Field master equation theory of the self-excited Hawkes process

Kiyoshi Kanazawa[1] and Didier Sornette[2,3,4]

[1]*Faculty of Engineering, Information and Systems, University of Tsukuba, Tennodai, Tsukuba, Ibaraki 305-8573, Japan*
[2]*ETH Zurich, Department of Management, Technology, and Economics, Zurich 8092, Switzerland*
[3]*Tokyo Tech World Research Hub Initiative, Institute of Innovative Research, Tokyo Institute of Technology, Tokyo 152-8550, Japan*
[4]*Institute of Risk Analysis, Prediction, and Management, Academy for Advanced Interdisciplinary Studies,*
Southern University of Science and Technology, Shenzhen 518055, China

A field theoretical framework is developed for the Hawkes self-excited point process with arbitrary memory kernels by embedding the original non-Markovian one-dimensional dynamics onto a Markovian infinite-dimensional one. The corresponding Langevin dynamics of the field variables is given by stochastic partial differential equations that are Markovian. This is in contrast to the Hawkes process, which is non-Markovian (in general) by construction as a result of its (long) memory kernel. We derive the exact solutions of the Lagrange-Charpit equations for the hyperbolic master equations in the Laplace representation in the steady state, close to the critical point of the Hawkes process. The critical condition of the original Hawkes process is found to correspond to a transcritical bifurcation in the Lagrange-Charpit equations. We predict a power law scaling of the probability density function (PDF) of the intensities in an intermediate asymptotic regime, which crosses over to an asymptotic exponential function beyond a characteristic intensity that diverges as the critical condition is approached. We also discuss the formal relationship between quantum field theories and our formulation. Our field theoretical framework provides a way to tackle complex generalization of the Hawkes process, such as nonlinear Hawkes processes previously proposed to describe the multifractal properties of earthquake seismicity and of financial volatility.

I. INTRODUCTION

The self-excited conditional Poisson process introduced by Hawkes [1–3] has progressively been adopted as a useful first-order model of intermittent processes with time (and space) clustering, such as those occurring in seismicity and financial markets. The Hawkes process was first used and extended in statistical seismology and remains probably the most successful parsimonious description of earthquake statistics [4–9]. More recently, the Hawkes model has known a burst of interest in finance (see, e.g., Ref. [10] for a short review), as it was realized that some of the stochastic processes in financial markets can be well represented by this class of models [11], for which the triggering and branching processes capture the herding nature of market participants (be they due to psychological or rational imitation of human traders or as a result of machine learning and adapting). In field of financial economics, the Hawkes process has been successfully involved in issues as diverse as estimating the volatility at the level of transaction data, estimating the market stability [11–13], accounting for systemic risk contagion, devising

optimal execution strategies, and capturing the dynamics of the full order book [14]. Another domain of intense use of the Hawkes model and its many variations is found in the field of social dynamics on the internet, including instant messaging and blogging such on Twitter [15] as well as the dynamics of book sales [16], video views [17], success of movies [18], and so on.

The present article, together with Ref. [19], can be considered as a sequel complementing a series of papers devoted to the analysis of various statistical properties of the Hawkes process [20–24]. These papers have extended the general theory of point processes [25] to obtain general results on the distributions of total number of events, total number of generations, and so on, in the limit of large time windows. Here, we consider the opposite limit of very small time windows, and characterize the distribution of "intensities," where the intensity $\nu(t)$ of the Hawkes process at time t is defined as the probability per unit time that an event occurs [more precisely, $\nu(t)dt$ is the probability that an event occurs between t and $t + dt$]. We propose a natural formulation of the Hawkes process in the form of a field theory of probability density functionals taking the form of a field master equation. This formulation is found to be ideally suited to investigate the hitherto ignored properties of the distribution of Hawkes intensities, which we analyze in depth in a series of increasingly sophisticated forms of the memory kernel characterizing how past events influence the triggering of future events.

This paper is organised as follows. Section II presents the Hawkes process in its original definition. In addition, we provide a comprehensive review of the previous literature on non-Markovian stochastic processes of diffusive transport developed in traditional statistical physics. Historically, the analytical properties of the generalized Langevin equation (GLE) have been intensively studied and we provide a brief review on the similarity and dissimilarity between the GLE and the Hawkes process from the viewpoint of the Markov embedding of the original non-Markovian one-dimensional dynamics onto a Markovian field dynamics. We then proceed to develop a stochastic Markovian partial differential equation equivalent to the Hawkes process in Sec. III. This is done first for the case where the memory kernel is a single exponential, then made of two exponentials, an arbitrary finite number of exponentials, and finally for general memory kernels. It is in Sec. III that the general field master equations are derived. Section IV presents the analytical treatment and provides the solutions of the master equations, leading to the derivation of the probability density function of the Hawkes intensities for the various above-mentioned forms of the memory kernel. We also discuss a formal relationship between quantum field theories and our formulation in Sec. V. Section VI summarizes and concludes by outlining future possible extensions of the formalism. These sections are complemented by seven Appendixes, in which the detailed analytical derivations are provided.

II. MODEL AND LITERATURE REVIEW

A. Notation

First, we explain our mathematical notation for stochastic processes. By convention, we denote stochastic variables with a hat symbol, such as $\hat{A}$, to distinguish them from the nonstochastic real numbers A, corresponding, for instance, to a specific realization of the random variable. The ensemble average of any stochastic variable $\hat{A}$ is also denoted by $\langle \hat{A} \rangle$. The probability density function (PDF) of any stochastic $\hat{A}$ is denoted by $P_t(\hat{A}(t) = A) = P_t(A)$. The PDF characterizes the probability that $\hat{A}(t) \in [A, A + dA]$ as $P_t(A)dA$. Using the notation of the PDF, the ensemble average reads $\langle \hat{A}(t) \rangle := \int A P_t(A) dA$.

We also make the following remark on notations used for our functional analysis. For any function $z(x) \in \mathcal{S}_F$ defined for $x \in \mathbf{R}_+ = (0, \infty)$ with a function space $\mathcal{S}_F$, we can consider a functional $f[\{z(x)\}_{x \in \mathbf{R}_+}]$, i.e., $f : \mathcal{S}_F \to \mathbf{R}$. Functionals in this paper are often abbreviated as $f[z] := f[\{z(x)\}_x]$, with the square brackets emphasized to distinguish from ordinary functions.

B. Definition of the Hawkes conditional Poisson process

The Hawkes process is the simplest self-excited point process, which describes with a linear intensity function $\hat{\nu}$ how past events influence the triggering of future events. Its structure is particularly well suited to address the general and important question occurring in many complex systems of disentangling the exogenous from the endogenous sources of observed activity. It has been and continues to be a very useful model in geophysical, social, and financial systems.
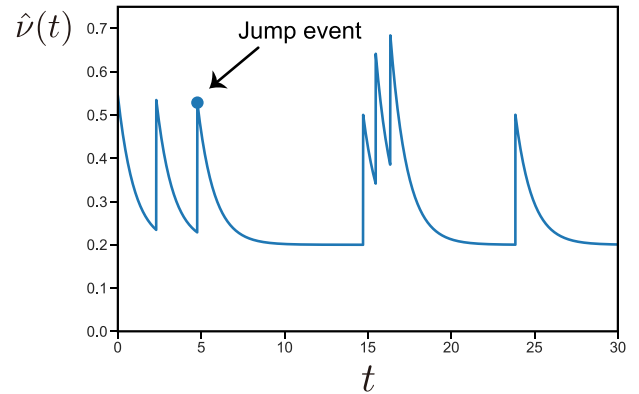

FIG. 1. Schematic representation of the Hawkes process (1) with an exponential kernel (13). The event occurring at a given time stamp is represented by a jump in the intensity $\hat{\nu}(t)$, the probability per unit time that a next event will occur.

The Hawkes process, as any other point process, deals with events ("points" along the time axis). The theory of point processes indeed considers events as being characterized by a time of occurrence but vanishing duration (the duration of events is very small compared to the interevent times). Thus, to a given event i is associated a time t_i of occurrence. In the case where one deals with spatial point processes, the event has also a position $\vec{r}_i$. A "mark" m_i can be included to describe the event's size, or its "fertility," i.e., the average number of events it can trigger directly.

The stochastic dynamics of the Hawkes process is defined as follows. Let us introduce a state variable $\hat{\nu}$, called the intensity. The intensity $\hat{\nu}$ is a statistical measure of the frequency of events per unit time (i.e., a shock occurs during $[t, t + dt)$ with the probability of $\hat{\nu}dt$). In the Hawkes process, the intensity satisfies the following stochastic sum equation (see Fig. 1):

$$\hat{\nu}(t) = \nu_0 + n \sum_{i=1}^{\hat{N}(t)} h(t - \hat{t}_i), \tag{1}$$

where ν_0 is the background intensity, $\{\hat{t}_i\}_i$ represent the time series of events, n is a positive number called the branching ratio, $h(t)$ is a normalized positive function [i.e., $\int_0^\infty h(t)dt = 1$], and $\hat{N}(t)$ is the number of events during the interval $[0, t)$ (called "counting process"). One often refers to $\hat{\nu}(t)$ as a conditional intensity in the sense that, conditional on the realized sequence of $\hat{N}(t) = k$ (with $k \geqslant 0$) events, the probability that the $(k + 1)$-th event occurs during $[t, t + dt)$, such that $\hat{t}_{k+1} \in [t, t + dt)$, is given by $\hat{\nu}(t)dt$. The pulse (or memory) kernel $h(t)$ represents the non-Markovian influence of a given event and is non-negative.

The branching ratio n is a very fundamental quantity, which is the average number of events of first generation ("daughters") triggered by a given event [8,25]. This definition results from the fact that the Hawkes model, as a consequence of the linear structure of its intensity (1), can be mapped exactly onto a branching process, making unambiguous the concept of generations: More precisely, a given realization of the Hawkes process can be represented by the set of all possible tree combinations, each of them weighted by a certain probability

derived from the intensity function [26]. The branching ratio is the control parameter separating three different regimes: (i) $n < 1$, subcritical; (ii) $n = 1$, critical; and (iii) $n > 1$, supercritical or explosive (with a finite probability). The branching ratio n can be shown to be also the fraction of events that are endogenous, i.e., that have been triggered by previous events [27].

C. Review of non-Markovian stochastic processes in the framework of the generalized Langevin equation

This subsection provides the background of previous methods for non-Markovian stochastic processes in statistical physics, by focusing on the diffusive dynamics of Brownian particles (e.g., see Refs. [28] for detailed reviews). While this class of physical stochastic processes exhibits dynamical characteristics that are quite different from those of the Hawkes processes, our framework can been formally related to such standard theories (in particular for the Markov embedding techniques for non-Markovian processes). We thus offer a comprehensive review to prepare the reader to better understand our theoretical developments. This subsection is written in a self-contained way and is not needed to understand our main results; readers only interested in our formulation can skip this section.

1. Markovian Langevin equations

a. Langevin equation. In the context of statistical physics, non-Markovian stochastic processes have been studied from the viewpoint of diffusion processes [29,30]. One of the typical diffusive models is the Langevin equation,

$$M\frac{d\hat{v}(t)}{dt} = -\frac{dU(\hat{x})}{d\hat{x}} - \gamma\hat{v}(t) + \hat{\eta}(t), \quad \frac{d\hat{x}(t)}{dt} = \hat{v}(t), \quad (2)$$

where $\hat{v}(t)$ and $\hat{x}(t)$ are the velocity and the position of the Brownian particle, M is its mass, $U(\hat{x})$ is the confining potential, γ is the viscous friction coefficient, and $\hat{\eta}(t)$ represents the thermal fluctuation modeled by the zero-mean white Gaussian noise. Equation (2) is one of the stochastic differential equations (SDE) first written down in the history of physics.

The noise term embodying the presence of thermal fluctuation satisfies the fluctuation-dissipation relation (FDR)

$$\langle\hat{\eta}(t)\hat{\eta}(t')\rangle = 2\gamma T\delta(t - t'), \quad (3)$$

where T is the temperature. In this paper, the Boltzmann constant is taken unity: $k_B = 1$. The FDR must hold for relaxation dynamics near equilibrium states because both viscous friction and thermal fluctuation come from the same thermal environment [30].

The standard analytical solution to this Langevin equation can be obtained via the time-evolution equation for the joint PDF $P_t(v, x)$, which is given by the Fokker-Planck (FP) equation [31,32]:

$$\frac{\partial P_t(v, x)}{\partial t} = \mathcal{L}_{FP}P_t(v, x),$$

$$\mathcal{L}_{FP} := -\frac{\partial}{\partial x}v + \frac{\gamma}{M}\left[\frac{\partial}{\partial v}\left(v + \frac{dU(x)}{dx}\right) + \frac{T}{M}\frac{\partial}{\partial v}\right]. \quad (4)$$

It is remarkable that the FP equation is always linear in terms of the PDF even if the Langevin equation has nonlinear terms in general. While there is a one-to-one correspondence between the Langevin equation and the FP equation, the FP equation is often analyzed because the standard methods based on linear algebra are available, such as the eigenfunction expansion.

The Langevin equation (2) can be interpreted as the equation of motion for a Brownian particle, obtained after integrating out the many degrees of freedom of the original microscopic dynamics except for those of the Brownian particles. For example, Eq. (2) can be systematically derived from the Hamiltonian dynamics of the Brownian particle surrounded by a dilute gas (see the kinetic frameworks [32–34] for examples).

b. Non-Markovian nature as a result of variable elimination. Let us focus on the case of a harmonic potential $U(\hat{x}) = \frac{1}{2}k\hat{x}^2$ and eliminate the velocity [30] to change the system descriptions from $(\hat{v}, \hat{x})$ to $\hat{x}$. By integrating out the velocity degree, we obtain the non-Markovian dynamics for the position $\hat{x}(t)$:

$$\frac{d\hat{x}(t)}{dt} = -\int_0^\infty ds K(s)\hat{x}(t - s) + \hat{\zeta}(t),$$

$$\hat{\zeta}(t) := \frac{1}{M}\int_0^\infty ds e^{-\gamma s/M}\hat{\eta}(t - s). \quad (5)$$

Here, the memory kernel $K(t) := ke^{-\gamma t/M}/M$ represents the retarded potential effect and the noise term $\hat{\zeta}(t)$ is the colored Gaussian noise with zero mean $\langle\hat{\zeta}(t)\rangle = 0$ and autocorrelation $\langle\hat{\zeta}(t)\hat{\zeta}(t')\rangle = (T/M)e^{-\gamma|t-t'|/M}$. Remarkably, the non-Markovian nature has appeared as a result of variable elimination. Here the non-Markovian version of the FDR holds $\langle\hat{\zeta}(t)\hat{\zeta}(t')\rangle = \langle x^2\rangle_{eq}K(|t - t'|)$ with equilibrium average $\langle x^2\rangle_{eq} := k_BT/k$.

2. Generalized non-Markovian Langevin equation

While the Markovian Langevin description (2) is reasonable for dilute thermal environments [35,36], such a Markovian description is not available for dense thermal environments, such as liquids [35,37,38]. Indeed, Eq. (2) is not valid even for a Brownian particle in water, which is one of the most historically important cases. For such cases, the Langevin description must be modified to accommodate non-Markovian effects originating from hydrodynamic interactions in liquids. The minimal model for such systems is given by the generalized Langevin equation (GLE) [29,30]:

$$M\frac{d\hat{v}(t)}{dt} = -\frac{dU(\hat{x})}{d\hat{x}} - \int_0^\infty ds K(s)\hat{v}(t - s) + \hat{\eta}(t), \quad (6)$$

where $K(t)$ is the memory kernel for viscous friction and $\hat{\eta}(t)$ is the thermal fluctuation modeled by a colored Gaussian noise with zero mean $\langle\hat{\eta}(t)\rangle = 0$. Since both viscous friction and thermal fluctuation share the same origin, they are related to each other via the FDR for non-Markovian processes:

$$\langle\hat{\eta}(t)\hat{\eta}(t')\rangle = TK(|t - t'|). \quad (7)$$

For typical three-dimensional liquid systems, the memory kernel has a long tail due to the hydrodynamic retardation

effect, such that $K(t) \propto t^{-3/2}$, which was verified by direct experiments in Refs. [35,37,38].

The GLE (6) can be derived from microscopic dynamics by rearrangement of the Liouville operator by the method of the projection operators [30,39]. Assuming that the initial state of the system is sufficiently close to equilibrium and that a sufficient number of macroscopic variables are accessible via experiments, the projection operator formalism provides a microscopic foundation of the non-Markovian stochastic processes with intuitive physical interpretation.

The solution of the GLE (6) is much harder to obtain than that of the Markovian Langevin equations due to its non-Markovian nature. However, due to the linearity of the dynamics and the Gaussianity of the thermal fluctuations [while they are specific to the GLE equation (6)], all the moments and correlation functions can be analytically calculated based on the Laplace transformation of the SDE [40] and response functions can be expanded as sums of exponentials through the residue theorems [41]. In addition, the recurrence methods of Ref. [42] are also available for this model.

3. Origin of the non-Markovian property and Markov embedding

We have seen that the non-Markovian property appears as the result of variable elimination through the example of Eq. (5). This shows that some non-Markovian systems with an exponential memory kernel can be converted back to a Markovian system by adding an auxiliary variable [30]. This procedure is called Markov embedding [43,44] and can be generalized to transform the GLE (6) into simultaneous Markovian SDEs when the kernel is given by a sum of exponentials:

$$K(t) = \sum_{k=1}^{K} \kappa_i e^{-t/\tau_i} \implies M \frac{d\hat{v}(t)}{dt} = \sum_{k=1}^{K} \hat{u}_k(t),$$

$$\frac{d\hat{u}_k(t)}{dt} = -\frac{\hat{u}_k(t)}{\tau_k} - \kappa_k \hat{v}(t) + \sqrt{\frac{2\kappa_k T}{\tau_k}} \hat{\xi}_k^{G}(t) \qquad (8)$$

with independent standard Gaussian noises $\hat{\xi}_k^{G}(t)$, satisfying $\langle \hat{\xi}^{G}(t)_k \rangle = 0$ and $\langle \hat{\xi}_k^{G}(t) \hat{\xi}_j^{G}(t') \rangle = \delta_{kj} \delta(t - t')$. Here we have taken the free potential case $U(\hat{x}) = 0$ for simplicity. We thus find that the exponential-sum memory case is Markovian by introducing the new system-variable set $\hat{\Gamma} := (\hat{v}, \hat{u}_1, \ldots, \hat{u}_K)$, while it was non-Markovian in the original variable representation $\hat{v}(t)$.

In this sense, whether a system is regarded as Markovian or non-Markovian crucially depends on which variable set is taken for the description of the system. This story is actually consistent with the projection operator formalism: While the original Hamiltonian dynamics obeys the Markovian dynamics (i.e., the time evolution of the phase-space distribution is given by the Liouville equation), the reduced dynamics of macroscopic variables obeys the non-Markovian Langevin dynamics due to integrating out irrelevant variables. Furthermore, a systematic method of Markov embedding [45] was proposed on the basis of the continued-fraction expansion [39].

4. Fokker-Planck descriptions of non-Markovian processes

Compared with Markovian stochastic processes, there are few mathematical techniques that can be applied to the Fokker-Plank (master) equations associated with general non-Markovian processes. One of the formal methods is to use the time-convolution Fokker-Planck equation for macroscopic variables $\boldsymbol{x}$: $\partial P_t(\boldsymbol{x})/\partial t = \int_0^t ds \mathcal{L}_{\text{GFP}}(\boldsymbol{x}, s) P_{t-s}(\boldsymbol{x})$, which was derived from the projection operator formalism [30]. Also, some specific class of non-Markovian SDEs can be formulated as a Fokker-Planck equation with time-dependent coefficients via the functional stochastic calculus [46]. While these equations are formally correct, they are not easy to exploit for practical calculations due to their genuine non-Markovian nature.

One of the most powerful approaches to tackle GLE is to use Markov embedding, thus making the system Markovian. This research direction was proposed by Ref. [45] and there are a variety of ways to select the auxiliary variables. By taking the selection in Eq. (8) according to Ref. [43], we obtain the complete Fokker-Planck equation for the GLE with the exponential-sum memory $K(t) = \sum_{k=1}^{K} \kappa_i e^{-t/\tau_k}$ in the absence of potential $U(\hat{x}) = 0$,

$$\frac{\partial P_t(\Gamma)}{\partial t} = \sum_{k=1}^{K} \left[-\frac{\partial}{\partial v} \frac{u_k}{M} + \frac{\partial}{\partial u_k} \left(\frac{u_k}{\tau_k} + \kappa_k v \right) + \frac{\kappa_k T}{\tau_k} \frac{\partial^2}{\partial u_k^2} \right] P_t(\Gamma) \qquad (9)$$

for the extended phase point $\Gamma := (v, u_1, \ldots, u_K)$ and the corresponding PDF $P_t(\Gamma)$.

