DIFFUSION APPROXIMATION OF MULTI-CLASS HAWKES PROCESSES: THEORETICAL AND NUMERICAL ANALYSIS

Julien Chevallier ⓘ,* *Université Grenoble Alpes*
Anna Melnykova ⓘ,* ** *Université de Cergy-Pontoise and Université Grenoble Alpes*
Irene Tubikanec ⓘ,*** *Johannes Kepler University Linz*

Abstract

Oscillatory systems of interacting Hawkes processes with Erlang memory kernels were introduced by Ditlevsen and Löcherbach (*Stoch. Process. Appl.*, 2017). They are piecewise deterministic Markov processes (PDMP) and can be approximated by a stochastic diffusion. In this paper, first, a strong error bound between the PDMP and the diffusion is proved. Second, moment bounds for the resulting diffusion are derived. Third, approximation schemes for the diffusion, based on the numerical splitting approach, are proposed. These schemes are proved to converge with mean-square order 1 and to preserve the properties of the diffusion, in particular the hypoellipticity, the ergodicity, and the moment bounds. Finally, the PDMP and the diffusion are compared through numerical experiments, where the PDMP is simulated with an adapted thinning procedure.

Keywords: Piecewise deterministic Markov processes; Hawkes processes; stochastic differential equations; diffusion processes; neuronal models; numerical splitting schemes

2010 Mathematics Subject Classification: Primary 60H35; 65C20
Secondary 65C30; 60G55; 60J25

Introduction

Fast and accurate simulation of a biological neuronal network is one of the most extensively studied problems in computational neuroscience. The general goal is to understand how information is processed and transmitted in the brain. One of the widely used approaches is to assume that the spike occurrences in a network are described by a point process. Poisson processes, as 'memoryless' Markovian processes, can account for neither a refractory period between two consecutive spikes nor the interaction between neurons, and are thus no proper candidates. Therefore, it is common to model the neuronal activity with nonlinear Hawkes processes, which are self-interacting point processes with a memory [10, 11, 21, 36, 39]. The price to pay for using Hawkes processes to model spiking activity is that their investigation is more difficult, since the Markovian theory cannot be directly applied.

However, for a certain class of memory kernels, the so-called Erlang kernels, the dynamics of the point process can be described by a piecewise deterministic Markov process (PDMP),

whose dimension is determined by the 'memory length' of the underlying Hawkes process [15]. This PDMP, also called a 'Markovian cascade of successive memory terms' in the literature, is a convenient framework to study the long-time behaviour of the particle system. In particular, it has been proved that it is positive Harris recurrent and converges to its unique invariant measure exponentially fast in Wasserstein distance [16, Theorems 1 and 2].

This Markovian cascade and its associated point process can be simulated thanks to the thinning procedure [35], which is a common way to simulate general point processes even without any Markovian assumption. This procedure yields an exact simulation algorithm but is costly to compute, especially when the number of neurons is large. This results from the fact that the computation time scales linearly with the number of neurons.

In the brain, neurons are clustered in populations with similar behaviours (excitatory, inhibitory, etc.). For instance, this has been observed in the visual cortex of mammals [8]. When the network size grows, but the proportion of neurons in each population remains constant, the PDMP can be approximated by a stochastic differential equation (SDE). In other words, the diffusion approximation theory makes it possible to replace the stochastic term, described by jumps in the PDMP, by a multidimensional Brownian motion. Passing from a Hawkes process to a diffusion process substantially simplifies the analysis of the system behaviour. First of all, the SDE framework allows one to study the limit behaviour, the large deviations, and the influence of the memory order, and to determine the conditions in which the phase transition occurs [15, 27]. Moreover, in this manuscript, precise first and second moment bounds of the process are derived, based on the SDE formulation. Also, the simulation of the diffusion process is much less computationally expensive than that of the PDMP, especially when the number of neurons is large. This results from the fact that the execution time for the SDE does not depend on the number of neurons. However, the SDE cannot be solved explicitly, and thus the construction of a reliable approximation scheme is required.

Note that the main difficulty does not lie in the construction of convergent numerical schemes. For example, standard methods such as the Euler–Maruyama or Milstein schemes converge in the mean-square sense when the time discretization step tends to zero. In practice, however, the solution is approximated with a strictly positive time step. As a consequence, even if the discrete solution is known to converge to the continuous process as the time step tends to zero, it does not follow that both processes share the same properties for a fixed discretization step. Thus, the approximation scheme should not be used to study the behaviour of the original model without further analysis of its qualitative properties. Constructing approximation schemes, which are not only convergent, but also preserve the properties of the model, constitutes the main difficulty.

In our case, the first challenge is that the diffusion term of the SDE is degenerate and that frequently applied numerical schemes, such as the Euler–Maruyama method, do not preserve the 'propagation-of-noise property' (formally known as *hypoellipticity*). Second, standard integrators may also fail to preserve second moment properties (see [30]), especially when the equation describes oscillatory dynamics, which is the case here. For example, in [20] it is proved that the Euler–Maruyama method does not preserve the second moment of linear stochastic oscillators. It is expected that this and similar negative results also extend to higher-dimensional and nonlinear stochastic oscillators; see, e.g., [3]. Even if higher-order Taylor approximation schemes may solve the problem of degenerate noise structure, they have two major drawbacks. They depend heavily on the dimension of the system (which is determined by a parameter in our model), and they commonly fail to preserve ergodic properties.

To overcome these problems, we construct numerical schemes based on the so-called splitting approach. This approach was first developed for ordinary differential equations (ODEs).

We refer to [4, 31] for an exhaustive discussion. For an extension to SDEs, see, e.g., [2, 3, 5, 24, 25, 33, 34, 37, 40]. The main idea of the numerical splitting approach is to decompose the system into explicitly solvable subequations and to find a proper composition of the derived explicit solutions. Such methods usually preserve the properties of the underlying model through the explicitly solved subparts.

The main contributions of this work can be divided into three steps. First, a strong error bound between the PDMP and the stochastic diffusion is proved. This complements the results presented in [15, 27]. Second, moment bounds of order one and two for the stochastic diffusion are derived. Third, simulation algorithms for the diffusion and the PDMP are provided. For the diffusion, two splitting schemes, based on the Lie–Trotter and the Strang approach [31, 41], are proposed. They are proved to converge with order one in the mean-square sense. Moreover, they are proved to preserve the ergodic property of the continuous process and to accurately reconstruct the moment bounds obtained in the second step. The simulation method for the PDMP is exact and based on the thinning procedure. In order to apply this method, global and local upper bounds are obtained. Their performances, with respect to the parameters of the model, are discussed.

This paper is organized as follows. In Section 1, the finite particle system, the corresponding PDMP, and the main notation are introduced. Section 2 is devoted to the stochastic diffusion and its properties. Section 3 presents the approximation schemes for the stochastic diffusion. Section 4 describes the simulation algorithm for the PDMP. Finally, Section 5 provides a numerical study illustrating the theoretical results.

1. Model and notation

The system considered in this paper consists of several populations of neurons, each of them representing a different functional group of neurons (layers in the visual cortex, pools of excitatory and inhibitory neurons in a network, etc.). This system is described by a multivariate counting process, which counts the spike occurrences. In a certain setting, it can be approximated by a stochastic diffusion in the large-population limit [15]. The resulting diffusion is the subject of study in Section 2.

1.1. Finite particle system

Let us consider a network consisting of K large populations of neurons, where the number of neurons in the kth population is denoted by N_k and the total number of neurons in the network is $N = N_1 + \cdots + N_K$. Let $Z_t^{k,n}$ represent the number of spikes of the nth neuron belonging to the kth population during the time interval $[0, t]$. The family of (simple) counting processes $\{(Z_t^{k,n})_{t \geq 0}, \ 1 \leq k \leq K, \ 1 \leq n \leq N_k\}$ is characterized by the intensity processes $(\lambda^{k,n}(t))_{t \geq 0}$, which are defined through the relation

$$\mathbb{P}\big(Z_t^{k,n} \text{ has a jump in } (t, \, t + dt] | \mathcal{F}_t\big) = \lambda^{k,n}(t)dt,$$

where $\mathcal{F}_t$ contains the information about the processes $(Z_t^{k,n})_{t \geq 0}$ up to time t. The interested reader is referred to [14] for the existence and formal definition of these processes. The mean-field framework considered here corresponds to intensities $\lambda^{k,n}(t)$ given by

$$\lambda^{k,n}(t) = f_k \left(\sum_{l=1}^{K} \frac{1}{N_l} \sum_{1 \leq m \leq N_l} \int_{(0,t)} h_{kl}(t - s) dZ_s^{l,m} \right), \tag{1}$$

where $\{h_{kl}: \mathbb{R}_+ \to \mathbb{R}\}$ is a family of *synaptic weight functions*, which model the influence of population l on population k. When h_{kl} is positive, the influence is excitatory. Conversely, when h_{kl} is negative, the influence is inhibitory. The function $f_k: \mathbb{R} \to \mathbb{R}_+$ is the *spiking rate function* of population k. The expression 'mean-field framework' refers to the fact that the intensity $\lambda^{k,n}(t)$ depends on the whole system only through the 'mean-field' behaviour of each population, namely $\frac{1}{N_l} \sum_{1 \le m \le N_l} dZ_s^{l,m}$. Furthermore, letting $N \to \infty$ (while keeping K fixed), we assume that $N_k/N \to p_k > 0$ for all k.

Throughout the paper we assume that the functions f_k satisfy the following conditions:

(A) The spiking rate functions f_k are strictly positive, Lipschitz-continuous, non-decreasing, and such that $0 < f_k \le f_k^{\max}$ for $k = 1, \ldots, K$, where $f_k^{\max}$ is a finite constant.

In this paper, Erlang-type memory kernels and a cyclic feedback system of interactions are considered. This means that for each k, population k is influenced only by population $k+1$, where we identify $K + 1$ with 1. In this case, all the memory kernels are null except the ones given by

$$h_{kk+1}(t) = c_k e^{-\nu_k t} \frac{t^{\eta_k}}{\eta_k!}, \tag{2}$$

where $c_k = \pm 1$. This constant determines whether the population has an *inhibitory* ($c_k = -1$) or *excitatory* ($c_k = +1$) effect. The parameter $\eta_k \ge 1$ is an integer determining the memory order for the interaction function from population $k+1$ to population k.

The parameters η_k and ν_k determine, intuitively, the typical delay of interaction and its time width. The delay of the influence of population $k+1$ on population k attains its maximum η_{k+1}/ν_{k+1} units back in time, and its mean is $(\eta_{k+1} + 1)/\nu_{k+1}$. The larger this ratio is, the more important 'old' events are. When the ratio is fixed (equal to τ), but both η_k and ν_k tend to infinity, h_{kk+1} tends to a Dirac mass in τ. This means that only one specific moment in time is important. The interested reader is referred to [15] and [27] for more details.

In this paper we are interested in the processes $\{(\bar{X}_t^{k,1})_{t \ge 0}, \ 1 \le k \le K\}$, which are the arguments of the function f_k in Equation (1) and are defined by

$$\bar{X}_t^{k,1} = \frac{1}{N_{k+1}} \sum_{1 \le m \le N_{k+1}} \int_{(0,t)} h_{kk+1}(t-s) dZ_s^{k+1,m}. \tag{3}$$

When the memory kernels are given in the form (2), the processes defined in (3) can be obtained as marginals of the process $(\bar{X}_t)_{t \ge 0} = \{(\bar{X}_t^{k,j})_{t \ge 0}, \ 1 \le k \le K, \ 1 \le j \le \eta_k + 1\}$ which solves the following system of dimension $\kappa = \sum_{k=1}^{K} (\eta_k + 1)$:

$$\begin{cases} d\bar{X}_t^{k,j} = \left[-\nu_k \bar{X}_t^{k,j} + \bar{X}_t^{k,j+1} \right] dt & \text{for } j = 1, \ldots, \eta_k, \\ d\bar{X}_t^{k,\eta_k+1} = -\nu_k \bar{X}_t^{k,\eta_k+1} dt + c_k d\bar{Z}_t^{k+1}, \\ \bar{X}_0 = x_0 \in \mathbb{R}^\kappa, \end{cases} \tag{4}$$

where

$$\bar{Z}_t^{k+1} = \frac{1}{N_{k+1}} \sum_{n=1}^{N_{k+1}} Z_t^{k+1,n},$$

with each $Z_t^{k+1,n}$ jumping at rate $f(\bar{X}_{t-}^{k+1,1})$; see [15] for more insight. This type of equation is sometimes called a Markovian cascade in the literature [15]. Throughout the paper we refer to the solution of (4) as the PDMP.

The process $(\bar{X}_t)_{t\geq 0}$ summarizes and averages the influence of the past events. This process, along with the firing rate functions f_k, determines the dynamics of $(Z_t^{k,n})_{t\geq 0}$, described by its intensity (1).

From a modelling point of view, the process $(\bar{X}_t^{k,1})_{t\geq 0}$ can be roughly regarded as the voltage membrane potential of any neuron in population k. Then, the probability that a neuron will emit a spike is given as a function of its membrane potential. To summarize, the processes with coordinates $(k, 1)$ defined by (3) describe the membrane potential in each population, whereas the other coordinates represent higher levels of memory for the process.

Note that the model presented so far starts with empty memory. The right-hand side of (1) is equal to $f_k(0)$ at time $t = 0$ or equivalently $x_0 = 0_\kappa$. However, one could easily generalize this to any initial condition x_0 in $\mathbb{R}^\kappa$, as is done in the rest of the paper. Moreover, the interested reader is referred to [16], where a more general model is studied numerically and theoretically for $K = 1$ population.

1.2. Notation

Now we focus on the case of two interacting populations of neurons ($K = 2$), consisting of N_1 and N_2 neurons, respectively. Taking $K = 2$ allows for an investigation of the interactions between populations of different sizes while avoiding heavy notation. Throughout the paper the following notation is used: $\mathbb{O}_{n\times m}$ denotes an $(n \times m)$-dimensional zero matrix, and 0_n denotes an n-dimensional zero vector. It is convenient to rewrite the system (4) in the matrix–vector form

$$d\bar{X}_t = A\bar{X}_t dt + \Gamma\, d\bar{Z}_t, \qquad \bar{X}_0 = x_0 \in \mathbb{R}^\kappa, \tag{5}$$

with

- $A \in \mathbb{R}^{\kappa \times \kappa}$ defined as

$$A = \begin{pmatrix} A_{\nu_1} & \mathbb{O}_{(\eta_1+1)\times(\eta_2+1)} \\ \mathbb{O}_{(\eta_2+1)\times(\eta_1+1)} & A_{\nu_2} \end{pmatrix}, \tag{6}$$

where A_{ν_k} is an $(\eta_k + 1) \times (\eta_k + 1)$ tri-diagonal matrix with lower diagonal equal to 0_{η_k}, diagonal equal to $(-\nu_k, \ldots, -\nu_k)$ and upper diagonal equal to $(1, \ldots, 1)$,

- $\Gamma \in \mathbb{R}^{\kappa \times 2}$ having zero coefficients everywhere, except for $\Gamma_{\eta_1+1,2} = c_1$ and $\Gamma_{\kappa,1} = c_2$,

- and $\bar{Z}_t = \left(\bar{Z}_t^1, \bar{Z}_t^2\right)^T$.

Throughout the paper the following convention is made. The coordinates of a generic vector x in $\mathbb{R}^\kappa$ are denoted by either $(x_i)_{i=1,\ldots,\kappa}$ or $(x^{k,j})_{k=1,2; j=1,\ldots,\eta_k+1}$, with the relations $i = j$ if $k = 1$ and $i = \eta_1 + 1 + j$ if $k = 2$. The second notation is usually preferred since each population is easily identified by the index k. For some generic function $g : \mathbb{R}^\kappa \to \mathbb{R}^\kappa$, the upper indices are used as follows: $(g(x))^{k,j}$. Moreover, it is sometimes more natural to consider some generic $\mathbb{R}^\kappa$-valued process x_t population-wise. Thus, it is split into two components $x_t^1 = (x_t^{1,1}, \ldots, x_t^{1,\eta_1+1}) \in \mathbb{R}^{\eta_1+1}$ and $x_t^2 = (x_t^{2,1}, \ldots, x_t^{2,\eta_2+1}) \in \mathbb{R}^{\eta_2+1}$, such that $x_t = (x_t^1, x_t^2)^T \in \mathbb{R}^\kappa$.

2. The limiting stochastic diffusion

In [15] it is proved that the limit behaviour of (5) can be approximated by the stochastic diffusion process $X = (X^1, X^2)^T \in \mathbb{R}^\kappa$, which is obtained as the unique strong solution of the SDE

$$dX_t = (AX_t + B(X_t))dt + \frac{1}{\sqrt{N}}\sigma(X_t)dW_t, \qquad X_0 = x_0, \tag{7}$$

where $W = (W^1, W^2)^T$ is a two-dimensional Brownian motion, and $x_0 \in \mathbb{R}^\kappa$ is a deterministic initial condition. The nonlinear part of the drift term $B : \mathbb{R}^\kappa \to \mathbb{R}^\kappa$ is given by

$$B(X) = (B^1(X^2), B^2(X^1))^T, \tag{8}$$

where $B^1 : \mathbb{R}^{\eta_2+1} \to \mathbb{R}^{\eta_1+1}$ and $B^2 : \mathbb{R}^{\eta_1+1} \to \mathbb{R}^{\eta_2+1}$ read as

$$B^1(X^2) = (0, \ldots, 0, c_1 f_2(X^{2,1})), \qquad B^2(X^1) = (0, \ldots, 0, c_2 f_1(X^{1,1})).$$

The diffusion component $\sigma : \mathbb{R}^\kappa \to \mathbb{R}^{\kappa \times 2}$ is given by

$$\sigma(X) = \begin{pmatrix} \sigma^1(X^2) \\ \sigma^2(X^1) \end{pmatrix}, \tag{9}$$

where $\sigma^1 : \mathbb{R}^{\eta_2+1} \to \mathbb{R}^{(\eta_1+1)\times 2}$ and $\sigma^2 : \mathbb{R}^{\eta_1+1} \to \mathbb{R}^{(\eta_2+1)\times 2}$ read as

$$\sigma^1(X^2) = \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & \frac{c_1}{\sqrt{p_2}}\sqrt{f_2(X^{2,1})} \end{pmatrix}, \qquad \sigma^2(X^1) = \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ \frac{c_2}{\sqrt{p_1}}\sqrt{f_1(X^{1,1})} & 0 \end{pmatrix}.$$

In other words, the jump term $\Gamma \, d\bar{Z}$, which determines the dynamics of the PDMP given in (5), is replaced by the sum of a nonlinear drift and a diffusion term.

