

# Hawkes processes: some key ideas, links to neuroscience and system identification

LÁSZLÓ GERENCSÉR, GYÖRGY PERCZEL, LORÁND ERŐSS,  
DÁNIEL FABÓ, ZSUZSANNA VÁGÓ

The primary objective of the present paper is to provide a brief survey of basic ideas and some key results of the theory of Hawkes processes that may serve as an interface for multidisciplinary research between mathematics and neuro-sciences, with hints to epidemiology. The class of Hawkes processes is considered as an appropriate model to describe the pattern of neuronal firing, see [56] [38]. As such, it is a prime candidate for detecting changes in brain activity, and consequently to predict epileptic seizures. Most of the advanced results will be presented via simple examples. In addition, we elaborate on links between the theory of Hawkes processes and linear stochastic systems, in particular system identification. A few new results in this context will be also presented with the outline of their proofs.

## 1. Recollection and salute

It is our honor to have the chance to contribute to this unique special issue dedicated to Tyrone Duncan on the occasion of his 80th birthday. As the first author I would like to take the chance in recalling my first meeting with Tyrone, which has never faded. It was 1981, a few months away from the introduction of martial law (*stan wojenny*) in Poland, when Bozenna visited me in SZTAKI in Budapest, Hungary. Tyrone was an accompanying person. At that time I was not aware of the fact that he was D of the DMZ equation, the fundamental result of the theory non-linear filtering, established in his PhD thesis [11], briefly summarized in [12]. Asking him on stochastic calculus on differentiable manifolds he said: it is very simple, you just roll the sphere along a Brownian path in the plane. Many year has passed since then, and

the topics of our discussion have changed to stochastic adaptive control [13], or stochastic systems driven by fractional Brownian motion, [14], the tune is unchanged. All are very simple. On top of all, his keen interest in applications and his impressive technical skills are paired with an exceptionally friendly, honest and helpful attitude towards his peers, which I enjoyed for many years. I would like to express my gratefulness for all, God bless you, Tyrone.

## 2. Introduction and motivation

According to the estimation of the World Health Organization (WHO) epilepsy is one of the most common neurological disease, with approximately 50 million patients world-wide. Its most characteristic features are seemingly spontaneous seizures 'due to abnormal excessive or synchronous neuronal activity in the brain' [17].

As no single, cellular cause of seizure generation has been identified, a network-level understanding of epileptic synchronization is in the focus of current research [47]. This pathological pattern may present itself in various forms with various severity from myoclonic jerks of a distinct group of muscles (focal seizures) through transient absence of consciousness to generalized, tonic-clonic seizures.

Epilepsy can be treated using anti-epileptic medications in the majority of cases. However, 30% of patients have epilepsy that is refractory to currently available medications. These patients may be candidates for surgery, that most often mean the resection of the epileptogenic focus.

Surgery must be preceded by a thorough exploration of the patient, with the primary objective of precisely localizing the focus of seizure-generation (seizure-onset zone, SOZ) and its relation to other brain areas. Various approaches have been proposed to explore the functional connectivity map of the epileptic brain, such as Granger causality-analysis, [18], [25].

Despite all efforts, some patients still live with frequent and disabling seizures. Unexpectedness of epileptic seizures is a key element with respect both to patient's quality of life and also current therapeutic approaches [52]. There is thus need for the prediction of seizures. The benefit of a reliable seizure prediction system would be twofold. On the one hand, a seizure-warning system could help patients to stay safe for the time of an

appending seizure and/or take an instantaneously acting medication. On the other hand, reliable predictions would provide a chance for developing closed-loop systems delivering electrical or pharmacological modulation to the brain in order to prevent the full evolution of the seizure. As such, numerous attempts have been made to accurately predict seizures, though only few methods performed beyond ‘above-chance’. Most of them are based on electroencephalography (EEG) acquired either from the scalp or using invasive electrodes placed on the surface or within the brain. For a thorough review on the proposed technologies and algorithms we refer to [19].

From the conceptual point of view the question arises if an altered brain-state just before seizures exists at all, and if it can be differentiated from the regular state of the brain. An affirmative answer to this was given among others in [10].

As for the available technology, for decades scalp electroencephalography (EEG) has been the gold standard in clinical practice for observing the neuronal activity of the brain, [54]. One of the academia and industry collaborations that did achieve ‘above-chance’ performance in predicting seizures was the Seizure Advisory System (NeuroVista, WA, USA), a long-term EEG recording and monitoring device, that used the classification of spectral features. The authors proved that seizures are not completely unanticipated, and studied the identification of seizure-prone brain-states [8].

The advent of implantable micro-electrodes even recording spikes of distinct action potentials (AP) generated by single neurons gave rise to the use of neuronal firing patterns as a substrate for seizure prediction. In [57] authors found that neurons changed their firing pattern before each seizure in a similar way, noting a phenomenon with high predictive value. Representing a sequence of AP-s of a neuron solely via its exact timing is an often-used procedure for data-compression based on the basic observation that AP-s are similar both in shape and amplitude, thus it is commonly accepted that their exact timing conveys the information to the afferent neurons. The availability of implantable micro-electrodes inspired the development of novel mathematical techniques, such as the application and adaptation of the theory of multivariate self-exciting point processes or Hawkes processes to model spike train dynamics, see [7], [48], [56], [38].

### 3. General theory of point processes

Point processes are widely used in various applications of mathematics. The most familiar example is the Poisson process, an excellent brief introduction to which, on general state-spaces, is given by [37]. In this paper we consider another important class of point processes, which are intimately related to the Poisson process on the real line, at least historically. These are the so-called Hawkes processes, the classical linear version of which can be defined as a Poisson cluster process, see [31], to be described below. But first we introduce a few general concepts.

A point process is defined by a sequence of events that are represented by a sequence of extended real-valued random variables  $(T_n)$  indexed by  $-\infty < n < +\infty$  such that  $T_0 \leq 0 < T_1$ , and  $T_n < T_{n+1}$  unless  $T_n = +\infty$  or  $T_{n+1} = -\infty$ . We assume that the sequence  $(T_n)$  has no finite limit points. The strict inequalities  $T_n < T_{n+1}$  imply that the point process is simple, i.e. for any fixed  $t$  the number of points in  $t$  is either 0 or 1, a.s.

Let,  $N((a, b])$ , for any  $a < b$ , denote the number of points in the interval  $(a, b]$ . Then  $N$  can be extended to a measure on  $\mathcal{B}(\mathbb{R})$ , the  $\sigma$ -algebra of Borel-measurable sets, and we get the counting measure denoted by  $N(\cdot)$ . We use the notation  $N_t$  to denote the number of points in the interval  $(0, t]$ . The point process  $(T_n)$  is perfectly described by its counting measure.

Let  $\mathcal{F}_t = \mathcal{F}_t^N$  denote the  $\sigma$ -algebra of events determined by the process preceding the time  $t$ , i.e. let  $\mathcal{F}_t^N = \sigma\{N(C), C \in \mathcal{B}(\mathbb{R}), C \subset (-\infty, t]\}$ , and let us call it the *internal history* at time  $t$ . The *future* of  $N(\cdot)$  can be defined in similar manner:  $\mathcal{F}_t^+ = \sigma\{N(C), C \in \mathcal{B}(\mathbb{R}), C \subset (t, +\infty)\}$ .

The finite dimensional distributions of a point process are given by the joint probabilities  $P(N(A_1) = r_1, \dots, N(A_k) = r_k)$  for any positive integer  $k$  and any collection of Borel sets  $A_i$  and non-negative integers  $r_i$ . A point process  $N(\cdot)$  is *stationary* if its finite dimensional distributions are invariant under translation of the time axis. Let the (backward) translation operator  $S_t$  be defined on the Borel sets of  $\mathbb{R}$  as  $S_t A := \{x : x + t \in A\}$ .

A key concept is the *intensity* of a point process. The most direct, although not quite precise way to define the intensity function of  $N(\cdot)$ , following [45], is this:

$$(1) \quad \lambda(t) = \lim_{\delta \downarrow 0} \frac{1}{\delta} P(N[t, t + \delta) > 0 \mid \mathcal{F}_t^N).$$

Choosing a left closed interval  $[t, t + \delta)$  is not accidental: it yields a predictable version of the intensity function. A rigorous definition is as follows: an  $\mathcal{F}_t = \mathcal{F}_t^N$ -progressively measurable process  $\lambda(\cdot)$  is called the  $\mathcal{F}_t$ -intensity of  $N$ , if for any  $a < b$  we have

$$(2) \quad \mathbb{E}[N((a, b]) \mid \mathcal{F}_a] = \mathbb{E}\left[\int_a^b \lambda(t) dt \mid \mathcal{F}_a\right] \quad \text{a.s.}$$

Let  $f(\cdot)$  be a non-negative,  $\mathcal{F}_t$ -progressively measurable process on  $\mathbb{R}$ , and define

$$(3) \quad \int_{-\infty}^{+\infty} f(t) dN(t) = \sum_{-\infty < i < +\infty} f(T_i).$$

Then the following holds true:

$$(4) \quad \mathbb{E} \int_{-\infty}^{+\infty} f(t) dN(t) = \mathbb{E} \int_{-\infty}^{+\infty} f(t) \lambda(t) dt.$$

The existence of an intensity implies that the point process, roughly speaking, behaves locally as a Poisson process. In other words changing the time-scale we get a homogeneous Poisson process. Formally: the increments  $U_k$  defined by

$$(5) \quad U_k := \int_{T_{k-1}}^{T_k} \lambda(t) dt$$

form an i.i.d sequence of r.v.-s with exponential distribution  $\text{Exp}(1)$ , see [2].

