Stability of Age Dependent Hawkes Processes

Ph.D. Thesis by

Mads Bonde Raad

raad@math.ku.dk

Department of Mathematical Sciences
University of Copenhagen

Submitted August 2019

Mads Bonde Raad
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5, 2100 Copenhagen
Denmark

Supervisors:	Prof. Susanne Ditlevsen University of Copenhagen, Denmark
	Prof. Eva Löcherbach Université Paris 1, France
Assessment Committee:	Prof. Thomas Mikosch (chair) University of Copenhagen, Denmark
	Prof. Carl Graham École Polytechnique, France
	Prof. Jesper Møller Aalborg university, Denmark

ISBN: 978-87-7078-894-6

Contents

Chapter 1

Preface

1.1 Abstract

The primary subject of this dissertation is Hawkes processes with a focus on stability and with neuroscience as a primary application. The Hawkes process has received much attention in the last decade for modeling events that exhibit self excitation - or inhibition. There are many examples of phenomena of interest which exhibit such behavior, including finance that propagate through a market giving rise to volatility clustering observations [2], interactions on social media [42], and pattern dependencies in DNA [39].

Hawkes processes are point processes where the intensity function is stochastic and allowed to depend on the past history, introducing memory in the temporal evolution of the stochastic process. We work with an extension to the ordinary Hawkes model called an age dependent Hawkes process, where the intensity of a unit may also depend on the time since its last jump.

The classical results concerning stability for Hawkes processes found in [5] assume bounds on the strength of the functional connectivity between units in the system. However, by taking inhibition into account, we will be able to produce new stability results for Hawkes processes. We shall also discuss stability for Hawkes processes from a regeneration point of view. We will give a constructive proof of a random time such that the incremented Hawkes process is a Hawkes process in itself independent of the past. Standard Markov chain techniques allow us to prove various asymptotic results for the Hawkes process. Finally, we end by studying the mean-field limit of age dependent Hawkes processes.

1.2 Dansk Resumé

Denne afhandling beskæftiger sig med Hawkes-processer med særligt fokus på stabilitet. Det primært tiltænkte anvendelsesområde for afhandlingens resultater er neuroscience. Det sidste årti har Hawkes-processer fået meget videnskabelig opmærksomhed som model til at beskrive fænomener, hvori der indgår selv-excitering eller -inhibering. Finansielle kriser [2], interaktion på sociale medier [42] og DNA mønstre [39] er alle eksempler på fænomener med en sådan adfærd.

Hawkes-processer er punktprocesser, hvor intensiteten er stokastisk og afhænger af fortiden. I denne afhandling arbejder vi med en udviddelse af den ordinære Hawkes-proces, hvor intensiteten også må afhænge af tiden siden sidste spring. Denne udvidede klasse af processer kalder vi Alders-afhængige Hawkes-processer.

De klassiske resultater om stabilitet af Hawkes-processer, først beskrevet i [5], antager, at den exiterende eller inhiberende styrke på tværs af enheder i netværket er begrænset. Ved at benytte inhiberingen i systemet er det muligt for os at formulere nye resultater for stabilitet af Hawkes-processer.

Deruover diskuteres regeneration af Hawkes-processer. Vi konstruerer en stokastisk tid, således at Hawkes-processen i tiden derefter er uafhængig af fortiden. Dette muliggør benyttelse af standard Markov-kæde-teknikker til at vise asymptotiske resultater for Hawkes-processer. Endeligt præsenterer vi mean-field-resultater for Alders-afhængige Hawkes-processer

1.3 Acknowledgements

I want to express my deep gratitude to my principal supervisor Susanne Ditlevsen for granting me the chance to become a Ph.D. student under her supervision, and I want to thank her for the supervision and guidance through these three years. I also want to thank my co-supervisor Eva Löcherbach, for the great supervision and for her tireless efforts in reading, correcting and improving my work. Furthermore, I thank her for the opportunity of working with her in Paris for a year. Most of my research is based on probability theory and stochastic processes. My knowledge of these topics has been founded by teachings and materials from UCPH. I learned Markov process theory from Anders Rønn Nielsen, stochastic process theory from Alexander Sokol and Measure theory from Ernst Hansen. I thank them and my other teachers throughout my studies. During my eight years studying and researching math I have been grateful for my friend and now colleague Martin, who has always been available for coffee and math discussions about Null-set problems. I've had the pleasure of sharing office with many great colleagues. I small-talk more than each of them, maybe more than 'em all combined. I thank all of them for enduring that, and for the fun and interesting time at the office. My deep and sincere gratitude goes to Anna Laksafoss for helping me typing this thesis, and making it readable. I could not have completed this dissertation without support from my family and my special one. They have been very supportive and shown me much patience. I thank them, and ask for forgiveness for the many times that I have paid attention to a math thought instead of the reality. I have two birds of the species Agapornis Personata named Hanne and Eddie. They have not been supportive, nor have they have they shown any patience.
I thank them anyways.

1.4 A Short Introduction

The age dependent Hawkes process (ADHP) is a class of processes including the well known nonlinear Hawkes process. A Hawkes process Z is defined by its conditional *intensity*. This object may be described in a less formal way by
$\lambda_t = P\left(Z \text{ jumps in } [t, t+dt) \mid \mathcal{F}_t\right)$. The defining property for a Hawkes process is that its intensity is a function of the aggregating integral $\int_0^{t-} h\left(t-s\right) dZ_s$ for some fixed weight function h. The extension which defines the ADHP is that the intensity is allowed to also depend on the *age* of Z, which is the process measuring the time passed since the last jump of Z. This makes it possible to directly model individual refractory periods in a system, such as neural networks on an individual level. The Hawkes process may be extended to an N-dimensional version in a natural way: The i'th unit has a conditional intensity that depends on a sum of aggregating integrals $\sum_{j=1}^{N} \int_0^t h_{ij}\left(t-s\right) dZ_s^j$. Often it is assumed that the relation between the integral and the intensity of the i'th unit is L_i-Lipschitz. In this case, a classic result from [5] states that a one-dimensional Hawkes process is stable if $L\left\|h\right\|_{\mathcal{L}^1} < 1$. A similar result exists for the multivariate process. An introduction to ADHPs are given in chapter 2, along with natural prerequisites such as the underlying measure theory.

The first result, presented in chapter 3, treats a specific submodel of ordinary multivariate Hawkes processes. The aim is to show that for Hawkes processes where inhibition is present, stability is not necessarily due to limited connectivity strength. Balance between exciting and inhibiting effects can also induce stability in the system. In the described model, the units in the network are divided into two homogeneous populations. One population is exciting the system, while the other population acts inhibiting on the system. Moreover, the specific choice of weight functions makes the memory processes Markovian, which allows us to study stability in the framework of Harris recurrence and invariant measures from Markov process theory. The main result is proven using Foster-Lyapunov theory, and gives a stability criteria which does not imply, nor is implied by, the classic stability results

The next result, presented in chapter 4, is about ADHPs with a refractory period. That is, we impose a strong bound on the intensity in a fixed window after each jump. Such behavior is seen in neurons, where the biophysical process of inducing an electrical impulse (a so-called "spike") forbids it from repeating such an impulse in the following $\approx$ 2ms. We are able to obtain classic stability results without imposing a sub-criticality bound on h. Indeed, there is a stationary and mixing version of the Hawkes process, and other Hawkes processes with regular starting conditions converge in variation towards it. The proof relies on constructing a random measure defined on the entire real line with the same dynamics as a Hawkes process. This has been done by [5] in the case where λ is bounded using Picard iteration. However, when including the age variables, it is no longer clear that the Picard-iterates exist. Moreover, a bound on the Picard-iteration is needed in order to obtain convergence. We use growth-bounds induced by the refractory period to produce a stationary process dominating the Picard-iterates.

In chapter 5 we study stability of Hawkes processes from another perspective, namely using regeneration. We study regeneration properties under both the classic stability criteria and the ones in chapter 4. We give a constructive proof of a regeneration time ρ for a Hawkes process, satisfying that the increment process $Z_{\rho+}$ is independent of the past. This work may be seen as an extension of the work done in [10] which treats the case where the weight function h is of bounded support. However, the problem changes qualitatively when h is of unbounded support since the aggregating integral $\int_0^t h\,(t-s)\,dZ_s$ may never regenerate. Instead, we utilize the driving Poisson random measure of the Hawkes process, which enjoys temporal and spatial independence. Using the Poisson measure, we are able to construct a scheme similar in spirit to Nummelin splitting constructions. Said in a colloquial way, we attempt to restart the Hawkes process at random (stopping) times. At each attempt, the independence properties of the Poisson measure allow us to flip a coin, deciding if the attempt is successful. The constructed regeneration time turns out to be a stopping time with respect to a filtration that is strictly larger than the filtration induced by Z. We also provide criteria for the existence of moments for the regeneration time. Finally, we use this to prove results about the Hawkes process, such as various CLTs.

Finally, in chapter 6, we discuss mean-field limits in a multipopulation framework, similar to what has been done in [13]. We consider a large Hawkes network and aim to take the $N \to \infty$ limit. We assume that there is a fixed number of populations, and that all units within a population are similar. We prove standard results such as propagation of chaos and asymptotic independence of a fixed finite subset of neurons.

1.5 Thesis Structure

The dissertation consists of the following:

- Chapter 2, section 1 gives an introduction to probability theory on measures, and is based on appendix material from [14] and [11]. Section 2-6 gives an introduction to the Hawkes process.
- Chapter 3 presents a class of Hawkes processes for which we present stability results. The material is based on ongoing work by Löcherbach and Raad.
- Chapter 4 is essentially the first half of [14] by Raad et al. concerning stability of Age Dependent Hawkes processes. The material is submitted for peer-review[1].
- Chapter 5 is essentially the article [38] by Raad about regeneration for Hawkes processes. The material is submitted for peer-review.
- Chapter 6 is essentially the second half of [14] by Raad et al. treating mean-field limits for Hawkes processes. The material is submitted for peer-review[1].
- Chapter 7 is an appendix based on the appendices of the above mentioned articles.

[1]Update November 2019: The material is to appear in the journal AIHP (B)

Chapter 2

Notation and Prerequisities

We denote the set of real numbers as $\mathbb{R}$ and the subset of strictly positive real numbers as $\mathbb{R}_+$. The set of strictly positive integers are denoted $\mathbb{N}$ and we define $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $\overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$. If $v = (v_1, \ldots, v_d)$ is a $d-$dimensional Euclidian vector for some $d \in \mathbb{N}$, then $|v| = \sum_{i=1}^{d} |v_i|$ denotes the 1-distance. We introduce a background probability space $(\Omega, \mathcal{F}, P)$ and assume that this measure space is completed. If $(Y, \mathcal{Y})$ is a measurable space then a measurable map $X : \Omega \to Y$ is called a *random variable* taking values in Y. For a closed interval I unbounded to the right, we denote a family of increasing σ-algebras $(\mathcal{F}_t)_{i \in I}$ as a *filtration* on the probabilty space. We shall always assume that a filtration is satisfying the usual hypothesis meaning that for all $t \in I$ $P(A) = 0$ implies $A \in \mathcal{F}_t$ and $\bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$. Whenever a new filtration is defined, it is implicitly assumed that we extend it to satisfy the usual conditions. A random variable X taking values in $\mathbb{R}$ is said to have q'th moment for some $q \geq 0$ if $\mathbb{E}|X^q| < \infty$, and it is said to have exponential moment if $\mathbb{E}\exp(cX) < \infty$ for some $c > 0$. Likewise, we say that a function $f : \mathbb{R} \to \mathbb{R}$ have q'th moment, respectively exponential moment, if $\int x^q f(x)\, dx < \infty$, respectively $\int \exp(cx) f(x)\, dx < \infty$. We recall the basic Stieltjes integration notation. Let $g : \mathbb{R} \to \mathbb{R}$ be a càdlàg function. The variation of g on a bounded interval I is given by

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

where $\mathcal{I}$ denotes the system of all finite sets of increasing indices $(x_i) \subset I$. The function g is said to be of finite variation, if the variation is finite on all bounded intervals. For such g there exists two singular $\sigma-$finite measures μ_g^+, μ_g^- s.t. $\mu_g := \mu_g^+ - \mu_g^-$ which satisfies $\mu_g (a,b] = g(b) - g(a)$. The variation measure $|dg| := \mu_g^+ + \mu_g^-$ satisfies $|dg| (a,b] = V_g (a,b]$. If $f : \mathbb{R} \to \mathbb{R}$ is a measurable function such that $\int |f| |d\mu_g| < \infty$ then we define the *Lebesgue-Stieltjes integral* as

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

see e.g. [22] or [26]. If ν is a measure on $\mathbb{R}^2$ we shall abbreviate the integral over a semi-closed box

$$\int \mathbb{1}\{(a,b]\}(x)\, \mathbb{1}\{(c,d]\}(y)\, f(x,y)\, d\nu(x,y),$$

as either

$$\int_a^b \int_c^d f(x,y)\, d\nu(x,y) \quad \text{or} \quad \int_a^b \int_c^d f(x,y)\ \nu(dx,dy).$$

2.1 Probability Theory on Measures

2.1.1 Measure Theory Spaces of Measures

The purpose of this section is to provide a brief overview of measure theory on the space M_E of boundedly finite measures defined on a c.s.m.s (Complete Separable Metric Space). Whenever a c.s.m.s. space (E,d) is introduced, it is always assumed to be equipped with the topology and borel-algebra induced by its metric, unless otherwise written.

First, assume that (E,d) is merely a metric space. Let M_E^0 be the space of *totally finite measures* on E, i.e. measures μ satisfying $\mu(E) < \infty$. Let us equip M^0 with a metric d^0. For a set $A \subset E$ let A^ε be the superset of A in E defined by all points with $< \epsilon$ d-distance to a point in A. Define the map $d^0 : M_E^0 \times M_E^0 \to [0,\infty)$ as

$$d^0(\mu,\nu) = \inf\{\varepsilon > 0\, \mu(A) \le \nu(A^\varepsilon) + \varepsilon \ and\ \nu(A) \le \mu(A^\varepsilon + \varepsilon)\}. \quad (2.1)$$

It may be verified that d^0 is indeed a distance function, and it is characterized by weak convergence:

$$\mu_n \overset{d^0}{\to} \mu \Leftrightarrow \forall f \in C_b(E) : \int f d\mu_n \to \int f d\mu \tag{2.2}$$

where $C_b(E)$ denotes the set of continuous bounded $E \to \mathbb{R}$ functions, see [11] Prop. A2.5II. Next, A measure ν on E is said to be *boundedly finite* if $\mathrm{Diam}(A) < \infty$ implies that $\nu(A) < \infty$ for $A \in \mathcal{B}_E$. Let M_E be the space of all boundedly-finite measures on $(E, \mathcal{B}_E)$. Clearly $M_E^0 \subset M_E$. We equip this space with the metric of local d^0 convergence in the standard way: Choose an arbitrary *origin point* $o \in E$. For a measure $\mu \in M_E$ and $r \geq 0$ denote $\mu^{(r)}$ as the measure μ restricted to the ball $B_{d^0}(o, r)$. Define the *weak-hat* metric $\hat{d}$ as

$$\hat{d}(\mu, \nu) = \int \exp(-r) \frac{d^0\left(\mu^{(r)}, \nu^{(r)}\right)}{1 + d^0\left(\mu^{(r)}, \nu^{(r)}\right)} dr. \tag{2.3}$$

It may be seen that this topology is characterized by

$$\mu_n \overset{\hat{d}}{\to} \mu \Leftrightarrow: \int f d\mu_n \to \int f d\mu \quad \forall f \in C_b(E) \ \text{ with } f \equiv 0 \text{ outside of a bounded set.}$$

Assume now that E is a c.s.m.s. space. It may be shown that the metric space $(M_E, \hat{d})$ is a c.s.m.s in its own right, and the topology is independent of the choice of o. Moreover, the space $\left(M_E^0, d^0\right)$ is a closed subspace of M_E. The Borel-algebra $\mathcal{M}_E$ is easily characterized by the projections $\Pi_A : M_E \ni \nu \mapsto \nu(A)$ in the sense that

$$\mathcal{M}_E = \sigma(\Pi_A, \ A \in \mathbb{D}), \tag{2.4}$$

where $\mathbb{D} \subset \mathcal{B}_E$ is a semi-ring of bounded sets. We will also need random measures on subspaces of c.s.m.s. spaces. Indeed if E^* is a c.s.m.s. and $E \subset E^*$ is a subspace, then the topology of $\left(M_E, \hat{d}\right)$ is equal to the one induced by identification with elements in $\{\mu \in M_{E^*} : \mu(E^c) = 0\}$ as a subspace of M_{E^*}. Finally, we introduce the subspace of all *counting measures*

$$M_E^c = \{\mu \in M_E : \forall A \in \mathcal{B}_E : \mu(A) \in \mathbb{N}_0 \cup \{\infty\}\}^1. \tag{2.5}$$

This space is seen to be a closed subspace, and therefore a c.s.m.s. in itself. We end this subsection by presenting a measurability result. While

[1]NB: the c superscript in M_E^c does not refer to the complement set, but to the word "counting".

we rarely refer explictly to this result, we mention it as it can be applied to show measurability for a broad range of processes discussed in this thesis.

Lemma 2.1.1.
Let D, E be complete and separable metric spaces, and let $H : D \times E \to \mathbb{R}_+$ be measurable. The section integral $F : M_E \times D \to \overline{\mathbb{R}}$

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

is $\mathcal{M}_E \times \mathcal{B}_D \to \mathcal{B}_{\overline{\mathbb{R}}}$ measurable. ○

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

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

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

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

proving that F is measurable. □

2.1.2 Random Measures

Let E^* be a c.s.m.s. space and let $E \subset E^*$ be a subspace. A measurable map
$\sigma : \Omega \to M_E$ is called a *random boundedly finite measure on E* or just a *random measure* on E for short. A particular interesting example is when $E = \mathbb{R} \times E_0$ where E_0 is a c.s.m.s. as considered in this paper. For such spaces, the first coordinate $t \in \mathbb{R}$ represents a time index. For $r \in \mathbb{R}$ the shift operator θ^{*r} defined by $\theta^{*r}(t, e) = (t + r, e)$ induces an automorphism on $M_{\mathbb{R}\times E_0}$ given by

$$(\theta^r \nu)(C) = \nu(\theta^{*r} C).$$

We also define the *increment measure* ν_{t+} on $[0,\infty) \times E_0$ as

$$\nu_{t+}(A) = \left(\theta^t \nu\right)(\mathbb{R}_+ \times E_0 \cap A), \quad A \in \mathcal{B}_{\mathbb{R}\times E_0}. \tag{2.6}$$

We stress that we intersect with $\mathbb{R}_+ \times E_0$ and not $[0,\infty) \times E_0$. For the sake of clear notation, we shall agree that $\theta^r \nu_{t+}(C) := (\theta^r(\nu_{t+}))(C)$. A random measure σ on $\mathbb{R} \times E_0$ is said to be *stationary* if the distribution is invariant under shift

$$\sigma \overset{D}{=} \theta^r \sigma$$

for all $r \in \mathbb{R}$. Stationarity is equivalent to having invariance of the finite dimensional distributions (fidi's) (Proposition 6.2.III of [11])

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

for all $r \in \mathbb{R}, n > 0, A_1, \ldots, A_n \in \mathcal{B}_{\mathbb{R}\times E_0}, B_1, \ldots, B_n \in \mathcal{B}_{\mathbb{R}_+}$. A stationary random measure σ is *mixing* if

$$P(\sigma \in V, \theta^r \sigma \in W) \overset{|r|\to\infty}{\to} P(\sigma \in V) P(\sigma \in W) \tag{2.7}$$

for all $V, W \in \mathcal{M}_{\mathbb{R}\times E_0}$. We refer to Chapter 10.2-10.3 of [11] for a thorough introduction to ergodic theory for random measures. As is the case for processes, mixing implies that $\mathcal{L}(\sigma)$ is ergodic w.r.t. the shift operator θ^r for all $r \in \mathbb{R}$, meaning that all invariant events have probability 0/1.

2.1.3 Poisson Random Measure

A principal example of a random measure is the *Poisson Random Measure*. These random measures shall serve as the randomness driving the Hawkes Process, similar to the role of a Brownian Motion in an SDE. While we define it on arbitrary spaces for completeness, we shall almost exclusively work with euclidian subspaces.

Definition 2.1.2.
Let E be a c.s.m.s. space. The random measure $\Pi : \Omega \to M_E$ is a Poisson Random Measure with mean measure $\nu \in M_E$ if

1. $\Pi(A) \sim Pois(\nu(A)), \ \forall A \in \mathcal{B}_E, \ \nu(A) < \infty$,

2. *$A_1, ..., A_m \in \mathcal{B}_E$ disjoint $\Rightarrow \Pi(A_1) \perp\!\!\!\perp ... \perp\!\!\!\perp \Pi(A_m)$.*

Let $E_0 \subset E$ be a subspace. We define the PRM on E_0 with mean measure ν as the PRM on E with mean measure $C \mapsto \nu(C \cap E_0)$. $\circ$

If E is euclidian it is assumed that the mean measure of a given PRM is the Lebesgue measure, unless otherwise mentioned. Let us state existence of the PRM immediately. The result may be found in example 9.2 b) and corollary 9.2 IV [11]

Theorem 2.1.3.
Let E be a c.s.m.s space, and let $\mu \in M_E$. There exists a unique PRM distribution on M_E. $\circ$

The following is a direct consequence of the definition.

Proposition 2.1.4.
Let E, K be c.s.m.s. spaces.

- *If $\mu_1, \mu_2 \in M_E$ and π_1, π_2 are independent PRMs on E with mean measure μ_1, μ_2, then $\pi_1 + \pi_2$ is a PRM on E with mean measure $\mu_1 + \mu_2$.*
- *Take $\mu \in M_E$ and let $\phi : E \to K$ be a measurable bijection. If π_E is a PRM on E with mean measure μ then $C \mapsto \pi\left(\phi^{-1}(C)\right)$ is a PRM on K with mean measure $C \mapsto \mu\left(\phi^{-1}(C)\right)$.*

$\circ$

Recall the previously defined shift operator θ^T on $M_{\mathbb{R}\times E_0}$ for $T \in \mathbb{R}$. The shift operator θ^{*T} on $\mathbb{R}$ is a bijection on $\mathbb{R} \times E_0$ and it is easily seen that it is preserves any measure $\mathbf{m} \times \nu$ where $\mathbf{m}$ is the Lebesgue measure and $\nu \in M_{E_0}$. In particular this implies that such a PRM is stationary and mixing.

Let now $(\mathcal{F}_t)_{t\in I}$ be a filtration. A PRM on $\mathbb{R} \times E_0$ with mean measure $\mu \in M_{\mathbb{R}\times E_0}$ is said to be a $\mathcal{F}_t-$PRM if for all $A \in \mathcal{B}_{\mathbb{R}\times E_0}, t \in I$ it holds that

$$\pi\left((-\infty, t] \times E_0 \cap A\right) \text{ is } \mathcal{F}_t \text{ measurable and } \pi\left((t, \infty) \times E_0 \cap A\right) \perp\!\!\!\perp \mathcal{F}_t. \tag{2.8}$$

Therefore, if π is a $\mathcal{F}_t$−PRM then it follows immediately that $\theta^T\pi$ is a $\mathcal{G}_t$-PRM with $\mathcal{G}_t = \mathcal{F}_{t+T}$. Also, PRMs satisfy a strong Markov property, which we present now. Notice that for any random time $\tau : \omega \to [0,\infty]$ and random measure Z the random variable $Z_{\tau+}\mathbb{1}(\tau < \infty)$ is a random measure as well. Also recall that If $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ are two sub-σ-algebras and $C \in \mathcal{F}$ is an event, then we say that $\mathcal{A}$ is conditionally independent of $\mathcal{B}$ given C if

$$P(A \cap B \cap C)P(C) = P(A \cap C)P(B \cap C), \quad \forall A \in \mathcal{A}, B \in \mathcal{B}, \tag{2.9}$$

and we denote this by $\mathcal{A} \perp\!\!\!\perp \mathcal{B} \mid C$. Notice that this simplifies to usual independence if $P(C) = 1$.

Theorem 2.1.5 (Strong Markov Property for PRMs).
Let $(\mathcal{F})_{t\in[0,\infty)}$ be a filtration and let τ be an $\mathcal{F}_t$-stopping time. Let π be an $\mathcal{F}_t$−PRM on $[0,\infty)\times E_0$ with mean measure $\mathbf{m}\times\nu$ where $\mathbf{m}$ is the Lebesgue measure and $\nu \in M_{E_0}$ The conditional distribution $\pi_{\tau+}\mathbb{1}(\tau < \infty) | \tau < \infty$ is also a PRM with mean measure $\mathbf{m} \times \mu$. Moreover, it holds that $\pi_{\tau+}\mathbb{1}(\tau < \infty) \perp\!\!\!\perp \mathcal{F}_\tau | \tau < \infty$. ∘

For a finite stopping time τ this reduces to the fact that $\pi_{\tau+}$ is a PRM independent of $\mathcal{F}_\tau$.

Proof.
The theorem is straightforward if τ is discrete-valued. In the general case, define
$\tau_n = \inf\{k2^{-n} : k2^{-n} > \tau\}$. These random times satisfy the following

- for all $n \in \mathbb{N}$, τ_n is a discrete-valued stopping time,
- $\infty > \tau_n(\omega) > \tau(\omega)$ if $\tau(\omega) < \infty$ and $\tau_n(\omega) = \tau(\omega) = \infty$ otherwise,
- $\tau_n \downarrow \tau$.

From monotonicity we get $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$ and $(\tau_n = \infty) = (\tau = \infty)$ so the discrete-case result applied to τ_n implies

$$\pi_{\tau_n+}\mathbb{1}(\tau < \infty) \perp\!\!\!\perp \mathcal{F}_\tau | \tau < \infty. \tag{2.10}$$

Notice that $\mathbb{D}_{[0,\infty)\times E} = \{(a,b] \times F : F \in \mathcal{B}_E \text{ is bounded and } a, b \in \mathbb{R}_+\}$ is a semi-ring generating $\mathcal{B}_{[0,\infty)\times E}$. Therefore the cylinder sets

$$\begin{aligned}\mathbb{D} = \{\ &\{\mu \in M_{[0,\infty)\times E} : (\mu(V_1), ..., \mu(V_m)) \in B\} : \\ &m \in \mathbb{N}, i \leq m, B \in \mathcal{B}_{\mathbb{R}^m}, V_i \in \mathbb{D}_{[0,\infty)\times E}\}\end{aligned}$$

form an intersection-stable generator for $\mathcal{M}_{[0,\infty)\times E}$. It suffices now to prove that

$$P(\pi_{\tau+} \in D, A, \tau < \infty) P(\tau < \infty) = P(\pi_{\tau+} \in D, \tau < \infty) P(A, \tau < \infty) \tag{2.11}$$

for all $D \in \mathbb{D}$, $A \in \mathcal{F}_\tau$. Take now a fixed combination of such sets, D, A, with
$D = ((\pi(V_1), ..., \pi(V_m)) \in B), V_i = (a_i, b_i] \times F_i \in \mathbb{D}_{[0,\infty)\times E}, B \in \mathcal{B}_{\mathbb{R}^n}$. There exists (random) n_0 such that $\pi_{\tau_n}(V_i) = \pi_\tau(V_i)$ for $n \geq n_0$,$i \leq m$. In particular

$$\mathbb{1}\{\pi_{\tau_n} \in D\} \xrightarrow{a.s.} \mathbb{1}\{\pi_\tau \in D\}.$$

From (2.10) we get that (2.11) holds with π_{τ_n+} in place of $\pi_{\tau+}$. By dominated convergence we obtain (2.11). □

As mentioned previously, we may induce new point processes from PRMs. To do that, let us recall that the predictable algebra $\mathcal{P}$ of the filtration $(\mathcal{F}_t)_{t\geq 0}$ is the algebra on $\Omega \times [0,\infty)$ generated by the sets

$$\{A \times \{0\} : A \in \mathcal{F}_0\} \cup \{A \times (s,\infty) : A \in \mathcal{F}_s, s \in [0,\infty)\}.$$

For a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}}$ we define it as the algebra induced by the sets $\{A \times (s,\infty) : A \in \mathcal{F}_s, s \in \mathbb{R}\}$. A process $t, \omega \mapsto X(t,\omega)$ is predictable if it is measurable w.r.t. to $\mathcal{P}$. A central result regarding stochastic integrals is that any $\mathcal{F}_t$-progressive process Z of finite variation and with $Z_0 = 0$ may be compensated by a predictable process Λ, such that $Z - \Lambda$ is a local martingale. The process Λ is called the *compensator* of Z. A more detailed explanation may be found in [29].

We now construct càdlàg finite variation processes by integrating a possibly random function H w.r.t. π. As with usual stochastic integrals w.r.t. progressive processes, it is important that the integrand field H is predictable.

Lemma 2.1.6.
Let $(\mathcal{F}_t)_{t\in\mathbb{R}}$ be a filtration, and let $\mathcal{P}$ be the predictable algebra w.r.t $(\mathcal{F}_t)$. Let π be an $\mathcal{F}_t$−PRM and assume that the field $H : \Omega \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is $\mathcal{P} \otimes \mathcal{B} \to \mathcal{B}$ measurable. Define

$$Z : t \mapsto \int_0^t \int_0^\infty H(s,z)\, \pi(dz, ds), \quad t > 0,$$

and assume that it almost surely does not explode; that is, for all $t > 0$,

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

almost surely.

1. *If H is bounded, then the compensator Λ of Z is given by*

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

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

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

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

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

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

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

$\circ$

Proof. The first point follows from [29], Theorem 1.8 of Chapter II, by using the localizing sequence

$$T_n = \inf\{t : \int_0^t \int_0^\infty |H(s,z)|\pi(dz,ds) \geq n\}, n \geq 1,$$

since

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

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

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

by monotone convergence. The third point is Lemma 1 in [5]. $\square$

The most important application for this thesis is when H is the field $(s,z) \mapsto \mathbb{1}\left\{z \in \left(\lambda_s^1, \lambda_s^2\right]\right\}$ for some predictable processes $0 \leq \lambda_s^1 \leq \lambda_s^2 < \infty, s \in I$ with $\mathbb{E}\left(\lambda_s^2 - \lambda_s^1\right) < \infty$. For this choice of H, Z induces a random counting measure defined by

$$Z\left(a,b\right] = \int_a^b \int_0^\infty z \in \mathbb{1}\left\{z \in \left(\lambda_s^1, \lambda_s^2\right]\right\} d\pi\left(s,z\right).$$

It is equal to π evaluated on the (random) subset of the plane, enclosed by $t=a, t=b$ and the intensities λ^1, λ^2 on the interval between a and b.

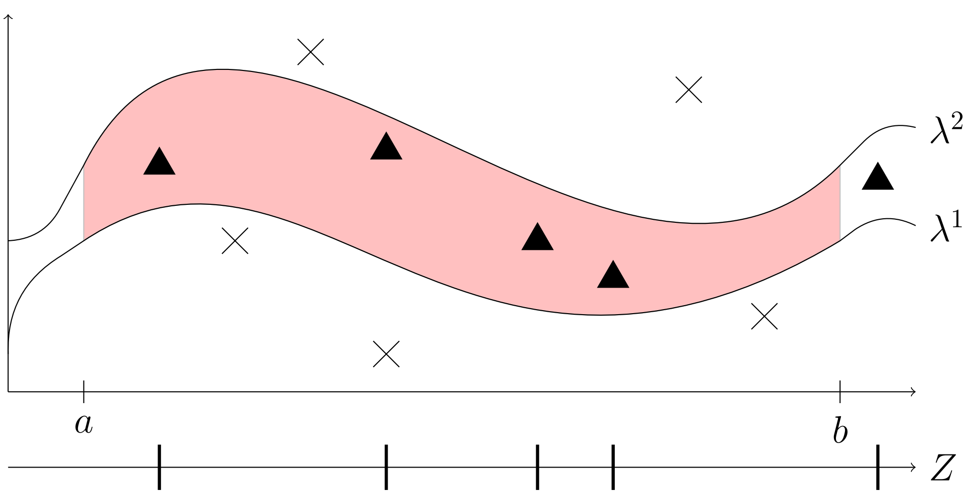

Figure 2.1: Illustration of the construction of Z, $Z[a,b]$. ▲ denotes atoms of π inside the λ^1,λ^2-band, while $\times$ are the atoms outside the band. Here $Z[a,b] = 4$.

Lemma 2.1.6 1) states that for $t_0 \in I$ the $\left(\mathcal{F}_{t+t_0}\right)_{t \geq 0}$-adapted process

$$t \mapsto Z\left(t_0, t_0 + t\right] - \int_{t_0}^{t+t_0} \lambda_s^2 - \lambda_s^1 ds$$

is a local martingale for all choices of t_0.

2.2 Core Assumptions and Definitions

The purpose of this section is to present some core definitions of mathematical objects. These may be seen as the building blocks of the Hawkes process.

The following definitions and assumptions are valid throughout the remaining of this dissertation

- π | π and $\pi^i, i \in \mathbb{N}$, are i.i.d. Poisson Random Measures (PRMs) on $\mathbb{R}_+ \times \mathbb{R}$ with Lebesgue intensity measure.
- $(\mathcal{F}_t)$ | We assume $(\mathcal{F}_t)_{t\geq 0}$ is a filtration such that π, π^i is an $\mathcal{F}_t$-PRM.
- h | Weight functions: For all $1 \leq i, j \leq N$, $h_{ij} : \mathbb{R}_+ \to \mathbb{R}$ is a locally integrable function.
- R_t | Initial signals: For all $1 \leq i \leq N$, $(R^i_t)_{t\geq 0}$ is an $\mathcal{F}_0 \otimes \mathcal{B}$ measurable process on $t \in \mathbb{R}_+$ such that $t \mapsto \mathbb{E}R^i_t$ is locally bounded.
- ψ | Rate functions: For all $1 \leq i \leq N$, $\psi^i : \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}_+$ is a measurable function which is L-Lipschitz in x when the age variables agree, and otherwise sub-linear in x, i.e.,

$$\left|\psi^i(x,a) - \psi^i(x',a')\right| \leq \begin{cases} L_i\left|x - x'\right| & \text{if } a = a' \\ L_i \max\left(\left|x'\right|, \left|x\right|\right) + c^{pre}_{\psi^i} & \text{if } a \neq a' \end{cases}, \tag{2.12}$$

  for some $L_1, ..., L_N, c^{pre}_{\psi^i} > 0$.
- A_0 | Initial ages $\left(A^i_0\right)_{i\in\mathbb{N}}$ are $\mathcal{F}_0-$measurable random variables with support in $\mathbb{R}_+$.