5. Field description for an infinite number of auxiliary variables

There are two classes for Markov embedding: One requires a finite number of auxiliary variables [e.g., the GLE with a finite discrete sum of exponential terms to describe the memory (8)], and the other one requires an infinite number of auxiliary variables. The latter class is essentially similar to the classical field theory of stochastic processes, represented by stochastic partial differential equations (SPDE). Indeed, the continuous version of the Markov embedding (8) can be expressed in terms of SPDEs as follows. For the continuous decomposition

$$K(t) = \int_0^{\infty} dx \kappa(x) e^{-t/x}, \qquad (10)$$

we obtain an equivalent Markov embedding representation in the absence of the potential $U(\hat{x}) = 0$,

$$\frac{d\hat{v}(t)}{dt} = \frac{1}{M} \int_0^{\infty} dx u(t, x),$$

$$\frac{\partial \hat{u}(t, x)}{\partial t} = -\frac{\hat{u}(t, x)}{x} - \kappa(x)\hat{v}(t) + \sqrt{\frac{2\kappa(x)T}{x}} \hat{\xi}^{G}(x, t), \quad (11)$$

with the spatial white Gaussian noise term $\hat{\xi}^{G}(t, x)$ satisfying $\langle \hat{\xi}^{G}(t, x) \rangle = 0$ and $\langle \hat{\xi}^{G}(t, x) \hat{\xi}^{G}(t', x') \rangle = \delta(x - x')\delta(t - t')$. This is a simple equation in terms of the time derivative but it can be regarded as a SPDE, since $\hat{u}(t, x)$ is spatially distributed over the auxiliary field $x \in (0, \infty)$. Indeed, this interpretation enables us to apply the functional calculus historically developed for the analytical solution of SPDEs [31].

TABLE I. Summary table to compare the generalized Langevin equation and the Hawkes process. Here FDR stands for the fluctuation-dissipation relation.

Model	GLE (exponential memory)	GLE (general memory)	Hawkes process (general memory)
Character	Diffusive transport	Diffusive transport	Point process
Fundamental equation	Fokker-Planck equation (9)	Field Fokker-Planck equation (12)	Field master equation (40)
Typical systems	Near equilibrium	Near equilibrium	Out-of-equilibrium
Phenomena	Relaxation	Relaxation	Critical bursts
FDR	Yes	Yes	No
Non-Markovian representation	SDE (6)	SDE (6)	SDE (1)
Markovian representation	SDE (8)	SPDE (11)	SPDE (36)
# of auxiliary variables	Finite	Infinite (field description)	Infinite (field description)

In the domain of mathematics dealing with SPDEs, the Fokker-Planck description is based on the functional calculus for the field variable (called the functional Fokker-Planck equation in Ref. [31]). The corresponding field Fokker-Planck equation can be derived by replacing the discrete sum with the functional derivative:

$$\frac{\partial P_t[\Gamma]}{\partial t} = \int_0^\infty dx \left[-\frac{\partial}{\partial v} \frac{u(x)}{M} + \frac{\delta}{\delta u(x)} \left(\frac{u(x)}{x} + \kappa(x)v \right) \right.$$
$$\left. + \frac{\kappa(x)T}{x} \frac{\delta^2}{\delta u^2(x)} \right] P_t[\Gamma] \qquad (12)$$

for the extended phase point $\Gamma := (v, \{u(x)\}_{x\in(0,\infty)})$ and the corresponding probability functional $P_t[\Gamma]$.

We note that the mathematical precise definition of such Fokker-Planck description has not been established yet [31]. One can see a formal divergent term in the FP field equation, such as $[\delta/\delta u(x)]u(x) = \delta(0)$. This divergence is irrelevant for physical observables as shown in Sec. V A 2 since the SPDE is linear. However, general nonlinear SPDEs can have divergences even for observables (see the example of the nonlinear stochastic reaction-diffusion model in Ref. [31], Chap. 13.3.3). According to convention, the safest interpretation is that all the procedures are implicitly discrete and the continuous notation is regarded as a useful abbreviation of the discrete underlying model. By introducing the finite lattice constant dx for the auxiliary variable x, we have the variable transformations $u_k \to u(x_k)dx$ and $\kappa_k \to \kappa(x_k)dx$. The functional derivative is then introduced as the formal limit of $\delta F[\Gamma]/\delta u(x_k) := \lim_{dx\to 0}(\partial F[\Gamma]/\partial u_k)/dx$ (see Ref. [31], Chap. 13.1.1). In this sense, the derivation of the field Fokker-Planck equation (12) here follows this convention: We first confirm that the discrete Markov embedding (8) works well and then generalize it to its general continuous version (11).

6. Relation to the non-Markovian Hawkes processes

The above historical discussion of non-Markovian diffusive process provides a clear guideline for different classes of stochastic processes, including the Hawkes process as follows. The summary highlighting dissimilarities and similarities between the GLE and the Hawkes process is presented in Table I.

a. Dissimilarities. The Hawkes process is an example of non-Markovian point processes, triggering finite-size jumps along a sample trajectory. This is in contrast to the non-Markovian Langevin equation, which is based on infinitesimal Gaussian noise to describe the diffusive local transport, and thus does not include trajectory jumps. This difference should be reflected in the form of the time-evolution equation of the PDF. Indeed, the time evolution of the PDF for the Hawkes process will be shown to obey the master equation (integrod-ifferential equations), while that for the Langevin dynamics obeys the Fokker-Planck equations (second-order derivative equations).

Another dissimilarity comes from the fact that the Hawkes process is an out-of-equilibrium model typically describing a branching process requiring immigrants or background events to drive the whole sequence of events, while the Langevin equations are near-equilibrium models in the sense that initial distributions of the thermal baths are characterized by small perturbations from the Gibbs distribution. Note that the physical validity of the projection operator formalism is not guaranteed in general for out-of-equilibrium systems [30] characterized by non-Gibbs initial distributions. This conceptual difference is important because the FDR (7) is not necessarily assumed for the Hawkes processes. Indeed, the master equation for the Hawkes process does not satisfy the detailed balance condition [31] (or the time-reversal symmetry) as shown later.

We should stress that the word "nonequilibrium" is used here for systems driven by external forces, which leads to non-Gibbs initial distributions for thermal baths in contact with the system. While we did not review these cases in detail, there are various understandings of the FDR. In the historical context of stochastic processes, the FDR is defined as being closely associated with the time-reversal symmetry of the stochastic processes. In this sense, one can mathematically prove that the Hawkes processes is actually "out of equilibrium," in the sense that the corresponding master equation does not satisfy the symmetry condition in Gardiner's textbook [31]. We note that such out-of-equilibrium processes are physically reasonable in general nonequilibrium setups. Indeed, the shot noise process [31] and the Lévy flights dynamics [47] can be observed in out-of-equilibrium systems (e.g., see Refs. [48,49] for their statistical physics derivation from microscopic dynamics), while they do not satisfy the detailed balance condition. In the historical context of the projection operators, the near-equilibrium condition is defined such that the initial distribution for the noise space is characterized by a linear response around the Gibbs distribution. In the context of fluctuation theorems [50], the FDR is derived from the assumptions of (1) time-reversal symmetry of the microscopic

dynamics and (2) validity of the Gibbs distribution for the thermal bath in contact with the target system. The common understanding is that the noise source (i.e., the thermal bath in contact with the systems) is characterized by a distribution close to the Gibbs distribution. Therefore, there should not be any confusion when considering the Hawkes process for which the FDR is irrelevant.

b. Similarities. The GLE (6) can be mapped onto a Markovian process (11) by adding a sufficient number of auxiliary variables. As we show below, the same Markov embedding technique is available even for the non-Markovian Hawkes processes (1), by adding a sufficient number of auxiliary variables. Since the memory kernel can be a continuous sum of exponential kernels, the most general description of the Hawkes process should be based on an infinite number of auxiliary variables. As discussed above, such systems are typically described as a classical field theory driven by stochastic terms and thus the dynamics of the original Hawkes process can be finally mapped onto an SPDE (36) and the field master equations (40).

III. MASTER EQUATIONS

We now formulate the master equation for the model (1) and provide its asymptotic solution around the critical point.

A. Markov embedding: Introduction of auxiliary variables

As discussed in Sec. II C 3, non-Markovian properties in many stochastic systems often arise as a result of variable eliminations. This suggests that, reciprocally, it might be possible to map a non-Markovian system onto a Markovian one by adding auxiliary variables (i.e., Markov embedding). While the Hawkes process (1) is non-Markovian in its original representation only based on the intensity $\hat{v}(t)$, it can be also transformed onto a Markovian process by selecting an appropriate set of system variables. In this section, we formulating such a Markov embedding procedure to derive the corresponding master equations.

B. The single exponential kernel case

1. Mapping to Markovian dynamics

Before developing the general framework for arbitrary memory kernel $h(t)$, we consider the simplest case of an

exponential memory kernel:

$$h(t) = \frac{1}{\tau} e^{-t/\tau}, \tag{13}$$

satisfying the normalization $\int_0^\infty h(t)dt = 1$. The decay time τ quantifies how long an event can typically trigger events in the future. This special case is Markovian as discussed in Refs. [51,52], because the lack of memory of exponential distributions ensures that the number of events after time t defines a Markov process in continuous time [53].

As shown in the example (5), some non-Markovian processes can be mapped onto a Markovian stochastic system if the memory function is exponential. Here we show that, likewise, this single exponential case (13) can be mapped onto an SDE driven by a state-dependent Markovian Poisson noise. By decomposing the intensity as

$$\hat{z} := \hat{v} - v_0, \tag{14}$$

let us consider the Langevin dynamics

$$\frac{d\hat{z}}{dt} = -\frac{1}{\tau}\hat{z} + \frac{n}{\tau}\hat{\xi}_{\hat{v}}^{\mathrm{P}} \tag{15}$$

with a state-dependent Poisson noise $\hat{\xi}_{\hat{v}}^{\mathrm{P}}$ with intensity given by $\hat{v} = \hat{z} + v_0$ and initial condition $\hat{z}(0) = 0$. The introduction of $\hat{z}$ is similar to the trick proposed in Ref. [54] for an efficient estimation of the maximum likelihood of the Hawkes process. By expressing the state-dependent Poisson noise as

$$\hat{\xi}_{\hat{v}}^{\mathrm{P}}(t) = \sum_{i=1}^{\hat{N}(t)} \delta(t - \hat{t}_i), \tag{16}$$

we obtain the formal solution of equation (15)

$$\hat{v}(t) = v_0 + \hat{z}(t) = v_0 + \frac{n}{\tau} \int_0^t dt' e^{-(t-t')/\tau} \hat{\xi}_{\hat{v}}^{\mathrm{P}}(t')$$

$$= v_0 + n \sum_{i=0}^{\hat{N}(t)} h(t - \hat{t}_i). \tag{17}$$

This solution shows that the SDE (15) is equivalent to the Hawkes process (1). Equation (15) together with (16) is therefore a short-hand notation for

$$\hat{z}(t+dt) - \hat{z}(t) = \begin{cases} -\frac{1}{\tau}\hat{z}(t)dt & \text{(no jump during } [t, t+dt); \text{ probability} = 1 - \hat{v}(t)dt) \\ \frac{n}{\tau} & \text{(jump in } [t, t+dt); \text{ probability} = \hat{v}(t)dt) \end{cases} \tag{18}$$

for the probabilistic time evolution during $[t, t + dt)$ [see Fig. 2(a) for a schematic representation]. Note that the event probability explicitly depends on $\hat{v}(t)$, which reflects the endogenous nature of the Hawkes process. This is the first example of the Markov embedding of the Hawkes process. Note that this procedure corresponds to Eq. (8) for the case of the GLE with $K = 1$.

2. Master equation

By introducing Eq. (15) together with (16), we have transformed a non-Markovian point process into a Markovian SDE. This allows us to derive the corresponding master equation for the probability density function (PDF) $P_t(z)$ of the excess intensity $\hat{z}$ (14),

$$\frac{\partial P_t(z)}{\partial t} = \frac{1}{\tau}\frac{\partial}{\partial z}zP_t(z) + [(v_0 + z - n/\tau)P_t(z - n/\tau) - (v_0 + z)P_t(z)], \tag{19}$$
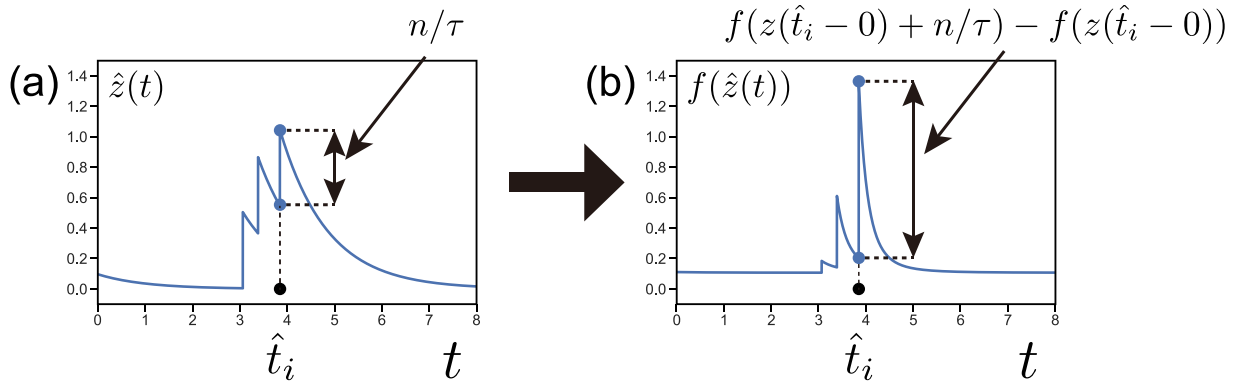
FIG. 2. (a) Schematic representation of a typical trajectory for $\hat{z}(t)$ defined by (14) and (15). A jump of size n/τ may occur in the time interval $[t, t + dt)$ with probability $\hat{v}(t)dt$. (b) Let us consider an arbitrary function $f(\hat{z})$. At the time $t = \hat{t}_i$ of the jump, there is a corresponding jump in the trajectory of $f(\hat{z})$, which is characterized by $df(\hat{z}(t)) := f(\hat{z}(\hat{t}_i - 0) + n/\tau) - f(\hat{z}(\hat{t}_i - 0))$. The plots shown here are based on a numerical simulation with the following parameters: $\tau = 1$, $n = 0.5$, $v_0 = 0.1$, and $f(z) = \exp[5(z - 1)] + 0.1$.

with the boundary condition

$$P_t(z)|_{z=0} = 0. \tag{20}$$

$P_t(z)dz$ is thus the probability that $\hat{z}(t)$ takes a value in the interval $\hat{z}(t) \in [z, z + dz]$ at time t. We note that this master equation after Markov embedding corresponds to the FP equation (9) for the case of the GLE with $K = 1$.

The master equation (19) is derived as follows. Let us consider an arbitrary function $f(\hat{z})$. Using (18), its time evolution during $[t, t + dt)$ is given by

$$f(\hat{z}(t + dt)) - f(\hat{z}(t)) = \begin{cases} -\frac{\hat{z}(t)}{\tau} \frac{\partial f(\hat{z}(t))}{\partial \hat{z}} dt & (\text{probability} = 1 - \hat{v}(t)dt) \\ f(\hat{z}(t) + n/\tau) - f(\hat{z}(t)) & (\text{probability} = \hat{v}(t)dt) \end{cases}, \tag{21}$$

as schematically illustrated in Fig. 2(b). By taking the ensemble average of both sides, we obtain

$$dt \int dz \frac{\partial P_t(z)}{\partial t} f(z) = \int dz P_t(z) \left[-\frac{z}{\tau} \frac{\partial f(z)}{\partial z} dt + (z + v_0)\{f(z + n/\tau) - f(z)\}dt \right],$$

$$\implies \int dz \frac{\partial P_t(z)}{\partial t} f(z) = \int dz f(z) \left[\frac{1}{\tau} \frac{\partial}{\partial z} z P_t(z) + \{(v_0 + z - n/\tau)P(z - n/\tau) - (v_0 + z)P(z)\} \right]. \tag{22}$$

This result (22) is obtained by (i) using the identity

$$\langle f(\hat{z}(t + dt)) - f(\hat{z}(t)) \rangle = \int dz [P_{t+dt}(z) - P_t(z)] f(z) = dt \int dz \frac{\partial P_t(z)}{\partial t} f(z), \tag{23}$$

(ii) by performing a partial integration of the first term in the right-hand side of Eq. (22) and (iii) by introducing the change of variable $z \to z - n/\tau$ for the second term. Since (22) is an identity holding for arbitrary $f(z)$, the integrants of the left-hand side and right-hand side must be equal for arbitrary $f(z)$, which yields the master equation (19).

Note that the above derivation of the master equation is not restricted to the exponential shape of the memory kernel. We are going to use the same derivation in the more complex examples discussed below. We also note that the master equation (19) does not satisfy the detailed balance condition of Ref. [31], which reflects the fact that the Hawkes process is a model for out-of-equilibrium systems.

C. Discrete sum of exponential kernels

1. Mapping to Markovian dynamics

The above formulation can be readily generalized to the case of a memory kernel expressed as a discrete sum of exponential functions:

$$h(t) = \frac{1}{n} \sum_{k=1}^{K} \frac{n_k}{\tau_k} e^{-t/\tau_k}. \tag{24}$$

In this case, each coefficient n_k quantifies the contribution of the kth exponential with memory length τ_k to the branching ratio $n = \sum_{k=1}^{K} n_k$, satisfying the normalization $\int_0^\infty h(t)dt = 1$. We note that this representation (24) is quite general, as it can approximate well the case of a power-law kernel with cutoff up to a constant [55].

Harris suggested the intuitive notion that it is possible to map this case to a Markovian dynamics if the state of the system at time t is made to include the list of the ages of all events [56]. The problem is that this conceptual approach is unworkable in practice due to the exorbitant size of the required information. By introducing an auxiliary age pyramid process, Ref. [57] identified some key components to add to the Hawkes process and its intensity to make the dynamics
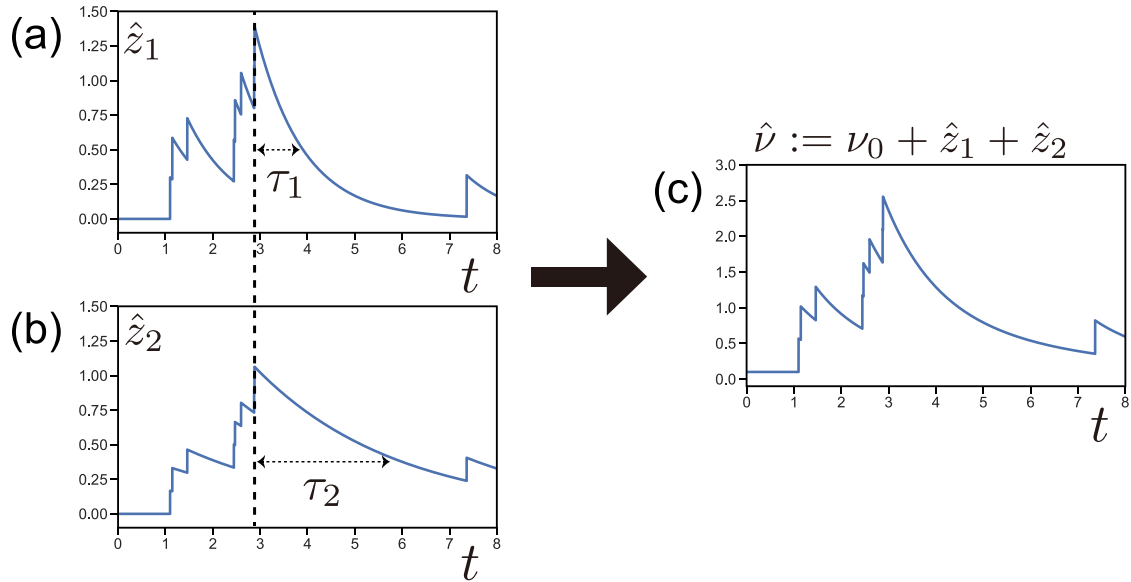
FIG. 3. Case where the memory kernel is the sum of two exponentials. Panels (a) and (b) show the schematic trajectories of the two excess intensities $\hat{z}_1$ and $\hat{z}_2$ and panel (c) that of the resulting total intensity $\hat{\nu} := \nu_0 + \hat{z}_1 + \hat{z}_2$. The parameters are $K = 2$, $\tau_1 = 1$, $n_1 = 0.3$, $\tau_2 = 3$, $n_2 = 0.5$ (and thus $n = 0.8$), and $\nu_0 = 0.1$.

Markovian. Here, in order to map model (1) onto a Markovian stochastic process, we propose a more straightforward Markov embedding approach, which generalized the previous case of a single exponential memory function. We decompose the intensity into a sum of K excess intensities $\{z_k\}_{k=1}^K$ as follows:

$$\hat{\nu}(t) = \nu_0 + \sum_{k=1}^K \hat{z}_k(t) . \tag{25}$$

Each excess intensity $\hat{z}_k$ is the solution of a Langevin equation driven by a state-dependent Markovian Poisson shot noise

$$\frac{d\hat{z}_k}{dt} = -\frac{\hat{z}_k}{\tau_k} + \frac{n_k}{\tau_k} \hat{\xi}_{\hat{\nu}}^{\mathrm{P}} . \tag{26}$$

Note that the same state-dependent Poisson noise $\hat{\xi}_{\hat{\nu}}^{\mathrm{P}}(t)$ defined by expression (16) acts on the Langevin equation for each excess intensity $\{\hat{z}_k\}_{k=1,\dots,K}$. In other words, each shock event impacts simultaneous the trajectories for all excess intensities $\{\hat{z}_k\}_{k=1,\dots,K}$ [see the vertical broken line in Figs. 3(a) and 3(b) and the resulting trajectory of $\hat{\nu}(t)$ in Fig. 3(c)]. Note that this Markov embedding corresponds to Eq. (8) for the case of the GLE.