Some of the key results in the theory of point processes were obtained back in the 60s and early 70s, such as the rigorous setup in [44], the spectral theory proposed in [1], the introduction of self-exciting point processes, [30], [29], [31], application of point processes in seismology, [60], or results on mixing, [39], [62]. An early pioneering and hardly known work on non-linear Hawkes processes is [36]. A key reference summarizing a number of basic results is the book [2], the author of which has been a major contributor in

the area since then, see [3], [4]. An extensive survey on the general theory and applications of point processes is the book [9]. A more focused introduction is provided by [32]. A number of relevant techniques has been elaborated in the theory of branching processes, see [33].

We note in passing that point-process representation of other variables relevant in epilepsy is also promising. As an example, interictal discharges (IID) are often observed on the seemingly normal EEG between seizures, see [35]. Also high frequency oscillations (HFO) increase both in number and in signal power prior to seizures [61].

#### 4. Linear Hawkes Processes

A simple point process  $N(\cdot)$  on the whole real line  $\mathbb{R}$  is called a linear Hawkes process if its intensity satisfies

$$(6) \quad \lambda(t) = \nu + \int_{-\infty}^{t-} h(t-u) dN(u), \quad \text{with } \nu > 0,$$

where  $\nu$  is a baseline intensity and  $h$  is a non-negative measurable function defined on  $\mathbb{R}_+$ . Due to the feedback connection between  $N(\cdot)$  and its own intensity, Hawkes processes are also called self-exciting point processes. Integration is taken until  $t-$  so that the jump at time  $t$  does not effect  $\lambda(t)$ , ensuring the predictability of  $\lambda(\cdot)$ .

The function Hawkes  $h(\cdot)$  has a variety of names depending on the application area, such as fertility rates, reproduction function, rate of infection or impulse response. We will settle for the last terminology used in the theory of linear systems. Taking a system’s point of view, it is appropriate to explicitly specify the feedback loop defining the pulse train  $(T_k)$ . Let us restrict attention to Hawkes processes with empty history, denoted by  $N^*$ , so that  $N^*((-\infty, 0)) = 0$ , and let  $T_0 = 0$ . Take an i.i.d sequence of r.v.-s  $(U_k)$ ,  $k \geq 1$  with exponential distribution  $\text{Exp}(1)$ , and, recalling the change of time scale, see (5), define  $T_k$ ,  $k \geq 1$  as the solution of the equation

$$(7) \quad \int_{T_{k-1}}^{T_k} \lambda(t) dt = U_k.$$

Recalling circuit theory a standard choice for  $h(\cdot)$  is a finite sum of exponentials. An alternative choice is an impulse response with finite memory, i.e. bounded support, suggested for modeling multiple spike trains, see [38]. Finite memory (multivariate) Hawkes processes seem to be appropriate also for mathematical epidemiology to model the spread of an infection with a finite period of incubation, such as COVID 19.

Hawkes processes have been introduced in [30] and [29], and found applications in many areas, such as seismology [60], finance [15] and neuroscience [56], [24] and [38]. For an extensive survey of application areas see [64].

It is by no means obvious that a point process  $N$  satisfying (6) exists. Assume that a stationary solution with finite mean intensity  $\bar{\lambda} = \mathbb{E}\lambda(t)$  does exist. Taking expectation of (6) we get

$$(8) \quad \bar{\lambda} = \nu + \int_{-\infty}^{t-} h(t-u) \bar{\lambda} du.$$

To ensure  $\bar{\lambda} > 0$  for all  $\nu > 0$  we need the following stability condition:

$$(9) \quad p := \int_0^{\infty} h(t) dt < 1,$$

under which we get for the mean intensity

$$(10) \quad \bar{\lambda} = \nu / \left( 1 - \int_0^{\infty} h(t) dt \right).$$

*Construction.* A classic construction of linear, stationary Hawkes-processes defined all over  $\mathbb{R}$  is obtained by an immigration-birth representation, see [31]. The events are produced recursively, starting with a homogeneous Poisson process  $N^0(\cdot)$  on  $\mathbb{R}$  with intensity  $\nu$ , considered as the generation with index 0 or the set of immigrants. Their arrival times are denoted by  $T_i^0$ , with  $-\infty < i < +\infty$ . Each immigrant arriving at time  $T$  produces children the dates of birth of which are distributed as a conditionally independent Poisson point process on  $[T, \infty)$  with density  $h(t-T)$ , see Kingman [37] for details. The procedure is then repeated from generation to generation over and over again, each off-spring acting independently of all others, resulting in a cluster with center at  $T = T^0$ , the counting process of which will be denoted by  $N^c(T, \cdot)$ .

Let the superposition of the processes  $N^c(T_i^0, \cdot)$ , be denoted by  $N(\cdot)$ . Now, in view of the stability condition  $p = \int_0^\infty h(t) dt < 1$  it is easily seen that  $N((a, b])$  is finite a.s. for any  $-\infty < a < b + \infty$ , in fact  $\mathbb{E}N((a, b]) = (b - a)/(1 - p)$ , and the process  $N(\cdot)$  satisfies the dynamics (6).

An often used alternative to the construction of a stationary linear Hawkes process is as follows: consider a linear Hawkes process  $N^*(\cdot)$  with *empty history*, i.e.  $N^*((-\infty, 0)) = 0$ , the intensity of which satisfies

$$(11) \quad \lambda_t^* = \nu + \int_0^{t-} h(t - u) dN^*(u), \quad t \geq 0.$$

The process  $N^*(\cdot)$  will be asymptotically stationary in the following sense: the restriction of the left shifted processes  $S_t N^*(\cdot)$  to  $\mathbb{R}^+$  will converge, for  $t \rightarrow \infty$ , to the restriction of a stationary linear Hawkes process to  $\mathbb{R}^+$ , denoted by  $N^+(\cdot)$ , in distribution. I.e. for any continuous function  $f$  with bounded support

$$(12) \quad \lim_{t \rightarrow \infty} \int_0^{+\infty} f(u) dS_t N^*(u) = \int_0^{+\infty} f(u) dS_t N(u).$$

A much stronger statement is given as Theorem 1 in [3].

An excellent user friendly survey on linear Hawkes processes is given in the PhD thesis [40].

When modeling neuronal firing with Hawkes processes an important feature is the neurons’ refractory period, during which a neuron can not generate its next AP. This can be taken into account with an appropriately chosen  $h(\cdot)$  combined with the use of nonlinear Hawkes processes, [6].

An example for a non-linear Hawkes process is a simple point process admitting an  $\mathcal{F}_t^N$ -intensity

$$(13) \quad \lambda(t) = (\nu + \int_{-\infty}^{t-} h(t - s) N(ds))_+,$$

where  $h : \mathbb{R}^+ \rightarrow \mathbb{R}$ , so that  $h$  may change sign, and  $x_+ = \max(x, 0)$ . We define the process with empty history in analogy with (11). Theorem 1 of [3] implies the following beautiful claim:

**Proposition 1.** *Assume that  $\int_0^\infty |h(t)| dt < 1$ . Then (13) has a unique weak solution, i.e. there is a unique stationary distribution for  $N$ , with finite*



average intensity. Moreover the process with empty history, satisfying (13) for  $t \geq 0$ , is asymptotically stationary.