As N goes to infinity, the diffusion term in (7) vanishes and the SDE transforms into an ODE of the form

$$dU_t = (AU_t + B(U_t))dt, \qquad U_0 = x_0. \tag{10}$$

The focus of this paper lies in the theoretical and numerical relations between the PDMP and its stochastic diffusion approximation. Thus, we do not address the properties of the ODE (10) in this work; we refer to [15] for related qualitative features and convergence results.

The rest of this section is organized as follows. First, we investigate how accurately the stochastic diffusion approximates the dynamics of the point process, proving a strong error bound between the PDMP (5) and the solution of the SDE (7). We then study the properties of the SDE (7), focusing on moment bounds.

2.1. Strong error bound between the limiting stochastic diffusion and the PDMP

Any error bound of the diffusion approximation is determined by two facts, namely the approximation of a compensated Poisson process by a Brownian motion and the approximation of N_k by $p_k N$. We get rid of the second approximation by considering the SDE (7) with $p_k = N_k/N$ and denote the solution of this equation by Y. By choosing different notation we stress the fact that, in contrast to X, the solution Y depends on the exact number of neurons

N_k and not on its proportion, obtained in the mean-field limit. The same convention is used in [15], where the following weak error bound is proved.

Theorem 1. ([15], Theorem 4.) *Grant Assumption (A), and suppose that all spiking functions f_k belong to the space C_b^5 of bounded functions having bounded derivatives up to order 5. Then there exists a constant C, depending only on f_1, f_2 and the bounds on their derivatives, such that for all $\varphi \in C_b^4(\mathbb{R}^\kappa, \mathbb{R})$ and $t \geq 0$,*

$$\sup_{x \in \mathbb{R}^k} \left| \mathbb{E}_x \varphi(\bar{X}_t) - \mathbb{E}_x \varphi(Y_t) \right| \leq Ct \frac{\|\varphi\|_{4,\infty}}{N^2}, \tag{11}$$

where $\mathbb{E}_x$ denotes the conditional expectation given that $\bar{X}_0 = Y_0 = x$.

In the following, we strengthen the above result, allowing for a comparison of trajectories of the PDMP and the stochastic diffusion.

Theorem 2. (Strong error bound.) *Grant Assumption (A), and let $\| \cdot \|_\infty$ denote the sup norm on $\mathbb{R}^\kappa$. For all $N > 0$, a solution $\bar{X}$ of (5) and a solution Y of (7) (with $p_k = N_k/N$) can be constructed on the same probability space, such that there exists a constant $C > 0$ such that, for all $T > 0$,*

$$\sup_{t \leq T} \|\bar{X}_t - Y_t\|_\infty \leq \Theta_N e^{CT} \frac{\log(N)}{N} \tag{12}$$

almost surely, where Θ_N is a random variable with exponential moments whose distribution does not depend on N. In particular,

$$\mathbb{E}\left[\sup_{t \leq T} \|\bar{X}_t - Y_t\|_\infty\right] \leq Ce^{CT} \frac{\log(N)}{N}. \tag{13}$$

The proof of Theorem 2 is mainly inspired by [23] and relies on two main ingredients, a strong coupling between the standard Poisson process and the Brownian motion and a sharp result on the modulus of continuity for the Brownian motion. All the material is postponed to Appendix A.1.

When comparing (11) and (13), one notices that there is an exchange between the expectation sign and the absolute value. There are two prices to pay for such an exchange: first, a slower convergence rate with respect to N; second, a faster divergence rate with respect to t (the exponential term comes from a Grönwall-type argument). In the following remark we make precise the bound on the error which is caused by directly using the parameter p_k instead of N_k/N.

Remark 1. Let Y denote a solution of (7) (with parameter p_k equal to N_k/N), and let X denote a solution of (7) with fixed values p_k. Following the proof of Theorem 2, one can show that

$$\sup_{t \leq T} \|X_t - Y_t\| \leq \Theta_N e^{CT} \left(\frac{\log(N)}{N} + \max_k \left\{ \frac{1}{\sqrt{p_k N}} \left(1 - \sqrt{p_k N/N_k} \right) \right\} \right),$$

so that the strong error bound stated in the theorem also holds for the non-modified SDE if $\sqrt{p_k N/N_k} - 1$ is of order $N^{-1/2}$ or of faster order.

Fortunately, for any fixed N, setting $N_1 = \lfloor p_1 N \rfloor$ and $N_2 = \lceil p_2 N \rceil$ ensures that $\sqrt{p_k N/N_k} - 1$ is of order $N^{-1} < N^{-1/2}$, which grants that Theorem 2 holds for the SDE (7).

Since the SDE (7) transforms into the ODE (10) as N goes to infinity, the strong error bound can be used to prove the convergence of the PDMP to the solution of the ODE. However, this is beyond the scope of this paper.

2.2. Properties of the stochastic diffusion

The solution process $(X_t)_{t\geq 0}$ of SDE (7) is positive Harris recurrent with invariant measure π which is of full support (see [27]). This means that the trajectories visit all sets in the support of the invariant measure infinitely often almost surely. More precisely, for any initial condition x_0 and measurable set A such that $\pi(A) > 0$, $\limsup_{t\to+\infty} \mathbf{1}_A(X_t) = 1$ almost surely. Furthermore, by following the arguments in [30], the technical results proven in [27] can be used to prove geometric ergodicity of $(X_t)_{t\geq 0}$ as stated in Proposition 1 below.

In order to state the geometric ergodicity of $(X_t)_{t\geq 0}$, let us first specify the Lyapunov function $G : \mathbb{R}^\kappa \to \mathbb{R}$ introduced in [15]:

$$G(x) = \sum_{k=1}^{K} \sum_{j=1}^{\eta_k+1} \frac{j}{\nu_k^{j-1}} J(x^{k,j}), \tag{14}$$

where J is some smooth approximation of the absolute value. In particular, $J(x) = |x|$ for all $|x| \geq 1$, and $\max\{|J'(x)|, |J''(x)|\} \leq J_c$ for all x, for some finite constant J_c.

Proposition 1. (Geometric ergodicity.) *Grant Assumption (A). Then the solution of the SDE (7) has a unique invariant measure π on $\mathbb{R}^\kappa$. For all initial conditions x_0 and all $m \geq 1$, there exist $C = C(m) > 0$ and $\lambda = \lambda(m) > 0$ such that, for all measurable functions $g : \mathbb{R}^\kappa \to \mathbb{R}$ such that $|g| \leq G^m$,*

$$|\mathbb{E}g(X_t) - \pi(g)| \leq CG(x_0)^m e^{-\lambda t} \qquad \forall t \geq 0.$$

Proof. The proof closely follows that of Theorem 3.2 in [30] and is based on Lyapunov and minorization conditions (the latter are implied by the existence of a smooth transition density and the irreducibility of the space).

(i) First we use the fact that G is a Lyapunov function for X [15, Proposition 5], i.e., that there exist $\alpha, \beta > 0$ such that

$$\mathcal{A}^X G(x) \leq -\alpha G(x) + \beta,$$

where $\mathcal{A}^X G(x)$ is the infinitesimal generator of (7).

(ii) Then we note that, from any initial condition x_0, for any time $T > 0$ and any open set O, the probability that X_T belongs to O is positive. This is ensured by the controllability of the system (7) (see Theorem 4 in [27]).

(iii) Finally, we note that the process $(X_t)_{t\geq 0}$ possesses a smooth transition density. Its existence is ensured by verifying the Hörmander condition, which is done in Proposition 7 of [15].

The rest of the proof follows that of [30, Theorem 3.2]: apply [30, Theorem 2.5] to some discrete-time sampling of the process and conclude by interpolation. $\square$

Also note that the rank of the diffusion matrix $\sigma\sigma^T$ is smaller than the dimension of the system (7). This means that the system is not elliptic. However, the specific cascade structure of the drift ensures that the noise is propagated through the whole system via the drift term, so that the diffusion is hypoelliptic in the sense of stochastic calculus of variations [13, 28]. We also note that the SDE (7) is semi-linear, with a linear term given by the matrix (6). Thus,

its solution can be written in the form of a convolution equation (see, among others, [29, Section 3]).

Proposition 2. *The unique solution of (7) satisfies*

$$X_t = e^{At}x_0 + \int_0^t e^{A(t-s)}B(X_s)ds + \frac{1}{\sqrt{N}}\int_0^t e^{A(t-s)}\sigma(X_s)dW_s. \tag{15}$$

Proof. Consider the process $Y_t = e^{-At}X_t$. By Itô's formula one obtains

$$d\left(e^{-At}X_t\right) = \left(-Ae^{-At}X_t + e^{-At}\left(AX_t + B(X_t)\right)\right)dt + \frac{e^{-At}}{\sqrt{N}}\sigma(X_t)dW_t$$

$$= e^{-At}B(X_t)dt + \frac{e^{-At}}{\sqrt{N}}\sigma(X_t)dW_t.$$

Integrating both parts yields

$$e^{-At}X_t = x_0 + \int_0^t e^{-As}B(X_s)ds + \frac{1}{\sqrt{N}}\int_0^t e^{-As}\sigma(X_s)dW_s.$$

Multiplying the expression by e^{At} gives the result. □

Note that from this form, it is straightforward to see that the diffusion term is of full rank. Intuitively, this ensures the hypoellipticity. Systems of type (15) are called stochastic Volterra equations [1].

Now we focus on first and second moment bounds. The following results are needed, in particular, to ensure the accuracy of the approximation scheme in Section 3. In the following remark we provide some purely computational results in order to ease the further analysis.

Remark 2. Because of the block structure of the matrix A introduced in (6), its matrix exponential e^{At} can be computed as

$$e^{At} = \begin{pmatrix} e^{A_{v_1}t} & \mathbb{O}_{(\eta_1+1)\times(\eta_2+1)} \\ \mathbb{O}_{(\eta_2+1)\times(\eta_1+1)} & e^{A_{v_2}t} \end{pmatrix},$$

where $e^{A_{v_k}t}$, $k = 1, 2$, is an $(\eta_k + 1) \times (\eta_k + 1)$ upper-triangular matrix given by

$$e^{A_{v_k}t} = e^{-v_k t}\begin{pmatrix} 1 & t & \frac{t^2}{2} & \cdots & \frac{t^{\eta_k}}{\eta_k!} \\ 0 & 1 & t & \cdots & \frac{t^{\eta_k-1}}{(\eta_k-1)!} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{16}$$

In further computations we will often use the vectors $e^{At}X_s$. The elements of $e^{At}X_s$ are given by the formula

$$\left(e^{At}X_s\right)^{k,j} = e^{-v_k t}\sum_{m=j}^{\eta_k+1}\frac{t^{m-j}}{(m-j)!}X_s^{k,m}. \tag{17}$$

Theorem 3. (First moment bounds of the diffusion process.) *Grant Assumption (A). The following bounds hold for the components of* $\mathbb{E}[X_t]$:

$$\mathcal{I}_{\min}^{k,j} \leq \mathbb{E}[X_t^{k,j}] \leq \mathcal{I}_{\max}^{k,j},$$

where

$$\mathcal{I}_{\min}^{k,j} = \left(e^{At}x_0\right)^{k,j} + \left[1 - e^{-tv_k} \sum_{l=0}^{\eta_k+1-j} \frac{(tv_k)^l}{l!}\right] \min\left\{0, \frac{c_k f_{k+1}^{\max}}{v_k^{(\eta_k+2-j)}}\right\},$$

$$\mathcal{I}_{\max}^{k,j} = \left(e^{At}x_0\right)^{k,j} + \left[1 - e^{-tv_k} \sum_{l=0}^{\eta_k+1-j} \frac{(tv_k)^l}{l!}\right] \max\left\{0, \frac{c_k f_{k+1}^{\max}}{v_k^{(\eta_k+2-j)}}\right\}.$$

Proof of Theorem 3. From Proposition 2 and Remark 2, it follows that the convolution-based representation of the kth population is given by

$$X_t^k = (e^{At}x_0)^k + \int_0^t e^{A_{v_k}(t-s)} B^k(X_s^{k+1}) ds + \frac{1}{\sqrt{N}} \int_0^t e^{A_{v_k}(t-s)} \sigma^k(X_s^{k+1}) dW_s.$$

Consequently, the jth components are given by

$$X_t^{k,j} = \underbrace{\left(e^{At}x_0\right)^{k,j}}_{=: T_1(t)} + \underbrace{\int_0^t c_k f_{k+1}(X_s^{k+1,1}) \frac{e^{-v_k(t-s)}}{(\eta_k+1-j)!} (t-s)^{\eta_k+1-j} ds}_{=: T_2(t)}$$

$$+ \underbrace{\frac{1}{\sqrt{N}} \int_0^t \frac{c_k}{\sqrt{p_{k+1}}} \sqrt{f_{k+1}(X_s^{k+1,1})} \frac{e^{-v_k(t-s)}}{(\eta_k+1-j)!} (t-s)^{\eta_k+1-j} dW_s^{k+1}}_{=: T_3(t)}.$$

Note that $\mathbb{E}[T_1(t)] = T_1(t)$ and $\mathbb{E}[T_3(t)] = 0$. It remains to consider $T_2(t)$. The fact that the intensity function is bounded by $0 < f_{k+1} \leq f_{k+1}^{\max}$ implies that

$$\min\{0, c_k\} \frac{f_{k+1}^{\max}}{(\eta_k+1-j)!} I^{k,j} \leq \mathbb{E}[T^2(t)] \leq \max\{0, c_k\} \frac{f_{k+1}^{\max}}{(\eta_k+1-j)!} I^{k,j},$$

where

$$I^{k,j} = \int_0^t e^{-v_k(t-s)} (t-s)^{\eta_k+1-j} ds.$$

Now, let us consider the integral $I^{k,j}$:

$$\int_0^t e^{-v_k(t-s)} (t-s)^{\eta_k+1-j} ds = t^{\eta_k+1-j} \int_0^t e^{-v_k t \frac{t-s}{t}} \left(\frac{t-s}{t}\right)^{\eta_k+1-j} ds.$$

Setting $z = \frac{t-s}{t}$ yields

$$t^{\eta_k+2-j} \int_0^1 e^{-v_k tz} z^{\eta_k+1-j} dz = \frac{(\eta_k + 1 - j)!}{v_k^{(\eta_k+2-j)}} \left[1 - e^{-v_k t} \sum_{l=0}^{\eta_k+1-j} \frac{(v_k t)^l}{l!} \right].$$

This gives the result. $\square$

Remark 3. Recalling (17) and using that

$$\lim_{t\to\infty} e^{-v_k t} \sum_{l=0}^{\eta_k+1-j} \frac{(tv_k)^l}{l!} = 0,$$

it follows from Theorem 3 that

$$\min\left\{0, \frac{c_k f_{k+1}^{\max}}{v_k^{(\eta_k+2-j)}}\right\} \leq \lim_{t\to\infty} \mathbb{E}[X_t^{k,j}] \leq \max\left\{0, \frac{c_k f_{k+1}^{\max}}{v_k^{(\eta_k+2-j)}}\right\}.$$

The derived moment bounds give some intuition on how the system behaves in the long run. Remarkably, depending on whether c_k is positive or negative, the trajectories of $(X_t^k)_{t\geq 0}$ are on average bounded by 0 from below or above, respectively. This is in agreement with the fact that the sign of c_k defines whether the corresponding neural population is excitatory ($c_k = +1$) or inhibitory ($c_k = -1$). Moreover, we may immediately see the effect of increasing the memory order η_k, depending on the constant v_k. When $v_k = 1$, the bounds for all j components are determined entirely by c_k and the bounds of the intensity functions. When $v_k < 1$ and $\eta_k \to \infty$, the first components, presenting the current state of the process, tend to infinity. Similarly, for $v_k > 1$, the trajectories are attracted to 0. Finally, note that the first moment bounds do not depend on the number of neurons in the system.