We observe that (2.12) implies that ψ is sublinear since

$$\psi^i(y,b) \leq \left|\psi^i(y,b) - \psi^i(0,0)\right| + \psi^i(0,0) \tag{2.13}$$

so with $c_\psi := \sum_{i=1}^N \left(c^{pre}_{\psi^i} + \psi^i(0,0)\right)$ we have

$$\psi^i(y,b) \leq c_\psi + L_i\left|y\right| \quad \forall y \in \mathbb{R}, b \in \mathbb{R}_+. \tag{2.14}$$

2.3 Age Dependent Hawkes Process

With the objects defined in the previous section, we are now able to present the age dependent Hawkes process (abbreviated ADHP). The aim of this section is to do that, and also to give a light introduction to this process. The emphasis will be on explaining the intuition behind different parameters and initialization processes. The only result presented here is proposition 2.3.2 stating that the ADHP in fact exists. First we define the ADHP

Definition 2.3.1 (The age dependent Hawkes process).
Let $N \in \mathbb{N}$ and let $(Z, X, A) = ((Z^i)_{1\leq i\leq N}, (X^i)_{1\leq i\leq N}, (A^i)_{1\leq i\leq N})$ be a triple consisting of an $N-$dimensional counting process Z, an $N-$dimensional predictable process X, and an $N-$dimensional adapted càglàd process A. The triple is an $N-$dimensional age dependent Hawkes process driven by $\pi^1, ..., \pi^N$ with weight functions $(h_{ij})_{i,j\leq N}$, rate functions $\left(\psi^i\right)_{i\leq N}$, initial ages $\left(A_0^i\right)_{i\leq N}$, and initial signals $\left(R^i\right)_{i\leq N}$ if almost surely all sample paths solve the system

$$\begin{aligned} Z_t^i &= \int_0^t \int_0^\infty \mathbb{1}\left\{z \leq \psi^i\left(X_s^i, A_s^i\right)\right\} \pi^i\left(dz, ds\right), \\ X_t^i &= \sum_{j=1}^N \int_0^{t-} h_{ij}\left(t-s\right) Z^j(ds) + R_t^i, \qquad (2.15) \\ A_t^i - A_0^i &= t - \int_0^{t-} A_s^i \, Z^i(ds), \end{aligned}$$

for $t \geq 0$. The intensity of Z^i is the predictable process $\lambda_t^i = \psi\left(X_t^i, A_t^i\right)$.

∘

The intensity has the interpretation that $P\left(Z^i \text{ jumps in } [t, t+dt) \,|\mathcal{F}_t\right) = \lambda_t^i \, dt$. The process has two parameters: A vector of *rate functions* $\left(\psi^i\right)$ and a matrix of *weight functions* (h_{ij}). It is initialized with an initial age A_0 and an initial signal R both of which are known at time 0. Notice that the compensator of Z^i is given by

$$\Lambda_t^i = \int_0^t \lambda_s^i ds \qquad (2.16)$$

meaning that $Z-\Lambda$ is a local martingale. In fact, we shall see in a moment that it is a true martingale. Also, for each fixed $t_0 \geq 0$ the process

$$\int_0^{t\wedge t_0-} h_{ij}\left(t_0-s\right) dZ_s^j - \int_0^{t\wedge t_0-} h_{ij}\left(t_0-s\right) \lambda_s^j ds \qquad (2.17)$$

is a martingale over $t \geq 0$. We immediately state existence of the Hawkes process

Proposition 2.3.2 (Existence of ADHP).
Almost surely, there is a unique sample path (Z, X, A) solving (2.15). Moreover
$\mathbb{E}\lambda_t^i < \infty$ *and* $\mathbb{E}\left|X_t^i\right| < \infty$ *for all* $t \geq 0$. ◦

Proof.
Let $h = \sum_{i,j=1}^N |h_{ij}|$ and $R = \sum_{i=1}^N \left|R^i\right|$. Let $L = \max\{L_1, ..., L_N, c_{\psi^1}, ..., c_{\psi^N}\}$ and consider the linear Hawkes processes

$$\begin{aligned}
\tilde{Z}_t^i &= \int_0^t \int_0^\infty \mathbb{1}\left\{z \leq L\left(1 + Y_s\right)\right\} \pi^i\left(dz, ds\right), 1 \leq i \leq N, \\
Y_t &= \sum_{j=1}^N \int_0^{t-} h\left(t - s\right) \tilde{Z}^j(ds) + R_t. \qquad (2.18)
\end{aligned}$$

It is well known that $\tilde{Z}_t^i$ is well defined and $\mathbb{E}Y_t < \infty$ (see e.g. proof of theorem 6 in [12]). Notice that $\tilde{Z}$ is driven by the same PRMs as Z. By induction over jump times of $\sum_{j=1}^N \tilde{Z}^j$ it follows that (2.15) has a unique solution satisfying $Z_t^i \leq \tilde{Z}_t^i$ and $\lambda_t^i \leq LY_t + L$ for all $t \in \mathbb{R}_+$, implying that Z does not explode. □

Below we explain some components of the ADHP.

The Rate Function $(x, a) \mapsto \psi^i\left(x, a\right)$ describes how the memory and the age influence the intensity of the ith unit. The existence of a non-exploding Hawkes process is generally ensured by assuming that ψ^i is sub-linear in x. Often, stronger assumptions such as a uniform bound on ψ^i are also imposed to prove basic properties. In this thesis we will always assume some sort of Lipschitz- and linear-growth-conditions.

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

$$A_t^i = \begin{cases} A_0^i + t, & \text{if } Z^i\left(0, t\right) = 0, \\ t - \sup\left\{s < t : \Delta Z_s^i > 0\right\}, & \text{otherwise,} \end{cases}$$

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

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

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

The Initial Signal R_t^i is a process assumed to be known at time $t = 0$. It should be thought of as a memory process which the process inherits from past time. Natural examples of R includes linear combination of "past" memory processes i.e. of the form $\int_0^t h^{past}(t-s)\, dZ^{past}$.

When $\psi^i(x,a) = \psi^i(x)$ i.e. the rate does not depend on the age, we obtain the ordinary Hawkes process. If moreover $\psi(x) = c_\psi + Lx_+$ for some suitable $L > 0$ we obtain the Linear Hawkes process.

Example 2.3.3.
Consider the rate function given by

$$\psi(x,a) = l(x)\varphi(x,a) \tag{2.19}$$

where l is increasing and L-Lipschitz and φ is bounded by 1. Moreover we assume that $x \mapsto l(x)\varphi(x,a)$ is L-Lipschitz for all $a \in \mathbb{R}_+$. For $\varphi \equiv 1$ we obtain the ordinary Hawkes process, so for general φ we may interpret the ADHP as an ordinary Hawkes process with rate function l, but inhibited by its own age process with a factor $\varphi(x,a)$. To show that it satisfies (2.12) take $x \le y$ and $a \ne b$ and see that

$$\psi(y,b) - \psi(x,a) = l(y)\varphi(y,b) - l(x)\varphi(x,a) \tag{2.20}$$

$$= (l(y) - l(x))\varphi(y,b) + l(x)(\varphi(y,b) - \varphi(x,a)). \tag{2.21}$$

If $x \ge 0$ then $l(x) \le l(0) + L(x-0)$ while $l(x) \le l(0)$ for $x < 0$. We use this, and the fact that $|\varphi(y,b) - \varphi(x,a)| \le 1$ to conclude

$$|\psi(y,b) - \psi(x,a)| \le L(y-x) + (l(0) + L|x|) \tag{2.22}$$

which fits into (2.12). The most principal example of φ is the simple $\mathbb{1}\{A \leq \delta)$ corresponding to a hard refractory period of length δ. Although this is a rather simple example, it is important due to its application for modelling neural spike-trains. We could also give a more complicated structure to the refractory period such as

$$\varphi(x,a) = \begin{cases} \frac{1-e^{-a}}{1 \vee l(x)} & a \leq \delta \\ 1-e^{-a} & a > \delta \end{cases}. \tag{2.23}$$

Notice that both of the above mentioned φ choices makes ψ increasing. As mentioned previously, if $\varphi \equiv 1$ then

$$\psi(x,a) = l(x) \tag{2.24}$$

and one obtains the ordinary Hawkes process. ○

2.4 Linear Hawkes Process

The linear Hawkes process was the first Hawkes process to be studied, originally by Alan G. Hawkes and David Oakes [24]. Here we assume that $h \geq 0$ and that the rate functions are independent of a and linear in x. The intensity becomes

$$\lambda_t^i = c_\psi^i + L^i \sum_{j=1}^{N} \int_0^{t-} h_{ij}(t-s)\, dZ_s^j + R_t^i. \tag{2.25}$$

The linear Hawkes process with non-negative weight is more convenient to study, mainly because it can be represented as a marked cluster measure $Z \in M_{[0,\infty)\times\mathcal{K}}$ where $\mathcal{K} = \{1, ..., N\}$

$$Z([0,t]\times\{i\}) = \sum_{j=1}^{N} \int_0^t Z^{\Pi_s^j, j}([0,t-s]\times\{i\})\, d\Pi_s^j. \tag{2.26}$$

Here $\Pi^j \mid \mathcal{F}_0$ is a Poisson process of intensity $c_{\psi^j} + R^j$, implying that the compensator is $t \mapsto \int_0^t c_{\psi^j} + R_s^j ds$, and $Z^{i,j} \overset{D}{=} Z^{0,j}$ are mutually independent Hawkes clusters which we define below.

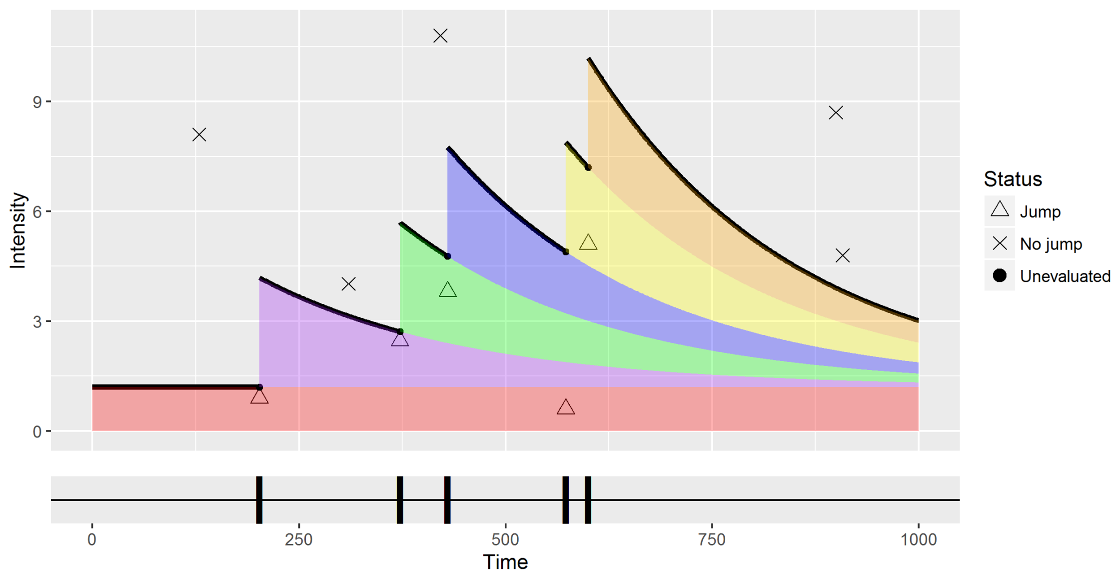

Figure 2.2: A simulation of a one dimensional linear Hawkes process. The colored chunks corresponds to i.i.d. inhomogeneous Poisson processes with intensity h.

Consider the *Galton-Watson* process $(G_i) = (G_{i,1}, ..., G_{i,N})$ on $\mathbb{N}^N$ defined through the following recursion started with $G_{i,0} = 1, G_{j,0} = 0$ for $i \neq j$ and

$$G_{i,n} = \sum_{m=1}^{N} \sum_{j=1}^{G_{i-1,m}} P_{i,j}^{n,m}. \tag{2.27}$$

Here, $P_{i,j}^{n,m}$ are mutually independent and for each pair n, m all variables $P_{i,j}^{n,m}$ are i.i.d. Poisson variables with mean $L_i \|h_{n,m}\|_{\mathcal{L}^1}$. Take mutually independent random variables $(X_{i,j,k}^{n,m})$, $k \in \mathbb{N}$ each with density $h_{n,m}(t)/\|h_{n,m}\|_{\mathcal{L}^1}$. Recall that the jump times of a Poisson process with intensity $h_{n,m}$ are distributed as $(X_{i,j,k}^{n,m})$, $k \leq P_{i,j}^{n,m}$.

The jump times of Z^{0,n_0} for a fixed $n_0 \leq N$ may be constructed as follows: Start with a jump $T_0^{0,n_0} = 0$ and inductively for $i \in \mathbb{N}$ do the following:

- For each $j \leq G_{i-1,m}$ and $n, m \leq N, k \leq P_{i,j}^{n,m}$ define[1] $T_{i,j,k}^{n} = X_{i,j,k}^{n,m} + T_{i-1,j}^{m}$.

[1] with the convention that $P_{i,j}^{n,m} = 0$ means that no new $T_{i,j,k}^{n}$ times are defined

- Re-index the jump times by concatenating over the indices j, k, into a sequence $\left(T^n_{i,j}\right)$, $j \leq G_{i,n}$ of i'th generation jump times.

See also [41] and [24] for more details on the cluster representation.

We may use the Galton-Watson process to control the growth of the clusters, due to the following well known result.

Proposition 2.4.1.
Let (G_i) be a Galton-Watson process with $G_0 = v$. Let $A \in \mathbb{R}^{N\times N}$ be the matrix $A_{nm} = L_n \left\|h_{nm}\right\|_{\mathcal{L}^1}$ and let ϱ be its spectral radius.

- *Assume $\varrho < 1$ (Sub-critical case). Then extinction occurs almost surely, i.e. $G_i = 0$ eventually. The extinction time has exponential moment. Moreover, Let $S_i = \sum_{j=0}^{i} G_j$. The total progeny $S := \lim_{i\to\infty} S_i$ has exponential moment.*
- *Assume $\varrho = 1$ (Critical case). Then extinction occurs almost surely.*
- *Assume $\varrho > 1$ (Super-critical case). Then there is a positive probability of no extinction and $\mathbb{E}G_i$ grows exponentially.*

$\circ$

For notational convenience we prove this is one dimension, meaning $N = 1$ and $\varrho = L\left\|h\right\|_{\mathcal{L}^1}$. The multivariate proof is similar.

Proof.
Define the filtration $(\mathcal{F}_i) = \sigma\left(G_j, S_j : j \leq i\right)$. We show that the process $M_i = \exp\left(pS_i + qG_i\right)$ is a supermartingale for appropriately chosen $p, q \in \mathbb{R}$. Recall that the moment generating function of a Poisson variable with mean $c \in \mathbb{R}$ is $\mathbb{R} \ni t \mapsto \exp\left(c\left(e^t - 1\right)\right)$.

$$
\begin{aligned}
\mathbb{E}\left(M_{i+1} \mid \mathcal{F}_i\right) &= \exp\left(pS_i\right)\mathbb{E}\left(\exp\left(\left(p+q\right)G_{i+1}\right) \mid \mathcal{F}_i\right) &(2.28)\\
&= \exp\left(pS_i\right)\mathbb{E}\left(\exp\left(\left(p+q\right)\sum_{j=1}^{G_i} P_{i+1,j}\right) \mid \mathcal{F}_i\right) &(2.29)\\
&= \exp\left(pS_i\right)\exp\left(G_i\varrho\left(e^{(p+q)} - 1\right)\right) &(2.30)\\
&= \exp\left(M_i\right)\exp\left(G_i\left(\varrho\left(e^{(p+q)} - 1\right) - q\right)\right). &(2.31)
\end{aligned}
$$

Take now $p, q > 0$ s.t. $q^{-1}\left(e^{(p+q)}-1\right) < \varrho^{-1}$. This is possible since $\lim_{q\to 0}\lim_{p\to 0} q^{-1}\left(e^{(p+q)}-1\right) = 1$. For such p, q M_n is a strict supermartingale. Consider the Markov chain (S_i^*, G_i^*), defined with same Markov kernel as (S_i, G_i) except that it is modified so that $\left(G_{i+1}^*, S_{i+1}^*\right) = (1, 0)$ whenever $(S_i^*, G_i^*) \in \mathbb{N}\times\{0\}$. The modified chain (S_i^*, G_i^*) is clearly irreducible and $\mathbb{N}\times\{0\}$ is an accessible atom for it. Choose now the Lyapunov function $V(s, g) = \exp(ps + qg)$. It follows from [34] theorem 15.0.1 that the chain (S_i^*, G_i^*) is positive recurrent and geometrically ergodic. The remaining claims of the first part also follow from theorem 15.0.1 in [34].

To prove the 2nd part and 3rd part, we notice that $M_i' = \varrho^{-i}G_i$ is a martingale since

$$\mathbb{E}\left(M_{i+1}' \mid \mathcal{F}_i\right) = \varrho^{-i-1}\mathbb{E}\left(\sum_{j=1}^{G_i} P_{i+1,j} \mid \mathcal{F}_i\right) = \varrho^{-i-1}\left(\varrho G_i\right) = M_i'. \tag{2.32}$$

The martingale convergence theorem gives that there is some random variable M_∞' s.t. $M_i' \overset{a.s.}{\to} M_\infty'$. Also, notice that $P\left(G_k = j\ \forall k > i\right) = 0$ for all $i, j \in \mathbb{N}$ so if G_i converge it must be towards 0.

In the critical case this implies that almost surely, $G_i = M_i' = 0$ eventually. In the super-critical case we conclude that $\mathbb{E}G_i = v\varrho^i$. To see that there is a positive probability of no extinction, pick $p = 0, q < 0$ s.t. $q^{-1}\left(e^q - 1\right) > \varrho^{-1}$. For these parameters, the derivations made in the first part gives that $M_i = 1 - \exp\left(qG_i\right)$ is a non-negative strict submartingale. Consider the chain (G_i^*) given as G_i but modified so what $G_{i+1} = v$ whenever $G_i = 0$. It follows from theorem 8.4.2. in [34] with Lyapunov function $V(x) = 1 - \exp(qx)$ that the modified chain (G_i^*) is transient. This implies that G_i^* hits v only finitely many times, before going to infinity. Thus, G_i has a positive probability of no extinction. □

While our focus is not on producing results specifically for the special case of linear Hawkes processes, we will present the stationary linear Hawkes process nevertheless. Let Π^j be a Poisson process on $\mathbb{R}$ and let $Z^{k,j}$, $k \le N$, $j \in \mathbb{Z}$, be mutually independent Hawkes clusters. The random measure defined by

$$Z\left((a,b]\times\{i\}\right)=\sum_{j=1}^{N}\int Z^{\Pi_s^j,j}\left((a-s,b-s]\times\{i\}\right)d\Pi_s^j,\quad a\leq b\in\mathbb{R} \tag{2.33}$$

is almost surely boundedly finite. Moreover it is stationary and ergodic. As we shall see later, this is a stronger result than what can be obtained for Nonlinear Hawkes processes, since we make no further assumptions on the tail of h_{nm} than integrability. In general, the tail of h is directly connected to the length of clusters

Proposition 2.4.2.
If $\varrho<1$ and $\int_0^\infty t^p h_{n,m}(t)dt<\infty$ for $p\geq 1$ and all $n,m\leq N$ then the extinction time has p-th moment. If also $\int_0^\infty \exp(ct)h_{n,m}(t)dt<\infty$ for all $n,m\leq N$ and some $c>0$, then the extinction time has exponential moment. ∘

For a proof see [41]. We also implicitly prove this in the proof of theorem 5.3.6 using random exchange processes (see appendix).

The described decomposition into i.i.d clusters makes the linear Hawkes process more convenient to analyse than the nonlinear Hawkes process which does not enjoy such decomposition. It also makes it faster and easier to simulate the Hawkes process in practice, which often counterweights its downside, being its simplicity compared to the more general Hawkes processes.

As mentioned previously, there is a relation between a general Hawkes process, and the linear Hawkes Process ensured by the fact that ψ is sublinear i.e. $\psi(x)\leq c_\psi+Lx$. If one removes all inhibition from the system (i.e. replace h with h_+) and replace ψ with the linear dominating function $x\mapsto c_\psi+Lx$, then the newly obtained linear Hawkes process Z^{lin} dominates Z in the sense that $Z^{lin}[t,t]\geq Z[t,t]$ for all $t\in\mathbb{R}_+$.

2.5 Markovian Hawkes Processes

In general, Hawkes processes are not Markovian. However, specific choices of weight functions will induce a Markovian system. An example of such weight functions are the scaled Erlang densities:

$$h_{ij}(x) = \frac{c_{ij}}{n_{ij}!} x^{n_{ij}} \exp(-\nu_{ij} x) \tag{2.34}$$

for $v_{ij}, \geq 0, c_{ij} \in \mathbb{R}, n_{ij} \in \mathbb{N}_0$. The Markovian system is discovered by splitting up the memory processes X^i into the contributions from each unit in the network $X^{ij} = X_0^{ij}$. Define in general

$$X_k^{ij}(t) = \frac{c_{ij}}{(n_{ij} - k)!} \int_0^{t-} t^{n_{ij}-k} \exp(-\nu_{ij} t)\, dZ_s^j \tag{2.35}$$

for $i, j \leq N$ $k \leq n_{ij}$. Using partial integration iteratively we obtain the system

$$dX_k^{ij}(t) = -\nu_{ij} X_k^{ij}(t)\, dt + X_{k+1}^{ij}(t)\, dt \quad k < n_{ij} \tag{2.36}$$

$$dX_{n_{ij}}^{ij}(t) = -\nu_{ij} X_{n_{ij}}^{ij}(t)\, dt + \sum_{j=1}^{n} c_{ij} \pi^j \left(\left[0, \psi^j \left(\sum_{l=1}^{N} X_0^{jl}(t), A_t^j \right) \right], dt \right) \tag{2.37}$$

$$dA_t^i = A_t^i dt - A_t^i \pi^i \left(\left[0, \psi^i \left(\sum_{j=1}^{N} X_0^{ij}(t), A_t^i \right) \right], dt \right). \tag{2.38}$$

If for all $i, j_1, j_2 \leq N$, it holds that $v_{ij_1} = v_{ij_2}, n_{ij_1} = n_{ij_2}$ then the memory process and its cascades $X_k^i = \sum_{j=1}^{N} X_k^{ij}$ themselves are Markovian. For $n_{ij} = 0$ the weight functions are exponential, which has been studied since the original works by Hawkes. The general Erlang case is more recent. In [13] it was studied how such systems in a large network are able to produce oscillating intensities.

2.6 A Brief Overview of Hawkes Process Results

Here we present a non-exhausting list of results relevant to Hawkes Processes and work presented in this thesis.

Stability

A central result for this dissertation, originally developed by Brémaud and Massoulié [5], states that ordinary Hawkes processes with limited connectivity strength are stable. More specifically, let Λ be the $(N \times N)$-matrix consisting of the $\mathcal{L}^1$-norms of the weight functions $\Lambda_{i,j} = L_i \left\| h_{i,j} \right\|_{\mathcal{L}^1}$. Assume that the spectral radius of Λ is strictly less than 1 and $\int_0^\infty t|h_{ij}(t)|dt < \infty$. Then there exist a Hawkes process living on the entire timeline and driven by π, meaning that

$$\begin{aligned} Z^i\left(t_1, t_2\right] &= \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}\left\{z \le \psi^i\left(X_s^i\right)\right\} \pi^i\left(dz, ds\right), \\ X_t^i &= \sum_{j=1}^N \int_{-\infty}^{t-} h_{ij}\left(t-s\right) Z^j(ds) \end{aligned} \tag{2.39}$$

for all $t \in \mathbb{R}$ and $t_1 \le t_2 \in \mathbb{R}$. Moreover, the solution is *compatible* to π in the sense that there is a map H satisfying $H\left(\theta^t \pi\right) = \theta^t Z$. In particular this implies that the stable process is stationary and ergodic. Moreover, it is shown that if Z^* is another ordinary Hawkes process initialized with some R^* satisfying $\mathbb{E}\int_0^\infty \left|R_s^*\right| ds < \infty$ and driven by same π, then Z and Z^* couples eventually i.e.

$$T := \inf\left\{t > 0 : \left(Z - Z^*\right)(t, \infty) = 0\right\} < \infty.$$

In a later article by Bremaud et al. [6] distribution results were shown for the coupling time T. In particular they showed that if $\int_0^t t^{p+1} h\left(t\right) dt < \infty$ for $p \ge 0$ then T has p'th moment. Also it was shown that T has exponential moment if h has an exponentially decaying tail.

CLT results

A FCLT result was found by Zhu [44] for a one-dimensional Hawkes process, when h is positive and decreasing. Define for $n \in \mathbb{N}$

$$B_t^n = n^{-1/2}\left(Z\left(0, nt\right] - \mathbb{E}Z\left(0, nt\right]\right). \tag{2.40}$$

It was proven that when $L\left\| h_{i,j} \right\|_{\mathcal{L}^1} \le 1$ and $\int_0^\infty th\left(t\right) dt < \infty$ there is $0 < \sigma < \infty$ s.t.

$$B^n \Rightarrow \sigma B, \tag{2.41}$$

where B is a Brownian motion, and " $\Rightarrow$ " denotes weak convergence in the D-space.

A time-average CLT result was found by Costa et al. in [10] in the case where h has bounded support. Take $A \geq \sup\{t > 0 : h(t) > 0\}$. Their approach is to study the Hawkes process as a Markov process. Because of the bounded support assumption the Hawkes process regenerates at $\rho = \inf\{t > 0 : Z[t - A, t] = 0\}$. In the Markov construction ρ can be seen as the return time to an atom. Here it was proven that the return times have exponential moments with exact coefficients. Although the results are stated for the linear Hawkes process, most results including the renewal results transfer to any Hawkes process where h has bounded support. From the Markov process the following CLT is established.

Assume that $G : M^c_{(-A,0]} \to \mathbb{R}$ is measurable and locally bounded, i.e. uniformly bounded on $\{\nu \in M^c_{(-A,0]} : \nu(-A, 0] \leq n\}$ for all n. Now define the probability measure μ and the variance σ^2 by

$$\mu = \frac{1}{\mathbb{E}\rho}\mathbb{E}\int_0^\rho G((\theta^t Z)_{|(-A,0]})dt, \quad \sigma^2 = \frac{1}{\mathbb{E}\rho}\mathbb{E}\left[\left(\int_0^\rho G((\theta^t Z)_{|(-A,0]})dt - \mu\right)^2\right]. \tag{2.42}$$

If $\sigma^2 \in (0, \infty)$ then the following holds

$$\sqrt{T}\left(\frac{1}{T}\int_0^T G((\theta^t Z)_{|(-A,0]})dt - \mu\right) \Rightarrow \mathcal{N}(0, \sigma^2) \quad \text{when } T \to \infty. \tag{2.43}$$

More results, including a concentration inequality may be found in the same paper.

Mean-field

Mean field limits have been studied for many types of multivariate dynamical systems, including Hawkes processes. The idea is to study the asymptotics for a multivariate Hawkes process for large N. In the limit, the dependence between two individual units should vanish in the overall correlation structure of the entire system. In [12] a mean-field limit was proven for Hawkes processes modelling a homogeneous population, where each weight function is the same for all units for fixed N, and given by $N^{-1}h$. More precisely, let $Z^{i,\infty}$ be i.i.d. inhomogeneous Poisson processes with intensity λ_t^∞ being the solution to

$$\lambda_t^\infty = f\left(\int_0^t h\left(t-s\right) d\Lambda_t\right) \tag{2.44}$$

and where $\Lambda_t := \int_0^t \lambda_s^\infty ds$ is the compensator of Z^∞. Let $i_1, ..., i_n \in \mathbb{N}$ be any finite set of indices. It holds that

$$\left(Z^{N,i_1}, \ldots, Z^{N,i_n}\right) \Rightarrow Z^{\infty,1} \otimes, \ldots, \otimes Z^{\infty,n}. \tag{2.45}$$

where " $\Rightarrow$ " denotes weak convergence in the D-space. An analogous mean-field limit was proven in [13] in a multi-class setup instead of a homogeneous one. A mean-field limit was proven for a homogeneous population for ADHPs in [7].

Chapter 3

A Multiclass Hawkes Process Stabilized by Inhibition

3.1 Introduction

In this chapter we consider a multivariate linear Hawkes process in a multi-class setup as in [13]. We consider two populations of possibly different sizes, such that one of the populations acts excitatory on the system, and the other population acts inhibitory on the system. The goal of this chapter is to present a class of such Hawkes processes with stable dynamics, but without assumptions on the spectral radius of the weight function matrix. Thus it illustrates how inhibition in a Hawkes system significantly affects the stability properties of the system.

3.2 Setup For This Chapter

We assume that we have a total of N units, represented by an N-dimensional Hawkes process. These units are divided into two populations marked with " $+$ " or " $-$ " signaling that the population acts excitatory or inhibitory on the system, respectively. Let $N^+, N^- \in \mathbb{N}$ be the number of units in each population with a total population of $N = N^+ + N^-$. The weight function from a unit in one population to a unit in another is assumed to be given by a decaying exponential and depends only on their respective populations. Thus, there are only 4 different weight functions,

$$h_{++}(t) = \frac{c_{++}}{N^+} \exp(-\nu_+ t), \quad h_{+-}(t) = \frac{c_{+-}}{N^+} \exp(-\nu_+ t),$$
$$h_{-+}(t) = \frac{c_{-+}}{N^-} \exp(-\nu_- t), \quad h_{--}(t) = \frac{c_{--}}{N^-} \exp(-\nu_- t).$$

where for example h_{+-} indicates the weight function from a unit in the excitatory group " + " to a unit in the inhibitory group " − "[1]. The coefficients of this system are the exponential leakage terms $\nu_+ > 0, \nu_- > 0$ and the synaptic weights $c_{++}, c_{+-}, c_{-+}, c_{--}$ satisfying that

$$c_{++} \geq 0, \; c_{+-} \geq 0, \; c_{--} \leq 0, \; c_{-+} \leq 0. \tag{3.1}$$

Put $R^{i+} = X_0^{i+} e^{-\nu_+ t}, R^{j-} = X_0^{j-} e^{-\nu_- t}$ for $i \leq N^+, j \leq N^-$. The multivariate linear Hawkes process with these parameters are given as

$$Z_t^{i+} = \int_0^t \int_0^\infty \mathbb{1}_{\{z \leq \psi_+^{i+}(X_{s-}^+)\}} \pi^{+i}(ds, dz), i \leq N^+ \tag{3.2}$$
$$Z_t^{j-} = \int_0^t \int_0^\infty \mathbb{1}_{\{z \leq \psi_-^{j-}(X_{s-}^-)\}} \pi^{-j}(ds, dz), j \leq N^- \tag{3.3}$$

$$X_t^+ = e^{-\nu_+ t} X_0^+ + c_{++} \sum_{i=1}^{N^+} \int_0^t e^{-\nu_+(t-s)} dZ_s^{i+} + c_{-+} \sum_{j=1}^{N^-} \int_0^t e^{-\nu_+(t-s)} dZ_s^{j-}, \tag{3.4}$$
$$X_t^- = e^{-\nu_- t} X_0^- + c_{+-} \sum_{i=1}^{N^+} \int_0^t e^{-\nu_-(t-s)} dZ_s^{i+} + c_{--} \sum_{j=1}^{N^-} \int_0^t e^{-\nu_-(t-s)} dZ_s^{j-}, \tag{3.5}$$

and jump rate functions ψ_+^i, ψ_-^j given by

$$\psi_+(x) = a_{+i} + \max(x, 0), \quad a_{+i} > 0, \tag{3.6}$$
$$\psi_-(x) = a_{-j} + \max(x, 0), \quad a_{-j} > 0. \tag{3.7}$$

As mentioned in the introduction (X^+, X^-) is a piecewise deterministic

[1] NB: This means that the indices of h_{ij} are opposite of the rest of the thesis, so that j is the recieving unit and i is the sending one.

Markov process. The generator of the Markov process is given by

$$Ag(x,y) = -\nu_+ x\partial_x g(x,y) - \nu_- y\partial_y g(x,y) \tag{3.8}$$
$$+\sum_{i=1}^{N^+} \psi_+^i(x)[g(x+\frac{c_{++}}{N^+}, y+\frac{c_{+-}}{N^+}) - g(x,y)] \tag{3.9}$$
$$+\sum_{j=1}^{N^-} \psi_-^j(y)[g(x+\frac{c_{-+}}{N^-}, y+\frac{c_{--}}{N^-}) - g(x,y)], \tag{3.10}$$

for any test function $g \in C^1$.

The classic stability results found in [5] for multivariate nonlinear Hawkes processes are stated in terms of the weight function matrix Λ. This is the $N \times N$ matrix with ij-th entry being $\Lambda_{ij} = L_i \left\|h_{ij}\right\|_{\mathcal{L}^1}$. The criteria is that the spectral radius of Λ should be strictly smaller than 1. In our system, the weight function matrix is given by the block matrix

$$\Lambda = \left(\begin{array}{c|c} \frac{c_{++}}{\nu_+} & \frac{|c_{-+}|}{\nu_+} \\ \hline \frac{c_{+-}}{\nu_-} & \frac{|c_{--}|}{\nu_-} \end{array}\right). \tag{3.11}$$

Notice that in (3.11), negative synaptic weights do only appear through their absolute values. This is due to the fact that using the Lipschitz continuity of the rate functions leads automatically to considering absolute values and does not enable us to make profit from the inhibitory action of c_{-+} and c_{--}. Obviously, having sufficiently fast decay, that is, $\min(\nu_+, \nu_-) >> 1$, is a sufficient condition for subcriticality.