### 5. Hawkes processes defined via state-space systems

Recalling the basics of the theory of linear systems a natural specialization of the above is obtained by considering Hawkes processes  $N(\cdot)$  with impulse response functions of the form  $h(t) = c^\top e^{At}b$ , where  $A$  is a stable square matrix of dimension, say,  $m$ , i.e. for the eigenvalues of  $A$  we have  $\Re\lambda_i(A) < 0$  for all  $i$ , and  $b$  and  $c$  are  $m$ -vectors. It is assumed that  $h(t) = c^\top e^{At}b \geq 0$  for all  $t \geq 0$ . Thus the intensity of the process is given by

$$(14) \quad \lambda(t) = \nu + \int_{-\infty}^{t-} c^\top e^{A(t-u)}b \, dN(u).$$

A stationary Hawkes process  $N(\cdot)$  satisfying (14) exists if and only if

$$(15) \quad g := \int_0^\infty c^\top e^{Av}b \, dv = -c^\top A^{-1}b < 1.$$

Under this condition the mean rate  $\bar{\lambda}$  will be given by  $\bar{\lambda} = \nu/(1 + c^\top A^{-1}b)$ . The condition  $h(t) = c^\top e^{At}b \geq 0$  is trivially satisfied if  $b, c \geq 0$  entry-wise, and  $A \geq 0$  entry-wise or at least  $A \geq 0$  is a Metzler matrix (in which all the off-diagonal components are nonnegative). But this is not a necessary condition at all: taking an arbitrary non-singular state-space transformation  $T$ , the triplet  $(TAT^{-1}, Tb, c^\top T^{-1})$  defines the same linear system with the same frequency transfer function given by  $G(s) = c^\top (sI - A)^{-1}b$ . For more details on this and related issues see [28], [58] and the monograph [16].

*Dynamics.* Following a standard line of thought define the state-process

$$(16) \quad x(t) = \int_{-\infty}^{t-} e^{A(t-u)}b \, dN(u).$$

Then the mapping from  $N(\cdot)$  to  $\lambda(\cdot)$  can be realized by the finite dimensional linear system, with  $-\infty < t < \infty$ ,

$$(17) \quad dx(t) = Ax(t) \, dt + b \, dN(t)$$

$$(18) \quad \lambda(t) = c^\top x(t) + \nu.$$

It is readily seen that the (left-continuous) state-process  $x(t)$  can be explicitly computed in any interval  $T_k < t \leq T_{k+1}$ , assuming that the initial value  $x(T_k+)$  is known: namely, in the interval  $T_k < t \leq T_{k+1}$  we have no jumps, and hence (17) reduces to  $dx(t) = Ax(t) dt$ .

The *simulation* of the above Hawkes process with empty history along the lines of [46] reduces to generating  $T_k$  recursively, starting at  $T_0 = 0$ , and with  $(T_k)$  obtained as the solution of

$$g(T) := \int_{T_{k-1}}^T \lambda(t) dt = \int_{T_{k-1}}^T [\nu + c^\top e^{A(t-T_{k-1})} \cdot x(T_{k-1}+)] dt = U_k,$$

where  $(U_k)$  is an i.i.d. sequence of r.v.-s with exponential distribution  $\text{Exp}(1)$ .

Note that the above simulation method readily generalizes to Hawkes processe, with empty history, for which  $h(t) = c^\top e^{At}b$  may take on negative values, compensated by a static non-linearity such as:

$$(19) \quad \lambda(t) = (\nu + \int_0^{t-} c^\top e^{A(t-u)}b dN(u))_+.$$

*Spectral methods.* The spectral theory for self-exciting point processes has been developed back in 1971, [29]. The state space realization, if it exists, yields a simple heuristic argument for determining the spectral density. First, let us define the auto-covariance function, with  $\delta(\tau)$  denoting the Dirac delta function,

$$(20) \quad r(\tau) = \bar{\lambda}\delta(\tau) + \mathbb{E} [ dN(t + \tau) dN(t)] / (dt)^2 - \bar{\lambda}^2.$$

Then the spectral density will be defined as the Fourier transform

$$(21) \quad f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\omega\tau} r(\tau) d\tau.$$

Introducing the integral operator  $G(s) = c^\top (sI - A)^{-1}b$  it is esily seen that we can rewrite the system dynamics and its inverse as

$$(22) \quad (dN(t) - \lambda(t) dt) = (1 - G(s)) (dN(t) - \bar{\lambda} dt)$$

$$(23) \quad (dN(t) - \bar{\lambda} dt) = (1 - G(s))^{-1} (dN(t) - \lambda(t) dt).$$

Let us focus on the second equation. Note that  $dN(t) - \lambda(t) dt$  is an infinitesimal martingale difference, hence its spectral density is constant, say  $\sigma^2$ . Since the constant term of  $|1 - G(i\omega)|^{-2}$  is 1, the spectral density of the l.h.s. will have a limit, for  $|\omega| \rightarrow \infty$ , equal to  $\sigma^2$ . Noting that the spectral density of  $(dN(t) - \bar{\lambda} dt)$  is exactly  $r(\tau)$  we conclude, by the definition of  $f(\omega)$ , that we must have  $\sigma^2 = \bar{\lambda}/2\pi$ , see [29] for details. Thus we get

$$(24) \quad f(\omega) = \frac{\bar{\lambda}}{2\pi} |1 - G(i\omega)|^{-2}.$$

## 6. Multivariate Hawkes processes

Applications in neurosciences call for a multivariate extension of Hawkes processes, where the firing patterns of individual neurons have mutual impact on their intensities. A dramatic deviation though is that the connection between two neurons may be either excitatory or inhibitory, resulting in an impulse response matrix with possibly negative elements.

Excitation and inhibition is realized as a transient positive or negative ‘deflection’ in the afferent neuron’s resting membrane potential. The time-course of such a deflection is often modeled using a bi-exponential function, such as  $h(t) = \frac{ab}{a-b}(e^{-bt} - e^{-at})$ , having a short rising phase and a long descent after its peak. From the neuroscience-perspective it is tempting to use these more realistic, physiologically founded impulse response functions in modeling.

Hawkes processes are also used to explore the functional connectivity graph of a neuronal network based on neuronal firing patterns recorded with implanted micro-electrodes. Based on the size of recording electrodes and anatomical considerations these neurons often effect each other indirectly. It follows that the impulse response function may biphasic, so that excitation on a shorter and inhibition on a longer path can occur, see [7], [38].

Ultimately, the linear model may yield, at least temporally, a negative intensity for some neurons. To by-pass this shortcoming of the model we must resort to some form of truncation, such as taking the positive part of the computed intensity. Thus we obtain a non-linear Hawkes model, the study of which is significantly more intricate, even in the univariate case.

For a start consider a multivariate linear Hawkes process defined as a family of simple point processes  $N_i$ ,  $1 \leq i \leq d$  without common points a.s., having respective  $\mathcal{F}_t$ -densities, with  $\mathcal{F}_t = \sigma \{ \cup \mathcal{F}_{it} \}$ ,

$$(25) \quad \lambda_i(t) = \nu_i + \sum_{j=1}^d \int_{-\infty}^{t-} h_{ij}(t-s) N_j(ds).$$

Taking expectation on both sides we get

$$(26) \quad \bar{\lambda}_i = \nu_i + \sum_{j=1}^d G_{ij} \bar{\lambda}_j \quad \text{with} \quad G_{ij} = \int_0^\infty h_{ij}(u) du.$$

It is then easily seen, using the Perron-Frobenius theorem, see [53], that a necessary and sufficient condition for the existence of a (unique) non-negative solution  $\bar{\lambda}$  for all possible values  $\nu = (\nu_i) \geq 0$  is that  $\rho(G) < 1$ , where  $\rho(\cdot)$  denotes the spectral radius. On the other hand, it can be shown that under the assumption  $\rho(G) < 1$  there exists a unique (in the weak sense) stationary multivariate linear Hawkes process satisfying the dynamics (25).

With a glance on epidemiology: if for all  $(i, j)$  the infection rate  $h_{ij}(\cdot)$  is

$$(27) \quad h_{ij}(u) = 0 \quad \text{or} \quad h_{ij}(u) = p \chi_{[0, T]}(u),$$

depending on the location (distance) of the agents  $i, j$ , with some fixed infection probability  $p$  and a fixed period of incubation  $T$ , then the verification of the stability condition  $\rho(G) < 1$  reduces to the computation of the spectral norm of the incidence matrix of the network.

As in the scalar case, a useful alternative is a multivariate linear Hawkes process  $N^*(\cdot)$  with empty history, i.e.  $N^*((-\infty, 0)) = 0$ , the intensity of which satisfies

$$(28) \quad \lambda_i^*(t) = \nu_i + \sum_{j=1}^d \int_0^{t-} h_{ij}(t-s) dN_j^*(s).$$

In complete analogy with the scalar case it can be shown that the restriction of the left shifted processes  $S_t N^*(\cdot)$  to  $\mathbb{R}^+$  converge to the restriction of the stationary linear Hawkes process  $N(\cdot)$  to  $\mathbb{R}^+$  in distribution. A nice introduction and survey on multivariate linear Hawkes processes can be found in the PhD thesis [40].

