0=0$, they have the same distribution. Since the intensity characterizes a point process, N also admits $\lambda(t,\mathcal{F}_{t-}^N)=f(S_{t-}^N)$ as an intensity on $(0,+\infty)$. Moreover, since N is a renewal process, it is non-explosive in finite time and so $\left(\lambda(t,\mathcal{F}_{t-}^N)\right)_{t>0}$ satisfies $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,a.s.}\right)$. □

B.2. *Generalized Wold processes*

In this Section, we suppose that there exists $k\geq 0$ such that the underlying point process N has intensity

$$\lambda\left(t,\mathcal{F}_{t-}^N\right)=f(S_{t-},A_t^1,...,A_t^k), \tag{B.3}$$

where f is a function and the A^i's are defined by Equation (2.2).

B.2.1. *Markovian property and the resulting PDE*

Let N be a point process of intensity given by (B.3). If $T_{-k}>-\infty$, its associated age process $(S_t)_t$ can be defined up to $t>T_{-k}$. Then let, for any integer $i\geq -k$,

$$\mathbb{A}_i=T_{i+1}-T_i=S_{T_{i+1}-} \tag{B.4}$$

and denote $(\mathcal{F}_i^{\mathbb{A}})_{i\geq -k}$ the natural filtration associated to $(\mathbb{A}_i)_{i\geq -k}$.

For any $t \geq 0$, and point process Π on $\mathbb{R}_+^2$, let us denote $\Pi_{\geq t}$ (resp. $\Pi_{>t}$) the restriction to $\mathbb{R}_+^2$ (resp. $(0,+\infty)\times\mathbb{R}_+$) of the point process Π shifted t time units to the left on the first coordinate. That is, $\Pi_{\geq t}(C\times D) = \Pi((t+C)\times D)$ for all $C \in \mathcal{B}(\mathbb{R}_+), D \in \mathcal{B}(\mathbb{R}_+)$ (resp. $C \in \mathcal{B}((0,+\infty))$).

Proposition B.2.

Let consider k a non-negative integer, f some non negative function on $\mathbb{R}_+^{k+1}$ and N a generalized Wold process of intensity given by (B.3)*. Suppose that $\mathbb{P}_0$ is such that $\mathbb{P}_0(T_{-k} > -\infty) = 1$ and that $\left(\lambda(t,\mathcal{F}_{t-}^N)\right)_{t>0}$ satisfies $\left(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,a.s.}\right)$. Then,*

(1) If $(X_t)_{t\geq 0} = \left((S_{t-}, A_t^1, ..., A_t^k)\right)_{t\geq 0}$, then for any finite non-negative stopping time τ, $(X_t^\tau)_{t\geq 0} = (X_{t+\tau})_{t\geq 0}$ is independent of $\mathcal{F}_{\tau-}^N$ given X_τ.

(2) the process $(\mathbb{A}_i)_{i\geq 1}$ given by (B.4) *forms a Markov chain of order k with transition measure given by*

$$\nu(\mathrm{d}x, y_1, ..., y_k) = f(x, y_1, ..., y_k)\exp\left(-\int_0^x f(z, y_1, ..., y_k)\mathrm{d}z\right)\mathrm{d}x. \quad \text{(B.5)}$$

If $T_0 = 0$ a.s., this holds for $(\mathbb{A}_i)_{i\geq 0}$.

If f is continuous then $\mathcal{G}$, the infinitesimal generator of $(X_t)_{t\geq 0}$, is given by

$$\forall \phi \in C^1(\mathbb{R}_+^{k+1}), \quad (\mathcal{G}\phi)(s, a_1, ..., a_k) = \\ \frac{\partial}{\partial s}\phi(s, a_1, ..., a_k) + f(s, a_1, ..., a_k)\left(\phi(0, s, a_1, ..., a_{k-1}) - \phi(s, a_1, ..., a_k)\right). \quad \text{(B.6)}$$

Proof. First, let us show the first point of the Proposition. Let Π be such that N is the process resulting of Ogata's thinning with Poisson measure Π. The existence of such a measure is assured by Proposition B.12. We show that for any finite stopping time τ, the process $(X_t^\tau)_{t\geq 0}$ can be expressed as a function of X_τ and $\Pi_{\geq\tau}$ which is the restriction to $\mathbb{R}_+^2$ of the Poisson process Π shifted τ time units to the left on the first coordinate. Let $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^{k+1}$. For all $t \geq 0$, let $Y_t = X_\tau + te_1$ and define

$$R_0 = \inf\left\{t \geq 0, \int_{[0,t]}\int_{x=0}^{f(Y_w)} \Pi_{\geq\tau}(\mathrm{d}w, \mathrm{d}x) = 1\right\}.$$

Note that R_0 may be null, in particular when τ is a jumping time of the underlying point process N. It is easy to check that R_0 can be expressed as a measurable function of X_τ and $\Pi_{\geq\tau}$. Moreover, it is clear that $X_{t\wedge R_0}^\tau = Y_{t\wedge R_0}$ for all $t \geq 0$. So, R_0 can be seen as the delay until the first point of the underlying process N after time τ. Suppose that R_p, the delay until the $(p+1)$th point, is constructed for some $p \geq 0$ and let us show how R_{p+1} can be constructed. For $t \geq R_p$, let $Z_t = \theta(X_{R_p}^\tau) + te_1$, where $\theta : (x_1, \ldots, x_{k+1}) \mapsto (0, x_1, \ldots, x_k)$ is a right shift operator

modelling the dynamics described by (2.3). Let us define

$$R_{p+1} = \inf\left\{ t > R_p, \int_{(R_p, R_p+t]} \int_{x=0}^{f(Z_w)} \Pi_{\geq \tau}(\mathrm{d}w, \mathrm{d}x) = 1 \right\}. \tag{B.7}$$

Note that for any $p \geq 0$, R_{p+1} cannot be null. It is coherent with the fact that the counting process $(N_t)_{t>0}$ only admits jumps with height 1. It is easy to check that R_{p+1} can be expressed as a measurable function of $\theta(X^\tau_{R_p})$ and $\Pi_{>\tau+R_p}$. It is also clear that $X^\tau_{t\wedge R_{p+1}} = Z_{t\wedge R_{p+1}}$ for all $t \geq R_p$. So, R_{p+1} can be seen as the delay until the $(p+2)$th point of the process N after time τ. By induction, $X^\tau_{R_p}$ can be expressed as a function of X_τ and $\Pi_{\geq\tau}$, and this holds for R_{p+1} and $X^\tau_{R_{p+1}}$ too.

To conclude, remark that the process $(X^\tau_t)_{t\geq 0}$ is a measurable function of X_τ and all the R_p's for $p \geq 0$. Thanks to the independence of the Poisson measure Π, $\mathcal{F}^N_{\tau-}$ is independent of $\Pi_{\geq\tau}$. Then, since $(X^\tau_t)_{t\geq 0}$ is a function of X_τ and $\Pi_{\geq\tau}$, $(X^\tau_t)_{t\geq 0}$ is independent of $\mathcal{F}^N_{\tau-}$ given X_τ which concludes the first point.

For Point (2), fix $i \geq 1$ and apply Point (1) with $\tau = T_i$. It appears that in this case, $R_0 = 0$ and $R_1 = \mathbb{A}_i$. Moreover, $R_1 = \mathbb{A}_i$ can be expressed as a function of $\theta(X_\tau)$ and $\Pi_{>\tau}$. However, $\theta(X_\tau) = (0, \mathbb{A}_{i-1}, \ldots, \mathbb{A}_{i-k})$ and $\mathcal{F}^{\mathbb{A}}_{i-1} \subset \mathcal{F}^N_{T_i}$. Since $\tau = T_i$, $\Pi_{>\tau}$ is independent of $\mathcal{F}^N_{T_i}$ and so $\mathbb{A}_i$ is independent of $\mathcal{F}^{\mathbb{A}}_{i-1}$ given $(\mathbb{A}_{i-1}, \ldots, \mathbb{A}_{i-k})$. That is, $(\mathbb{A}_i)_{i\geq 1}$ forms a Markov chain of order k.

Note that if $T_0 = 0$ a.s. (in particular it is non-negative), then one can use the previous argumentation with $\tau = 0$ and conclude that the Markov chain starts one time step earlier, i.e. $(\mathbb{A}_i)_{i\geq 0}$ forms a Markov chain of order k.

For (B.5),$R_1 = \mathbb{A}_i$, defined by (B.7), has the same distribution as the first point of a Poisson process with intensity $\lambda(t) = f(t, \mathbb{A}_{i-1}, \ldots, \mathbb{A}_{i-k})$ thanks to the thinning Theorem. Hence, the transition measure of $(\mathbb{A}_i)_{i\geq 1}$ is given by (B.5).