The purpose of this note is to show how the presence of sufficiently strong negative synaptic weights helps stabilizing the process without imposing such a subcriticality condition, in particular, without imposing ν_+, ν_- being large. To the best of our knowledge, no such study has been proposed in the literature. [5] gives an attempt in this direction but does only deal with the case when c_{+-} and c_{-+} are of the same sign (see Theorem 6 in [5]).

In the following, we shall write

$$c^*_{++} := c_{++} - \nu_+, \; c^*_{--} := c_{--} - \nu_-.$$

Notice that c^*_{++} could be interpreted as the net increase of X^+ due to self-interactions of X^+ with itself. c^*_{--} is always negative.

Assumption 3.1.
We assume the following inequalities

$$c_{++}^* + c_{--}^* < 0, \tag{3.12}$$
$$(c_{++}^* - c_{--}^*)^2 < 4c_{+-}|c_{-+}|, \tag{3.13}$$
$$c_{++}^* - c_{--}^* > 0. \tag{3.14}$$

○

Assumption 3.2.
We assume that $\nu_+ = \nu_-$ and $(c_{++}, c_{+-}), (c_{-+}, c_{--})$ are linearly independent. ○

The main theorem in this chapter is that assumption 3.1 and assumption 3.2 imply positive Harris recurrence together with a strong mixing result. The important assumption is assumption 3.1, which makes sure that the system is balanced. Notice that assumption 3.1 does not imply - nor is implied by - that the spectral radius of Λ is strictly smaller than 1. For example, if assumption 3.1 is satisfied for some parameters $(c_{++}, c_{+-}, c_{-+}, c_{--}, \nu, \nu)$ such that $c_{++}+c_{--} < 0$ then for all $C > 1$ the parameters $(Cc_{++}, Cc_{+-}, Cc_{-+}, Cc_{--}, \nu, \nu)$ satisfies assumption 3.1 as well. But the associated offspring matrix Λ_C of the scaled parameters is equal to $C\Lambda$ and thus the spectral radius is also scaled by C. Assumption 3.2 could be loosened or even removed, at the cost of more complicated proofs, and more technical formulations for the results.

In order to state the main result, let

$$\bar{V}(z) := V(z) + 1 \text{ and } \|\mu\|_{\bar{V}} := \sup_{g:|g|\leq \bar{V}} |\mu(g)|,$$

for any probability measure μ on $\mathcal{B}_{\mathbb{R}^2}$. Moreover, for $t > 0$ and $z = (x, y) \in \mathbb{R}^2$, we write $P_t(z, \cdot)$ for the transition semigroup of the process, defined through
$P_t(z, A) = \mathbb{E}_z(1_A(X_t))$.

Theorem 3.2.1.
Grant assumption 3.1 and assumption 3.2.

- *Recurrence: The process $X_t := (X_t^+, X_t^-)$ is positive recurrent in the sense of Harris. In particular, it possesses a unique invariant probability measure μ.*

Geometric Ergodicity: There exists a quadratic polynomial $\overline{V}$ *and constants*
$c_1, c_2 > 0$ *such that for all* $z \in \mathbb{R}^2$,

$$\|P_t(z,\cdot) - \mu\|_{\bar{V}} \leq c_1 \bar{V}(z) e^{-c_2 t}. \tag{3.15}$$

∘

3.3 Proof of Theorem 3.2.1

This section is devoted to the proof of the above theorem.

A Lyapunov Function for X_t

The first result shows that if the cross-interactions, that is, influence from X^+ to X^- and vice versa, are sufficiently strong, then – under mild additional assumptions – it is possible to construct a Lyapunov function for the system that mainly profits from the inhibitory part of the jumps.

Proposition 3.3.1.
Assume assumption 3.1. Define the function $V : \mathbb{R}^2 \to \mathbb{R}$

$$V(x,y) := \begin{cases} V_{++}(x,y) := c_{+-}x^2 - c_{-+}y^2 - (c^*_{++} - c^*_{--})xy & x \in \mathbb{R}_+, y \in \mathbb{R}_+ \\ V_{+-}(x,y) := c_{+-}x^2 + qy^2 - (c^*_{++} - c^*_{--})xy & x \in \mathbb{R}_+, y \in \mathbb{R}_- \\ V_{-+}(x,y) := px^2 - c_{-+}y^2 - (c^*_{++} - c^*_{--})xy & x \in \mathbb{R}_-, y \in \mathbb{R}_+ \\ V_{--}(x,y) := px^2 + qy^2 - (c^*_{++} - c^*_{--})xy & x \in \mathbb{R}_-, y \in \mathbb{R}_- \end{cases}$$

with p *so small such that*

$$-(c^*_{++} - c^*_{--})(c_{--} - \nu_+ - \nu_-) + 2pc_{-+} > 0$$

and q *so large such that*

$$(c^*_{++} - c^*_{--})[\nu_+ + \nu_- - c_{++}] + 2qc_{+-} > 0 \text{ and } 4pq > (c^*_{++} - c^*_{--})^2.$$

Then $\lim_{|x|+|y|\to\infty} V(x,y) = \infty$ *and there exists positive constants* $\kappa, c, K > 0$ *such that*

$$AV(x,y) \leq -\kappa V(x,y) + c\mathbb{1}_{\{|x|+|y| \geq K\}}. \tag{3.16}$$

∘

Proof. We calculate $AV(x,y) = A^1V(x,y) + A^2V(x,y)$, with

$$A^1V(x,y) = -\nu_+\partial_x V(x,y) - \nu_-\partial_y V(x,y)$$

and A^2 the jump part of the generator.

Part 1.1 Suppose first that $x \geq |c_{-+}|/N^-, y \geq |c_{--}|/N^-$. Then

$$AV(x,y) = A^1V_{++}(x,y)+A^2V_{++}(x,y) = a_{++}x^2+b_{++}xy+d_{++}y^2+L_{++}(x,y),$$

where L_{++} is a polynomial of degree 1. A straightforward calculus shows that

$$\begin{array}{rcl} a_{++} & = & c_{+-}(c^*_{++} + c^*_{--}), \\ b_{++} & = & -(c^*_{++} - c^*_{--})(c^*_{++} + c^*_{--}) \\ d_{++} & = & -c_{-+}(c^*_{++} + c^*_{--}), \end{array}$$

proving that

$$AV(x,y) = (c^*_{++} + c^*_{--})V(x,y) + L_{++}(x,y),$$

implying that there exists $K, \kappa > 0$ such that

$$AV(x,y) \leq -\kappa V(x,y)$$

for all $x > K, y > K$, since $c^*_{++} + c^*_{--} < 0$ by assumption.

Part 1.2 Suppose now that $0 \leq x < |c_{-+}|/N^-$ and $y \geq |c_{--}|/N^-$. Then a jump of any inhibitory neuron will lead to a change $x \mapsto x + c_{-+}/N^- < 0$. In this case we obtain

$$\begin{aligned} & AV(x,y) \\ =& AV_{++}(x,y) + \sum_{j=1}^{N^-}(a_{-j} + y)\left[V_{-+}(x + \frac{c_{-+}}{N^-}, y + \frac{c_{--}}{N^-}) - V_{++}(x + \frac{c_{-+}}{N^-}, y + \frac{c_{-+}}{N^-})\right]. \end{aligned}$$

But

$$\left|V_{-+}(x + \frac{c_{-+}}{N^-}, y + \frac{c_{--}}{N^-}) - V_{++}(x + \frac{c_{-+}}{N^-}, y + \frac{c_{--}}{N^-})\right| \leq C,$$

and therefore

$$AV(x,y) \leq AV_{++}(x,y) + L(y),$$

where $L(y)$ is a monomial in y.

The other case $0 \leq y < |c_{--}|/N^-$ and $x \geq |c_{-+}|/N^-$ is treated analogously.

Part 2.1 Suppose now that $x \geq |c_{-+}|/N^-, y \leq -c_{+-}/N^+$. Then

$$AV(x,y) = A^1 V_{+-}(x,y) + A^2 V_{+-}(x,y) = a_{+-}x^2 + b_{+-}xy + d_{+-}y^2 + L_{+-}(x,y),$$

where L_{+-} is a polynomial of degree 1. We obtain

$$\begin{array}{rcl} a_{+-} & = & c_{+-}(c^*_{++} + c^*_{--}), \\ b_{+-} & = & (c^*_{++} - c^*_{--})(\nu_+ + \nu_- - c_{++}) + 2qc_{+-} \\ d_{+-} & = & -2\nu_- q. \end{array}$$

Since $b_{+-} > 0$ by choice of q, this implies that for a suitable positive constant $\kappa > 0$,

$$AV(x,y) \leq -\kappa V(x,y) + L_{+-}(x,y),$$

which allows to conclude as before.

Part 2.2 The cases $x \geq |c_{-+}|/N^-, 0 \geq y > -c_{+-}/N^+$ or $0 \leq x < |c_{-+}|/N^-$, $y \leq -c_{+-}/N^+$ are treated analogously to Part 1.2.

Part 3 Suppose now that $x \leq -c_{++}/N^+$, $y \geq -c_{--}/N^-$. Then

$$AV(x,y) = A^1 V_{-+}(x,y) + A^2 V_{-+}(x,y) = a_{-+}x^2 + b_{-+}xy + d_{-+}y^2 + L_{-+}(x,y),$$

where L_{-+} is a polynomial of degree 1 and where

$$\begin{array}{rcl} a_{-+} & = & -2\nu_+ p, \\ b_{-+} & = & (c^*_{++} - c^*_{--})(\nu_+ + \nu_- - c_{--}) + 2pc_{-+} \\ d_{-+} & = & -c_{-+}(c^*_{++} + c^*_{--}). \end{array}$$

Notice that by choice of p, $b_{-+} > 0$. The conclusion of this part follows analogously to the previous parts 1.1 and 2.1.

Part 4 Suppose finally that $x \leq -c_{++}/N^+, y \leq -c_{+-}/N^+$. Then

$$AV(x,y) = A^1 V_{--}(x,y) + A^2 V_{--}(x,y) = a_{--}x^2 + b_{--}xy + d_{--}y^2 + L_{--}(x,y),$$

where L_{--} is a polynomial of degree 1 and where

$$\begin{array}{rcl} a_{--} & = & -2\nu_+ p, \\ b_{--} & = & (c^*_{++} - c^*_{--})(\nu_+ + \nu_-) \\ d_{--} & = & -2\nu_- q, \end{array}$$

leading to the same conclusion as in the previous parts. $\square$

As a consequence of Proposition 3.3.1, the process X_t is stable in the sense that it necessarily possesses invariant probability measures, maybe several of them.

Next we prove the following useful property.

Proposition 3.3.2.
The process X is a Feller process, that is, for any $f : \mathbb{R}^2 \to \mathbb{R}$ which is bounded and continuous, we have that the map given by

$$\mathbb{R}^2 \ni (x, y) \mapsto \mathbb{E}_{(x,y)} f(X_t) = P_t f(x, y)$$

is continuous. ○

The proof of this result follows from classical arguments, see e.g. the proof of Proposition 4.8 in [27], or [28]. Next, recall that a set $C \in \mathbb{R}^2$ is *petite* for X if there is a probability measure μ on $\mathbb{R}_+$ and a positive measure ν on $\mathbb{R}^2$ such that

$$\int P_x^t(A) d\mu(t) \geq \nu(A) \quad \forall x \in C, \ A \in \mathcal{B}^2. \tag{3.17}$$

When the probability measure μ is the dirac measure the criteria becomes existence of ν and $t \geq 0$ such that

$$P_x^t(A) \geq \nu(A) \quad \forall x \in C, \ A \in \mathcal{B}^2. \tag{3.18}$$

Proposition 3.3.3.
Grant assumption 3.2. It holds that all compact sets in $\mathbb{R}^2$ are petite for the Markov process X. In particular X is a φ-irreducible T-chain. ○

At the cost of a more complicated proof, it is possible to remove assumption 3.2 and still obtain a $T-$chain. See [9] lemma 6.3 for details.

Proof. We prove it for $N^+ = N^- = 1$ for notational convenience. It will suffice to prove that any fixed box $H = \left\{x, y \in \mathbb{R}^2 : |x| \leq M^+, |y| \leq M^-\right\}$ is petite. We do so by proving that there is some time $T > 0$ s.t.

$$P_x^T(A) \geq \beta m_2 (A \cap A_{T,H})$$

where $A_{T,H}$ is some subset of $\mathbb{R}^2$ depending on H and T. To do so, set $E \in \mathcal{F}$ as the event

- $\pi^+ (0, T] \times [a_+, M^+ + a_+ + c_{++}] = 0$

- $\pi^{+}(0,T] \times [0,a_{+}] = 1$
- $\pi^{-}(0,T] \times [a_{-}, M^{-} + a_{-} + c_{+-}] = 0$
- $\pi^{-}(0,T] \times [0,a_{-}] = 1$

Define the substochastic kernel for $x \in \mathbb{R}^2$

$$\begin{aligned} Q_x^T(A) =& P(E \cap (X_T \in A) \,|\, X_0 = x) \\ =& P(E)\, P((X_T \in A) \,|\, \mathbb{1}\{E\} = 1, X_0 = x). \end{aligned}$$

It is seen that the conditional law of $X_T|E, X_0$ is equal to the law of

$$Y_T^x = xe^{-\nu T} + e^{-\nu U_+} \begin{pmatrix} c_{++} \\ c_{+-} \end{pmatrix} + e^{-\nu U_-} \begin{pmatrix} c_{-+} \\ c_{--}. \end{pmatrix} \tag{3.19}$$

where the two jump-times U_+, U_- are independent uniform variables on $[0,T]$. Since $C = \begin{pmatrix} c_{++} & c_{-+} \\ c_{+-} & c_{--} \end{pmatrix}$ is invertible and the law of $\left(e^{-\nu U_+}, e^{-\nu U_-}\right)$ is equivalent with the Lebesgue measure on $\left[e^{-\nu T}, 1\right]^2$ the Jacobi transformation theorem gives that the law of Y_T^x has density

$$f_x : u \mapsto |det\ C|^{-1}\, f \circ C^{-1}\left(u - xe^{-\nu T}\right),$$

where f is the density of $\left(e^{-\nu U_+}, e^{-\nu U_-}\right)$. The density is positive on the interior of its support

$$supp\,(Y_T^x) = e^{-\nu T} x + C\left(\left[e^{-\nu T}, 1\right]^2\right).$$

Since C is a homeomorphism, it is an open mapping. Thus we can find balls
$B(v_0, r) \subset B(v_0, 2r)$ contained in $C\left(\left[e^{-\nu T}, 1\right]^2\right)$ for all $T > 1$. Take now T so large that $e^{-\nu T} \sup_{v \in H} \|Cv\| < r$. For such T and $x \in H$ we have

$$\overline{B}(v_0, r) \subset e^{-\nu T} Cx + B(v_0, 2r) \subset supp\,(Y_T^x).$$

Note now that $H \times \overline{B}(v_0, r) \ni x, v \mapsto f_x(v)$ is continuous so positivity of the density gives $\inf_{x \in H, v \in \overline{B}(v_0, r)} f_x(v) := \alpha > 0$.

We therefore conclude that

$$Q_x^T(A) \geq P(E)\, \alpha m_2 (A \cap B(v_0, r)) \tag{3.20}$$

for all $x \in H$. This proves the desired. $\square$

We do now dispose of all ingredients to conclude the proof of Theorem 3.2.1.

Proof of Theorem 3.2.1. The combination of our two results and theorem 4.2 [33] gives that X is positive Harris. Theorem 6.1 from same reference gives the geometric ergodicity. □

3.4 Discussion And Further Research Topics

There are two natural ways to generalize this work. Either one could increase the number of populations or one could consider the more general Erlang weight functions instead of the exponential weight functions. Both generalizations lead to generators of a specific structure. Fix $d \in \mathbb{N}$ and let $a \in \mathbb{R}_+^d$, $A^c, A^J \in \mathbb{R}^{d\times d}$. Define x_+ as the componentwise rectifier $\left(x_+\right)_i = \max\left(0, x_i\right)$. Let Δ^J be the functional which maps a function $V : \mathbb{R}^d \to \mathbb{R}$ to a function $\mathbb{R}^d \to \mathbb{R}^d$ and is given by

$$\Delta_j^J V\left(x\right) = V\left(A^J e_j + x\right) - V\left(x\right),$$

where e_j is the j'th unit vector. The generator corresponding to either of the above mentioned generalizations can be shown to be on the form

$$AV\left(x\right) = \nabla V\left(x\right) A^c x + \Delta^J V\left(x\right)\left(x_+ + a\right). \tag{3.21}$$

Below we will study the Erlang generalization further. We keep two populations consisting of a single neuron each for notational convenience. We consider weight functions

$$\begin{aligned} h_{++}\left(t\right) &= \frac{c_{++}}{\eta_+!}\exp\left(-\nu_+ t\right), \quad & h_{+-}\left(t\right) &= \frac{c_{+-}}{\eta_+!}\exp\left(-\nu_+ t\right), \\ h_{-+}\left(t\right) &= \frac{c_{-+}}{\eta_-!}\exp\left(-\nu_- t\right), \quad & h_{--}\left(t\right) &= \frac{c_{--}}{\eta_-!}\exp\left(-\nu_- t\right), \end{aligned}$$

where $\eta_+, \eta_- \in \mathbb{N}$. The rate functions are kept unchanged. As mentioned in the introduction this induces a Markovian system of dimension

$d := \eta_+ + \eta_- + 2$ given by

$$dX_k^+(t) = -\nu_+ X_k^+(t)\,dt + X_{k+1}^+(t)\,dt \quad k < \eta_+, \tag{3.22}$$
$$dX_{\eta_+}^+(t) = -\nu_+ X_{\eta_+}^+(t)\,dt + Z^+ c_{++} + Z^- c_{-+}, \tag{3.23}$$
$$dX_k^-(t) = -\nu_- X_k^-(t)\,dt + X_{k+1}^-(t)\,dt \quad k < \eta_-, \tag{3.24}$$
$$dX_{\eta_-}^-(t) = -\nu_- X_{\eta_-}^-(t)\,dt + Z^+ c_{+-} + Z^- c_{--}. \tag{3.25}$$

Let us study the Lyapunov functions for this system among the quadratic forms

$$V : z \mapsto z^T P z, \quad z \in \mathbb{R}^d,\ P \in \mathbb{R}^{d\times d},\ P > 0. \tag{3.26}$$

Direct calculations as in the previous chapter shows that the generator of V is given by

$$GV(z) = z^T((A^c)^T P + PA^c)z + z_+^T(\left(A^J\right)^T P + PA^J)z_+ + L(x,y), \tag{3.27}$$

where $L(x,y)$ is polynomial of first degree and

$$A_{ij}^c = \begin{cases} -\nu_+ & i = j \le \eta_+ + 1 \\ -\nu_- & i = j > \eta_+ + 1 \\ 1 & i+1 = j \text{ and } i \ne \eta_+ + 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$A_{ij}^J = \begin{cases} c_{++} & j = 1 \text{ and } i = \eta_+ + 1 \\ c_{-+} & j = \eta_+ + 2 \text{ and } i = \eta_+ + 1 \\ c_{+-} & j = 1 \text{ and } i = \eta_+ + \eta_- + 2 \\ c_{--} & j = \eta_+ + 2 \text{ and } i = \eta_+ + \eta_- + 2 \\ 0 & \text{otherwise} \end{cases}.$$

Let us restrict our attention to the case $\nu := \nu_+ = \nu_-$ and $\eta := \eta_+ = \eta_-$. For $(z_1, z_{\eta+2}) \in [0,\infty)^2$ the resulting matrix $A = A^c + A^J$ is given by

$$A = \begin{array}{c@{}l}
\left[\begin{array}{cccccccccc}
-\nu & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \ddots & \ddots & & & & & & & \vdots \\
\vdots & & \ddots & \ddots & & & & & & \vdots \\
0 & \cdots & 0 & -\nu & 1 & 0 & \cdots & \cdots & \cdots & 0 \\
c_{++} & 0 & \cdots & 0 & -\nu & c_{-+} & 0 & \cdots & \cdots & 0 \\
0 & & & & & -\nu & 1 & & & \vdots \\
\vdots & & & & & & \ddots & \ddots & & \vdots \\
\vdots & & & & & & & -\nu & \ddots & 0 \\
0 & & & & & & & & \ddots & 1 \\
c_{+-} & 0 & \cdots & \cdots & 0 & c_{--} & 0 & \cdots & 0 & -\nu
\end{array}\right]
&
\begin{array}{l}
\left.\vphantom{\begin{array}{c}1\\1\\1\\1\\1\end{array}}\right\}\eta+1 \\
\left.\vphantom{\begin{array}{c}1\\1\\1\\1\\1\end{array}}\right\}\eta+1
\end{array}
\\
\begin{array}{l}\quad\underbrace{\hphantom{c_{+-}\;\; 0 \;\; \cdots \;\; \cdots \;\; 0}}_{\eta+1}\hphantom{c_{--}\;\; 0 \;\; \cdots \;\; 0 \;\; -\nu}\end{array} &
\end{array}$$

The sparse structure of A makes it possible to analyze the dynamics in this region. Define

$$\xi_i = \frac{1}{2}\left(c_{++} + c_{--} + \omega_i\right), \quad i = 1, 2 \tag{3.28}$$

where ω_i are the complex square roots of $(c_{++} - c_{--})^2 + 4c_{+-}c_{-+}$. Under assumption 3.1 ω_i is truly complex and the eigenvalues of A are given by

$$\lambda_{i,k} = -\nu + \zeta_{i,k}, \quad i = 1, 2,\ k = 1, ..., \eta + 1 \tag{3.29}$$

where $\zeta_{i,k}$ are the kth complex $(\eta+1)$-th root of ξ_i. With $\xi_i = |\xi_i| e^{i\theta_i}$ the largest real part of $\lambda_{i,k}$ may be written as $-\nu + |\xi_i|^{\frac{1}{\eta+1}} \cos\left(\frac{\theta_i}{\eta+1}\right)$. Thus A is Hurwitz if and only if

$$|\xi_i|^{\frac{1}{\eta+1}} \cos\left(\frac{\theta_i}{\eta+1}\right) < \nu. \tag{3.30}$$

This proves that X cannot exhibit transient behavior in a way where X^+, X^- stays in the first quadrant eventually. However, unlike the exponential case we do not have a simple expression for the Lyapunov function, making it a difficult task to extend the Lyapunov function from $z \in \mathbb{R}^d$ s.t. $(z_1, z_{\eta+2}) \in \mathbb{R}^2_+$ to the entire space $\mathbb{R}^d$.

It would be of interest to investigate whether assumption 3.1 and (3.30) are sufficient to establish positive Harris recurrence for the Erlang system. Also, it would be of interest to relate Erlang systems which are recurrent due to balancing criteria, with the systems that are known to produce oscillatory behavior, see [13].

We end this discussion by mentioning that it is possible to produce an algebraic expression of a Lyapunov function for A when $(z_1, z_{\eta+2}) \in \mathbb{R}_+^2$. It is well known that

$$P := \int_0^\infty \exp(A^T t) \exp(At) dt \tag{3.31}$$

is a positive definite matrix inducing a valid Lyapunov function for X given by $V(z) = z^T P z$. To find $\exp(At)$, define $B := (A + \nu I)$ and $C = \begin{pmatrix} c_{++} & c_{-+} \\ c_{+-} & c_{--} \end{pmatrix}$, and note that $\exp(At) = \exp(Bt)\exp(-\nu t)$. For a pair i, j define $i^* = 1$ if $i \leq \eta + 1$ and 2 otherwise, and j^* is defined likewise. Looking at the structure of B^n for $n \in \mathbb{N}$ it may be shown that the (i, j)th entry of $\exp(Bt)$ may be written as the (i^*, j^*)th entry of

$$C^{r_1} \sum_{l=1}^{\infty} \frac{C^l}{(l\eta + r_2)!} t^{l\eta + r_2}, \tag{3.32}$$

where $r_1, r_2 < \eta$ are appropriately chosen depending on i, j. The power series corresponds to analytic functions[2] f_{ij} which may be inserted into Sylvester's formula to derive an expression for $\exp(Bt)$. One may insert such an expression back into (3.31) to obtain the Lyapunov function.

[2]More precisely the functions are roots composed with generalized hyperbolic functions.

Chapter 4

Stability for ADHP's With a Refractory Period

4.1 Introduction

In this chapter we discuss stability of the age dependent Hawkes process. The main assumption is a post-jump bound on the intensity, corresponding to a strong self-inhibition for a short time interval after a spike. This models the refractory period. We do not impose any a priori bounds on the intensities. Within this sub-model, we are able to prove stability properties for the $N-$dimensional Hawkes process. The results we obtain are similar to what has been shown for ordinary nonlinear Hawkes processes in [5] and recently in [10]. This last paper is however entirely devoted to the study of weight functions which are of compact support giving rise to explicit regeneration points when the process comes back to the all zero measure. Compared to these studies, it turns out that the natural self-inhibition by the age processes eliminates the need of controlling the Lipschitz constant of ψ, and we do not need any restriction on the support of the weight functions. We also discuss which starting conditions (that is, which form of an initial process) will ensure coupling to the invariant process. These results are collected within our first main theorem, Theorem 4.3.1.

During the proof of the stability properties, other interesting properties of the model are discussed, such as a nice-behaving domination of the intensities (Lemma 4.4.1).

4.2 Setup For This Chapter

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

$$\theta^r Z(C) := Z(r + C) := Z(\{r + x : x \in C\}), \tag{4.1}$$

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

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

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

Moreover, for simplicity, we may and will take a constant $L \geq \max(L_1, ..., L_n, c_{\psi^1}, ..., c_{\psi^n})$ large enough so that $L \geq 1$. Then (2.12) and (5.5) imply the simpler inequalities for ψ^i,

$$\psi^i\,(y, b) \leq L(1 + |y|) \quad \forall y \in \mathbb{R}, b \in \mathbb{R}_+. \tag{4.2}$$

$$\left|\psi^i\,(x,a) - \psi^i\,(x',a')\right| \leq \begin{cases} L\,|x - x'|\,, & \text{if } a = a' \\ L\,(\max(|x'|\,,|x|) + 1)\,, & \text{if } a \neq a' \end{cases}. \tag{4.3}$$

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

Assumption 4.1.
1. There exists K and $\delta > 0$ such that

$$\psi^i\,(x,a) \leq K \text{ for all } 1 \leq i \leq N, a \in [0,\delta], x \in \mathbb{R}. \tag{4.4}$$

2. There exists $x^*, a^*, c > 0$ such that for all $|x| \leq x^*, a \geq a^*$ and for all $1 \leq i \leq N$,
$$\psi^i(x, a) \geq c > 0. \tag{4.5}$$
3. We suppose that
$$[0, \infty[\ni t \mapsto \overline{h}_{ij}(t) \in \mathcal{L}^1 \cap \mathcal{L}^2 \text{ and } [0, \infty[\ni t \mapsto t h_{ij}(t) \in \mathcal{L}^1. \tag{4.6}$$

$\circ$

Notice that (4.6) implies that
$$\overline{h} := \sum_{i,j=1}^{N} \overline{h}_{ij} \in \mathcal{L}^1 \cap \mathcal{L}^2 \tag{4.7}$$
is a decreasing function that dominates h_{ij} for all $i, j \leq N$.

Remark 4.2.1.
The existence of K, δ in (4.4) excludes instantaneous bursting by imposing a bound on the immediate post-jump intensity. Moreover, the existence of x^*, a^*, c in (4.5) ensures that no unit will eventually stop spiking. A main example of rate functions that satisfy this assumption are those inducing absolute refractory periods as given in example 2.3.3.

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

Throughout this chapter we use the following notation. For $K > 0$ as in (4.4) above, we denote the PRMs
$$\pi_K(ds) := \pi(ds, [0, K]), \quad \pi_K^i(ds) := \pi^i(ds, [0, K]) \text{ and } \pi_{NK} := \sum_{i=1}^{N} \pi_K^i. \tag{4.8}$$

Example 4.2.2 (Hawkes processes with Erlang weight functions). Weight functions given by Erlang kernels are widely used in the modeling literature to describe delay in the information transmission. They are given by
$$h_{ij}(t) = c_{ij} t^{n_{ij}} e^{-\nu_{ij} t}, t \geq 0,$$

where $c_{ij} \in \mathbb{R}, \nu_{ij} > 0$ and $n_{ij} \in \mathbb{N} \cup \{0\}$ are fixed constants. The order of the delay is given by n_{ij}. The delay of the influence of particle j on particle i is distributed and taking its maximum absolute value at n_{ij}/ν_{ij} time units back in time. The sign of c_{ij} indicates if the influence is inhibitory or excitatory, and the absolute value of c_{ij} scales how strong the influence is. All h_{ij} clearly satisfy (4.6). ◦

The main result of this chapter shows existence of a unique stationary $N-$dimensional age dependent Hawkes process following the dynamics of (2.15). In order to state the result, we first introduce the notion of *compatibility* (see e.g. [5]). Let $M_{\mathbb{R}_-}$ be the set of all bounded measures defined on $\mathbb{R}_-$ equipped with the weak-hat metric and the associated Borel $\sigma-$algebra $\mathcal{M}_{\mathbb{R}_-}$ (see Appendix for details). We shall say that Z is *compatible* (to $\pi^1, \ldots, \pi^N$) if there is a measurable map $H : M_{\mathbb{R}_-}^N \to M_{\mathbb{R}_-}$ such that for all $t \in \mathbb{R}$,

$$\left(\theta^t Z\right)_{|\mathbb{R}_-} = H\left(\left(\theta^t \pi^1, \ldots, \theta^t \pi^N\right)_{|\mathbb{R}_-}\right). \tag{4.9}$$

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

Remark 4.2.3.
Note that if $Z^1, \ldots Z^n$ are compatible random measures, then $\left(Z^1, \ldots Z^n\right)$ is a stationary and ergodic n-tuple of random measures. ◦

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

$$\begin{aligned} Z^i(t_1, t_2] &= \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}\left\{z \leq \psi^i\left(X_s^i, A_s^i\right)\right\} \pi^i(ds, dz), \\ X_t^i &= \sum_{j=1}^N \int_{-\infty}^{t-} h_{ij}(t-s)\, Z^j(ds) \quad t \in \mathbb{R}, \\ A_{t_2}^i - A_{t_1}^i &= t_2 - t_1 - \int_{t_1}^{t_2-} A_t^i\, Z^i(dt), \end{aligned} \tag{4.10}$$

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

4.3 Main Result

The main result of this chapter is

Theorem 4.3.1.
Grant Assumption 4.1.

1. *There exists an N-dimensional age dependent Hawkes process Z on $\mathbb{R}$, compatible to $\left(\pi^1, \ldots, \pi^N\right)$.*

2. *Let $\check{Z}$ be another $N-$dimensional age dependent Hawkes process with the same weight functions $(h_{ij})_{i,j\leq N}$ and driven by the same PRMs $\left(\pi^1, \ldots, \pi^N\right)$, following the dynamics (2.15), that is; starting at time 0 with arbitrary initial ages $(\check{A}^i_0)_{i\leq N}$ and initial signals $\left(R^i\right)_{i\leq N}$ such that*
$$\mathbb{E}\int_0^\infty \left|R^i_s\right| ds < \infty \tag{4.11}$$
for all $1 \leq i \leq N$. Then almost surely, $\check{Z}$ and Z couple eventually, i.e.,
$$\exists\, t_0 \in \mathbb{R}_+ : \quad \check{Z}_{|[t_0,\infty)} = Z_{|[t_0,\infty)}.$$

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

$\circ$

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

Corollary 4.3.2.
Grant Assumption 4.1. Let (Z, X, A) be the stationary age dependent Hawkes process and let $\check{Z}$ be as in Item 2. of Theorem 4.3.1. Let moreover $f : M_T \to \mathbb{R}$ be any measurable function such that
$$\mu(f) := \mathbb{E}f((Z_{|\,(-T,0]})) < \infty. \tag{4.12}$$

Then

$$\frac{1}{t}\int_0^t f((\theta^s \check{Z}_{|(-T,0]}))ds = \frac{1}{t}\int_0^t f((\check{Z}_{|(s-T,s]}))ds \to \mu(f) \tag{4.13}$$

almost surely, as $t \to \infty$. ∘

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

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

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

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

which implies (4.13). □

4.4 Proof of Main Result

This section is devoted to the proof of Theorem 4.3.1, and we will thus work under Assumption 4.1. The proof follows the ideas of Theorem 4 in [5]. The proof of the existence part relies on the Picard iteration

$$\begin{aligned} X_t^{n,i} &= \sum_{j=1}^{N} \int_{-\infty}^{t-} h_{ij}(t-s)\, Z^{n-1,j}(ds), \\ Z^{n,i}(t_1, t_2] &= \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}\left\{z \leq \psi^i\left(X_s^{n,i}, A_s^{n,i}\right)\right\} \pi^i(ds, dz), \quad t_1 < t_2 \in \mathbb{R}, \end{aligned} \tag{4.14}$$

where $A^{n,j}$ is the age process of $Z^{n,j}$. We initialize the iteration with $Z^{0,i} \equiv \pi_K^i, X^0 \equiv 0$. Before we can prove convergence of this iteration we need to address the following issues.

- We need to produce an integrable intensity $\hat{\lambda}$ that a priori dominates the intensities $\psi^i(X_t^{n,i}, A_t^{n,i})$. This is done in Proposition 4.4.3.
- Using the Lipschitz part of (4.3), we will construct events $E_t \in \mathcal{F}_t$ for all $t \in \mathbb{R}$ such that $A_t^{n,i} = A_t^{n+1,i}$ on E_t, for all $n \in \mathbb{N}$, $i \leq N$. This is done in Lemma 4.4.4.