In order to model both excitatory or inhibitory connections we should remove the restriction  $h_{ij}(t) \geq 0$  for all  $t \geq 0$ , and ensure the non-negativity of the intensity by inserting a static non-linearity such as  $\varphi(x) = x_+$ ,

$$(29) \quad \lambda_i(t) = (\nu_i + \sum_{j=1}^K \int_{-\infty}^{t-} h_{ij}(t-s) dN_j(s))_+$$

with  $x_+$  being taken coordinate-wise. This model was used in [38] to describe neuronal activity, in particular to model spike trains. A remarkable corollary of Theorem 7 of [3] is the following result:

**Proposition 2.** *Assume that the matrix  $G$  with entries  $G_{ij} = \int_0^\infty |h_{ij}(t)| dt$  has spectral radius strictly less than 1,  $\rho(G) < 1$ . Then there is a unique weak solution of (29), i.e there is a unique stationary law for  $N$ , with finite average intensity, satisfying the dynamics (29). Moreover, the solution of (29) with empty history for  $t \geq 0$  is asymptotically stationary.*

If the support of the transfer function  $h$  is compact, in other words the point process  $N$  has bounded memory, then a much stronger stability result holds: the point process defined via (28) modified so that an arbitrary history on  $\mathbb{R}^-$  is given, is asymptotically stable, even in total variation, appropriately defined, and the convergence is exponentially fast, see Theorem 5, [3].

The rate of convergence to the stationary version of the multivariate non-linear Hawkes process was obtained under increasingly weaker conditions in [43], [4], [55] and in the arXiv paper [34]. An excellent survey on non-linear Hawkes processes, providing also a number of new results on large deviations, is given in the PhD thesis [64].

## 7. Estimation

*General theory.* First we give a brief summary of some of the basic ideas in the statistical analysis of point processes developed in [45]. Assume that the point process generating our data belongs to a parametric family of stationary point processes parametrized by  $\theta$  with  $\theta \in \Theta \subset \mathbb{R}^p$  having intensities  $\lambda(t, \theta)$ . Having  $N(\cdot)$  observed in  $[0, T]$  the negative log-likelihood is given by

$$(30) \quad L(T, \theta) = \int_0^T \lambda(t, \theta) dt - \int_0^T \log \lambda(t, \theta) dN(t),$$

modulo constants. This is purely a theoretical tool, since  $\lambda(t, \theta)$ , being a conditional expectation on the infinite past, is not available.

Under very general, and fairly complex, conditions the following basic identities are established in [45], which in turn are used to prove convergence in probability of the ML estimator:

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial}{\partial \theta} L(T, \theta) \right]_{\theta=\theta^*} &= \mathbb{E} \left[ \int_0^T \frac{\partial}{\partial \theta} \lambda(t, \theta) \left( dt - \frac{dN(t)}{\lambda(t, \theta)} \right) \right]_{\theta=\theta^*} = 0, \\ R^* &:= \frac{1}{T} \mathbb{E} \left[ \frac{\partial^2}{\partial \theta^2} L(T, \theta) \right]_{\theta=\theta^*} = \mathbb{E} \left[ \frac{1}{\lambda^2(T, \theta)} \frac{\partial}{\partial \theta} \lambda(T, \theta) \frac{\partial}{\partial \theta} \lambda(T, \theta)^\top \right]_{\theta=\theta^*}. \end{aligned}$$

A simple, but fundamental result is the following Lemma 2 of [45]:

**Lemma 3.** *Let the point process  $N(\cdot)$ , having intensity  $\lambda(t)$ , be stationary, ergodic and absolutely continuous w.r.t. the standard Poisson process on any finite interval. Let  $\xi(t)$ ,  $t \geq 0$  be a stationary predictable process with finite second order moment. Then we have a.s.*

$$(31) \quad \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \xi(t) \left( dt - \frac{dN(t)}{\lambda(t)} \right) = 0.$$

*ML method for Hawkes processes.* Adapting the above general theory to Hawkes processes consider the problem of identification of their dynamics given the counting data  $N(t)$  for  $0 \leq t \leq T$ . We assume that the impulse response function  $h(\cdot)$  belongs to a known family of functions  $h(\cdot, \theta)$ , with  $\theta$  belonging to an open set  $\Theta \subset \mathbb{R}^p$ , with  $\theta^*$  denoting the true parameter. We may have in mind the impulse response functions of linear systems with a rational transfer function of fixed degree. For any fixed feasible  $\theta$  compute the assumed intensity, with empty history, by

$$(32) \quad \lambda^*(t, \theta) = \nu + \int_0^{t-} \gamma(\theta, t-u) dN(u) = \nu + \sum_{0 \leq T_i < t} \gamma(\theta, t - T_i).$$

Accordingly, we obtain the conditional negative log-likelihood as follows:

$$(33) \quad L^*(T, \theta) = \int_0^T \lambda^*(t, \theta) dt - \int_0^T \log \lambda^*(t, \theta) dN(t).$$

*Exponential impulse response.* The computation of the assumed intensity  $\lambda^*(t, \theta)$  for any given parameter  $\theta$  is in most cases practically not feasible.

A notable exception is a Hawkes process with exponential impulse response function, say  $h(t) = ce^{at}$  with  $c > 0$  and  $a < 0$ , see [46]. Thus, denoting the true parameters by  $\nu^*$ ,  $a^*$  and  $c^*$  the intensity  $\lambda(\cdot)$  is assumed to satisfy

$$(34) \quad \lambda(t) = \nu^* + \int_{-\infty}^{t-} c^* e^{a^*(t-u)} dN(u).$$

A stationary Hawkes process  $N(\cdot)$  satisfying (34) exists if and only if

$$(35) \quad g^* := \int_0^\infty c^* e^{a^*v} v dv = -c^*/a^* < 1 \quad \text{or} \quad a^* + c^* < 0,$$

and the mean rate  $\bar{\lambda}$  will given by  $\bar{\lambda} = \nu^*/(1 - g^*) = \nu^*/(1 + c^*/a^*)$ .

Let us now take a feasible triplet of parameters  $\theta = (\nu, a, c)$  such that  $\nu, c > 0$  and  $a, a + c < 0$ , and define, as in (32),

$$(36) \quad \lambda^*(t, \theta) = \nu + \int_0^{t-} ce^{a(t-u)} dN(u) = \nu + \sum_{0 \leq T_i < t} ce^{a(t-T_i)}.$$

A remarkable property of the exponential impulse response function is that the intensity reconstructed above can be explicitly computed in any interval  $T_k < t < T_{k+1}$ , assuming that the initial value  $\lambda^*(\theta, T_k+)$  is known. Indeed, since we have no jumps for  $T_k < t < T_{k+1}$ , we have

$$\lambda^*(t, \theta) - \nu = e^{a(t-T_k)}(\lambda^*(\theta, T_k+) - \nu).$$

Accordingly, the computation of the conditional negative log-likelihood function  $L^*(T, \theta)$  given in (32) is practically feasible. This above observation is reinforced by considering the dynamics

$$(37) \quad d(\lambda^*(t, \theta) - \nu) = a(\lambda^*(t, \theta) - \nu) dt + c dN(t),$$

and noting that in the interval  $T_k < t < T_{k+1}$  we have  $dN(t) = 0$ .

From (37) the dynamics of the gradient of  $\lambda^*(t, \theta)$  is readily obtained:

$$(38) \quad d\lambda_a^*(t, \theta) = a\lambda_a^*(t, \theta) dt + (\lambda^*(t, \theta) - \nu) dt$$

$$(39) \quad d\lambda_c^*(t, \theta) = a\lambda_c^*(t, \theta) dt + dN(t)$$

$$(40) \quad d\lambda_\nu^*(t, \theta) = a\lambda_\nu^*(t, \theta) dt - a dt.$$

Note that the derivative  $\lambda_\nu^*(t, \theta)$  is not affected by the observations at all, neither directly or indirectly. In fact,  $\lambda_\nu^*(t, \theta) = 1$  for all  $t$ , a trivial consequence of the (off-line) definition of  $\lambda^*(t, \theta)$ . On the other hand,  $\lambda_a^*(t, \theta)$  and  $\lambda_c^*(t, \theta)$  are obtained as the outputs of a stable  $AR(1)$  system, with inputs  $(\lambda^*(t, \theta) - \nu)$  and  $dN(t)$ , respectively.

Taking into account the results of the forthcoming Section 8, in particular Theorems 6 and 7, we can conclude that the gradient process  $\lambda_\theta^*(t, \theta)$  is  $L$ -mixing in the sense given in Section 8. It follows that a powerful technology of system-identification is applicable, leading to strong approximation results for the ML estimator  $\hat{\theta}_T$ , see (49) below, and [21] for further details.

It has been noted already in [46] that a single exponential impulse response function can be replaced by a sum of exponential functions, but no connection to system theory was mentioned. Next we discuss the applicability of a compact framework provided by the theory of linear systems.

*Rational transfer functions.* Consider a more general, finite dimensional linear dynamics defining a linear Hawkes process  $N(\cdot)$  implicitly by the state-space system (14) or (17) and (18), where the triplet  $(A, b, c)$  is parametrized by a parameter  $\theta$ . Let  $\theta \in \Theta \subset \mathbb{R}^p$ , with  $\Theta$  being an open domain, and let the impulse response functions be denoted by  $h(t, \theta) = c(\theta)^\top e^{A(\theta)t} b(\theta)$ .