2. Master equation

As the set of SDEs for $\hat{z} := (\hat{z}_1, \hat{z}_2, \dots, \hat{z}_K)^{\mathrm{T}}$ are standard Markovian stochastic processes, we obtain the corresponding master equation:

$$\frac{\partial P_t(z)}{\partial t} = \sum_{k=1}^K \frac{\partial}{\partial z_k} \frac{z_k}{\tau_k} P_t(z) + \left\{ \left[\nu_0 + \sum_{k=1}^K (z_k - n_k/\tau_k) \right] P_t(z - h) \right.$$

$$\left. - \left[\nu_0 + \sum_{k=1}^K z_k \right] P_t(z) \right\}. \tag{27}$$

The jump-size vector is given by $h := (n_1/\tau_1, n_2/\tau_2, \dots, n_K/\tau_K)^{\mathrm{T}}$. The PDF $P_t(z)$ obeys the following boundary condition,

$$P_t(z)|_{z \in \partial \mathbf{R}_+^K} = 0 , \tag{28}$$

on the boundary $\partial \mathbf{R}_+^K := \{z | z_k = 0 \text{ for some } k\}$. Equation (27) can be derived following the procedure used for the single exponential case that led us to the master equation (19) (see Appendix A 1 for an explicit derivation). Note that this master equation corresponds to the FP equation (9) for the case of the GLE.

3. Laplace representation of the master equation

The master equation (27) takes a simplified form under the Laplace representation,

$$\tilde{P}_t(s) := \mathcal{L}_K[P_t(z); s] , \tag{29}$$

where the Laplace transformation in the K-dimensional space is defined by

$$\mathcal{L}_K[f(z); s] := \int_0^\infty dz e^{-s \cdot z} f(z) \tag{30}$$

with volume element $dz := \prod_{k=1}^K dz_k$. The wave vector $s := (s_1, \dots, s_K)^{\mathrm{T}}$ is the conjugate of the excess intensity vector $z := (z_1, \dots, z_K)^{\mathrm{T}}$.

The Laplace representation of the master equation (27) is then given by

$$\frac{\partial \tilde{P}_t(s)}{\partial t} = -\sum_{k=1}^K \frac{s_k}{\tau_k} \frac{\partial \tilde{P}_t(s)}{\partial s_k} + (e^{-h \cdot s} - 1) \left(\nu_0 - \sum_{k=1}^K \frac{\partial}{\partial s_k} \right) \tilde{P}_t(s). \tag{31}$$

Then, the Laplace representation (29) of $P_t(z)$, which is the solution of (31), allows us to obtain the Laplace representation

$\tilde{Q}_t(s)$ of the intensity PDF $P_t(\nu)$ according to

$$\tilde{Q}_t(s) := \mathcal{L}_1[P_t(\nu); s] = \langle e^{-s(\nu_0 + \sum_{k=1}^K \hat{z}_k)} \rangle$$
$$= e^{-\nu_0 s} \tilde{P}_t(\boldsymbol{s} = (s, s, \dots,)^{\mathrm{T}}). \quad (32)$$

D. General kernels

1. Mapping to Markovian dynamics

The above formulation can be generalized to general forms of the memory kernel. Let us decompose the kernel as a continuous superposition of exponential kernels,

$$h(t) = \frac{1}{n} \int_0^\infty \frac{n(x)}{x} e^{-t/x} dx, \quad n = \int_0^\infty n(x) dx, \quad (33)$$

where we have introduced the set of continuous auxiliary variables $x \in \boldsymbol{R}_+$. This decomposition satisfies the normalization condition $\int_0^\infty h(t) dt = 1$. Here we use the notation x for these auxiliary variables to emphasize the formal connection with the usual field theory of classical stochastic (or quantum) systems. Indeed, the observables will be defined over the "field" $x \in \boldsymbol{R}_+$. The function $n(x)$ quantifies the contribution of the xth exponential with memory length x to the branching ratio. We can then interpret $n(x)/n$ as a normalized distribution of timescales present in the memory kernel of the Hawkes process. As we show below, an important condition for solvability will be the existence of its first-order moment

$$\frac{\alpha}{n} := \langle \tau \rangle := \int_0^\infty x \frac{n(x)}{n} dx < \infty. \quad (34)$$

This condition (34) means that $n(x)$ should decay faster than $1/x^2$ at large x's. Hence, the representation (33) implies that the memory kernel has to decay at large times faster than $1/t^2$. This covers situations where the variance of the timescales embedded in the memory kernel diverges. But this excluded the cases $h(t) \sim 1/t^{1+\theta}$ with $0 < \theta < 1$ that are relevant to the Omori law for earthquakes [16,17] and to the response to social shocks [20]. This case $0 < \theta < 1$ for which α diverges needs to be treated separately and this is beyond the content of the present work.

We then decompose the intensity of the Hawkes process as a continuous sum of excess intensities $\hat{z}(t, x)$

$$\hat{\nu}(t) = \nu_0 + \int_0^\infty dx \hat{z}(t, x). \quad (35)$$

Each excess intensities $\hat{z}(t, x)$ is the solution of the following dynamical equation,

$$\frac{\partial \hat{z}(t, x)}{\partial t} = -\frac{\hat{z}(t, x)}{x} + \hat{\eta}(t, x), \quad \hat{\eta}(t, x) := \frac{n(x)}{x} \hat{\xi}_{\hat{\nu}}^{\mathrm{P}}, \quad (36)$$

where, as for the previous case of a discrete sum of exponentials, the same state-dependent Poisson noise $\hat{\xi}_{\hat{\nu}}^{\mathrm{P}}(t)$ defined by expression (16) acts on the Langevin equation for each excess intensity $\hat{z}(t, x)$. While Eq. (36) is a simple equation only related to the time derivative, this equation can be regarded as an SPDE since $\hat{z}(t, x)$ is distributed over the auxiliary variable

field $x \in (0, \infty)$. Furthermore, this interpretation allows us to use functional calculus, historically developed for the solution of SPDEs (such as the stochastic reaction-diffusion equations [31]). Note that this SPDE corresponds to Eq. (11) for the case of the GLE.

The set of SPDEs (36) expresses the fact that the continuous field of excess intensity $\{\hat{z}(t, x)\}_{x \in \boldsymbol{R}_+}$ tends to relax to zero, but they are intermittently simultaneously shocked by the shared shot noise term $\hat{\xi}_{\hat{\nu}}^{\mathrm{P}}$, with an x-dependent jump size $n(x)/x$. This is in contrast to the SPDE representation (11) for the GLE (6), where the noise term at x has no correlation with that at another point x': $\langle \hat{\xi}^{\mathrm{G}}(t, x) \hat{\xi}^{\mathrm{G}}(t, x') \rangle = 0$ for $x \neq x'$. In a more conventional expression, we can rewrite the long-range nature of the spatial correlation in $\hat{\eta}(t, x)$ as

$$\langle \hat{\eta}(t, x) \hat{\eta}(t, x') \rangle_{ss} = \mathcal{K}(x, x') \delta(t - t'),$$
$$\mathcal{K}(x, x') = \frac{n(x)n(x')}{xx'} \langle \hat{\nu} \rangle_{ss}. \quad (37)$$

If $K(x, x')$ was a δ function, the SPDE could be regarded as a noninteracting infinite variable systems. This long-range nature means that all the excess intensity $\hat{z}(t, x)$ at different points are strongly correlated through this noise term, even though the SPDE (36) has no spatial derivatives.

2. Field master equation

The master equation corresponding to the SDE (36) can be derived by following the same procedure presented for the simple exponential case and for the discrete sum of exponentials. There is, however, a technical difference since the state of the system is now specified by the continuous field variable $\{\hat{z}(t, x)\}_{x \in \boldsymbol{R}_+}$. Thus, the probability density function is replaced with the probability density functional $P[\{\hat{z}(t, x) = z(x)\}_{x \in \boldsymbol{R}_+}] = P_t[\{z(x)\}_{x \in \boldsymbol{R}_+}]$. In other words, the probability that the system state is in the state specified by $\{z(x)\}_{x \in \boldsymbol{R}_+}$ at time t is characterized by $P_t[\{z(x)\}_{x \in \boldsymbol{R}_+}] \mathcal{D}z$ with functional integral volume element $\mathcal{D}z$.

We use the notational convention that any mapping with square bracket $A[\{f(x)\}_{x \in \boldsymbol{R}_+}]$ indicates that the map A is a functional of $\{f(x)\}_{x \in \boldsymbol{R}_+}$. In addition, we sometimes abbreviate the functional $P_t[\{z(x)\}_{x \in \boldsymbol{R}_+}]$ by $P_t[z] := P_t[\{z(x)\}_{x \in \boldsymbol{R}_+}]$ for the sake of brevity.

The presence of a continuous field variable leads to several technical issues, such as in the correct application of the Laplace transform. The functional Laplace transformation $\mathcal{L}_{\mathrm{path}}$ of an arbitrary functional $f[z]$ is defined by a functional integration (i.e., a path integral):

$$\mathcal{L}_{\mathrm{path}}[f[z]; s] := \int \mathcal{D}z e^{-\int_0^\infty dx s(x) z(x)} f[z]. \quad (38)$$

This allows us to define the Laplace representation of the probability density functional by

$$\tilde{P}_t[s] := \mathcal{L}_{\mathrm{path}}[P_t[z]; s] \quad (39)$$

for an arbitrary nonnegative function $\{s(\tau)\}_{\tau \in \boldsymbol{R}_+}$.

As the natural extension of Eq. (27), the master equation for the probability density functional is given by

$$\frac{\partial P_t[z]}{\partial t} = \int_0^\infty dx \frac{\delta}{\delta z(x)} \frac{z(x)}{x} P_t[z] + \left\{ \left[\nu_0 + \int_0^\infty (z - n/x) dx \right] P_t[z - n/x] - \left[\nu_0 + \int_0^\infty z dx \right] P_t[z] \right\} \quad (40)$$

with the boundary condition

$$P_t[z]|_{z \in \partial \boldsymbol{R}_+^\infty} = 0, \qquad (41)$$

where the boundary of the function space $\partial \boldsymbol{R}_+^\infty := \{z | z(x) = 0$ for some $x \in [0, \infty)\}$. This field master equation after Markov embedding corresponds to the FP field equation (12) for the GLE.

Interpretation. As discussed in Sec. II C 5, one of the safest interpretations is to regard this functional description as the formal continuous limit from the discrete description in Sec. III C. By introducing a finite lattice interval $dx > 0$ and rewriting $\tau_k \to x_k$, $\hat{z}_k(t) \to \hat{z}(t, x_k)dx$, and $n_k \to n(x_k)dx$, Eqs. (24)–(26) can be rewritten as

$$h(t) = \frac{1}{n} \sum_{k=1}^{K} \frac{n(x_k)}{x_k} e^{-t/x_k} dx, \quad \hat{v}(t) = v_0 + \sum_{k=1}^{K} \hat{z}(t, x_k)dx, \quad \frac{\partial z(t, x_k)}{\partial t} = -\frac{\hat{z}(t, x_k)}{x_k} + \frac{n(x_k)}{x_k} \hat{\xi}_{\hat{v}}^P(t). \qquad (42)$$

The master equation (27) is rewritten as

$$\frac{\partial P_t(z)}{\partial t} = \sum_{k=1}^{K} dx \left[\frac{1}{dx} \frac{\partial}{\partial z(x_k)} \right] \frac{z(x_k)}{x_k} P_t(z) + \left\{ v_0 + \sum_{k=1}^{K} (z(x_k) - n(x_k)/x_k) dx \right\} P_t(z - h) - \left\{ v_0 + \sum_{k=1}^{K} z(x_k) dx \right\} P_t(z). \qquad (43)$$

According to the convention in Ref. [31], we take a formal limit $K \to \infty$ and $dx \to 0$ to apply the formal replacement

$$\int_0^\infty dx[...] := \lim_{dx \to 0} \sum_{k=1}^{K} dx[...], \quad \frac{\delta}{\delta z(x)} := \lim_{dx \to 0} \frac{1}{dx} \frac{\partial}{\partial z(x_k)}. \qquad (44)$$

The master equation (40) for the field variables $\{\hat{z}(t, x)\}_{x \in (0, \infty)}$ is thus derived. While this derivation is based on a formal limit of a discrete description, Eq. (40) can be also derived by direct continuous operations based on the functional Taylor expansions (see Appendix A 2 for the detailed derivation based on functional calculus).

We remark that the SPDE (36) is linear and serious divergence problems did not appear at least for our main results on physical observables. In addition, our final results have been confirmed to be robust for both discrete and continuous cases as shown later. This strategy is consistent with the safe prescription suggested in Ref. [31]: In the beginning, the functional descriptions for field variables should be based on a discrete formulation. The continuous description should be introduced afterward as a formal limit of zero-lattice intervals. In this sense, our analysis has successfully avoided the delicate divergence problems on general SPDEs.

3. Laplace representation of the master equation

In the functional Laplace representation (39), the master equation (40) takes the following simple first-order functional differential equation:

$$\frac{\partial \tilde{P}_t[s]}{\partial t} = v_0 \left(e^{-\int_0^\infty dx' s(x') n(x')/x'} - 1 \right) \tilde{P}_t[s]$$

$$- \int_0^\infty dx \left(e^{-\int_0^\infty dx' s(x') n(x')/x'} - 1 + \frac{s(x)}{x} \right) \frac{\delta \tilde{P}_t[s]}{\delta s(x)}. \qquad (45)$$

4. General formulation

All the above forms of the memory kernel can be unified by remarking that the variable transformation (33) is equivalent

to a Laplace transform, since it can be rewritten as

$$h(t) = \frac{1}{n} \int_0^\infty \frac{1}{s} n \left(\frac{1}{s} \right) e^{-st} ds = \frac{1}{n} \mathcal{L}_1 \left[\frac{1}{s} n \left(\frac{1}{s} \right); t \right]$$

$$\iff \frac{n(x)}{n} = \frac{1}{x} \mathcal{L}_1^{-1}[h(t); s] \Big|_{s = 1/x}. \qquad (46)$$

This allows us to reformulate the several examples discussed above in a unified way presented in Table II.

IV. SOLUTION

A. Main results

In Sec. III, we have derived the master equations and their Laplace representations for the Hawkes processes with arbitrary memory kernels. Remarkably, the Laplace representations are first-order partial (functional) differential equations. Because first-order partial (functional) differential equations can be formally solved by the method of characteristics (see Appendix C for a brief review), various analytical properties of the Hawkes process can be studied in detail.

In this section, we present novel properties of the Hawkes process unearthed from the solution of the master equations by the method of characteristics. In particular, we focus on the behavior of the PDF of the steady-state intensity near the

TABLE II. Examples of various memory kernel $h(t)$ and corresponding $n(x)$ defined in expression (33).

Case	$h(t)$	$n(x)$
Single exponential kernel	$\frac{1}{\tau_1} e^{-t/\tau_1}$	$n_1 \delta(x - \tau_1)$
Discrete superposition of exponential kernel	$\frac{1}{n} \sum_{k=1}^{K} \frac{n_k}{\tau_k} e^{-t/\tau_k}$	$\sum_{k=1}^{K} n_k \delta(x - \tau_k)$
Power-law kernel ($\beta \geqslant 0$)	$\frac{1}{\tau^*} \frac{\beta}{(1 + t/\tau^*)^{\beta+1}}$	$\frac{n}{x} \left(\frac{\tau^*}{x} \right)^\beta \frac{e^{-\tau^*/x}}{\Gamma(\beta)}$

critical point $n = 1$. Under the condition of the existence of the first-order moment (34), an asymptotic analysis of the master equations shows that the PDF $P_{ss}(\nu) := \lim_{t \to \infty} P_t(\nu)$ exhibits a power-law behavior with a nonuniversal exponent:

$$P_{ss}(\nu) \propto \frac{1}{\nu^{1-2\nu_0\alpha}}, \quad \text{with } \alpha = n\langle\tau\rangle \tag{47}$$

for large ν, up to an exponential truncation, which is pushed toward $\nu \to \infty$ as $n \to 1$. As the tail exponent is smaller than 1, the steady-state PDF $P_{ss}(\nu)$ is not renormalizable without the exponential cutoff. However, the characteristic scale of the exponential tail diverges as the system approaches the critical point $n = 1$, and the power-law tail (47) can be observed over many orders of magnitude of the intensity for near-critical systems, as we illustrate below.

The parameter $\alpha = n\langle\tau\rangle$ entering in the expression of tail exponent in expression (47) has been defined by expression (34). Since ν_0 is the background intensity of the Hawkes intensity as defined in (1), the exponent of $P_{ss}(\nu)$ depends on $\nu_0\alpha = n\nu_0\langle\tau\rangle$, which is n times the average number of background events (or immigrants) occurring during a time equal to the average time scale $\langle\tau\rangle$ of the memory kernel. Thus, the larger the memory $\langle\tau\rangle$, the larger is the background intensity ν_0, and the larger the branching ratio n, the smaller is the exponent $1 - 2\nu_0\alpha$. Note that $1 - 2\nu_0\alpha$ can even turn negative for $\nu_0\alpha > 1/2$, which corresponds to a nonmonotonous PDF $P_{ss}(\nu)$, which first grows according to the power law (47) before decaying exponentially at very large ν's.

In simple terms, the PDF (47) describes the distribution of the number νdt of events in the limit of infinitely small time windows $[t, t + dt]$. We should contrast this limit to the other previously studied limit of infinitely large and finite but very large time windows. Standard results of branching processes (of which the Hawkes model is a subset) give the total number of events generated by a given triggering event (see Ref. [21] for a detailed derivation). In Eq. (1), this corresponds to counting all the events over an infinitely large time window that are triggered by a single source event $\nu_0 = \delta(t)$ occurring at the origin of time. Reference [22] demonstrates the distribution of "seismic rates" in the limit of large time windows which, in our current formulation, corresponds to the distribution of $N(t) := \int_t^{t+T} \nu(\tau)d\tau$, in the limit of large T's. The corresponding probability density distributions are totally different from (47), which corresponds to the other limit $T \to 0$.

We derive our main result (47) first for the single exponential form of the memory kernel, then for the discrete sum of exponentials, and then for the general case.

B. Single exponential kernel

As the first example, we focus on the single exponential kernel (13). While this special case is analytically tractable without the need to refer to the master equation approach [52], we nevertheless derive its exact solution via the master equation approach, because the methodology will be readily generalized to the more complex cases.

1. Steady-state solution

Let us first study the steady solution of the PDF $P_{ss}(\nu)$. By setting $K = 1$ in Eq. (31), we obtain the expression of the Laplace transform of the steady state $\tilde{P}_{ss}(s) := \int_0^\infty d\nu e^{-s\nu} P_{ss}(z)$ of the master equation (19) in the form of a first-order ordinary differential equation

$$\left(e^{-ns/\tau} - 1 + \frac{s}{\tau}\right)\frac{d\tilde{P}_{ss}(s)}{ds} = \nu_0(e^{-ns/\tau} - 1)\tilde{P}_{ss}(s). \tag{48}$$

By solving this equation, we obtain the exact steady solution below the critical point $n < 1$,

$$\log\tilde{Q}_{ss}(s) = -s\nu_0 + \log\tilde{P}_{ss}(s) = -\frac{\nu_0}{\tau}\int_0^s \frac{s\,ds}{e^{-ns/\tau} - 1 + s/\tau} \tag{49}$$

with the renormalization condition

$$\int_0^\infty P_{ss}(\nu) = \tilde{Q}_{ss}(s = 0) = 1. \tag{50}$$

a. Near the critical point. Let us evaluate the asymptotic behavior of $\tilde{Q}_{ss}(s)$ for large ν by assuming that the system is in the near-critical state, such that

$$\epsilon := 1 - n \ll 1. \tag{51}$$

By performing an expansion in the small parameter ϵ, we obtain an asymptotic formula for small s (large ν),

$$\log\tilde{Q}_{ss}(s) \simeq -\frac{\nu_0}{\tau}\int_0^s \frac{ds}{\epsilon/\tau + s/2\tau^2} = -2\nu_0\tau\log\left(1 + \frac{s}{2\tau\epsilon}\right), \tag{52}$$

which implies a power-law behavior with a nonuniversal exponent, up to an exponential truncation:

$$P_{ss}(\nu) \propto \nu^{-1+2\nu_0\tau}\,e^{-2\tau\epsilon\nu}, \tag{53}$$

for large ν. This is a special case of expression (47) obtained for $n \to 1$ and $\langle\tau\rangle = \tau$.

It is remarkable that the power-law exponent is less than 1, and thus the PDF is not renormalizable without the exponential truncation. This means that the power-law scaling actually corresponds to an intermediate asymptotics of the PDF, according to the classification of Barenblatt [58]. In addition, while this scaling can be regarded as a heavy "tail" for $2\nu_0\tau < 1$, the exponent can be negative when $2\nu_0\tau > 1$ (i.e., the PDF is a power-law increasing function until the exponential tail takes over and ensure the normalization of the PDF).