- We need to ensure that the $i-$th iteration is well defined, i.e. for a given $X^{n,i}$ there exist $Z^{n,i}$, $A^{n,i}$ such that (4.14) is satisfied. This is done in Lemma 4.4.5.

Finally we combine these results to complete the proof of Theorem 4.3.1. We start with the following useful result which provides bounds on the intensities.

Lemma 4.4.1.
Let K, δ be the constants from (4.4). Let X be a predictable stochastic process and assume that (Z, A) solves the system

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

where A is the age process of Z and where ψ satisfies (4.3) and (4.4). Suppose moreover that (4.6) is satisfied. Then almost surely, for any $1 \leq i, j \leq N$, $t_1 \leq t_2$,

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

is well-defined and

$$\left|Y_{ij}\left(t_1, t_2\right)\right| \leq \sum_{k=0}^{\infty} \overline{h}_{ij}\left(t_2 - t_1 + A_{t_1} + k\delta\right) + \int_{-\infty}^{t_1 - A_{t_1}} \overline{h}_{ij}\left(t_2 - s\right) \pi_K\left(ds\right). \tag{4.15}$$

Moreover,

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

for all t. ○

Corollary 4.4.2.
If we suppose in addition that $\bar{h}(0) < \infty$, then

$$\mathbb{E} Y_{ij}\left(t, t\right) \leq K \int_0^{\infty} \bar{h}(u) du + \sum_{k \geq 0} \bar{h}(k\delta) < \infty. \tag{4.16}$$

○

Proof of Lemma 4.4.1. For any $i, j \leq N$, $t \leq t_2$ we have

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

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

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

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

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

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

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

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

$\square$

We now construct the dominating intensity, as mentioned in the start of the section. Recall that $L \geq 1$ is the Lipschitz constant appearing in (4.3) and let K, δ be the constants from (4.4), we suppose w.l.o.g. that $K \geq c$, where c is the lower bound from (4.5).

Proposition 4.4.3.
Let $C \geq \max\left\{1 + \sum_{k\geq 1} \bar{h}(k\delta), K\right\}$. There exists a compatible process $(\hat{Z}, \hat{A}, \hat{\lambda})$ which is defined for any $t \in \mathbb{R}$ by

$$\hat{\lambda}_t = L\left(C + \int_{-\infty}^{t-} \bar{h}(t-s)\pi_{NK}(ds) + \bar{h}(\hat{A}_t)\right), \tag{4.17}$$

where

$$\hat{Z}(t_1, t_2] = \sum_{i=1}^{N} \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}\{z \leq \hat{\lambda}_s\}\pi^i(ds, dz) \tag{4.18}$$

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

$$\mathbb{E}(\hat{\lambda}_t) < \infty. \tag{4.19}$$

$\circ$

Proof. By construction, $\hat{\lambda}_t \geq K$ for all t, and therefore, any jump time τ of π^i_K is also a jump of $\hat{Z}$. Hence, at τ, the age process $\hat{A}_t$ is reset to 0. It is therefore possible to construct a unique solution to (4.17) on $t \in (\tau, \infty)$. This solution is non-exploding since the process is stochastically dominated by a classical linear Hawkes process Z' having intensity $L\left(C + \int_{-\infty}^{\tau-} \bar{h}(t-s)\pi_{NK}(ds) + 2\int_{\tau}^{t-} \bar{h}(t-s)Z'(ds)\right)$ which is non-exploding by Proposition 2.3.2 since $\bar{h} \in \mathcal{L}^1$. A solution on the entire real line may be constructed by pasting together the solutions constructed in between the successive jump times of π_{NK}. It is unique and compatible by construction.

It remains to prove that $\mathbb{E}(\hat{\lambda}_t) < \infty$. Due to stationarity, it is sufficient to prove that $\mathbb{E}\bar{h}(\hat{A}_0) < \infty$. Also from stationarity, writing T_1 for the first jump time of $\hat{Z}$ after time 0, it follows that $\mathcal{L}(T_1) = \mathcal{L}(\hat{A}_0)$. It follows from Lemma 2.1.6 in the Appendix that

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

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

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

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

We now proceed to the construction of events E_t which a priori will serve by coupling the age processes in the Picard iteration. Indeed, Assumption (4.5) will enable us to construct common jumps for any two point processes Z^1, Z^2 having intensity $\psi(\tilde{X}_t^1, A_t^1)$ and $\psi(\tilde{X}_t^2, A_t^2)$, where A_t^i is the age process of Z^i, and $\tilde{X}_t^i$ is a predictable process such that

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

for $i = 1, 2$.

Fix some $p > a^*$ such that

$$\sum_{k\geq 1} L\bar{h}(p+k\delta) < \frac{x^*}{3N}, \tag{4.20}$$

where a^* and x^* are given in (4.5), and fix some $M > LC$ where L and C are as in (4.17). Then necessarily $M \geq K \geq c$. Introduce for all $t \in \mathbb{R}$ the events

$$\begin{aligned}&E_t^1 := \\ &\{\pi^1(ds,[0,c]) \text{ has a unique jump } \tau^1 \text{ in } (t-2Np+p, t-2Np+2p])\} \\ &\cap \bigcap_{j=1}^N \left\{\int_{t-2Np}^{\tau^1-}\int_{\mathbb{R}_+} \mathbb{1}\{z \leq M\}\pi^j(ds,dz) = 0\right\} \\ &\cap \bigcap_{j=1}^N \left\{\int_{\tau^1}^{t-2Np+2p}\int_{\mathbb{R}_+} \mathbb{1}\{z \leq M+2L\bar{h}(s-\tau^1)\}\pi^j(ds,dz) = 0\right\}\end{aligned} \tag{4.21}$$

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

$$\begin{aligned}
&E_t^i := \\
&\{\pi^i(ds, [0,c]) \text{ has a unique jump } \tau^i \text{ in } (t-2Np+2(i-1)p+p, t-2Np+2ip])\} \\
&\cap \bigcap_{j=1}^{N} \left\{ \int_{t-2Np+2(i-1)p}^{\tau^i -} \int_{\mathbb{R}_+} \mathbb{1}\{z \le M + 2L\bar{h}(s-(t-2Np+2(i-1)p))\}\pi^j(ds,dz) = 0 \right\} \\
&\cap \bigcap_{j=1}^{N} \left\{ \int_{\tau^i}^{t-2Np+2ip} \int_{\mathbb{R}_+} \mathbb{1}\{z \le M + 2L\bar{h}(s-\tau^i)\}\pi^j(ds,dz) = 0 \right\}, \qquad (4.22)
\end{aligned}$$

where the constant c is given in (4.5). This event splits the interval $(t-2Np, t)$ up in intervals of length $2p$, where the ith truncated PRM has exactly one jump in the second part, and no other events (of truncated PRMs) occur.

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

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

and put

$$E_t := \bigcap_{i=0}^{N} E_t^i. \tag{4.23}$$

The event E_{2Np} is illustrated in Figure 4.1, for $N = 2$ and $\overline{h}\,(t) \approx t^{-0.4}$. The grey area is the relevant part for the truncated PRMs.

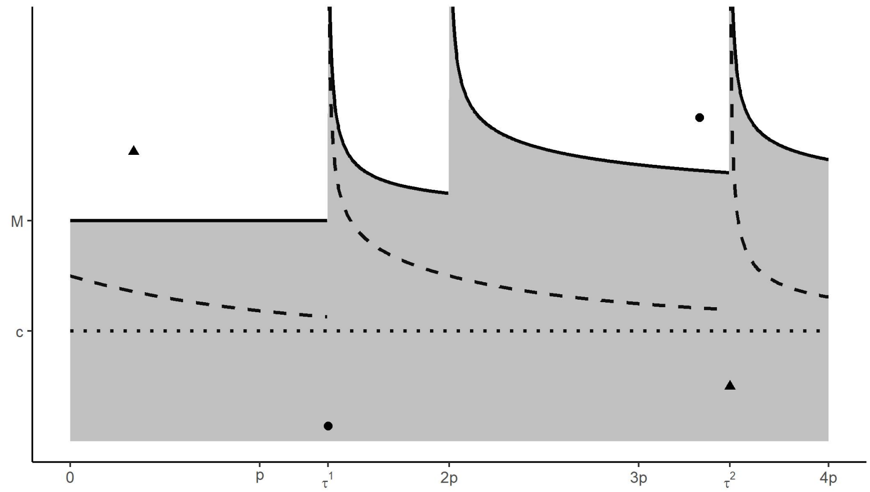

Figure 4.1: An illustration of the event E_{2Np} with $N = 2$ and $\overline{h}(t) \approx t^{-0.4}$. The figure shows a superposition of $\pi^1(\bullet), \pi^2(\blacktriangle)$, including the jump times τ^1 and τ^2, and three curves: **The dotted curve** is the constant c. **The dashed curve** is the intensity process $\hat{\lambda}$. **The solid curve** is enclosing the area (in grey) of the plane that is relevant to the event E_{2Np}, and it is given by M for $0 < s \leq \tau^1, M + 2L\overline{h}(s - \tau^1)$ for $\tau^1 < s \leq 2p$, $M + 2L\overline{h}(s - 2p)$ for $2p < s \leq \tau^2$ and $M + 2L\overline{h}(s - \tau^2)$ for $\tau^2 < s \leq 4p$.

The main feature of the event E_{2Np} with $N = 2$ is the fact that the process is forced to have some regeneration events during the intervals $]p, 2p]$ and $]3p, 4p]$. Indeed, on these intervals, the corresponding age processes will have values larger than $p > a^*$, and the associated memory processes will be bounded by x^*, such that we can use (4.5).

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

$$Y_t := \int_{-\infty}^{t-\hat{A}_t} \overline{h}(t-s)\pi_{NK}(ds) + \sum_{k=0}^{\infty} \overline{h}(\hat{A}_t + k\delta). \tag{4.24}$$

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

Lemma 4.4.4.
On E_t, for all $1 \leq i \leq N$, each measure $\pi^i(ds, [0, c])$ has a jump at time $\tau^i \in (t - 2Np, t)$ such that

$$\int_{t-2Np}^{t} \int_{\mathbb{R}_+} \mathbb{1}\{s \neq \tau^i, z \leq \hat{\lambda}_s\}\pi^i(ds, dz) = 0, \tag{4.25}$$

$$\hat{\lambda}_{\tau^i} \leq \hat{\lambda}_{t-2Np} + \frac{2i}{3N} x^*, \tag{4.26}$$

$$Y_{\tau^i} \leq \frac{2i}{3N} x^*. \tag{4.27}$$

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

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

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

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

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

as well as (4.26) and (4.27) hold for $i \in \{0, \ldots, j\}$ in the event E_t. The induction start is trivial, so assume that the assertion is true up to $j-1$. Notice that by the induction assumption

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

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

□

The next result ensures that for a well-behaving process X^i there exist couples $\left(Z^i, A^i\right)$ such that Z^i has intensity $\psi^i\left(X^i, A^i\right)$ and A^i is the age of Z^i. The proof relies on a Picard iteration of (4.10) that alternately updates $\left(X^i\right)_{i \leq N}$ and $\left(Z^i, A^i\right)_{i \leq N}$.

Lemma 4.4.5.
Let $(\hat{Z}, \hat{A}, \hat{\lambda})$ be as in Proposition 4.4.3 and let $\left(X_t^i\right)_{t \in \mathbb{R}}, 1 \leq i \leq N$,

be compatible and predictable stochastic processes satisfying that almost surely,

$$|X_t^i| \leq Y_t, \tag{4.28}$$

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

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

$$Z^i(B) = \int_B \int_0^\infty \mathbb{1}\left\{z \leq \psi^i\left(X_s^i, A_s^i\right)\right\} \pi^i(ds, dz), \quad \forall B \in \mathcal{B}(\mathbb{R}), \tag{4.29}$$

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

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

$$Z^{n,i}(t_1, t_2] = \int_{t_1}^{t_2} \int_0^\infty \mathbb{1}\left\{z \leq \psi^i\left(X_s^i, A_s^{n-1,i}\right)\right\} \pi^i(ds, dz), \quad t_1 < t_2 \in \mathbb{R}, \tag{4.30}$$

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

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

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

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

$$\underline{Z}^i[t] = \liminf_n Z^{ni}[t], \; \overline{Z}^i[t] = \limsup_n Z^{ni}[t]$$

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

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

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

To prove this claim, consider the event

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

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

We now prove that $P(G_0) > 0$ by showing that $E_0 \subset G_0$, where E_0 was defined in (4.23) above (that is, we choose $t = 0$).

The assumption $|X_t^i| \leq Y_t$ implies that $\lambda_t^{n,i} \leq \hat{\lambda}_t$ for all i, n and t. Lemma 4.4.4 implies that on E_t we have $\hat{A}_{\tau^i} \geq a^*$, and therefore also $A_{\tau^i}^{n,i} \geq a^*$. Moreover, (4.27) implies that $|X_{\tau^i}^i| \leq x^*$. Therefore, (4.5) implies

$$\lambda_{\tau^i}^{n,i} \geq c \tag{4.31}$$

for all n, i. As a consequence, at time τ^i, all Z^{ni} have a common jump. From (4.25) it follows that $Z^{ni}(\tau^i, 0) = 0$, and therefore, $A_0^{n,i} = -\tau^i$. In particular, they are all equal. We may now conclude that on E_0, $Z_{|\mathbb{R}_+}^{n,i}$ is a constant sequence over n, for all i. In particular, we have $\tilde{Z}(0, \infty) = 0$. To conclude the proof, we have proven that $E_0 \subset G_0$, and thus

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

implying the result. □

We are now ready to prove Theorem 4.3.1.

Proof of Theorem 4.3.1. The proof follows the ideas of the proof of Theorem 4 in [5]. First we construct a stationary solution to (4.10). For this sake we consider the Picard iteration

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

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

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

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

$$L(1 + |X^{n,i}_t|) \leq \hat{\lambda}_t \quad \forall t \in \mathbb{R}, \tag{4.32}$$

for all n, i, where $\hat{\lambda}$ is defined in (4.17) above. The induction start is trivial. Suppose now that the assertion holds for $n-1$. We apply Lemma 4.4.5 with $X^i = X^{n,i}$, and show that the conditions of this Lemma are met, then well-posedness, ergodicity and stationarity of $Z^{n,i}, A^{n,i}$ follow.

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

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

We apply Lemma 4.4.1 to each of the N terms within the above sum and obtain

$$\int_{-\infty}^{t-A^{n-1,j}_t} h_{ij}(t-s) Z^{n-1,j}(ds) \leq \sum_{k \geq 0} \bar{h}_{ij}(A^{n-1,j}_t + k\delta) + \int_{-\infty}^{t-A^{n-1,j}_t} \bar{h}_{ij}(t-s) \pi^j_K(ds). \tag{4.33}$$

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

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

which is (4.28). Finally, since A^{n-1}, Z^{n-1} are compatible, it is straightforward to show that Z^n is compatible as well.

Define now

$$\underline{\lambda}^i_t = \liminf_{n \to \infty} \psi^i\left(X^{n,i}_t, A^{n,i}_t\right), \quad \overline{\lambda}^i_t = \limsup_{n \to \infty} \psi^i\left(X^{n,i}_t, A^{n,i}_t\right), 1 \leq i \leq N.$$

Note that by (4.2) and (4.32), $\underline{\lambda}^i_t \leq \overline{\lambda}^i_t \leq \limsup_{n \to \infty} L(1 + |X^{n,i}_t|) \leq \hat{\lambda}_t$. So almost surely, $\underline{\lambda}^i, \overline{\lambda}^i$ have finite sample paths. Note also that they are limits of predictable processes (see Lemma 2.1.1 in Appendix), and thus they are predictable as well. Define also

$$\tilde{Z}^i[t] = \limsup_n Z^{n,i}[t] - \liminf_n Z^{n,i}[t] = \pi^i\left(\{t\} \times (\underline{\lambda}^i_t, \overline{\lambda}^i_t]\right),$$

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

$$\tilde{Z} = \sum_{j=1}^{N} \tilde{Z}^j \tag{4.34}$$

is almost surely the trivial measure. It will follow that $Z^{n,j}$, and thus also $X^{n,i}, A^{n,i}$ converge. Moreover, it is straight forward to check that the limit variables solve (4.10).

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

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

Consider the events E_t defined in (4.23) above as well as

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

Using the functionals obtained previously, it follows that

$$\{\tilde{Z}\,(t,\infty) = 0\} = \{\theta^t(\pi^i)_{i=1}^N \in V\}$$

for some $V \in \mathcal{M}$ and thus $\{\tilde{Z}\,(t,\infty) = 0$ infinitely often$\}$ is an invariant set, and thus also a 0/1 event. As before this implies that $P\left(\tilde{Z}\,(\mathbb{R}) = 0\right) = 1$ if $P\,(G_0 \cap E_0) > 0$.

To prove that $P(G_0 \cap E_0) > 0$, note that we have $\lambda^{n,i}_t = \psi(X^{n,i}_t, A^{n,i}_t) \leq \hat{\lambda}_t$ and $|X^{n,i}_t| \leq Y_t$. Therefore, the same arguments as those exposed in the proof of Lemma 4.4.5, show that on E_0, we have $A^{n,i}_0 = A^{m,i}_0$; for all n, m and i, that is, the age variables are all equal at time 0. Moreover, on G_0, either no jumps happen any more, or they happen conjointly, and so the Lipschitz criterion (4.3) ensures the bound

$$\left|\overline{\lambda}^i_t - \underline{\lambda}^i_t\right| \leq L \lim_{n\to\infty} \sup_{m,k\geq n} \left|X^m_t - X^k_t\right| \leq L \int_{-\infty}^{0-} h\,(t-s)\,\tilde{Z}(ds) := \bar{X}_t,$$

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

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

Note that Lemma 2.1.6 in Appendix reveals that the compensator of the integral-sum above is

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

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

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

To prove that the right hand side is positive, it suffices to show that the double integral inside the exponential is almost surely finite. Notice that by construction, $\tilde{Z}_t \leq \hat{Z}_t$, and recall from Proposition 4.4.3 that $\hat{\lambda}$ is stationary with $\mathbb{E}\hat{\lambda}_0 < \infty$. After taking expectation,

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

This proves the desired result.

We now prove the coupling part. This will be done in two steps. First suppose that $|R_t|$ is bounded by a constant C_R. We suppose w.l.o.g. that $\hat{\lambda}$ defined in (4.17) is such that also $C \geq C_R$.

Let $\left(\left(\check{Z}^i\right)_{i \leq N}, \left(\check{X}^i\right), \left(\check{A}^i\right)_{i \leq N}\right)$ be the N-dimensional age dependent Hawkes process with initial conditions $\left(\check{A}_0^i\right), \left(R^i\right)$ and define the process

$\check{\lambda}^i_t := \psi(\check{X}^i_t, \check{A}^i_t)$. Then we clearly have that

$$\check{\lambda}^i_0 \leq \hat{\lambda}_0$$

for all i, and it can be shown inductively over the successive jumps of $\hat{Z}$, using Lemma 4.4.1, that this inequality is preserved over time, that is,

$$\check{\lambda}^i_t \leq \hat{\lambda}_t \tag{4.35}$$

for all $t \geq 0$. Now introduce

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

and put

$$E'_t := E_t \cap \check{E}^0_t. \tag{4.36}$$

Let $\left(\tau^i\right)_{i \leq N}$ be the jump times from Lemma 4.4.4. We necessarily have that on E'_t,

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

for all $1 \leq i \leq N$. As a consequence, (4.27) implies that

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

Due to (4.35), we have $\check{A}^i_{\tau^i} \geq \hat{A}_{\tau^i} \geq a^*$. Therefore, using (4.5), we conclude that

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

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

$$\begin{aligned} \sum_{i=1}^N \left|\psi^i\left(X^i_t, A^i_t\right) - \psi^i\left(\check{X}^i_t, \check{A}^i_t\right)\right| &\leq L \sum_{i=1}^N \left|X^i_t - \check{X}^i_t\right| \\ &\leq L \left(\int_{-\infty}^{0-} h\left(t-s\right) Z(ds) + |R_t| + \int_0^{t-} h\left(t-s\right) \tilde{Z}(ds) \right), \end{aligned}$$

which holds on E'_t.

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

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

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

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

As a consequence,

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

where

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

is stationary and ergodic, and where

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

Clearly, $D_t \to 0$ as $t \to \infty$ almost surely. Now apply Lemma 7.1.1 in Appendix with $U_t := \mathbb{1}E_t e^{-LC_t}$, $r_t := \mathbb{1}E_t e^{-LC_t} - \mathbb{1}E_t' e^{-LC_t - LD_t}$. Clearly, U_t is ergodic and satisfies $P(U_t > 0) > 0$. To see that $r_t \to 0$ almost surely as $t \to \infty$, it suffices to prove that $\mathbb{1}E_t' - \mathbb{1}E_t \to 0$ almost surely, as $t \to \infty$, which is equivalent to proving that $\mathbb{1}\check{E}_t^0 \to 1$ almost surely. But this follows from

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

which follows from (4.11).

This finishes the first part of the proof of the coupling result. Finally, suppose that the initial process only satisfies (4.11). Take then a sequence $(\check{Z}^m, \check{X}^m, \check{A}^m)$ of $N-$dimensional age dependent Hawkes processes with starting condition $(\check{A}^i_0)_{i\le n}$ and initial processes $(-m \vee R^i \wedge m)_{i\le N}$. Write $\check{\lambda}^{m,i}_t = \psi^i(\check{X}^{m,i}_t, \check{A}^{m,i}_t)$ for the associated intensity and define $\check{\lambda}^{m\wedge m+1,i}_t := \check{\lambda}^{m,i}_t \wedge \check{\lambda}^{m+1,i}_t$. Denote $R^{m,i}_t := -m \vee R^i_t \wedge m$. As in the proof of Proposition 2.3.2, we have that

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

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

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

implying that almost surely, $\check{Z}^m = \check{Z}$ on $\mathbb{R}_+$ for sufficiently large m. Since $\check{Z}^m$ and Z couple eventually almost surely, this proves the coupling part. It remains to prove uniqueness of the stationary solution. Let (Z', X', A') be another age dependent Hawkes process on $t \in \mathbb{R}$ compatible to $\pi^1, \ldots, \pi^N$. Lemma 4.4.1 gives the inequality

$$\left|X'^i_t\right| \le \sum_{j=1}^N \left(\sum_{k\ge 0} \bar{h}_{ij}(A'^j_t + k\delta) + \int_{-\infty}^{t-A'^j_t} \bar{h}_{ij}(t-s)\pi^j_K(ds) \right).$$

Let τ be a jump of $\hat{Z}$. Using the above inequality, it is shown inductively over future jumps of $\hat{Z}$ that $\hat{\lambda}_t \ge \left|\lambda'^i\right|$ for all $t \in]\tau, \infty[$. Thus it follows that $\hat{\lambda}_t \ge \left|\lambda'^i\right|$ for all $t \in \mathbb{R}$. Note that the (Z', X', A') system may be written in terms of Definition 2.3.1 with initial signals $R'^i_t := \sum_{j=1}^N \int_\infty^0 h_{ij}(t-s)\, dZ'^j_s$. The same arguments as in (4.35) give

$$\mathbb{E} \int_0^\infty \left|R'^i_s\right| ds \le \mathbb{E} \int_0^\infty \int_{-\infty}^{0-} h(s-u)\, \hat{Z}(du)\, ds < \infty.$$

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

$$P\left(\exists\, t_0 \in \mathbb{R} : \quad Z'_{|[t_0,\infty)} = Z_{|[t_0,\infty)}\right) = P \bigcup_{n=-\infty}^{\infty} \left(Z'_{|[n,\infty)} = Z_{|[n,\infty)}\right) = 1. \tag{4.37}$$

Since Z and Z' are both compatible, it follows that (Z, Z') is compatible and therefore also stationary. Thus, the events $\left\{Z'_{|[n,\infty)} = Z_{|[n,\infty)}\right\}$ have the same probability for all $n \in \mathbb{Z}$, and from (4.37) it follows that the probability is equal to 1. This proves that $Z = Z'$ almost surely. $\square$

4.5 Age Dependent Hawkes Processes With Erlang Weight Functions

Here we show how the above results can be applied for weight functions given by Erlang kernels as in Example 4.2.2, and consider a one-dimensional ($N = 1$) age dependent Hawkes process (Z, X, A), solution of

$$\begin{aligned} Z_t &= \int_0^t \int_0^\infty \mathbb{1}\left\{z \leq \psi\left(X_s, A_s\right)\right\} \pi\left(ds, dz\right), \\ X_t &= \int_0^{t-} h\left(t-s\right) Z(ds) + R_t, \\ A_t &= A_0 + t - \int_0^{t-} A_s Z(ds), \end{aligned} \tag{4.38}$$

where

$$h(t) = b\frac{t^n}{n!}e^{-\nu t}, \tag{4.39}$$

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

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

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

The process (X_t, A_t) is not Markov, but it is well-known (see e.g. [13]) that it can be completed to a Markovian system $(X_t^{(0)} := X_t, X_t^{(1)}, \ldots, X_t^{(n)}, A_t)$, by introducing the auxiliary processes

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

for all $0 \le k \le n$. By [13], these satisfy the system of coupled differential equations, driven by the PRM π, given by

$$
\begin{aligned}
dX_{t+}^{(k)} &= -\nu X_{t+}^{(k)} dt + X_{t+}^{(k+1)} dt, \quad 0 \le k < n, && (4.40)\\
dX_{t+}^{(n)} &= -\nu X_{t+}^{(n)} dt + b \int_0^\infty \mathbb{1}\{z \le \psi(X_t^{(0)}, A_t)\} \pi(dt, dz), && (4.41)
\end{aligned}
$$

and

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

for $t \ge 0$. Evidently, h satisfies (4.6). We suppose that $\psi(x,a)$ satisfies (4.3) and we strengthen (4.5) to the following assumption.

Assumption 4.2.
$\psi(x,a)$ is continuous in x and a; and $\psi(x,a) \ge c > 0$ for all x, a with $a \ge a^*$. $\circ$

Then the following result strengthens Theorem 4.3.1 in this Markovian setting.

Theorem 4.5.1.
Grant Assumptions 4.1 and 4.2. Then the process $(X_t^{(0)}, X_t^{(1)}, \dots, X_t^{(n)}, A_t)$ *is positively recurrent in the sense of Harris having unique invariant probability measure* μ. $\circ$

Proof. Step 1. By Lemma 4.4.1 and Corollary 4.4.2, $t \mapsto \mathbb{E}(\lambda_t) = \mathbb{E}(\psi(X_t, A_t))$ and $t \mapsto \mathbb{E}(|X_t|) = \mathbb{E}(|X_t^{(0)}|)$ are bounded on $\mathbb{R}$. By the same argument, also $t \mapsto \mathbb{E}(|X_t^{(k)}|)$ is bounded for $1 \le k \le n$. Therefore, $(X_t^{(0)}, X_t^{(1)}, \dots, X_t^{(n)})$ is a $1-$ultimately bounded Feller process (the Feller property follows from the continuity of ψ), see e.g. [35].

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

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

where

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

and

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

implying inequality (6) of [35]. Thus, by Theorem 1 of [35], the Markov processes $(X_t^{(0)}, \ldots, X_t^{(n)}, A_t)_{t\geq 0}$ possesses invariant probability measures μ (not necessarily unique ones).[1]

Step 2. We shall now use the coupling property proved in Theorem 4.3.1 to prove uniqueness of the invariant measure μ. In what follows we write (Z, X, A) for the stationary version of (4.38), which exists according to Theorem 4.3.1. Moreover, we write $(\tilde{Z}, \tilde{X}, \tilde{A})$ for a version of (4.38) starting at time $t = 0$ from an arbitrary initial age a_0 and an initial configuration $x_0 = (x_0^{(0)}, \ldots, x_0^{(n)})$ with $x_0^{(k)} = b\int_{-\infty}^0 \frac{(-s)^{n-k}}{(n-k)!} e^{-\nu s} z(ds)$. Write

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

Note that

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

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

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

showing that

$$\lim_{t\to\infty} |X_t - \tilde{X}_t| = 0 \tag{4.42}$$

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

$$\lim_{t\to\infty} |X_t^{(k)} - \tilde{X}_t^{(k)}| = 0 \tag{4.43}$$

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

Step 3. Finally, to prove the Harris recurrence of the process $(X_t^{(0)}, \ldots, X_t^{(n)}, A_t)$, we rely on the following local Doeblin lower bound.

[1] As a matter of fact, this provides a different approach to prove the existence of a stationary version of the age dependent Hawkes process.

It states that for all $(x^{**}, a^{**}) \in \mathbb{R}^{n+1} \times \mathbb{R}_+$, there exist $R > 0$, an open set $I \subset \mathbb{R}^{n+1} \times \mathbb{R}_+$ and a constant $\beta \in (0,1)$, such that for any $T > (n+2)a^*$,

$$P_T((x_0, a_0), \cdot) \geq \beta \mathbb{1}_C(x_0, a_0) U(\cdot), \tag{4.44}$$

where $C = B_R((x^{**}, a^{**}))$ is the (open) ball of radius R centered at (x^{**}, a^{**}), and where U is the uniform measure on I. This lower bound follows easily adapting the proof of Theorem 3 in [15] to our framework.

We may apply the above result with $(x^{**}, a^{**}) \in supp(\mu)$ where μ is the (unique) invariant measure of the process. Then for the stationary version of the process, $(X_t^{(0)}, \ldots, X_t^{(n)}, A_t) \in B_{R/2}((x^{**}, a^{**})$ infinitely often. Then (4.42) and (4.43) imply that also $(\tilde{X}_t^{(0)}, \ldots, \tilde{X}_t^{(n)}, \tilde{A}_t) \in B_R(x^{**}, a^{**}) = C$ infinitely often, almost surely. The classical regeneration technique, see e.g. [36], allows to conclude that indeed the process is positively recurrent in the sense of Harris. $\square$

Chapter 5

Renewal Time Points For Hawkes Processes

5.1 Introduction, Purpose, and Results

In this chapter we consider a one-dimensional Hawkes process. It may be an ordinary nonlinear type as in [5], or an age dependent one as in the previous chapter. The aim is to discuss stability of Hawkes processes from a renewal perspective. When h is of compact support and Z is an ergodic linear Hawkes process, it will happen infinitely often that $Z\left[t - \operatorname{supp}\left(h\right), t\right] = 0$, at which point a renewal occurs. It was shown in [19] that these renewal times have exponential moment under certain regularity assumptions. However, when the weight function h does not have compact support, it is no longer straightforward to find timepoints where the past can be eliminated. In this chapter we show how to construct such renewal times. The procedure is not unlike the Athreya-Ney technique for Markov Chains in the sense that we wait for some stopping-time α_0 to occur, which may be interpreted as a minorization criteria. Here we let random variables independent of Z decide whether we obtain a renewal α_0 at this point, or we jump to a new state of Z by moving time forward to a stopping time τ_1. This procedure is repeated, until a renewal has occured after a random number of iterations η. The renewal time α_η will be a stopping time w.r.t. the enlarged filtration induced by Z and the independent decisions.

The renewal approach to discussing stability of Hawkes processes turn out to be beneficial for establishing a number of key results for Hawkes processes. Here we give a brief overview of the results:

- It is well known from [5] that two ordinary Hawkes processes driven by the same Poisson random measure with sufficiently fast decaying initial signals couple eventually. The coupling time is bounded by the renewal time α_η. We use this to formulate moment results for the coupling time in terms of the distribution $h\,dt$. These results are in agreement with known results about coupling times for ordinary Hawkes processes found in [6]. Moreover, α_η is constructed explicitly so that it can be simulated.

- We prove a CLT for processes of the time average type:

$$t^{-1/2} \int_0^t H(Z_{|[s-D,s]})ds \Rightarrow N\left(\mu, \sigma^2\right) \tag{5.1}$$

  for appropiate $\mu \in \mathbb{R}, \sigma > 0$. This was done for the linear Hawkes processes in [19] assuming compact support of h, and for such h our results coincide.

- We prove a functional CLT for Hawkes processes. This was done for ordinary Hawkes processes in [44] with slightly weaker integrability assumptions on h compared to what we impose. However, we do not need positivity of h, nor do we need that h itself is decreasing.

5.2 Setup In This Chapter

For convenience in proofs we extend the definition of Hawkes processes to the D-delayed ADHP for $D \geq 0$, which is essentially an ADHP where the intensity is killed until time D.

Definition 5.2.1.
Let $D \geq 0$. We shall say that Z^* is a D-delayed ADHP with weight function h, rate function ψ, initial signal R^* and initial age A_0^* if $Z^*[t,t] = 0$ for $t \in [0, D]$ and Z^*_{D+} is an ADHP (driven by π_{D+}) with parameters (h, ψ), initial age $A_0^* + D$ and signal $t \mapsto R^*_{t+D}$. ○

When $D = 0$ we obtain the regular ADHP. If it also holds that $\psi(x, a)$ does not depend on a, then it is the ordinary nonlinear Hawkes process. If moreover

$\psi(x) = \psi(0) + Lx_+$ is linear, then we obtain the linear Hawkes process. We shall submit to the following restriction on ψ.

Assumption 5.1.

- ψ is increasing in x, a.
- We strengthen the Lipschitz assumption of ψ to the following: For all $x \leq y \in \mathbb{R}$ and $a, b \in \mathbb{R}_+$ it holds that

$$\psi(y,b) - \psi(x,a) \leq \begin{cases} L(y-x) & a = b \\ L(y-x) + \left(c_\psi^{pre} + Lx_+\right) g(a \wedge b) & a \neq b \end{cases} \tag{5.2}$$

where g is a decreasing function bounded by 1, and $L, c_\psi^{pre} > 0$.