Now that $(X_t)_{t\geq 0}$ is Markovian, one can compute its infinitesimal generator. Suppose that f is continuous and let $\phi \in C^1_b(\mathbb{R}^{k+1}_+)$, The generator of $(X_t)_{t\geq 0}$ is defined by $\mathcal{G}\phi(s, a_1, \ldots, a_k) = \lim_{h\to 0^+} \frac{P_h - \mathrm{Id}}{h}\phi$, where

$$\begin{aligned} P_h\phi\left(s, a_1, \ldots, a_k\right) &= \mathbb{E}\left[\phi\left(X_h\right) | X_0 = (s, a_1, \ldots, a_k)\right] \\ &= \mathbb{E}\left[\phi\left(X_h\right) \mathbb{1}_{\{N([0,h])=0\}} \middle| X_0 = (s, a_1, \ldots, a_k)\right] \\ &\quad + \mathbb{E}\left[\phi\left(X_h\right) \mathbb{1}_{\{N([0,h])>0\}} \middle| X_0 = (s, a_1, \ldots, a_k)\right] \\ &= E_0 + E_{>0}. \end{aligned}$$

The case with no jump is easy to compute,

$$E_0 = \phi\left(s + h, a_1, \ldots, a_k\right)\left(1 - f\left(s, a_1, \ldots, a_k\right) h\right) + o(h), \tag{B.8}$$

thanks to the continuity of f. When h is small, the probability to have more than two jumps in $[0, h]$ is a $o(h)$, so the second case can be reduced to the case with

exactly one random jump (namely T),

$$\begin{aligned} E_{>0} &= \mathbb{E}\left[\phi\left(X_h\right)\mathbb{1}_{\{N([0,h])=1\}}\middle|X_0=(s,a_1,\dots,a_k)\right]+o(h)\\ &= \mathbb{E}\left[\phi\left(\theta(X_0+T)+(h-T)e_1\right)\mathbb{1}_{\{N\cap[0,h]=\{T\}\}}\middle|X_0=(s,a_1,\dots,a_k)\right]+o(h)\\ &= \mathbb{E}\left[\left(\phi\left(0,s,a_1,\dots,a_{k-1}\right)+o(1)\right)\mathbb{1}_{\{N\cap[0,h]=\{T\}\}}\middle|X_0=(s,a_1,\dots,a_k)\right]+o(h)\\ &= \phi\left(0,s,a_1\dots,a_{k-1}\right)\left(f\left(s,a_1,\dots,a_k\right)h\right)+o\left(h\right), \end{aligned} \tag{B.9}$$

thanks to the continuity of ϕ and f. Gathering (B.8) and (B.9) with the definition of the generator gives (B.6). □

B.2.2. *Sketch of proof of Proposition 4.1*

Let N be the point process construct by Ogata's thinning of the Poisson process Π and U_k be as defined in Proposition 4.1. By an easy generalisation of Proposition 3.1, one can prove that on the event Ω of probability 1, where Ogata's thinning is well defined, and where $T_0<0$, U_k satisfies $(\mathcal{P}_{Fubini})$, (4.3) and on $\mathbb{R}_+\times\mathbb{R}_+^{k+1}$, the following system in the weak sense

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)U_k\left(\mathrm{d}t,\mathrm{d}s,\mathrm{d}\mathbf{a}\right)+\left(\int_{x=0}^{f(s,a_1,\dots,a_k)}\Pi\left(\mathrm{d}t,\mathrm{d}x\right)\right)U_k\left(t,\mathrm{d}s,\mathrm{d}\mathbf{a}\right)=0,$$

$$U_k\left(\mathrm{d}t,0,\mathrm{d}s,\mathrm{d}a_1,...,\mathrm{d}a_{k-1}\right)=\int_{a_k\in\mathbb{R}}\left(\int_{x=0}^{f(s,a_1,\dots,a_k)}\Pi\left(\mathrm{d}t,\mathrm{d}x\right)\right)U_k\left(t,\mathrm{d}s,\mathrm{d}\mathbf{a}\right),$$

with $\mathrm{d}\mathbf{a}=\mathrm{d}a_1\times...\times\mathrm{d}a_k$ and initial condition $U^{in}=\delta_{(-T_0,A_0^1,\dots,A_0^k)}$.

Similarly to Proposition 3.2, one can also prove that for any test function φ in $\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$, $\mathbb{E}\left[\int\varphi(t,s,\mathbf{a})U_k(t,\mathrm{d}s,\mathrm{d}\mathbf{a})\right]$ and $\mathbb{E}\left[\int\varphi(t,s,\mathbf{a})U_k(\mathrm{d}t,s,\mathrm{d}\mathbf{a})\right]$ are finite and one can define $u_k(t,\mathrm{d}s,\mathrm{d}\mathbf{a})$ and $u_k(\mathrm{d}t,s,\mathrm{d}\mathbf{a})$ by, for all φ in $\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$,

$$\int\varphi(t,s,\mathbf{a})u_k(t,\mathrm{d}s,\mathrm{d}\mathbf{a})=\mathbb{E}\left[\int\varphi(t,s,\mathbf{a})U_k(t,\mathrm{d}s,\mathrm{d}\mathbf{a})\right],$$

for all $t\geq 0$, and

$$\int\varphi(t,s,\mathbf{a})u_k(\mathrm{d}t,s,\mathrm{d}\mathbf{a})=\mathbb{E}\left[\int\varphi(t,s,\mathbf{a})U_k(\mathrm{d}t,s,\mathrm{d}\mathbf{a})\right],$$

for all $s\geq 0$. Moreover, $u_k(t,\mathrm{d}s,\mathrm{d}\mathbf{a})$ and $u_k(\mathrm{d}t,s,\mathrm{d}\mathbf{a})$ satisfy $(\mathcal{P}_{Fubini})$ and one can define $u_k(\mathrm{d}t,\mathrm{d}s,\mathrm{d}\mathbf{a})=u_k(t,\mathrm{d}s,\mathrm{d}\mathbf{a})\mathrm{d}t=u_k(\mathrm{d}t,s,\mathrm{d}\mathbf{a})\mathrm{d}s$ on $\mathbb{R}_+^2$, such that for any test function φ in $\mathcal{M}_{c,b}(\mathbb{R}_+^{k+2})$,

$$\int\varphi(t,s,\mathbf{a})u_k(\mathrm{d}t,\mathrm{d}s,\mathrm{d}\mathbf{a})=\mathbb{E}\left[\int\varphi(t,s,\mathbf{a})U_k(\mathrm{d}t,\mathrm{d}s,\mathrm{d}\mathbf{a})\right],$$

quantity which is finite. The end of the proof is completely analogous to the one of Theorem 3.3.

B.3. *Linear Hawkes processes*

B.3.1. *Cluster decomposition*

Proposition B.3. *Let g be a non negative $L^1_{loc}(\mathbb{R}_+)$ function and h a non negative $L^1(\mathbb{R}_+)$ function such that $\|h\|_1 < 1$. Then the branching point process N is defined as $\cup_{k=0}^{\infty} N_k$ the set of all the points in all generations constructed as follows:*

- *Ancestral points are N_{anc} distributed as a Poisson process of intensity g; $N_0 := N_{anc}$ can be seen as the points of generation 0.*
- *Conditionally to N_{anc}, each ancestor $a \in N_{anc}$ gives birth, independently of anything else, to children points $N_{1,a}$ according to a Poisson process of intensity $h(. - a)$; $N_1 = \cup_{a \in N_{anc}} N_{1,a}$ forms the first generation points.*

Then the construction is recursive in k, the number of generations:

- *Denoting N_k the set of points in generation k, then conditionally to N_k, each point $x \in N_k$ gives birth, independently of anything else, to children points $N_{k+1,x}$ according to a Poisson process of intensity $h(. - x)$; $N_{k+1} = \cup_{x \in N_k} N_{k+1,x}$ forms the points of the $(k+1)$th generation.*

This construction ends almost surely in every finite interval. Moreover the intensity of N exists and is given by

$$\lambda(t, \mathcal{F}^N_{t-}) = g(t) + \int_0^{t-} h(t-x) N(\mathrm{d}x).$$

This is the cluster representation of the Hawkes process. When $g \equiv \nu$, this has been proved in [20]. However up to our knowledge this has not been written for a general function g.

Proof. First, let us fix some $A > 0$. The process ends up almost surely in $[0, A]$ because there is a.s. a finite number of ancestors in $[0, A]$: if we consider the family of points attached to one particular ancestor, the number of points in each generation form a sub-critical Galton Watson process with reproduction distribution, a Poisson variable with mean $\int h < 1$ and whose extinction is consequently almost sure.

Next, to prove that N has intensity

$$H(t) = g(t) + \int_0^{t-} h(t-x) N(\mathrm{d}x),$$

we exhibit a particular thinning construction, where on one hand, N is indeed a branching process by construction as defined by the proposition and, which, on the other hand, guarantees that Ogata's thinning project the points below $H(t)$. We can always assume that $h(0) = 0$, since changing the intensity of Poisson process in the branching structure at one particular point has no impact. Hence $H(t) = g(t) + \int_0^t h(t-x) N(\mathrm{d}x)$.

The construction is recursive in the same way. Fix some realisation Π of a Poisson process on $\mathbb{R}^2_+$.

For N_{anc}, project the points below the curve $t \to g(t)$ on $[0, A]$. By construction, N_{anc} is a Poisson process of intensity $g(t)$ on $[0, A]$. Note that for the identification (see Theorem B.11) we just need to do it on finite intervals and that the ancestors that may be born after time A do not have any descendants in $[0, A]$, so we can discard them, since they do not appear in $H(t)$, for $t \leq A$.

Enumerate the points in $N_{anc} \cap [0, A]$ from T_1 to $T_{N_{0,\infty}}$.

- The children of T_1, N_{1,T_1}, are given by the projection of the points of Π whose ordinates are in the strip $t \mapsto (g(t), g(t)+h(t-T_1)]$. As before, by the property of spatial independence of Π, this is a Poisson process of intensity $h(. - T_1)$ conditionally to N_{anc}.
- Repeat until $T_{N_{0,\infty}}$, where $N_{1,T_{N_{0,\infty}}}$ are given by the projection of the points of Π whose ordinates are in the strip $t \mapsto (g(t) + \sum_{i=1}^{N_{0,\infty}-1} h(t - T_i), g(t) + \sum_{i=1}^{N_{0,\infty}} h(t-T_i)]$. As before, by the property of independence of Π, this is a Poisson process of intensity $h(. - T_{N_{0,\infty}})$ conditionally to N_{anc} and because the consecutive strips do not overlap, this process is completely independent of the previous processes (N_{1,T_i})'s that have been constructed.

Note that at the end of this first generation, $N_1 = \cup_{T \in N_{anc}} N_{1,T}$ consists of the projection of points of Π in the strip $t \mapsto (g(t), g(t) + \sum_{i=1}^{N_{0,\infty}} h(t - T_i)]$. They therefore form a Poisson process of intensity $\sum_{i=1}^{N_{0,\infty}} h(t-T_i) = \int h(t-u)N_{anc}(du)$, conditionally to N_{anc}.