Assume that for all feasible choices of  $\theta$  we have  $h(t, \theta) \geq 0$  for all  $t$ , and  $g(\theta) := \int_0^\infty h(t, \theta) dt < 1$  holds. Let  $\nu$  be treated as an independent parameter with the only restriction that  $\nu > 0$ . Let the true values, corresponding to the data generating mechanism be denoted by  $\theta^*$  and  $\nu^*$ . Take a pair of feasible values  $\theta$  and  $\nu$ , and consider the state-space system

$$(41) \quad dx^*(t, \theta) = A(\theta)x^*(t, \theta) dt + b(\theta) dN(t) \quad x(0, \theta) = 0,$$

$$(42) \quad \lambda^*(t, \theta) = c^\top(\theta)x^*(t, \theta) + \nu.$$

Now, between two successive events, say for  $T_k < t \leq T_{k+1}$ , we have  $dN(t) = 0$ , and thus the assumed state-process  $x^*(t, \theta)$  and the assumed conditional intensity process  $\lambda^*(t, \theta)$  can be computed explicitly as

$$\lambda^*(t, \theta) = \nu + c^\top(\theta) e^{A(\theta)(t-T_k)} x^*(T_k+, \theta).$$



Let  $\eta^\top = (\theta^\top, \nu)$ . The gradient processes  $x_\theta^*(T_k^+, \theta)$  and  $\lambda_\eta^*(t, \theta, \nu)$  are trivially obtained from (41) and (42), respectively. Thus we get a computationally feasible ML equation

$$(43) \quad \frac{\partial}{\partial \eta} L(T, \eta) = \int_0^T \frac{\partial}{\partial \eta} \lambda^*(t, \eta) \left( dt - \frac{dN(t)}{\lambda^*(t, \eta)} \right) dt = 0.$$

*Identifiability.* The above state-space model of a Hawkes process is (locally) identifiable if the Fisher matrix

$$(44) \quad R^* = \mathbb{E} \left[ \frac{1}{\lambda^2(T, \eta)} \frac{\partial}{\partial \eta} \lambda(T, \eta) \frac{\partial}{\partial \eta} \lambda(T, \eta)^\top \right]_{|\eta=\eta^*}$$

is non-singular. A minimal condition for this is that the impulse response function  $h(t, \theta)$  itself is identifiable in the sense that for any  $v \in \mathbb{R}^p$ ,  $v \neq 0$  we have  $v^\top h_\theta(\cdot, \theta) \neq 0$ . Equivalently, in frequency domain we would require that  $v^\top G_\theta(\cdot, \theta) \neq 0$ .

A standard parameterization to achieve identifiability is to take a canonical state-space realization, such as the one obtained by writing for the transfer function  $1 - c^\top (sI - A)^{-1} b$ , see (22), in ARMA form as  $P(s)/Q(s)$ , where the polynomials  $P(s)$  and  $Q(s)$  are relative prime, with leading coefficients 1, and then taking a state-space realization, where the state transition matrix is a companion matrix defined by  $Q(s)$ .

Now note that  $R^*$  has a block-diagonal structure with blocks

$$(45) \quad R_{11} = \mathbb{E} \left[ \frac{1}{\lambda^2(T, \theta)} \lambda_\theta(T, \theta) \lambda_\theta(T, \theta)^\top \right]_{|\theta=\theta^*} \quad \text{and} \quad R_{22} = 1.$$

Noting that  $\lambda^2(t, \theta) \geq \nu^2 > 0$  for all  $t$  it is sufficient to show that

$$(46) \quad \bar{R}_{11} = \mathbb{E} [ \lambda_\theta(T, \theta) \lambda_\theta(T, \theta)^\top ]_{|\theta=\theta^*}$$

is non-singular. Writing briefly  $\lambda(\cdot, \theta, \nu) = \nu + G(\cdot, \theta) dN$  we get

$$(47) \quad \lambda_\theta(\cdot, \theta, \nu) = G_\theta(\cdot, \theta) dN.$$

At this point we need the spectral theory for self-exciting point processes as discussed above. Letting  $f(\omega)$  denote the spectral density of  $N$ , we have

$$(48) \quad \bar{R}_{11} = \int_{-\infty}^{+\infty} G_{\theta}(i\omega, \theta) f(\omega) G_{\theta}^{\top}(-i\omega, \theta) d\omega.$$

Recalling that  $f(\omega) = \frac{\bar{\lambda}}{2\pi} |1 - G(i\omega)|^{-2} \neq 0$  for all  $\omega$ , see (24) and also [29], we conclude that  $\bar{R}_{11}$  is non-singular iff  $\tilde{R}_{11} = \int_{-\infty}^{+\infty} G_{\theta}(i\omega, \theta) G_{\theta}^{\top}(i\omega, \theta) d\omega$  is non-singular, which is known to be the case if the parametrization is given by the coefficients of  $P(s)$  and  $Q(s)$ , see [42].

Now the remarkable fact, to be discussed below in Section 8, see Theorems 6, 7, is that, under the above ARMA-parametrization  $\lambda^*(t, \theta, \nu)$  and its derivatives are  $L$ -mixing in the sense of [20]. Thus the arsenal of system identification for linear stochastic systems is applicable, and the analysis of the ML estimate for the above Hawkes processes can be carried out exactly along the lines of thought followed for the prediction error estimate for ARMA systems in [21]. Thus we arrive at the following result, where the error term  $O_M(T^{-1/2})$  has  $L_p$ -moments decreasing like  $T^{-1/2}$  for any  $p \geq 1$ :

$$(49) \quad \hat{\eta}_T - \eta^* = -(R^*)^{-1} \int_0^T \frac{\partial}{\partial \eta} \lambda(\eta^*, t) \left( dt - \frac{dN(t)}{\lambda(\eta^*, t)} \right) dt + O_M(T^{-1/2}).$$

As for the accuracy of these estimates are concerned it is well-known that in the vase of AR(1) processes the signal-to-noise ratio, and hence the accuracy of the estimator improves in the vicinity of the stability boundary. This is not quite the case for Hawkes processes, in particular the accuracy for the baseline intensity does not improve, since the corresponding block of the Fisher matrix is  $R_{22} = 1$ . However, the diffusion limit of an appropriately re-scaled intensity process has been established in [49].

To conclude this section let us mention a few specifications imposed by neuroscience. When analyzing human recordings, there is debate in the literature regarding the length of a preictal period representing an already altered brain-state prior to seizure-onset, but in general time-windows ranging from a few minutes to half-an-hour are considered, which are followed by the transition to the seizure itself. The latter transition is generally accepted to happen on the timescale of seconds leading to very low sample-size.

A specific problem in fitting multivariate Hawkes processes, in an effort to reconstruct functional connectivity graphs, is the enormous number of

parameters to be estimated. A sparse connectivity graph with a significantly smaller number of parameters can be obtained using LASSO, see [38].

## 8. Mixing

A stationary point process is (strongly) mixing if for all  $A, B \in \mathcal{F}$

$$(50) \quad \lim_{t \rightarrow \infty} P(A \cap S_{-t}B) = P(A)P(B),$$

see [63]. A wide class of mixing point processes is given in Theorem 6, [62]:

**Proposition 4.** *A stationary cluster process is mixing if the cluster center process is mixing.*

The proof is based on Theorem 8 of Westcott, [63] characterizing mixing for a stationary point process in terms of its probability generating functional defined, for a suitable class of functions  $\xi$ , as

$$(51) \quad G[\xi] = \mathbb{E} [\prod_i \xi(T_i)],$$

where the  $T_i$ -s are the positions of the events. Equivalently,

$$(52) \quad G[\xi] = \mathbb{E} \left[ \int \log \xi(t) \, dN(t) \right].$$

Historically, a basic class of functions to be considered was  $V$ , denoting the set of measurable functions having bounded support, moreover satisfying  $0 \leq \xi(t) \leq 1$  for all  $t$ , see [44], [1], [59]. Extending and complementing the results of [39], presented without proof, the following result, given as Theorem 8. of [63], is a nice characterization of mixing:

**Proposition 5.** *A stationary point process with probability generating functional  $G[\xi]$  is mixing if and only if for any pair of functions  $\xi_1, \xi_2$ , such that  $1 - \xi_1, 1 - \xi_2 \in V$  we have*

$$(53) \quad \lim_{t \rightarrow \infty} G[\xi_1 S^t \xi_2] = G[\xi_1] \cdot G[\xi_2].$$

In another setting, let us recall that while strong mixing is a central concept in the theory of stationary processes, it turned out that for the

statistical analysis of stationary processes generated by linear stochastic systems an alternative mixing concept is significantly more useful. This is the notion of  $L$ -mixing, see [20], a special case of which is widely used in the system identification literature under the name of exponentially stable processes, going back to [41] and [51]. See also [42] and [5].