Theorem 4. (Second moment bounds of the diffusion process.) *Grant Assumption (A). The following bounds hold for* $\mathbb{E}[(X_t^{k,j})^2]$:

$$\mathbb{E}[(X_t^{k,j})^2] \leq \left(\left(e^{At} x_0\right)^{k,j}\right)^2 + 2\left(e^{At} x_0\right)^{k,j} \max\left\{0, \frac{c_k f_{k+1}^{\max}}{(\eta_k + 1 - j)!} I_1^{k,j}(t)\right\}$$

$$+ f_{k+1}^{\max} \left(\frac{c_k}{(\eta_k + 1 - j)!}\right)^2 \left(\sqrt{f_{k+1}^{\max}} I_1^{k,j}(t) + \sqrt{\frac{I_2^{k,j}(t)}{N \cdot p_{k+1}}}\right)^2,$$

where $I_u^{k,j}(t)$, $u = 1, 2$, *are defined as*

$$I_u^{k,j}(t) := \int_0^t e^{-uv_k(t-s)} (t-s)^{u(\eta_k+1-j)} ds$$

$$= \frac{(u(\eta_k + 1 - j))!}{(uv_k)^{u(\eta_k+1-j)+1}} \left[1 - e^{-utv_k} \sum_{l=0}^{u(\eta_k+1-j)} \frac{(utv_k)^l}{l!} \right].$$
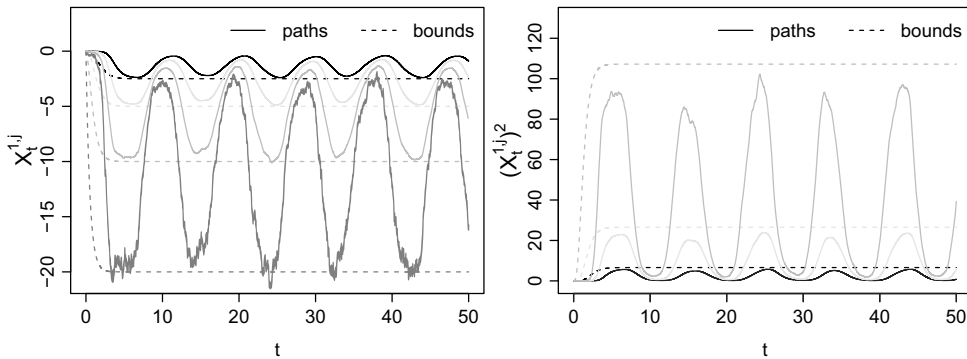
FIGURE 1. First (left panel) and second (right panel) moment bounds with respective trajectories of the inhibitory population $k = 1$. The rate function f_2 is given in Section 5. The parameters are $\eta_1 = 3$, $\nu_1 = 2$, $N = 20$, and $p_2 = 1/2$.

The proof of Theorem 4 is similar to that of Theorem 3 and is postponed to Appendix A.2.

Remark 4. Theorem 4 gives the following asymptotic bounds:

$$\lim_{t \to \infty} \mathbb{E}[(X_t^{k,j})^2] \leq f_{k+1}^{\max} \left(\frac{c_k}{(\eta_k + 1 - j)!} \right)^2 \left(\sqrt{f_{k+1}^{\max}} C_1^{k,j} + \sqrt{\frac{C_2^{k,j}}{N \cdot p_{k+1}}} \right)^2,$$

where

$$C_u^{k,j} := \lim_{t \to \infty} I_u^{k,j}(t) = \frac{(u(\eta_k + 1 - j))!}{(u\nu_k)^{u(\eta_k + 1 - j) + 1}}.$$

Note that for $N \to \infty$, the bound obtained in Theorem 4 equals the square of the bound for the first moment, derived in Theorem 3. This is in agreement with the fact that the stochastic system (7) transforms into an ODE as N increases [15]. In other words, its diffusion coefficient tends to $\mathbb{O}_{\kappa \times 2}$ as N tends to infinity.

In Figure 1, the first and second moment bounds, derived in Theorem 3 and Theorem 4, respectively, are illustrated. In the left panel, we plot four sample trajectories (solid lines) of an inhibitory population X^1 and their lower first moment bounds (dashed lines). The main variable $X^{1,1}$ and its lower moment bound are depicted in black. The remaining three trajectories $X^{1,j}$, $j = 2, 3, 4$, are auxiliary variables. They (and their corresponding bounds) are depicted in different shades of grey. We see that the trajectories can exceed the theoretical bounds, especially when the effect of noise is large. On average, the trajectories stay within the bounds. In the right panel, we plot the square of the first three components of X^1 (and their second moment bounds), omitting the fourth in order to stay within an easily interpretable scale. We conclude that the bounds are fairly precise for the parameter setting under consideration.

3. Numerical splitting schemes for the stochastic diffusion

The solution of the system (7) cannot be written in an explicit form, and thus a numerical approximation is required. Let $[0, T]$ with $T > 0$ be the time interval of interest, and consider the discretization $(t_i)_{i=0,\dots,i_{\max}}$ given by $t_i = i\Delta$, where $\Delta = T/i_{\max}$. In the following, $\tilde{X}_{t_i}$ denotes a numerical realization of the diffusion process, evaluated at the discrete time points.

We derive and investigate two numerical schemes based on the splitting approach. The goal of this method is to divide the equation into explicitly solvable subequations and to compose the obtained explicit solutions in a proper way. Usually, the choice of the subsystems is not unique. Here, because of the specific structure of the SDE (7), we split it into the subsystems

$$dX_t^{[1]} = AX_t^{[1]}dt, \tag{18}$$

$$dX_t^{[2]} = B(X_t^{[2]})dt + \frac{1}{\sqrt{N}}\sigma\left(X_t^{[2]}\right)dW_t. \tag{19}$$

Both subsystems are explicitly solvable. The first one is a linear ODE whose flow is given by $\psi_t^{[1]} : x \mapsto e^{At}x$. For the second one, recall that B and σ are given by (8) and (9), respectively. It is easy to see that all components of $X^{[2]}$, except for two ($X^{[2],1,\eta_1+1}$ and $X^{[2],2,\eta_2+1}$), have null derivative. Moreover, the drift and diffusion coefficients of $X^{[2],1,\eta_1+1}$ depend only on $X^{[2],2,1}$, and those of $X^{[2],2,\eta_2+1}$ depend only on $X^{[2],1,1}$. Hence, the respective explicit (stochastic) flows are given by

$$\psi_t^{[1]}(x) := e^{At}x,$$

$$\psi_t^{[2]}(x) := x + tB(x) + \frac{\sqrt{t}}{\sqrt{N}}\sigma(x)\xi,$$

where $\xi = (\xi^1, \xi^2)^T$ is a two-dimensional standard normal vector. Then, the Lie–Trotter and the Strang compositions of flows [31, 41] are given by

$$\tilde{X}_{t_{i+1}}^{LT} = \left(\psi_\Delta^{[1]} \circ \psi_\Delta^{[2]}\right)\left(\tilde{X}_{t_i}^{LT}\right) = e^{A\Delta}\left(\tilde{X}_{t_i}^{LT} + \Delta B(\tilde{X}_{t_i}^{LT}) + \frac{\sqrt{\Delta}}{\sqrt{N}}\sigma(\tilde{X}_{t_i}^{LT})\xi_i\right), \tag{20}$$

$$\tilde{X}_{t_{i+1}}^{ST} = \left(\psi_{\frac{\Delta}{2}}^{[1]} \circ \psi_\Delta^{[2]} \circ \psi_{\frac{\Delta}{2}}^{[1]}\right)\left(\tilde{X}_{t_i}^{ST}\right)$$

$$= e^{A\Delta}\tilde{X}_{t_i}^{ST} + \Delta e^{A\frac{\Delta}{2}}B(e^{A\frac{\Delta}{2}}\tilde{X}_{t_i}^{ST}) + \frac{\sqrt{\Delta}}{\sqrt{N}}e^{A\frac{\Delta}{2}}\sigma(e^{A\frac{\Delta}{2}}\tilde{X}_{t_i}^{ST})\xi_i, \tag{21}$$

respectively, with $\tilde{X}_0^{LT} = \tilde{X}_0^{ST} = x_0$ and $(\xi_i)_{i=1,\dots,i_{max}}$ independent and identically distributed. The two splitting schemes (20) and (21) define numerical solutions of the SDE (7). Note that by setting $\sigma(x) \equiv \mathbb{0}_{\kappa\times 2}$, both schemes can be used for simulating the ODE (10).

For the sake of simplicity, we focus on the Lie–Trotter splitting (20) in the subsequent analysis, since its representation is more intuitive. Thus, throughout Section 3 we set $\tilde{X} \equiv \tilde{X}^{LT}$. However, similar results can also be obtained for the more evolved Strang approach (21).

Remark 5. Note that thanks to the matrix exponential entering the diffusion terms in (20) and (21), the noise propagates through all components of the system at each time step. In other words, the conditional variance matrix Σ is of full rank and is given by

$$\Sigma\left[\tilde{X}_{t_{i+1}}|\tilde{X}_{t_i}\right] := \frac{\Delta}{N}e^{A\Delta}\sigma(\tilde{X}_{t_i})\sigma^T(\tilde{X}_{t_i})\left(e^{A\Delta}\right)^T.$$

This can be regarded as a discrete analogue of the hypoellipticity of the continuous process, a property that the approximation methods based on the Itô–Taylor expansion of the infinitesimal generator of (7) (see [22]) do not preserve.

3.1. Strong convergence in the mean-square sense

We now focus on the convergence in the mean-square sense and show that the numerical solutions obtained via the splitting approach converge to the process as the time step $\Delta \to 0$ with order 1. The frequently applied Euler–Maruyama scheme usually converges with mean-square order $1/2$ if the noise is multiplicative [22, 33], as is the case for the system (7). In the following result, thanks to the specific structure of the noise component, we show that the Euler–Maruyama scheme coincides with the Milstein scheme, which is known to converge with mean-square order 1. This result is then used to establish the convergence order of the splitting scheme.

Theorem 5. (Mean-square convergence of the splitting scheme.) *Grant Assumption (A). Let $\tilde{X}_{t_i}$ denote the numerical method defined by (20) at time point t_i and starting from x_0. Then $\tilde{X}_{t_i}$ is mean-square convergent with order 1; i.e., there exists a constant $C > 0$ such that*

$$\left(\mathbb{E} \left[\left\| X_{t_i} - \tilde{X}_{t_i} \right\|^2 \right] \right)^{\frac{1}{2}} \leq C\Delta$$

for all time points t_i, $i = 1, \ldots, i_{\max}$, where $\| \cdot \|$ denotes the Euclidean norm.

Proof of Theorem 5. Let us denote by $\tilde{X}^{EM}$ a numerical solution of the SDE (7) obtained via the Euler–Maruyama method; that is,

$$\tilde{X}_{t_{i+1}}^{EM} = \tilde{X}_{t_i}^{EM} + \Delta \left(A \tilde{X}_{t_i}^{EM} + B(\tilde{X}_{t_i}^{EM}) \right) + \frac{\sqrt{\Delta}}{\sqrt{N}} \sigma(\tilde{X}_{t_i}^{EM}) \xi_i. \tag{22}$$

First, we show that the Euler–Maruyama method, when applied to the system (7), coincides with the Milstein scheme, which is known to converge with mean-square order 1. To do so, we define the vector x by $x = (x^1, \ldots, x^\kappa)$, where $\kappa = \eta_1 + \eta_2 + 2$. Further, we recall that the jth component, $j = 1, \ldots, \kappa$, of the Milstein scheme differs from the jth component of the Euler–Maruyama scheme (22) only by the following additional term:

$$\sum_{m_1, m_2 = 1}^{2} \sum_{l=1}^{\kappa} \sigma^{l, m_1} \frac{\partial \sigma^{j, m_2}}{\partial x^l} I(m_1, m_2),$$

where $\sigma^{j, m}$ denotes the value of the element at the jth row and the mth column of the diffusion matrix σ at time t_i, and

$$I(m_1, m_2) := \int_{t_i}^{t_{i+1}} \int_{t_i}^{s_2} dW_{s_1}^{m_1} dW_{s_2}^{m_2}.$$

Now note that the term $\partial \sigma^{j, m_2} / \partial x^l$ is different from 0 only for $j = \eta_1 + 1$, $m_2 = 1$, $l = \eta_1 + 2$ and for $j = \eta_1 + \eta_2 + 2$, $m_1 = 2$, $l = 1$. However, σ^{l, m_1} equals 0 for those values of l. Thus, the above double sum equals 0 and the Euler–Maruyama method coincides with the Milstein scheme. This implies that

$$\| X_{t_i} - \tilde{X}_{t_i}^{EM} \|_{L^2} \leq C\Delta, \tag{23}$$

where $\| \cdot \|_{L^2} := \left(\mathbb{E}[\| \cdot \|^2] \right)^{1/2}$ denotes the L^2 norm and C is some generic constant. For the second part, we provide a proof similar to the one presented in [32]. Applying the triangle inequality yields that

$$\| X_{t_i} - \tilde{X}_{t_i} \|_{L^2} \leq \| X_{t_i} - \tilde{X}_{t_i}^{EM} \|_{L^2} + \| \tilde{X}_{t_i}^{EM} - \tilde{X}_{t_i} \|_{L^2}.$$

Given $X_{t_i} := x$, let us denote by $\tilde{X}^{EM}_{t_{i+1}}(x, t_i)$ and $\tilde{X}_{t_{i+1}}(x, t_i)$ the one-step approximation of the Euler–Maruyama and splitting scheme, respectively. For instance, $\tilde{X}^{EM}_{t_{i+1}}(x, t_i)$ is given by Equation (22) with $\tilde{X}^{EM}_{t_i}$ replaced by x. By the definition of the matrix exponent, i.e. $e^{A\Delta} := I + \Delta A + \frac{\Delta^2}{2} A^2 + O(\Delta^3)$, and by recalling (20), we obtain that

$$
\begin{aligned}
\tilde{X}^{EM}_{t_{i+1}}(x, t_i) - \tilde{X}_{t_{i+1}}(x, t_i) &= x + \Delta A x + \Delta B(x) + \frac{\sqrt{\Delta}}{\sqrt{N}} \sigma(x) \xi_i \\
&\quad - e^{A\Delta} \left(x + \Delta B(x) + \frac{\sqrt{\Delta}}{\sqrt{N}} \sigma(x) \xi_i \right) \\
&= x + \Delta A x + \Delta B(x) + \frac{\sqrt{\Delta}}{\sqrt{N}} \sigma(x) \xi_i \\
&\quad - x - \Delta B(x) - \frac{\sqrt{\Delta}}{\sqrt{N}} \sigma(x) \xi_i \\
&\quad - \Delta A x - \Delta^2 A B(x) - \frac{\Delta^{\frac{3}{2}}}{\sqrt{N}} A \, \sigma(x) \xi_i + R \\
&= -\Delta^2 A B(x) - \frac{\Delta^{\frac{3}{2}}}{\sqrt{N}} A \, \sigma(x) \xi_i + R,
\end{aligned}
$$

where R is a corresponding stochastic remainder. Consequently, we get that

$$
\left\| \mathbb{E} \left[\tilde{X}^{EM}_{t_{i+1}}(x, t_i) - \tilde{X}_{t_{i+1}}(x, t_i) \right] \right\| = O(\Delta^2),
$$

$$
\left\| \tilde{X}^{EM}_{t_{i+1}}(x, t_i) - \tilde{X}_{t_{i+1}}(x, t_i) \right\|_{L^2} = O(\Delta^{\frac{3}{2}}).
$$

Recalling (23), the result follows from the fundamental theorem on the mean-square order of convergence; see Theorem 1.1. in [33].   □

Theorem 5 states that as $\Delta \to 0$, the approximated solution $(\tilde{X}_{t_i})_{i=0,\dots,i_{\max}}$ converges to the true process $(X_t)_{t \geq 0}$ in the mean-square sense. In practice, however, fixed time steps $\Delta > 0$ are required. Thus, there is not yet any guarantee that the constructed numerical solutions share the same properties as the true solution of (7). For these reasons, we additionally study the ability of $(\tilde{X}_{t_i})_{i=0,\dots,i_{\max}}$ to preserve the derived first and second moment bounds and the geometric ergodicity of the SDE (7).

Note also that, unlike the case of ODE systems [19], for stochastic equations the theoretical order of convergence usually cannot be increased by using the Strang composition instead of the Lie–Trotter approach. In practice, however, the Strang splitting often performs better than the Lie–Trotter method; see e.g. [3, 7]. This is also confirmed by our numerical experiments in Section 5.

3.2. Moment bounds of the approximated process

We are now interested in studying the qualitative properties of the splitting schemes for fixed time steps $\Delta > 0$. We start by illustrating that the constructed splitting schemes preserve

the convolution-based structure of the model derived in Proposition 2. Using the one-step approximation (20) and performing back iteration yields

$$\tilde{X}_{t_i} = e^{At_i}x_0 + \Delta \sum_{l=1}^{i} e^{At_l} B(\tilde{X}_{t_{i-l}}) + \frac{\sqrt{\Delta}}{\sqrt{N}} \sum_{l=1}^{i} e^{At_l} \sigma(\tilde{X}_{t_{i-l}})\xi_{i-l}. \tag{24}$$

Note that the first term on the right side of (15) is preserved exactly. Moreover, the sums in (24) correspond to approximations of the integrals in (15) using the left-point rectangle rule. The expression (24) allows one to derive moment bounds for the numerical process in a similar fashion to that presented for the continuous process in the previous section.