The characteristic scale of the exponential truncation is defined by

$$\nu_{\text{cut}} := \frac{1}{2\tau\epsilon} = \frac{1}{2\tau(1-n)}, \tag{54}$$

which diverges as the system approaches to the critical condition $n = 1$. This means that (i) if the system is in a near-critical state $\epsilon \ll 1$ and (ii) the background intensity is sufficiently small $\nu_0 < 1/(2\tau)$, one can actually observe the power-law intermediate asymptotics for a wide range $\nu \ll \nu_{\text{cut}} = O(\epsilon^{-1})$, up to the exponential truncation.
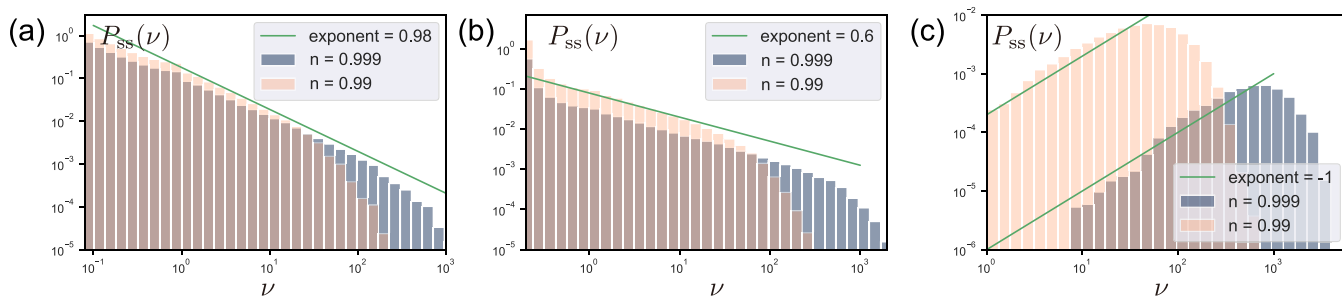
FIG. 4. Numerical evaluation of the steady state PDF of the intensity $\hat{\nu}$ for the following parameter sets near the critical point: $n = 0.999$ (blue bars) and $n = 0.99$ (red bars). The theoretical power law is shown by the green straight line. (a) Background intensity $\nu_0 = 0.01$, relaxation time $\tau = 1$, leading to the power-law exponent 0.98. (b) $\nu_0 = 0.2$, $\tau = 1$, leading to the power-law exponent 0.6. (c) $\nu_0 = 1.0$, $\tau = 1$, leading to the negative (i.e., growing) power-law exponent -1.0. For all simulations, the sampling time interval and total sampling time are $dt = 0.001$ and $T_{\text{tot}} = 10000$ with the initial condition $\hat{z}(0) = 0$. The initial 10% of the sample trajectories were discarded from the statistics.

b. Numerical verification. We now present numerical confirmations of our theoretical prediction, in particular for the intermediate asymptotics as shown in Fig. 4. There is a well-developed literature on the numerical simulation of Hawkes process [59,60]. We have used an established simulation package for PYTHON called "tick" (version 0.6.0.0) for 32-thread parallel computing. The total simulation time was 10^4 and the sampling time interval was 0.001. We note that the initial 10% of the sampled trajectories were discorded for initialization.

For small background intensity $\nu_0 \ll 1/\tau$, we obtain an approximate universal exponent -1. For $2\nu_0\tau < 1$, we observe a decaying power law of exponent $1 - 2\nu_0\tau$, while we observe a growing power law for $2\nu_0\tau > 1$. The power-law intermediate asymptotics is truncated by the exponential function, as predicted and also discussed in Ref. [54] (albeit with the error of missing the 1 in the exponent and thus failing to describe the intermediate asymptotics), ensuring the normalization of the PDF of the Hawkes intensities.

2. Time-dependent solution

We now present the exact solution of the time-dependent master equation. In the Laplace representation, the dynamics of the PDF of the intensities is given by the following first-order PDE,

$$\frac{\partial \tilde{P}_t(s)}{\partial t} + \left(e^{-ns/\tau} - 1 + \frac{s}{\tau}\right)\frac{\partial \tilde{P}_t(s)}{\partial s} = \nu_0(e^{-ns/\tau} - 1)\tilde{P}_t(s). \tag{55}$$

This equation can be solved by the method of characteristics (see Appendix C for a brief review). The corresponding Lagrange-Charpit equations are given by

$$\frac{ds}{dt} = e^{-ns/\tau} - 1 + \frac{s}{\tau},$$

$$\frac{d\Phi}{dt} = \nu_o(e^{-ns/\tau} - 1), \quad \text{with} \quad \Phi := \log \tilde{P}. \tag{56}$$

These equations can be solved explicitly,

$$t = \mathcal{F}(s) + C_1, \quad \Phi = \nu_0 s - \frac{\nu_0}{\tau}\int_0^s \frac{s'ds'}{e^{-ns'/\tau} - 1 + s'/\tau} + C_2 \tag{57}$$

with

$$\mathcal{F}(s) := \int_{s_0}^s \frac{ds'}{e^{-ns'/\tau} - 1 + s'/\tau} + C_1. \tag{58}$$

C_1 and C_2 are constants of integration and s_0 is a positive constant chosen to satisfy several convenient properties discussed below.

a. Summary of the properties of $\mathcal{F}$. We present several analytical properties of $\mathcal{F}(s)$ (see Appendix D for their proof):

(α1) $\mathcal{F}(s)$ is a monotonically increasing function by choosing $s_0 > 0$ appropriately.

(α2) The inverse function $\mathcal{F}^{-1}(s)$ can be defined uniquely. In addition, for the subcritical case $n < 1$, the following properties hold true: (α3) s_0 can be set to any positive value.

(α4) $\lim_{s\to+0} \mathcal{F}(s) = -\infty$.

(α5) $\lim_{s\to\infty} \mathcal{F}(s) = +\infty$.

(α6) $\mathcal{F}(s)$ can take all real values: $\mathcal{F}(s) \in (-\infty, \infty)$ for $s > 0$.

b. Regularization of $\mathcal{F}(s)$. In the following, we assume the subcritical condition $n < 1$. It is useful to decompose $\mathcal{F}(s)$ into regular and singular parts:

$$\mathcal{F}(s) = \underbrace{\int_{s_0}^s \frac{ds}{e^{-ns/\tau} - 1 + s/\tau} - \int_{s_0}^s \frac{ds}{(1-n)s/\tau} + \frac{\tau}{1-n}\log\frac{s}{s_0}}_{\text{totally zero as an identity}} = \underbrace{\frac{\tau}{1-n}\int_{s_0}^s ds \frac{1 - e^{-ns/\tau} - ns/\tau}{s(e^{-ns/\tau} - 1 + s/\tau)}}_{\text{regular part}} + \underbrace{\frac{\tau}{1-n}\log\frac{s}{s_0}}_{\text{singular part}}, \tag{59}$$

where the regular part is well defined even for $s_0 \to 0$. This expression is useful since the divergent factor in $\mathcal{F}(s)$ can be renormalized into the integral constant C_1, such that

$$C_1 - \frac{\tau}{1-n} \log s_0 \to C_1. \tag{60}$$

We then take the formal limit $s_0 \to 0$ and use the following regularized expression for the subcritical condition $n < 1$,

$$\mathcal{F}(s) = \mathcal{F}_{\mathrm{R}}(s) + \frac{\tau}{1-n} \log s,$$

$$\mathcal{F}_{\mathrm{R}}(s) := \frac{\tau}{1-n} \int_0^s ds \frac{1 - e^{-ns/\tau} - ns/\tau}{s(e^{-ns/\tau} - 1 + s/\tau)}, \tag{61}$$

where the s_0 dependence is removed as the result of the renormalization. The regular part has no singularity at $s \simeq 0$ and reads $\mathcal{F}_{\mathrm{R}}(s) \simeq -n^2 s/\{2\tau(1-n)\}$.

c. Explicit solution. Building on the above, we now provide the solution of the master equation (55). According to the method of characteristics (see Appendix C for a brief review), the general solution is given by

$$C_2 = \mathcal{H}(C_1) \tag{62}$$

with an arbitrary function $\mathcal{H}(\cdot)$. The time-dependent solution $\log \tilde{Q}_t(s) = \log \tilde{P}_t(s) - v_0 s$ is thus given by

$$\log \tilde{Q}_t(s) = -\frac{v_0}{\tau} \int_0^s \frac{sds}{e^{-ns/\tau} - 1 + s/\tau} + \mathcal{H}(t - \mathcal{F}(s)). \tag{63}$$

The function $\mathcal{H}(\cdot)$ is determined by the initial condition.

Let us assume that the initial PDF and its Laplace representation are given by $P_{t=0}(v)$ and $\tilde{P}_{t=0}(s)$, respectively. Then, we obtain

$$\mathcal{H}(-\mathcal{F}(s)) = \log \tilde{P}_{t=0}(s) + \frac{v_0}{\tau} \int_0^s \frac{sds}{e^{-ns/\tau} - 1 + s/\tau} \tag{64}$$

or equivalently,

$$\mathcal{H}(x) = \log \tilde{P}_{t=0}(S(x)) + \frac{v_0}{\tau} \int_0^{S(x)} \frac{sds}{e^{-ns/\tau} - 1 + s/\tau},$$

$$S(x) = \mathcal{F}^{-1}(-x). \tag{65}$$

Note that the time-dependent solution (63) is consistent with the steady solution (49)

$$\lim_{t \to \infty} \tilde{Q}_t(s) = \tilde{Q}_{ss}(s), \tag{66}$$

since $\lim_{x \to +\infty} \mathcal{H}(x) = 0$ (see Appendix E 1 for the proof). We also note that, from the time-dependent solution (63), we can derive the dynamics of the intensity $\hat{v}(t)$ for finite t as

$$\langle \hat{v}(t) \rangle = v_{\mathrm{ini}} e^{-(1-n)t/\tau} + v_0 \frac{1 - e^{-nt/\tau}}{1-n} \tag{67}$$

with the initial condition $\hat{v}(0) = v_{\mathrm{ini}}$ (see Appendix E 2 for the derivation). This expression (67) shows that the mean intensity converges at long times $t \to +\infty$ to $\langle \hat{v}(t) \rangle \to v_0/(1-n)$, which is a well-known result [8,25]. Expression (67) also shows that an initial impulse decays exponentially with a renormalised time decay $\tau/(1-n)$, which is also consistent with previous reports [18]. This diverging timescale $\tau/(1-n)$, as $n \to 1$, reflects the occurrence of all the generations of triggered events that renormalize the "bare" memory function into a "dressed" memory kernel with much longer memory.

d. Asymptotic relaxation dynamics for large t. The time-dependent asymptotic solution is given for large t by

$$\log \tilde{P}_t(s) \simeq -\frac{v_0}{\tau} \int_{S(t,s)}^s \frac{sds}{e^{-ns/\tau} - 1 + s/\tau} + \log \tilde{P}_{t=0}(S(t,s)),$$

$$S(t,s) = s \exp\left[-\frac{1-n}{\tau}(t - \mathcal{F}_{\mathrm{R}}(s)) \right], \tag{68}$$

assuming that $t \gg \mathcal{F}(s)$ for a given s. As a corollary of this formula, an asymptotic prediction for the distribution conditional on the initial intensity is given by the following formula,

$$P(v,t|v_{\mathrm{ini}}, t = 0) = \mathcal{L}_1^{-1}[\tilde{P}_t(s); t], \quad \log \tilde{P}_{t=0}(s) = -v_{\mathrm{ini}} s, \tag{69}$$

for large t. Note that asymptotic convergence of these formulas is not uniform in terms of s; indeed, the convergence of the Laplace representation for large s is slower than that for small s.

3. Another derivation of the power law exponent: Linear stability analysis of the Lagrange-Charpit equation

We have presented both the steady-state and time-dependent solutions of the master equations, based on exact or asymptotic methods. While these formulations are already clear, here we revisit the power-law behavior (53) of the steady-state PDF and present another derivation based on the linear stability analysis of the Lagrange-Charpit equation, which has the advantage of being generalizable to memory kernels defined as superposition of exponential functions. Indeed, while the derivation based on the exact solution (49) is clear and powerful, it is not easy to extend this kind of calculation to general cases, such as superposition of exponential kernels. In contrast, the derivation that we now present can be extended to arbitrary forms of the memory kernel of the Hawkes processes, as will be shown later. Moreover, we have found additional distinct derivations of (53) and we refer the interested reader to Appendix B.

While the steady-state master equation (48) is a ordinary differential equation which can be solved exactly, let us consider its corresponding Lagrange-Charpit equations,

$$\frac{ds}{dl} = -e^{-ns/\tau} + 1 - \frac{s}{\tau}, \quad \frac{d}{dl} \log \tilde{P}_{ss} = -v_0(e^{-ns/\tau} - 1), \tag{70}$$

where we introduce the parameter l of the characteristic curve. These equations can be regarded as describing a "dynamical system" in terms of the auxiliary "time" l. This formulation is useful because the well-developed theory of dynamical systems is applicable even to more general cases as shown later.

a. Subcritical condition $n < 1$. Let us first focus on the subcritical case $n < 1$ and consider the expansion of Eq. (70) around $s = 0$, which leads to

$$\frac{ds}{dl} \simeq -\frac{1-n}{\tau} s - \frac{n^2 s^2}{2\tau^2} + \cdots, \quad \frac{d}{dl} \log \tilde{P}_{ss} \simeq \frac{nv_0}{\tau} s + \cdots. \tag{71}$$

The corresponding flow of this effective dynamical system $s(l)$ along the s axis as a function of "time" l is illustrated in

(a)

$-2\tau\epsilon \quad 0$

Subcritical condition
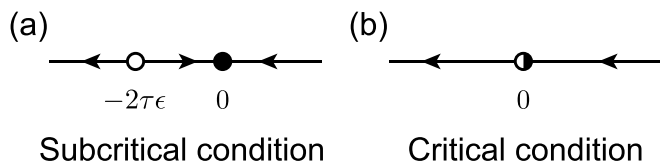
(b)

0

Critical condition

FIG. 5. Schematic illustration of the vector field $V(s) := ds/dt = -e^{-ns/\tau} + 1 - s/\tau$ along the s dimension as a function of "time" l. (a) Near critical condition $\epsilon := 1 - n \ll 1$, two fixed points exist at $s = 0$ (attractor) and $s \simeq -2\tau\epsilon$ (repeller). (b) At the critical condition $n = 1$, the repeller merges with the attractor, which corresponds to a transcritical bifurcation.

Fig. 5. Near the critical condition $\epsilon = 1 - n \ll 1$, this "dynamical system" has two fixed points $V(s) = 0$ at

$$s = 0, \quad s \simeq -2\tau\epsilon. \tag{72}$$

The former is a stable attractor whereas the latter is an unstable repeller [see Fig. 5(a)]. Remarkably, the critical condition $n = 1$ for the Hawkes process corresponds to the condition of a transcritical bifurcation [i.e., the repeller merges with the attractor; see Fig. 5(b)] for the "dynamical system" described by the Lagrange-Charpit equations. This picture is useful because it can be straightforwardly generalized to more general memory kernels $h(t)$, as shown later.

Let us neglect the subleading contribution to obtain the general solution as

$$s = e^{-(1-n)(l-l_0)/\tau}, \quad \log \tilde{P}_{ss} \simeq \frac{nv_0}{\tau} \int dl\, s(l) + C, \tag{73}$$

with constants of integration l_0 and C. In the following, we set the initial "time" (i.e., the initial point on the characteristic curve) as $l_0 = 0$. We then obtain

$$\log \tilde{P}_{ss} \simeq -\frac{nv_0 s}{1 - n} + C \tag{74}$$

with constant of integration C. This constant is fixed by the condition of normalization of the PDF, given by $\log \tilde{P}_{ss} = 0$ for $s = 0$, which imposes $C = 0$. We thus obtain

$$\log \tilde{Q}_{ss}(s) = -v_0 s + \log \tilde{P}_{ss}(s) \simeq -\frac{v_0 s}{1 - n}, \tag{75}$$

which is consistent with the asymptotic mean intensity in the steady state [see the long time limit of Eq. (67)].

b. At criticality $n = 1$. For $n = 1$, the lowest-order contribution in the Lagrange-Charpit equation is given by

$$\frac{ds}{dl} \simeq -\frac{s^2}{2\tau^2} \implies s = \frac{2\tau^2}{l - l_0} \tag{76}$$

with constant of integration l_0. In the following, we set $l_0 = 0$ as the initial point on the characteristic curve. We then obtain

$$\log \tilde{P}_{ss} = \frac{v_0}{\tau} \int dl\, s(l) + C \simeq -2v_0\tau \log|s| + C \tag{77}$$

with the constant of integration C. The constant is an "divergent" constant since it has to compensate the diverging logarithm to ensure that $\log P_{ss}(s = 0) = 0$. This divergent constant appears as a result of neglecting the ultraviolet (UV) cutoff for small s (which corresponds to neglecting the exponential tail of the PDF of intensities). By ignoring the

divergent constant C, we obtain the intermediate asymptotics,

$$P_{ss}(v) \propto v^{-1+2v_0\tau}, \tag{78}$$

which recovers the leading power law intermediate asymptotic (53), which is a special case of the general solution (47).

C. Double exponential kernel

We now consider the case where the memory function (24) is made of $K = 2$ exponential functions. Since the Laplace representation of the master equation is still a first-order partial differential equation, its solution can be formally obtained by the method of characteristics (see Appendix C for a short review). Unfortunately, the time-dependent Lagrange Charpit equation cannot be exactly solved in explicit form anymore. We therefore focus on the steady-state solution of the master equation, with a special focus on the regime close to the critical point. We develop the stability analysis of the Lagrange-Charpit equations following the same approach as in Sec. IV B 3.

Let us start from the Lagrange-Charpit equations, which are given by

$$\frac{ds_1}{dl} = -e^{-(n_1 s_1/\tau_1 + n_2 s_2/\tau_2)} + 1 - \frac{s_1}{\tau_1}, \tag{79a}$$

$$\frac{ds_2}{dl} = -e^{-(n_1 s_1/\tau_1 + n_2 s_2/\tau_2)} + 1 - \frac{s_2}{\tau_2}, \tag{79b}$$

$$\frac{d\Phi}{dl} = -v_0(e^{-(n_1 s_1/\tau_1 + n_2 s_2/\tau_2)} - 1) \quad \text{with} \quad \Phi := \log \tilde{P}_{ss}, \tag{79c}$$

and l is the auxiliary "time" parameterizing the position on the characteristic curve. Let us develop the stability analysis around $s = 0$ (i.e., for large v's) for this pseudodynamical system.

a. Subcritical case $n < 1$. Assuming $n := n_1 + n_2 < 1$, let us first consider the linearized dynamics of system (79) as

$$\frac{ds}{dl} \simeq -Hs, \quad \frac{d\Phi}{dl} \simeq v_0 Ks \tag{80}$$

with

$$s := \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}, \quad H := \begin{pmatrix} \frac{1-n_1}{\tau_1} & \frac{-n_2}{\tau_2} \\ \frac{-n_1}{\tau_1} & \frac{1-n_2}{\tau_2} \end{pmatrix}, \quad K := \left(\frac{n_1}{\tau_1}, \frac{n_2}{\tau_2} \right). \tag{81}$$

Regarding this system as a dynamical system with the auxiliary "time" l, its qualitative dynamics can be illustrated by its phase space depicted in Fig. 6. In the subcritical case $n < 1$, the origin $s = (0, 0)$ is "attractive" since all the eigenvalues of H are positive [Fig. 6(a)].

Let us introduce the eigenvalues λ_1, λ_2 and eigenvectors e_1, e_2 of H, such that

$$P := (e_1, e_2), \quad P^{-1}HP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}. \tag{82}$$

Because all eigenvalues are real (see Appendix F 1 for the proof), we denote $\lambda_1 \leqslant \lambda_2$. The determinant of H is given by
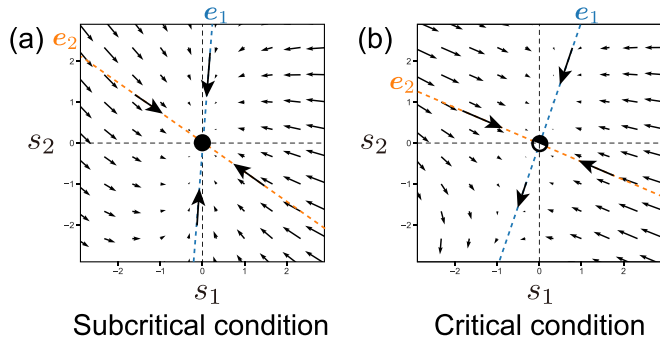
$$\det H = \frac{1 - n}{\tau_1 \tau_2}. \tag{83}$$

FIG. 6. Qualitative representation of the Lagrange-Charpit equations in phase space. By rewriting $ds/dk := V(s) \simeq -Hs$, the "velocity" vector field $V(s)$ is plotted in the phase space (s_1, s_2). (a) Subcritical case with $(\tau_1, \tau_2, n_1, n_2) = (1, 3, 0.3, 0.1)$, showing that $s = 0$ is a stable attractor. (b) Critical case with $(\tau_1, \tau_2, n_1, n_2) = (1, 3, 0.3, 0.7)$, showing that the e_1 direction is marginal in terms of the linear stability analysis (i.e., the repeller merges with the attractor, which corresponds to a transcritical bifurcation in dynamical systems).