Let  $\mathcal{G}_t^-$  and  $\mathcal{G}_t^+$ , for  $-\infty < t < +\infty$ , be a non-decreasing and a non-increasing family of  $\sigma$ -algebras respectively, such that  $\mathcal{G}_t^-$  and  $\mathcal{G}_t^+$  are independent for all  $t$ . We say that a stochastic process  $x = (x_t)$ ,  $-\infty < t < +\infty$  is  $L$ -mixing if for all  $q \geq 1$  it is  $L_q$ -bounded, and defining for any  $\tau \geq 0$

$$(54) \quad \gamma_q(\tau) = \sup_{-\infty < t < +\infty} E^{1/q} |x_t - \mathbb{E}[x_t | \mathcal{G}_{t-\tau}^-]|^q,$$

we have  $\Gamma_q := \int_0^\infty \gamma_q(\tau) d\tau < \infty$ , see [20].

The objective of this section is to present the basic ideas for establishing  $L$ -mixing for Hawkes processes with at most exponentially decaying impulse response functions. A key point in this is an unconventional definition of past and future of  $N(\cdot)$ . Let

$$(55) \quad N^-(t, \cdot) := \cup_{T \leq t} N^c(T, \cdot) \quad \text{and} \quad N^+(t, \cdot) := \cup_{T > t} N^c(T, \cdot).$$

Then, obviously  $N(\cdot) = N^-(t, \cdot) \cup N^+(t, \cdot)$  and  $N^-(t, \cdot)$  and  $N^+(t, \cdot)$  are independent. Let us define the past and future  $\sigma$ -algebras

$$(56) \quad \mathcal{G}_t^- = \sigma\{N^-(t, A) : A \in \mathcal{B}(\mathbb{R})\}$$

$$(57) \quad \mathcal{G}_t^+ = \sigma\{N^+(t, A) : A \in \mathcal{B}(\mathbb{R})\}.$$

The relation with the standard definition of past and future of a point process  $N(\cdot)$  is as follows: we have  $\mathcal{F}_t^- \subset \mathcal{G}_t^-$  and  $\mathcal{G}_t^+ \subset \mathcal{F}_t^+$ . Obviously,  $\mathcal{G}_t^-$  and  $\mathcal{G}_t^+$  are independent for all  $t$ .

**Theorem 6.** *Assume that the impulse response function of a Hawkes process  $N(\cdot)$  decays exponentially,  $0 \leq h(t) \leq ce^{at}$  with some  $a < 0$ . Then for any  $\Delta > 0$  the process  $N((t - \Delta, t])$  is  $L$ -mixing w.r.t. the pair  $(\mathcal{G}_t^-, \mathcal{G}_t^+)$ . In fact, for any integer  $q \geq 1$  and any  $a' > a$  we have for  $\tau \geq 0$ , with  $C = C(a')$ ,*

$$(58) \quad \sup_{-\infty < t < +\infty} E^{1/q} |N((t - \Delta, t]) - \mathbb{E}[N((t - \Delta, t]) | \mathcal{G}_{t-\tau}^+]|^q \leq Ce^{a'\tau/q}.$$

The condition  $h(t) \leq ce^{at}$  is not very restrictive and it is certainly satisfied if the underlying linear system has finite memory, i.e. the support of  $h$

is bounded, see [38]. The outline of the proof, based on extending results for tail estimates for a cluster process in [50], will be given in the next section.

The above result establishes a link between the statistical theory of Hawkes processes and system identification for linear stochastic systems. Let us consider a Hawkes process the impulse response function of which  $h(\cdot)$  belongs to a known family of functions  $h(\cdot, \theta)$ , with  $\theta \in \Theta \subset \mathbb{R}^p$ , where  $\Theta$  is an open set. Assume that for all  $\theta \in \Theta$  we have  $h(t, \theta) \leq ce^{at}$  with  $a < 0$ . For the (full) maximum-likelihood method we fix  $\theta \in \Theta$ , and compute the assumed intensity process, based on the observed data  $dN(u)$ :

$$(59) \quad \lambda(t, \theta) = \int_{-\infty}^{t-} h(t-u, \theta) dN(u).$$

**Theorem 7.** *Assume that the conditions of Theorem 6 are satisfied for  $h(\cdot, \theta)$  uniformly in  $\theta$ . Then the process  $\lambda(\cdot, \theta)$  is  $L$ -mixing w.r.t. the pair  $(\mathcal{G}_t^-, \mathcal{G}_t^+)$ , uniformly in  $\theta$ . In fact, for any integer  $q \geq 1$  and any  $a' > a$  we have, with some constant  $C = C(a')$ ,*

$$(60) \quad \sup_{-\infty < t < +\infty} \mathbb{E}^{1/q} |\lambda(t, \theta) - \mathbb{E}[\lambda(t, \theta) | \mathcal{G}_{t-\tau}^+]|^q \leq Ce^{a'\tau/q}.$$

*Outline of the proof of Theorem 7.* The proposition is plausible in light of the methodology of system identification. Namely, as for any  $\Delta > 0$  the process  $N((t-\Delta, t])$  is  $L$ -mixing w.r.t. the pair  $(\mathcal{G}_t^-, \mathcal{G}_t^+)$ , and  $h(\cdot, \theta)$  is exponentially decaying, it follows by [20] that the process

$$(61) \quad \lambda^\Delta(t, \theta) = \int_{-\infty}^{t-} h(t-u, \theta) N((u-\Delta, u]) du$$

is  $L$ -mixing w.r.t. the pair  $(\mathcal{G}_t^-, \mathcal{G}_t^+)$ . Writing

$$N((u-\Delta, u]) = \int_{-\infty}^{+\infty} \chi_{(u-\Delta, u]}(v) dN(v),$$

and applying a suitable version of Fubini’s theorem we get

$$(62) \quad \lambda^\Delta(t, \theta) = \int_{-\infty}^{t-} \bar{h}(t-v, \theta) dN(v),$$

where  $\bar{h}(t, \theta) = \int_{-\infty}^t h(t-u, \theta) \chi_{(0, \Delta]}(u) du = \int_{t-\Delta}^t h(t-u, \theta) du$ . Note that  $\lambda^\Delta(t, \theta)$  has a representation which is almost identical to that of  $\lambda(t, \theta)$ ,

and thus with bit of reverse engineering we get the claim. A direct proof is obtained by adapting the proof of Theorem 6.

### 9. Outline of proof of Theorem 6

As we have already indicated, the proof is based on extending results for tail estimates for a cluster process given in [50]. Let  $N^c(\cdot) = N^c(0, \cdot)$  denote the counting process associated with the cluster having an ancestor at time 0. Consider the number of points in the semi-infinite interval  $[\tau, +\infty)$  denoted by  $N^c([\tau, +\infty))$  for any  $\tau \geq 0$ . Let  $p := \int_0^\infty h(t) dt < 1$ . Following [50] define the logarithmic probability generating function, for  $\zeta > 0$ , as

$$(63) \quad U(\tau, \zeta) = \log [ \mathbb{E} ( e^{\zeta N^c([\tau, +\infty))} ) ].$$

**Theorem 8.** *Assume that the impulse response function is exponentially decaying in  $\tau$ , say  $h(t) \leq ce^{at}$  with  $a < 0$ . Let  $0 < \zeta \leq p - \log p - 1$ . Then  $U(\cdot, \zeta)$  is finite, and it is exponentially decreasing with respect to  $\tau$ : in fact for any  $a' > a$  we have  $U(\tau, \zeta) \leq Ce^{a'\tau}$  with some constant  $C = C(a')$ .*

*Outline of proof of Theorem 8:* It is well-known, and in fact can be easily seen taking into account the reproductive nature of a cluster and Campbell’s theorem, see [37], that  $U(\tau, \zeta)$  satisfies the following equation: for any  $\tau > 0$

$$(64) \quad U(\tau, \zeta) = \int_0^\tau [ e^{U(\tau-t, \zeta)} - 1 ] h(t) dt + (e^{U(0, \zeta)} - 1) \int_\tau^\infty h(t) dt.$$

Incidentally, we get for  $\tau = 0$ , by taking the limit, and noting that at time 0 the ancestor is born, the well-known equality

$$(65) \quad U(0, \zeta) = \zeta + p[e^{U(0, \zeta)} - 1].$$

This equation provides a justification for the range of  $\zeta \leq p - \log p - 1$ : namely, for given  $0 < p < 1$  the above equation has a solution in  $U = U(0, \zeta)$  exactly for  $\zeta \leq p - \log p - 1$ .