For generation $k + 1$ replace in the previous construction N_{anc} by N_k and $g(t)$ by $g(t) + \sum_{j=0}^{k-1} \int h(t - u)dN_j(u)$. Once again we end up for each point x in N_k with a process of children $N_{k+1,x}$ which is a Poisson process of intensity $h(t - x)$ conditionally to N_k and which is totally independent of the other $N_{k+1,y}$'s. Note also that as before, $N_{k+1} = \cup_{x \in N_k} N_{k+1,x}$ is a Poisson process of intensity $\int h(t - u)N_k(du)$, conditionally to $N_0, ..., N_k$.

Hence we are indeed constructing a branching process as defined by the proposition. Because the underlying Galton Watson process ends almost surely, as shown before, it means that there exists a.s. one generation N_{k^*} which will be completely empty and our recursive contruction ends up too.

The main point is to realize that at the end the points in $N = \cup_{k=0}^{\infty} N_k$ are exactly the projection of the points in Π that are below

$$t \mapsto g(t) + \sum_{k=0}^{\infty} \int h(t - u)N_k(du) = g(t) + \sum_{k=0}^{\infty} \int_0^t h(t - u)N_k(du)$$

hence below

$$t \mapsto g(t) + \int_0^t h(t - u)N(du) = H(t).$$

Moreover $H(t)$ is $\mathcal{F}_t^N$ predictable. Therefore by Theorem B.11, N has intensity $H(t)$, which concludes the proof. □

A cluster process N_c, is a branching process, as defined before, which admits intensity $\lambda(t, \mathcal{F}_{t-}^{N_c}) = h(t) + \int_0^{t-} h(t-z)N_c(\mathrm{d}z)$. Its distribution only depends on the function h. It corresponds to the family generated by one ancestor at time 0 in Proposition B.3. Therefore, by Proposition B.3, a Hawkes process with empty past ($N_- = \emptyset$) of intensity $\lambda(t, \mathcal{F}_{t-}^{N}) = g(t) + \int_0^{t-} h(t-z)N(\mathrm{d}z)$ can always be seen as the union of N_{anc} and of all the $a + N_c^a$ for $a \in N_{anc}$ where the N_c^a are i.i.d. cluster processes.

For a Hawkes process N with non empty past, N_-, this is more technical. Let N_{anc} be a Poisson process of intensity g on $\mathbb{R}_+$ and $\left(N_c^V\right)_{V\in N_{anc}}$ be a sequence of i.i.d. cluster processes associated to h. Let also

$$N_{>0} = N_{anc} \cup \left(\bigcup_{V\in N_{anc}} V + N_c^V \right). \tag{B.10}$$

As we prove below, this represents the points in N that do not depend on N_-. The points that are depending on N_- are constructed as follows independently of $N_{>0}$. Given N_-, let $\left(N_1^T\right)_{T\in N_-}$ denote a sequence of independent Poisson processes with respective intensities $\lambda_T(v) = h(v-T)\mathbb{1}_{(0,\infty)}(v)$. Then, given N_- and $\left(N_1^T\right)_{T\in N_-}$, let $\left(N_c^{T,V}\right)_{V\in N_1^T, T\in N_-}$ be a sequence of i.i.d. cluster processes associated to h. The points depending on the past N_- are given by the following formula as proved in the next Proposition:

$$N_{\leq 0} = N_- \cup \left(\bigcup_{T\in N_-} N_1^T \cup \left(\bigcup_{V\in N_1^T} V + N_c^{T,V} \right) \right). \tag{B.11}$$

Proposition B.4. *Let $N = N_{\leq 0} \cup N_{>0}$, where $N_{>0}$ and $N_{\leq 0}$ are given by* (B.10) *and* (B.11)*. Then N is a linear Hawkes process with past given by N_- and intensity on $(0,\infty)$ given by $\lambda(t, \mathcal{F}_{t-}^{N}) = g(t) + \int_{-\infty}^{t-} h(t-x)N(\mathrm{d}x)$, where g and h are as in Proposition B.3.*

Proof. Proposition B.3 yields that $N_{>0}$ has intensity

$$\lambda_{N_{>0}}(t, \mathcal{F}_{t-}^{N_{>0}}) = g(t) + \int_0^{t-} h(t-x)N_{>0}(\mathrm{d}x), \tag{B.12}$$

and that, given N_-, for any $T \in N_-$, $N_H^T = N_1^T \cup \left(\bigcup_{V\in N_1^T} V + N_c^{T,V}\right)$ has intensity

$$\lambda_{N_H^T}(t, \mathcal{F}_{t-}^{N_H^T}) = h(t-T) + \int_0^{t-} h(t-x)N_H^T(\mathrm{d}x), \tag{B.13}$$

Moreover, all these processes are independent given N_-. For any $t \geq 0$, one can note that $\mathcal{F}_t^{N_{\leq 0}} \subset \mathcal{G}_t := \mathcal{F}_0^{N_-} \vee \left(\bigvee_{T\in N_-} \mathcal{F}_t^{N_H^T}\right)$, and so $N_{\leq 0}$ has intensity

$$\lambda_{N_{\leq 0}}(t, \mathcal{G}_{t-}) = \sum_{T\in N_-} \lambda_{N_H^T}(t, \mathcal{F}_{t-}^{N_H^T}) = \int_{-\infty}^{t-} h(t-x)N_{\leq 0}(\mathrm{d}x) \tag{B.14}$$

on $(0, +\infty)$. Since this last expression is $\mathcal{F}_t^{N_{\leq 0}}$-predictable, by page 27 in [3], this is also $\lambda_{N_{\leq 0}}(t, \mathcal{F}_{t-}^{N_{\leq 0}})$. Moreover, $N_{\leq 0}$ and $N_{>0}$ are independent by construction and, for any $t \geq 0$, $\mathcal{F}_t^N \subset \mathcal{F}_t^{N_{\leq 0}} \vee \mathcal{F}_t^{N_{>0}}$. Hence, as before, N has intensity on $(0, +\infty)$ given by

$$\lambda(t, \mathcal{F}_{t-}^N) = \lambda(t, \mathcal{F}_{t-}^{N_{\leq 0}}) + \lambda(t, \mathcal{F}_{t-}^{N_{>0}}) = g(t) + \int_{-\infty}^{t-} h(t-x)N(\mathrm{d}x). \qquad \square$$

B.3.2. *A general result for linear Hawkes processes*

The following proposition is a consequence of Theorem 3.3 applied to Hawkes processes with general past N_-.

Proposition B.5. *Using the notations of Theorem 3.3, let N be a Hawkes process with past before 0 given by N_- of distribution $\mathbb{P}_0$ and with intensity on $\mathbb{R}_+$ given by*

$$\lambda(t, \mathcal{F}_{t-}^N) = \mu + \int_{-\infty}^{t-} h(t-x)N(dx),$$

where μ is a positive real number and h is a non-negative function with support in $\mathbb{R}_+$ such that $\int h < 1$. Suppose that $\mathbb{P}_0$ is such that

$$\sup_{t \geq 0} \mathbb{E}\left[\int_{-\infty}^{0} h(t-x)N_-(dx)\right] < \infty. \tag{B.15}$$

Then, the mean measure u defined in Proposition 3.2 satisfies Theorem 3.3 and moreover its integral $v(t,s) := \int_s^\infty u(t, d\sigma)$ is a solution of the system (4.14)–(4.15) where v^{in} is the survival function of $-T_0$, and where $\Phi = \Phi_{\mathbb{P}_0}^{\mu,h}$ is given by $\Phi_{\mathbb{P}_0}^{\mu,h} = \Phi_+^{\mu,h} + \Phi_{-,\mathbb{P}_0}^{\mu,h}$, with $\Phi_+^{\mu,h}$ given by (4.17) and $\Phi_{-,\mathbb{P}_0}^{\mu,h}$ given by,

$$\forall\, s,t \geq 0, \quad \Phi_{-,\mathbb{P}_0}^{\mu,h}(t,s) = \mathbb{E}\left[\int_{-\infty}^{t-} h(t-z)N_{\leq 0}(dz) \,\middle|\, N_{\leq 0}\left([t-s,t)\right) = 0\right]. \tag{B.16}$$

Moreover, (4.20) holds.

B.3.3. *Proof of the general result of Proposition B.5*

Before proving Proposition B.5, we need some technical preliminaries.

Events of the type $\{S_{t-} \geq s\}$ are equivalent to the fact that the underlying process has no point between $t-s$ and t. Therefore, for any point process N and any real numbers $t, s \geq 0$, let

$$\mathcal{E}_{t,s}(N) = \{N \cap [t-s, t) = \emptyset\}. \tag{B.17}$$

Various sets $\mathcal{E}_{t,s}(N)$ are used in the sequel and the following lemma, whose proof is obvious and therefore omitted, is applied several times to those sets.

Lemma B.6. *Let Y be some random variable and $I(Y)$ some countable set of indices depending on Y. Suppose that $(X_i)_{i\in I(Y)}$ is a sequence of random variables which are independent conditionally on Y. Let $A(Y)$ be some event depending on Y and $\forall\, j \in I(Y)$, $B_j = B_j(Y, X_j)$ be some event depending on Y and X_j. Then, for any $i \in I(Y)$, and for all sequence of measurable functions $(f_i)_{i\in I(Y)}$ such that the following quantities exist,*

$$\mathbb{E}\left[\sum_{i\in I(Y)} f_i(Y, X_i)\middle| A\#B\right] = \mathbb{E}\left[\sum_{i\in I(Y)} \mathbb{E}\left[f_i(Y, X_i)\middle| Y, B_i\right]\middle| A\#B\right],$$

where $\mathbb{E}\left[f_i(Y, X_i)\middle| Y, B_i\right] = \frac{\mathbb{E}\left[f_i(Y,X_i)\mathbb{1}_{B_i}\middle| Y\right]}{\mathbb{P}(B_i|\, Y)}$ *and* $A\#B = A(Y) \cap \left(\bigcap_{j\in I(Y)} B_j\right)$.