This means that the zero eigenvalue $\lambda_1 = 0$ appears at the critical point $n = 1$. Below the critical point $n < 1$, all the eigenvalues are positive $(\lambda_1, \lambda_2 > 0)$. For $n < 1$, the dynamics can be rewritten as

$$\frac{d}{dl} P^{-1} s = -\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} P^{-1} s \implies s(l) = P \begin{pmatrix} e^{-\lambda_1(l-l_0)} \\ e^{-\lambda_2(l-l_0)}/C_1 \end{pmatrix} \tag{84}$$

with constants of integration l_0 and C_1. We can assume $l_0 = 0$ as the initial point of the characteristic curve without loss of generality. Integrating the second equation in (80), we obtain

$$\Phi = v_0 K \int s(l) dl + C_2 = -v_0 K P \begin{pmatrix} 1/\lambda_1 & 0 \\ 0 & 1/\lambda_2 \end{pmatrix} P^{-1} s + C_2$$
$$= -v K H^{-1} s + C_2. \tag{85}$$

The general solution is given by

$$\mathcal{H}(C_1) = C_2 \tag{86}$$

with a function $\mathcal{H}$ determined by the initial condition on the characteristic curve. Let us introduce

$$\bar{s} := P^{-1} s = \begin{pmatrix} \bar{s}_1 \\ \bar{s}_2 \end{pmatrix} \implies C_1 = (\bar{s}_1)^{\lambda_2/\lambda_1} (\bar{s}_2)^{-1}. \tag{87}$$

This means that the solution is given by the following form:

$$\Phi(s) = -v K H^{-1} s + \mathcal{H}((\bar{s}_1)^{\lambda_2/\lambda_1} (\bar{s}_2)^{-1}). \tag{88}$$

Because of the renormalization of the PDF, the following relation

$$\lim_{s \to 0} \Phi(s) = 0 \tag{89}$$

must hold for any path in the (s_1, s_2) space ending on the origin (limit $s \to 0$). Let us consider the specific limit such that $\bar{s}_1 \to 0$ with $\bar{s}_2 = x^{-1}(\bar{s}_1)^{\lambda_2/\lambda_1}$ for an arbitrary positive x:

$$\lim_{\bar{s}_1 \to 0} \Phi(s) = \mathcal{H}(x). \tag{90}$$

Since the left-hand side (LHS) is zero for any x, the function $\mathcal{H}(\cdot)$ must be identically zero. With $\Phi := \log \tilde{P}_{ss}$ as defined in (79), this leads to

$$\log \tilde{P}_{ss}(s) = -v K H^{-1} s. \tag{91}$$

By substituting with the special $s = (s_1 = s, s_2 = s)^T$, we obtain

$$\log \tilde{Q}_{ss}(s) = -v_0 s + \Phi_{t=\infty}(s(1,1)^T) \simeq -\frac{v_0}{1-n} s \tag{92}$$

for small s, which recovers expression (75) derived above.

b. Critical case $n = 1$. In this case, the eigenvalues and eigenvectors of H are given by

$$\lambda_1 = 0, \quad \lambda_2 = \frac{n_1 \tau_1 + n_2 \tau_2}{\tau_1 \tau_2}, \quad e_1 = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix}, \quad e_2 = \begin{pmatrix} -n_2 \\ n_1 \end{pmatrix}. \tag{93}$$

This means that the eigenvalue matrix and its inverse matrix are respectively given by

$$P = \begin{pmatrix} \tau_1 & -n_2 \\ \tau_2 & n_1 \end{pmatrix}, \quad P^{-1} = \frac{1}{\alpha} \begin{pmatrix} n_1 & n_2 \\ -\tau_2 & \tau_1 \end{pmatrix},$$
$$\alpha := \det P = \tau_1 n_1 + \tau_2 n_2. \tag{94}$$

This value of α is the special case for two exponentials of the general definition (34). Accordingly, let us introduce

$$X = (X, Y)^T = P^{-1} s, \iff X = \frac{n_1 s_1 + n_2 s_2}{\alpha},$$
$$Y = \frac{-\tau_2 s_1 + \tau_1 s_2}{\alpha}. \tag{95}$$

We then obtain

$$\frac{dX}{dl} = 0, \quad \frac{dY}{dl} = -\lambda_2 Y \tag{96}$$

at the leading linear order in expansions in powers of X and Y. Since the first linear term is zero in the dynamics of X, corresponding to a transcritical bifurcation for the Lagrange-Charpit equations (79), we need to take into account the second-order term in X, namely

$$e^{-(n_1 s_1/\tau_1 + n_2 s_2/\tau_2)} \simeq 1 - X + \frac{X^2}{2} + n_1 n_2 \left(\frac{1}{\tau_1} - \frac{1}{\tau_2} \right) Y$$
$$+ O(XY, X^2 Y, Y^2), \tag{97}$$

where we have dropped terms of the order Y^2, XY, and $X^2 Y$. We then obtain the dynamical equations at the transcritical bifurcation [see Fig. 6(b)] to leading order

$$\frac{dY}{dl} \simeq -\lambda_2 Y, \quad \frac{dX}{dl} \simeq -\frac{X^2}{2\alpha}, \tag{98}$$

whose solutions are given by

$$X(l) = \frac{2\alpha}{l - l_0}, \quad Y(l) = C_1 e^{-\lambda_2(l-l_0)} \tag{99}$$

with constants of integration l_0 and C_1. We can assume $l_0 = 0$ as the initial point on the characteristic curve. Remarkably, only the contribution along the X axis is dominant for the large l limit (i.e., $|X| \gg |Y|$ for $l \to \infty$), which corresponds to the
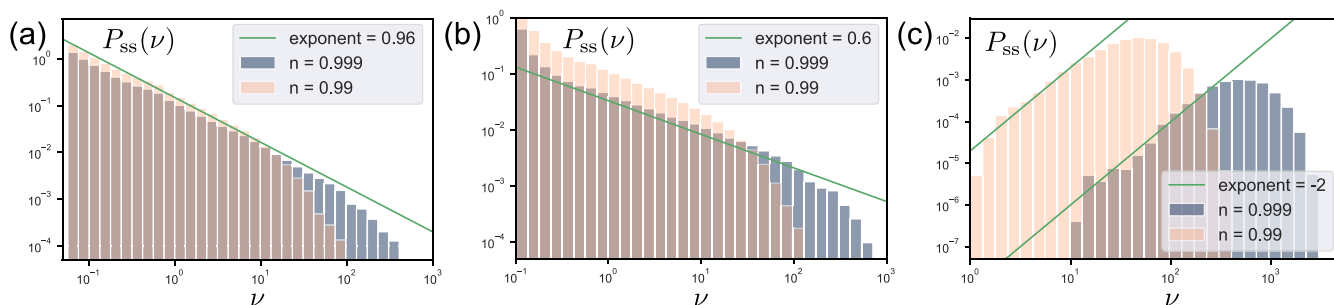
FIG. 7. Numerical evaluation of the steady-state PDF of the Hawkes intensity $\hat{\nu}$ for the double exponential case $K = 2$ (24) with $(\tau_1, \tau_2) = (1, 3)$, $(n_1, n_2) = (0.5, 0.499)$, or $(n_1, n_2) = (0.5, 0.49)$, near the critical point: (a) Background intensity $\nu_0 = 0.01$, leading to the power law exponent 0.96. (b) $\nu_0 = 0.1$, leading to the power law exponent 0.6. (c) $\nu_0 = 0.75$, leading to the negative (i.e., growing PDF) power law exponent -2.0. For all simulations, the sampling time interval and total sampling time are $dt = 0.001$ and $T_{\text{tot}} = 10000$ from the initial condition $\hat{z}(0) = 0$. The initial 10% of the sample was discarded from the statistics.

asymptotic limit $s \rightarrow 0$. We then obtain

$$\Phi \simeq \nu_0 \int dl \left(\frac{n_1 s_1(l)}{\tau_1} + \frac{n_2 s_2(l)}{\tau_2} \right)$$

$$\simeq -2\nu_0 \alpha \log |X| + \frac{\nu_0 n_1 n_2}{\lambda_2} \left(\frac{1}{\tau_1} - \frac{1}{\tau_2} \right) Y + C_2 \quad (100)$$

with constant of integration C_2. The general solution is given by

$$\mathcal{H}(C_1) = C_2 \quad (101)$$

with a function $\mathcal{H}$, which is determined by the initial condition. Considering that

$$C_1 = Y \exp \left[\frac{2\lambda_2 \alpha}{X} \right], \quad (102)$$

the solution is given by the following form:

$$\Phi(s) = -2\nu_0 \alpha \log |X| + \frac{\nu_0 n_1 n_2}{\lambda_2} \left(\frac{1}{\tau_1} - \frac{1}{\tau_2} \right) Y$$

$$+ \mathcal{H} \left(Y \exp \left[\frac{2\lambda_2 \alpha}{X} \right] \right). \quad (103)$$

Because we have neglected the UV cutoff for small s, there is an artificial divergent term $-2\nu_0 \alpha \log |X|$ for small X. Except for this divergent term, $\Phi(s)$ must be constant for $s \rightarrow 0$. The function $\mathcal{H}(\cdot)$ is thus constant because

$$\lim_{y \rightarrow 0} [\Phi(s) + 2\nu_0 \alpha \log |X|] = \mathcal{H}(Z) = \text{const.} \quad (104)$$

with the choice of $X = 2\lambda_2 \alpha / \log(Z/Y)$ for any positive constant Z. Therefore, we obtain the steady solution

$$\log \tilde{P}_{ss}(s) \simeq -2\nu_0 \alpha \log |X| + \frac{\nu_0 n_1 n_2}{\lambda_2} \left(\frac{1}{\tau_1} - \frac{1}{\tau_2} \right) Y \quad (105)$$

for small X and Y, by ignoring the UV cutoff and the constant contribution. This recovers the power law formula of the intermediate asymptotics of the PDF of the Hawkes intensities:

$$\log \tilde{Q}_{ss}(s) := -\nu_0 s + \log \tilde{P}_{ss}(s, s) \simeq -2\nu_0 \alpha \log |s| \quad (s \sim 0)$$

$$\iff P(\nu) \sim \nu^{-1+2\nu_0 \alpha} \quad (\nu \rightarrow +\infty), \quad (106)$$

with $\alpha = \tau_1 n_1 + \tau_2 n_2$ as defined in (94).

c. Numerical verification. We have numerically confirmed our theoretical prediction (106), a special case of (47) for a memory kernel with two exponentials, as shown in Fig. 7. The main properties are the same as those shown in Fig. 4, implying that our prediction is verified for memory kernels with one and two exponentials.

D. Discrete superposition of exponential kernels

We now consider the case where the memory kernel is the sum of an arbitrary finite number K of exponentials according to expression (24). Our treatment follows the method presented for the case $K = 2$.

The corresponding Lagrange-Charpit equations read

$$\frac{ds_k}{dl} = -e^{-\sum_{j=1}^{K} n_j s_j / \tau_j} + 1 - \frac{s_k}{\tau_k},$$

$$\frac{d\Phi}{dl} = -\nu_0 (e^{-\sum_{j=1}^{K} n_j s_j / \tau_j} - 1). \quad (107)$$

The derivation of the PDF of the Hawkes intensities boils down to a stability analysis of these equations around $s = 0$ in the neighborhood of the critical condition $n = 1$.

a. Subcritical case $n < 1$. We linearize the Lagrange-Charpit equations to obtain

$$\frac{ds}{dl} \simeq -Hs, \quad \frac{d\Phi}{dl} \simeq \nu_0 Ks \quad (108)$$

with

$$H := \begin{pmatrix} \frac{1-n_1}{\tau_1}, & -\frac{n_2}{\tau_2}, & \cdots & -\frac{n_K}{\tau_K} \\ -\frac{n_1}{\tau_1}, & \frac{1-n_2}{\tau_2}, & \cdots & -\frac{n_K}{\tau_K} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{n_1}{\tau_1}, & -\frac{n_2}{\tau_2}, & \cdots & \frac{1-n_K}{\tau_K} \end{pmatrix}, \quad K := \left(\frac{n_1}{\tau_1}, \ldots, \frac{n_K}{\tau_K} \right).$$

$$(109)$$

Considering that all eigenvalues $\{\lambda_k\}_{k=1,\ldots,K}$ of H are real (see Appendix F 1 for its proof), we order them according to $\lambda_i < \lambda_j$ for $i < j$. We denote the corresponding eigenvectors as $\{e_k\}_{k=1,\ldots,K}$. The matrix H can thus be diagonalized as

follows:

$$
\boldsymbol{P} := (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_K), \quad \boldsymbol{P}^{-1}\boldsymbol{H}\boldsymbol{P} = \begin{pmatrix} \lambda_1, & 0, & \ldots & 0 \\ 0, & \lambda_2, & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \ldots & \lambda_K \end{pmatrix}.
$$
(110)

The critical case $n = 1$ corresponds to the existence of a zero eigenvalue. Therefore, at the critical point, the determinant of $\boldsymbol{H}$ is zero (see Appendix F 2 for the derivation of the explicit form of its determinant):

$$
\det \boldsymbol{H} = \frac{1 - \sum_{k=1}^{K} n_k}{\prod_{k=1}^{K} \tau_k} = 0 \quad \Longleftrightarrow \quad n := \sum_{k=1}^{K} n_k = 1.
$$
(111)

Following calculations similar those presented in to Sec. IV B, we obtain

$$
\Phi(\boldsymbol{s}) \simeq -\nu_0 \boldsymbol{K}\boldsymbol{H}^{-1}\boldsymbol{s},
$$
(112)

where the inverse matrix $\boldsymbol{H}^{-1}$ is explicitly given in Appendix F 3. We finally obtain

$$
\log \tilde{Q}_{ss}(s) = -\nu_0 s + \Phi(s(1, \ldots, 1)^T) = \frac{-\nu_0}{1 - n} s,
$$
(113)

again recovering (92) and (75) derived above.

b. Critical case $n = 1$. At the critical point, the smallest eigenvalue of $\boldsymbol{H}$ is zero ($\lambda_1 = 0$). By direct substitution, its corresponding eigenvector is

$$
\boldsymbol{e}_1 = (\tau_1, \ldots, \tau_K)^T,
$$
(114)

as seen from

$$
\boldsymbol{H}\boldsymbol{e}_1 = \begin{pmatrix} \frac{1-n_1}{\tau_1}, & -\frac{n_2}{\tau_2}, & \ldots & -\frac{n_K}{\tau_K} \\ -\frac{n_1}{\tau_1}, & \frac{1-n_2}{\tau_2}, & \ldots & -\frac{n_K}{\tau_K} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{n_1}{\tau_1}, & -\frac{n_2}{\tau_2}, & \ldots & \frac{1-n_K}{\tau_K} \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_K \end{pmatrix}
$$

$$
= \begin{pmatrix} 1-n \\ 1-n \\ \vdots \\ 1-n \end{pmatrix} = \boldsymbol{0} \quad \text{for } n = 1.
$$
(115)

We now introduce a new set of variables (i.e., representation based on the eigenvectors)

$$
\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \ldots \\ X_K \end{pmatrix} := \boldsymbol{P}^{-1}\boldsymbol{s}, \qquad \boldsymbol{P}^{-1} = \begin{pmatrix} \boldsymbol{g}_1^T \\ \boldsymbol{g}_2^T \\ \ldots \\ \boldsymbol{g}_K^T \end{pmatrix}.
$$
(116)

The linearized Lagrange-Charpit equations are given by

$$
\frac{dX_1}{dl} \simeq 0, \qquad \frac{dX_j}{dl} \simeq -\lambda_j X_j \quad \text{for } j \geqslant 2.
$$
(117)

Similarly to Eq. (99), the leading-order contribution comes from the X_1 direction because $|X_1| \gg |X_j|$ for $j \geqslant 2$ in the asymptotic limit $l \to \infty$. We thus neglect other contribution by assuming $X_j \sim 0$ for $j \geqslant 2$. It is therefore necessary to take the second-order contribution along the X_1 direction,

$$
e^{-\sum_{k=1}^{K} n_k s_k / \tau_k} - 1 = -\sum_{k=1}^{K} \frac{n_k s_k}{\tau_k} + \frac{1}{2}\left(\sum_{k=1}^{K} \frac{n_k s_k}{\tau_k}\right)^2 + \cdots.
$$
(118)

We note that X_1 is given by

$$
X_1 = \boldsymbol{g}_1 \cdot \boldsymbol{s} = \frac{1}{\alpha}\sum_{k=1}^{K} n_k s_k, \qquad \boldsymbol{g}_1 = \left(\frac{n_1}{\alpha}, \ldots, \frac{n_K}{\alpha}\right)^T,
$$
(119)

where $\alpha := \sum_{k=1}^{K} \tau_k n_k$, which is a special case for a discrete sum of exponentials of the general definition (34).

Taking the derivative of (119) and using Eq. (107), we obtain

$$
\frac{dX_1}{dl} = \frac{1}{\alpha}\sum_{k=1}^{K} n_k \frac{ds_k}{dl} = 0 - \frac{1}{2\alpha}\left(\sum_{k=1}^{K} \frac{n_k s_k}{\tau_k}\right)^2 + \cdots.
$$
(120)

This means that $\boldsymbol{g}_1$ is a correct representation. Note that the value of α given by (34) ensures consistency with the following identify:

$$
\boldsymbol{P}^{-1}\boldsymbol{P} = \begin{pmatrix} \boldsymbol{g}_1^T \\ \boldsymbol{g}_2^T \\ \ldots \\ \boldsymbol{g}_K^T \end{pmatrix}(\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_K) = \begin{pmatrix} n_1/\alpha, & n_2/\alpha, & \ldots & n_K/\alpha \\ \bigcirc, & \bigcirc, & \ldots & \bigcirc \\ \vdots & \vdots & \ddots & \vdots \\ \bigcirc, & \bigcirc, & \ldots & \bigcirc \end{pmatrix} \begin{pmatrix} \tau_1, & \bigcirc, & \ldots & \bigcirc \\ \tau_2, & \bigcirc, & \ldots & \bigcirc \\ \vdots & \vdots & \ddots & \vdots \\ \tau_2, & \bigcirc, & \ldots & \bigcirc \end{pmatrix} = \begin{pmatrix} 1, & 0, & \ldots & 0 \\ 0, & 1, & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \ldots & 1 \end{pmatrix},
$$
(121)

where $\bigcirc$ represents some unspecified value. Since the contribution of $X_2, \ldots X_K$ can be ignored for the description of the leading behavior along X_1, let us set $X_2 = X_3 = \cdots = X_K = 0$, which leads

$$
\boldsymbol{s} = \boldsymbol{P}\boldsymbol{X} \simeq (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_K)\begin{pmatrix} X_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = X\boldsymbol{e}_1 = \begin{pmatrix} X_1 \tau_1 \\ X_1 \tau_2 \\ \vdots \\ X_1 \tau_K \end{pmatrix}.
$$
(122)

We thus obtain the second-order contribution along the X_1 axis by ignoring nonlinear contribution from $X_2, \ldots, X_K$:

$$\frac{dX_1}{dl} \simeq -\frac{X_1^2}{2\alpha}. \tag{123}$$

With calculations that follow step by step those in Sec. IV B, we obtain

$$\log \tilde{Q}_{ss}(s) \simeq -2\nu_0 \alpha \log |s|$$

$$\iff P(\nu) \sim \nu^{-1+2\nu_0\alpha}, \quad \text{with } \alpha := \sum_{k=1}^{K} n_k \tau_k. \tag{124}$$

This recovers the power law formula of the intermediate asymptotics of the PDF of the Hawkes intensities given by (47).

$$\int_0^\infty dx \left(e^{-\int_0^\infty dx' s(x')n(x')/x'} - 1 + \frac{s(x)}{x} \right) \frac{\delta \Phi[s]}{\delta s(x)} = \nu_0 (e^{-\int_0^\infty dx' s(x')n(x')/x'} - 1). \tag{126}$$

The corresponding Lagrange-Charpit equations are the following partial-integro equations,

$$\frac{\partial s(l; x)}{\partial l} = 1 - e^{-\int_0^\infty dx' s(x')n(x')/x'} - \frac{s(x)}{\tau}, \quad \frac{\partial \Phi(l)}{\partial l} = -\nu_0 (e^{-\int_0^\infty dx' s(x')n(x')/x'} - 1), \tag{127}$$

where l is the curvilinear parameter indexing the position along the characteristic curve. We now perform the stability analysis of Eq. (127) in the neighborhood of $s = 0$ close to the critical condition $n = 1$.

a. Subcritical case $n < 1$. We linearize the Lagrange-Charpit equation (127) to obtain

$$\frac{\partial s(l; x)}{\partial l} = -\int_0^\infty dx' H(x, x') s(x'),$$

$$\frac{\partial \Phi(l)}{\partial l} = \nu_0 \int_0^\infty dx' K(x') s(x') \tag{128}$$

with

$$H(x, x') := \frac{\delta(x - x') - n(x')}{x'}, \quad K(x') := \frac{n(x')}{x'}. \tag{129}$$

Let us introduce the eigenvalues $\lambda \geqslant \lambda_{\min}$ and eigenfunctions $e(x; \lambda)$, satisfying

$$\int_0^\infty dx' H(x, x') e(x'; \lambda) = \lambda e(x; \lambda). \tag{130}$$

Appendix G 1 shows that all the eigenvalues are real. The inverse matrix of $H(x, x')$, denoted by $H^{-1}(x, x')$, can be explicitly obtained as shown in Appendix G 2. Since the inverse matrix $H^{-1}(x, x')$ has a singularity at $n = 1$, the critical condition of this Hawkes process is given by $n = 1$ as expected. Using calculations that are analogous to those in Sec. IV B, we obtain

$$\Phi[s] \simeq -\nu_0 \int_0^\infty dx \int_0^\infty dx' K(x) H^{-1}(x, x') s(x'), \tag{131}$$

from which we state that

$$\log \tilde{Q}_{ss}(s) = -\nu_0 s + \Phi[s\mathbf{1}(x)] = \frac{-\nu_0}{1-n} s, \tag{132}$$

where $\mathbf{1}(x)$ is an indicator function defined by $\mathbf{1}(x) = 1$ for any x. This recovers (75), (92), and (113) derived above.