From the above we get after fairly standard manipulations that the Laplace-transform  $\mathcal{L}U(s, \zeta)$  is well defined for  $\Re s > a$ . Taking into account that  $U(\tau, \zeta)$  is non-negative and monotone non-increasing for  $\tau > 0$ , the claim is implied by the following elementary lemma:

**Lemma 9.** *Let  $U(\tau), \tau > 0$  be a non-negative, monotone non-increasing function such that its Laplace-transform  $\mathcal{L}U(s)$  is well defined for  $\Re s > a$  for some  $a < 0$ . Then  $U(\cdot)$  is exponentially decreasing: for any  $a' > a$  we have  $U(\tau) \leq Ce^{a'\tau}$  with some constant  $C = C(a')$ .*

Extending the above, let us take the superposition of all clusters with ancestors  $T_i^0 < 0$ , the counting process of which will be denoted by  $N(\mathbb{R}_-, \cdot)$ :

$$(66) \quad N(\mathbb{R}_-, \cdot) = \sum_{T_i^0 < 0} N^c(T_i^0, \cdot).$$

Obviously the terms within the sum are conditionally independent given the set of  $T_i^0$ -s. Let  $\tau > 0$  and let us consider the tail process  $N(\mathbb{R}_-, [\tau, +\infty))$ . Let  $\zeta > 0$  and define the logarithmic probability generating function:

$$(67) \quad V(\tau, \zeta) = \log [ \mathbb{E} (e^{\zeta N(\mathbb{R}_-, [\tau, +\infty))}) ], \quad \tau > 0.$$

**Theorem 10.** *Under the conditions of Theorem 8 the function  $V(\cdot, \zeta)$  is finite for all  $0 < \zeta \leq p - \log p - 1$ , and it is exponentially decreasing in  $\tau$ : in fact for any  $a' > a$  we have  $V(\tau, \zeta) \leq Ce^{a'\tau}$  with some constant  $C = C(a')$ .*

The proof of Theorem 10 is obtained from Theorem 8 by a direct application of Campbell’s theorem, see [37].

To conclude the outline of the proof of Theorem 6 let  $\tau > 0$  and let us consider the count  $N[\tau - \Delta, \tau)$ . Obviously, it is measurable w.r.t. the past of  $\tau$  defined as  $\mathcal{G}_\tau$ . Let us decompose  $N[\tau - \Delta, \tau)$  as

$$(68) \quad N[\tau - \Delta, \tau) = N(\mathbb{R}_-, [\tau - \Delta, \tau)) + N(\mathbb{R}_+, [\tau - \Delta, \tau)).$$

Here  $N(\mathbb{R}_+, [\tau - \Delta, \tau))$  serves as a  $\mathcal{G}_0^+$ -measurable approximation to the count  $N[\tau - \Delta, \tau)$ , in fact,

$$(69) \quad N(\mathbb{R}_+, [\tau - \Delta, \tau)) = \mathbb{E}[N((\tau - \Delta, \tau)] | \mathcal{G}_0^+].$$

The error of this approximation,  $N(\mathbb{R}_-, [\tau - \Delta, \tau))$ , was shown to be exponentially small in the sense of Theorem 10. Shifting the process  $N$  to the left by an amount of  $\tau$  we get a  $\mathcal{G}_{-\tau}^+$ -measurable approximation to  $N[-\Delta, 0)$ , the error of which is exponentially small in the sense of Theorem 10. The claim now follows by elementary inequalities between exponential moments and  $L_q$ -norms.

## 10. Discussion and conclusion

The extension of the results of the last section to multivariate Hawkes processes seems to be straightforward. Extension non-linear Hawkes processes considered in this paper is, however wide open. An additional practical advantage of having  $L$ -mixing established is that this opens up the avenue to the analysis of recursive, on-line estimators of Hawkes processes as proposed in [23], using the methodology of [22].

A new level of difficulty is to model neuronal activity based on data acquired from electroencephalography (EEG). The shortcoming of this method is that it represents the average of filtered electrical activity of billions of neurons, with a temporal resolution of tens of milliseconds. In simplified mathematical terms our observation process is of the form

$$(70) \quad \mu_i(t) = \sum_{j=1}^d \int_{-\infty}^{t-} k_{ij}(t-s) N_j(ds),$$

where the range of the indices  $i$  is several order of magnitudes smaller than the range of the indices  $j$ , identifying individual neurons. While the basic mathematical technology for non-linear filtering in a fairly general setting is available, see [27], [26], its adaptation to the current situation seems to be far from being trivial.

## 11. Acknowledgement

This research was partially supported by the National Research, Development and Innovation Office of Hungary, Grant no: EFOP-3.6.2.-16-2017-00015, title: "Extending the Hungarian Service Network for Mathematics in Industry and Innovations.

## References

- [1] M. S. Bartlett. *The spectral analysis of point processes*. Journal of the Royal Statistical Society: Series B (Methodological), **25**(2):264–281, 1963.



- [2] P. Brémaud. Point processes and queues: martingale dynamics, volume 50. Springer Verlag, New York, 1981.
- [3] P. Brémaud and L. Massoulié. *Stability of nonlinear Hawkes processes*. Annals of Probability, **24**(3):1563–1588, 1996.
- [4] P. Brémaud, G. Nappo and G. Torrisi. *Rate of convergence to equilibrium of marked Hawkes processes*. Journal of Applied Probability, **39**(1):123–136, 2002.
- [5] P. E. Caines. Linear stochastic systems, volume 77 of *Classics in Applied Mathematics*. SIAM, 2018.
- [6] J. Chevallier. *Mean-field limit of generalized Hawkes processes*. Stochastic Processes and their Applications, **127**(12):3870–3912, 2017.
- [7] E. Chornoboy, L. Schramm and A. Karr. *Maximum likelihood identification of neural point process systems*. Biological Cybernetics, **59**(4-5):265–275, 1988.
- [8] M. J. Cook, T. J. O’Brien, S. F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi et al. *Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study*. The Lancet. Neurology, **12**(6):563–71, 2013.
- [9] D. J. Daley and D. Vere-Jones. An introduction to the theory of point processes: volume II: General theory and structure. Probability and Its Applications. Springer, New York, 2007.
- [10] D. Duncan, R. Talmon, H. P. Zaveri and R. R. Coifman. *Identifying preseizure state in intracranial EEG data using diffusion kernels*. Math. Biosci. Eng., **10**(3):579–590, 2013.
- [11] T. E. Duncan. Probability densities for diffusion processes with applications to nonlinear filtering theory and detection theory. Ph.D. thesis, Stanford University, 1967.
- [12] T. E. Duncan. *Evaluation of likelihood functions*. Information and Control, **13**(1):62–74, 1968.
- [13] T. E. Duncan, L. Guo and B. Pasik-Duncan. *Adaptive continuous-time linear quadratic Gaussian control*. IEEE Transactions on Automatic Control, **44**(9):1653–1662, 1999.

- [14] T. E. Duncan, Y. Hu and B. Pasik-Duncan. *Stochastic calculus for fractional Brownian motion I. Theory*. SIAM Journal on Control and Optimization, **38**(2):582–612, 2000.
- [15] P. Embrechts, T. Liniger and L. Lin. *Multivariate Hawkes processes: an application to financial data*. Journal of Applied Probability, **48**(A):367–378, 2011.
- [16] L. Farina and S. Rinaldi. *Positive linear systems: theory and applications*, volume 50. John Wiley & Sons, 2011.
- [17] R. S. Fisher, W. Van Emde Boas, W. Blume, C. Elger, P. Genton, P. Lee et al. *Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)*. Epilepsia, **46**(4):470–472, 2005.
- [18] C. Flamm, A. Graef, S. Pirker, C. Baumgartner and M. Deistler. *Influence analysis for high-dimensional time series with an application to epileptic seizure onset zone detection*. Journal of Neuroscience Methods, **214**(1):80–90, 2013.
- [19] K. Gadhomi, J.-M. Lina, F. Mormann and J. Gotman. *Seizure prediction for therapeutic devices: A review*. Journal of Neuroscience Methods, (29):1–13, 2015.
- [20] L. Gerencsér. *On a class of mixing processes*. Stochastics, **26**(3):165–191, 1989.
- [21] L. Gerencsér. *On the martingale approximation of the estimation error of ARMA parameters*. Systems & Control Letters, **15**(5):417–423, 1990.
- [22] L. Gerencsér. *Rate of convergence of recursive estimators*. SIAM Journal on Control and Optimization, **30**(5):1200–1227, 1992.
- [23] L. Gerencsér, C. Matias, B. Torma, Z. Vágó and B. Weiss. *Self-exciting point processes with applications in finance and medicine*. In Proc. of the International Symposium on Mathematical Theory of Networks and Systems (MTNS 2008), page 10. 2008.
- [24] F. Gerhard, M. Deger and W. Truccolo. *On the stability and dynamics of stochastic spiking neuron models: Nonlinear Hawkes process and point process GLMs*. PLoS Computational Biology, **13**(2):1–31, 2017.
- [25] C. W. Granger. *Investigating causal relations by econometric models and cross-spectral methods*. Econometrica: journal of the Econometric Society, pages 424–438, 1969.