The following lemma is linked to Lemma 4.2.

Lemma B.7. *Let N be a linear Hawkes process with no past before time 0 (i.e. $N_- = \emptyset$) and intensity on $(0,\infty)$ given by $\lambda(t, \mathcal{F}^N_{t-}) = g(t) + \int_0^{t-} h(t-x)N(\mathrm{d}x)$, where g and h are as in Proposition B.3 and let for any $x, s \geq 0$*

$$\begin{cases} L^{g,h}_s(x) = \mathbb{E}\left[\displaystyle\int_0^x h(x-z)N(\mathrm{d}z)\middle| \mathcal{E}_{x,s}(N)\right] \\ G^{g,h}_s(x) = \mathbb{P}\left(\mathcal{E}_{x,s}(N)\right), \end{cases}$$

Then, for any $x, s \geq 0$,

$$L^{g,h}_s(x) = \int_{s\wedge x}^x \left(h\left(z\right) + L^{h,h}_s(z)\right) G^{h,h}_s(z) g(x-z)\,\mathrm{d}z, \tag{B.18}$$

and

$$\log(G^{g,h}_s(x)) = \int_0^{(x-s)\vee 0} G^{h,h}_s(x-z)g(z)\mathrm{d}z - \int_0^x g(z)\mathrm{d}z. \tag{B.19}$$

In particular, $(L^{h,h}_s, G^{h,h}_s)$ is in $L^1 \times L^\infty$ and is a solution of (4.11)-(4.12).

Proof. The statement only depends on the distribution of N. Hence, thanks to Proposition B.4, it is sufficient to consider $N = N_{anc} \cup \left(\cup_{V\in N_{anc}} V + N^V_c\right)$.

Let us show (B.18). First, let us write $L^{g,h}_s(x) = \mathbb{E}\left[\sum_{X\in N} h(x-X)\middle|\mathcal{E}_{x,s}(N)\right]$. and note that $L^{g,h}_s\left(x\right) = 0$ if $x \leq s$. The following decomposition holds

$$L^{g,h}_s(x) = \mathbb{E}\left[\sum_{V\in N_{anc}} \left(h(x-V) + \sum_{W\in N^V_c} h(x-V-W)\right)\middle|\mathcal{E}_{x,s}(N)\right].$$

According to Lemma B.6 and the following decomposition,

$$\mathcal{E}_{x,s}(N) = \mathcal{E}_{x,s}(N_{anc}) \cap \left(\bigcap_{V\in N_{anc}} \mathcal{E}_{x-V,s}(N^V_c)\right), \tag{B.20}$$

let us denote $Y = N_{anc}$, $X_V = N_c^V$ and $B_V = \mathcal{E}_{x-V,s}(N_c^V)$ for all $V \in N_{anc}$. Let us fix $V \in N_{anc}$ and compute the conditional expectation of the inner sum with respect to the filtration of N_{anc} which is

$$\mathbb{E}\left[\sum_{W\in N_c^V} h(x-V-W)\middle| Y, B_V\right] = \mathbb{E}\left[\sum_{W\in N_c} h((x-V)-W)\middle|\mathcal{E}_{x-V,s}(N_c)\right] = L_s^{h,h}(x-V), \tag{B.21}$$

since, conditionally on N_{anc}, N_c^V has the same distribution as N_c which is a linear Hawkes process with conditional intensity $\lambda(t, \mathcal{F}_{t-}^{N_c}) = h(t) + \int_0^{t-} h(t-z)N_c(\mathrm{d}z)$. Using the conditional independence of the cluster processes with respect to N_{anc}, one can apply Lemma B.6 and deduce that

$$L_s^{g,h}(x) = \mathbb{E}\left[\sum_{V\in N_{anc}} \big(h(x-V) + L_s^{h,h}(x-V)\big)\middle|\mathcal{E}_{x,s}(N)\right]$$

The following argument is inspired by Moller [28]. For every $V \in N_{anc}$, we say that V has mark 0 if V has no descendant or himself in $[x-s, x)$ and mark 1 otherwise. Let us denote N_{anc}^0 the set of points with mark 0 and $N_{anc}^1 = N_{anc} \setminus N_{anc}^0$. For any $V \in N_{anc}$, we have $\mathbb{P}\left(V \in N_{anc}^0 \middle| N_{anc}\right) = G_s^{h,h}(x-V)\mathbb{1}_{[x-s,x)^c}(V)$, and all the marks are chosen independently given N_{anc}. Hence, N_{anc}^0 and N_{anc}^1 are independent Poisson processes and the intensity of N_{anc}^0 is given by $\lambda(v) = g(v)G_s^{h,h}(x-v)\mathbb{1}_{[x-s,x)^c}(v)$. Moreover, the event $\left\{N_{anc}^1 = \emptyset\right\}$ can be identified to $\mathcal{E}_{x,s}(N)$and

$$\begin{aligned} L_s^{g,h}(x) &= \mathbb{E}\left[\sum_{V\in N_{anc}^0} \big(h(x-V) + L_s^{h,h}(x-V)\big)\middle| N_{anc}^1 = \emptyset\right] \\ &= \int_{-\infty}^{x-} \left(h\left(x-w\right) + L_s^{h,h}(x-w)\right) g(w)G_s^{h,h}(x-w)\mathbb{1}_{[x-s,x)^c}(w)\mathrm{d}w \\ &= \int_0^{(x-s)\vee 0} \left(h\left(x-w\right) + L_s^{h,h}(x-w)\right) G_s^{h,h}(x-w)g(w)\,\mathrm{d}w, \end{aligned}$$

where we used the independence between the two Poisson processes. It suffices to substitute w by $z = x - w$ in the integral to get the desired formula. Since $G_s^{h,h}$ is bounded, it is obvious that $L_s^{h,h}$ is L^1.

Then, let us show (B.19). First note that if $x < 0$, $G_s^{g,h}(x) = 1$. Next, following (B.20) one has $G_s^{g,h}(x) = \mathbb{E}\left[\mathbb{1}_{\mathcal{E}_{x,s}(N_{anc})}\prod_{X\in N_{anc}}\mathbb{1}_{\mathcal{E}_{x-X,s}(N_c^X)}\right]$. This is also

$$\begin{aligned} G_s^{g,h}(x) &= \mathbb{E}\left[\mathbb{1}_{N_{anc}\cap[x-s,x)=\emptyset}\prod_{V\in N_{anc}\cap[x-s,x)^c}\mathbb{1}_{\mathcal{E}_{x-V,s}(N_c^V)}\right], \\ &= \mathbb{E}\left[\mathbb{1}_{N_{anc}\cap[x-s,x)=\emptyset}\prod_{V\in N_{anc}\cap[x-s,x)^c} G_s^{h,h}(x-V)\right], \end{aligned}$$

by conditioning with respect to N_{anc}. Since $N_{anc} \cap [x-s,x)$ is independent of $N_{anc} \cap [x-s,x)^c$, this gives

$$G_s^{g,h}(x) = \exp(-\int_{x-s}^{x} g(z)\mathrm{d}z)\mathbb{E}\left[\exp\left(\int_{[x-s,x)^c} \log(G_s^{h,h}(x-z))N_{anc}(\mathrm{d}z)\right)\right].$$

This leads to $\log(G_s^{g,h}(x)) = -\int_{x-s}^{x} g(z)\mathrm{d}z + \int_{[x-s,x)^c}(G_s^{h,h}(x-z)-1)g(z)\mathrm{d}z$, thanks to Campbell's Theorem [23]. Then, (B.19) clearly follows from the facts that if $z > x > 0$ then $G_s^{h,h}(x-z) = 1$ and $g(z) = 0$ as soon as $z < 0$. □

Proof of Lemma 4.2 In turn, we use a Banach fixed point argument to prove that for all $s \geq 0$ there exists a unique couple $(L_s, G_s) \in L^1(\mathbb{R}_+) \times L^\infty(\mathbb{R}_+)$ solution to these equations. To do so, let us first study Equation (4.11) and define $T_{G,s} : L^\infty(\mathbb{R}_+) \to L^\infty(\mathbb{R}_+)$ by $T_{G,s}(f)(x) := \exp\left(\int_0^{(x-s)\vee 0} f(x-z)h(z)\mathrm{d}z - \int_0^x h(z)\mathrm{d}z\right)$. The right-hand side is well-defined since $h \in L^1$ and $f \in L^\infty$. Moreover we have

$$T_{G,s}(f)(x) \leq e^{\|f\|_{L^\infty}\left(\int_0^{(x-s)\vee 0} h(z)\mathrm{d}z - \int_0^x h(z)\mathrm{d}z\right)} \leq e^{(\|f\|_{L^\infty}-1)\int_0^{(x-s)\vee 0} h(z)\mathrm{d}z}.$$

This shows that $T_{G,s}$ maps the ball of radius 1 of L^∞ into itself, and more precisely into the intersection of the positive cone and the ball. We distinguish two cases:
– If $x < s$, then $T_{G,s}(f)(x) = \exp(-\int_0^x h(z)\mathrm{d}z)$ for any f, thus, the unique fixed point is given by $G_s : x \mapsto \exp(-\int_0^x h(z)\mathrm{d}z)$, which does not depend on $s > x$.
– And if $x > s$, the functional $T_{G,s}$ is a $k-$contraction in $\{f \in L^\infty(\mathbb{R}_+), \|f\|_{L^\infty} \leq 1\}$, with $k \leq \int_0^\infty h(z)\mathrm{d}z < 1$, by convexity of the exponential. More precisely, using that for all x, y, $|e^x - e^y| \leq e^{\max(x,y)}|x-y|$ we end up with, for $\|f\|, \|g\|_{L^\infty} \leq 1$,

$$\begin{aligned}\left|T_{G,s}(f)(x) - T_{G,s}(g)(x)\right| &\leq e^{-\int_0^x h(z)\mathrm{d}z} e^{\int_0^{x-s} h(z)\mathrm{d}z} \|f-g\|_{L^\infty} \int_0^{(x-s)} h(z)\mathrm{d}z \\ &\leq \|f-g\|_{L^\infty} \int_{\mathbb{R}_+} h(z)\mathrm{d}z.\end{aligned}$$

Hence there exists only one fixed point G_s that we can identify with $G_s^{h,h}$ given in Proposition B.7 and $G_s^{h,h}$ being a probability, G_s takes values in $[0,1]$.