E. General case

We are now prepared to study the general case where the memory kernel of the Hawkes process is a continuous superposition of exponential functions (33). Introducing the steady-state cumulant functional

$$\Phi[s] := \log \tilde{P}_{ss}[s], \tag{125}$$

and from the master equation in its functional Laplace representation Eq. (45), we obtain the following first-order functional differential equation in the steady state,

b. Critical case $n = 1$. At criticality, the smallest eigenvalue vanishes: $\lambda_{\min} = 0$. Indeed, we obtain the zero eigenfunction

$$e(x; \lambda = 0) = x, \tag{133}$$

which can be checked by direct substitution:

$$\int_0^\infty dx H(x, x') e(x'; \lambda = 0) = \int_0^\infty dx \frac{\delta(x - x') - n(x')}{x'} x'$$

$$= 1 - n = 0, \quad \text{for } n = 1. \tag{134}$$

We now introduce a set of variables to obtain a new representation based on the eigenfunctions,

$$s(x) = \sum_\lambda e(x; \lambda) X(\lambda) \iff X(\lambda) = \int_0^\infty dx e^{-1}(\lambda; x) s(x), \tag{135}$$

with the inverse matrix $e^{-1}(\lambda; x)$ satisfying

$$\int_0^\infty dx e^{-1}(\lambda; x) e(x; \lambda') = \delta_{\lambda, \lambda'}. \tag{136}$$

We assume the existence of the inverse matrix, which is equivalent to the assumption that the set of all eigenfunctions is complete. $H(x, x')$ can be diagonalized:

$$\int_0^\infty dx \int_0^\infty dx' e^{-1}(\lambda; x) H(x, x') e(x'; \lambda') = \lambda \delta_{\lambda, \lambda'}. \tag{137}$$

We then obtain the linearized Lagrange-Charpit equations,

$$\frac{\partial X(\lambda)}{\partial l} \simeq -\lambda X(\lambda). \tag{138}$$

The dominant contribution comes from the vanishing eigenvalue. We therefore focus on $X(0)$ by setting $X(\lambda) = 0$ for

$\lambda > 0$. We then form the expansion

$$
e^{-\int_0^\infty dx' s(x') n(x')/x'} - 1 = -\int_0^\infty dx' \frac{n(x')s(x')}{x'}
$$
$$
+ \frac{1}{2}\left(\int_0^\infty dx' \frac{n(x')s(x')}{x'}\right)^2 + \cdots.
$$
(139)

The explicit representation of $x(0)$ is given by

$$
X(\lambda = 0) = \int_0^\infty dx\, e^{-1}(\lambda = 0; x)s(x)
$$
$$
= \frac{1}{\alpha}\int_0^\infty dx\, n(x)s(x), \quad e^{-1}(\lambda = 0; x) = \frac{n(x)}{\alpha},
$$
(140)

where α is defined by expression (34). Expression (140) can be checked to be valid by direct substitution since, from Eq. (127), we have

$$
\frac{\partial X(0)}{\partial l} = \int_0^\infty dx\, e^{-1}(0; x)\frac{\partial s(x)}{\partial l}
$$
$$
= 0 - \frac{1}{2\alpha}\left(\int_0^\infty dx \frac{n(x)s(x)}{x}\right)^2 + \cdots, \quad (141)
$$

showing that the first-order contribution is actually null in this representation. The parameter α (34) has the property to ensure the consistency with the following identity:

$$
\int_0^\infty dx\, e^{-1}(\lambda = 0; x)e(x; \lambda = 0) = \frac{1}{\alpha}\int_0^\infty dx\, n(x)x
$$
$$
= \delta_{\lambda=0, \lambda'=0} = 1. \quad (142)
$$

Since we ignore the contribution from $X(\lambda)$ except for $\lambda = 0$, let us set $X(\lambda) = 0$ for $\lambda > 0$, which yields

$$
s(x) = \sum_\lambda e(x; \lambda)X(\lambda) = e(x; 0)X_0 = X_0 x, \quad (143)
$$

where we have written $X_0 := X(0)$. We then obtain the second-order contribution along the X_0 axis,

$$
\frac{\partial X_0}{\partial l} \simeq -\frac{X_0^2}{2\alpha}. \quad (144)
$$

From calculations mimicking those in Sec. IV B, we obtain

$$
\log \tilde{Q}_{ss}(s) \simeq -2\nu_0\alpha \log |s| \iff P(\nu) \sim \nu^{-1+2\nu_0\alpha},
$$
$$
\alpha := \int_0^\infty dx\, n(x)x. \quad (145)
$$

This recovers the power law formula of the intermediate asymptotics of the PDF of the Hawkes intensities given by (47).

V. DISCUSSION: FORMAL RELATION TO QUANTUM FIELD THEORY

We have formulated the non-Markovian Hawkes process as a classical field theory associated with stochastic excitation. While the formulation is entirely classical, it is interesting to point out its formal relationship with quantum field theory.

A. Formal equivalence between the generalized Langevin equation and quantum field theory

We first discuss the formal relationship between the GLE and quantum field theory. In the beginning, let us study the case of the discrete-exponential-sum memory kernel (8). It is well known that the Fokker-Planck (FP) or master equations have a structure that is quite similar to that of the Schrödinger equation [61], and here we reformulate the Fokker-Planck or master equations according to this classical idea.

1. Schrödinger-like representation for the Fokker-Planck equation

After rewriting $v \to \Phi$, $u_k \to \phi_k$, and $|P_t\rangle := \int d\Phi \prod_{k=1}^K d\phi_k P_t(\Phi, \phi_1, \ldots, \phi_K)|\Phi, \phi_1, \ldots, \phi_K\rangle$, the FP equation can be rewritten in a quantum-mechanics-like form:

$$
\frac{\partial}{\partial t}|P_t\rangle = H_{\mathrm{orgn}}|P_t\rangle,
$$
$$
H_{\mathrm{orgn}} := \sum_{k=1}^K \left[\frac{-i}{M}\Pi\phi_k + i\pi_k\left(\frac{\phi_k}{\tau_k} + \kappa_k\Phi\right) - \frac{\kappa_k T}{\tau_k}\pi_k^2\right]
$$
(146)

with the "momentum" operators $\Pi := -i\partial/\partial\Phi$ and $\pi_k := -i\partial/\partial\phi_k$ satisfying commutative relations

$$
[\Phi, \Pi] = i, \quad [\phi_k, \pi_{k'}] = i\delta_{kk'}. \quad (147)
$$

While the Hamiltonian H_{org} is non-Hermitian (i.e., the evolution operator is not self-adjoint), this quantum-mechanical formulation is sometimes useful since analytical methods developed in quantum mechanics are available and can be formally transposed.

In particular, on the condition that the detailed balance is satisfied, there exists a mathematically better mapping [62]. The steady solution of the FP is given by

$$
P_{ss}(\Phi, \phi_1, \ldots, \phi_K) \propto \exp\left[-\frac{1}{T}\left(\frac{M}{2}\Phi^2 + \sum_{k=1}^K \frac{1}{2\kappa_k}\phi_k^2\right)\right].
$$
(148)

Here we make a transformation $|\psi_t\rangle := P_{ss}^{-1/2}|P_t\rangle$ to obtain

$$
-\frac{\partial}{\partial t}|\psi_t\rangle = (H_H + H_A)|\psi_t\rangle,
$$
$$
H_H := \sum_{k=1}^K \left[\frac{\pi_k^2}{2m_k} + \frac{m_k\omega_k^2}{2}\phi_k^2 - \frac{\omega_k}{2}\right],
$$
$$
H_A := i\sum_{k=1}^K \left[\kappa_k\Phi\pi_k - \frac{1}{M}\phi_k\Pi\right] \quad (149)
$$

with $m_k := \tau_k/(2T\kappa_k)$ and $\omega_k := 1/\tau_k$. Here, H_H and H_A are Hermitian and anti-Hermitian operators, respectively (i.e., $H_H^\dagger = H_H$ and $H_A^\dagger = -H_A$). Equation (149) can be regarded as a non-Hermitian Schrödinger equation based on the harmonic-oscillator Hamiltonian H_H. This decomposition is known to be mathematically useful in particular for the eigenfunction expansions [62].

We can introduce the creation and annihilation operators and their commutative relation

$$a_k^\dagger := \sqrt{\frac{m_k \omega_k}{2}} \left(\phi_k - \frac{i}{m_k \omega_k} \pi_k \right),$$

$$a_k := \sqrt{\frac{m_k \omega_k}{2}} \left(\phi_k + \frac{i}{m_k \omega_k} \pi_k \right), \quad [a_k, a_{k'}^\dagger] = \delta_{k,k'} \quad (150)$$

to lead the Hamiltonian of the harmonic oscillators

$$H_H = \sum_{k=1}^{K} \omega_k a_k^\dagger a_k. \quad (151)$$

2. Schrödinger-like representation for the field Fokker-Planck equation

This calculation can be generalized to the stochastic field theory of the Fokker-Planck equation. Indeed, by introducing the field operator and the corresponding commutative relation

$$\pi(x) := -i \frac{\delta}{\delta \phi(x)}, \quad [\phi(x), \pi(x')] = i\delta(x - x'), \quad (152)$$

we obtain the operator form of the FP field equation for the state vector $|P_t\rangle := \int d\Phi \mathcal{D}\phi P_t(\Phi, \phi)|\Phi, \phi\rangle$:

$$\frac{\partial}{\partial t} |P_t\rangle = H_{\text{orgn}} |P_t\rangle,$$

$$H_{\text{orgn}} := \int dx \left[\frac{-i}{M} \Pi \phi(x) + i\pi(x) \left(\frac{\phi(x)}{x} + \kappa(x)\Phi \right) \right.$$

$$\left. - \frac{\kappa(x)T}{x} \pi(x)^2 \right]. \quad (153)$$

The FP field equation can be transformed into a non-Hermitian quantum-field theory composed of harmonic oscillators on the field:

$$-\frac{\partial}{\partial t} |\psi_t\rangle = (H_H + H_A)|\psi_t\rangle \quad (154)$$

with Hermitian and anti-Hermitian operators

$$H_H := \int dx \left[\frac{\pi^2(x)}{2m(x)} + \frac{m(x)\omega^2(x)}{2} \phi^2(x) - \frac{\omega(x)}{2} \delta(0) \right],$$

$$H_A := i \int dx \left[\kappa(x)\Phi\pi(x) - \frac{1}{M}\phi(x)\Pi \right] \quad (155)$$

by defining

$$|\psi_t\rangle := P_{ss}^{-1/2}|P_t\rangle,$$

$$P_{ss}(\Phi, \{\phi(x)\}_x) \propto \exp\left[-\frac{1}{T} \left(\frac{M}{2} \Phi^2 + \int dx \frac{1}{2\kappa(x)} \phi^2(x) \right) \right],$$

$$m(x) := \frac{x}{2T\kappa(x)}, \quad \omega(x) := \frac{1}{x}. \quad (156)$$

While the divergent term $\delta(0)$ appears due to the singularity of the commutative relation (152) at $x = x'$, this term is irrelevant to the physical observables. Indeed, by introducing the field creation and annihilation operators

$$a^\dagger(x) := \sqrt{\frac{m(x)\omega(x)}{2}} \left[\phi(x) - \frac{i}{m(x)\omega(x)} \pi(x) \right],$$

$$a(x) := \sqrt{\frac{m(x)\omega(x)}{2}} \left[\phi(x) + \frac{i}{m(x)\omega(x)} \pi(x) \right] \quad (157)$$

satisfying the field commutative relation

$$[a(x), a^\dagger(x')] = \delta(x - x'), \quad (158)$$

we obtain the field harmonic-oscillator representation

$$H_H = \int dx \omega(x) a^\dagger(x) a(x), \quad (159)$$

where the divergent term $\delta(0)$ cancels out of the final results. This technical procedure is essentially the same as that in quantum electrodynamics, where renormalization of the energy is required to avoid divergence by removing the zero-point energy of harmonic oscillators.

The formal correspondence between the FP field equation and non-Hermitian quantum field theory itself is not surprising. For example, it is known that a stochastic chemical reaction system, characterized by an SPDE, is formally equivalent to non-Hermitian quantum field theory (see Sec. I E in Ref. [63]). Since the formal relationship between classical and quantum mechanics is a recent popular topic in terms of non-Hermitian physics [64], it might be interesting to further seek this mathematical relationship in understanding general non-Markovian processes.

B. Quantum-field-like representation for the Hawkes process

We can apply this formal procedure to the field master equation for the Hawkes process. Let us apply the functional Kramers-Moyal expansion to obtain

$$\frac{\partial P_t[z]}{\partial t} = \left\{ \int dx \frac{\delta}{\delta z(x)} \frac{z(x)}{x} + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \left[\int dx \frac{n(x)}{x} \frac{\delta}{\delta z(x)} \right]^k \left[\nu_0 + \int z(x') dx' \right] \right\} P_t[z], \quad (160)$$

where we have used the functional Taylor expansion

$$\left[\nu_0 + \int \left(z(x') - \frac{n(x')}{x'} \right) dx' \right] P_t \left[z - \frac{n}{x} \right] = \sum_{k=0}^{\infty} \frac{1}{k!} \left[-\int dx \frac{n(x)}{x} \frac{\delta}{\delta z(x)} \right]^k \left[\nu_0 + \int z(x') dx' \right] P_t[z]$$

$$= \exp\left[-\int dx \frac{n(x)}{x} \frac{\delta}{\delta z(x)} \right] \left[\nu_0 + \int z(x') dx' \right] P_t[z]. \quad (161)$$

We then introduce the field operators and the corresponding commutative relations,

$$\phi(x) := z(x), \quad \pi(x) := -i\frac{\delta}{\delta\phi(x)}, \quad [\phi(x), \pi(x')] = i\delta(x - x'). \tag{162}$$

We then obtain the Schrödinger-like representation for the state vector $|P_t\rangle := \int \mathcal{D}\phi P_t[\phi]|\phi\rangle$

$$\frac{\partial}{\partial t}|P_t\rangle = H|P_t\rangle \tag{163}$$

with non-Hermitian Hamiltonian H defined by

$$H := \int \frac{dx}{x}\pi(x)\phi(x) + \left\{\exp\left[\frac{1}{i}\int dx \frac{n(x)}{x}\pi(x)\right] - 1\right\}\left(\nu_0 + \int \phi(x')dx'\right). \tag{164}$$

Since the Hamiltonian includes infinite-order "momentum" operators, the Hamiltonian is classified as a nonlocal operator, which reflects trajectory jumps in the point processes. A similar nonlocal Hamiltonian representation for the master equation can be seen in Ref. [65] in the context of path-integral representations of Lévy processes. In this sense, the field master equation can be formally regarded as a non-Hermitian quantum field theory without locality. The exponential operator $T[y] := \exp[-i\int dxy(x)\pi(x)]$ naturally appears because it is a translation operator: $T[y]P_t[z] = P_t[z - y]$.

We note that the Hamiltonian reduces to a local operator if we can approximately truncate the Kramers-Moyal expansion up to the second order. While the validity of such approximation is not obvious for this linear Hawkes process, we can actually formulate such a formal approximation by generalizing the system size expansion for the field master equation in the case of nonlinear Hawkes processes [66].

VI. CONCLUSION

We have presented an analytical framework of the Hawkes process for an arbitrary memory kernel, based on the master equation governing the behavior of auxiliary field variables. We have derived systematically the corresponding functional master equation for the auxiliary field variables. While the Hawkes point process is non-Markovian by construction, the introduction of auxiliary field variables provides a formulation in terms of linear stochastic partial differential equations that are Markovian. For the case of a memory kernel decaying as a single exponential, we presented the exact time-dependent and steady-state solutions for the probability density function (PDF) of the Hawkes intensities, using the Laplace representation of the master equation. For memory kernels represented as arbitrary sums of exponential (discrete and continuous sums), we derived the asymptotic solutions of the Lagrange-Charpit equations for the hyperbolic master equations in the Laplace representation in the steady state, close to the critical point $n = 1$ of the Hawkes process, where n is the branching ratio. Our theory predicts a power law scaling of the PDF of the intensities in an intermediate asymptotics regime, which

crosses over to an asymptotic exponential function beyond a characteristics intensity that diverges as the critical condition is approached ($n \to 1$). The exponent of the PDF is nonuniversal and a function of the background intensity ν_0 of the Hawkes intensity and of the parameter $\alpha = n\langle\tau\rangle$, where $\langle\tau\rangle$ is the first-order moment of the distribution of timescales of the memory function of the Hawkes process. We found that the larger the memory $\langle\tau\rangle$, the larger the background intensity ν_0, and the larger the branching ratio n, the smaller is the exponent $1 - 2\nu_0\alpha$ of the PDF of Hawkes intensities.

This work provides the basic analytical tools to analyze Hawkes processes from a different angle than hitherto developed and will be useful to study more general and complex models derived from the Hawkes process. For instance, it is straightforward to extend our treatment to the case where each event has a mark quantifying its impact or "fertility," thus defining the more general Hawkes process with intensity $\hat{\nu}(t) = \nu_0 + n\sum_{i=1}^{\hat{N}(t)}\hat{\rho}_i h(t - \hat{t}_i)$ with independent and identically distributed random numbers $\{\hat{\rho}_i\}_i$. Our framework is also well suited to nonlinear generalisations of the Hawkes process, for instance with the intensity taking the form $\hat{\nu}(t) = g(\hat{\omega}(t)) > 0$, where the auxiliary variable $\hat{\omega}$ is given by $\hat{\omega}(t) = \omega_0 + n\sum_{i=1}^{\hat{N}(t)}\hat{\rho}_i h(t - \hat{t}_i)$ and where the times $\{\hat{t}_i\}_i$ of the events are determined from the intensity $\hat{\nu}(t)$. In this nonlinear version, the positivity of $\hat{\omega}(t)$ and ρ_i are not required anymore. This nonlinear Hawkes process is more complex than the linear Hawkes process but our framework can be applied to derive its most important analytical properties [66]. We note that such nonlinear Hawkes process include several models that have been proposed in the past, with applications to explain the multifractal properties of earthquake seismicity [67] and financial volatility [68].

APPENDIX A: EXPLICIT DERIVATION OF THE MASTER EQUATIONS

1. Derivation of Eqs. (27) and (31)

Given the dynamical equations (26) for the excess intensities $\{\hat{z}_k\}_{k=1,\dots,K}$, which are short-hand notations for the dynamics given by Eq. (18), for an arbitrary function $f(\hat{z})$, its stochastic time evolution therefore reads

$$df(\hat{z}(t)) = f(\hat{z}(t+dt)) - f(\hat{z}(t)) = \begin{cases} -\sum_{k=1}^{K} \frac{\hat{z}_k}{\tau_k}\frac{\partial f(\hat{z})}{\partial \hat{z}_k}dt & \text{(no jump during } [t, t+dt); \text{ probability} = 1 - \hat{\nu}(t)dt) \\ f(\hat{z}(t) + \boldsymbol{h}) - f(\hat{z}(t)) & \text{(jump in } [t, t+dt); \text{ probability} = \hat{\nu}(t)dt) \end{cases} \tag{A1}$$

with jump size vector $\boldsymbol{h}$ and Hawkes intensity $\hat{\nu}$, defined by

$$\boldsymbol{h} := \left(\frac{n_1}{\tau_1}, \frac{n_2}{\tau_2}, \dots, \frac{n_K}{\tau_K}\right)^{\mathrm{T}}, \quad \hat{\nu}(t) := \nu_0 + \sum_{k=1}^{K} \hat{z}_k(t). \tag{A2}$$

Taking the ensemble average of both sides of (A1) and after partial integration of the left-hand side, we get

$$\int d\boldsymbol{z} f(\boldsymbol{z})\frac{\partial P_t(\boldsymbol{z})}{\partial t}dt = \int d\boldsymbol{z}\left[-\sum_{k=1}^{K}\frac{z_k}{\tau_k}\frac{\partial f(\boldsymbol{z})}{\partial z_k}dt + \left(\nu_0 + \sum_{k=1}^{K}z_k\right)dt\{f(\boldsymbol{z}+\boldsymbol{h}) - f(\boldsymbol{z})\}\right]P_t(\boldsymbol{z}). \tag{A3}$$

After partial integration of the right-hand side and making the change of variable $\boldsymbol{z} + \boldsymbol{h} \to \boldsymbol{z}$, we obtain

$$\int d\boldsymbol{z} f(\boldsymbol{z})\frac{\partial P_t(\boldsymbol{z})}{\partial t} = \int d\boldsymbol{z}\left[\sum_{k=1}^{K}\frac{\partial}{\partial z_k}\frac{z_k}{\tau_k}P(\boldsymbol{z}) + \left\{\nu_0 + \sum_{k=1}^{K}(z_k - h_k)\right\}P(\boldsymbol{z}-\boldsymbol{h}) - \left\{\nu_0 + \sum_{k=1}^{K}z_k\right\}P(\boldsymbol{z})\right]f(\boldsymbol{z}). \tag{A4}$$

Since this is an identify for an arbitrary $f(\boldsymbol{z})$, we obtain Eq. (27).