Theorem 6. (First moment bounds of the approximated process.) *Grant Assumption (A). The following bounds hold for the components of* $\mathbb{E}[\tilde{X}_{t_i}]$:

$$\tilde{\mathcal{I}}_{\min}^{k,j} \leq \mathbb{E}[\tilde{X}_{t_i}^{k,j}] \leq \tilde{\mathcal{I}}_{\max}^{k,j},$$

where

$$\tilde{\mathcal{I}}_{\min}^{k,j} = \left(e^{At_i}x_0\right)^{k,j} + \Delta \sum_{l=0}^{i} e^{-\nu_k t_l} t_l^{\eta_k+1-j} \min\left\{0, \frac{c_k f_{k+1}^{\max}}{(\eta_k+1-j)!}\right\},$$

$$\tilde{\mathcal{I}}_{\max}^{k,j} = \left(e^{At_i}x_0\right)^{k,j} + \Delta \sum_{l=0}^{i} e^{-\nu_k t_l} t_l^{\eta_k+1-j} \max\left\{0, \frac{c_k f_{k+1}^{\max}}{(\eta_k+1-j)!}\right\}.$$

Proof of Theorem 6. From Remark 2 and (24), it follows that

$$\tilde{X}_{t_i}^k = (e^{At_i}x_0)^k + \Delta \sum_{l=1}^{i} e^{A\nu_k t_l} B^k(\tilde{X}_{t_{i-l}}^{k+1}) + \frac{\sqrt{\Delta}}{\sqrt{N}} \sum_{l=1}^{i} e^{A\nu_k t_l} \sigma^k(\tilde{X}_{t_{i-l}}^{k+1})\xi_{i-l}.$$

Consequently, the *j*th components are given by

$$\tilde{X}_{t_i}^{k,j} = \underbrace{(e^{At_i}x_0)^{k,j}}_{=T_1(t_i)} + \underbrace{\frac{1}{(\eta_k+1-j)!}\Delta \sum_{l=1}^{i} c_k f_{k+1}(\tilde{X}_{t_{i-l}}^{k+1,1})e^{-\nu_k t_l} t_l^{\eta_k+1-j}}_{=\tilde{T}_2(t)}$$

$$+ \underbrace{\frac{1}{(\eta_k+1-j)!}\frac{\sqrt{\Delta}}{\sqrt{N}} \sum_{l=1}^{i} \frac{c_k}{\sqrt{p_{k+1}}}\sqrt{f_{k+1}(\tilde{X}_{t_{i-l}}^{k+1,1})}e^{-\nu_k t_l} t_l^{\eta_k+1-j}\xi_{i-l}^{k+1}}_{=\tilde{T}_3(t_i)}.$$

Note that $\mathbb{E}[T_1(t_i)] = T_1(t_i)$ and $\mathbb{E}[\tilde{T}_3(t)] = 0$. The fact that the intensity function is bounded by $0 < f_{k+1} \leq f_{k+1}^{\max}$ implies the result. $\qquad\square$

Note that the bounds obtained in Theorem 6 equal those derived in Theorem 3, up to replacing the integrals (calculated in the proof of Theorem 3) by left Riemann sums. The accuracy of this
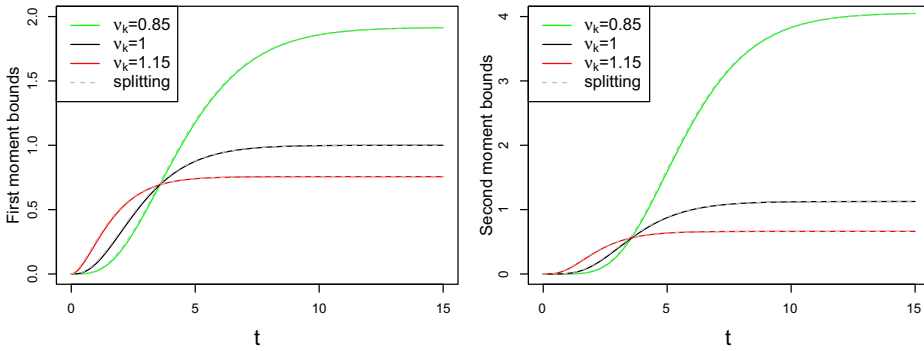
FIGURE 2. First (left panel) and second (right panel) moment bounds of the excitatory population $k = 2$ for different values of v_2. The moment bounds for the diffusion are in solid lines and the moment bounds for the splitting scheme are in dashed lines. The bound of the rate function is fixed to $f_1^{\max} = 1$. The parameters are $\eta_2 = 3$, $N = 100$, $p_1 = 1/2$, and the time step $\Delta = 0.1$ is used.

approximation depends on the step size Δ. Under reasonably small choices of Δ, the bounds are preserved accurately for all t_i. This is illustrated in the left panel of Figure 2, where we plot the first moment bound of the process (main variable of an excitatory population) and that of the approximated process, derived in Theorem 3 and Theorem 6, respectively. Different choices of v_k are compared, and $\Delta = 0.1$ is used for the bound of the approximated process.

The following corollary gives intuition for the long-time behaviour of the bounds.

Corollary 1.

(i) *The following bounds hold for the components of* $\mathbb{E}[\tilde{X}_{t_i}]$ *as* $i \to \infty$ *(with* Δ *fixed):*

$$\Delta^{\kappa^{k,j}+1} Li_{-\kappa^{k,j}} \left(e^{-v_k\Delta}\right) \min\left\{0, \frac{f_{k+1}^{max} c_k}{\kappa^{k,j}!}\right\} \leq \lim_{i\to\infty} \mathbb{E}[\tilde{X}_{t_i}^{k,j}]$$

$$\leq \Delta^{\kappa^{k,j}+1} Li_{-\kappa^{k,j}} \left(e^{-v_k\Delta}\right) \max\left\{0, \frac{f_{k+1}^{max} c_k}{\kappa^{k,j}!}\right\},$$

where $\kappa^{k,j} := \eta_k + 1 - j$, *and* $Li_{-\kappa^{k,j}} \left(e^{-v_k\Delta}\right)$ *is a polylogarithm function, which can be written as*

$$Li_{-\kappa^{k,j}} \left(e^{-v_k\Delta}\right) = (-1)^{\kappa^{k,j}+1} \sum_{l=0}^{\kappa^{k,j}} l!\, S(\kappa^{k,j}+1, l+1) \left(\frac{-1}{1-e^{-v_k\Delta}}\right)^{l+1},$$

with $S(\kappa^{k,j}+1, l+1)$ *denoting the Stirling numbers of the second kind* [38].

(ii) *The following bounds hold for the components of* $\mathbb{E}[\tilde{X}_{t_i}]$ *as* $i \to \infty$ *and* $\Delta \to 0$:

$$\min\left\{0, \frac{c_k f_{k+1}^{max}}{v_k^{\kappa^{k,j}+1}}\right\} \leq \lim_{\Delta\to 0} \lim_{i\to\infty} \mathbb{E}[\tilde{X}_{t_i}^{k,j}] \leq \max\left\{0, \frac{c_k f_{k+1}^{max}}{v_k^{\kappa^{k,j}+1}}\right\}.$$

Proof. (i) The zero bound is trivial. Considering

$$\lim_{i\to\infty}\sum_{l=0}^{i}e^{-\nu_k t_l}t_l^{\kappa^{k,j}} = \lim_{i\to\infty}\sum_{l=0}^{i}e^{-\nu_k l\Delta}(l\Delta)^{\kappa^{k,j}}$$

$$= \lim_{i\to\infty}\Delta^{\kappa^{k,j}}\sum_{l=0}^{i}e^{-\nu_k l\Delta}l^{\kappa^{k,j}} = \Delta^{\kappa^{k,j}}Li_{-\kappa^{k,j}}\left(e^{-\nu_k\Delta}\right)$$

gives the result. The explicit form of the function is given in [43].

(ii) Let us rewrite once again the expression included in the limit:

$$(-1)^{\kappa^{k,j}+1}\lim_{\Delta\to 0}\frac{(\nu_k\Delta)^{\kappa^{k,j}+1}}{\nu_k^{\kappa^{k,j}+1}\kappa^{k,j}!}\sum_{m=0}^{\kappa^{k,j}}m!\,S(\kappa^{k,j}+1,m+1)\left(\frac{-1}{1-e^{-\nu_k\Delta}}\right)^{m+1}$$

$$= \lim_{\Delta\to 0}\left[\frac{S(\kappa^{k,j}+1,\kappa^{k,j}+1)\kappa^{k,j}!}{\nu_k^{\kappa^{k,j}+1}\kappa^{k,j}!}\left(\frac{\nu_k\Delta}{1-e^{-\nu_k\Delta}}\right)^{\kappa^{k,j}+1}\right.$$

$$\left. - \Delta\frac{S(\kappa^{k,j}+1,\kappa^{k,j})(\kappa^{k,j}-1)!}{\nu_k^{\kappa^{k,j}}\kappa^{k,j}!}\left(\frac{\nu_k\Delta}{1-e^{-\nu_k\Delta}}\right)^{\kappa^{k,j}}+O(\Delta^2)\right].$$

Note that

$$\lim_{\Delta\to 0}\left(\frac{\nu_k\Delta}{1-e^{-\nu_k\Delta}}\right)=1.$$

This implies that in the limit $1/\nu_k^{\kappa^{k,j}+1}$ is the only remaining term, since the rest converges to 0 as $\Delta\to 0$. This gives the result. $\square$

In the first part of Corollary 1, the sums in Theorem 6 are calculated explicitly as $i\to\infty$. This limit is described by polylogarithm functions. Note that the zero bounds derived in Remark 3, i.e. the upper bounds for the inhibitory population ($c_k=-1$) and the lower bounds for the excitatory population ($c_k=+1$), are preserved exactly by the splitting scheme for all times t_i and for all choices of $\Delta>0$. Moreover, the lower bounds for the inhibitory population and the upper bounds for the excitatory population are preserved accurately as $i\to\infty$, provided that Δ is reasonably small. Indeed, as $i\to\infty$ and $\Delta\to 0$ (second part of Corollary 1), the bounds coincide with the ones obtained in Remark 3.

Theorem 7. (Second moment bounds of the approximated process.) *Grant Assumption (A). Each component of* $\mathbb{E}[(\tilde{X}_t^k)^2]$ *is bounded by*

$$\mathbb{E}[(\tilde{X}_t^{k,j})^2]\le\left(\left(e^{At}x_0\right)^{k,j}\right)^2 + 2\left(e^{At}x_0\right)^{k,j}\max\left\{0,\frac{c_k f_{k+1}^{\max}}{(\eta_k+1-j)!}\tilde{I}_1^{k,j}(t)\right\}$$

$$+f_{k+1}^{\max}\left(\frac{c_k}{(\eta_k+1-j)!}\right)^2\left(\sqrt{f_{k+1}^{\max}}\,\tilde{I}_1^{k,j}(t)+\sqrt{\frac{\tilde{I}_2^{k,j}(t)}{N\cdot p_{k+1}}}\right)^2,$$

where $\tilde{I}_u^{k,j}(t)$, $u = 1, 2$, *are defined as*

$$\tilde{I}_u^{k,j}(t) := \Delta \sum_{l=0}^{i} e^{-uv_k t_l} t_l^{u(\eta_k+1-j)}.$$

Proof of Theorem 7. The proof repeats the proof of Theorem 4, up to replacing integrals $I_u^{k,j}(t)$ by sums $\tilde{I}_u^{k,j}(t)$. □

As before, the second moment bounds obtained for the splitting scheme equal those derived for the true process in Theorem 4, except that the integrals are replaced by the corresponding Riemann sums. Using the same arguments as in the proof of Corollary 1, we conclude that the second moment bounds are also preserved accurately by the splitting scheme for reasonable choices of the time step Δ. A comparison of the theoretical and discrete second moment bounds is provided in the right panel of Figure 2.

3.3. Geometric ergodicity of the approximated process

Finally, our aim is to prove that the splitting scheme preserves the ergodic property of the underlying process, in the spirit of [3, 30], providing a discrete analogue of Proposition 1. The main step is to establish a discrete Lyapunov condition for the approximated solution $(\tilde{X}_{t_i})_{i=0,\ldots,i_{\max}}$. This is granted by the following lemma.

Lemma 1. (Lyapunov condition for the approximated process.) *Grant Assumption (A). The functional $\tilde{G}$ given by*

$$\tilde{G}(x) = \sum_{k=1}^{2} \sum_{j=1}^{\eta_k+1} \frac{j}{v_k^{j-1}} \left| x^{k,j} \right|$$

is a Lyapunov function for $\tilde{X}$; i.e., there exist constants $\alpha \in [0, 1)$ and $\beta \geq 0$ such that

$$\mathbb{E}\left[\tilde{G}(\tilde{X}_{t_{i+1}}) | \tilde{X}_{t_i} \right] \leq \alpha \tilde{G}(\tilde{X}_{t_i}) + \beta.$$

Proof. We bound the approximated solution obtained via (20) from above by a sum of three terms, thanks to the triangle inequality:

$$\tilde{G}(\tilde{X}_{t_{i+1}}) = \tilde{G}\left(e^{A\Delta}\tilde{X}_{t_i} + \Delta e^{A\Delta} B(\tilde{X}_{t_i}) + \frac{\sqrt{\Delta}}{\sqrt{N}} e^{A\Delta}\sigma(\tilde{X}_{t_i})\xi_i \right)$$

$$\leq \underbrace{\tilde{G}\left(e^{A\Delta}\tilde{X}_{t_i} \right)}_{T_1} + \underbrace{\Delta\tilde{G}\left(e^{A\Delta}B(\tilde{X}_{t_i}) \right)}_{T_2} + \underbrace{\frac{\sqrt{\Delta}}{\sqrt{N}}\tilde{G}\left(e^{A\Delta}\sigma(\tilde{X}_{t_i})\xi_i \right)}_{T_3}.$$

Note that the term T_2, as well as the expectation of T_3, is bounded by a constant depending on $f_k^{\max}$, so that $\mathbb{E}[T_2|\tilde{X}_{t_i}] + \mathbb{E}[T_3|\tilde{X}_{t_i}] \leq \beta$, and $\beta > 0$ since we consider the absolute value. Further, using the formulas (16)–(17), we can expand T_1 as follows:

$$\tilde{G}\left(e^{A\Delta}\tilde{X}_{t_i} \right) = \sum_{k=1}^{2} e^{-v_k\Delta} \sum_{j=1}^{\eta_k+1} \frac{j}{v_k^{j-1}} \left| \sum_{m=j}^{\eta_k+1} \frac{\Delta^{m-j}}{(m-j)!} \tilde{X}_{t_i}^{k,m} \right|$$

$$\leq \sum_{k=1}^{2} e^{-v_k\Delta} \sum_{j=1}^{\eta_k+1} \frac{j}{v_k^{j-1}} \sum_{m=j}^{\eta_k+1} \frac{\Delta^{m-j}}{(m-j)!} \left| \tilde{X}_{t_i}^{k,m} \right|$$

$$= \sum_{k=1}^{2} e^{-v_k \Delta} \left[\sum_{j=1}^{\eta_k+1} \frac{j}{v_k^{j-1}} \left| \tilde{X}_{t_i}^{k,j} \right| + \Delta \sum_{j=2}^{\eta_k+1} \frac{j-1}{v_k^{j-2}} \left| \tilde{X}_{t_i}^{k,j} \right| + \cdots + \frac{\Delta^{\eta_k}}{\eta_k!} \left| \tilde{X}_{t_i}^{k,\eta_k+1} \right| \right]$$

$$= \sum_{k=1}^{2} e^{-v_k \Delta} \left[\sum_{j=1}^{\eta_k+1} \frac{j}{v_k^{j-1}} \left| \tilde{X}_{t_i}^{k,j} \right| + \Delta v_k \sum_{j=2}^{\eta_k+1} \frac{j-1}{v_k^{j-1}} \left| \tilde{X}_{t_i}^{k,j} \right| + \cdots + \frac{(v_k \Delta)^{\eta_k}}{\eta_k!} \frac{1}{v_k^{\eta_k}} \left| \tilde{X}_{t_i}^{k,\eta_k+1} \right| \right].$$

Note that, since $v_k > 0$, for all $m \geq 1$ it holds that

$$\sum_{j=m}^{\eta_k+1} \frac{(j-m+1)}{v_k^j} \left| \tilde{X}_{t_i}^{k,j} \right| \leq \sum_{j=1}^{\eta_k+1} \frac{j}{v_k^j} \left| \tilde{X}_{t_i}^{k,j} \right| = \tilde{G} \left(\tilde{X}_{t_i}^{k} \right).$$

Thus, we have

$$\tilde{G} \left(e^{A\Delta} \tilde{X}_{t_i} \right) \leq \sum_{k=1}^{2} \left(e^{-v_k \Delta} \sum_{j=0}^{\eta_k} \frac{(v_k \Delta)^j}{j!} \right) \tilde{G} \left(\tilde{X}_{t_i}^{k} \right).$$

Let

$$\alpha = \max_{k} \left(e^{-v_k \Delta} \sum_{j=0}^{\eta_k} \frac{(v_k \Delta)^j}{j!} \right).$$

Since η_k is finite, we get $\alpha < 1$, which implies the result. $\qquad\square$

Note that the statement of Lemma 1 holds without any assumption on the time step Δ. Also, the Lyapunov function is the same as for the continuous process up to smoothing the absolute value (see (14)). With a discrete Lyapunov condition established, the ergodicity is conditioned on two further technical steps. First, the transition probability of two (or more) consecutive steps, given by the recursive relation (20), must have a smooth transition density. This fact is granted by the hypoellipticity of the scheme (see Remark 5).

Second, the irreducibility condition must hold. This means that any point $y \in \mathbb{R}^\kappa$ can be reached from any starting point $x \in \mathbb{R}^\kappa$ in a fixed number of steps. In other words, we need a discrete-time analogue of Theorem 4 in [27], granting the controllability of the SDE (7). This is provided by the following lemma, which is proved in Appendix A.3.