∘

The first point is very natural. Indeed, the intuitive interpretation that the sign of h implies excitation/inhibition requires that ψ is increasing in x. For our purpose, age acts as an inhibitory effect throught the refractory period, so it makes sense to have ψ increasing, g decreasing as well. The second point prevents that initial disyncrony of two hawkes processes i.e. $A_0^1 \neq A_0^2$ will induce a large and persisting difference of their intensities, no matter the similarity of their histories. We observe that (5.2) implies that ψ is sublinear since

$$\psi(y,b) \leq \psi(0,b) \pm \psi(0,0) \leq c_\psi^{pre} g(0) + \psi(0,0) \quad \text{if } y < 0, b \in \mathbb{R}_+ \tag{5.3}$$

$$\psi(y,b) = \psi(y,b) \pm \psi(0,0) \leq Ly + c_\psi^{pre} g(0) + \psi(0,0) \quad \text{if } y \geq 0, b \in \mathbb{R}_+ \tag{5.4}$$

so with $c_\psi := c_{\psi^i}^{pre} + \psi(0,0)$ we have

$$\psi(y,b) \leq c_\psi + Ly_+ \quad \forall y \in \mathbb{R}, b \in \mathbb{R}_+. \tag{5.5}$$

Let now Z^* be the ADHP that we wish to obtain a regeneration point for. It is well known that depending on the parameters h, ψ, Hawkes processes can either be in the subcritical regime where $\limsup_{t\to\infty} Z^*(0,t]/t < \infty$ or in the supercritical regime where the limit is ∞. To succeed we must ensure that Z^* is in the subcritical regime. We shall treat two different setups that will ensure this. The first setup assumes that $\int_{\mathbb{R}_+} h^+(s)\,ds < L^{-1}$. For the Linear Hawkes process with

$$\psi_L(x) := c_\psi + Lx_+ \tag{5.6}$$

this has the interpretation that each direct child of a parent jump induces < 1 new child on average. The second setup assumes that the ADHP has a refractory period, i.e. the intensity is bounded for a period after each jump (this includes the case where ψ is uniformly bounded).

Setup (Ordinary Hawkes Process).
We assume that $\int_{\mathbb{R}_+} h_+(s)\,ds < L^{-1}$ where $h_+(t) = \max(h(t), 0)$. ◦

Setup (Age Dependent Hawkes Process).
There exists $K \geq 0, \delta \in \{1/n : n \in \mathbb{N}\}$ s.t.[1]

$$\psi(x, a) \leq K \text{ for } a \in [0, \delta], x \in \mathbb{R}. \tag{5.7}$$

◦

We shall establish a renewal time for each of these setups. While some variables will vary slightly in their definitions for each setup, the approach is similar so the renewal time will be constructed simultaneously. We shall refer to the two setups above as setup (O) or (AD) respectively.

Example 5.2.2.
Consider the rate function given by

$$\psi(x, a) = l(x)\varphi(x, a) \tag{5.8}$$

where l is increasing and L-Lipschitz and φ is bounded by 1 for all $x \in \mathbb{R}$, $a \in \mathbb{R}_+$. Moreover, we assume that $x \mapsto l(x)\varphi(x, a)$ is L-lipschitz and that φ converges to 1 in the sense that there is a function $g : \mathbb{R}_+ \to [0, 1]$ decreasing towards 0 and satisfying $1 - \varphi(x, a) \leq g(a)$ for all $x \in \mathbb{R}, a \in \mathbb{R}_+$. For $\varphi \equiv 1$ we obtain the ordinary Hawkes process, so for general φ we may interpret the ADHP as an ordinary Hawkes process with rate function l, but inhibited by its own age process with a factor $\varphi(x, a)$.

[1] It is merely for mathematical convenience that we restrict δ to reciprocal integers instead of arbitrary $\delta \in \mathbb{R}_+$ in setup (AD).

To show that it satisfies (5.2) we take arbitrary $x \leq y$ and $a \neq b$ and obtain

$$\psi(y,b) - \psi(x,a) = l\,(y)\,\varphi\,(y,b) - l\,(x)\,\varphi\,(x,a) \tag{5.9}$$

$$= (l\,(y) - l\,(x))\,\varphi\,(y,b) + l\,(x)\,(\varphi\,(y,b) - \varphi\,(x,a)) \tag{5.10}$$

If $x \geq 0$ then $l\,(x) \leq l\,(0) + L\,(x-0)$ while $l\,(x) \leq l\,(0)$ for $x < 0$. We use this and the fact that $|\varphi\,(y,b) - \varphi\,(x,a)| \leq 1 - \varphi\,(x, a \wedge b) \leq g\,(a \wedge b)$ to conclude

$$\psi(y,b) - \psi(x,a) \leq L\,(y-x) + (l\,(0) + Lx_+)\,g\,(a \wedge b) \tag{5.11}$$

which fits into (5.2). The most principal example of φ is the simple $\mathbb{1}\,\{A \leq \delta)$ corresponding to a hard refractory period, and in this case Z^* is in setup (AD). Although this is a rather simple example, it is important due to its application for modelling neural spike-trains. We could also give a more complicated structure to the refractory period such as

$$\varphi\,(x,a) = \begin{cases} \frac{1-e^{-a}}{1 \vee l(x)} & a \leq \delta \\ 1 - e^{-a} & a > \delta \end{cases}. \tag{5.12}$$

Notice that both of the above mentioned φ choices makes ψ increasing.

As mentioned previously, if $\varphi \equiv 1$ then

$$\psi(x,a) = l(x) \tag{5.13}$$

and one obtains the ordinary Hawkes process. We may assume that $\|h_+\|_{\mathcal{L}^1} < L^{-1}$ in which case the parameters fit under setup (O), or we can assume ψ is bounded in which case it fits under setup (AD). ∘

For each setup, we impose two assumptions. The first one restricts the randomness in the initial signal.

Assumption 5.2.
There is an a.s. finite $\mathcal{F}_t$-stopping time α_0 and a deterministic decreasing function $r : \mathbb{R}_+ \to [0,\infty)$ such that for all $t > \alpha_0$

$$\left| \int_0^{\alpha_0} h\,(t-s)\,dZ_s^* + R_t^* \right| \leq r(t - \alpha_0). \tag{5.14}$$

∘

The next assumption puts integrability assumptions on r, h, g. It will be split in two. One where h, r have power tails, and one where they have exponential tails.

Assumption 5.3.
Let $\gamma : [0,\infty) \to [0,\infty)$ be an increasing and right continuous function and define
$\overline{h}(t) := \sup_{s\geq t} |h(s)|$. We assume that $\overline{h} \in \mathcal{L}^1_{loc}$ and either (A) or (B) below holds.

Assumption 3 (A):
There exists $p \geq 0$ s.t.

- $t \mapsto t^p r(t) \in \mathcal{L}^1, t \mapsto t^p g(t) \in \mathcal{L}^1,$ and $t \mapsto t^{p+1}\gamma(t+1)\,\overline{h}(t) \in \mathcal{L}^1$.
- Under setup (O) we assume

$$\liminf_{t\to\infty} \frac{\gamma(t)}{c_h^{-1}(p+1)\ln_+ t} > 1 \tag{5.15}$$

  where $c_h = L\|h_+\|_{\mathcal{L}^1} - \ln(L\|h_+\|_{\mathcal{L}^1}) - 1$. Under (AD) we merely assume

$$\liminf_{t\to\infty} \frac{\gamma(t)}{\ln_+ t} > 0. \tag{5.16}$$

Assumption 3 (B):

- The functions r, g and $t \mapsto \gamma(t+1)\,\overline{h}(t)$ have exponential moments.
- We assume that

$$\liminf_{t\to\infty} \frac{\gamma(t)}{t} > 0.$$

○

Remark 5.2.3.
A few remarks on the introduced variables and assumptions:

1. *For all results to come in this article we shall implicitly assume assumption 5.1, assumption 5.2 and assumption 5.3 (A), unless otherwise stated. We will state explicitly which setup we work under. Assumption 5.3 (B) is clearly stronger than the (A) version for any choice of p, and we state explicitly when we work under assumption (B) instead of (A).*

2. *The map $\overline{h}$ is the smallest decreasing function dominating h. As in the previous chapter, we put integrability assumptions on $\overline{h}$ which is slightly more restrictive than if they were put on h. It turns out to be advantagous to work with a decreasing weight function and the restriction is, at least in the belief of the author, of small consequence for practical applications.*

3. *If $\int_0^\infty t^{p+1} \ln_+ t \; \overline{h}(t)\, dt < \infty$ then the choice $\gamma(t) = C \ln_+(t)$ for large C satisfies the parts of assumption 5.3 (A) relevant to $\overline{h}, \gamma$. Likewise, if $\overline{h}$ has exponential moment, then the choice $\gamma(t) = Ct$ for any $C > 0$ satisfies assumption 5.3 (B).We allow γ to be chosen freely because it may change the speed of computation in an actual simulation of the renewal-times. Recall that since γ is right continuous, the generalized inverse $\gamma^{-1}(t) := \inf\{s \geq 0 : \gamma(s) \geq t\}$ satisfies*

$$y \leq \gamma(t) \Leftrightarrow \gamma^{-1}(y) \leq t. \tag{5.17}$$

$\circ$

We now define some key functions to be used in the construction of a regeneration time, and with assumption 5.3 we immediately determine their integrability properties. Define

$$\begin{aligned} f(t_1, t_2) = &\left(1 + \delta^{-1} + \gamma(0)\right) \overline{h}(t_1) \\ &+ \int_0^{t_2} \left(1 + \delta^{-1} + \gamma(s+1)\right) \overline{h}(t_1 + s)\, ds + r(t_1) \end{aligned} \tag{5.18}$$

(with convention $\delta = \infty$ in the (O)-system). For convenience we write $f(t)$ instead of $f(t, \infty)$. Define also

$$F^{pre}(t) = 2Lf(t) + c_\psi g(t) \tag{5.19}$$

$$F(t) = \mathbb{1}\{t \leq D\}\left(c_\psi + Lf(t)\right) + \mathbb{1}\{t > D\} F^{pre}(t) \tag{5.20}$$

Proposition 5.2.4.
Consider the maps $t \mapsto f(t), F^{pre}(t), F(t)$ *defined above. Under either setup and assumption 5.3(A) these functions have p'th moment, and under assumption 5.3(B) these functions have exponential moments.* ∘

An important example for which assumption 5.2 is satisfied is the stationary ADHP:

Example 5.2.5.
The classical method of studying stability of Hawkes processes, due to Brémaud and Massoulié [5], has been to find a solution

$$\begin{aligned} Z^I(a,b] &= \int_a^b \int_0^\infty \mathbb{1}\{z \le \lambda_s^I\} d\pi(s,z) \\ X_t^I &= \int_{-\infty}^{t-} h(t-s)\, dZ_s^I \\ \lambda_s^I &= \psi(X_t, A_t) \end{aligned}$$

such that $\theta^t Z^I = H^I(\theta^t \pi)$ for some suitable map $H^I : M^c_{\mathbb{R}\times\mathbb{R}_+} \to M^c_{\mathbb{R}}$. Note that Z^I is an ADHP with $R_t^I = \int_{-\infty}^{0-} h(t-s)\, dZ_s^I$. See [5] and the previous chapter for criterias of existence. In both of the cited papers, it is proven that when Z^I exists, it is stationary and ergodic, and if Z^* is another Hawkes process driven by π, with a signal R^* satisfying $\mathbb{E}\int_0^\infty |R_s^*|\, ds < \infty$, then Z^* couples with Z^I eventually. We shall see in theorem 5.3.3 that there is a suitable choice of α_0,r such that assumption 5.2 is satisfied for Z^I. In proposition 5.5.1 we prove that the coupling time has p'th moment and even exponential moment under assumption 5.3(B). ∘

5.3 Constructing a Renewal Time Point for a Hawkes Process

In this section we are given an ADHP Z^*. The goal is to prove the main result theorem 5.3.2 which gives a random time ρ satisfying that $Z^*_{\rho+} \perp\!\!\!\perp Z_{|(0,\rho]}$. This is done by introducing a point process Z which regenerates at stopping times α_n. Then, by using the specific construction of α_n and Z, we are able to throw a biased coin deciding whether α_n should be the renewal time point for Z^* or not.

The first step is to construct Z. Given either of the two setups, we shall simultaneously define Z, the sequences of stopping times $(\tau_n), (\alpha_n)$, and intensities $\underline{\lambda}, \overline{\lambda}$ as a system. While the system is defined slightly different for each of the two setups, they are very similar. Only the α's differs in the definition, depending on whether we discuss the (O) system or the (AD) system. Recall the split PRMs $\pi^{\uparrow}, \pi^{\downarrow}$ defined in the first section of the appendix. We also note that we make use of the convention $\inf\{\emptyset\} = \infty$.

The system is defined as follows: Fix $D \geq 0$ and define $\overline{\lambda}_t = \underline{\lambda}_t = 0$ for $t \in (0, \alpha_0]$ and $Z(0, \alpha_0] = 0$. For $n \in \mathbb{N}$ s.t. $\alpha_{n-1} < \infty$ we define Z as the D-delayed ADHP driven by π, with parameters h, ψ and initial conditions $A_{\alpha_{n-1}+} = 0$, $R_t = -f(t - \alpha_{n-1})$. Let λ be its intensity. We set $\underline{\lambda}_t, \overline{\lambda}_t = 0$ if $t \in (\tau_n, \alpha_n]$ and

$$\underline{\lambda}_t = \lambda_t \tag{5.21}$$

$$\overline{\lambda}_t = \underline{\lambda}_t + F(t - \alpha_{n-1}) \tag{5.22}$$

when $t \in (\alpha_{n-1}, \tau_n]$. Moreover, we set

$$\tau_n = \inf\left\{t > \alpha_{n-1} : \int_{\alpha_{n-1}}^{t} \int_0^{\infty} \mathbb{1}\left\{z \in \left(\underline{\lambda}_s, \overline{\lambda}_s\right]\right\} d\pi(s, z) \geq 1\right\} \tag{5.23}$$

$$= \inf\left\{t > \alpha_{n-1} : \int_{\alpha_{n-1}}^{t} \int_0^{\infty} \mathbb{1}\left\{z \leq F(t - \alpha_{n-1})\right\} d\pi^{\uparrow \underline{\lambda}, \overline{\lambda}}(s, z) \geq 1\right\}. \tag{5.24}$$

If $\tau_n = \infty$ we set $\alpha_n = \infty$ in either setup. Otherwise, under setup (AD) we choose

$$\alpha_n = \alpha_{n-1} + \inf\{i > \lceil \tau_n - \alpha_{n-1} \rceil : \theta^{i-j} N_{\alpha_{n-1}+}(-1, 0] \leq \gamma(j), j = 0, \dots, i-1\}, \tag{5.25}$$

where N is the K-Poisson process driven by $\pi^{\downarrow \underline{\lambda}, \overline{\lambda}}$. For setup (O), let $\varsigma_{\tau_n - \alpha_{n-1}}$ be the Dirac-measure on $\tau_n - \alpha_{n-1}$ and let $Z^{n,pre}$ be the linear Hawkes process driven by $\pi^{\downarrow \underline{\lambda}, \overline{\lambda}}_{\alpha_{n-1}+}$ with weight function h_+, rate function ψ_L and initial signal $R_t = f(t) + h_+(t - (\tau_n - \alpha_{n-1})) \mathbb{1}\{t > \tau_n - \alpha_{n-1}\}$. Define $Z^n = \varsigma_{\tau_n - \alpha_{n-1}} + Z^{n,pre}$ and put

$$\alpha_n = \alpha_{n-1} + \inf\Bigg\{ i > \lceil \tau_n - \alpha_{n-1} \rceil : \\ \int_0^i h_+\,(t-s)\,dZ_s^n \le f(t-i,i-1) - r(t-i)\ \forall t > i \text{ and } Z^n\,(0,i] \le \int_0^i \gamma\,(s+1)\,ds \Bigg\}.$$

Remark 5.3.1.
We notice some properties of the system above.

1. *Z is a well-defined $\mathcal{F}_t$-progressive process on $\mathbb{R}_+$ and $\lambda, \underline{\lambda}, \overline{\lambda}$ are $\mathcal{F}_t$-predictable.*
2. *The system remains unchanged for any choice of initial conditions R^*, A_0^* as long as α_0 from assumption 5.2 remains unchanged.*
3. *By theorem 7.3.1 $\pi^{\downarrow\underline{\lambda},\overline{\lambda}}, \pi^{\uparrow\underline{\lambda},\overline{\lambda}}$ are $\mathcal{F}_t - PRMs$. In particular*
$$(\alpha_n - \alpha_{n-1}, \tau_n - \alpha_{n-1}) \perp\!\!\!\perp \mathcal{F}_{\alpha_{n-1}} \mid (\alpha_{n-1} < \infty).$$
4. *The process is reversible in the sense that $\underline{\lambda}_{\alpha_n+s}, \overline{\lambda}_{\alpha_n+s}$ may be computed from*
$\left(\pi^{\downarrow\underline{\lambda},\overline{\lambda}}_{\alpha_n+}\right)_{|(0,s]}, \left(\pi^{\uparrow\underline{\lambda},\overline{\lambda}}_{\alpha_n+}\right)_{|(0,s]}$.
That is, there is a map $H : M^c_{\mathbb{R}^2_+} \times M^c_{\mathbb{R}^2_+} \times \mathbb{R}_+ \to M^c_{\mathbb{R}^2_+}$ satisfying
$$H\left(\left(\pi^{\downarrow\underline{\lambda},\overline{\lambda}}_{\alpha_n+}\right)_{|(0,s]}, \left(\pi^{\uparrow\underline{\lambda},\overline{\lambda}}_{\alpha_n+}\right)_{|(0,s]}, s\right) = (\pi_{\alpha_n+})_{|(0,s]}, \tag{5.26}$$
for all $n \in \mathbb{N}_0$ s.t. $\alpha_n < \infty$.

$\circ$

The rest of this section is dedicated to presenting our main result theorem 5.3.2, and its related results 5.3.3 - 5.3.6. Before we state it, we colloquially explain the essence of the result. The purpose of the α's is to have points in time, where the intensity contribution from the past of Z^* may be replaced by something deterministic. More preciesly α_n should satisfy

$$\left|\int_0^{\alpha_n} h\,(t-s)\,dZ_s^* + R_t^*\right| \le f(t-\alpha_n), \quad t > \alpha_n, n \in \mathbb{N} : \alpha_n < \infty. \tag{5.27}$$

We prove that this property holds for both setups in theorem 5.3.3. The inequality (5.27) combined with the properties of ψ gives that $\lambda \leq \lambda^*$, at least locally in time after α_n. We will also be able to control the difference $\lambda^* - \lambda$ locally after α_n, and establish that Z mimics Z^* in that same interval. In fact, the purpose of τ_{n+1} is to act as a conservative right end of an interval starting at α_n on which Z and Z^* are equal. All this will be proved in proposition 5.3.4. We then proceed to study the distributions of τ_n, α_n. In theorem 5.3.5 we study $\tau_{n+1} - \alpha_n | \alpha_n < \infty$ and prove $P(\tau_{n+1} = \infty, \alpha_n < \infty) > 0$ along with moment properties of the distribution $\tau_{n+1} - \alpha_n | \tau_{n+1} < \infty$. In theorem 5.3.6 we investigate the law of $\alpha_n - \tau_n | \tau_n < \infty$. Here we prove that $P\left(\alpha_n - \tau_n < \infty | \tau_n < \infty\right) = 1$, and we characterize its moments. Combining these results implies that

$$\eta := \inf\left\{n \in \mathbb{N}_0 : \tau_{n+1} = \infty\right\} \tag{5.28}$$

is finite almost surely, and we will be able to show that

$$\rho := \alpha_\eta + D \tag{5.29}$$

is a point of regeneration for Z^*. In fact, we have $Z^*\left(\rho - D, \rho\right] = 0$ so $Z^*_{|(0,\rho]} \perp\!\!\!\perp Z^*_{\alpha_\eta +}$, i.e. there is an overlap of length D. We characterize the integrability properties of ρ by applying the previously mentioned results concerning α_n, τ_n.

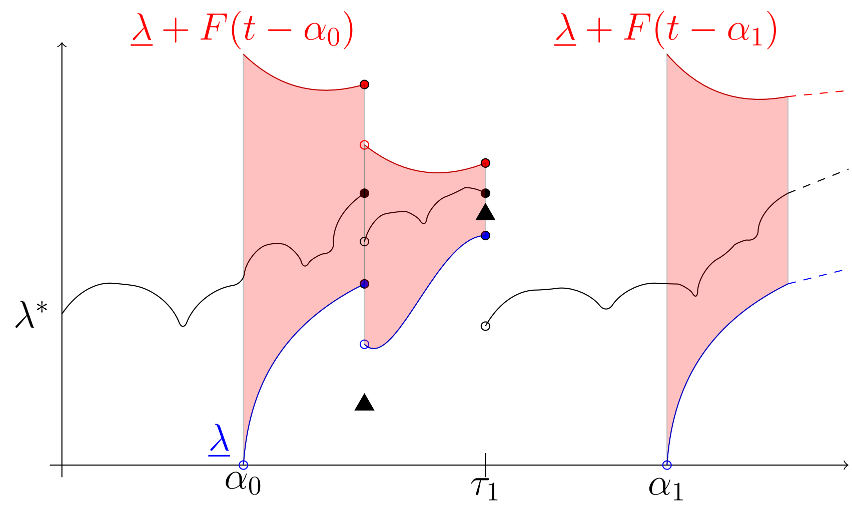

Figure 5.1: An illustration of the system with $D = 0$. The points of the PRM π are depicted by ($\blacktriangle$), and the three intensities $\underline{\lambda}, \lambda^*, \overline{\lambda}$ are colored in blue, black and red respectively. By definition of the stopping times τ_1, τ_2, the red bands contain no π-jumps.

The precise result is as follows: Define $\mathcal{F}_t^* = \sigma\left(\mathcal{F}_t, \pi^{\uparrow\underline{\lambda},\bar{\lambda}}\right)$. It is clear that $(\mathcal{F}_t^*)$ defines a filtration and without changing notation, we extend it to satisfy the usual hypothesis.

Theorem 5.3.2.
Grant either setup (AD) or (O).

1. *The random time $\rho = \alpha_\eta + D$ is an a.s. finite $\mathcal{F}_t^*$-stopping time.*

2. *The random measure $\pi^{\downarrow\underline{\lambda},\bar{\lambda}}$ is an $\mathcal{F}_t^*$-PRM and hence $\pi^{\downarrow\underline{\lambda},\bar{\lambda}} \perp\!\!\!\perp \mathcal{F}_\rho^*$. Moreover, we have $Z_{\rho+}^* = Z_{\rho+}$ and independent of $Z_{|(0,\rho]}^*$. In particular, $Z_{\rho+}^*$ is distributed as an ADHP with weight h, rate ψ, initial age D and signal $t \mapsto -f(t+D)$.*

3. *It holds that $\mathbb{E}(\rho - \alpha_0)^p < \infty$. Under assumption 5.3(B) it holds that $\rho - \alpha_0$ has exponential moment.*

$\circ$

To prove Theorem 5.3.2 we establish the results below and combine them in the end. The proofs of these results, and the main result, may be found in the proof section.

Theorem 5.3.3.

1. *Consider setup (AD). It holds that $\alpha_n, n \in \mathbb{N}$ satisfies (5.27). Moreover, assume Z^I from example 5.2.5 exists, and assume only that assumption 5.1 and assumption 5.3 (A) holds with $r = f$. Then assumption 5.2 is satisfied for $Z^* = Z^I$ and*

$$\alpha_0^I = \inf\left\{i > 0 : \theta^{i-j} N(-1, 0] \leq \gamma(j), j \geq 0\right\}. \tag{5.30}$$

2. *Consider setup (O) and recall ψ_L from (5.6). It holds that $\alpha_n, n \in \mathbb{N}$ satisfies (5.27). Moreover, assume Z^I from example 5.2.5 exists, and assume only that assumption 5.1 and assumption 5.3 (A) holds with $r = f$. Define Z^0 as the stationary linear Hawkes process driven by π with weight/rate h_+, ψ_L (see [5] theorem 1 and remark 8). Then assumption 5.2 is satisfied with*

$$\alpha_0^I = \inf\{i > 0 : \int_{-\infty}^{i} h_+(t-s)\, dZ_s^0 \leq f(t-i) - r(t-i) \ \ \forall t > i\}. \tag{5.31}$$

○

Proposition 5.3.4.
Under either setup it holds a.s. for all $n \in \mathbb{N}_0$ such that $\alpha_n < \infty$ and $t \in (\alpha_n, \tau_{n+1}]$ that

$$Z_{|(\alpha_n,\tau_{n+1})} = Z^*_{|(\alpha_n,\tau_{n+1})} \tag{5.32}$$

and

$$0 \leq \lambda^*_t - \underline{\lambda_t} \leq F(t - \alpha_n). \tag{5.33}$$

○

Theorem 5.3.5.
Under either setup it holds that

$$P\left(\tau_n - \alpha_{n-1} = \infty | \alpha_{n-1} < \infty\right) = \exp\left(- \|F\|_{\mathcal{L}^1}\right) \tag{5.34}$$

$$P\left(\tau_n - \alpha_{n-1} \leq t | \tau_n < \infty\right) = \frac{1 - \exp\left(- \int_0^t F\left(s\right) ds\right)}{1 - \exp\left(- \|F\|_{\mathcal{L}^1}\right)}. \tag{5.35}$$

In particular the conditional distribution $\tau_n - \alpha_{n-1} | \tau_n < \infty$ has p'th moment. Under assumption 5.3 (B) it has exponential moment. ○

It turns out that the α_n's defined above may be analyzed using a discrete Markov chain. In fact one may rewrite α_0^I and $\alpha_n - \tau_n$ as return times to state 0 for a specific Random Exchange process (see Appendix). This yields precise distribution results as given in the next proposition.

Theorem 5.3.6.
Under either setup it holds that α_0^I and $(\alpha_n - \tau_n) \mid (\tau_n < \infty)$ have p'th moment. If also assumption 5.3 (B) holds then these laws have exponential moment.

○

5.4 Hawkes Processes In a Markov Chain Framework

In this section we first apply theorem 5.3.2 iteratively to obtain consecutive renewal time points ρ_i, which partition Z^* into independent bits.

Afterwards, we construct a Markov chain that contains the information of Z^*, and where the ρ_i's acts as the return times to an atom. The purpose is to use Markov chain theory to obtain results for Z^*, which we do in the next section.

Choose $D > 0$ and set $\rho_0 := \rho, \pi^0 := \pi, \overline{\pi}^0 := \overline{\pi}$ and $\pi^1 = \pi^{\downarrow\lambda,\overline{\lambda}}_{\rho_0+}$. We also introduce another auxiliary PRM $\overline{\pi}^1$ independent of $\pi^0, \overline{\pi}^0, \pi^1$. Note that $\pi^1, \overline{\pi}^1$ are $\mathcal{F}^1_t$-PRMs where $\mathcal{F}^1_t = \sigma(\mathcal{F}^*_\rho, \pi^1_{|(0,t]}, \overline{\pi}^1_{|(0,t]})$. It holds that $Z^*_{\rho+}$ is an ADHP driven by π^1 with $R : t \mapsto -f\,(t + D)$, $A_0 = D$. In particular $\alpha^1_0 = 0$ satisfies (5.27) already. Thus, $\pi^1, \overline{\pi}^1$ and α^1_0 induces a new renewal system as in section 3. From there we obtain sequences $\left(\alpha^1_i\right), \left(\tau^1_i\right)$ and two new PRMs which we denote $\pi^{\downarrow 1}, \pi^{\uparrow 1}$. Applying theorem 5.3.2 on this system gives a new renewal time ρ_1.

Continuing this way gives sequences $\left(\pi^i\right), \left(\overline{\pi}^i\right), \left(\pi^{\uparrow i}\right), \left(\pi^{\downarrow i}\right)$, (ρ_i). If we set $\varrho_i = \sum_{j=0}^{i} \rho_j$, we have that

$$B \mapsto Z^*\left((\varrho_{i-1} - D, \varrho_i] \cap (B + \varrho_{i-1})\right), \quad B \in \mathcal{B}_{(-D,\infty)} \tag{5.36}$$

are i.i.d. for $i \in \mathbb{N}$, each being the Hawkes process initialized with $A_0 = D$, $R : t \mapsto -f\,(t + D)$ and driven by π^i.

In fact, we can study Z^* from a Markov chain perspective. Define

$$\rho^-(t) = \sup\left\{s < t : \exists i \in \mathbb{N}_0 : s = \varrho_i\right\}, \tag{5.37}$$

$$J\,(n) = \left|\{j \in \mathbb{N}_0 : \varrho_j < n\}\right|, \tag{5.38}$$

and $A^\rho_t = t - \rho^-(t)$. Consider the stochastic processes on the state-space $M^c_{\mathbb{R}^2_+} \times M^c_{\mathbb{R}^2_+} \times \mathbb{N}$,

$$\Phi^{pre}_n = \left(\pi^{\downarrow J(n)}_{|(0,A^\rho_n]}, \pi^{\uparrow J(n)}, A^p_n\right) \tag{5.39}$$

$$\Phi_n = \theta^{\rho_0}\Phi^{pre}_n \tag{5.40}$$

for $n \in \mathbb{N}$. In this framework, Theorem 5.3.2 states that $\Phi_n \perp\!\!\!\perp \Phi^{pre}_0, \ldots, \Phi^{pre}_{\rho_0}$. Using (5.26) we may construct a map $H^* : M^c_{\mathbb{R}^2_+} \times M^c_{\mathbb{R}^2_+} \times \mathbb{N} \to M^c_{(-D,0]}$ satisfying $H^*(\Phi_n) = (\theta^{\rho_0+n}Z^*)_{|(-D,0]}$. Also, by construction of ρ_i, the indicator function $J\,(n + \rho_0 + 1) - J\,(n + \rho_0)$ may be written as some map $H_J\,(\Phi_n)$. It follows that Φ is a Markov chain with an atom

$$\Xi = \{\mu, \nu, n : H_J\,(\mu, \nu, n) = 1\}. \tag{5.41}$$

Consider the subspace

$$\mathbb{X} = \left\{ \mu, \nu, n : \mu, \nu \text{ are simple }, P^{\Phi}_{\mu,\nu,n} \left(\Phi \text{ hits } \Xi \text{ eventually} \right) = 1 \right\} \quad (5.42)$$

and let P^{Φ} be the kernel of Φ. By definition of $\mathbb{X}$, any chain with kernel P^{Φ} started in $\mathbb{X}$ eventually hits Ξ, and by theorem 5.3.2 Φ almost surely returns to Ξ once hitting it. Thus $\mathbb{X}$ is an absorbing state for chain Φ, and we shall from now on always refer to Φ, P^{Φ} as the restricted kernel to $\mathbb{X}$ (see proposition 4.2.4. [34]). Since Ξ is an accesible atom for this chain, it is irreducible (Prop. 5.1.1 [34]) and aperiodic. The return time to Ξ is distributed as ρ_1, so by Kac's theorem it follows that P^{Φ} is positive with invariant law $\tilde{P}$ for $p \geq 1$. Not surprisingly it turns out that $\tilde{P}$ agrees with Z^I from example 5.2.5 with parameters h, ψ, whenever they both exist.

Proposition 5.4.1.
Assume that $p \geq 1$ and Z^I from example 5.2.5 exist with coupling properties as given in the example. It holds that $H^* \left(\tilde{P} \right) \overset{D}{=} Z^I_{|(-D,0]}$. ∘

5.5 Applications

In this section we apply the Markov chain construction from section 5.4 to establish asymptotic results for Z^*. We show distribution results of the coupling time. Then we present a functional CLT, a time-average CLT, and a LIL for Z^*.

5.5.1 Bound On The Coupling Time

In [5] it was shown that two ordinary Hawkes processes, started with different initial conditions couple under regularity conditions. In the previous chapter we showed a similar statement for ADHPs. Our construction of ρ confirms these results and provide moment results for these coupling times. More preciesly, we have the following proposition.

Proposition 5.5.1.
Let Z^{*1}, Z^{*2} be two π-driven ADHPs initialized with A_0^{*1}, A_0^{*2} and signals R_t^{*1}, R_t^{*2}. Assume that α_0 satisifies assumption 5.2 for both measures simultaneously.

Define the coupling time $T = \inf\left\{t > 0 : \left|Z^{*1} - Z^{*2}\right|(t,\infty) = 0\right\}$. *It holds that* $T \leq \rho$ *and in particular that* $(T - \alpha_0)_+$ *has p'th moment. Under assumption 5.3 (B),* $(T - \alpha_0)_+$ *has exponential moment.* ∘

Proof.
By theorem 5.3.2 2) we have $Z^{*1}_{\rho+} = Z^{*2}_{\rho+} = Z_{\rho+}$. The moment results follows from theorem 5.3.5 and theorem 5.3.6. □

5.5.2 Asymptotics

The Markov chain Φ_n from (5.39) can be used to establish various asymptotic results for Z^* in a general setting. Let $G : M^c_{\mathbb{R}^2_+} \times M^c_{\mathbb{R}^2_+} \times \mathbb{N} \to \mathbb{R}$ be a measurable function which we normalize with $\overline{G} = G - \tilde{P}G$ where $\tilde{P}$ is the invariant measure from proposition 5.4.1. We shall discuss asymptotic results of the sum

$$S_n\left(\overline{G}\right) = \sum_{k=1}^{n} \overline{G}\left(\Phi^{pre}_k\right) \tag{5.43}$$

provided of course that G is a function s.t. $G\left(\Phi^{pre}_n\right)$ is a well-defined variable for all $n \in \mathbb{N}$. Define

$$\tilde{S}_n\left(\overline{G}\right) = S_{n+\rho_0} - S_{\rho_0}. \tag{5.44}$$

Theorem 5.5.2.
Assume that $p \geq 2$. *Define* $\mu_\rho = \mathbb{E}\rho_1$ *and*

$$\sigma^2 = \mu_\rho^{-1}\mathbb{E}\tilde{S}_{\rho_1}\left(\overline{G}\right)^2. \tag{5.45}$$

Assume that σ^2 *is finite and nonzero.*

1. *The following CLT holds*

$$n^{-1/2}S_n\left(\overline{G}\right) \Rightarrow N\left(0,\sigma^2\right). \tag{5.46}$$

2. *The following LIL holds: Almost surely,*

$$\liminf_{n\to\infty} \frac{S_n\left(\overline{G}\right)}{\sqrt{2\sigma^2 n \ln\ln n}} = -1, \quad \limsup_{n\to\infty} \frac{S_n\left(\overline{G}\right)}{\sqrt{2\sigma^2 n \ln\ln n}} = 1. \tag{5.47}$$

Define $S_t\left(\overline{G}\right)$ *for* $t \in (n, n+1)$ *as the linear interpolation between* $S_n\left(\overline{G}\right)$ *and* $S_{n+1}\left(\overline{G}\right)$*, and put*

$$B_t^n = \frac{1}{\sqrt{n\sigma^2}} S_{nt}\left(\overline{G}\right). \tag{5.48}$$

The following functional CLT holds

$$B_{(\cdot)}^n \Rightarrow B_{(\cdot)} \tag{5.49}$$

where B *is the 1-dim Brownian motion, and the convergence* ($\Rightarrow$) *is in distribution in the* D *space on* $[0, \infty)$*. See [4] chapter 3.* $\circ$

Proof.
A direct application of theorem 17.2.2, theorem 17.4.4, and section 17.4.3 [34] on Φ_n gives the desired results for $\tilde{S}_n$ in place of S_n. Standard arguments extend these results to hold for S_n as well. $\square$

Example 5.5.3.
We give two applications of theorem 5.5.2.