Analogously, we define the functional $T_{L,s} : L^1(\mathbb{R}_+) \to L^1(\mathbb{R}_+)$ by $T_{L,s}(f)(x) := \int_{s\wedge x}^x \left(h(z) + f(z)\right) G_s(z)h(x-z)\,\mathrm{d}z$, and it is easy to check that $T_{L,s}$ is well-defined as well. We similarly distinguish the two cases:
– If $x < s$, then the unique fixed point is given by $L_s(x) = 0$.
– And if $x > s$, thus $T_{L,s}$ is a $k-$contraction with $k \leq \int_0^\infty h(y)dy < 1$ in $L^1((s,\infty))$ since $\|G_s\|_{L^\infty} \leq 1$:

$$\begin{aligned}\|T_{L,s}(f) - T_{L,s}(g)\|_{L^1} &= \int_s^\infty \Big| \int_s^x \big(f(z) - g(z)\big) G_s(z) h(x-z) \mathrm{d}z \Big| \mathrm{d}x \\ &\leq \|G_s\|_{L^\infty} \int_s^\infty \int_v^\infty \big|f(z) - g(z)\big| h(x-z) \mathrm{d}x \mathrm{d}z \\ &= \|G_s\|_{L^\infty} \|f-g\|_{L^1((s,\infty))} \int_0^\infty h(y) dy.\end{aligned}$$

In the same way, there exists only one fixed point $L_s = L_s^{h,h}$ given by Proposition B.7. In particular $L_s(x \leq s) \equiv 0$.

Finally, as a consequence of Equation (4.12) we find that if L_s is the unique fixed point of $T_{L,s}$, then $\|L_s\|_{L^1(\mathbb{R}_+)} \leq \frac{(\int_0^\infty h(y)\ dy)^2}{1-\int_0^\infty h(y)\ dy}$ and therefore L_s is uniformly bounded in L^1 with respect to s.

Lemma B.8. *Let N be a linear Hawkes process with past before time 0 given by N_- and intensity on $(0,\infty)$ given by $\lambda(t, \mathcal{F}_{t-}^N) = \mu + \int_{-\infty}^{t-} h(t-x)N(\mathrm{d}x)$, where μ is a positive real number and h is a non-negative function with support in $\mathbb{R}_+$, such that $||h||_{L^1} < 1$. If the distribution of N_- satisfies* (B.15) *then* $(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp})$ *is satisfied.*

Proof. For all $t > 0$, let $\overline{\lambda}(t) = \mathbb{E}\left[\lambda(t, \mathcal{F}_{t-}^N)\right]$. By Proposition B.4, $\overline{\lambda}(t) = \mathbb{E}\left[\mu + \int_0^{t-} h(t-x)N_{>0}(\mathrm{d}x)\right] + \mathbb{E}\left[\int_{-\infty}^{t-} h(t-x)N_{\leq 0}(\mathrm{d}x)\right]$ which is possibly infinite.

Let us apply Proposition B.7 with $g \equiv \mu$ and $s = 0$, the choice $s = 0$ implying that $\mathcal{E}_{t,0}(N_{>0})$ is of probability 1. Therefore

$$\mathbb{E}\left[\mu + \int_0^{t-} h(t-x)N_{>0}(\mathrm{d}x)\right] = \mu\left(1 + \int_0^t (h(x) + L_0(x))\mathrm{d}x\right),$$

where $(L_0, G_0 = 1)$ is the solution of Lemma 4.2 for $s = 0$, by identification of Proposition B.7. Hence $\mathbb{E}\left[\mu + \int_0^{t-} h(t-x)N_{>0}(\mathrm{d}x)\right] \leq \mu(1 + ||h||_{L^1} + ||L_0||_{L^1})$. On the other hand, thanks to Lemma B.9, we have

$$\mathbb{E}\left[\int_{-\infty}^{t-} h(t-x)N_{\leq 0}(\mathrm{d}x)\right] = \mathbb{E}\left[\sum_{T\in N_-}\left(h(t-T) + \int_0^t \left[h(t-x) + L_0(t-x)\right] h(x-T)\mathrm{d}x\right)\right].$$

Since all the quantities are non negative, one can exchange all the integrals and deduce that

$$\mathbb{E}\left[\int_{-\infty}^{t-} h(t-x)N_{\leq 0}(\mathrm{d}x)\right] \leq M(1 + ||h||_{L^1} + ||L_0||_{L^1}),$$

with $M = \sup_{t\geq 0} \mathbb{E}\left[\int_{-\infty}^0 h(t-x)N_-(\mathrm{d}x)\right]$ which is finite by assumption. Hence, $\overline{\lambda}(t) \leq (\mu + M)(1 + ||h||_{L^1} + ||L_0||_{L^1})$, and therefore $(\mathcal{A}_{\lambda,loc}^{\mathbb{L}^1,exp})$ is satisfied. □

Proof of Proposition B.5 First, by Proposition B.4

$$\begin{aligned}&\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\middle|\, S_{t-}\geq s\right]=\\&\quad\mu+\mathbb{E}\left[\int_0^{t-}h(t-z)N_{>0}(\mathrm{d}z)\middle|\mathcal{E}_{t,s}(N)\right]+\mathbb{E}\left[\int_{-\infty}^{t-}h(t-z)N_{\leq 0}(\mathrm{d}z)\middle|\mathcal{E}_{t,s}(N)\right]\\&=\mu+\mathbb{E}\left[\int_0^{t-}h(t-z)N_{>0}(\mathrm{d}z)\middle|\mathcal{E}_{t,s}(N_{>0})\right]+\mathbb{E}\left[\int_{-\infty}^{t-}h(t-z)N_{\leq 0}(\mathrm{d}z)\middle|\mathcal{E}_{t,s}(N_{\leq 0})\right]\end{aligned}$$

By Lemma B.7, we obtain $\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\middle|\, S_{t-}\geq s\right]=\mu+L_s^{\mu,h}(t)+\Phi^h_{-,\mathbb{P}_0}(t,s)$. Identifying by Lemma 4.2, $L_s=L_s^{h,h}$ and $G_s=G_s^{h,h}$, we obtain

$$\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\middle|\, S_{t-}\geq s\right]=\Phi_+^{\mu,h}(t,s)+\Phi^h_{-,\mathbb{P}_0}(t,s).$$

Hence $\Phi^{\mu,h}_{\mathbb{P}_0}(t,s)=\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\middle|\, S_{t-}\geq s\right]$.

Lemma B.8 ensures that the assumptions of Theorem 3.3 are fulfilled. Let u and $\rho^{\mu,h}_{\mathbb{P}_0}=\rho_{\lambda,\mathbb{P}_0}$ be defined accordingly as in Theorem 3.3. With respect to the PDE system, there are two possibilities to express $\mathbb{E}\left[\lambda(t,\mathcal{F}_{t-}^N)\mathbb{1}_{\{S_{t-}\geq s\}}\right]$. The first one involves $\rho_{\lambda,\mathbb{P}_0}$ and is $\mathbb{E}\left[\rho^{\mu,h}_{\mathbb{P}_0}(t,S_{t-})\mathbb{1}_{S_{t-}\geq s}\right]$, whereas the second one involves $\Phi^{\mu,h}_{\mathbb{P}_0}$ and is $\Phi^{\mu,h}_{\mathbb{P}_0}(t,s)\mathbb{P}\left(S_{t-}\geq s\right)$.

This leads to $\int_s^{+\infty}\rho^{\mu,h}_{\mathbb{P}_0}(t,x)u(t,\mathrm{d}x)=\Phi^{\mu,h}_{\mathbb{P}_0}(t,s)\int_s^{+\infty}u(t,\mathrm{d}x)$, since $u(t,\mathrm{d}s)$ is the distribution of S_{t-}. Let us denote $v(t,s)=\int_s^{+\infty}u(t,\mathrm{d}x)$: this relation, together with Equation (3.10) for u, immediately gives us that v satisfies Equation (4.14) with $\Phi=\Phi^{\mu,h}_{\mathbb{P}_0}$. Moreover, $\int_0^{+\infty}u(t,\mathrm{d}x)=1$, which gives us the boundary condition in (4.15).

B.3.4. *Study of the general case for $\Phi^h_{-,\mathbb{P}_0}$ in Proposition B.5*

Lemma B.9. *Let consider h a non-negative function with support in $\mathbb{R}_+$ such that $\int h<1$, N_- a point process on $\mathbb{R}_-$ with distribution $\mathbb{P}_0$ and $N_{\leq 0}$ defined by* (B.11). *If $\Phi^h_{-,\mathbb{P}_0}(t,s):=\mathbb{E}\left[\int_{-\infty}^{t-}h(t-z)N_{\leq 0}(\mathrm{d}z)\middle|\,\mathcal{E}_{t,s}(N_{\leq 0})\right]$, for all $s,t\geq 0$, then,*

$$\Phi^h_{-,\mathbb{P}_0}(t,s)=\mathbb{E}\left[\sum_{T\in N_-}\left(h(t-T)+K_s(t,T)\right)\middle|\mathcal{E}_{t,s}(N_{\leq 0})\right],\tag{B.22}$$

where $K_s(t,u)$ is given by (4.13).