We derive the corresponding Laplace representation (31) as follows: Let us apply the Laplace transform to both sides of Eq. (27),

$$\mathcal{L}_K\left[\frac{\partial P_t(\boldsymbol{z})}{\partial t}\right] = \mathcal{L}_K\left[\sum_{k=1}^{K}\frac{\partial}{\partial z_k}\frac{z_k}{\tau_k}P(\boldsymbol{z}) + \left\{\nu_0 + \sum_{k=1}^{K}(z_k - h_k)\right\}P(\boldsymbol{z}-\boldsymbol{h}) - \left\{\nu_0 + \sum_{k=1}^{K}z_k\right\}P(\boldsymbol{z})\right]. \tag{A5}$$

The left-hand side is given by

$$\mathcal{L}_K\left[\frac{\partial P_t(\boldsymbol{z})}{\partial t}\right] = \frac{\partial \tilde{P}_t(\boldsymbol{s})}{\partial t}. \tag{A6}$$

For the right-hand side, let us consider the following two relations:

$$\mathcal{L}_K\left[\frac{\partial}{\partial z_k}\frac{z_k}{\tau_k}P(\boldsymbol{z})\right] = \int d\boldsymbol{z} e^{-\boldsymbol{s}\cdot\boldsymbol{z}}\frac{\partial}{\partial z_k}\frac{z_k}{\tau_k}P(\boldsymbol{z}) = \int \prod_{i|i\neq k}dz_i \int dz_k e^{-\boldsymbol{s}\cdot\boldsymbol{z}}\frac{\partial}{\partial z_k}\frac{z_k}{\tau_k}P(\boldsymbol{z})$$

$$= \int \prod_{i|i\neq k}dz_i\left\{\left[\frac{z_k}{\tau_k}P(\boldsymbol{z})\right]_{z_k=0}^{z_k=\infty} + s_k\int dz_k e^{-\boldsymbol{s}\cdot\boldsymbol{z}}\frac{z_k}{\tau_k}P(\boldsymbol{z})\right\} = s_k\int \prod_{i|i\neq k}dz_i \int dz_k e^{-\boldsymbol{s}\cdot\boldsymbol{z}}\frac{z_k}{\tau_k}P(\boldsymbol{z})$$

$$= s_k\int \prod_{i|i\neq k}dz_i\left(-\frac{1}{\tau_k}\frac{\partial}{\partial s_k}\right)\int dz_k e^{-\boldsymbol{s}\cdot\boldsymbol{z}}P(\boldsymbol{z}) = -\frac{s_k}{\tau_k}\frac{\partial}{\partial s_k}\tilde{P}_t(\boldsymbol{s}), \tag{A7}$$

where we have used the partial integration on the second line and have used the boundary condition (28) on the third line, and

$$\mathcal{L}_K\left[\left\{\nu_0 + \sum_{k=1}^{K}(z_k - h_k)\right\}P(\boldsymbol{z}-\boldsymbol{h})\right] = \int d\boldsymbol{z} e^{-\boldsymbol{s}\cdot\boldsymbol{z}}\left\{\nu_0 + \sum_{k=1}^{K}(z_k - h_k)\right\}P(\boldsymbol{z}-\boldsymbol{h})$$

$$= e^{-\boldsymbol{s}\cdot\boldsymbol{h}}\int d\boldsymbol{z} e^{-\boldsymbol{s}\cdot(\boldsymbol{z}-\boldsymbol{h})}\left\{\nu_0 + \sum_{k=1}^{K}(z_k - h_k)\right\}P(\boldsymbol{z}-\boldsymbol{h}) = e^{-\boldsymbol{s}\cdot\boldsymbol{h}}\int d\boldsymbol{z} e^{-\boldsymbol{s}\cdot\boldsymbol{z}}\left\{\nu_0 + \sum_{k=1}^{K}z_k\right\}P(\boldsymbol{z})$$

$$= e^{-\boldsymbol{s}\cdot\boldsymbol{h}}\left\{\nu_0 - \sum_{k=1}^{K}\frac{\partial}{\partial s_k}\right\}\int d\boldsymbol{z} e^{-\boldsymbol{s}\cdot\boldsymbol{z}}P(\boldsymbol{z}) = e^{-\boldsymbol{s}\cdot\boldsymbol{h}}\left\{\nu_0 - \sum_{k=1}^{K}\frac{\partial}{\partial s_k}\right\}\tilde{P}_t(\boldsymbol{s}), \tag{A8}$$

where we have applied the change of variable $\boldsymbol{z} - \boldsymbol{h} \to \boldsymbol{z}$ on the second line. By applying these two relations to the right-hand side of Eq. (A5), we obtain Eq. (31).

2. Derivation of Eq. (40)

The Hawkes intensity $\hat{v}$ is defined by

$$\hat{v}(t) := v_0 + \int_0^\infty \hat{z}(t, x)dx, \tag{A9}$$

in terms of the continuous field of excess intensities $\{\hat{z}(t, x)\}_{x \in (0,\infty)}$. For an arbitrary functional $f[\hat{z}]$, let us consider its stochastic time evolution:

$$df[\hat{z}] = f[\{\hat{z}(t + dt, x)\}_x] - f[\{\hat{z}(t, x)\}_x] = \begin{cases} -dt \int_0^\infty dx \frac{\hat{z}(t,x)}{x} \frac{\delta f[\hat{z}]}{\delta \hat{z}(x)} & \text{(no jump during } [t, t + dt); \text{ probability} = 1 - \hat{v}(t)dt) \\ f[\hat{z} + n/x] - f[\hat{z}] & \text{(jump in } [t, t + dt); \text{ probability} = \hat{v}(t)dt) \end{cases}, \tag{A10}$$

where we have used the the functional Taylor expansion

$$f[z + \eta] - f[z] = \sum_{k=0}^\infty \frac{1}{k!} \int dx_1 \ldots dx_k \frac{\delta^k f[z]}{\delta z(x_1) \ldots \delta z(x_k)} \eta(x_1) \ldots \eta(x_k) \tag{A11}$$

up to first order. Taking the ensemble average of both sides of (A10) yields

$$\int \mathcal{D}z f[z] \frac{\partial P_t[z]}{\partial t} dt = \int \mathcal{D}z \left[-\int_0^\infty dx \frac{z(x)}{x} \frac{\delta f[z]}{\delta z(x)} dt + \left(v_0 + \int_0^\infty z(x)dx \right) dt \{f[z + n/x] - f[z]\} \right] P_t[z]. \tag{A12}$$

By partial integration and with the change of variable $z + n/x \to z$, we obtain

$$\int \mathcal{D}z f[z] \frac{\partial P_t[z]}{\partial t} = \int \mathcal{D}z \left[\int_0^\infty dx \frac{\delta}{\delta z} \frac{z}{x} P_t[z] + \left\{ v_0 + \int_0^\infty \left(z - \frac{n}{x} \right) dx \right\} P_t[z - n/x] - \left\{ v_0 + \int_0^\infty z dx \right\} P[z] \right] f[z]. \tag{A13}$$

Since this is an identify for arbitrary $f[z]$, we obtain Eq. (40).

APPENDIX B: DERIVATIONS OF THE POWER LAW PDF OF HAWKES INTENSITIES FOR THE EXPONENTIAL MEMORY KERNEL

Here, we provide two different derivations of the power law PDF (53) of Hawkes intensities for the exponential memory kernel (13).

1. Introduction of a UV cutoff

We now investigate the steady solution of the master equation (19) for the probability density function (PDF) $P_t(z)$ of the excess intensity $\hat{z}$ (14), at the critical point $n = 1$.

Let us introduce a UV cutoff s_{uv} to address the singularity at $s \to 0$ so that we can express

$$\tilde{Q}_{ss}(s) \simeq \exp \left[-\frac{v_0}{\tau} \int_{s_{uv}}^s \frac{s ds}{e^{-s/\tau} - 1 + s/\tau} \right]$$
$$= \exp \left[-v_0(s - s_{uv}) - v_0 \tau \log \left(\frac{e^{-s/\tau} - 1 + s/\tau}{e^{-s_{uv}/\tau} - 1 + s_{uv}/\tau} \right) \right]. \tag{B1}$$

Recall that $\log \tilde{Q}_{ss}(s) = -s v_0 + \log \tilde{P}_{ss}(s)$ and $\tilde{P}_{ss}(s)$ is the Laplace transform of the steady state $\tilde{P}_{ss}(s) := \int_0^\infty dv e^{-sv} P_{ss}(z)$ of the master equation (19). The introduction of this UV cutoff s_{uv} amounts to introducing a cutoff in the memory function $h(t)$ at large timescale [i.e., there exists t_{cut} such that $h(t)$ is negligible for $t > t_{cut}$]. The validity of this approximation is confirmed by considering the time-dependent solution (see Sec. IV B 2). At the critical point $n = 1$, it has

an asymptotic form for small $s_{uv} < s \ll \tau$,

$$\log \tilde{Q}_{ss}(s) \simeq -\frac{v_0}{\tau} \int_{s_{uv}}^s \frac{s ds}{e^{-s/\tau} - 1 + s/\tau}$$
$$\sim -\frac{v_0}{\tau} \int_{s_{uv}}^s \frac{2\tau^2 ds}{s} = -2 v_0 \tau \log \frac{s}{s_{uv}}, \tag{B2}$$

which implies the power-law relation for the tail distribution:

$$P_{ss}(v) \propto v^{-1 + 2 v_0 \tau} \tag{B3}$$

for $0 \ll v \ll v_{max} := 1/s_{uv}$.

2. Kramers-Moyal approach

We can derive relation (53) using the Kramers-Moyal expansion of the master equation (19). Let us consider the expansion

$$(v_0 + z - n/\tau)P_t(z - n/\tau) = \sum_{k=0}^\infty \frac{1}{k!} \left(\frac{-n}{\tau} \right)^k \frac{\partial^k}{\partial z^k} (v_0 + z) P_t(z). \tag{B4}$$

By truncating the series at the second order, we obtain the Fokker-Planck equation at the critical point $n = 1$ in the steady state:

$$\left[-\frac{v_0}{\tau} \frac{\partial}{\partial z} + \frac{1}{2\tau^2} \frac{\partial^2}{\partial z^2} (v_0 + z) \right] P_{ss}(z) \simeq 0, \tag{B5}$$

for $z \to \infty$. We thus obtain an asymptotic formula

$$P_{ss}(z) \sim (v_0 + z)^{-1 + 2 v_0 \tau}, \quad \text{for large } z. \tag{B6}$$

This solution is consistent with the truncation of the Kramers-Moyal series at the second order, which consists in removing negligible higher order terms. For large $l \geqslant 3$, indeed, we obtain

$$\left| \frac{\partial^2}{\partial z^2} (\nu_0 + z) P_{ss}(z) \right| \gg \left| \frac{\partial^l}{\partial z^l} (\nu_0 + z) P_{ss}(z) \right|, \quad \text{for large } z.$$

(B7)

APPENDIX C: ELEMENTARY SUMMARY OF THE METHOD OF CHARACTERISTICS

The method of characteristics is a standard method to solve first-order PDEs. Here we focus on linear first-order PDEs that are relevant to the derivation of the PDF of Hawkes intensities. Let us consider the following PDE:

$$a(x, y, z) \frac{\partial z(x, y)}{\partial x} + b(x, y, z) \frac{\partial z(x, y)}{\partial y} = c(x, y, z). \quad \text{(C1)}$$

According to the method of characteristics, we consider the corresponding Lagrange-Charpit equations:

$$\frac{dx}{dl} = -a(x, y, z), \quad \text{(C2a)}$$

$$\frac{dy}{dl} = -b(x, y, z), \quad \text{(C2b)}$$

$$\frac{dz}{dl} = -c(x, y, z), \quad \text{(C2c)}$$

with the parameter l encoding the position along the characteristic curves. These equations are equivalent to an invariant form in terms of l

$$\frac{dx}{a(x, y, z)} = \frac{dy}{b(x, y, z)} = \frac{dz}{c(x, y, z)}. \quad \text{(C3)}$$

Let us write their formal solutions as $C_1 = F_1(x, y, z)$ and $C_2 = F_2(x, y, z)$ with constants of integration C_1 and C_2. The general solution of the original PDE (C1) is given by

$$\phi(F_1(x, y, z), F_2(x, y, z)) = 0 \quad \text{(C4)}$$

with an arbitrary function $\phi(C_1, C_2)$, which is determined by the initial or boundary condition of the PDE (C1). This method can be readily generalized to systems with many variables.

APPENDIX D: ANALYTICAL DERIVATION OF SOME MAIN PROPERTIES OF $\mathcal{F}(s)$ (58)

Here, we derive properties $(\alpha 1)$–$(\alpha 6)$ of $\mathcal{F}(s)$ (58). First, the following relations hold true:

$$s > \frac{\tau}{n} \log n \implies \frac{d}{ds} (e^{-ns/\tau} - 1 + s/\tau)$$
$$= -\frac{n}{\tau} e^{-ns/\tau} + \frac{1}{\tau} > 0 \quad \text{(D1)}$$

and

$$\lim_{s \to \infty} (e^{-ns/\tau} - 1 + s/\tau) = \infty. \quad \text{(D2)}$$

These relations guarantee that there exists $s_0 > 0$ such that $e^{-ns/\tau} - 1 + s/\tau > 0$ for $s > s_0$. Therefore, the integrand is

positive definite

$$\frac{1}{e^{-ns/\tau} - 1 + s/\tau} > 0 \quad \text{(D3)}$$

for $s > s_0$ by choosing an appropriate positive s_0. Since the integrand is positive definite, the statement $(\alpha 1)$ is proved. As a corollary of $(\alpha 1)$, the statement $(\alpha 2)$ is proved.

We next study $\mathcal{F}(s)$ in the subcritical regime ($n < 1$). For $n < 1$, the statement $(\alpha 3)$ is true because $e^{-ns/\tau} - 1 + s/\tau > 0$ for any $s > 0$ and $(\tau/n) \log n < 0$. The statement $(\alpha 4)$ is true because, for $0 < s < s_0 \ll \tau/n$, we obtain

$$\mathcal{F}(s) \simeq \frac{\tau}{1-n} \log \frac{s}{s_0} \to -\infty \quad \text{(D4)}$$

for $s \to +0$. The statement $(\alpha 5)$ is correct because

$$\mathcal{F}(s) \simeq \int_{c_0}^{s} \frac{ds'}{s'/\tau} + c_1 = \tau \log s + c_1 \to +\infty \quad \text{(D5)}$$

for $s \to +\infty$ with constants c_0 and c_1. As a corollary of $(\alpha 1)$, $(\alpha 4)$, and $(\alpha 5)$, the statement $(\alpha 6)$ is proved.

APPENDIX E: ANALYTICAL DERIVATION OF THE TIME-DEPENDENT SOLUTION (63)

1. Consistency check 1: Convergence to the steady solution

Let us check that the time-dependent solution (63) is consistent with the steady solution (49) for $n < 1$. To prove this, it is sufficient to show that

$$\lim_{x \to \infty} \mathcal{H}(x) = 0. \quad \text{(E1)}$$

Using $(\alpha 1)$, $(\alpha 4)$, and $(\alpha 5)$, we obtain

$$\lim_{x \to \infty} S(x) = \lim_{x \to -\infty} \mathcal{F}^{-1}(x) = 0, \quad \text{(E2)}$$

and thus

$$\lim_{x \to \infty} \mathcal{H}(x) = \lim_{S \to 0} \left[\log \tilde{P}_{t=0}(S) + \frac{\nu_0}{\tau} \int_0^S \frac{s \, ds}{e^{-ns/\tau} - 1 + s/\tau} \right]$$
$$= 0. \quad \text{(E3)}$$

2. Consistency check 2: Relaxation dynamics of the average $\hat{\nu}(t)$ at finite times

Let us now study the dynamics of the average intensity $\hat{\nu}(t)$ via the time-dependent formula (63). Below the critical point, we can use the renormalized expression (61). Then the integral $\mathcal{F}(s)$ can be asymptotically evaluated for small $0 \leqslant s \ll \tau/n$:

$$\mathcal{F}(s) \simeq \frac{\tau}{1-n} \log s \to -\infty \quad (s \to 0). \quad \text{(E4)}$$

This means that the argument $x(t, s) = t - \mathcal{F}(s)$ shows the divergence

$$x(t, s) = t - \mathcal{F}(s) \simeq t - \frac{\tau}{1-n} \log s \to +\infty. \quad \text{(E5)}$$

From Eq. (E4), the inverse function shows the asymptotic behavior for large x:

$$S(x) = \mathcal{F}^{-1}(-x) \simeq \exp\left[-\frac{1-n}{\tau} x \right]. \quad \text{(E6)}$$

By substituting the relation (E5), we obtain

$$
\begin{aligned}
S(x(t,s)) &= \mathcal{F}^{-1}(-x(t,s)) \\
&\simeq \exp\left[-\frac{1-n}{\tau}\left(t - \frac{\tau}{1-n}\log s\right)\right] \\
&= se^{-(1-n)t/\tau}.
\end{aligned}
\tag{E7}
$$

We now assume the initial condition $\hat{v}(0) = v_{\text{ini}}$, or equivalently $\log \tilde{P}_{t=0}(s) = -v_{\text{ini}}s$. From Eq. (63), we thus obtain the relaxation dynamics for the tail $s \simeq 0$,

$$
\begin{aligned}
\log \tilde{P}_t(s) &\simeq -se^{-(1-n)t/\tau}v_{\text{ini}} - \frac{v_0}{\tau}\int_{se^{-(1-n)t/\tau}}^{s}\frac{s\,ds}{e^{-ns/\tau}-1+s/\tau} \\
&\simeq -\left[v_{\text{ini}}e^{-(1-n)t/\tau} + \frac{v_0(1-e^{-nt/\tau})}{1-n}\right]s,
\end{aligned}
\tag{E8}
$$

which means that the average of $\hat{v}(t)$ is given by

$$
\begin{aligned}
\langle\hat{v}(t)\rangle &= -\frac{d}{ds}\log \tilde{P}_t(s)\Big|_{s=0} \\
&= v_{\text{ini}}e^{-(1-n)t/\tau} + \frac{v_0(1-e^{-nt/\tau})}{1-n}.
\end{aligned}
\tag{E9}
$$

3. Derivation of the asymptotic formula (68)

Let us derive the asymptotic relaxation formula (68) for sufficiently large t, satisfying

$$
t \gg \mathcal{F}(s) \tag{E10}
$$

for a given s. Under such a condition, the asymptotic relation for the inverse function $\mathcal{F}^{-1}(-x(t,s))$ is available as Eq. (E6) with $x(t,s) = t - \mathcal{F}(s)$. We then obtain

$$
\begin{aligned}
\mathcal{F}^{-1}(-x(t,s)) &\simeq \exp\left[-\frac{1-n}{\tau}\left(t - \mathcal{F}_R(s) - \frac{\tau}{1-n}\log s\right)\right] \\
&= s\exp\left[-\frac{1-n}{\tau}(t - \mathcal{F}_R(s))\right]
\end{aligned}
\tag{E11}
$$

for sufficiently large t. By substituting this into Eq. (63), we obtain Eq. (68).

APPENDIX F: PROOFS OF MATHEMATICAL PROPERTIES OF H (109)

Here, we summarize the proofs of the main mathematical properties of H (109) for arbitrary values of K.

1. Proof of that its eigenvalues are real

All eigenvalues of H are real numbers for the following reasons. H can be symmetrized as $\bar{H}$, defined by

$$
\bar{H} := AHA^{-1} = \begin{pmatrix}
\frac{1-n_1}{\tau_1}, & \sqrt{\frac{n_1 n_2}{\tau_1 \tau_2}}, & \cdots & \sqrt{\frac{n_1 n_K}{\tau_1 \tau_K}} \\
\sqrt{\frac{n_2 n_1}{\tau_2 \tau_1}}, & \frac{1-n_2}{\tau_2}, & \cdots & \sqrt{\frac{n_2 n_K}{\tau_2 \tau_K}} \\
\vdots & \vdots & \ddots & \vdots \\
\sqrt{\frac{n_K n_1}{\tau_K \tau_1}}, & \sqrt{\frac{n_K n_2}{\tau_K \tau_2}}, & \cdots & \frac{1-n_K}{\tau_K}
\end{pmatrix},
$$

$$
A := \begin{pmatrix}
\sqrt{\frac{n_1}{\tau_1}}, & 0, & \cdots & 0 \\
0, & \sqrt{\frac{n_2}{\tau_2}}, & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0, & 0, & \cdots & \sqrt{\frac{n_2}{\tau_2}}
\end{pmatrix}.
\tag{F1}
$$

Indeed, by representing all the matrices by their elements $\bar{H} := (\bar{H}_{ij})$, $H := (H_{ij})$, and $A := A_{ij}$, we obtain

$$
\begin{aligned}
\bar{H}_{ij} &= \sum_{k,l}A_{ik}H_{kl}A_{lj}^{-1} = \sum_{k,l}\sqrt{\frac{n_i}{\tau_i}}\delta_{ik}\left(\frac{\delta_{kl}}{\tau_k} - \frac{n_l}{\tau_l}\right)\sqrt{\frac{\tau_j}{n_j}}\delta_{lj} \\
&= \frac{\delta_{ij} - \sqrt{n_i n_j}}{\sqrt{\tau_i \tau_j}}.
\end{aligned}
\tag{F2}
$$

We therefore obtain

$$
He_i = \lambda_i e_i \iff \bar{H}(Ae_i) = \lambda_i(Ae_i), \tag{F3}
$$

implying that any eigenvalue of H is the same as that of $\bar{H}$. Because $\bar{H}$ is a symmetric matrix, all the eigenvalues of $\bar{H}$ are real. Therefore, all the eigenvalues of H are also real.