- [26] B. Grigelionis. *On nonlinear filtering theory and absolute continuity of measures, corresponding to stochastic processes*. In Proceedings of the Second Japan-USSR Symposium on Probability Theory, pages 80–94. Springer, 1973.
- [27] B. Grigelionis. *Stochastic nonlinear filtering equations and semimartingales*. In Nonlinear Filtering and Stochastic Control, pages 63–99. Springer, 1982.
- [28] H. M. Härdin and J. H. Van Schuppen. *Observers for linear positive systems*. Linear Algebra and its Applications, **425**(2-3):571–607, 2007.
- [29] A. G. Hawkes. *Point spectra of some mutually exciting point processes*. Journal of the Royal Statistical Society: Series B (Methodological), **33**(3):438–443, 1971.
- [30] A. G. Hawkes. *Spectra of some self-exciting and mutually exciting point processes*. Biometrika, **58**(1):83–90, 1971.
- [31] A. G. Hawkes and D. Oakes. *A cluster process representation of a self-exciting process*. Journal of Applied Probability, **11**(3):493–503, 1974.
- [32] M. Jacobsen. *Point process theory and applications: marked point and piecewise deterministic processes*. Springer Science & Business Media, 2006.
- [33] P. Jagers. *Branching processes with biological applications*. Wiley, 1975.
- [34] D. Karabash. *On stability of Hawkes process*. arXiv:1201.1573, 2012.
- [35] P. J. Karoly, D. R. Freestone, R. Boston, D. B. Grayden, D. Himes, K. Leyde et al. *Interictal spikes and epileptic seizures: Their relationship and underlying rhythmicity*. Brain, **139**(4):1066–1078, 2016.
- [36] J. Kerstan. *Teilprozesse Poissonscher prozesse*. In Trans. Third Prague Conf. Information Theory, Statist. Decision Functions, Random Processes (Liblice, 1962), volume 377, page 7. 1964.
- [37] J. F. C. Kingman. *Poisson processes*, volume 6 of *Encyclopedia of Biostatistics*. Wiley Online Library, 2005.
- [38] R. C. Lambert, C. Tuleau-Malot, T. Bessaih, V. Rivoirard, Y. Bouret, N. Leresche et al. *Reconstructing the functional connectivity of multiple spike trains using Hawkes models*. Journal of Neuroscience Methods, **297**:9–21, 2018.

- [39] V. Leonov. *The use of the characteristic functional and semi-invariants in the ergodic theory of stationary processes*. In Doklady Akademii Nauk USSR, volume 133, pages 523–526. 1960.
- [40] T. J. Liniger. Multivariate Hawkes processes. Ph.D. thesis, ETH Zurich, 2009.
- [41] L. Ljung. *On consistency and identifiability*. In Stochastic Systems: Modeling, Identification and Optimization, I, pages 169–190. Springer, 1976.
- [42] L. Ljung and T. Söderström. Theory and practice of recursive identification. MIT press, 1983.
- [43] L. Massoulié. *Stability results for a general class of interacting point processes dynamics, and applications*. Stochastic Processes and Their Applications, **75**(1):1–30, 1998.
- [44] J. E. Moyal et al. *The general theory of stochastic population processes*. Acta Mathematica, **108**:1–31, 1962.
- [45] Y. Ogata. *The asymptotic behaviour of maximum likelihood estimators for stationary point processes*. Annals of the Institute of Statistical Mathematics, **30**(1):243–261, 1978.
- [46] T. Ozaki. *Maximum likelihood estimation of Hawkes’ self-exciting point processes*. Annals of the Institute of Statistical Mathematics, **31**(1):145–155, 1979.
- [47] J. T. Paz and J. R. Huguenard. *Microcircuits and their interactions in epilepsy: Is the focus out of focus?* Nature Neuroscience, **18**(3):351–359, 2015.
- [48] V. Pernice, B. Staude, S. Cardanobile and S. Rotter. *Recurrent interactions in spiking networks with arbitrary topology*. Physical review E, **85**(3):031.916, 2012.
- [49] V. Prokaj and B. Torma. *Identification of almost unstable Hawkes processes*. Publicationes Mathematicae Debrecen, **78**(1):8, 2011.
- [50] P. Reynaud-Bouret and E. Roy. *Some non asymptotic tail estimates for Hawkes processes*. Bulletin of the Belgian Mathematical Society-Simon Stevin, **13**(5):883–896, 2007.

- [51] J. Rissanen and P. Caines. *The strong consistency of maximum likelihood estimators for ARMA processes*. The Annals of Statistics, **7**(2):297–315, 1979.
- [52] A. Schulze-Bonhage and A. Kühn. *Unpredictability of Seizures and the Burden of Epilepsy*. In Seizure Prediction in Epilepsy: From Basic Mechanisms to Clinical Applications, pages 1–10. 2008.
- [53] E. Seneta. Non-negative matrices and Markov chains. Springer Science & Business Media, 2006.
- [54] W. O. Tatum, G. Rubboli, P. W. Kaplan, S. M. Mirsatari, K. Radhakrishnan, D. Gloss et al. *Clinical utility of EEG in diagnosing and monitoring epilepsy in adults*. Clinical Neurophysiology, **129**(5):1056–1082, 2018.
- [55] G. Torrisi. *A class of interacting marked point processes: rate of convergence to equilibrium*. Journal of Applied Probability, **39**(1):137–160, 2002.
- [56] W. Truccolo. *From point process observations to collective neural dynamics: Nonlinear Hawkes process GLMs, low-dimensional dynamics and coarse graining*. Journal of Physiology-Paris, **110**(4):336–347, 2016.
- [57] W. Truccolo, J. A. Donoghue, L. R. Hochberg, E. N. Eskandar, J. R. Madsen, W. S. Anderson et al. *Single-neuron dynamics in human focal epilepsy*. Nature Neuroscience, **14**(5):635–641, 2011.
- [58] J. Van Den Hof. *Realization of positive linear systems*. Linear Algebra and its Applications, **256**:287–308, 1997.
- [59] D. Vere-Jones. *Some applications of probability generating functionals to the study of input-output streams*. Journal of the Royal Statistical Society: Series B (Methodological), **30**(2):321–333, 1968.
- [60] D. Vere-Jones. *Stochastic models for earthquake occurrence*. Journal of the Royal Statistical Society: Series B (Methodological), **32**(1):1–45, 1970.
- [61] S. A. Weiss, C. Alvarado-Rojas, A. Bragin, E. Behnke, T. Fields, I. Fried et al. *Ictal onset patterns of local field potentials, high frequency oscillations, and unit activity in human mesial temporal lobe epilepsy*. Epilepsia, **57**(1):111–21, 2016.

- [62] M. Westcott. *On existence and mixing results for cluster point processes*. Journal of the Royal Statistical Society: Series B (Methodological), **33**(2):290–300, 1971.
- [63] M. Westcott. *The probability generating functional*. Journal of the Australian Mathematical Society, **14**(4):448–466, 1972.
- [64] L. Zhu. Nonlinear Hawkes processes. Ph.D. thesis, New York University, 2013.

INSTITUTE FOR COMPUTER SCIENCE AND CONTROL (SZTAKI)  
BUDAPEST, HUNGARY  
AND SZÉCHENYI ISTVÁN UNIVERSITY, GYŐR, HUNGARY.  
*E-mail address:* `gerencser.laszlo@sztaki.hu`.

FACULTY OF INFORMATION TECHNOLOGY AND BIONICS,  
PÁZMÁNY PÉTER CATHOLIC UNIVERSITY (PPKE ITK), BUDAPEST, HUNGARY  
AND SZÉCHENYI ISTVÁN UNIVERSITY, GYŐR, HUNGARY.  
*E-mail address:* `perczel.gyorgy.miklos@itk.ppke.hu`

NATIONAL INSTITUTE OF CLINICAL NEUROSCIENCES (OKITI),  
BUDAPEST, HUNGARY,  
AND FACULTY OF INFORMATION TECHNOLOGY AND BIONICS,  
PÁZMÁNY PÉTER CATHOLIC UNIVERSITY (PPKE ITK), BUDAPEST, HUNGARY  
*E-mail address:* `l.g.eross@gmail.com`

NATIONAL INSTITUTE OF CLINICAL NEUROSCIENCES (OKITI),  
BUDAPEST, HUNGARY,  
AND FACULTY OF INFORMATION TECHNOLOGY AND BIONICS,  
PÁZMÁNY PÉTER CATHOLIC UNIVERSITY (PPKE ITK), BUDAPEST, HUNGARY  
*E-mail address:* `fabo.daniel@gmail.com`.

FACULTY OF INFORMATION TECHNOLOGY AND BIONICS,  
PÁZMÁNY PÉTER CATHOLIC UNIVERSITY (PPKE ITK), BUDAPEST, HUNGARY,  
AND SZÉCHENYI ISTVÁN UNIVERSITY, GYŐR, HUNGARY.  
*E-mail address:* `vago.zsuzsa@itk.ppke.hu`