Proof. Following the decomposition given in Proposition B.4, one has

$$\begin{aligned}\Phi^h_{-,\mathbb{P}_0}(t,s)=\mathbb{E}\Bigg[\sum_{T\in N_-}\Bigg(&h(t-T)\\&+\sum_{V\in N_1^T}\left(h(t-V)+\sum_{W\in N_c^{T,V}}h(t-V-W)\right)\Bigg)\Bigg|\mathcal{E}_{t,s}(N_{\leq 0})\Bigg],\end{aligned}$$

where $\mathcal{E}_{t,s}(N_{\leq 0}) = \mathcal{E}_{t,s}(N_-) \bigcap_{T' \in N_-} \left(\mathcal{E}_{t,s}(N_1^T) \bigcap_{V' \in N_1^T} \mathcal{E}_{t-V',s}(N_c^{V'}) \right)$. Let us fix $T \in N_-$, $V \in N_1^T$ and compute the conditional expectation of the inner sum with respect to N_- and N_1^T. In the same way as for (B.21) we end up with

$$\mathbb{E}\left[\sum_{W \in N_c^{T,V}} h(t-V-W) \middle| N_-, N_1^T, \mathcal{E}_{t-V,s}(N_c^{T,V}) \right] = L_s^{h,h}(t-V),$$

since, conditionally on N_- and N_1^T, $N_c^{T,V}$ has the same distribution as N_c. Using the conditional independence of the cluster processes $(N_c^{T,V})_{V \in N_1^T}$ with respect to $\left(N_-, (N_1^T)_{T \in N_-}\right)$, one can apply Lemma B.6 with $Y = \left(N_-, (N_1^T)_{T \in N_-}\right)$ and $X_{(T,V)} = N_c^{T,V}$ and deduce that

$$\Phi^h_{-,\mathbb{P}_0}(t,s) = \mathbb{E}\left[\sum_{T \in N_-} \left(h(t-T) + \sum_{V \in N_1^T} \big(h(t-V) + L_s^{h,h}(t-V) \big) \right) \middle| \mathcal{E}_{t,s}(N_{\leq 0}) \right].$$

Let us fix $T \in N_-$ and compute the conditional expectation of the inner sum with respect to N_- which is

$$\Gamma := \mathbb{E}\left[\sum_{V \in N_1^T} \big(h(t-V) + L_s^{h,h}(t-V) \big) \middle| N_-, A_{t,s}^T \right], \tag{B.23}$$

where $A_{t,s}^T = \mathcal{E}_{t,s}(N_1^T) \cap \left(\bigcap_{V' \in N_1^T} \mathcal{E}_{t-V',s}(N_c^{T,V'}) \right)$. For every $V \in N_1^T$, we say that V has mark 0 if V has no descendant or himself in $[t-s,t)$ and mark 1 otherwise. Let us denote $N_1^{T,0}$ the set of points with mark 0 and $N_1^{T,1} = N_1^T \setminus N_1^{T,0}$.

For any $V \in N_1^T$, $\mathbb{P}\left(V \in N_1^{T,0} \middle| N_1^T \right) = G_s^{h,h}(t-V) \mathbb{1}_{[t-s,t)^c}(V)$ and all the marks are chosen independently given N_1^T. Hence, $N_1^{T,0}$ and $N_1^{T,1}$ are independent Poisson processes and the intensity of $N_1^{T,0}$ is given by $\lambda(v) = h(v-T)\mathbb{1}_{[0,\infty)}(v) G_s^{h,h}(t-v)\mathbb{1}_{[t-s,t)^c}(v)$. Moreover, $A_{t,s}^T$ is the event $\left\{ N_1^{T,1} = \emptyset \right\}$, so

$$\begin{aligned}
\Gamma &= \mathbb{E}\left[\sum_{V \in N_1^{T,0}} \big(h(t-V) + L_s^{h,h}(t-V) \big) \middle| N_-, \left\{ N_1^{T,1} = \emptyset \right\} \right] \\
&= \int_{-\infty}^{t-} \left[h(t-v) + L_s^{h,h}(t-v) \right] h(v-T) \mathbb{1}_{[0,\infty)}(v) G_s^{h,h}(t-v) \mathbb{1}_{[t-s,t)^c}(v) dv \\
&= K_s(t,T).
\end{aligned}$$

Using the independence of the cluster processes, one can apply Lemma B.6 with $Y = N_-$ and $X_T = \left(N_1^T, (N_c^{T,V})_{V \in N_1^T} \right)$ and (B.22) clearly follows. □

Lemma B.10. *Under the assumptions and notations of Proposition B.5 and Lemma 4.2, the function $\Phi^h_{-,\mathbb{P}_0}$ of Proposition B.5 can be identified with* (4.18) *under* $(\mathcal{A}^1_{N_-})$ *and with* (4.19) *under* $(\mathcal{A}^2_{N_-})$ *and* (B.15) *is satisfied in those two cases.*

Proof. Using
Lemma B.9, we have $\Phi^h_{-,\mathbb{P}_0}(t,s) = \mathbb{E}\left[\sum_{T\in N_-}(h(t-T)+K_s(t,T))\middle|\mathcal{E}_{t,s}(N_{\leq 0})\right]$.
Under $\left(\mathcal{A}^1_{N_-}\right)$. On the one hand, for every $t \geq 0$,

$$\mathbb{E}\left[\int_{-\infty}^0 h(t-x)N_-(\mathrm{d}x)\right] = \mathbb{E}\left[h(t-T_0)\right] = \int_{-\infty}^0 h(t-t_0)f_0(t_0)\mathrm{d}t_0 \leq ||f_0||_{L^\infty}\int_0^\infty h(y)dy,$$

hence $\mathbb{P}_0$ satisfies (B.15). On the other hand, since N_- is reduced to one point T_0, $\Phi^h_{-,\mathbb{P}_0}(t,s) = \frac{1}{\mathbb{P}\left(\mathcal{E}_{t,s}(N_{\leq 0})\right)}\mathbb{E}\left[(h(t-T_0)+K_s(t,T_0))\mathbb{1}_{\mathcal{E}_{t,s}(N_{\leq 0})}\right]$, using the definition of the conditional expectation. First, we compute $\mathbb{P}(\mathcal{E}_{t,s}(N_{\leq 0}|T_0)$. To do so, we use the decomposition $\mathcal{E}_{t,s}(N_{\leq 0}) = \{T_0 < t-s\}\cap\mathcal{E}_{t,s}(N_1^{T_0})\cap\left(\bigcap_{V\in N_1^{T_0}}\mathcal{E}_{t-V,s}(N_c^{T_0,V})\right)$ and the fact that, conditionally on $N_1^{T_0}$, for all $V \in N_1^{T_0}$, $N_c^{T_0,V}$ has the same distribution as N_c to deduce that

$$\mathbb{E}\left[\mathbb{1}_{\mathcal{E}_{t,s}(N_{\leq 0})}\middle|T_0\right] = \mathbb{1}_{T_0<t-s}\mathbb{E}\left[\mathbb{1}_{\mathcal{E}_{t,s}(N_1^{T_0})}\middle|T_0\right]\mathbb{E}\left[\prod_{V\in N_1^{T_0}\cap[t-s,t)^c}G_s(t-V)\middle|T_0\right],$$

because the event $\mathcal{E}_{t,s}(N_1^{T_0})$ involves $N_1^{T_0}\cap[t-s,t)$ whereas the product involves $N_1^{T_0}\cap[t-s,t)^c$, both of those processes being two independent Poisson processes. Their respective intensities are $\lambda(x) = h(x-T_0)\mathbb{1}_{[(t-s)\vee 0,t)}(x)$ and $\lambda(x) = h(x-T_0)\mathbb{1}_{[0,(t-s)\vee 0)}(x)$, so we end up with

$$\begin{cases}\mathbb{E}\left[\mathbb{1}_{\mathcal{E}_{t,s}(N_1^{T_0})}\middle|T_0\right] = \exp\left(-\int_{t-s}^t h(x-T_0)\mathbb{1}_{[0,\infty)}(x)\mathrm{d}x\right)\\ \mathbb{E}\left[\prod_{V\in N_1^{T_0}\cap[t-s,t)^c}G_s(t-V)\middle|T_0\right] = \exp\left(-\int_0^{(t-s)\vee 0}\left[1-G_s(t-x)\right]h(x-T_0)\mathrm{d}x\right).\end{cases}$$

The product of these two last quantities is exactly $q(t,s,T_0)$ given by (4.13). Note that $q(t,s,T_0)$ is exactly the probability that T_0 has no descendant in $[t-s,t)$ given T_0. Hence, $\mathbb{P}\left(\mathcal{E}_{t,s}(N_{\leq 0})\right) = \int_{-\infty}^{0\wedge(t-s)}q(t,s,t_0)f_0(t_0)\mathrm{d}t_0$ and (4.18) clearly follows.
Under $\left(\mathcal{A}^2_{N_-}\right)$. On the one hand, for any $t \geq 0$,

$$\mathbb{E}\left[\int_{-\infty}^0 h(t-x)N_-(\mathrm{d}x)\right] = \mathbb{E}\left[\int_{-\infty}^0 h(t-x)\alpha\mathrm{d}x\right] \leq \alpha\int_0^\infty h(y)dy,$$

hence $\mathbb{P}_0$ satisfies (B.15). On the other hand, since we are dealing with a Poisson process, we can use the same argumentation of marked Poisson processes as in the proof of Lemma B.7. For every $T \in N_-$, we say that T has mark 0 if T has no descendant or himself in $[t-s,t)$ and mark 1 otherwise. Let us denote N_-^0 the set of points with mark 0 and $N_-^1 = N_- \setminus N_-^0$. For any $T \in N_-$, we have

$$\mathbb{P}\left(T \in N_-^0\middle|N_-\right) = q(t,s,T)\mathbb{1}_{[t-s,t)^c}(T),$$

and all the marks are chosen independently given N_-. Hence, N_-^0 and N_-^1 are independent Poisson processes and the intensity of N_-^0 is given by

$$\lambda(z) = \alpha \mathbb{1}_{z\leq 0}\, q(t,s,z)\mathbb{1}_{[t-s,t)^c}(z)$$

Moreover, $\mathcal{E}_{t,s}(N_{\leq 0}) = \left\{N_-^1 = \emptyset\right\}$. Hence,

$$\Phi^h_{-,\mathbb{P}_0}(t,s) = \mathbb{E}\left[\sum_{T\in N_-^0} (h(t-T) + K_s(t,T))\middle| N_-^1 = \emptyset\right]$$

which gives (4.19) thanks to the independence of N_-^0 and N_-^1.