1. Take $G(\mu) = \int_{-D}^{0} \phi(s)\, dH^*(\mu)(s)$ where $\phi : (-D, 0] \to \mathbb{R}$ is a bounded map. In particular, when $\phi \equiv 1, D = 1$, we obtain

$$S_n\left(\overline{G}\right) = Z^*(0, n] - \mathbb{E}Z^I(0, n]. \tag{5.50}$$

It is straightforward to show a sufficient criteria for $\sigma^2 < \infty$ is that either $p > 2$ in the (O) setup or $p \geq 2$ in the (AD) setup. Also it is easy to see that for non-degenerate choices of h, ψ we have $\sigma^2 > 0$.

2. Fix $m \in \mathbb{N}$, and let $T : M^c_{|(-m,0]} \to \mathbb{R}$ be a measurable map. Take $D = m+1$ and $G(\mu) = \int_0^1 T\left(\left(\theta^s H^*(\mu)\right)_{|(-m,0]}\right) ds$. These choices lead to

$$S_n\left(\overline{G}\right) = \int_0^n T\left(\theta^s Z^*_{|(s-m,s]}\right) ds - n\mathbb{E}T\left(Z^I_{|(-m,0]}\right). \tag{5.51}$$

To obtain $\sigma^2 < \infty$ we need a growth condition depending on the setup:

$$\text{(AD)} \quad \exists c, C > 0 : |T(\mu)| \le C \exp(c\mu(-m, 0]) \tag{5.52}$$

$$\text{(O)} \quad \exists C > 0 \quad : |T(\mu)| \le C \left| \frac{\mu(-m, 0]}{1 + \ln(\mu(-m, 0])} \right|^{p/2-1}. \tag{5.53}$$

Consider first setup (AD). Recall that

$$Z^*_{\rho_0+}(n - D, n] = Z_{\rho_0+}(n - D, n] \le D\delta^{-1} + \pi^{\downarrow\lambda,\overline{\lambda}}_{\rho_0+}(n - D, n] \times [0, K]. \tag{5.54}$$

It follows that

$$\sup_{s \in (n-1,n]} \left| T\left(\theta^{-s} Z^*_{|(s-m,s]}\right) \right| \le C \exp\left(cD\delta^{-1}\right) \exp\left(c\pi^{\downarrow\lambda,\overline{\lambda}}_{\rho_0+}(n - D, n] \times [0, K]\right) := Y_n$$

For $j = 1 \dots D, n \in \mathbb{N}$ define $\tilde{Y}^j_n = Y_n$ if $n \equiv j \bmod D$, and otherwise set $\tilde{Y}^j_n$ as an i.i.d. copy of Y_1. Indeed by the C_p-inequality there is some possibly larger $C > 0$ s.t.

$$\tilde{S}_{\rho_1}(G)^2 \le C \sum_{j=1}^{D} \left(\sum_{n=1}^{\rho_1} \tilde{Y}^j_n \right)^2 \tag{5.55}$$

Notice that for each fixed $j = 1, \dots, D$, the sequence $\tilde{Y}^j_n, n \in \mathbb{N}$ is i.i.d. It follows from theorem 5.2, chapter 1 in [21] that $\sigma^2 < \infty$.

Consider now setup (O), and take $\gamma(t) = C \ln_+ t$ for C so large that (5.15) is satisfied. Note that $x \mapsto x/\ln(x + 1)$ is increasing for $x > 0$ so

$$\sup_{s \in (\varrho_0, \varrho_1]} \left| T\left(\theta^{-s} Z^*_{|(s-m,s]}\right) \right| \le C \left| \frac{Z^*(\varrho_0, \varrho_1]}{\ln(1 + Z^*(\varrho_0, \varrho_1])} \right|^{p/2-1}. \tag{5.56}$$

Notice that for our choice of γ, we have the inequality

$$\int_0^x \gamma(t + 1)\, dt \le Cx \ln(x + 1)$$

for a possibly larger constant $C > 0$. From the definition of α_n it follows that

$$\sup_{s\in(\varrho_0,\varrho_1]}\left|T\left(\theta^{-s}Z^*_{|(s-m,s]}\right)\right|\leq C_T\left|\frac{C\rho_1\ln(\rho_1+1)}{\ln(1+C\rho_1\ln(\rho_1+1))}\right|^{p/2-1}\leq C_TC\rho_1^{p/2-1} \tag{5.57}$$

again for a possibly larger constant $C>0$. From here it follows that $\sigma^2<\infty$. One would have to check $\sigma^2>0$ for the given T, but for most practical applications, this is a triviality. ∘

5.6 Discussion and Outlook

In the following, we shall discuss generalizations and limitations of the presented results, and suggest further research topics.

Multivariate Hawkes Processes

It is straight forward to generalize the regeneration procedure to a multivariate Hawkes Process with N units (see [12],[13] or chapter 2 for an introduction to these). One should split each π_i , $i\leq N$ into $\pi^{\uparrow i},\pi^{\downarrow i}$ for $i=1,...,N$ - analogous to what was done in the start of section 3.1. The τ_n^i's should be generalized in the obvious way, while α_n should be modified so that it ensures that $\sum_{i=1}^N\left|\int_0^{\alpha_n}h_{ij}(t-s)\,dZ_s^i+R_t^i\right|\leq f(t-\alpha_n)$. In setup (AD) this is achieved by substituting $\pi^{\downarrow\underline{\lambda},\overline{\lambda}}$ in (5.25),(5.31) with $\sum_{i=1}^N\pi^{\downarrow i}$ which will be a PRM with mean intensity $N\ dzds$. In setup (O), the clusters Z^i should be dominating linear N-dimensional Hawkes processes. While the total progeny distribution of Z^i is no longer Borel distributed, it is well known that it has exponential moment, which is sufficient to complete the proof.

Stability For More General Setups

A significant observation is that the setups (AD) and (O), essentially only affect the construction of ρ through the choice of α's and f. For other and more general setups in the univariate or multivariate case, one may adapt this procedure to establish stability regimes. For example it might be a method to explore other multivariate systems where inhibition from either the weight or the age have a potential effect on the stability regime.

Optimizing the Regeneration Scheme And Improving Results

These results establish a regeneration scheme for weight functions h s.t. $\int_0^\infty |h(t)|\, t^{p+1} \ln_+ t dt < \infty$. However, invariant solutions to Z exist already for h with first moment i.e. $\int_0^\infty t\,|h(t)|\, dt < \infty$. Also, the CLT result for ordinary Hawkes processes by Zhu [44], assuming that h decreasing and positive, only requires that $t \mapsto th(t)$ is integrable. This corresponds to $p = 0$, instead of $p \geq 2$ which we require in theorem 5.5.2. These facts indicate that there may exist renewal times with better moment properties, than those discussed in this chapter.

Also, the main result in [19], developed simultaneously with the results presented in this paper, proves existence of a regeneration point with improved integrability properties, in the case where the Hawkes process is linear with positive weight functions. Moreover, the exponential moment results in [19] are more refined, primarily due to the use of renewal times of a certain M/G/∞ queue with a deterministic service time instead of random exchange processes.

Implementation and Practical Computation

While this chapter focuses on the theoretical development of regeneration times, the method is constructive and ρ may be simulated in either setup. It would be of interest to study the efficiency of this algorithm.

5.7 Proofs

5.7.1 Proofs of Section 2 Results

Proof of Proposition 5.2.4

To show the claimed result for f under assumption 5.3 (A), substitute the inner variable with $u = s + t$ and apply Tonelli to obtain

$$\int_0^\infty \int_0^\infty t^p \gamma(s+1)\,\overline{h}(t+s)\,dsdt = \int_0^\infty \int_t^\infty t^p \gamma(u-t+1)\,\overline{h}(u)\,dudt \tag{5.58}$$

$$= \int_0^\infty \int_0^u t^p \gamma(u-t+1)\,\overline{h}(u)\,dtdu \tag{5.59}$$

$$\leq (p+1)^{-1} \int_0^\infty u^{p+1} \gamma(u+1)\,\overline{h}(u)\,du \tag{5.60}$$

which proves the desired. Under assumption 5.3 (B) it is straightforward to show that f has exponential moment. The claimed result for F^p, F follows immediately.

5.7.2 Proofs of Section 3.1 Results

Proof of Theorem 5.3.3
1. We claim for $n \in \mathbb{N}$ that

$$\left| \int_0^{\alpha_n} h(t-s)\,dZ_s^* + R_t^* \right| \leq f(t-\alpha_n) \quad \forall t > \alpha_n \tag{5.61}$$

if $\alpha_n < \infty$. We prove this by induction over $n \in \mathbb{N}_0$. The induction start $n = 0$ is per assumption 5.2. To prove the induction step, we split the integral of interest

$$\left| \int_0^{\alpha_n} h(t-s)\,dZ_s^* + R_t^* \right| \leq \left| \int_{\alpha_{n-1}}^{\alpha_n} h(t-s)\,dZ_s^* \right| + \left| \int_0^{\alpha_{n-1}} h(t-s)\,dZ_s^* + R_t^* \right|. \tag{5.62}$$

By the induction assumption, the 2nd term is bounded by $f(t-\alpha_{n-1})$ for all $t > \alpha_{n-1}$. The first term above can be split up to whether jumps of Z^* happen when $A_t^* \leq \delta$ or not

$$\int_{\alpha_{n-1}}^{\alpha_n} \mathbb{1}\{A_s \leq \delta\}\,h(t-s)\,dZ_s^* + \int_{\alpha_{n-1}}^{\alpha_n} \mathbb{1}\{A_s > \delta\}\,h(t-s)\,dZ_s^*. \tag{5.63}$$

Consider the first term. By the (AD) criteria we have

$$\int_a^b \mathbb{1}\{A_s \le \delta\}\, dZ_s^* \le \pi^K(a,b] := \int_{(a,b\}\times[0,K]} d\pi(s,z)\,. \tag{5.64}$$

On the interval $t \in (\alpha_{n-1}, \tau_n)$, we have per definition of τ_n that $N[t,t] \ge \pi^K[t,t]$ while for $t \in (\tau_n, \alpha_n]$ we have $\pi^K[t,t] = N[t,t]$. It follows that for $t > \alpha_n$, we have

$$\int_{\alpha_{n-1}}^{\alpha_n} \mathbb{1}\{A_s \le \delta\}\, h(t-s)\, dZ_s^* \tag{5.65}$$

$$\le \int_{\alpha_{n-1}}^{\alpha_n} \overline{h}(t-s)\, d\pi_s^K \tag{5.66}$$

$$\le \overline{h}(t-\tau_n) + \int_{\alpha_{n-1}}^{\alpha_n} \overline{h}(t-s)\, dN_s \tag{5.67}$$

$$\le \overline{h}(t-\alpha_n) + \sum_{i=0}^{\alpha_n-\alpha_{n-1}-1} \gamma(i)\, \overline{h}(t-\alpha_n+i) \tag{5.68}$$

$$\le \int_0^{\alpha_n-\alpha_{n-1}-1} \gamma(s+1)\, \overline{h}(t-\alpha_n+s)\; ds + (1+\gamma(0))\, \overline{h}(t-\alpha_n)\,. \tag{5.69}$$

For the second integral in (5.63), recall that $\delta^{-1} \in \mathbb{N}$ and notice that the interdistance between jumps of Z^* with $A_s > \delta$ is at least δ per definition. We obtain the bound

$$\int_{\alpha_{n-1}}^{\alpha_n} \mathbb{1}\{A_s > \delta\}\, h(t-s)\, dZ_s^* \tag{5.70}$$

$$\le \sum_{i=0}^{(\alpha_n-\alpha_{n-1})\delta^{-1}-1} \overline{h}(t-\alpha_n+i\delta) \tag{5.71}$$

$$\le \sum_{j=0}^{(\alpha_n-\alpha_{n-1})-1} \delta^{-1}\overline{h}(t-\alpha_n+j) \tag{5.72}$$

$$\le \int_0^{\alpha_n-\alpha_{n-1}-1} \delta^{-1}\overline{h}(t-\alpha_n+s)\; ds + \delta^{-1}\overline{h}(t-\alpha_n)\,. \tag{5.73}$$

The sum of the two right-hand sides of (5.69) and (5.73) are less than $f(t-\alpha_n, \alpha_n-\alpha_{n-1}-1) - r(t-\alpha_n)$. The induction claim now follows by inserting this back into (5.62).

To prove that α_0^I satisfies assumption 5.2, repeat the proof above with $-\infty$ in place of α_{n-1} and α_0^I in place of α_n. We omit the details.

2. Consider now the (O) setup. We claim again for $n \in \mathbb{N}$ that

$$\left| \int_0^{\alpha_n} h\left(t-s\right) dZ_s^* + R_t^* \right| \leq f\left(t-\alpha_n\right) \quad \forall t > \alpha_n. \tag{5.74}$$

As before we proceed by induction over $n \in \mathbb{N}_0$. The induction start follows from assumption 5.2. Assume now the claim holds for $n-1$. By the induction assumption and the definition of f we have

$$\left| \int_0^{\alpha_n} h\left(t-s\right) dZ_s^* + R_t^* \right| \leq \left| \int_{\alpha_{n-1}}^{\alpha_n} h\left(t-s\right) dZ_s^* \right| + f\left(t-\alpha_{n-1}\right). \tag{5.75}$$

It is seen per induction over jumps of Z_n, that

$$Z_n\left[s,s\right] \geq Z^*\left[\alpha_{n-1}+s, \alpha_{n-1}+s\right]$$

for all $s \in (\alpha_{n-1}, \alpha_n]$. Hence

$$\begin{aligned} &\left| \int_{\alpha_{n-1}}^{\alpha_n} h\left(t-s\right) dZ_s^* \right| \\ \leq &\left| \int_0^{\alpha_n - \alpha_{n-1}} h\left(t-s\right) dZ_s^n \right| \\ \leq &f(t-\alpha_n, \alpha_n - \alpha_{n-1} - 1) - r\left(t-\alpha_n\right) \end{aligned} \tag{5.76}$$

inserting back into (5.75), and using the definition of f gives the desired result.

The statement about α_0^I is a direct implication of the claim that $Z^I\left[t,t\right] \leq Z^0\left[t,t\right]$ for all $t \in \mathbb{R}$ almost surely. To prove this claim, let Z^1, Z^2 be Hawkes processes with weight h_+, h and rate functions ψ_L, ψ respectively, with common intialization $R \equiv 0, A_0 = 0$. By the coupling property from example 5.2.5 We have almost surely that $Z^0_{|[t,\infty)} = Z^1_{|[t,\infty)}, Z^I_{|[t,\infty)} = Z^2_{|[t,\infty)}$ for t large enough. On the other hand, per induction over jumps of Z^1 it is straightforward to prove that $Z^1[t,t] \geq Z^2[t,t]$ for all $t \in \mathbb{R}_+$ so it follows that almost surely $Z^0[t,t] \geq Z^I[t,t]$ eventually. The claim now follows from the fact that $\left(Z^0, Z^I\right)$ is stationary and ergodic.

□

Proof of Proposition 5.3.4

Define $\tilde{Z} = |d(Z^* - Z)|$ and let $\tilde{A}$ be its age process. We have for $t \in (\alpha_{n-1}, \tau_n]$

$$X_t^* - X_t = \int_{\alpha_{n-1}}^{t-} h(t-s)\, d(Z_s^* - Z_s) + \int_0^{\alpha_{n-1}} h(t-s)\, dZ_s^* + R^*(t) - f(t - \alpha_{n-1}). \tag{5.77}$$

Applying (5.27) gives

$$\left| \int_0^{\alpha_{n-1}} h(t-s)\, dZ_s^* + R^*(t) - f(t-\alpha_{n-1}) \right| \leq 2f(t - \alpha_{n-1}). \tag{5.78}$$

Define for $n \in \mathbb{N}$

$$\tau_n^* = \inf\left\{ t > \alpha_{n-1} : \tilde{Z}[t,t] = 1 \right\}. \tag{5.79}$$

We claim that $\tau_n^* \geq \tau_n$ almost surely. Notice that $A_t^* \geq A_t$ for $t \in (\alpha_{n-1}, \tau_n^*]$. For all $t \in (\alpha_{n-1}, (\alpha_{n-1} + D) \wedge \tau_n^*]$ we combine (5.27) and (5.5) to see that

$$0 = \lambda_t \leq \lambda_t^* \leq c_\psi + Lf(t - \alpha_{n-1}) \leq F(t - \alpha_{n-1}) \tag{5.80}$$

which shows that $\tau_n^* \geq \tau_n \wedge (\alpha_{n-1} + D)$. For all t in the (possibly empty) interval $(\alpha_{n-1} + D, \tau_n^*]$ we have

$$\begin{aligned} X_t^* &= \int_{\alpha_{n-1}}^{t-} h(t-s)\, dZ_s^* + \int_0^{\alpha_{n-1}} h(t-s)\, dZ_s^* + R_t^* && (5.81) \\ &\geq \int_{\alpha_{n-1}}^{t-} h(t-s)\, dZ_s^* - f(t - \alpha_{n-1}) && (5.82) \\ &= X_t. && (5.83) \end{aligned}$$

Since ψ is increasing we get the following inequality

$$\psi(X_t, A_t) \leq \psi(X_t^*, A_t^*) \quad \text{for all } t \in (\alpha_{n-1} + D, \tau_n^*]. \tag{5.84}$$

If $A_t^* > A_t$ the definition of τ_n^* gives that $A_t^*, A_t \geq t - \alpha_{n-1}$. Therefore (5.77) and (5.78) give

$$\begin{aligned} \psi(X_t^*, A_t^*) - \psi(X_t, A_t) &\leq L(X_t^* - X_t) + c_\psi g(t - \alpha_{n-1}) && (5.85)\\ &\leq 2Lf(t - \alpha_{n-1}) + c_\psi g(t - \alpha_{n-1}) && (5.86)\\ &\leq F(t - \alpha_{n-1}). && (5.87)\end{aligned}$$

Likewise, if $A_t^* = A_t$ we have

$$\psi(X_t^*, A_t^*) - \psi(X_t, A_t) \leq L(X_t^* - X_t). \tag{5.88}$$

By definition of τ_n, this implies $\tau_n^* \geq \tau_n$. Thus, in between two consecutive τ stopping times, the two Hawkes processes agree. □

Proof of Theorem 5.3.5

The results (5.34),(5.35) are straightforward consequences of the strong Markov Property and [5] lemma 1. The density of the conditional distribution $\tau_n - \alpha_{n-1} | \tau_n < \infty$ is proportional to

$$F(t) \exp\left(-\int_0^t F(s)\, ds\right) \leq F(t) \tag{5.89}$$

which shows the desired moment results for the distribution . □

Proof of Theorem 5.3.6

To structure the proof, we discuss the following four variables, in written order

1. α_0^I in the (AD) setup
2. α_n in the (AD) setup
3. α_0^I in the (O) setup
4. α_n in the (O) setup

1. To study α_0^I, we introduce the process M_i for $i \geq 0$ given by

$$\begin{aligned} M_i &= \inf\left\{m \geq 0 : \theta^{i-j} N(-1, 0] \leq \gamma(j + m), j \geq 0\right\} && (5.90)\\ &= \inf\left\{m \geq 0 : \theta^k N(-1, 0] \leq \gamma(i + m - k), k \leq i\right\}. && (5.91)\end{aligned}$$

In words, the value M_i is the minimum number of units the function $j \mapsto \gamma(i+j)$ must be shifted before it bounds π^K on all intervals $(i-j-1, i-j]$, $j \in \mathbb{N}_0$, see fig. 5.2.

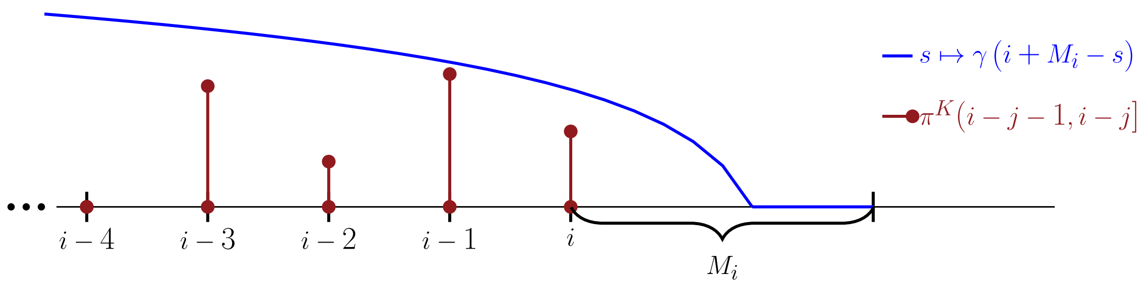

Figure 5.2: An illustration of M_i.

We see that $\alpha_0 = \inf\{i > 0 : M_i = 0\}$ i.e α_0 is the first time M_i hits 0. Notice that M_0 is a well-defined random variable by a Borel-Cantelli argument. Observe also that M is in fact an RE-process with update scheme

$$M_i = (M_{i-1} - 1) \vee \lceil \gamma^{-1}(N(i-1, i]) \rceil. \tag{5.92}$$

By corollary 7.4.3 M has a unique invariant distribution μ and since $N \overset{d}{=} \theta^1 N$ it follows that $M_0 \overset{d}{=} M_1 \sim \mu$. The result follows from corollary 7.4.3 part 2.

2. Notice that $\alpha_n - \tau_n$ has p'th-moment / exponential moment iff $\alpha_n - \lceil \tau_n \rceil$ has as well. For any realization such that $\tau_n < \infty$ we may write

$$\alpha_n - \lceil \tau_n \rceil = \inf\left\{i > 0 : \theta^{i-j+\lceil \tau_n - \alpha_{n-1} \rceil} N_{\alpha_{n-1}+}(-1, 0] \leq \gamma(j), j = 0, \ldots, i + \lceil \tau_n - \alpha_{n-1} \rceil - 1\right\}. \tag{5.93}$$

We now proceed as previously. Define M_i' for $i \geq 0$ as the process

$$M_i' = \inf\left\{m \geq 0 : \theta^{i-j+\lceil \tau_n - \alpha_{n-1} \rceil} N_{\alpha_{n-1}+}(-1, 0] \leq \gamma(j+m), j = 0, \ldots, i + \lceil \tau_n - \alpha_{n-1} \rceil - 1\right\}. \tag{5.94}$$

Notice that $\alpha_n - \lceil \tau_n \rceil$ is the first time M_i' hits 0. From theorem 7.3.1 we have $\pi^{\uparrow\lambda,\bar{\lambda}}_{\alpha_{n-1}+} \perp\!\!\!\perp \pi^{\downarrow\lambda,\bar{\lambda}}_{\alpha_{n-1}+}$ and thus $\tau_n - \alpha_{n-1}$ is independent of $N_{\alpha_{n-1}+}$. Observe also that M' is an RE-process defined by

$$M_i' = (M_{i-1}' - 1) \vee \lceil \gamma^{-1}\left(N_{\lceil \tau_n \rceil +}(i-1, i]\right) \rceil. \tag{5.95}$$

To study the distribution of M'_0 note that it may be described as $M''_{\alpha_n - \lceil \tau_n \rceil}$ where M'' is another RE-process defined by $M''_0 = 0$ and

$$M''_i = (M''_{i-1} - 1) \vee \lceil \gamma^{-1} \left(N_{\alpha_{n-1}+} (i-1, i] \right) \rceil. \tag{5.96}$$

Let P^* be the conditional distribution of $\lceil \tau_n - \alpha_{n-1} \rceil$ given $\tau_n < \infty$. If ϕ is a positive increasing function and $\left(P^k_x\right)$ is the k-step Markov kernel for M'' then

$$\mathbb{E}\left(\phi\left(M'_0\right)|\tau_n < \infty\right) = \int \int \phi(y)\, dP^k_0(y)\, dP^*(k) \tag{5.97}$$
$$\leq \int \phi(y)\, \mu(y) + \int \int \phi(y)\, d\left|P^k_0 - \mu\right|(y)\, dP^*(k). \tag{5.98}$$

From theorem 14.1.4 [34] it follows that if $\mu(\phi) < \infty$, then the second term above is finite. The desired result now follows from theorem 5.3.5 and corollary 7.4.3.

3. We now analyze α^I_0 under setup (O). Here we utilize that the law of the stationary Linear Hawkes process Z^0 has a cluster process representation which we now describe; Let N be a Poisson process on $\mathbb{R}$ with intensity c_ψ and for $i \in \mathbb{Z}$ let (Z_i) be independent Hawkes processes with weight/rate $h_+, x \mapsto Lx$ and initialized with a single jump at $t = 0$ (i.e. $Z_i[0,0] = 1$ and $R^i_t = h_+(t)$). Define now

$$\overline{Z}(A) = \sum_{i \in \mathbb{Z}} C_i(A) \tag{5.99}$$

where C_i is a random measure given by $C_i(A) = Z_i(A - s_i)$ and s_i is the i'th jump of N for $i \in \mathbb{Z}$ (with convention $s_1, s_2, ...$ being the first jumps after $t = 0$, and $s_0, s_{-1}, ...$ being the most recent jumps before $t = 0$). Then $\overline{Z}$ is distributed as the stationary linear Hawkes process with weight/rate h_+/ψ_L. See [11],[41] for more details on this construction. It follows that $\alpha^I_0 \overset{D}{=} \tilde{\alpha}_0$ where

$$\tilde{\alpha}_0 = \inf\{i > 0 : \left| \int_{-\infty}^{i} h_+(t-s)\, d\overline{Z}_s \right| \leq f(t-i) - r(t-i), \forall t > i\}. \tag{5.100}$$

We shall use the following fact, coming from section 1.1 in [41] and the proof of proposition 1.2 in same refrence: We may assume that the clusters Z^i are constructed s.t. there is i.i.d. $(W_i, X_{i,j})$, also independent of N such that $W_i \perp\!\!\!\perp X_{i,j}$, $X_{i,j} \sim \|h_+\|_{\mathcal{L}^1}^{-1} h_+(t)\, dt$ and

$$Z_i(\mathbb{R}) = Z_i(-\infty, Y_i] = W_i, \tag{5.101}$$

where $Y_i = \sum_{j=1}^{W_i} X_{i,j}$ and W_1 is distributed as the *total progeny* of a Poisson branching process, with mean offspring $\|h_+\|_{\mathcal{L}^1}^{-1}$.

By the Otter-Dwass formula (see [16]) The p.m.f. of W_1 is

$$p_W(n) = \frac{n^{n-1}}{\|h_+\|_{\mathcal{L}^1} n!} e^{n\left(-\|h_+\|_{\mathcal{L}^1} + \ln\|h_+\|_{\mathcal{L}^1}\right)}. \tag{5.102}$$

The stirling approximation for $n!$ gives that

$$\mathbb{E}\exp(cW) < \infty \iff c \le c_h = \|h_+\|_{\mathcal{L}^1} - \ln_+ \|h_+\|_{\mathcal{L}^1} - 1 \tag{5.103}$$

If assumption 5.3 (B) holds, take any $c_0 > 1$. Otherwise, let $c_0 > 1$ be a constant satisfying $\gamma(t) \ge c_0 (p+1) c_h^{-1} \ln_+ t$ for t sufficiently large. Define $\gamma^*(t) = c_0^{-1}\gamma\left(c_0^{-1}t - 1\right)$ when $c_0^{-1}t \ge 1$ and $\gamma^*(t) = 0$ otherwise.

With $N_t := N[0,t]$ for $t > 0$ and $-N(t,0)$ otherwise we define

$$\overline{Y}_i = \max_{l=N_{i-1}+1}^{N_i} \{Y_l\}, \quad \overline{W}_i = \sum_{l=N_{i-1}+1}^{N_i} W_l, \tag{5.104}$$

$$\overline{\alpha}_0 = \inf\{i > 0 : \overline{Y}_{i-j} \le \left(1 - c_0^{-1}\right) j,\ \overline{W}_{i-j} \le \gamma^*(j),\ j \ge 0\}. \tag{5.105}$$

We claim that $\tilde{\alpha}_0 \le \overline{\alpha}_0$. This follows from the calculations for $t > \overline{\alpha}_0$

$$\int_{-\infty}^{\overline{\alpha}_0} h_+(t-s) d\overline{Z}_s = \sum_{j=0}^{\infty} \int_{\overline{\alpha}_0 - j - 1}^{\overline{\alpha}_0 - j} \int_s^t h_+(t-u) dC_{N_s}(u) dN_s \tag{5.106}$$

Note that (5.101) implies that for all clusters C_{N_s} with $s \in (\overline{\alpha_0} - j - 1, \overline{\alpha_0} - j]$ we have $Supp\left(C_{N_s}\right) \subset \left[s, s + \overline{Y}_{\overline{\alpha_0} - j}\right]$. By the inequalities obtained from the definition of $\overline{\alpha}_0$ we have $\left[s, s + \overline{Y}_{\overline{\alpha_0} - j}\right] \subset \left[s, \overline{\alpha_0} - c_0^{-1} j\right]$. Since $h_+ \le \overline{h}$ and $\overline{h}$ is decreasing we conclude that $\overline{h}(t-u) \le \overline{h}\left(t - \overline{\alpha}_0 + c_0^{-1} j\right)$ for all $u \in Supp\,(C_{N_s})\,, s \in (\overline{\alpha_0} - j - 1, \overline{\alpha_0} - j]\,, t > \alpha_0$. Inserting this into (5.106) gives

$$\int_{-\infty}^{\overline{\alpha}_0} h_+(t-s)d\overline{Z}_s \leq \sum_{j=0}^{\infty} \overline{h}(t - \overline{\alpha}_0 + c_0^{-1}j) \int_{\overline{\alpha}_0 - j - 1}^{\overline{\alpha}_0 - j} C_{N_s}[s, \overline{\alpha}_0 - c_0^{-1}j]dN_s. \tag{5.107}$$

$$\leq \sum_{j=0}^{\infty} \overline{h}(t - \overline{\alpha}_0 + c_0^{-1}j)\overline{W}_{\overline{\alpha}_0 - j} \tag{5.108}$$

Again, by the inequalities defining $\overline{\alpha}_0$ we get

$$\int_{-\infty}^{\overline{\alpha}_0} h_+(t-s)d\overline{Z}_s \leq \int_0^{\infty} \overline{h}(t - \overline{\alpha}_0 + c_0^{-1}s)\gamma^*(s+1)ds + \overline{h}(t - \overline{\alpha}_0)\gamma^*(0) \tag{5.109}$$

$$\leq \int_0^{\infty} \overline{h}(t - \overline{\alpha}_0 + s)\gamma(s+1)ds + \overline{h}(t - \overline{\alpha}_0)\gamma(0) \tag{5.110}$$

$$\leq f(t - \overline{\alpha}_0) - r(t - \overline{\alpha}_0) \tag{5.111}$$

which proves the claim.

To describe the moment of $\overline{\alpha}_0$ define

$$M_i = \inf\left\{m \geq 0 : \overline{Y}_{i-j} \leq \left(1 - c_0^{-1}\right)(j+m)\,,\ \overline{W}_{i-j} \leq \gamma^*(j+m)\,, j \geq 0\right\}. \tag{5.112}$$

Indeed, $M_0 < \infty$ by a Borel-Cantelli argument. As before $\overline{\alpha}_0$ is the return time to 0 for the RE-process M_i with update-variables

$$\lceil \left(\left(1 - c_0^{-1}\right)^{-1} \overline{Y}_i\right) \vee (\gamma^*)^{-1}\left(\overline{W}_i\right) \rceil$$

and started at M_0, carrying the invariant distribution of M. Under assumption 5.3 (A) we have $(\gamma^*)^{-1}(t) \leq \exp\left(c_h\,(p+1)^{-1}\,t\right)$ for t large. It follows that the update variables have $(p+1)$th moment and the starting distribution has p'th moment, so it follows from corollary 7.4.3 that $\overline{\alpha}_0$ has p'th moment. Likewise, under assumption 5.3 (B) the update variables have exponential moment and the starting distribution has exponential moment as well. It follows from corollary 7.4.3 that $\overline{\alpha_0}$ has exponential moment.

4. We now analyze α_n under setup (O). To outline the similarity we shall re-use some of the notation from the previous proof, but for slightly modified random variables.