Lemma 2. (Irreducibility condition.) *Grant Assumption (A). Let $\eta^* = \max_k\{\eta_k\}$. Then, for all $x, y \in \mathbb{R}^\kappa$, there exists some sequence of two-dimensional vectors $(\xi_i)_{i=1,\ldots,\eta^*+1}$ such that*

$$y = \underbrace{\left(\psi_\Delta[\xi_{\eta^*+1}] \circ \cdots \circ \psi_\Delta[\xi_1] \right)}_{\eta^*+1} (x),$$

where ψ_Δ denotes one step of the scheme defined by (20), and the notation $[\,\cdot\,]$ is introduced to stress the dependency on the vectors $(\xi_i)_{i=1,\ldots,\eta^*+1}$.

Lemmas 1 and 2, combined with the hypoellipticity of the scheme, gives the following result, which is analogous to Theorem 7.3 in [30].

Theorem 8. (Geometric ergodicity.) *Grant Assumption (A). Then the process $(\tilde{X}_{t_i})_{i=0,\ldots,i_{max}}$ has a unique invariant measure π^Δ on $\mathbb{R}^\kappa$. For all initial conditions x_0 and all $m \geq 1$, there exist $\tilde{C} = C(m, \Delta) > 0$ and $\tilde{\lambda} = \tilde{\lambda}(m, \Delta) > 0$ such that, for all measurable functions $g : \mathbb{R}^\kappa \to \mathbb{R}$ such that $|g| \leq \tilde{G}^m$,*

$$\forall i = 0, \ldots, i_{max}, \qquad \left| \mathbb{E}g(\tilde{X}_{t_i}) - \pi^\Delta(g) \right| \leq \tilde{C}\tilde{G}(x_0)^m e^{-\tilde{\lambda} t_i}.$$

4. Thinning procedure for the simulation of the PDMP

In this section we explain the simulation method for the multidimensional point process characterized by the intensities (1). This part is motivated by the fact that the simulation of the PDMP, unlike that of the stochastic diffusion, can be exact. By that, we mean that the result of the simulation is a realization of $(\bar{X}_t)_{t\geq 0}$. In comparison, the result of the simulation of the diffusion $(X_t)_{t\geq 0}$ is in fact the discrete-time process $(\tilde{X}_{t_i})_{i=0,\ldots,i_{\max}}$. This allows us to compare the PDMP (5) with the stochastic diffusion defined through (7), which we treat via the property-preserving splitting scheme introduced in the previous section.

We choose the thinning procedure which dates back to [25] and [26]. It is based on the rejection principle and relies on the following fact. In order to simulate a point process Z according to the stochastic intensity λ_t, it is sufficient to simulate some (dominating) point process $\tilde{Z}$ with (dominating) predictable piecewise constant intensity $\tilde{\lambda}$ such that $\lambda_t \leq \tilde{\lambda}_t$. During the simulation of $\tilde{Z}$, each new simulated spiking time $\tilde{T}$ for $\tilde{Z}$ is kept as a point of Z with probability $\lambda_{\tilde{T}}/\tilde{\lambda}_{\tilde{T}}$ (independently from every other point). Otherwise, $\tilde{T}$ is discarded. The efficiency of the thinning procedure is highly related to the sharpness of the upper bound $\tilde{\lambda}$. The sharper the bound, the fewer rejections are made and the more efficient the procedure is.

Note that the case $\eta_k = 0$ corresponds to the exponential kernel. The simulation of Hawkes processes with an exponential kernel has been widely studied, and there exist several implemented packages, e.g. for the software R. Apart from the thinning procedure, other exact simulation algorithms are available; see, in particular, [12]. To the best of our knowledge, the only reference for the case when $\eta_k \geq 1$ is [16]. The aim of the current section is to generalize the algorithm presented in the above-mentioned work to the case of multiple populations and to provide a more efficient upper bound $\tilde{\lambda}$. In particular, our approach allows for an efficient handling of rapidly increasing intensity functions.

4.1. Choice of an upper bound for the intensity

If $\bar{Z}_t = 0$, i.e. in the absence of any spike, it follows from (5) that $\bar{X}$ evolves as a linear ODE with matrix A, so that $\bar{X}_t = e^{At}x_0$. In particular, for all neurons $n = 1, \ldots, N_k$, it follows that

$$\lambda_t^{k,n} = f_k((e^{At}x_0)^{k,1}). \tag{25}$$

One possible choice for the dominating intensity $\tilde{\lambda}$ in the thinning procedure is to provide an upper bound of (25) which holds for all $t \geq 0$. A straightforward candidate for such a bound is provided in the following lemma.

Lemma 3. *For any $x \in \mathbb{R}^\kappa$, let $\Phi_k(x) = \sup_{t\geq 0} (e^{At}x)^{k,1}$. Then*

$$\Phi_k(x) \leq \tilde{\Phi}_k(x) = \max_{j=1,\ldots,\eta_k+1} \left\{ 0, \frac{x^{k,j}}{\nu_k^{j-1}} \right\}.$$

Proof. The explicit expression for $(e^{At}x)^{k,1}$ is given in (17); that is,

$$(e^{At}x)^{k,1} = e^{-\nu_k t} \left(x^{k,1} + tx^{k,2} + \cdots + \frac{t^{\eta_k}}{\eta_k!} x^{k,\eta_k+1} \right).$$

Setting $y_j = x^{k,j}/(v_k)^{j-1}$, one gets

$$(e^{At}x)^{k,1} = e^{-v_k t}\left(y_1 + tv_k y_2 + \cdots + \frac{(tv_k)^{\eta_k}}{\eta_k!}y_{\eta_k+1}\right) \leq \max_k\{0, y_k\}e^{-v_k t}g(t).$$

The result follows from the fact that $g(t) = 1 + tv_k + \cdots + (tv_k)^{\eta_k}/\eta_k! \leq e^{v_k t}$. $\qquad\square$

Remark 6. Another possible choice of a uniform bound, similar to the one given in Lemma 3, is provided in [16]. Their method, adapted to our case, gives

$$\Phi_k(x) \leq e \max\left\{1, \left(\frac{\eta_k}{ev_k}\right)^{\eta_k}\right\}\max_j\{x^{k,j}\},$$

which is larger, and thus less efficient, than the bound proposed in Lemma 3.

Since the functions f_k are non-decreasing, the upper bound of $(e^{At}x)^{k,1}$ given in Lemma 3 provides the bound

$$\tilde{f}_k(x) = f_k(\tilde{\Phi}_k(x)) \tag{26}$$

on the intensity. However, there is no guarantee that this bound is sharp. In most practical cases (especially when the functions f_k are increasing fast), the procedure rejects a vast majority of the simulated points. Hence, a more efficient approach, based on the computation of the critical points of the function $(e^{At}x)^{k,1}$, is proposed. Furthermore, instead of considering a bound for any $t > 0$, we choose a fixed time step $\tilde{\Delta} > 0$ (such that one spike is likely to occur in the interval $[0, \tilde{\Delta}]$) and compute $\Phi_k^{\tilde{\Delta}}(x) = \sup_{0 \leq t \leq \tilde{\Delta}}(e^{At}x)^{k,1}$ instead of $\Phi_k(x)$. This choice has no impact on the precision of the simulation. It only influences the sharpness of the bound used in the method and thus its computational efficiency.

Lemma 4. *For any $x \in \mathbb{R}^\kappa$, it holds that*

$$\Phi_k^{\tilde{\Delta}}(x) = \max_{0 < t_c < \tilde{\Delta}}\left\{x^{k,1}, \left(e^{At_c}x\right)^{k,1}, \left(e^{A\tilde{\Delta}}x\right)^{k,1}\right\},$$

where the maximum is taken over the critical points t_c of $t \mapsto (e^{At}x)^{k,1}$, which are the solutions of the equation

$$\left(-v_k x^{k,1} + x^{k,2}\right) + \cdots + \left(-v_k x^{k,\eta_k} + x^{k,\eta_k+1}\right)\frac{(t_c)^{\eta_k-1}}{(\eta_k-1)!} + \left(-v_k x^{k,\eta_k+1}\right)\frac{(t_c)^{\eta_k}}{(\eta_k)!} = 0.$$

Proof. The result follows from the computation of the time derivative of $(e^{At}x)^{k,1}$. $\qquad\square$

The critical points in Lemma 4 are given by polynomial roots, which can be accurately computed numerically. In most practical cases, the computational cost of the polynomial roots is compensated by the efficiency gained in the rejection method. Finally, let us define the upper-bound intensity function by

$$\tilde{f}_k^{\tilde{\Delta}}(x) = f_k(\Phi_k^{\tilde{\Delta}}(x)). \tag{27}$$

Note that when the population is inhibitory ($c_k = -1$), the naive upper bound $\tilde{f}_k$ is constant with respect to time, because all the coordinates of $\bar{X}^1$ are always negative and the bound given by Lemma 3 is 0. Thus, $\tilde{f}_k \equiv f_k(0)$. Of course, such a bound is not sharp in general. However, it is interesting to see how the two upper bounds $\tilde{f}_k$ and $\tilde{f}_k^{\tilde{\Delta}}$ behave for a particular realization of the intensity process for excitatory populations. Figure 3 gives a comparison of the paths of $\tilde{f}_2$ and
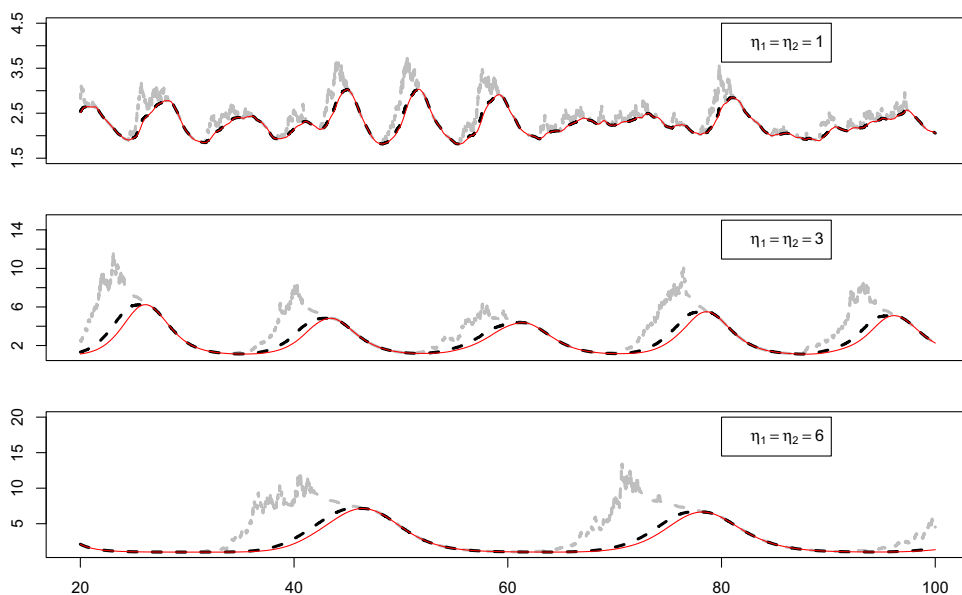
FIGURE 3. Intensity and intensity bounds for the second population (excitatory), $t \in [20, 100]$. Red solid line: true intensity λ_t^2. Black dashed line: $\tilde{f}_2^{\tilde{\Delta}}(\bar{X}_t)$. Grey dashed line: $\tilde{f}_2(\bar{X}_t)$. Intensity functions f_1 and f_2 are given in (28); $\nu_1 = \nu_2 = 0.9$, $N_1 = N_2 = 50$.

$\tilde{f}_2^{\tilde{\Delta}}$ for the excitatory population (with $\tilde{\Delta} \equiv 1$). We observe that both bounds are fairly precise when the potential $\bar{X}_t^2$ (or, respectively, the intensity process) is decreasing. On these intervals the differences between the three trajectories are negligible. However, the accuracy of $\tilde{f}_2$ drops drastically on the intervals where the intensity grows. In general, the higher the amplitude of the oscillations, the poorer the performance of the naive bound. This is particularly visible when illustrated on systems with high memory order ($\eta_k = 3$ or 6). For $\eta_k = 1$ both bounds perform well; however, $\tilde{f}_k^{\tilde{\Delta}}$ is clearly closer to the true process. The influence of the bound ($\tilde{f}_2$ or $\tilde{f}_2^{\tilde{\Delta}}$) on the execution time is discussed in Section 5.

4.2. Simulation algorithm

Now let us detail the recursive procedure, which is summarized in Algorithm 1. We choose a discrete time step $\tilde{\Delta}$ and a stopping time $t_{\max}$, and we fix the initial values $t_0 = 0$ and $\bar{X}_0 = x_0 \in \mathbb{R}^\kappa$. Let us assume that the procedure's current step is i, with current time t_i and potential value $\bar{X}_i$. Let us explain how t_{i+1} and $\bar{X}_{i+1}$ are obtained. One simulates two independent exponential variables τ_1 and τ_2 with respective parameters $N_k \tilde{f}_k^{\tilde{\Delta}}(x)$ (one for each population). They represent the waiting times to the next spikes of the dominant process $\tilde{Z}$ for each respective population. Then two cases may occur:

1. If $\min\{\tau_1, \tau_2\} > \tilde{\Delta}$, no spike occurs in the interval $[t_i, t_i + \tilde{\Delta}]$. We update $t_{i+1} = t_i + \tilde{\Delta}$ and $\bar{X}_{i+1} = e^{A\tilde{\Delta}} \bar{X}_i$.

2. If $\tau = \min\{\tau_1, \tau_2\} \leq \tilde{\Delta}$, then the dominating point process $\tilde{Z}$ emits a spike at time $t^* = t_i + \tau$. Let us denote by k^* the population with the smallest waiting time; that is, $\tau = \tau_{k^*}$. It remains to decide whether t^* is also a spiking time for the process Z.

Algorithm 1: Simulation of model (1) with $K = 2$ populations.

Input: intensity functions f_1 and f_2; integers N_1, N_2, η_1, and η_2; real numbers
c_1, c_2, ν_1, ν_2, $\tilde{\Delta}$, and $t_{\max}$; real vector $x_0 \in \mathbb{R}^\kappa$.

Output: point processes $(Z^{k,n})_{k=1,2;\, n=1,\dots,N_k}$, Markovian cascade process $\bar{X}$,
and intensity processes $(\lambda^k)_{k=1,2}$.

Initialization: $t \leftarrow 0$, $x \leftarrow x_0$;

while $t < t_{\max}$ **do**

> $\tilde{\lambda}_k \leftarrow \tilde{f}_k^{\tilde{\Delta}}(x)$ for $k = 1, 2$;
>
> draw $\tau_k \sim \mathcal{E}(N_k \tilde{\lambda}_k)$ for $k = 1, 2$;
>
> $\tau \leftarrow \min_k \tau_k$ and $k^* \leftarrow \arg\min_k \tau_k$;
>
> **if** $\tau > \tilde{\Delta}$ **then**
>
> (1) > > $t \leftarrow t + \tilde{\Delta}$ and $x \leftarrow e^{A\tilde{\Delta}} x$;
>
> **else**
>
> > $t \leftarrow t + \tau$, $x \leftarrow e^{A\tau} x$, $\lambda_{k^*} \leftarrow f_{k^*}(x)$;
> >
> > draw $U \sim \mathcal{U}([0,1])$;
> >
> > **if** $U < \lambda_{k^*}/\tilde{\lambda}_{k^*}$ **then**
> >
> > (2) > > > draw $n \sim \mathcal{U}(\{1, \dots, N_{k^*}\})$ and add t to the list $Z^{k^*,n}$;
> > >
> > > $x \leftarrow x + \Gamma \mathbb{1}(k^*)$;
> > >
> > > add x to the list $\bar{X}$ and $\lambda_k = f_k(x)$ to the list λ^k for $k = 1, 2$;
> >
> > **else**
> >
> > (3) > > > do nothing;

If not, this point is discarded. We draw a uniform variable U on $[0, 1]$ and define the threshold R:

$$R := \frac{f_{k^*}\left(e^{A\tau}\bar{X}_i\right)}{\tilde{f}_{k^*}^{\tilde{\Delta}}(\bar{X}_i)}; \qquad R \in [0, 1] \text{ by the definition of } \tilde{f}_k^{\tilde{\Delta}}(x).$$

- If $U \geq R$, then t^* is discarded, i.e., no spike occurs in the interval $[t_i, t^*]$. We update $t_{i+1} = t^*$ and $\bar{X}_{i+1} = e^{A\tau}\bar{X}_i$.

- If $U < R$, then t^* is kept, i.e., we add t^* to the list of one neuron of population k^* chosen uniformly at random. We update $t_{i+1} = t^*$ and $\bar{X}_{i+1} = e^{A\tau}\bar{X}_i + \Gamma \mathbb{1}(k^*)$, where $\mathbb{1}(k^*) = (\mathbb{1}_{k^*=1}, \mathbb{1}_{k^*=2})^T$.

Finally, the execution is stopped once $t_i \geq t_{\max}$, i.e. once the time horizon of interest is reached. As output the algorithm returns a list of the spiking times of each neuron and the values of the

processes $\bar{X}$ and λ at the spiking times. At this stage it is clear why it is important to have a sharp upper bound. The closer the threshold R is to 1, the fewer points are rejected.

Algorithm 1 is most efficient when every iteration of the **while** loop enters the condition (**2**). Of course, that ideal case does not occur in practice. When lowering the value of $\tilde{\Delta}$, the number of loops satisfying the condition (**3**) decreases, because the dominating intensity $\tilde{\lambda}$ is getting smaller. On the other hand, the number of loops fulfilling the condition (**1**) increases, because the exponentially distributed times have greater chances of being larger than $\tilde{\Delta}$. The calibration of $\tilde{\Delta}$ is a difficult problem which is not addressed here. In practice, it is observed that the execution time is not very sensitive to the value of $\tilde{\Delta}$. The main bottleneck of the thinning algorithm is the sharpness of the intensity bound. When the intensity functions are exponential, the execution time is more than halved with the bound of Lemma 4 compared to the bound of Lemma 3. This is illustrated in the right panel of Figure 8.