B.3.5. *Proof of Propositions 4.3 and 4.4*

Since we already proved Proposition B.5 and Lemma B.10, to obtain Proposition 4.3 it only remains to prove that $\Phi^{\mu,h}_{\mathbb{P}_0} \in L^\infty(\mathbb{R}_+^2)$, to ensure uniqueness of the solution by Remark 4.1. To do so, it is easy to see that the assumption $h \in L^\infty(\mathbb{R}_+)$ combined with Lemma 4.2 giving that $G_s \in [0,1]$ and $L_s \in L^1(\mathbb{R}_+)$ ensures that $\Phi^{\mu,h}_+, q$ and K_s are in $L^\infty(\mathbb{R}_+)$. In turn, this implies that $\Phi^h_{-,\mathbb{P}_0}$ in both (4.18) and (4.19) is in $L^\infty(\mathbb{R}_+)$, which concludes the proof of Proposition 4.3.

Proof of Proposition 4.4 The method of characteristics leads us to rewrite the solution v of (4.14)–(4.15) by defining $f^{in} \equiv v^{in}$ on $\mathbb{R}_+$, $f^{in} \equiv 1$ on $\mathbb{R}_-$ such that

$$v(t,s) = \begin{cases} f^{in}(s-t)e^{-\int_{(t-s)\vee 0}^{t} \Phi(y,s-t+y)\,dy}, \text{ when } s \geq t \\ f^{in}(s-t)e^{-\int_{(s-t)\vee 0}^{s} \Phi(y+t-s,y)\,dy}, \text{ when } t \geq s. \end{cases} \tag{B.24}$$

Let $\mathbb{P}_0^M$ be the distribution of the past given by $\left(\mathcal{A}^1_{N_-}\right)$ and $T_0 \sim \mathcal{U}([-M-1,-M])$. By Proposition 4.3, let v_M be the solution of System (4.14)–(4.15) with $\Phi = \Phi^{\mu,h}_{\mathbb{P}_0^M}$ and $v^{in} = v_M^{in}$, (i.e. the survival function of a uniform variable on $[-M-1,-M]$). Let also v_M^∞ be the solution of System (4.14)–(4.15) with $\Phi = \Phi^{\mu,h}_{\mathbb{P}_0^M}$ and $v^{in} \equiv 1$, and v_∞ the solution of (4.21)-(4.22). Then

$$\|v_M - v^\infty\|_{L^\infty((0,T)\times(0,S))} \leq \|v_M - v_M^\infty\|_{L^\infty((0,T)\times(0,S))} + \|v_M^\infty - v^\infty\|_{L^\infty((0,T)\times(0,S))}.$$

By definition of v_M^{in}, it is clear that $v_M^{in}(s) = 1$ for $s \leq M$, so that Formula (B.24) implies that $v_M(t,s) = v_M^\infty(t,s)$ as soon as $s-t \leq M$ and so $\|v_M - v_M^\infty\|_{L^\infty((0,T)\times(0,S))} = 0$ as soon as $M \geq S$.

To evaluate the distance $\|v_M^\infty - v^\infty\|_{L^\infty((0,T)\times(0,S))}$, it remains to prove that $e^{-\int_0^t \Phi^h_{-,\mathbb{P}_0^M}(y,s-t+y)\,dy} \to 1$ uniformly on $(0,T)\times(0,S)$ for any $T>0$, $S>0$. For this, it suffices to prove that $\Phi^h_{-,\mathbb{P}_0^M}(t,s) \to 0$ uniformly on $(0,T)\times(0,S)$. Since q given by (4.13) takes values in $[\exp(-2||h||_{L^1}),1]$, (4.18) implies

$$\Phi^h_{-,\mathbb{P}_0^M}(t,s) \leq \frac{\int_{-\infty}^{0\wedge(t-s)} (h(t-t_0) + K_s(t,t_0))\,\mathbb{1}_{[-M-1,-M]}(t_0)\mathrm{d}t_0}{\int_{-\infty}^{0\wedge(t-s)} \exp(-2||h||_{L^1})\mathbb{1}_{[-M-1,-M]}(t_0)\mathrm{d}t_0}.$$

Since $||G_s||_{L^\infty} \leq 1$, L_s and h are non-negative, it is clear that

$$K_s(t,t_0) \leq \int_0^{+\infty} \left[h(t-x) + L_s(t-x)\right] h(x-t_0)\mathrm{d}x,$$

and so

$$\begin{aligned}\int_{-M-1}^{-M} K_s(t,t_0)\mathrm{d}t_0 &\leq \int_0^{+\infty} \left[h(t-x) + L_s(t-x)\right] \left(\int_{-M-1}^{-M} h(x-t_0)\mathrm{d}t_0\right) \mathrm{d}x \\ &\leq \int_M^{\infty} h(y)dy \int_0^{+\infty} \left[h(t-x) + L_s(t-x)\right] \mathrm{d}x \\ &\leq \int_M^{\infty} h(y)dy \left[||h||_{L^1} + ||L_s||_{L^1}\right].\end{aligned}$$

Hence, for M large enough $\Phi^h_{-,\mathbb{P}_0^M}(t,s) \leq \frac{\int_M^\infty h(y)dy\left[||h||_{L^1}+||L_s||_{L^1}\right]}{\exp(-2||h||_{L^1})} \to 0$, uniformly in (t,s) since L_s is uniformly bounded in L^1, which concludes the proof.

B.4. *Thinning*

The demonstration of Ogata's thinning algorithm uses a generalization of point processes, namely the marked point processes. However, only the basic properties of simple and marked point processes are needed (see [3] for a good overview of point processes theory). Here $(\mathcal{F}_t)_{t>0}$ denotes a general filtration such that $\mathcal{F}_t^N \subset \mathcal{F}_t$ for all $t > 0$, and not necessarily the natural one, i.e. $(\mathcal{F}_t^N)_{t>0}$.

Theorem B.11. *Let Π be a $(\mathcal{F}_t)$-Poisson process with intensity 1 on $\mathbb{R}_+^2$. Let $\lambda(t,\mathcal{F}_{t-})$ be a non-negative $(\mathcal{F}_t)$-predictable process which is L^1_{loc} a.s. and define the point process N by $N(C) = \int_{C\times\mathbb{R}_+} \mathbf{1}_{[0,\lambda(t,\mathcal{F}_{t-})]}(z)\,\Pi(\mathrm{d}t\times\mathrm{d}z)$, for all $C \in \mathcal{B}(\mathbb{R}_+)$. Then N admits $\lambda(t,\mathcal{F}_{t-})$ as a $(\mathcal{F}_t)$-predictable intensity. Moreover, if λ is in fact $\left(\mathcal{F}_t^N\right)$-predictable, i.e. $\lambda(t,\mathcal{F}_{t-}) = \lambda(t,\mathcal{F}_{t-}^N)$, then N admits $\lambda(t,\mathcal{F}_{t-}^N)$ as a $\left(\mathcal{F}_t^N\right)$-predictable intensity.*

Proof. The goal is to apply the martingale characterization of the intensity (Chapter II, Theorem 9 in [3]). We cannot consider Π as a point process on $\mathbb{R}_+$ marked in $\mathbb{R}_+$ (in particular, the point with the smallest abscissa cannot be defined). However, for every $k \in \mathbb{N}$, we can define $\Pi^{(k)}$, the restriction of Π to the points with ordinate smaller than k, by $\Pi^{(k)}(C) = \int_C \Pi(\mathrm{d}t\times\mathrm{d}z)$ for all $C \in \mathcal{B}(\mathbb{R}_+\times[0,k])$. Then $\Pi^{(k)}$ can be seen as a point process on $\mathbb{R}_+$ marked in $E_k := [0,k]$ with intensity kernel $1.\mathrm{d}z$ with respect to $(\mathcal{F}_t)$. In the same way, we define $N^{(k)}$ by

$$N^{(k)}(C) = \int_{C\times\mathbb{R}_+} \mathbf{1}_{z\in[0,\lambda(t,\mathcal{F}_{t-})]}\,\Pi^{(k)}(\mathrm{d}t\times\mathrm{d}z) \quad \text{for all} \quad C \in \mathcal{B}(\mathbb{R}_+).$$

Let $\mathcal{P}(\mathcal{F}_t)$ be the predictable σ-algebra (see page 8 of [3]).

Let us denote $\mathcal{E}_k = \mathcal{B}([0,k])$ and $\tilde{\mathcal{P}}_k(\mathcal{F}_t) = \mathcal{P}(\mathcal{F}_t)\otimes\mathcal{E}_k$ the associated marked predictable σ-algebra.