Notice that $Z^n[t,t]$ is dominated by the Hawkes process Z' with the same weight function h_+, rate function ψ_L and initial signal $R'_t = f(t-\lfloor t \rfloor) + h_+(t-(\tau_n - \alpha_{n-1}))$. The law of Z' has a cluster process representation given as follows:
Let N be an inhomogeneous Poisson process with intensity $c_\psi + Lf(t-\lfloor t \rfloor)$, and let $\xi \sim (\tau_n - \alpha_{n-1}) \mid (\tau_n < \infty)$. Define Z_i as before, and let C_i, C^ξ be the mutually independent clusters given by $C_i(A) = Z_i(A - s_i)$ where s_i is the ith jump of N for $i \in \mathbb{N}$, and $C^\xi(A) = Z_{-1}(A - \xi)$. Define

$$\overline{Z}(A) = \sum_{i=0}^{\infty} C_i(A) + C^\xi(A), \quad A \in \mathcal{B}_{\mathbb{R}_+}. \tag{5.113}$$

Then $\overline{Z} \overset{D}{=} Z'$ and hence also

$$\begin{aligned} &(\alpha_n - \alpha_{n-1} \mid \tau_n < \infty) &(5.114)\\ &\overset{D}{\leq} \inf \Bigg\{ i > \lceil \xi \rceil : \\ &\left| \int_0^i h_+(t-s)\, d\overline{Z}_s \right| \leq f(t-i, i-1) - r(t-i), \forall t > i, \overline{Z}[0,i] \leq \int_0^i \gamma(s+1)\, ds \Bigg\} &(5.115) \end{aligned}$$

where $\overset{D}{\leq}$ denotes inequality in the usual stochastic order(see [3] chapter 2). As before, we have i.i.d. variables $(W_i, X_{i,j}), (W^\xi, X^\xi_j)$ and also independent of N, ξ such that $X_{i,j} \sim \|h_+\|^{-1}_{\mathcal{L}^1} h(t)\, dt$, $W_i \sim Z_0(\mathbb{R})$ and

$$Z_i(\mathbb{R}) = Z_i(-\infty, Y_i] = W_i, \quad \forall i \in \mathbb{N} \tag{5.116}$$

$$Z_{-1}(\mathbb{R}) = Z_{-1}(-\infty, Y^\xi] = W^\xi \tag{5.117}$$

where $Y_i = \sum_{j=1}^{W_i} X_{i,j}$ and $Y^\xi = \sum_{j=1}^{W^\xi} X^\xi_j$. Define now

$$\overline{Y}_i = \max_{l=N_{i-1}+1}^{N_i} \{Y_l\}, \quad \overline{W}_i = \sum_{l=N_{i-1}+1}^{N_i} W_l \tag{5.118}$$

and notice that $(\overline{Y}_i, \overline{W}_i)$ for $i \in \mathbb{N}$ is an i.i.d. sequence. Define now c_0, γ^* as previously and set

$$\overline{\zeta} = \inf\{i > 0 : \overline{Y}_{i+\lceil\xi\rceil-j} \leq \left(1 - c_0^{-1}\right) j,\ \overline{W}_{i+\lceil\xi\rceil-j} \leq \gamma^* (j)\ \ \forall j \in \{0, ..., i + \lceil\xi\rceil - 1\} \setminus \{i\}, \tag{5.119}$$

$$\overline{Y}_{\lceil\xi\rceil} \vee Y^{\xi} \leq \left(1 - c_0^{-1}\right) i,\ \ \overline{W}_{\lceil\xi\rceil} + W^{\xi} \leq \gamma^* (i)\}. \tag{5.120}$$

Define now the random time ζ s.t. $\zeta + \lceil\xi\rceil$ is equal to (5.115). We claim that $\zeta \leq \overline{\zeta}$. To prove this, we apply similar arguments as in the part of **3**) where we showed that
$\tilde{\alpha}_0 \leq \overline{\alpha}_0$. By construction of $\overline{Y}_i$'s and $\overline{\zeta}$, it holds for all
$s \in \left(\overline{\zeta} + [\xi] - j - 1, \overline{\zeta} + \lceil\xi\rceil - j\right)$ that $Supp\left(C_{N_s}\right) \subset \left(-\infty, \overline{\zeta} + \lceil\xi\rceil - c_0^{-1} j\right]$ and this implies

$$\int_0^{\lceil\xi\rceil+\overline{\zeta}} h_+(t-s) d\overline{Z}_s \tag{5.121}$$

$$\leq \sum_{j=0, j\neq\overline{\zeta}}^{\lceil\xi\rceil+\overline{\zeta}-1} \overline{h}(t - \left(\lceil\xi\rceil + \overline{\zeta}\right) + c_0^{-1} j) \overline{W}_{\lceil\xi\rceil+\overline{\zeta}-j} \tag{5.122}$$

$$+\overline{h}(t - \left(\lceil\xi\rceil + \overline{\zeta}\right) + c_0^{-1}\overline{\zeta}) \left(\overline{W}_{\lceil\xi\rceil} + W^{\xi}\right)$$

$$\leq \int_0^{\lceil\xi\rceil+\overline{\zeta}-1} \overline{h}(t - \left(\lceil\xi\rceil + \overline{\zeta}\right) + c_0^{-1} s) \gamma^*(s+1) ds + \overline{h}(t - \left(\lceil\xi\rceil + \overline{\zeta}\right)) \gamma^*(0) \tag{5.123}$$

$$\leq f(t - \left(\lceil\xi\rceil + \overline{\zeta}\right), \lceil\xi\rceil + \overline{\zeta} - 1) - r(t - (\overline{\zeta} + \lceil\xi\rceil)). \tag{5.124}$$

Likewise, the support constraint on C_{N_s} mentioned above must imply that

$$\overline{Z}\left(0, \overline{\zeta} + \lceil\xi\rceil\right] = \sum_{j=0}^{\overline{\zeta}+[\xi]-1} W_j \leq \sum_{j=0}^{\overline{\zeta}+\lceil\xi\rceil-1} \gamma^* (j) \leq \int_0^{\overline{\zeta}+\lceil\xi\rceil-1} \gamma (s+1) \, ds. \tag{5.125}$$

This proves our claim. To analyze $\overline{\zeta}$ define

$$M_i = \inf\left\{m \geq 0 : \overline{Y}_{i-j} \leq \left(1 - c_0^{-1}\right)(j+m),\ \overline{W}_{i-j} \leq \gamma^* (j+m),\ j = 0, ..., i-1\right\}. \tag{5.126}$$

This is an RE-process started at $M_0 = 0$ and with update variables $\lceil \max\{(\gamma^*)^{-1}(\overline{W}_i), (1 - c_0^{-1})^{-1} \overline{Y}_i\} \rceil$. As in the proof of **2)** it follows that

$M_{[\xi]-1}$ has p'th moment, and under assumption 5.3(B) it has exponential moment. The same therefore holds for the variable

$$M_0' := (M_{[\xi]-1} - 1) \vee \lceil \max\{(\gamma^*)^{-1}(\overline{W}_{[\xi]} + W^\xi), (1 - c_0^{-1})^{-1}\overline{Y}_{[\xi]} \vee Y^\xi\} \rceil. \tag{5.127}$$

Now realize that $\overline{\zeta}$ is the return time to 0 for the RE-process with update variables

$$\lceil \left(\left(1 - c_0^{-1}\right)^{-1} \overline{Y}_i \right) \vee (\gamma^*)^{-1}(\overline{W}_i) \rceil \tag{5.128}$$

and started at M_0'. The desired result now follows from corollary 7.4.3 $\square$

Proof of Theorem 5.3.2

1. To prove that ρ is a $\mathcal{F}^*$ stopping time, we notice that

$$(\rho \leq t) = (\alpha_\eta \leq t - D) = \bigcup_{i=0}^{\lfloor t-D \rfloor} (\alpha_\eta = i). \tag{5.129}$$

So it suffices to show $(\alpha_\eta = i) \in \mathcal{F}_i^*$. Indeed, this is true since

$$(\alpha_\eta = i) = \bigcup_{k=0}^{\infty} (\alpha_k = i) \cap \left(\int_i^\infty \mathbb{1}\{z \leq F(s)\} d\pi^{\uparrow \underline{\lambda}, \overline{\lambda}} \right). \tag{5.130}$$

2. By applying theorem 7.3.1 for each fixed t we see that $\pi^{\downarrow \underline{\lambda}, \overline{\lambda}}$ is an $\mathcal{F}_t^*$-PRM. The Strong Markov Property gives that $\pi_{\rho+}^{\downarrow \underline{\lambda}, \overline{\lambda}}$ is a PRM independent of $\mathcal{F}_\rho^*$. Notice now that $Z_+, \lambda_{+t} := \lambda_{\rho+t}$ is exactly the point process and intensity of the ADHP driven by $\pi_{\rho+}^{\downarrow \underline{\lambda}, \overline{\lambda}}$ with initial age D and signal $R : t \mapsto -f\,(t + D)$. That is, $Z_{\rho+}$ is entirely generated by $\pi_{\rho+}^{\downarrow \underline{\lambda}, \overline{\lambda}}$ and hence $Z_{\rho+} \perp\!\!\!\perp \mathcal{F}_\rho^*$. Notice that $Z_{\rho+} = Z_{\rho+}^*$ by proposition 5.3.4 and that $Z_{|(0,\rho]}^* \subset \mathcal{F}_\rho^*$. It now follows that $Z_{\rho+}^* \perp\!\!\!\perp Z_{|(0,\rho]}^*$.

3. We introduce an i.i.d. sequence of random variables (β_i) with distribution $\beta_1 \sim \alpha_1 - \alpha_0 | \tau_1 - \alpha_0 < \infty$. We then introduce another sequence of random variables given by

$$\tilde{\beta}_i = \begin{cases} \alpha_i - \alpha_{i-1} & i \leq \eta \\ \beta_i & i > \eta. \end{cases} \tag{5.131}$$

Recall that $(\alpha_j - \alpha_{j-1})_{1 \le j \le i}$ are conditionally i.i.d. given $\tau_i < \infty$. From this we see that $\tilde{\beta}$ is an i.i.d. sequence of variables distributed as β_1, and which is also independent of η. We may now write

$$\alpha_\eta - \alpha_0 = \sum_{i=1}^{\eta} \alpha_i - \alpha_{i-1} \tag{5.132}$$

$$= \sum_{i=1}^{\eta} \tilde{\beta}_i. \tag{5.133}$$

From theorem 5.3.5 1) it follows that η is distributed as a negative binomial, and in particular it has exponential moment. To study the distribution of $\tilde{\beta}_1$, we use that

$$\alpha_1 - \alpha_0 = (\alpha_1 - \tau_1) + (\tau_1 - \alpha_0)\,. \tag{5.134}$$

Theorem 5.3.5 and theorem 5.3.6 give that β has p'th moment so the desired result follows from theorem 5.2, chapter 1 [21]. Under assumption 5.3(B), one may conclude the proof by writing for small $c > 0$

$$\mathbb{E}\exp(c\alpha_\eta) = \mathbb{E}\exp(c\alpha_0)\ \mathbb{E}\prod_{i=0}^{\eta}\exp(c\tilde{\beta}_i) \tag{5.135}$$

$$= \mathbb{E}\exp(c\alpha_0)\ \mathbb{E}\left[\mathbb{E}\left(\prod_{i=0}^{\eta}\exp(c\tilde{\beta}_i)\ \Big|\ \eta\right)\right] \tag{5.136}$$

$$= \mathbb{E}\exp(c\alpha_0)\ \mathbb{E}\left(\mathbb{E}\exp(c\tilde{\beta}_1)\right)^{\eta}. \tag{5.137}$$

Since η has exponential moment, the above expression is finite for small c. □

5.7.3 Proofs of Section 3.2 Results

Proof of Proposition 5.4.1
The proof is a coupling argument. Consider the Hawkes process Z^* driven by a fixed PRM π and started with $R : t \mapsto f(t + D)\,, A_0 = D$. Then assumption 5.2 is satisfied with $\alpha_0 = 0$, $r = -f$. Thus, defining Φ as in (5.39) we have a coupling in the sense that $(\theta^n Z^*)_{|(-D,0]} = H^*(\Phi_n)$ for

all $n > \rho_0$. Since $\tilde{P}$ is the invariant distribution of Φ the markov chain converges in total variation and hence also $H^* (\Phi_n) \Rightarrow H^* \left(\tilde{P}\right)$. On the other hand, we have by the coupling property of Z^I that almost surely, there is a random integer n_0 s.t.
$(\theta^n Z^*)_{|(-D,0]} = \left(\theta^n Z^I\right)_{|(-D,0]}$ for all $n \geq n_0$. It follows that $H^* (\Phi^n) = (\theta^n Z^*)_{(-D,0]} \Rightarrow Z^I_{(-D,0]}$ and the desired result follows from uniqueness of limits. $\square$

Chapter 6

Mean-field Limit and Propagation of Chaos

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

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

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

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

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

where r_t^k is a suitable limit of the initial processes in population k. In theorem 6.2.2, we discuss criteria under which such a system exists. Our second main theorem, theorem 6.3.1, shows that this system will indeed be a limit process for the generalized Hawkes processes for $N \to \infty$. We also discuss in lemma 6.3.2 how robust the system is to adjusting the weight functions. Not only is this robustness a good model feature in itself, but it also allows approximation of an arbitrary age dependent Hawkes process, using weight functions with better features. Examples are weight functions given by Erlang densities or exponential polynomials which induce Markovian systems, see [13].

6.1 Setup in This Chapter

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

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

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

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

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

$$\frac{N_k}{N} \overset{N\to\infty}{\to} p_k.$$

For a fixed $N \in \mathbb{N}$, we re-index the N-dimensional age dependent Hawkes process of (2.15) as

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

where the superscript kj denotes the jth unit in population k. The weight function from the ith unit of population l to the jth unit of population k is given by $N^{-1}h_{kl}$. Moreover, all units within the same population have the same spiking rate ψ^k. By taking L, c_ψ larger, we can and will assume that (4.2) and (4.3) hold in the chapter as well. Finally, we assume that all units in population k have the same initial signal R^k, and that the initial ages are interchangeable in groups and mutually independent in and between groups. With this set of parameters, the age dependent Hawkes process (Z, X, A) from (2.15) defined on $t \in \mathbb{R}_+$ becomes

$$
\begin{aligned}
Z_t^{kj} &= \int_0^t \int_0^\infty \mathbb{1}\left\{z \leq \psi^k\left(X_s^k, A_s^{kj}\right)\right\} d\pi^{kj}(ds, dz), \quad j \leq N_k, k \leq \mathcal{K}, \\
X_t^k &= \frac{1}{N} \sum_{l=1}^{\mathcal{K}} \sum_{j=1}^{N_l} \int_0^{t-} h_{kl}(t-s) Z^{lj}(ds) + R_t^k, \qquad k \leq \mathcal{K},
\end{aligned}
\tag{6.1}
$$

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

Model Observations

1. Suppose that the initial ages $(A_0^i)_{i \in \mathbb{N}}$ are exchangeable. Then the symmetry of the system gives interchangeability between units within the same population, i.e., $Z^{kj} \overset{\mathcal{L}}{=} Z^{ki}$ for $i, j \leq N_k, k \leq \mathcal{K}$.

2. In the mean-field setup, all units within a population k share the same memory process X^k.

6.2 The Limit System

We propose a limit system for $N \to \infty$. To pursue this goal, take finite variation functions $t \mapsto \alpha_t^k$, locally bounded functions $t \mapsto \beta_t^k$, and PRMs

π^k for $k \leq \mathcal{K}$, and consider the stochastic convolution equation

$$\begin{aligned} \phi_t^k &= \int_0^t \mathbb{E}\psi^k\left(x_s^k, A_s^k\right) ds, \\ x_t^k &= \sum_{l=1}^{\mathcal{K}} p_l \int_0^t h_{kl}\left(t-s\right) d\alpha_s^l + \beta_t^k, \\ A_t^k - A_0^k &= t - \int_0^{t-}\int_0^\infty A_s^k \mathbb{1}\left\{z \leq \psi^k\left(x_s^k, A_s^k\right)\right\}\pi^k\left(ds, dz\right), \end{aligned} \tag{6.2}$$

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

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

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

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

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

for large N, where $(Z^1, \ldots, Z^{\mathcal{K}})$ denotes the limit process such that each Z^k describes the jump activity of a typical unit belonging to population k. This relation invites the idea that the memory process for $N \to \infty$, $t \mapsto x_t$ should satisfy the integral system (6.2) with $\phi_t^k = \alpha_t^k = \mathbb{E}Z_t^k$ and $\beta_t^k = \mathbb{E}R_t^{k1}$. This motivates the following result.

Lemma 6.2.1.
Let $\beta_t = \left(\beta_t^k\right)_{k\leq\mathcal{K}}$ be measurable and locally bounded. There is a unique function α such that $\alpha = \phi$, where (ϕ, x, A) is the solution to (6.2). Moreover, ϕ is continuous and x is bounded on $[0, T]$ by a constant C which depends on $h := \sum_{k,j}|h_{kj}|$, $\|\beta\|_T$, T and L. $\circ$

The proof is given in the proof section succeeding this one. Once this lemma is established, we can ensure existence of the limit process.

Theorem 6.2.2.
Let $\beta_t = \left(\beta_t^k\right)_{k\leq K}$ be measurable and locally bounded. There is a unique solution (Z, A) to the integral equation

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

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

$\circ$

Proof. Let $(\phi, x, A)_{k\leq\mathcal{K}}$ be the tuple given in Lemma 6.2.1. Define the counting process

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

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

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

Defining $\tilde{x}_t^k = \sum_{l=1}^{\mathcal{K}} p_l \int_0^t h_{kl}\left(t-s\right) d\mathbb{E}\tilde{Z}_u^l + \beta_t^k$ and $\tilde{\phi}_t^k = \int_0^t \mathbb{E}\psi^k\left(\tilde{x}_s^k, \tilde{A}_s^k\right) ds$, we note that $\mathbb{E}\tilde{Z}_t^k = \tilde{\phi}_t^k$. Thus, if we insert $\alpha = \tilde{\phi}$ in (6.2), $\left(\tilde{\phi}, \tilde{x}, \tilde{A}\right)$ is a solution, and hence the uniqueness part of Lemma 6.2.1 gives that $\left(\tilde{\phi}, \tilde{x}, \tilde{A}\right) = (\phi, x, A)$ and thus also $Z = \tilde{Z}$. $\square$

6.3 Large Network Asymptotics and Weight Approximations

In this section we couple the N-dimensional Hawkes process with the limit system proposed in the previous section. This coupling implies that the finite-dimensional system converges to the limit system. The result is traditionally named *Propagation of Chaos*, a typical result within mean-field theory. Specifically for Hawkes processes, there are several variants of this result. Some of the recent results may be found in [7], [12] and in [13].

Framework for Propagation of Chaos

We will first introduce a set of assumptions.

Assumption 6.1.

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

$$\int_0^t \mathbb{E}\left|R_s^{Nk} - r_s^k\right| ds \to 0 \text{ as } N \to \infty, \tag{6.3}$$

  for all $t \geq 0$.
- The initial ages $A_0^{ki}, 1 \leq k \leq \mathcal{K}, 1 \leq i < \infty$ are i.i.d.
- The weight functions $h_{kl}^N : \mathbb{R}_+ \to \mathbb{R}$ satisfy $h_{kl}^N \to h_{kl}$ as $N \to \infty$ locally in $\mathcal{L}^1$, where $h_{kl} \in \mathcal{L}^2_{loc}$ for all $1 \leq k, l \leq \mathcal{K}$.

○

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

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

given by (6.1), driven by (π^{kj}), with weight functions $\left(N^{-1}h_{kl}^N\right)$, spiking rate $\left(\psi^k\right)$ and initial processes $\left(R^{Nk}\right)$.

Applying theorem 6.2.2 with weight functions (h_{kl}) and initial functions $\beta^k = r^k$, we obtain, for any $1 \leq k \leq \mathcal{K}$ and for all $i \in \mathbb{N}$, a solution (Z^{ki}, X^k, A^{ki}) to the equation

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

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

Theorem 6.3.1 (Propagation of Chaos).
Consider the framework described above and grant Assumptions 6.1. Then for all $t \geq 0$,

$$\mathbb{E}\left|d\left(Z_t^{Nki} - Z_t^{ki}\right)\right| \to 0, \text{ for } N \to \infty, \tag{6.4}$$

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

$$\left(Z^{Nk_1i_1}, \ldots, Z^{Nk_ni_n}\right)_{t\geq 0} \overset{wk}{\to} \left(Z^{k_1i_1}, \ldots, Z^{k_ni_n}\right)_{t\geq 0}$$

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

To prove this theorem we shall need the following lemma.

Lemma 6.3.2.
Let $(h_{kl})_{1\leq k,l\leq \mathcal{K}}, (\tilde{h}_{kl})_{1\leq k,l\leq \mathcal{K}}$ *be sets included in a family* $\mathcal{E}$ *of real-valued functions defined on* $\mathbb{R}_+$ *which is uniformly integrable on* $[0, T]$. *Define* $(Z, X, A), (\tilde{Z}, \tilde{X}, \tilde{A})$ *as the* $N-$*dimensional age dependent Hawkes process with weight functions* $(N^{-1}h_{kl})_{1\leq k,l\leq \mathcal{K}}$, $(N^{-1}\tilde{h}_{kl})_{1\leq k,l\leq \mathcal{K}}$, *rate functions* $\left(\psi^k\right)_{k\leq \mathcal{K}}$, *and with initial conditions* A_0, R^k. *There exists* $C > 0$ *depending on the family* $\mathcal{E}$, *on* $T, L, \mathcal{K}$ *and on* $\sup_{k\leq \mathcal{K}} \left\|\mathbb{E}R^k\right\|_T$ *(but not on* N*) such that*

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

for all $t \leq T$. ∘

The proofs of theorem 6.3.1 and lemma 6.3.2 may be found in the next section.

Remark 6.3.3.
The result shows that finitely many units will be asymptotically independent for $N \to \infty$. ∘

6.4 The Mean-Field Limit In the Case of a Hard Refractory Period

In this section we consider the mean-field limit of age dependent Hawkes processes with one single population ($\mathcal{K} = 1$) and a weight function given by an Erlang kernel as in Example 4.2.2, that is,

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

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

Assumption 6.2.

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

$\circ$

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

$$\phi_t = \int_0^t \mathbb{E}(\psi(x_s, A_s))ds = \int_0^t \bar{\lambda}_s ds, \tag{6.5}$$

where

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

denotes the expected number of jumps up to time t of a typical unit in the limit system. As in Section 4.5 above, we write $x^{(0)} := x$, and we add auxiliary variables $x^{(i)}, 1 \leq i \leq n$ to obtain the system

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

together with

$$\begin{aligned} dx_t^{(0)} &= -\nu x_t^{(0)}dt + x_t^{(1)}dt, \\ &\vdots \\ dx_t^{(n-1)} &= -\nu x_t^{(n-1)}dt + x_t^{(n)}dt, \\ dx_t^{(n)} &= -\nu x_t^{(n)}dt + bd\phi_t = -\nu x_t^{(n)}dt + b\bar{\lambda}_t dt. \end{aligned} \tag{6.6}$$

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

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

Due to Assumption 6.2, we have the following

Proposition 6.4.1.

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

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

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

where $A_0 \sim \mu_0(da)$. ○

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

Corollary 6.4.2.
For any starting law $\mu_0(da)$, $t \mapsto p_t$ is continuous on $(\delta,\infty]$, and thus, taking into account (6.6)*, $C^1((2\delta,\infty[,\mathbb{R})$, solving*

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

○

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

Corollary 6.4.3.
If $\mu_0(da) = \mu_0(a)da$, with $\mu_0 \in C(\mathbb{R},\mathbb{R}_+)$, then for all $t<\delta$,

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

is continuous and thus, taking into account (6.6)*, $C^1([0,\delta),\mathbb{R})$. In particular, on $[0,\delta)$, $t\mapsto p_t$ solves*

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

By induction, this implies that $t \mapsto p_t$ is continuous on $\mathbb{R}_+$ and C^1 on (δ,∞), with

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

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

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

$\circ$

Let us now look for possible stationary solutions of (6.6). At equilibrium, we necessarily have that

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

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

$$dA_{t+} = dt - A_t \int_{\mathbb{R}_+} 1_{\{z \leq \psi(x^{**}, A_t)\}} \pi(dt, dz). \tag{6.7}$$

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

Assumption 6.3.
For all x, there exists $r(x) \geq 0$ such that $\psi(x, a)$ is lower bounded for all $a \geq r(x)$. $\circ$

The stationary distribution of (6.7) is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_+$, having the density (see Proposition 21 of [17])

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

on $\mathbb{R}_+$, where κ is chosen such that $\int_0^\infty g_{x^{**}}(a) da = 1$. Recall that $\bar{\lambda}_t = \frac{d\phi_t}{dt}$ denotes the (expected) jump rate of the limit system at time t. Then at equilibrium, the total jump rate is constant and given by $\bar{\lambda}_t = \bar{\lambda}$. From (6.5) we get that

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

where we have used the change of variables

$$y = \int_0^a \psi(x^{**}, z) dz, dy = \psi(x^{**}, a) da.$$

As a consequence,

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

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

$$\frac{1}{\bar{\lambda}} = \int_0^\infty \exp\left(- \int_0^a \psi\left(\frac{b}{\nu^{n+1}} \bar{\lambda}, z \right) dz \right) da. \tag{6.8}$$

Here we have used that at equilibrium

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

which follows from (6.6).

Proposition 6.4.4.
Suppose that $x \mapsto \psi(x, a)$ is strictly increasing for any fixed $a \geq 0$ and that $b < 0$. There exists a unique solution λ^* to (6.8). ∘

Recall that we suppose that $\psi(x, a) = f(x) \mathbb{1}\{a \geq \delta\}$, for some $\delta > 0$. We calculate the right hand side of (6.8) and obtain the fixed point equation

$$\int_0^\infty \exp\left(- \int_0^a \psi\left(\frac{b}{\nu^{n+1}} \bar{\lambda}, z \right) dz \right) da = \delta + \frac{1}{f(\frac{b}{\nu^{n+1}} \bar{\lambda})} = \frac{1}{\bar{\lambda}}. \tag{6.9}$$

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

$$\frac{1}{\bar{\lambda}} = \delta + \frac{1}{f(\bar{\lambda} \int_0^\infty h(t) dt)} \tag{6.10}$$

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

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

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

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

In the excitatory case when $\int_0^\infty h(t)dt > 0$, to ensure that the fixed point equation (6.8) has a solution, suppose that f is strictly increasing and bounded from above and below, away from zero. Then the function

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

is a strictly decreasing function mapping $[0, \infty]$ onto $[\frac{1}{f(0)}, \frac{1}{f(\infty)}]$. Therefore, there is exactly one fixed point solution of (6.8), and $\delta \mapsto \bar{\lambda}(\delta)$ is again decreasing.

6.5 Proofs of Section 6 Results

The purpose of this section is to prove lemma 6.2.1, theorem 6.3.1 and lemma 6.3.2. To do so, we shall often make use of the convolution version of Gronwall's lemma, see lemma 7.2.1 in the appendix.

Proof of Lemma 6.2.1. It suffices to show that a unique solution exists on $[0, T]$, for arbitrary $T \geq 0$. In the following proof, $C := C_T$ will denote a dynamic constant depending on the parameters described in the lemma. It need not represent the same constant from line to line, nor from equation to equation.

First we prove existence of a solution to (6.2) with $\phi_t = \alpha_t$ using Picard-iteration. For $n \in \mathbb{N}$ define $(\phi^n, x^n, A^n) = \left(\phi^{n,k}, x^{n,k}, A^{n,k}\right)_{k \leq \mathcal{K}}$ as follows. Initialize the system for $n = 0$ by putting $\left(\phi^{0,k}, x^{0,k}, A^{0,k}\right) \equiv (0, 0, A_0)$. For general $n \in \mathbb{N}, n \geq 1$, the triple (ϕ^n, x^n, A^n) is defined as the solution to (6.2) with $\alpha = \phi^{n-1}$. Inductively it is seen that these processes are well-defined. Recall that $h = \sum_{k,l=1}^{\mathcal{K}} |h_{kl}|$. Using (4.2) we bound x^n by

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

It follows from lemma 7.2.1 that there exists a constant $C > 0$ which bounds all $\|x^n\|_T, n \in \mathbb{N}$. Using this upper bound on x^n, we also bound the difference of two consecutive solutions. Define

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

The Lipschitz property of ψ and the bound on x^n yield

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

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

$$\sum_{k=1}^{\mathcal{K}} P \left(\left\| A^{n+1,k} - A^{n,k} \right\|_t > 0 \right) \tag{6.11}$$

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

$$\leq \delta_t^{n+1}, \tag{6.12}$$

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

$$\delta_t^{n+1} \leq C \int_0^t \left| x_s^{n+1} - x_s^n \right| \, ds. \tag{6.13}$$

Moreover, lemma 22 of [12] gives

$$\begin{aligned} \int_0^t \left| x_s^{n+1} - x_s^n \right| ds &\leq \sum_{l=1}^{\mathcal{K}} \int_0^t \int_0^s h(s-u) \left| d \left(\phi_u^{n,l} - \phi_u^{n-1,l} \right) \right| ds \\ &\leq \int_0^t h(t-s) \delta_s^n ds. \end{aligned} \tag{6.14}$$

It therefore follows from lemma 7.2.1 that for all $1 \leq k \leq K$,

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

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

$$P\left(A^n_{s\le T} \neq A^{n+1}_{s\le T} \; i.o.\right) = P\left(\bigcap_{m\in\mathbb{N}} \bigcup_{n\ge m} \left\{\left\|A^n - A^{n+1}\right\|_T > 0\right\}\right)$$
$$\le \lim_{m\to\infty} \sum_{n\ge m}^{\infty} \delta_T^{n+1} = 0.$$

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

We need to show that the limit triple (ϕ, x, A) satisfies (6.2) with $\phi_t = \alpha_t$. Recall that $x \mapsto \psi^k(x,a)$ is continuous for fixed $a \in \mathbb{R}_+$. Since A^n reaches its limit in finitely many iterations, and ψ is continuous in x for fixed a, we obtain $\lim_{n\to\infty} \psi^k\left(x^{nk}_s, A^{nk}_s\right)$ exists for all $s \le T$ almost surely. By dominated convergence and (4.2), it follows that $\mathbb{E}\psi^k\left(x^{n,k}_s, A^{n,k}_s\right)$ converges as well. Therefore, once again by dominated convergence,

$$\phi^k_t = \lim_{n\to\infty} \phi^{n,k}_t = \lim_{n\to\infty} \int_0^t \mathbb{E}\psi^k\left(x^{n,k}_s, A^{n,k}_s\right) ds = \int_0^t \mathbb{E}\psi^k\left(x^k_s, A^k_s\right) ds,$$

that is, ϕ satisfies (6.2). One shows similarly that

$$\begin{aligned}
x^k_t &= \sum_{l=1}^{\mathcal{K}} \lim_{n\to\infty} \int_0^t h_{kl}(t-s)d\phi^{n,l}_s + \beta^k_t \\
&= \sum_{l=1}^{\mathcal{K}} \lim_{n\to\infty} \int_0^t h_{kl}(t-s)\mathbb{E}\psi^l(x^{nl}_s, A^{n,l}_s)ds + \beta^k_t \\
&= \sum_{l=1}^{\mathcal{K}} \int_0^t h_{kl}(t-s)\mathbb{E}\psi^l(x^l_s, A^l_s)ds + \beta^k_t \\
&= \sum_{l=1}^{\mathcal{K}} \int_0^t h_{kl}(t-s)d\phi^l_s + \beta^k_t,
\end{aligned}$$

and x satisfies (6.2) as well. For the age process, notice that the càglàd process

$$\varepsilon(t) = \sum_{k=1}^{\mathcal{K}} \int_0^{t-} \int_0^{\infty} \mathbb{1}\left\{z = \psi^k\left(x^k_s, A^k_s\right)\right\} \pi^k(ds, dz)$$

has a compensator which is equal to zero for all $t \geq 0$, almost surely, by lemma 2.1.6. Therefore, $\varepsilon_t = 0$ for all $t \geq 0$ almost surely. This implies that with probability 1, $\mathbb{1}\left\{z \leq \psi^k\left(x_s^{nk}, A_s^k\right)\right\}$ converges π^k–a.e. to $\mathbb{1}\left\{z \leq \psi^k\left(x_s^k, A_s^k\right)\right\}$ for all $k \leq \mathcal{K}$. As a consequence,

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

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

To prove uniqueness, we assume that $\left(\tilde{\phi}, \tilde{x}, \tilde{A}\right)$ also solves (6.2) with

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

Define

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

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

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

From Gronwall it follows that $\delta \equiv 0$, and therefore also that $x = \tilde{x}$ on $[0, T]$. From (6.2) it follows immediately $\phi = \tilde{\phi}$ and $A = \tilde{A}$ almost surely. □

Proof of Lemma 6.3.2. Throughout this proof, C is a dynamic constant with dependencies as declared in the theorem. Define the functions $h = \sum_{k,l} |h_{kl}|, \tilde{h} = \sum_{k,l} \left|\tilde{h}_{kl}\right|$. First we prove that the memory processes

$\mathbb{E}|X_t|, \mathbb{E}|\tilde{X}_t|$ are bounded on $[0,T]$ by a suitable constant C. Note that

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

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

$$\int_0^T \sum_{k,l=1}^{\mathcal{K}}\left|f_{kl}\right|(s)\,\mathbb{1}\left\{\sum_{k,l=1}^{\mathcal{K}}\left|f_{kl}\right|(s)>b\right\}ds < 2^{-1} \tag{6.15}$$

for all choices of $(f_{kl})\subset\mathcal{E}$. It follows from lemma 7.2.1 that $\mathbb{E}\|X\|_T\leq C$ for a suitable C. The same argument shows that also $\mathbb{E}\|\tilde{X}\|_T\leq C$. Define the total variation measure $\delta_t=\sum_{k=1}^{\mathcal{K}}\mathbb{E}\left|d\left(Z_t^{k1}-\tilde{Z}_t^{k1}\right)\right|$. We may write

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

As in the proof of lemma 6.2.1 we apply Markov's inequality to achieve

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

We insert this inequality into (6.16) to get

$$\delta_t\leq C\left(\int_0^t\mathbb{E}\left|\tilde{X}_s-X_s\right|ds+\int_0^t\delta_s ds\right). \tag{6.16}$$

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

$$
\begin{aligned}
&\left|\int_0^{s-} h_{kl}(s-u)\,dZ_u^{lj} - \int_0^{s-} \tilde{h}_{kl}(s-u)\,d\tilde{Z}_u^{lj}\right| \\
&\leq \int_0^{s-} \left|h_{kl} - \tilde{h}_{kl}\right|(s-u)\,dZ_u^{lj} + \int_0^{s-} \left|\tilde{h}_{kl}\right|(s-u)\left|d\left(Z_u^{lj} - \tilde{Z}_u^{lj}\right)\right| \\
&\leq \int_0^{s-} \gamma(s-u)\,d\sum_{l=1}^{\mathcal{K}} Z_u^{lj} + \int_0^{s-} \tilde{h}(s-u)\left|d\sum_{l=1}^{\mathcal{K}}\left(Z_u^{lj} - \tilde{Z}_u^{lj}\right)\right|. \qquad (6.17)
\end{aligned}
$$

We take expectation and apply Lemma 22 of [12] to obtain

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

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

$$
\begin{aligned}
&\sum_{k=1}^{\mathcal{K}}\int_0^t \mathbb{E}\left|\tilde{X}_s^k - X_s^k\right| ds \\
&\leq \sum_{k=1}^{\mathcal{K}}\int_0^t N^{-1}\sum_{l=1}^{\mathcal{K}}\sum_{j=1}^{N_l}\mathbb{E}\left|\int_0^{s-} h_{kl}(s-u)\,dZ_u^{lj} - \int_0^{s-} \tilde{h}_{kl}(s-u)\,d\tilde{Z}_u^{lj}\right| ds \\
&\leq C\left(\int_0^t \gamma(s)\,ds + \int_0^t \left|\tilde{h}(t-s)\right|\delta_s ds\right). \qquad (6.18)
\end{aligned}
$$

Inserting inequality (6.18) into (6.16), we obtain

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

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

□

Proof of Theorem 6.3.1. Let $(\tilde{Z}^N, \tilde{X}^N, \tilde{A}^N)$ be the $N-$dimensional age dependent Hawkes process induced by the same parameters as (Z^N, X^N, A^N), except the weight functions h_{kl} instead of h_{kl}^N.