2. Determinant

Here, we derive the determinant $\det H$ for arbitrary values of K. Let us recall the following identities, showing the invariance of determinants:

$$
\det H = \det\begin{pmatrix}
a_1 \\
a_2 \\
\vdots \\
a_j \\
\vdots \\
a_K
\end{pmatrix} = \det\begin{pmatrix}
a_1 \\
a_2 \\
\vdots \\
a_j + ca_k \\
\vdots \\
a_K
\end{pmatrix}.
\tag{F4}
$$

This implies

$$
\begin{aligned}
\det H &= \det\begin{pmatrix}
a_1 \\
a_2 \\
a_3 \\
\vdots \\
a_K
\end{pmatrix} = \det\begin{pmatrix}
a_1 \\
a_2 - a_1 \\
a_3 \\
\vdots \\
a_K
\end{pmatrix} = \det\begin{pmatrix}
a_1 \\
a_2 - a_1 \\
a_3 - a_1 \\
\vdots \\
a_K
\end{pmatrix} \\
&= \cdots = \det\begin{pmatrix}
a_1 \\
a_2 - a_1 \\
a_3 - a_1 \\
\vdots \\
a_K - a_1
\end{pmatrix} := \det\begin{pmatrix}
a'_1 \\
a'_2 \\
a'_3 \\
\vdots \\
a'_K
\end{pmatrix}
\end{aligned}
\tag{F5}
$$

and

$$
\det \boldsymbol{H} = \det \begin{pmatrix} \boldsymbol{a}'_1 \\ \boldsymbol{a}'_2 \\ \vdots \\ \boldsymbol{a}'_K \end{pmatrix} = \det \begin{pmatrix} \boldsymbol{a}'_1 + n_2\boldsymbol{a}'_2 \\ \boldsymbol{a}'_2 \\ \vdots \\ \boldsymbol{a}'_K \end{pmatrix} = \det \begin{pmatrix} \boldsymbol{a}'_1 + n_2\boldsymbol{a}'_2 + n_3\boldsymbol{a}'_3 \\ \boldsymbol{a}'_2 \\ \vdots \\ \boldsymbol{a}'_K \end{pmatrix} = \cdots = \det \begin{pmatrix} \boldsymbol{a}'_1 + \sum_{k=2}^{K} n_k\boldsymbol{a}'_k \\ \boldsymbol{a}'_2 \\ \vdots \\ \boldsymbol{a}'_K \end{pmatrix} \tag{F6}
$$

with constants $\{n_k\}_k$. Using these relations, the determinant of $\boldsymbol{H}$ is given by

$$
\det \boldsymbol{H} = \det \begin{pmatrix} -n_1/\tau_1 + 1/\tau_1, & -n_2/\tau_2, & \ldots, & -n_K/\tau_K \\ -n_1/\tau_1, & -n_2/\tau_2 + 1/\tau_2, & \ldots, & -n_K/\tau_K \\ \vdots & \vdots & \ddots & \vdots \\ -n_1/\tau_1, & -n_2/\tau_2, & \ldots, & -n_K/\tau_K + 1/\tau_K \end{pmatrix} \begin{matrix} \leftarrow \boldsymbol{a}_1 \\ \leftarrow \boldsymbol{a}_2 \\ \vdots \\ \leftarrow \boldsymbol{a}_K \end{matrix}
$$

$$
= \det \begin{pmatrix} (1 - n_1)/\tau_1, & -n_2/\tau_2, & \ldots & -n_K/\tau_K \\ -1/\tau_1, & 1/\tau_2, & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1/\tau_1, & 0, & \ldots & 1/\tau_K \end{pmatrix} \begin{matrix} \leftarrow \boldsymbol{a}'_1 & = \boldsymbol{a}_1 \\ \leftarrow \boldsymbol{a}'_2 & = \boldsymbol{a}_2 & - & \boldsymbol{a}_1 \\ \vdots \\ \leftarrow \boldsymbol{a}'_K & = \boldsymbol{a}_K & - & \boldsymbol{a}_1 \end{matrix}
$$

$$
= \det \begin{pmatrix} (1 - \sum_{k=1}^{K} n_k)/\tau_1, & 0, & \ldots & 0 \\ -1/\tau_1, & 1/\tau_2, & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1/\tau_1, & 0, & \ldots & 1/\tau_K \end{pmatrix} \begin{matrix} \leftarrow \boldsymbol{a}''_1 & = \boldsymbol{a}'_1 & + & \sum_{k=2}^{K} n_k\boldsymbol{a}'_k \\ \leftarrow \boldsymbol{a}''_2 & = \boldsymbol{a}'_2 \\ \vdots \\ \leftarrow \boldsymbol{a}''_K & = \boldsymbol{a}'_K \end{matrix}
$$

$$
= \frac{1 - \sum_{k=1}^{K} n_k}{\tau_1 \ldots \tau_K}. \tag{F7}
$$

3. Inverse matrix

Here we derive the inverse matrix of $\boldsymbol{H}$ for arbitrary values of K. The inverse matrix is derived from the method of row reduction:

$$
\begin{pmatrix} -n_1/\tau_1 + 1/\tau_1, & -n_2/\tau_2, & \ldots & -n_K/\tau_K & \bigg| & 1, & 0, & \ldots, & 0 \\ -n_1/\tau_1, & -n_2/\tau_2 + 1/\tau_2, & \ldots & -n_K/\tau_K & \bigg| & 0, & 1, & \ldots, & 0 \\ \vdots & \vdots & \ddots & \vdots & \bigg| & \vdots & \vdots & \ddots & \vdots \\ -n_1/\tau_1, & -n_2/\tau_2, & \ldots, & -n_K/\tau_K + 1/\tau_K & \bigg| & 0, & 0, & \ldots, & 1 \end{pmatrix} \begin{matrix} \leftarrow \boldsymbol{b}_1 \\ \leftarrow \boldsymbol{b}_2 \\ \vdots \\ \leftarrow \boldsymbol{b}_K \end{matrix}
$$

$$
\rightarrow \begin{pmatrix} (1 - n_1)/\tau_1, & -n_2/\tau_2, & \ldots & -n_K/\tau_K & \bigg| & 1, & 0, & \ldots & 0 \\ -1/\tau_1, & 1/\tau_2, & \ldots & 0 & \bigg| & -1, & 1, & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \bigg| & \vdots & \vdots & \ddots & \vdots \\ -1/\tau_1, & 0, & \ldots & 1/\tau_K & \bigg| & -1, & 0, & \ldots & 1 \end{pmatrix} \begin{matrix} \leftarrow \boldsymbol{b}'_1 & = \boldsymbol{b}_1 \\ \leftarrow \boldsymbol{b}'_2 & = \boldsymbol{b}_2 & - & \boldsymbol{b}_1 \\ \vdots \\ \leftarrow \boldsymbol{b}'_K & = \boldsymbol{b}_K & - & \boldsymbol{b}_1 \end{matrix}
$$

$$
\rightarrow \begin{pmatrix} (1 - n)/\tau_1, & 0, & \ldots & 0 & \bigg| & 1 - \sum_{k=2}^{K} n_k, & n_2, & \ldots & n_K \\ -1/\tau_1, & 1/\tau_2, & \ldots & 0 & \bigg| & -1, & 1, & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \bigg| & \vdots & \vdots & \ddots & \vdots \\ -1/\tau_1, & 0, & \ldots & 1/\tau_K & \bigg| & -1, & 0, & \ldots & 1 \end{pmatrix} \begin{matrix} \leftarrow \boldsymbol{b}''_1 & = \boldsymbol{b}'_1 & + & \sum_{k=2}^{K} n_k\boldsymbol{b}'_k \\ \leftarrow \boldsymbol{b}''_2 & = \boldsymbol{b}'_2 \\ \vdots \\ \leftarrow \boldsymbol{b}''_K & = \boldsymbol{b}'_K \end{matrix}
$$

$$
\rightarrow \begin{pmatrix} 1, & 0, & \ldots & 0 & \bigg| & \tau_1 + \tau_1 n_1/(1 - n), & \tau_1 n_2/(1 - n), & \ldots & \tau_1 n_K/(1 - n) \\ -\tau_2/\tau_1, & 1, & \ldots & 0 & \bigg| & -\tau_2, & \tau_2, & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \bigg| & \vdots & \vdots & \ddots & \vdots \\ -\tau_K/\tau_1, & 0, & \ldots & 1 & \bigg| & -\tau_K, & 0, & \ldots & \tau_K \end{pmatrix} \begin{matrix} \leftarrow \boldsymbol{b}'''_1 & = \frac{\tau_1}{(1 - n)}\boldsymbol{b}''_1 \\ \leftarrow \boldsymbol{b}'''_2 & = \tau_2\boldsymbol{b}''_2 \\ \vdots \\ \leftarrow \boldsymbol{b}'''_K & = \tau_K\boldsymbol{b}''_K \end{matrix}
$$

$$\rightarrow \begin{pmatrix} 1, & 0, & \dots & 0 \\ 0, & 1, & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \dots & 1 \end{pmatrix} \begin{array}{l} \tau_1 + \tau_1 n_1/(1-n), & \tau_1 n_2/(1-n), & \dots & \tau_1 n_K/(1-n) \\ \tau_2 n_1/(1-n), & \tau_2 + \tau_2 n_2/(1-n), & \dots & \tau_2 n_K/(1-n) \\ \vdots & \vdots & \ddots & \vdots \\ \tau_K n_1/(1-n), & \tau_K n_2/(1-n), & \dots & \tau_K + \tau_K n_K/(1-n) \end{array} \begin{array}{l} \leftarrow \boldsymbol{b}_1'''' & = \boldsymbol{b}_1''' \\ \leftarrow \boldsymbol{b}_2'''' & = \boldsymbol{b}_2''' + \frac{\tau_2}{\tau_1}\boldsymbol{b}_1''' \\ \vdots \\ \leftarrow \boldsymbol{b}_K'''' & = \boldsymbol{b}_K''' + \frac{\tau_2}{\tau_1}\boldsymbol{b}_1''' \end{array}$$

$$\tag{F8}$$

which implies

$$\boldsymbol{H}^{-1} = \begin{pmatrix} \tau_1 + \tau_1 n_1/(1-n), & \tau_1 n_2/(1-n), & \dots & \tau_1 n_K/(1-n) \\ \tau_2 n_1/(1-n), & \tau_2 + \tau_2 n_2/(1-n), & \dots & \tau_2 n_K/(1-n) \\ \vdots & \vdots & \ddots & \vdots \\ \tau_K n_1/(1-n), & \tau_K n_2/(1-n), & \dots & \tau_K + \tau_K n_K/(1-n) \end{pmatrix} \tag{F9}$$

or equivalently

$$H_{ij}^{-1} = \tau_i \delta_{ij} + \frac{\tau_i n_j}{1-n} \tag{F10}$$

in the representation by matrix elements. As a check of the above calculation, we can directly confirm the following relation, defining the inverse matrix:

$$\boldsymbol{H}\boldsymbol{H}^{-1} = \boldsymbol{I} \iff \sum_{j=1}^{K} H_{ij} H_{jk}^{-1} = \sum_{j=1}^{K} \left(-\frac{n_j}{\tau_j} + \frac{1}{\tau_j}\delta_{ij} \right) \left(\tau_j \delta_{jk} + \frac{\tau_j n_k}{1-n} \right) = \delta_{ik}. \tag{F11}$$

The inverse matrix has a singularity at $n = 1$, which corresponds to the critical regime of the Hawkes process.

APPENDIX G: PROOFS OF THE MATHEMATICAL PROPERTIES OF $H(x, x')$ (129)

1. Proof that the eigenvalues are real

Considering the analogy to the eigenvalue problem of the finite-dimensional matrix $\boldsymbol{H}$ (109), it is obvious by taking the continuous limit that all the eigenvalues of $H(x, x')$ are positive. As an Appendix, we remark that all eigenvalues can be proved real for $H(x, x')$ by making some specific assumptions. For example, let us assume that $n(x)$ has a finite cutoff, such that

$$n(x) = \begin{cases} \tilde{n}(x) & (x < \tau^*) \\ 0 & (x \geqslant \tau^*) \end{cases} \tag{G1}$$

with a positive continuous function $\tilde{n}(x) > 0$ and a cutoff $\tau^* > 0$. In this case, the eigenvalue problem for $H(x, x')$ can be rewritten as

$$\int_0^\infty dx' H(x, x') e(x'; \lambda) = \int_0^{\tau^*} H(x, x') e(x'; \lambda) = \lambda e(x; \lambda). \tag{G2}$$

While $H(x, x')$ itself is not a symmetric kernel, $H(x, x')$ can be symmetrized by introducing

$$\bar{H}(x, x') := \frac{\delta(x - x') - \sqrt{n(x)n(x')}}{\sqrt{xx'}}, \tag{G3}$$

such that

$$\int_0^{\tau^*} dx' \bar{H}(x, x') \bar{e}(x; \lambda) = \lambda \bar{e}(x; \lambda), \quad \bar{e}(x; \lambda) := \sqrt{\frac{n(x)}{x}} e(x; \lambda) \tag{G4}$$

or equivalently,

$$\int_0^{\tau^*} dx' \sqrt{\frac{n(x)n(x')}{xx'}} \bar{e}(x; \lambda) = (1/x - \lambda) \bar{e}(x; \lambda). \tag{G5}$$

This implies that all the eigenvalues of $H(x, x')$ are identical to those of $\bar{H}(x, x')$. Since Eq. (G5) is a homogeneous Fredholm integral equation of the second kind with a continuous and symmetric kernel $\sqrt{n(x)n(x')/(xx')}$ and with a finite interval $[0, \tau^*]$, all the eigenvalues of $\bar{H}(x, x')$ are real according to the Hilbert-Schmidt theory [69]. Therefore, all the eigenvalues of $H(x, x')$ are also real.

2. Inverse matrix

The inverse matrix of $H(x, x')$ is given by

$$H^{-1}(x, x') := x \left\{ \delta(x - x') + \frac{n(x')}{1-n} \right\}. \tag{G6}$$

Indeed, we verify that

$$\int_0^\infty dx' H(x, x') H^{-1}(x', x'')$$
$$= \int_0^\infty dx' \frac{\delta(x - x') - n(x')}{x'} x' \left\{ \delta(x' - x'') + \frac{n(x'')}{1-n} \right\}$$
$$= \delta(x - x''). \tag{G7}$$

The inverse matrix has a singularity at $n = 1$, corresponding to the critical regime of the Hawkes process.

[1] A. Hawkes, J. Royal Stat. Soc., Ser. B **33**, 438 (1971).

[2] A. Hawkes, Biometrika **58**, 83 (1971).

[3] A. Hawkes and D. Oakes, J. Appl. Prob. **11**, 493 (1974).

[4] Y. Y. Kagan and L. Knopoff, J. Geophys. Res. **86**, 2853 (1981).

[5] Y. Y. Kagan and L. Knopoff, Science **236**, 1563 (1987).

[6] Y. Ogata, J. Am. Stat. Assoc. **83**, 9 (1988).

[7] Y. Ogata, Pure Appl. Geophys. **155**, 471 (1999).

[8] A. Helmstetter and D. Sornette, J. Geophys. Res. **107**, 2237 (2002).

[9] S. Nandan, G. Ouillon, D. Sornette, and S. Wiemer, Seismol. Res. Lett. **90**, 1650 (2019).

[10] A. G. Hawkes, Quant. Finance **18**, 193 (2018).

[11] V. Filimonov and D. Sornette, Phys. Rev. E **85**, 056108 (2012).

[12] V. Filimonov and D. Sornette, Quant. Finance **15**, 1293 (2015).

[13] S. Wheatley, A. Wehrli, and D. Sornette, Quant. Finance **19**, 1165 (2019).

[14] E. Bacry, I. Mastromatteo, and J.-F. Muzy, Market Microstr. Liquidity **1**, 1550005 (2015).

[15] Q. Zhao, M. A. Erdogdu, H. Y. He, A. Rajaraman, and J. Leskovec, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2015), pp. 1513.

[16] D. Sornette, F. Deschatres, T. Gilbert, and Y. Ageon, Phys. Rev. Lett. **93**, 228701 (2004).

[17] R. Crane and D. Sornette, Proc. Nat. Acad. Sci. USA **105**, 15649 (2008).

[18] J. V. Escobar and D. Sornette, PLoS One **10**, e0116811 (2015).

[19] K. Kanazawa and D. Sornette, Phys. Rev. Lett., arXiv:2001.01195.

[20] A. I. Saichev and D. Sornette, Phys. Rev. E **70**, 046123 (2004).

[21] A. Saichev, A. Helmstetter, and D. Sornette, Pure Appl. Geophys. **162**, 1113 (2005).

[22] A. I. Saichev and D. Sornette, Eur. Phys. J. B **49**, 377 (2006).

[23] A. Saichev and D. Sornette, J. Geophys. Res. **112**, B04313 (2007).

[24] A. Saichev and D. Sornette, Phys. Rev. E **89**, 012104 (2014).

[25] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes* (Springer, Heidelberg, 2003), Vol. 1.

[26] J. Zhuang, Y. Ogata, and D. Vere-Jones, J. Am. Stat. Assoc. **97**, 369 (2002).

[27] A. Helmstetter and D. Sornette, Geophys. Res. Lett. **30**, 1576 (2003).

[28] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990); R. Metzler and J. Krafter, Phys. Rep. **339**, 1 (2000).

[29] R. Kubo, M. Toda, and N. Hashitsume, *Statsitical Physics II*, 2nd ed. (Springer-Verlag, Berlin, 1991).

[30] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).

[31] C. W. Gardiner, *Handbook of Stochastic Methods*, 4th ed. (Springer, Berlin, 2009).

[32] N. G. van Kampen, Stochastic Processes in Physics and Chemistry, 3rd ed. (Elsevier, Amsterdam, 2007).

[33] J.-P. Hansen and I. McDonald, *Theory of Simple Liquids*, 3rd ed. (Academic Press, Amsterdam, 2006).

[34] H. Spohn, Rev. Mod. Phys. **52**, 569 (1980).

[35] T. Li and M. G. Raizen, Ann. Phys. (Berlin) **525**, 281 (2013).

[36] T. Li, S. Kheifets, D. Medellin, and M. G. Raizen, Science **328**, 1673 (2010).

[37] R. Huang, I. Chavez, K. M. Taute, B. Lukić, S. Jeney, M. G. Raizen, and E.-L. Florin, Nat. Phys. **7**, 576 (2011).

[38] T. Franosch, M. Grimm, M. Belushkin, F. M. Mor, G. Foffi, L. Forró, and S. Jeney, Nature (London) **478**, 85 (2011).

[39] H. Mori, Prog. Theor. Phys. **33**, 423 (1965); **34**, 399 (1965).

[40] R. Morgado, F. A. Oliveira, G. G. Batrouni, and A. Hansen, Phys. Rev. Lett. **89**, 100601 (2002).

[41] J.-D. Bao, Y.-Z. Zhuo, F. A. Oliveira, and P. Hänggi, Phys. Rev. E **74**, 061111 (2006).

[42] M. H. Lee, Phys. Rev. B **26**, 2547 (1982).

[43] I. Goychuk, Phys. Rev. E **80**, 046125 (2009)

[44] R. Kupferman, J. Stat. Phys. **114**, 291 (2004).

[45] F. Marchesoni and P. Grigolini, J. Chem. Phys. **78**, 6287 (1983); M. Ferrario and P. Grigolini, J. Math. Phys. **20**, 2567 (1979).

[46] P. Hänggi and P. Jung, Adv. Chem. Phys. **89**, 239 (1995).

[47] J. Klafter and I. M. Sokolov, *First Steps in Random Walks: From Tools to Applications* (Oxford University Press, Oxford, UK, 2011).

[48] K. Kanazawa, T. G. Sano, T. Sagawa, and H. Hayakawa, Phys. Rev. Lett. **114**, 090601 (2015).

[49] K. Kanazawa, T. G. Sano, A. Cairoli, and A. Baule, Nature (London) **579**, 364 (2020).

[50] C. Jarzynski, J. Stat. Phys. **98**, 77 (2000).

[51] D. Oakes, J. Appl. Prob. **12**, 69 (1975).

[52] A. Dassios and H. Zhao, Adv. Appl. Probab. **43**, 814 (2011).

[53] D. Sornette and L. Knopoff, Bull. Seismol. Soc. Am. **87**, 789 (1997).

[54] J.-P. Bouchaud, J. Bonart, J. Donier, and M. Gould, *Trades, Quotes, and Prices* (Cambridge University Press, Cambridge, UK, 2018), Sec. 9.3.4.

[55] S. J. Hardiman, N. Bercot, and J.-P. Bouchaud, Eur. Phys. J. B **86**, 442 (2013).

[56] T. E. Harris, *The Theory of Branching Processes* (Springer, Berlin, 1963).

[57] A. Boumezoued, Adv. Appl. Probab. **48**, 463 (2016).

[58] G. I. Barenblatt, *Scaling, Self-similarity, and Intermediate Asymptotics* (Cambridge University Press, Cambridge, UK, 1996).

[59] A. Dassios and H. Zhao, Electron. Commun. Probab. **18**, 1 (2013).

[60] D. Harte, J. Stat. Software **35**, 1 (2010).

[61] K. Kawasaki, Prog. Theor. Phys. **51**, 1064 (1974); H. Leschke and M. Schmutz, Z. Phys. B **27**, 85 (1977); M. Doi, J. Phys. A: Math. Gen. **9**, 1465 (1976); **9**, 1479 (1976); L. Peliti, J. Phys. **46**, 1469 (1985).

[62] H. Risken, *The Fokker-Planck Equation*, 2nd ed. (Springer-Verlag, Berlin, 1989).

[63] G. Ódor, Rev. Mod. Phys. **76**, 663 (2004).

[64] Y. Ashida, Z. Gong, and M. Ueda, arXiv:2006.01837.

[65] H. Kleinert, *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics*, 5th ed. (World Scientific, Singapore, 2009).

[66] K. Kanazawa and D. Sornette, Phys. Rev. Lett. **125**, 138301 (2020).

[67] D. Sornette and G. Ouillon, Phys. Rev. Lett. **94**, 038501 (2005).

[68] V. A. Filimonov and D. Sornette, EPL **9**, 46003 (2011).

[69] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists* (Academic Press, San Diego, 1995).