5. Numerical experiments

We now describe a simulation study illustrating the theoretical results discussed in the previous sections. It consists of two steps. First, we study the performance of the proposed splitting schemes. More precisely, we compare the Lie–Trotter (20) and Strang (21) splitting schemes with the Euler–Maruyama approximation. We report sample paths and empirical densities and comment also on the first and second moments. This step follows the numerical study in [3], and shows that the Strang splitting performs best. Second, we compare the diffusion process (simulated with the property-preserving Strang splitting scheme) to the PDMP, varying the number of neurons N. In particular, when comparing the long-time behaviour of the processes (see Figure 10), we show that the diffusion approximation is less and less accurate as $t \to +\infty$. This confirms the results obtained in Theorems 1 and 2.

Following the work of [15], throughout this section we use the following intensity functions:

$$f_1(x) = \begin{cases} 10e^x & \text{if } x < \log 20), \\ \frac{400}{1+400e^{-2x}} & \text{if } x \geq \log(20), \end{cases} \qquad f_2(x) = \begin{cases} e^x & \text{if } x < \log(20), \\ \frac{40}{1+400e^{-2x}} & \text{if } x \geq \log(20). \end{cases} \qquad (28)$$

Further, we fix the parameters $c_1 = -1$, $c_2 = 1$ and consider $N_1 = N_2$. Unless stated differently, throughout this section the initial condition is fixed to $x_0 = 0_\kappa$. The parameter p_k is then defined as N_k/N. The fact that $c_1 c_2 < 0$ ensures that the population shows an oscillatory behaviour, for certain parameters ν_k and η_k (see [15] for further details).

5.1. Comparison of the Euler–Maruyama method and the splitting schemes

In this section we are interested in comparing the performance of the splitting schemes with that of the frequently applied Euler–Maruyama method, for varying time steps Δ. The parameter values $\nu_1 = \nu_2 = 1$, $\eta_1 = 3$, $\eta_2 = 2$, $N_1 = N_2 = 50$ are used, and the dimension of the system is thus $\kappa = 7$. Except for the density and mean-square convergence plots, we consider the time interval [0,100]. Unless stated otherwise, we plot the variables $X_t^{k,1}$ for $k = 1, 2$ in black and the remaining $\eta_1 + \eta_2$ auxiliary memory variables in grey.

5.1.1 *Illustration of the mean-square convergence order.* We start our study by comparing the convergence rates of the Euler–Maruyama method and the Lie–Trotter (20) and Strang (21)
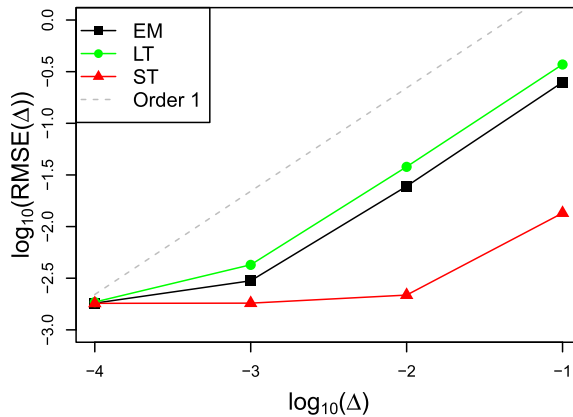
FIGURE 4. Mean-square order of convergence. The reference solution is obtained with the Euler–Maruyama method and the small time step $\Delta = 10^{-5}$. The numerical solutions are calculated for $\Delta = 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$. The time t^* is 1, and $M = 10^3$.

splitting schemes. The root-mean-square error (RMSE), approximating the left side of the equation in Theorem 5, is defined as

$$\text{RMSE}(\Delta) := \left(\frac{1}{M} \sum_{l=1}^{M} \| X_{t^*}^{(l)} - \tilde{X}_{t^*}^{(l)} \|^2 \right)^{1/2},$$

where $X_{t^*}^{(l)}$ and $\tilde{X}_{t^*}^{(l)}$ denote the values at a fixed time t^* of the lth simulated trajectory of the true process and its numerical approximation, respectively. The integer M is the total number of simulated differences. The value of the true process $X_{t^*}^{(l)}$ is obtained from the Euler–Maruyama scheme, using the small time step $\Delta = 10^{-5}$. The number of simulations is fixed to $M = 10^3$ and $t^* = 1$.

We report the RMSE in Figure 4, where the x-axis corresponds to the logarithm (base 10) of the time step Δ and the y-axis corresponds to the logarithm (base 10) of the RMSE. The theoretical rate of convergence obtained in Theorem 5 (all considered schemes converge with order 1) is confirmed empirically. While the Lie–Trotter splitting and the Euler–Maruyama scheme show a similar RMSE for varying Δ, the RMSE obtained for the Strang splitting method is significantly smaller for all Δ under consideration, implying higher efficiency for that scheme. We stress, however, that from the fact that the rate of convergence is the same, it does not follow that they share the same qualitative properties when the step size Δ is fixed.

5.1.2 *Illustration of the qualitative properties of the splitting schemes.* We now illustrate how the proposed splitting schemes preserve the structure (e.g. the moments and the underlying invariant distribution) of the process, even for large values of Δ, while the Euler–Maruyama method may fail to do so. We start by studying sample trajectories (see Figure 5). All three methods yield comparable performance when $\Delta = 0.01$. For $\Delta = 0.5$, the Euler–Maruyama scheme preserves the oscillations, but does not preserve the amplitude. The behaviour of the inhibitory population is less accurately approximated than that of the excitatory one. This problem worsens as Δ increases further to 0.7. An interesting observation is that, for time steps Δ that are not 'small enough', the Euler–Maruyama scheme may not preserve the mean of the
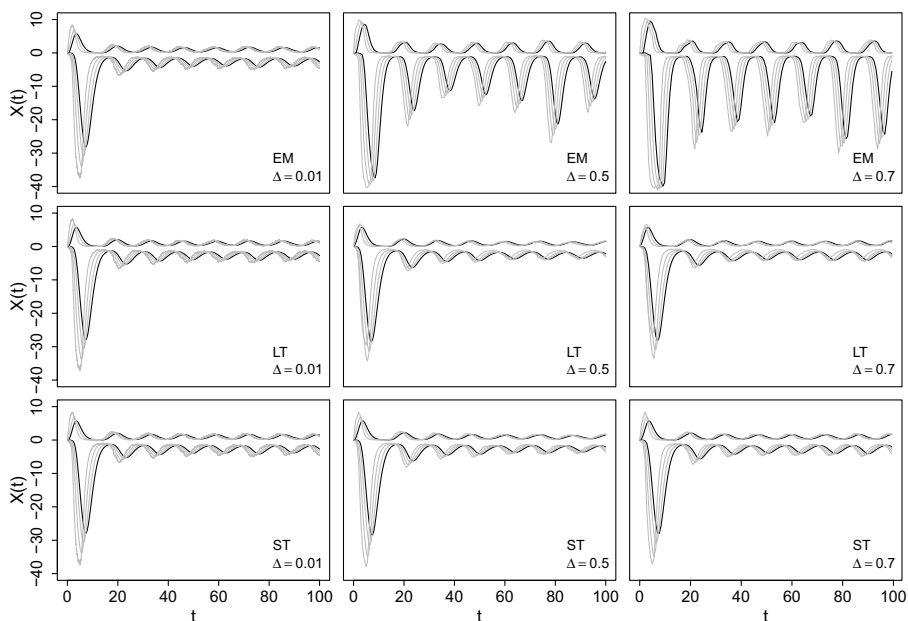
FIGURE 5. Sample trajectories of the system, simulated with the Euler–Maruyama scheme (top), the Lie–Trotter splitting scheme (middle), and the Strang splitting scheme (bottom), for varying Δ.

process (see also Figure 7). Indeed, it has been observed that the Euler–Maruyama method preserves the first but not the second moment (see e.g. [3, 20]). In other words, the amplitude of the oscillations grows, but the mean is unchanged. In our case, however, since the trajectories are bounded by 0 from below or above (depending on the sign of c_k), the increased amplitude also introduces a bias in the first moment. Thus, the Euler–Maruyama approximation of the system (7) preserves neither the first nor the second moment. In contrast, the Lie–Trotter and Strang splitting schemes show comparably good performance. However, the Lie–Trotter splitting is less accurate in reproducing the delay between the current state of the process (black line) and the memory variables (grey lines) in the beginning of the interval, where the amplitude of the oscillations is large (see also Figure 7).

The difference between the schemes becomes clearer as we look at the phase portrait of the system (Figure 6). We observe again that both splitting schemes yield satisfactory approximations (for all Δ under consideration), the Strang approach slightly outperforming the Lie–Trotter method. In contrast, the phase portrait obtained with the Euler–Maruyama approximation fails to reproduce the behaviour of the process for $\Delta = 0.5$ or $\Delta = 0.7$.

Similar conclusions can be drawn from Figure 7, where we visualize the marginal densities of the process. Each visualized density is estimated with a standard kernel density estimator, based on a simulated long-time trajectory ($T = 10^5$) of each variable of the process. We observe again that the Euler–Maruyama method may not preserve the mean of the process (red dashed vertical lines). Moreover, the Euler–Maruyama scheme may even suggest a transition from a unimodal to a bimodal density as Δ increases.

5.2. Comparison of the PDMP and the diffusion

Now we are interested in comparing the PDMP process $\bar{X}$, simulated with the thinning algorithm detailed in Section 4, and the stochastic diffusion X, simulated with the
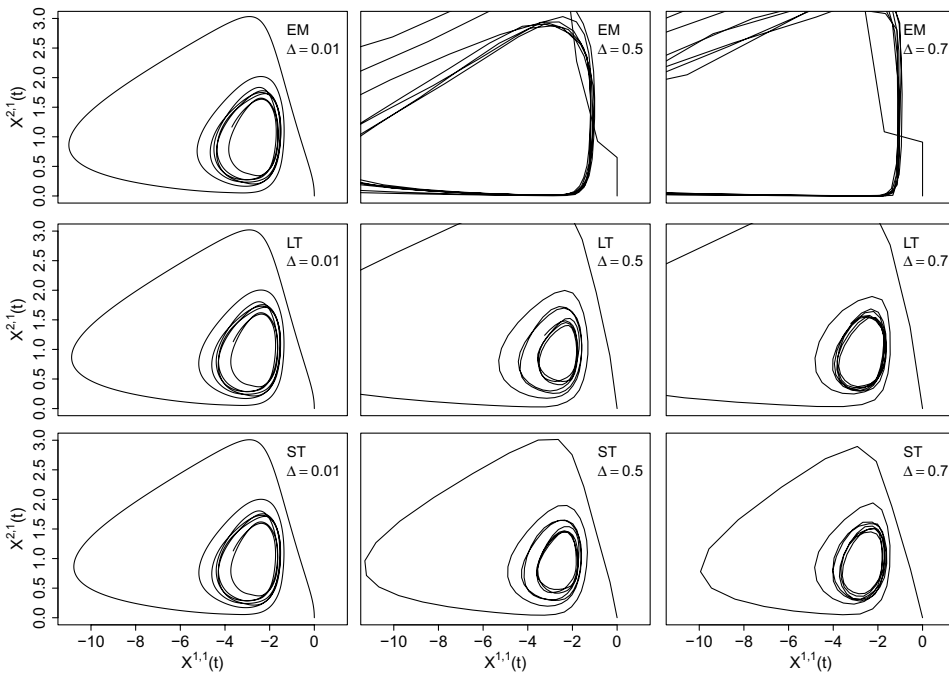
FIGURE 6. Phase portrait of the main variables, simulated with the Euler–Maruyama scheme (top), the Lie–Trotter splitting scheme (middle), and the Strang splitting scheme (bottom), for varying Δ and $x_0 = (0, 0, -3.5, -4, 0, 1.3, 1.1)$.

property-preserving Strang splitting scheme introduced in Section 3. We simulate the trajectories of the stochastic diffusion process with the Strang splitting scheme, since it showed the best performance in the previous section.

5.2.1 *Execution time.* As a first step we consider the execution time. We compare the numerical cost of the simulation of the process $\bar{X}$ with two different intensity bounds (based on Lemmas 3 and 4) to that of the simulation of the diffusion X with the Strang splitting scheme.

We set $t_{\max} = 100$ and vary the total number of neurons, taking $N = 20, 50, 100, 150, 200$ and $N_1 = N_2$. In the case of the diffusion simulation, the parameter N does not influence the computational cost. Thus, we report the execution time for the diffusion simulation only for $N = 200$, taking $\Delta = 0.1$, and report it as a reference value. The time step $\tilde{\Delta}$ for the thinning procedure is defined in an adaptive way within the **while** loop of Algorithm 1. In each step we use the last computed value of the intensities λ_k and set $\tilde{\Delta}$ equal to $(N_1\lambda_1 + N_2\lambda_2)^{-1}$. This choice takes into account the scaling with respect to the number of neurons and the dynamics of the intensities. For instance, $\bar{X}^{2,1}$ roughly belongs to $[0, 2]$ (see Figure 10), so that the intensity of population roughly belongs to $[1, 7]$ (with the intensity functions defined in (28)).

In Figure 8, two different sets of intensity functions, 'truncated linear' ones and 'exponential' ones, are studied. The plots show, in seconds, the mean execution times (over 100 realizations) required to simulate the process on the interval $[0, t_{\max}]$, using the bounds $\tilde{f}(x)$ and $\tilde{f}^{\Delta}(x)$ (defined in (26) and (27), respectively).

Note that there is almost no difference in the performance of the algorithm with different bounds in the linear case (left panel of Figure 8). That means that the bound obtained in Lemma 3 is sharp enough. Note also that since $f_k^{\max} = 10$, there occur only a few spikes and
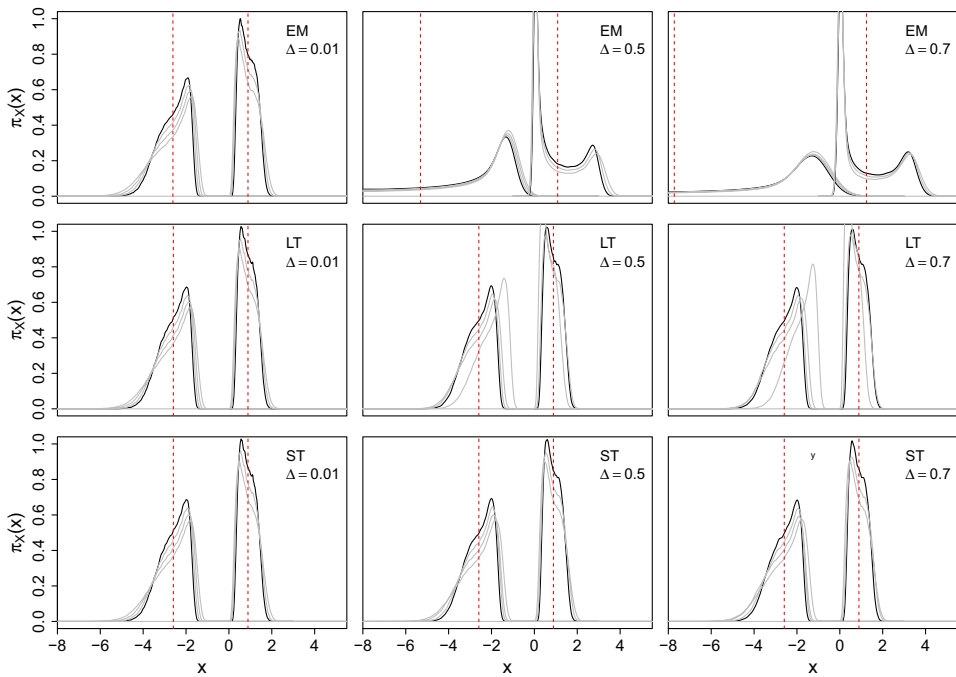
FIGURE 7. Empirical density of the system, simulated with the Euler–Maruyama scheme (top), the Lie–Trotter splitting scheme (middle), and the Strang splitting scheme (bottom), for varying Δ and $T = 10^5$. The red dashed vertical lines denote the mean of the main variables.
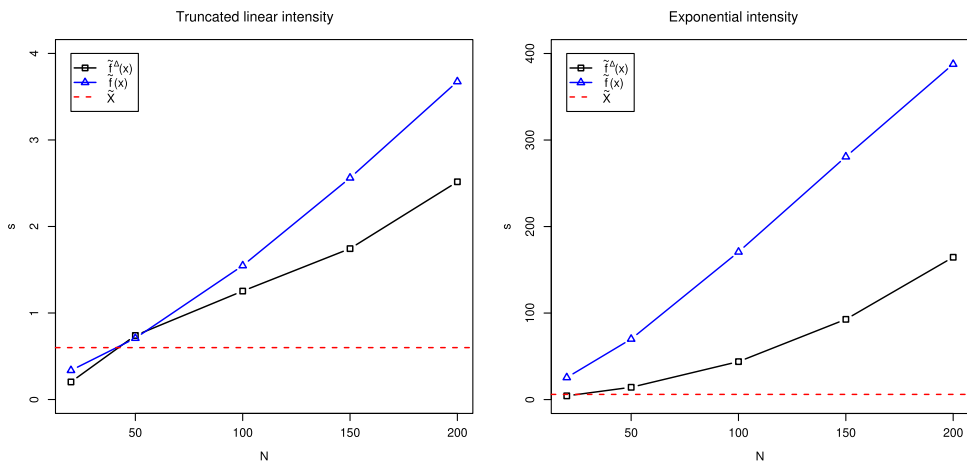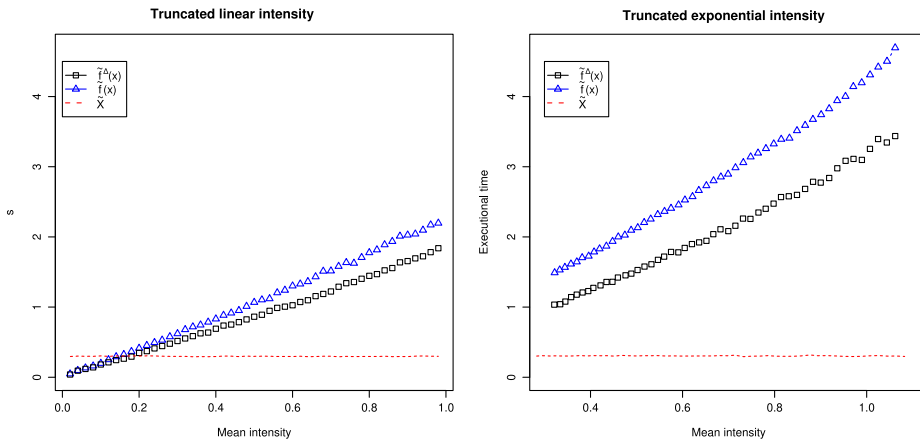

FIGURE 8. Mean execution time for the PDMP (solid line) and the stochastic diffusion (dashed line) simulation for $t_{\max} = 100$ over 100 realizations. Left panel: $f_1(x) = f_2(x) = \min\{1 + x\mathbb{1}_{[x>0]}, 10\}$. Right panel: f_1 and f_2 are given by Equation (28). The rest of the parameters are given in the beginning of Section 5.