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

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

The first term converges by lemma 6.3.2, and so it remains to prove convergence of δ_t^N. This part of the proof follows closely the proof given by Chevallier in [7], but we include it here for completeness. Let C be a dynamic constant depending on $p_k, L, T, \mathcal{K}, \|r\|_T$ and (h_{kl}). We use the symbol $\varepsilon(N)$ for any function depending on the same parameters as C, and N such that $\varepsilon(N) \overset{N\to\infty}{\rightarrow} 0$. Recall that $\|x\|_T$ is bounded by C sufficiently large by lemma 6.2.1.

As in the proof of lemma 6.2.1 we obtain

$$\delta_t^N \leq C\left(\int_0^t \mathbb{E}\left|\tilde{X}_s^N - x_s\right| ds + \int_0^t \delta_s^N ds\right). \tag{6.19}$$

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

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

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

$$\begin{aligned}
& \int_0^t \left(\mathbb{E} \left| \tilde{X}_s^{Nk} - x_s^k \right| - \mathbb{E} \left| R_s^{Nk} - r_s^k \right| \right) ds \leq \varepsilon(N) \\
& + \int_0^t N^{-1} \sum_{l=1}^{\mathcal{K}} \mathbb{E} \left| \sum_{j=1}^{N_l} \int_0^{s-} h_{kl}(s-u)\, d\left(\phi_u^l - \Lambda_u^{lj}\right) \right| ds \qquad (6.20) \\
& + \int_0^t N^{-1} \sum_{l=1}^{\mathcal{K}} \mathbb{E} \left| \sum_{j=1}^{N_l} \int_0^{s-} h_{kl}(s-u)\, d\left(\Lambda_u^{lj} - Z_u^{lj}\right) \right| ds \qquad (6.21) \\
& + \int_0^t N^{-1} \sum_{l=1}^{\mathcal{K}} \mathbb{E} \left| \sum_{j=1}^{N_l} \int_0^{s-} h_{kl}(s-u)\, d\left(Z_u^{lj} - \tilde{Z}_u^{Nlj}\right) \right| ds \qquad (6.22) \\
:= \quad & \varepsilon(N) + B_t^{1k} + B_t^{2k} + B_t^{3k}.
\end{aligned}$$

We now proceed to bound $B^i := \sum_{k=1}^{\mathcal{K}} B^{ik}$, $i \leq 3$. Define $h = \sum_{k,l=1}^{\mathcal{K}} |h_{kl}|$. Rewrite ϕ and Λ in terms of their densities, and thereby obtain a bound for the inner-most sum in (6.20) for $s \in [0,t]\,, l \leq \mathcal{K}$, which is given by

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

Notice that the sum consists of i.i.d. terms, so we may apply Cauchy-Schwarz to bound it by $\sqrt{N_l \operatorname{Var}(\psi(x_u^l, A_u^{lj}))}$, which is bounded for $u \in [0,T]$ by $\sqrt{N_l} C \left(1 + \left\| r^l \right\|_T\right)$ using (4.2). Insert this into (6.20) to see that

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

For B_t^2, recall that $(Z^{lj} - \Lambda^{lj})_j$ are i.i.d. for fixed l. By Cauchy-Schwarz, we obtain a bound for the inner-most sum of (6.21)

$$N_l^{1/2} \sqrt{\operatorname{Var} \int_0^s h_{kl}(s-u)\, d\left(Z_u^{l1} - \Lambda_u^{l1}\right)}\,. \qquad (6.23)$$

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

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

Then I is a martingale, and

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

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

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

for all $t \leq T$. For B^3 the triangle inequality, and lemma 22 [12] gives

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

We plug the bounds for B_1, B_2 and B_3 into (6.19) to obtain

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

Applying lemma 7.2.1

$$\delta_t^N \leq \varepsilon (N) + C \int_0^T \mathbb{E} \left| R_s^N - r_s \right| ds = \varepsilon (N) \tag{6.24}$$

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

$\square$

Chapter 7

Appendix

7.1 Point Process Results

The following result from [5] gives a sufficient criteria for two processes to couples.

Lemma 7.1.1 (Lemma 5 of [5]).
Assume that X, Y *are* $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$*-progressive and that for all* $s \geq 0$

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

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

7.2 Analysis results

We shall need the following version of Gronwall's lemma which has been proven in [12]. Recall that for any function $g : \mathbb{R}_+ \to \mathbb{R}$ and any $T > 0$, we have introduced $\|g\|_T = \sup_{t\leq T} |g(t)|$.

Lemma 7.2.1 (Lemma 23 of [12]).
Let $h : \mathbb{R}_+ \to \mathbb{R}_+$ *be locally integrable and* $g : \mathbb{R}_+ \to \mathbb{R}_+$ *be locally bounded. Let* $T \geq 0$.
1. Let u *be a locally bounded nonnegative function satisfying* $u_t \leq g_t + \int_0^t h\left(t-s\right) u_s ds$ *for all* $t \in [0, T]$*. If* $b > 0$ *satisfies that*

$$\int_0^T h\left(s\right) \mathbb{1}\left\{h\left(s\right) \geq b\right\} ds < \frac{1}{2}, \tag{7.1}$$

then $\|u\|_T \leq 2e^{2bT}\|g\|_T =: C_T \|g\|_T$.
2. Let (u^n) *be a sequence of locally bounded nonnegative functions such that* $u_t^{n+1} \leq g_t + \int_0^t h(t-s)\, u_s^n ds$ *for all* $t \in [0,T]$. *Then*

$$\sup_n \|u^n\|_T \leq C_T\left(\|g\|_T + \left\|u^0\right\|_T\right).$$

Moreover, if the inequality is satisfied with $g \equiv 0$, *then* $\sum_n u^n$ *converges uniformly on* $[0,T]$. ○

7.3 Splitting Two PRMs

Let $\pi, \overline{\pi}$ be two independent PRMs on $\mathbb{R}_+^2$. For any two functions $f_1 : \mathbb{R} \to [0,\infty)$, $f_2 : \mathbb{R} \to [0,\infty]$, such that $f_1 \leq f_2$ we define for $B \in \mathcal{B}_{\mathbb{R}_+^2}$

$$\pi^{\downarrow f_1,f_2}(B) = \int_B \mathbb{1}\left\{z \notin [f_1, f_2)\right\}\, d\pi(s,z) + \int_B \mathbb{1}\left\{z \in [f_1, f_2)\right\}\, d\overline{\pi}(s,z). \tag{7.2}$$

$$\pi^{\uparrow f_1,f_2}(B) = \int \mathbb{1}\left\{(s,z) : (s, z - f_1(s)) \in B,\ z \leq f_2(s)\right\}\, d\pi(s,z) \tag{7.3}$$

$$+ \int \mathbb{1}\left\{(s,z) : (s, z - f_1(s)) \in B,\ z > f_2(s)\right\}\, d\overline{\pi}(s,z). \tag{7.4}$$

It may be shown directly that both of the above set functions are PRMs with $\pi^{\downarrow f_1,f_2} \perp\!\!\!\perp \pi^{\uparrow f_1,f_2}$.

Theorem 7.3.1 shows that these independence properties generalize to when f_1, f_2 are predictable, but random, intensities.

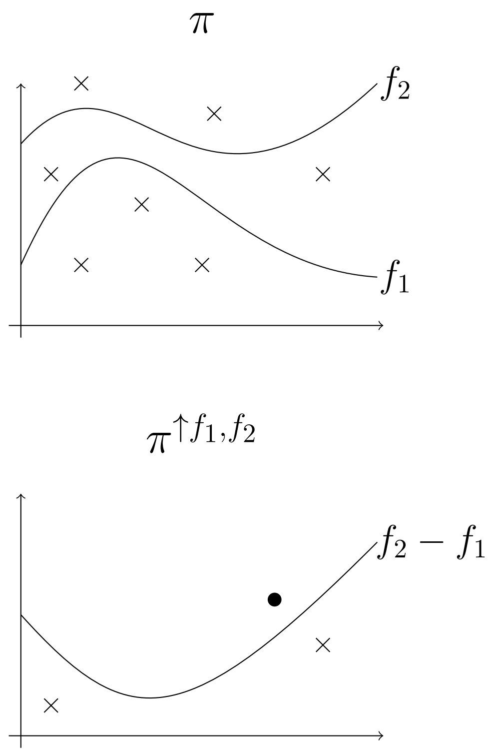

Figure 7.1: A figure illustrating how $\pi^{\uparrow f_1,f_2}$ and $\pi^{\downarrow f_1,f_2}$ are created from $\pi, \overline{\pi}$. Notice that $\pi^{\downarrow f_1,f_2}$ contains the part of π *below* f_1 while $\pi^{\uparrow f_1,f_2}$ contains an area part immediately *above* it.

Theorem 7.3.1.
Let $(\mathcal{F}_t)_{t\in[0,\infty)}$ be a filtration and $\pi, \overline{\pi}$ be two independent $\mathcal{F}_t$-PRMs on $\mathbb{R}_+$. For
$t \in [0,\infty)$, let $\lambda_t \leq \lambda'_t$ be $\mathcal{F}_t$-predictable processes taking values in $[0,\infty), [0,\infty]$, respectively. Define $\mathcal{F}^*_t = \sigma\left(\mathcal{F}_t, \pi^{\uparrow\lambda,\lambda'}\right)$. It holds that $\pi^{\uparrow\lambda,\lambda'}, \pi^{\downarrow\lambda,\lambda'}$ are PRMs such that $\pi^{\uparrow\lambda,\lambda'} \perp\!\!\!\perp \pi^{\downarrow\lambda,\lambda'} \perp\!\!\!\perp \mathcal{F}_0$. $\circ$

Proof.
The proof will be done in several steps.

Step 1.
Let $(t_i)_{i\in\mathbb{N}_0}$ be a fixed partition where $0 < t_{i-1} < t_i$. $t_i \overset{i\to\infty}{\to} \infty$. Assume Y_i^1, Y_i^2 is $\mathcal{F}_{t_{i-1}}$−measurable, taking values in a finite state space $\mathcal{Y} \subset \mathbb{R}$, and assume that

$$\lambda_t = \sum_{i=1}^{\infty} Y_i^1 1_{]t_{i-1},t_i]}(t). \quad \lambda'_t = \sum_{i=1}^{\infty} Y_i^2 1_{]t_{i-1},t_i]}(t). \tag{7.5}$$

Fix $m \in \mathbb{N}$ and take $k_j^i, l_j^i \in \mathbb{N}$ and mutually disjoint, bounded $B_j^i, C_j^i \in \mathcal{B}_{(t_{i-1},t_i \times \mathbb{R}_+]}$ for $i = 1, \ldots, n \;\; j = 1, \ldots, m$. Define

$$F_i = \bigcap_{j=1}^{m} \left(\pi^{\downarrow \lambda, \lambda'} \left(B_j^i \right) = k_j^i \right), \tag{7.6}$$

$$G_i = \bigcap_{j=1}^{m} \left(\pi^{\uparrow \lambda, \lambda'} \left(C_j^i \right) = l_j^i \right). \tag{7.7}$$

Take $E \in \mathcal{F}_0$. and set $E_i = F_i \cap G_i, E = \bigcap_{i=0}^{n} E_i$. It is sufficient to show that the projection E has the correct distribution, i.e.:

$$P(E) = P(E_0) \prod_{i=1}^{n} \prod_{j=1}^{m} \mathcal{P}\left(\int_{B_j^i} ds, k_j^i \right) \mathcal{P}\left(\int_{C_j^i} ds, l_j^i \right) \tag{7.8}$$

where $\mathcal{P}(c, \cdot)$ is the Poisson density with mean c. This will be proved using induction. The induction claim over $N = 0, ..., n$ is that

$$P(E_0) = \mathbb{E}\mathbb{1}\left\{ E_0 \cap E_1 \cap \cdots \cap E_{n-N} \right\} \prod_{i=n-N+1}^{n} \prod_{j=1}^{m} \mathcal{P}\left(\int_{B_j^i} ds, k_j^i \right) \mathcal{P}\left(\int_{C_j^i} ds, l_j^i \right). \tag{7.9}$$

The induction start $N = 0$ is clear (where the empty product is 1 per convention). Assume that the claim holds for some $N - 1$. Since $E_N \perp\!\!\!\perp \mathcal{F}_{t_{N-1}} | \lambda_{t_N}, \lambda'_{t_N}$, we may write $P\left(E_N | \mathcal{F}_{t_{N-1}}\right) = P\left(E_N | \lambda_{t_N}, \lambda'_{t_N}\right)$ and for $c \leq d \in \mathcal{Y}$

$$P\left(E_N | \left(\lambda_{t_N}, \lambda'_{t_N}\right) = (c, d)\right) \tag{7.10}$$

$$= P \bigcap_{j=1}^{m} \left(\int_{B_j^N} \mathbb{1}\left\{ z \notin [c, d) \right\} \, d\pi(s, z) + \int_{B_j^N} \mathbb{1}\left\{ z \in [c, d) \right\} \, d\overline{\pi}(s, z) = k_j^i \right) \tag{7.11}$$

$$\cdot P \bigcap_{j=1}^{m} \left(\int_{c + C_j^N} \mathbb{1}\left\{ z \in [c, d) \right\} \, d\pi(s, z) + \int_{c + C_j^N} \mathbb{1}\left\{ z \notin [c, d) \right\} \, d\overline{\pi}(s, z) = l_j^i \right). \tag{7.12}$$

Straightforward calculations show that indeed

$$P\left(E_N \middle| \left(\lambda_{t_N}, \lambda'_{t_n}\right) = (c,d)\right) = \prod_{j=1}^{m} \mathcal{P}\left(\int_{B_j^N} ds, k_j^N\right) \mathcal{P}\left(\int_{C_j^N} ds, l_j^N\right). \tag{7.13}$$

Notice that the right side does not depend on c, d implying that $E_N \perp\!\!\!\perp \mathcal{F}_{t_{N-1}}$. Thus, by conditioning (7.9) w.r.t. $\mathcal{F}_{t_{N-1}}$ and inserting this result, the induction step follows and the proof is completed.

Step 2.
Assume now that λ, λ' is bounded and continuous in t for all discrete measures. We use some dyadic approximation by putting

$$\lfloor x \rfloor_n := \sup_k \{k2^{-n} : k2^{-n} < x\}$$

$$\lambda_t^n = \lfloor \lambda_{\lfloor t \rfloor_n} \rfloor_n, \quad \lambda_t'^n = \lfloor \lambda'_{\lfloor t \rfloor_n} \rfloor_n.$$

Then as $n \to \infty$,

$$\lambda_t^n \to \lambda_t, \quad \lambda_t'^n \to \lambda'_t \tag{7.14}$$

for all $t \in [0, T]$. Almost surely, the graphs of λ, λ' are $\pi, \overline{\pi}$ null-sets. It follows that almost surely

$$\forall A \in \mathcal{B}^2, Leb(A) < \infty : \pi^{\uparrow \lambda^n, \lambda'^n}(A) \to \pi^{\uparrow \lambda, \lambda'}(A), \quad \pi^{\downarrow \lambda^n, \lambda'^n}(A) \to \pi^{\downarrow \lambda, \lambda'}(A). \tag{7.15}$$

It is now straightforward to prove the claim by applying step 1 for each n.

Step 3.
Assume the same set up as before, except for continuity in t. Define for $n \in \mathbb{N}$

$$\lambda_t^n = n \int_{t-\frac{1}{n}}^{t} \lambda_s ds, \quad \lambda_t'^n = n \int_{t-\frac{1}{n}}^{t} \lambda'_s ds. \tag{7.16}$$

Note that the processes

$$t \mapsto \int_0^t \lambda_s ds, \quad t \mapsto \int_0^t \lambda'_s ds \tag{7.17}$$

is Lipschitz continuous since λ, λ' is bounded. By Rademacher's theorem there is a Lebesque-full set on which the above map is differentiable. It follows that almost surely, λ^n, λ'^n converges almost everywhere in t to λ, λ'. The remaining part of step 2 is similar to step 3.

Step 4.
Assume now that λ, λ' are given as in the assumptions. One may define $\lambda_n = \lambda \wedge n$, $\lambda'_n = \lambda' \wedge n$, and repeat the procedure from the previous steps to complete the proof, which we leave to the reader. □

7.4 The Random Exchange Process

The purpose of this section is to study the Markov Chain given by

$$M_i = (M_{i-1} - 1) \vee X_i \tag{7.18}$$

where M_0, X_i are non-negative and mutually independent variables such that (X_i) are i.i.d. This process is going under the name Random Exchange Process with constant decrements. RE-processes have been treated in [25] where it was shown that M is positive recurrent when X has finite expectation. See also [43] for a null-recurrence characterization. We are interested in moments of the return time $\sigma = \inf \{n > 0 : M_n \leq 0\}$, and moments of the invariant distribution μ. To the best knowledge of this author, there has been no published result about such.

Let F, S be the distribution function and survival function of X_1. Let $q \geq 0$ be a real number. Clearly the transition kernel of M_i is given by

$$\begin{aligned} P_x\left((a,b]\right) &= P\left(X \in (a,b]\right) \quad \forall b > a > x-1 \\ P_x\left(\{x-1\}\right) &= F\left(x-1\right). \end{aligned}$$

Let $\phi : [0, \infty) \to [0, \infty)$ denote an increasing and differentiable function. Valid choices of ϕ include $\phi(x) = x^{q+1}$ and $\phi(x) = \exp(cx)$.

Theorem 7.4.1.
Assume that ϕ is convex and $\mathbb{E}\phi(X) < \infty$. Then M_i is positive recurrent with stationary distribution μ and $\int \phi'(y-1)\, d\mu(y) < \infty$. ◦

Proof.
We use a Lyapunov argument.

$$\begin{aligned} & \mathbb{E}_x \phi(M_1) - \phi(x) && (7.19) \\ &= \int_{(x-1,\infty)} \phi(y)\, dF(y) + \phi(x-1) F(x-1) - \phi(x) \\ &= \int_{(x-1,\infty)} \phi(y)\, dF(y) + (\phi(x-1) - \phi(x)) F(x-1) - S(x-1)\phi(x). && (7.20) \end{aligned}$$

Notice that the first term above converges to 0 for $x \to \infty$. The second term can be controlled using the mean value theorem. Indeed, for x so large that $F(x-1) \geq 2^{-1}$ we have

$$(\phi(x-1) - \phi(x)) F(x-1) \leq -\frac{1}{2}\phi'(x-1) \tag{7.21}$$

We may apply proposition 14.1.1 and theroem 14.2.3 from [34] with $f(m) = \frac{1}{2}\phi'(x-1)$ to conclude the desired result. □

Note that μ must satisfy $\mu[0,x] = F(x)\mu[0,x+1]$, and in turn it satisfies

$$\mu([0,x]) = \prod_{k=0}^{\infty} F(x+k) \tag{7.22}$$

whenever it exists. In particular when X has support on $\mathbb{N}_0$ we have

$$\mu([0,n]) = \prod_{k=n}^{\infty} F(k). \tag{7.23}$$

We now discuss the hitting time σ. We need an intermediate result that gives a peculiar relation between the return time to 0, and the hitting time given general distributions for a RE-process, when the update variables have support on $\mathbb{N}_0$.

Theorem 7.4.2.
Assume that X has support on $\mathbb{N}_0$. Define for $i, j \in \mathbb{N}_0$ $e_{i,j} = \mathbb{E}_i \phi(\sigma + j)$. Assume that $e_{i,j} < \infty$ for all i, j. For any probability measure ν on $\mathbb{N}_0$ it holds that

$$\mathbb{E}_\nu \phi(\sigma) = \phi(0) + \nu(0)\left(\mathbb{E}_0 \phi(\sigma) - \phi(0)\right) + \sum_{i=0}^{\infty} \left(\mathbb{E}_1 \phi(\sigma + i) - \phi(i)\right) S_\nu(i) \frac{\mu(0)}{\mu[0, i]} \tag{7.24}$$

where S_ν is the survival function of ν. ○

Proof.
note that

$$\mathbb{E}_\nu \phi(\sigma) = \sum_{i=1}^{\infty} \nu(i)\, e_{i0}. \tag{7.25}$$

We claim that

$$e_{i,j} - e_{i-1,j} = F(i-2)\left(e_{i-1,j+1} - e_{i-2,j+1}\right), \quad i \geq 2. \tag{7.26}$$

This follows from coupling two Markov chains M_k^i, M_k^{i-1} started at $i, i-1$ respectively, and sharing the same i.i.d. update sequence (X_k). The two processes are equal for all $k \geq 1$ if $X_1 \geq i-1$ while $M_1^i = i-1, M_1^{i-1} = i-2$ if $X_1 \leq i-2$, implying eq. (7.26). For all $i \geq 1$ we obtain

$$\begin{aligned} e_{ij} &= e_{i-1,j} + \left(e_{1,j+i-1} - \phi(j + l - 1)\right) \prod_{k=0}^{i-2} F(k) && (7.27) \\ &= \phi(j) + \sum_{l=1}^{i} \left(e_{1,j+l-1} - \phi(j+l-1)\right) \prod_{k=0}^{l-2} F(k) && (7.28) \\ &= \phi(j) + \sum_{l=1}^{i} \left(e_{1,j+l-1} - \phi(j+l-1)\right) \frac{\mu(0)}{\mu[0, l-1]} && (7.29) \end{aligned}$$

with convention $\prod_{i=0}^{-1} = 1$. Inserting this into (7.25) gives

$$\mathbb{E}_\nu \phi(\sigma) = S_\nu(0)\,\phi(0) + \nu(0)\,\mathbb{E}_0 \phi(\sigma) + \sum_{i=1}^{\infty} \nu(i) \sum_{l=1}^{i} \left(\mathbb{E}_1 \phi(\sigma + l - 1) - \phi(l-1)\right) \frac{\mu(0)}{\mu[0, l-1]}. \tag{7.30}$$

The result follows from adding $\pm\nu(0)\,\phi(0)$ and interchanging the two sums. □

Corollary 7.4.3.

- *If* $\mathbb{E}X^{q+1} < \infty$ *then an invariant distribution* μ *of* M *exists and* $\int x^q d\mu(x) < \infty$. *Also, if there is* $c_X > 0$ *such that* $\exp(c_X X) < \infty$ *then* $\int \exp(cx)\, d\mu(x)$ *for all* $c < c_X$ *and* M *is geometrically ergodic.*

- *Assume still that* $\mathbb{E}X^{q+1} < \infty$. *It holds that* $\mathbb{E}_\nu \sigma^{q^* \wedge (q+1)} < \infty$ *for all* $q^* \geq 0$ *and all initial measures* ν *with* q^**'th moment. Moreover, if* ν *and* X *have exponential moment then so has* σ.

○

Proof.
The first point follows directly from theorem 7.4.1.

To prove the second point notice that we can without loss of generalization assume X, M_0 has support on $\mathbb{N}_0$ by replacing X and M_0 with $\lceil X \rceil$ and $\lceil M_0 \rceil$.

We start with the power-moment case. For $z \in \mathbb{R}$ write $e^z_{i,j}$ for the variable $e_{i,j}$ with $\phi(x) = x^z$. Write $q = r + n$, $r \in [0,1)$. We show by induction over $m = 0, \ldots, n+1$ that $\mathbb{E}_\nu \sigma^{q_m \wedge q^*} < \infty$ with $q_m = r + m$, and for all $q^* \geq 0$ and measures ν with q^*'th moment.

The induction start $n = 0$; if $r = 0$ or $q^* = 0$ the claim is trivial. Otherwise, note that $e^r_{0,0} \leq e^1_{0,0} < \infty$ by Kac's theorem. We can apply theorem 7.4.2 and the mean value theorem to obtain

$$\mathbb{E}_\nu \sigma^{q^* \wedge r} \leq C + C \sum_{i=0}^{\infty} i^{q^* \wedge r - 1} S_\nu(i) < \infty \tag{7.31}$$

where $C > 0$ is sufficiently large.
Assume now that the induction claim holds for some $m \leq n$. Since $q_m \leq q$, and π have q'th moment, we can use the induction assumption to see that $\mathbb{E}_\pi \sigma^{q_m} < \infty$. It is well known that $P_\pi(\sigma = i) = P_0(\sigma \geq i)$ for $i \geq 0$ (see section 10.3.1 [34]). It follows that $e^{q_m+1}_{0,0} < \infty$ and hence also $e^{q_m+1}_{i,j} < \infty$. We may now apply theorem 7.4.2 and the mean value theorem to obtain

$$\mathbb{E}_\nu \sigma^{(q_m+1) \wedge q^*} \leq C + C \sum_{i=0}^{\infty} i^{(q_m+1) \wedge q^* - 1} S_\nu(i) < \infty \tag{7.32}$$

where $C > 0$ is sufficiently large.

Consider now the case where $\int \exp(c_\nu y)\, d\nu(y) < \infty$. For $c < c_\nu$ we consider the function $\phi(x) = \exp(cx)$. Combining theorem 7.4.1 above and theorem 15.0.1 ii) in [34] we get that $\mathbb{E}_0 \phi(\inf\{j > 0 : M_j \leq K\}) < \infty$ for some $K > 0$ large and small $c > 0$. It follows that $e_{0,0} < \infty$ for a possibly smaller c and hence also $e_{i,j} < \infty$ for $i, j \in \mathbb{N}_0$. By dominated convergence, we can choose C large so that

$$\forall i \in \mathbb{N}_0 : \mathbb{E}_1 \sigma \phi'(\sigma + i) \leq C \exp(c' i) \tag{7.33}$$

for all $c' > c$. Choose now $c' \in (c, c_\nu)$, and combine theorem 7.4.2 with the above inequality to obtain

$$\mathbb{E}_\nu \exp(c\sigma) \leq C + C \sum_{i=0}^{\infty} S_\nu(i) \exp(c' i), \tag{7.34}$$

again for a possibly larger C. An application of Markovs inequality gives the desired result. $\square$

Bibliography

[1] Athreya, K. B., Lahiri, S. N. Measure Theory and Probability Theory. *Springer* (2006).

[2] Bacry, E., Muzy, J.F. Hawkes model for price and trades high-frequency dynamics. *Quantitative finance*, 14 (2014), 1–20.

[3] Belzounce, F., Riquelme, C., Mulero, J. An Introduction to Stochastic Orders 1st Edition *Academic Press* (2015)

[4] Billingsley, P. Convergence of Probability Measures, 2nd Edition *Wiley*, (1999).

[5] Brémaud, P., Massoulié, L. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3) (1996) 1563-1588.

[6] Brémaud, P., Nappo, G., Torrisi, L. Rate of Convergence to Equilibrium of Marked Hawkes Processes *Journal of Applied Probability*, Vol. 39, No. 1 (2002), 123-136.

[7] Chevallier, J. Mean-field limit of generalized Hawkes processes. *Stoc. Proc. and their Appl.* 127 (2017), 3870-3912.

[8] Chevallier, J., Caceres, MJ., Doumic, M., Reynaud-Bouret, P. Microscopic approach of a time elapsed neural model. *Math. Mod. & Meth. Appl. Sci.*, 25(14) (2015) 2669–2719.

[9] Clinet, S., and Yoshida, N. Statistical inference for ergodic point processes and application to Limit Order Book. *Stoch. Proc. Appl*, 127 (2017), 1800-1839.

[10] Costa, M, Graham, C., Marsalle, L, Tran, V.C. Renewal in Hawkes processes with self-excitation and inhibition. arxiv.org/abs/1801.04645, (2018).

[11] Daley, D., Vere-Jones, D. An Introduction to the Theory of Point Processes. *Springer*, (1988).

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

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

[14] Ditlevsen, S., Löcherbach, E., Raad, M. B. Stability and Mean-Field Limits of Age Dependent Hawkes Processes. arxiv.org/abs/1806.06370, (2019).

[15] Duarte, A., Löcherbach, E., Ost, G. Stability, convergence to equilibrium and simulation of non-linear Hawkes Processes with memory kernels given by the sum of Erlang kernels. Submitted, arxiv.org/abs/1610.03300, (2017).

[16] Dwass, M. The Total Progeny in a Branching Process and a Related Random Walk *Journal of Applied Probability* Vol. 6 No. 3 (1969), 682-686.

[17] Fournier, N., Löcherbach, E. A toy model of interacting neurons. *Annales de l'I.H.P.* 52 (2016), 1844–1876.

[18] Gerhard, F., Deger, M. and Truccolo, W. On the stability and dynamics of stochastic spiking neuron models: Nonlinear Hawkes process and point process GLMs *PLOS Computational Biology*, 13(2) (2017): e1005390. doi:10.1371/journal.pcbi.1005390.

[19] Graham, C. Regenerative properties of the linear Hawkes process with unbounded memory. arXiv preprint arXiv:1905.11053, (2019).

[20] Graham, C., Robert, P. Interacting multi-class transmissions in large stochastic systems. *Ann. Appl. Prob. 19*, 6 (2009) 2334–2361.

[21] Gut, A. Stopped Random Walks *Springer-Verlag New York*, (2009).

[22] Halmos, P.R. Measure theory. *Springer Berlin New York*, (1974).

[23] Hawkes, A. G. Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, 58 (1971) 83-90.

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

[25] Helland, I., Nilsen, T. On a General Random Exchange Model. *Journal of Applied Probability* 13(4) (1976) 781-790.

[26] Sokol, A. An Introduction to Stochastic Integration with Respect to General Semimartingales. Lecture notes. `http://alexandersokol.dk/`

[27] Höpfner, R., Löcherbach, E. Statistical models for Birth and Death on a Flow: Local absolute continuity and likelihood ratio processes. *Scandinavian Journal of Statistics*, 26(1):107–128, 1999.

[28] Ikeda, N., Nagasawa,M., Watanabe, S. A construction of Markov processes by piecing out. *Proc. Japan Acad.*, 42(4):370–375, 1966.

[29] Jacod, J., Shiryaev, A.N. Limit theorems for stochastic processes. Second edition, *Springer-Verlag, Berlin*, (2003).

[30] Löcherbach, E. Loukianova, D., On Nummelin splitting for continuous time Harris recurrent Markov processes and application to kernel estimation for multi-dimensional diffusions. *Stoch. Proc. Appl.*, 118:1301–1321, 2008.

[31] Meyn, S. P., Tweedie, R. L. Stability of Markovian processes I : Criteria for discrete-time chains. *Adv. Appl. Probab.*, 24:542–574, 1992.

[32] Meyn, S. P., Tweedie, R. L. Stability of Markovian processes II : Stability of Markovian Processes II: Continuous-Time Processes and Sampled Chains. *Adv. Appl. Probab.*, 25:487–513, 1993.

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

[34] Meyn, S. P., Tweedie, R. L. Markov chains and stochastic stability. *Springer-Verlag, London*, (1993). Available at: probability.ca/MT

[35] Miyahara, Y. Invariant measures of ultimately bounded stochastic processes. *Nagoya Math. J.*, 49 (1973), 149–153.

[36] Nummelin, E. A splitting technique for Harris recurrent Markov chains. *Z. Wahrscheinlichkeitstheorie Verw. Geb.* 43 (1978), 309–318.

[37] Pakdaman, K., Perthame, B., Salort, D. Relaxation and self-sustained oscillations in the time elapsed neuron network model. arxiv.org/abs/1109.3014, (2011).

[38] Raad, M. B. Renewal Time Points for Hawkes Processes. arXiv preprint arXiv:1906.02036, (2019)

[39] Reynaud-Bouret, P., Schbath, S. Adaptive estimation for Hawkes processes; application to genome analysis. *The Annals of Statistics*, 5 (2010), 2781–2822.

[40] Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. Inference of functional connectivity in Neurosciences via Hawkes processes. 1st IEEE Global Conference on Signal and Information Processing, Austin, Texas (2013)

[41] Reynaud-Bouret, P., Roy, E. Some non asymptotic tail estimates for Hawkes processes. *Bulletin of the Belgian Mathematical Society-Simon Stevin*, 13(5), (2007) 883-896.

[42] Zadeh, A. H., Sharda, R. Hawkes Point Processes for Social Media Analytics. In *"Reshaping Society through Analytics, Collaboration, and Decision Support: Role of Business Intelligence and Social Media"*, *Springer International Publishing*, 2015.

[43] Zerner, M. P. Recurrence and transience of contractive autoregressive processes and related Markov chains. *Electronic Journal of Probability*, 23 (2018).

[44] Zhu, L. Central limit theorem for nonlinear Hawkes processes. *Journal of Applied Probability*, 50(3) (2013), 760–771.