For any fixed z in E, $\{(u,\omega) \in \mathbb{R}_+ \times \Omega \text{ such that } \lambda(u, \mathcal{F}_{u-})(\omega) \geq z\} \in \mathcal{P}(\mathcal{F}_t)$ since λ is predictable. If $\Gamma_k = \{(u,\omega,z) \in \mathbb{R}_+ \times \Omega \times E_k,\ \lambda(u,\mathcal{F}_{u-})(\omega) \geq z\}$, then

$$\Gamma_k = \bigcap_{n\in\mathbb{N}^*} \bigcup_{q\in\mathbb{Q}_+} \{(u,\omega) \in \mathbb{R}_+ \times \Omega,\ \lambda(u,\mathcal{F}_{u-})(\omega) \geq q\} \times \left(\left[0, q+\frac{1}{n}\right] \cap E_k\right).$$

So, $\Gamma_k \in \tilde{\mathcal{P}}_k(\mathcal{F}_t)$ and $\mathbf{1}_{z\in[0,\lambda(u,\mathcal{F}_{u-})]\cap E_k}$ is $\tilde{\mathcal{P}}_k(\mathcal{F}_t)$-measurable. Hence, one can apply the Integration Theorem (Chapter VIII, Corollary 4 in [3]). So,

$$(X_t)_{t\geq 0} := \left(\int_0^t \int_{E_k} \mathbf{1}_{z\in[0,\lambda(u,\mathcal{F}_{u-})]} \bar{M}^{(k)}(du \times \mathrm{d}z)\right)_{t\geq 0} \text{ is a } (\mathcal{F}_t)\text{-local martingale}$$

where $\bar{M}^{(k)}(du \times \mathrm{d}z) = \Pi^{(k)}(du \times \mathrm{d}z) - \mathrm{d}z\mathrm{d}u$. In fact,

$$X_t = N_t^{(k)} - \int_0^t \min(\lambda(u,\mathcal{F}_{u-}), k)\,\mathrm{d}u.$$

Yet, $N_t^{(k)}$ (respectively $\int_0^t \min(\lambda(u,\mathcal{F}_{u-}),k)\,\mathrm{d}u$) is non-decreasingly converging towards N_t (resp. $\int_0^t \lambda(u,\mathcal{F}_{u-})\mathrm{d}u$). Both of the limits are finite a.s. thanks to the local integrability of the intensity (see page 27 of [3]). Thanks to monotone convergence we deduce that $\left(N_t - \int_0^t \lambda(u,\mathcal{F}_{u-})\mathrm{d}u\right)_{t\geq 0}$ is a $(\mathcal{F}_t)$-local martingale. Then, thanks to the martingale characterization of the intensity, N_t admits $\lambda(t,\mathcal{F}_{t-})$ as an $(\mathcal{F}_t)$-intensity. The last point of the Theorem is a reduction of the filtration. Since $\lambda(t,\mathcal{F}_{t-}) = \lambda(t,\mathcal{F}^N_{t-})$, it is a fortiori $\left(\mathcal{F}^N_t\right)$-progressive and the desired result follows (see page 27 in [3]). □

This final result can be found in [4].

Proposition B.12 (Inversion Theorem).

Let $N = \{T_n\}_{n>0}$ be a non explosive point process on $\mathbb{R}_+$ with $\left(\mathcal{F}^N_t\right)$-predictable intensity $\lambda_t = \lambda(t,\mathcal{F}^N_{t-})$. Let $\{U_n\}_{n>0}$ be a sequence of i.i.d. random variables with uniform distribution on $[0,1]$. Moreover, suppose that they are independent of $\mathcal{F}^N_\infty$. Denote $\mathcal{G}_t = \sigma(U_n, T_n \leq t)$. Let $\hat{N}$ be an homogeneous Poisson process with intensity 1 on $\mathbb{R}^2_+$ independent of $\mathcal{F}_\infty \vee \mathcal{G}_\infty$. Define a point process $\bar{N}$ on $\mathbb{R}^2_+$ by

$$\bar{N}((a,b]\times A) = \sum_{n>0} \mathbb{1}_{(a,b]}(T_n)\,\mathbb{1}_A\left(U_n\lambda(T_n,\mathcal{F}^N_{T_n-})\right) + \int_{(a,b]}\int_{A-[0,\lambda(t,\mathcal{F}^N_{t-})]} \hat{N}(\mathrm{d}t\times\mathrm{d}z)$$

for every $0 \leq a < b$ and $A \subset \mathbb{R}_+$.

Then, $\bar{N}$ is an homogeneous Poisson process on $\mathbb{R}^2_+$ with intensity 1 with respect to the filtration $(\mathcal{H}_t)_{t\geq 0} = \left(\mathcal{F}_t \vee \mathcal{G}_t \vee \mathcal{F}^{\hat{N}}_t\right)_{t\geq 0}$.

References

1. M. Bossy, N. Champagnat, et al. Markov processes and parabolic partial differential equations. *Encyclopedia of Quantitative Finance*, pages 1142–1159, 2010.
2. O. Boxma, D. Perry, W. Stadje, and S. Zacks. A markovian growth-collapse model. *Advances in applied probability*, pages 221–243, 2006.

3. P. Brémaud. *Point processes and queues*. Springer-Verlag, New York, 1981. Martingale dynamics, Springer Series in Statistics.
4. P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3):1563–1588, 1996.
5. D.R. Brillinger, H.L. Bryant, and J.P. Segundo. Identification of synaptic interactions. *Biol. Cybernetics*, 22:213–228, 1976.
6. E.N. Brown, R. Barbieri, V. Ventura, R.E. Kass, and L.M. Frank. The time rescaling theorem and its application to neural spike train analysis. *Neural Computation*, 14(2):325–346, 2002.
7. N. Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 8:183–208, 2000.
8. N. Brunel and V. Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation*, 11:1621–1671, 1999.
9. M. J Cáceres, J. A Carrillo, and Benoît Perthame. Analysis of nonlinear noisy integrate&fire neuron models: blow-up and steady states. *The Journal of Mathematical Neuroscience*, 1(1):1–33, 2011.
10. M. J. Cáceres, J. A. Carrillo, and L. Tao. A numerical solver for a nonlinear fokker-planck equation representation of neuronal network dynamics. *J. Comp. Phys.*, 230:1084–1099, 2011.
11. M. J Cáceres and B. Perthame. Beyond blow-up in excitatory integrate and fire neuronal networks: refractory period and spontaneous activity. *Journal of theoretical biology*, 350:81–89, 2014.
12. José A. Cañizo, José A. Carrillo, and Sílvia Cuadrado. Measure solutions for some models in population dynamics. *Acta Appl. Math.*, 123:141–156, 2013.
13. E.S. Chornoboy, L.P. Schramm, and A.F. Karr. Maximum likelihood identification of neural point process systems. *Biol. Cybernetics*, 59:265–275, 1988.
14. B. Cloez. Limit theorems for some branching measure-valued processes. *arXiv:1106.0660v2*, 2012.
15. A. Compte, N. Brunel, P. S. Goldman-Rakic, and X.-J. Wang. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex 10*, 10:910–923, 2000.
16. D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*, volume 2. Springer, 1988.
17. F. Delarue, J. Inglis, S. Rubenthaler, and E Tanré. Global solvability of a networked integrate-and-fire model of McKean-Vlasov type. *Annals of Applied Probability*, 2012. to appear.
18. S. Ditlevsen and A. Samson. Estimation in the partially observed stochastic morris-lecar neuronal model with particle filter and stochastic approximation methods. *Annals of Applied Statistics*, to appear.
19. M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, in press, 2014.
20. A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.
21. P. Jahn, R. W. Berg, J. Hounsgaard, and S. Ditlevsen. Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *J. Comput. Neurosci.*, 31:563–579, 2011.
22. P. Jahn, R.W. Berg, J. Hounsgaard, and S. Ditlevsen. Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *Journal of Computational Neuroscience*, 31:563–579, 2011.
23. J.F.C. Kingman. *Poisson processes*. Oxford Science Publications, 1993.

24. C.D. Lai. An example of Wold's Point Processes with Markov-Dependent Intervals. *Journal of Applied Probability*, 15(4):748–758, 1978.
25. P. A. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.
26. M. Mattia and P. Del Giudice. Population dynamics of interacting spiking neurons. *Phys. Rev. E*, 66:051917, 2002.
27. P. Mattila. *Geometry of sets and measures in Euclidean spaces: fractals and rectifiability*. Number 44 in Cambridge studies in advanced mathematics. Cambridge University Press, 1999.
28. J. Møller and J. G. Rasmussen. Perfect simulation of Hawkes processes. *Advances in Applied Probability*, pages 629–646, 2005.
29. Y. Ogata. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.
30. A Omurtag, Knight B. W., and L. Sirovich. On the simulation of large populations of neurons. *J. Comp. Neurosci.*, 8:51–63, 2000.
31. K. Pakdaman, B. Perthame, and D. Salort. Dynamics of a structured neuron population. *Nonlinearity*, 23(1):55–75, 2010.
32. 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.
33. 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*, 4(14):1–26, 2014.
34. B. Perthame. *Transport equations in biology*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2007.
35. J. Pham, K. Pakdaman, J. Champagnat, and J.-F. Vibert. Activity in sparsely connected excitatory neural networks: effect of connectivity. *Neural Networks*, 11(3):415–434, 1998.
36. J.W. Pillow, J. Shlens, L. Paninski, A. Sher, A.M. Litke, E.J. Chichilnisky, and E.P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454:995–999, 2008.
37. G. Pipa, S. Grün, and C. van Vreeswijk. Impact of spike train autostructure on probability distribution of joint spike events. *Neural Computation*, 25:1123–1163, 2013.
38. C. Pouzat and A. Chaffiol. Automatic spike train analysis and report generation. an implementation with R, R2HTML and STAR. *Journal of Neuroscience Methods*, pages 119–144, 2009.
39. A. Renart, N. Brunel, and X.-J. Wang. Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks. In Jianfeng Feng, editor, *Computational Neuroscience: A comprehensive approach*. Chapman & Hall/CRC Mathematical Biology and Medicine Series, 2004.
40. 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*, 4(3), 2014.
41. P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Inference of functional connectivity in neurosciences via Hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*, 2013, Austin Texas.
42. L. Sirovich, A Omurtag, and K. Lubliner. Dynamics of neural populations: Stability and synchrony. *Network: Computation in Neural Systems*, 17:3–29, 2006.
43. V. Ventura, R. Carta, R.E. Kass, S.N. Gettner, and C.R. Olson. Statistical analysis of temporal evolution in single-neuron firing rates. *Biostatistics*, 3(1):1–20, 2002.