FIGURE 9. Mean execution time for the PDMP and the stochastic diffusion (dashed line) simulation for $t_{\max} = 50$ over 100 realizations. Left panel: $f_1(x) = f_2(x) = \min\{a + x\mathbb{1}_{[x>0]}, 40\}$. Right panel: $f_1(x) = f_2(x) = \min\{a + e^x, 40\}$, where the parameter a varies between 0.02 and 1 with a step of 0.02. The rest of the parameters are given in the beginning of Section 5; $N_1 = N_2 = 25$.

the process is simulated very fast. However, in the case of an exponential intensity (right panel of Figure 8), the execution time drastically increases. The process is simulated at least twice as fast with the local bound. The main reason is that the local bound $\tilde{f}_k^\Delta(x)$ rejects around 2% of spikes, while the bound $\tilde{f}_k(x)$ rejects around 90%. In general, we can conclude that the execution time depends linearly on the number of neurons for both the local and the general bound. Disregarding the bound chosen, neither algorithm can compete with the time required for simulating the diffusion. For $\Delta = 0.1$ and $T = 100$, the average running time with the exponential firing rate function is equal to 0.598s (with standard deviation 0.12s). For the linear one it is 0.597s (with standard deviation 0.15s). Thus, the execution time for the diffusion approximation does not depend on the firing rates.

Obviously, the execution time of the thinning algorithm depends on the number of spikes of the system per time unit, which is well approximated by the temporal mean intensity (formally defined by $T^{-1} \int_0^T \lambda_s^{k,n} ds$). Hence, for a fixed number of neurons N, a modification of the intensity functions f_k changes the execution time. In Figure 9, we report the mean execution time over 100 repetitions of the experiment, plotted against the temporal mean value of the inhibitory population for a fixed number of neurons $N_1 = N_2 = 25$. For simplicity, we take the same intensity function for both populations, but we add a baseline intensity parameter a in order to control the temporal mean intensity: more precisely, $f_1(x) = f_2(x) = \min\{a + x\mathbb{1}_{[x>0]}, 40\}$ for the truncated linear case, and $f_1(x) = f_2(x) = \min\{a + e^x, 40\}$ for the truncated exponential case. As a increases, the temporal mean intensity increases and so does the execution time: the latter even increases linearly with respect to the mean intensity, whatever bound method is used for the algorithm. Comparing the two plots in Figure 9, it appears that the truncated exponential case is computationally more expensive (the values and the slopes are larger for both bounds). Once again, this comes from the fact that the proportion of rejected spikes is larger in the truncated exponential case. In both cases, the dependency on the temporal mean value is linear (which further confirms the results presented in Figure 8). Finally, notice that the execution time for the stochastic diffusion approximation does not depend on the mean intensity.
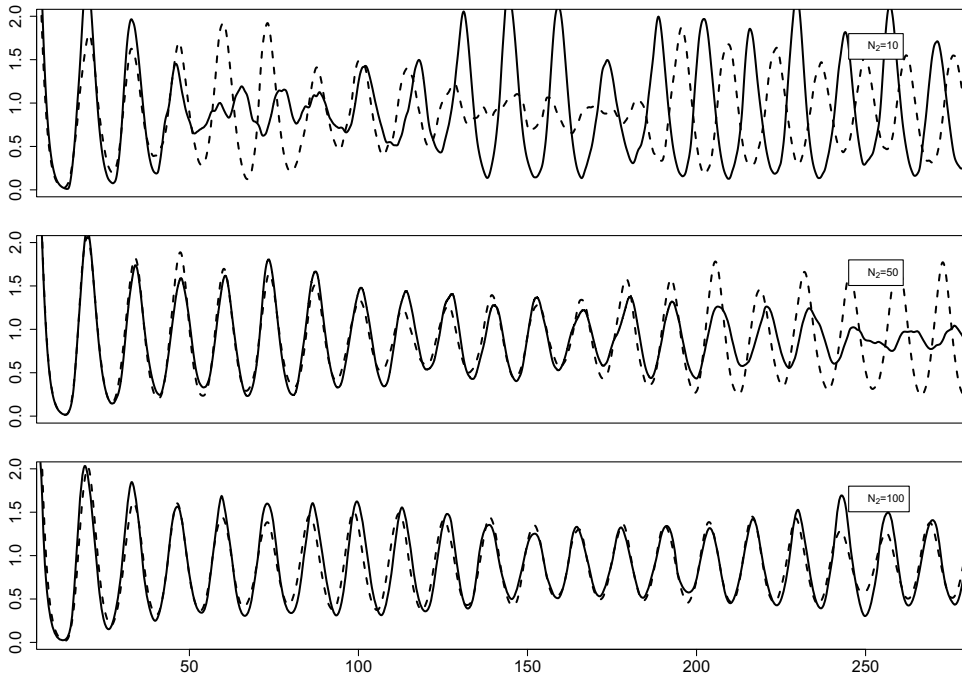
FIGURE 10. Sample trajectories of the PDMP and the diffusion for varying N (excitatory population). Solid line: main variable of $\bar{X}$. Dashed line: main variable of X (simulated with the Strang splitting scheme, using $\Delta = 0.01$).

A summary of the performances of both frameworks (stochastic diffusion and PDMP), with respect to the parameters of the model, is given below.

- In both cases, the execution time increases as the dimension of the system grows, i.e. as η_k increases.

- For the stochastic diffusion, the execution time depends, in a linear manner, on the step size Δ.

- For the PDMP, the execution time mainly depends, in a linear manner, on the temporal mean intensities of the two populations. They in turn depend linearly on the number N of neurons. They also depend, in a complex nonlinear manner, on the parameters v_k, η_k, and f_k.

- The computational cost of the simulation of the stochastic diffusion does not depend on the number of neurons N or the temporal mean intensity. Hence, it should be preferred to the simulation of the PDMP unless the number of neurons N or the temporal mean intensity is small.

5.2.2 *Qualitative properties.* It remains to determine whether the stochastic diffusion can really capture the behaviour of the underlying PDMP. To get an intuitive idea of how different processes behave when the number of neurons changes, we look at some sample trajectories. We take one realization of the PDMP and the stochastic diffusion process on a time interval of length $T = 300$ and plot them in Figure 10, cutting the short interval of time in the very
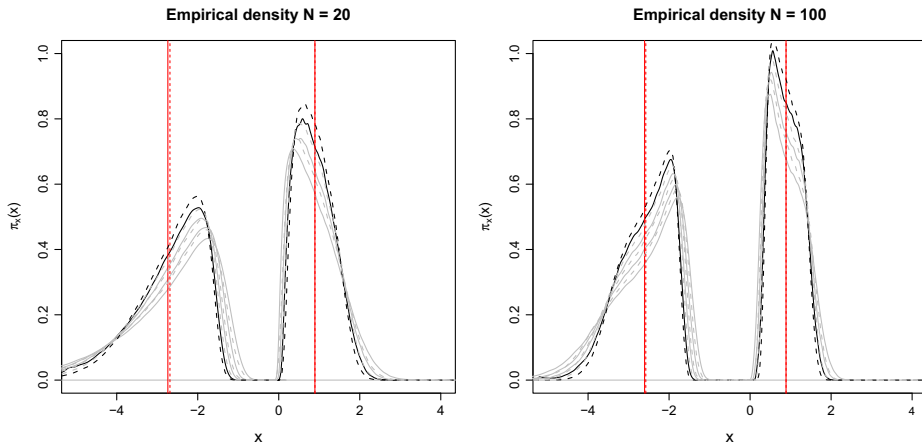
FIGURE 11. Empirical density of PDMP and the diffusion for $N = 20$ (left) and $N = 100$ (right). Solid line: $\bar{X}$. Dashed line: X (simulated with the splitting scheme, $\Delta = 0.1$, $T = 10^5$). The red solid and dashed vertical lines denote the mean of the respective main variables.

beginning in order to observe the process in its oscillatory regime. For simplicity, we focus only on the second (excitatory) population. The trajectories in the top panel are simulated with $N_2 = 10$, those in the middle panel with $N_2 = 50$, and those in the bottom panel with $N_2 = 100$.

Let us mention that Figure 10 is not an illustration of Theorem 2. Indeed, the trajectories are not coupled in such a way that (12) is satisfied. To our knowledge, there is no such result in the literature, and the coupling involved in the proof of Theorem 2 is not explicit. However, the figure illustrates the fact that the fluctuations of both trajectories vanish as N goes to infinity and that both converge to the solution of the ODE (10).

As a final step, we are interested in the long-time behaviour of the processes. We simulate both processes ($\bar{X}$ and X) on a long-time interval, taking $T = 10^5$, and report the respective marginal empirical densities in Figure 11. The densities of the PDMP are plotted with solid lines and those of the diffusion with dashed lines. Even for small N, the difference between the densities is negligible and their means are almost overlapping. As the number of neurons N increases, we observe that the empirical densities converge to some compactly supported distribution. Note that the mean-field limit is given by the ODE (10) as illustrated in Figure 10. Thus we expect that the support of the limit distribution is given by the amplitude of the solution of the ODE.

Conclusions

This work is thought to complement the papers [15] and [16]. First, we bridge the gap between the PDMP (5) and the solution of the SDE (7) by proving a strong error bound on the distance between the two. Second, moment bounds of the diffusion process are derived.

Furthermore, since the SDE (7) cannot be solved explicitly, two approximation schemes, based on the Lie–Trotter and the Strang splitting approaches, are proposed. They are proved to converge with mean-square order 1 and to preserve the properties of the model. In particular, the advantage of the proposed approximation methods is that they make full use of the matrix exponential e^{At}, which describes the flow of the PDMP (5). Thanks to this we are able to propagate the noise through all components of the system, thus preserving its hypoellipticity. Moreover, we show that the splitting schemes accurately reproduce the derived first and second

moment bounds and that they preserve the ergodicity of the continuous process, even for large values of the discretization step Δ.

These properties are particularly important when embedding the numerical scheme, for instance, into a statistical inference procedure. For example, maximum likelihood estimation techniques require the existence of a non-degenerate covariance matrix of the discretized process. For simulation-based inference methods (see [7]), the performance of the Euler–Maruyama method may be acceptable for 'small enough' time steps. However, the use of smaller time steps drastically increases the computational cost, making inference based on the Euler–Maruyama method computationally infeasible. Moreover, even for arbitrarily small time steps there is no guarantee that the Euler–Maruyama scheme preserves the model properties.

In addition, an exact simulation procedure for the PDMP is proposed. A sharp upper bound is provided, in order to yield an efficient procedure, and its performance is compared to that of the procedure given in [16]. When the number of neurons increases, the computational cost required for the PDMP simulation rises rapidly and cannot compete with the simulation of the stochastic diffusion (especially for the simulation via the splitting scheme).

The PDMP and the stochastic diffusion show similar behaviour. In particular, they possess matching empirical densities. Thus, we conclude that the stochastic diffusion describes the behaviour of the original neuronal model with very good precision and at negligible computational cost, compared to the PDMP.

To conclude, let us point out possible extensions of the presented work. First, it would be interesting to generalize the approach to the system with a non-cyclic feedback. Note that, if the coordinates directly perturbed by a Brownian motion exclusively depend on non-directly perturbed components, then the generalization of the proposed numerical schemes, based on the subsystems (18)–(19), is straightforward. But the formal proofs would involve more tedious computations. Second, to be as general as possible, it would be interesting to study the system behaviour in the case of unbounded intensity functions. In this case, there is no guarantee that the PDMP converges to the solution of the SDE in a mean-field limit. Even if this limit exists, proving the ergodic property and the moment bounds of the resulting process would require very different tools of stochastic and numerical analysis.

Acknowledgements

Appendix A. Proofs

A.1 Proof of Theorem 2

The proof of Theorem 2 is mainly based on two lemmas, which are stated before the proof. The first lemma concerns the coupling between a Poisson process and a Brownian motion. Its proof can be found in [17, Section 5.5] (the exponential moments can be deduced from the proof of Corollary 5.5.5).

Lemma 5. *A standard Poisson process* $(\Pi_t)_{t\geq0}$ *and a standard one-dimensional Brownian motion* $(B_t)_{t\geq0}$ *can be constructed on the same probability space, such that*

$$\sup_{t\geq0} \frac{|\Pi_t - t - B_t|}{\log(2\vee t)} \leq \Xi < \infty$$

almost surely, where Ξ *is a random variable having exponential moments.*

The second lemma concerns the modulus of continuity for the Brownian motion. It is stated in [23] without a proof. Hence, for the sake of completeness, we provide a proof which is an adaptation of the arguments presented in the appendix of [18].

Lemma 6. *Let B be a standard Brownian motion and T a positive time. Then*

$$M := \sup_{t,s\leq T} \frac{|B_t - B_s|}{\sqrt{|t-s|(1+\log(T/|t-s|))}}$$

is a finite random variable such that M^2 *has finite exponential moments.*

Proof. Thanks to the scaling properties of Brownian motion, it is sufficient to prove the statement for $T = 1$. In accordance with [18], let $c > 1$ and define two increasing functions Ψ and μ by

$$\Psi(x) = e^{x^2/2} - 1 \quad \text{and} \quad \mu(x) = \sqrt{cx},$$

for all $x \geq 0$. Now let ξ be the random variable defined by

$$\xi = \int_0^1 \int_0^1 \Psi\left(\frac{|B_t - B_s|}{\mu(|t-s|)}\right) dt ds.$$

The Garsia–Rodemich–Rumsey inequality [42, Theorem 2.1.3] implies that

$$|B_t - B_s| \leq 8 \int_0^{|t-s|} \Psi^{-1}\left(\frac{4\xi}{x^2}\right) \mu'(x) dx,$$

with $\Psi^{-1}(y) = \sqrt{2\log(1+y)}$ and $\mu'(x) = (\sqrt{c}/2)x^{-1/2}$. Yet, for $0 < x < 1$,

$$\Psi^{-1}\left(\frac{4\xi}{x^2}\right) = \sqrt{2}\sqrt{\log(4\xi+x^2)+2\log(1/x)} \leq \sqrt{2}\sqrt{\log(4\xi+1)} + 2\sqrt{\log(1/x)}.$$

Combining the last two equations, one gets, for all h,

$$\sup_{|t-s|\leq h} |B_t - B_s| \leq 4\sqrt{2c}\sqrt{\log(4\xi+1)} \int_0^h \frac{dx}{\sqrt{x}} + 8\sqrt{c}\int_0^h \sqrt{\log(1/x)}\frac{dx}{\sqrt{x}}. \tag{29}$$

The second term can be bounded thanks to

$$\int_0^h \sqrt{\log(1/x)}\frac{dx}{\sqrt{x}} = \int_0^h \left(\sqrt{\log(1/x)} - \frac{1}{\sqrt{\log(1/x)}}\right) + \frac{1}{\sqrt{\log(1/x)}}\frac{dx}{\sqrt{x}}$$

$$\leq 2\sqrt{h\log(1/h)} + 4\sqrt{h},$$

using (when $h > e^{-1}$) the fact that

$$\int_{e^{-1}}^{h} \frac{1}{\sqrt{x \log{(1/x)}}}dx = \int_{e^{-1}}^{h} 2\sqrt{x}\frac{1}{x\sqrt{\log{(1/x)}}}dx \leq 2\sqrt{h}(1 - \sqrt{\log{(1/h)}}).$$

Going back to Equation (29), for some constant C which does not depend on c, one has that

$$\sup_{|t-s|\leq h} |B_t - B_s| \leq C\sqrt{c} \left(\sqrt{\log{(4\xi + 1)}} + 1 \right) \sqrt{h(1 + \log{(1/h)})}.$$

Note that the random variable M defined in the statement of the lemma satisfies

$$M = \sup_{0<h<1} \frac{\sup_{|t-s|\leq h} |B_t - B_s|}{\sqrt{h(1 + \log{(1/h)})}} \leq C\sqrt{c} \left(\sqrt{\log{(4\xi + 1)}} + 1 \right),$$

so that

$$\mathbb{E}\left[e^{\alpha M^2} \right] \leq \mathbb{E}\left[e^{2\alpha c C^2 (\log{(4\xi+1)}+1)} \right] \leq e^{2\alpha c C^2} \mathbb{E}\left[(4\xi + 1)^{2\alpha c C^2} \right].$$

To conclude, we refer to the control of the moments of ξ given in the appendix of [18]. This states in particular that $\mathbb{E}[(4\xi + 1)^{2\alpha c C^2}]$ is finite as soon as $2\alpha c C^2 < c$, which is granted if we take α small enough. $\qquad\square$

Before going through the proof of the Theorem, let us give an alternative representation of Equation (5) and a sketch of the proof. Thanks to the time-change property of point processes (see [6, Section II.6], for instance), there exist two independent standard (i.e. with rate equal to one) Poisson processes Π^1 and Π^2 such that $\bar{Z}_t^k = N_k^{-1} \Pi_{\bar{\Lambda}_t^k}^k$, where $\bar{\Lambda}_t^k$ is the integrated intensity of $\bar{Z}_t^k$, that is,

$$\bar{\Lambda}_t^k = N_k \int_0^t f_k(\bar{X}_s^{k+1,1})ds.$$

This time-change property is an analogous martingale property to the time-change property for diffusions. The integrated form of (5) is then given by

$$\bar{X}_t = x_0 + \int_0^t A\bar{X}_s ds + c\bar{Z}_t = \int_0^t A\bar{X}_s ds + c \begin{pmatrix} N_1^{-1} \Pi_{\bar{\Lambda}_t^1}^1 \\ N_2^{-1} \Pi_{\bar{\Lambda}_t^2}^2 \end{pmatrix}. \tag{30}$$

In a similar way, the SDE can be written with respect to two time-changed Brownian motions. The general idea of the proof is then to couple the standard Poisson processes Π^k with the Brownian motions.

Proof of Theorem 2. It is more convenient to first prescribe the Brownian motions and then couple them with Poisson processes. That is exactly how we proceed below. Let Y be the solution of (7) with respect to some two-dimensional Brownian motion $W = (W^1, W^2)^T$. Thanks to the time-change property of the Brownian motion (see [17, Theorem 2.12], for instance), we let B^k be the Brownian motion defined by

$$B_t^k = \int_0^{\tau^k(t)} \sqrt{N_k f_k(Y_s^{k,1})}dW_s^k,$$

where $\tau^k(t)$ is the stopping time satisfying

$$t = N_k \int_0^{\tau^k(t)} f_k(Y_s^{k,1}) ds.$$

Then Y can be written as follows:

$$Y_t = x_0 + \int_0^t AY_s ds + \int_0^t \Gamma \begin{pmatrix} f_2(Y_s^{2,1}) \\ f_1(Y_s^{1,1}) \end{pmatrix} ds + \Gamma \begin{pmatrix} N_1^{-1} B_{\Lambda_t^1}^1 \\ N_2^{-1} B_{\Lambda_t^2}^2 \end{pmatrix}, \tag{31}$$

where

$$\Lambda_t^k = N_k \int_0^t f_k(Y_s^{k,1}) ds.$$

We are now in a position to use the coupling with Poisson processes. Let Π^k be the Poisson process given by Lemma 5 with associated random variable Ξ_k. Let $\bar{X}$ be defined as in (30). Then

$$\bar{X}_t = x_0 + \int_0^t A\bar{X}_s ds + \int_0^t \Gamma \begin{pmatrix} f_2(\bar{X}_s^{2,1}) \\ f_1(\bar{X}_s^{1,1}) \end{pmatrix} ds + \Gamma \begin{pmatrix} N_1^{-1} B_{\bar{\Lambda}_t^1}^1 + R_t^1 \\ N_2^{-1} B_{\bar{\Lambda}_t^2}^2 + R_t^2 \end{pmatrix}, \tag{32}$$

where

$$R_t^k = \frac{1}{N_k} \left(\Pi_{\bar{\Lambda}_t^k}^k - \bar{\Lambda}_t^k - B_{\bar{\Lambda}_t^k}^k \right).$$

Thanks to Lemma 5,

$$|R_t^k| \le \frac{1}{N_k} \Xi_k \log(2 \vee \bar{\Lambda}_t^k) \le \Xi_k \left(\frac{\log N_k}{N_k} + \frac{\log t}{N_k} + \frac{1}{N_k} \right)$$

$$\le C \Xi_k \left(\frac{\log N}{N} + \frac{\log t}{N} + \frac{1}{N} \right), \tag{33}$$

for some constant C, where we used that $\bar{\Lambda}_t^k \le N_k t f_k^{max}$ and N/N_k is bounded for N and N_k large enough.

Let us write $G^N(t) = \sup_{s \le t} N\|\bar{X}_s - Y_s\|$; here and below, $\|\cdot\|$ denotes the sup norm on $\mathbb{R}^\kappa$. Combining (31) and (32), as well as using the Lipschitz continuity of f_k (with respect to constant L_k), we have

$$\|\bar{X}_t - Y_t\| \le \int_0^t \|A(\bar{X}_s - Y_s)\| ds + \max\{|c_1|L_2, |c_2|L_1\} \int_0^t \|\bar{X}_s - Y_s\| ds$$

$$+ \max_k \left\{ |c_{k-1}| \left(N_k^{-1} \left| B_{\bar{\Lambda}_t^k}^k - B_{\Lambda_t^k}^k \right| + |R_t^k| \right) \right\}.$$

Then, since the operator norm $\|A\|$ corresponding to the sup norm is finite, Grönwall's lemma yields

$$G^N(T) \le C_1 \max_k \left\{ \sup_{t \le T} \left| B_{\bar{\Lambda}_t^k}^k - B_{\Lambda_t^k}^k \right| + N|R_t^k| \right\} e^{C_2 T} \tag{34}$$

for two deterministic constants C_1 and C_2 which depend on neither N nor T. Hence, it only remains to estimate the Brownian increments. This can be done via the modulus of

continuity of Brownian motion. Indeed, for $t \leq T$, $\bar{\Lambda}_t^k$ and Λ_t^k are bounded by NTf_k^{max}, so Lemma 6 gives

$$\left| B_{\bar{\Lambda}_t^k}^k - B_{\Lambda_t^k}^k \right| \leq M_k \sqrt{|\bar{\Lambda}_t^k - \Lambda_t^k|\,(1 + \log{(Nf_k^{max}T/|\bar{\Lambda}_t^k - \Lambda_t^k|)})},$$

where M_k is some random variable defined in the lemma. For all $a > 0$, the function $x \mapsto \sqrt{x(1 + \log{(a/x)})}$ is increasing for $0 < x \leq a$, and Lipschitz continuity of f_k gives

$$|\bar{\Lambda}_t^k - \Lambda_t^k| = N \left| \int_0^t f_k(\bar{X}_s^{k,1}) - f_k(Y_s^{k,1})ds \right| \leq C \int_0^t G^N(s)ds \leq CTG^N(T),$$

so that

$$\left| B_{\bar{\Lambda}_t^k}^k - B_{\Lambda_t^k}^k \right| \leq M_k \sqrt{CTG^N(T)(1 + \log{(Nf_k^{max}/CG^N(T))})}.$$

On the event where $G^N(T) < 1$, (12) holds. If $G^N(T) \geq 1$, then the equation above implies

$$\left| B_{\bar{\Lambda}_t^k}^k - B_{\Lambda_t^k}^k \right| \leq M_k \sqrt{CTG^N(T)(1 + \log{(Nf_k^{max}/C)})},$$

and so coming back to (34) one has

$$G^N(T) \leq C_1 \left(M\sqrt{CT(1 + \log{(Nf^{max}/C)})}\sqrt{G^N(T)} + N \max_k \sup_{t \leq T} |R_t^k| \right) e^{C_2 T},$$

with $f^{max} = \max\{f_1^{max}, f_2^{max}\}$ and $M = \max\{M_1, M_2\}$. The inequality above is of order 2 with respect to $x = \sqrt{G^N(T)}$. Yet the positive values of x such that $p(x) = x^2 + bx + c$ is negative are such that $x^2 \leq b^2 - c$. Hence,

$$G^N(T) \leq C \left(M^2 T(1 + \log{(Nf^{max}/C)}) + N \max_k \sup_{t \leq T} |R_t^k| \right) e^{2C_2 T}.$$

Finally, (12) follows from the control of $|R_t^k|$ given by (33). □

A.2 Proof of Theorem 4

Recall that the components of the process X^k are given by

$$X_t^{k,j} = \left(e^{At}X_0 \right)^{k,j} + \int_0^t c_k f_{k+1}(X_s^{k+1,1}) \frac{e^{-\nu_k(t-s)}}{(\eta_k + 1 - j)!}(t-s)^{\eta_k + 1 - j}ds$$

$$+ \frac{1}{\sqrt{N}} \int_0^t \frac{c_k}{\sqrt{p_{k+1}}} \sqrt{f_{k+1}(X_s^{k+1,1})} \frac{e^{-\nu_k(t-s)}}{(\eta_k + 1 - j)!}(t-s)^{\eta_k + 1 - j}dW_s^{k+1}.$$

Squaring the above expression yields

$$\left(X_t^{k,j} \right)^2 = T_1(t) + T_2(t) + T_3(t) + T_4(t) + T_5(t) + T_6(t),$$

where

$$T_1 = \left((e^{At} x_0)^{k,j} \right)^2,$$

$$T_2 = \left(\int_0^t c_k f_{k+1}(X_s^{k+1,1}) \frac{e^{-\nu_k(t-s)}}{(\eta_k + 1 - j)!} (t-s)^{\eta_k+1-j} ds \right)^2,$$

$$T_3 = \left(\frac{1}{\sqrt{N}} \int_0^t \frac{c_k}{\sqrt{p_{k+1}}} \sqrt{f_{k+1}(X_s^{k+1,1})} \frac{e^{-\nu_k(t-s)}}{(\eta_k + 1 - j)!} (t-s)^{\eta_k+1-j} dW_s^{k+1} \right)^2,$$

$$T_4 = 2(e^{At} x_0)^{k,j} \int_0^t c_k f_{k+1}(X_s^{k+1,1}) \frac{e^{-\nu_k(t-s)}}{(\eta_k + 1 - j)!} (t-s)^{\eta_k+1-j} ds,$$

$$T_5 = 2 \int_0^t c_k f_{k+1}(X_s^{k+1,1}) \frac{e^{-\nu_k(t-s)}}{(\eta_k + 1 - j)!} (t-s)^{\eta_k+1-j} ds$$
$$\cdot \frac{1}{\sqrt{N}} \int_0^t \frac{c_k}{\sqrt{p_{k+1}}} \sqrt{f_{k+1}(X_s^{k+1,1})} \frac{e^{-\nu_k(t-s)}}{(\eta_k + 1 - j)!} (t-s)^{\eta_k+1-j} dW_s^{k+1},$$

$$T_6 = 2(e^{At} x_0)^{k,j} \frac{1}{\sqrt{N}} \int_0^t \frac{c_k}{\sqrt{p_{k+1}}} \sqrt{f_{k+1}(X_s^{k+1,1})} \frac{e^{-\nu_k(t-s)}}{(\eta_k + 1 - j)!} (t-s)^{\eta_k+1-j} dW_s^{k+1}.$$

First, we note that $\mathbb{E}[T_6(t)] = 0$ and that $\mathbb{E}[T_1(t)] = T_1(t)$. Since the intensity function is bounded by $0 < f_{k+1} \leq f_{k+1}^{\max}$, we have that

$$\mathbb{E}[T_4(t)] \leq \max \left\{ 0, \frac{c_k f_{k+1}^{\max}}{(\eta_k + 1 - j)!} \right\} 2(e^{At} x_0)^{k,j} \int_0^t e^{-\nu_k(t-s)} (t-s)^{\eta_k+1-j} ds.$$

Furthermore, applying Itô's isometry gives

$$\mathbb{E}[T_3(t)] = \frac{c_k^2}{N p_{k+1} ((\eta_k + 1 - j)!)^2} \int_0^t \mathbb{E}[f_{k+1}(X_s^{k+1,1})] e^{-2\nu_k(t-s)} (t-s)^{2(\eta_k+1-j)} ds.$$

Using again the fact that $f_{k+1} < f_{k+1}^{\max}$ results in

$$\mathbb{E}[T_3(t)] \leq \frac{1}{N} \frac{c_k^2}{p_{k+1}} \frac{f_{k+1}^{\max}}{((\eta_k + 1 - j)!)^2} \int_0^t e^{-2\nu_k(t-s)} (t-s)^{2(\eta_k+1-j)} ds.$$

Moreover, since f_{k+1} is bounded, we have that $(T_2(t))^2$ is also bounded, and thus it follows from the proof of Theorem 3 that

$$\mathbb{E}[T_2(t)] \leq \left(\frac{c_k f_{k+1}^{\max}}{(\eta_k + 1 - j)!} \right)^2 \left(\int_0^t e^{-\nu_k(t-s)} (t-s)^{(\eta_k+1-j)} ds \right)^2.$$

Applying the Cauchy–Schwarz inequality gives

$$\mathbb{E}[T_6(t)] \leq 2 \left(\mathbb{E}[T_2(t)] \mathbb{E}[T_3(t)] \right)^{1/2}.$$

Combining the above results and using that

$$\int_0^t e^{-2\nu_k(t-s)} (t-s)^{2(\eta_k+1-j)} ds = \frac{(2(\eta_k + 1 - j))!}{(2\nu_k)^{2(\eta_k+1-j)+1}} \left[1 - e^{-2\nu_k t} \sum_{l=0}^{2(\eta_k+1-j)} \frac{(2\nu_k t)^l}{l!} \right]$$

proves the statement.

A.3. Proof of Lemma 2

In order to rely on a linear control problem, we decouple the two populations and treat the nonlinear interactions in a second step, as is done in [27] for the continuous-time framework. Let us rewrite the numerical scheme (20) as given by the one-step mapping ψ_Δ defined by

$$\psi_\Delta[\xi](x) = e^{A\Delta}\left(x + \Delta B(x) + \frac{\sqrt{\Delta}}{\sqrt{N}}\sigma(x)\xi\right) = \begin{pmatrix}\psi_\Delta[\xi](x)_1 \\ \psi_\Delta[\xi](x)_2\end{pmatrix},$$

where

$$\psi_\Delta[\xi](x)_k = e^{A_{v_k}\Delta}x^k + \left(\Delta c_k f_{k+1}(x^{k+1,1}) + \frac{\sqrt{\Delta}}{\sqrt{N}}\frac{c_k}{\sqrt{p_{k+1}}}\sqrt{f_{k+1}(x^{k+1,1})}\xi^{k+1}\right)b_k,$$

with

$$b_k = e^{A_{v_k}\Delta}(0,\ldots,0,1)^T = \left(\frac{\Delta^{\eta_k}}{\eta_k!}, \frac{\Delta^{\eta_k-1}}{(\eta_k-1)!}, \ldots, 1\right)^T \in \mathbb{R}^{\eta_k+1}.$$

Now let us study the following discrete dynamical systems: $x^k(0) = x^k$, and for all $t \in \mathbb{N}$,

$$x^k(t+1) = e^{A_{v_k}\Delta}x^k(t) + b_k u^k(t+1), \tag{35}$$

where $(u^k(t))_{t\in\mathbb{N}^*}$ is a sequence of real numbers that will be specified below. This system is controllable as soon as b_k, $e^{A_{v_k}\Delta}b_k$, ..., $e^{\eta_k A_{v_k}\Delta}b_k$ are linearly independent (see Theorem 6.D1 in [9]). For all $j = 0, \ldots, \eta_k$, we have

$$e^{jA_{v_k}\Delta}b_k = \left(\frac{((j+1)\Delta)^{\eta_k}}{\eta_k!}, \frac{((j+1)\Delta)^{\eta_k-1}}{(\eta_k-1)!}, \ldots, 1\right)^T.$$

Yet $\{1, X, \ldots, X^{\eta_k}/\eta_k!\}$ is a basis of the vector space of polynomials with degree at most η_k, which ensures linear independence. The controllability of the system means that for all $x^k, y^k \in \mathbb{R}^{\eta_k+1}$, there exists some sequence of real numbers $(u^k(t))_{t=1,\ldots,\eta^*+1}$ such that $x^k(\eta^*+1) = y^k$, where x^k is inductively defined by (35). In the following, we use the notation $x(t) = (x^1(t), x^2(t))^T$. Now, let x and y be as in the statement of Lemma 2, and let $x = (x^1, x^2)^T$ and $y = (y^1, y^2)^T$. According to the first step of the proof, let $(u^k(t))_{t=1,\ldots,\eta^*+1}$ be such that $x^k(\eta^*+1) = y^k$ and define, for all $t = 1, \ldots, \eta^*+1$,

$$\xi^k(t) = \frac{u^k(t) - \Delta c_{k+1}f_k(x^{k,1}(t))}{\frac{\sqrt{\Delta}}{\sqrt{N}}\frac{c_k}{\sqrt{p_{k+1}}}\sqrt{f_k(x^{k,1}(t))}},$$

in such a way that

$$u^k(t) = \Delta c_{k+1}f_k(x^{k,1}(t)) + \frac{\sqrt{\Delta}}{\sqrt{N}}\frac{c_k}{\sqrt{p_{k+1}}}\sqrt{f_k(x^{k,1}(t))}\xi^k(t).$$

Substituting $u^k(t+1)$ in (29) and writing $\xi_t = \xi(t)$ yields $x^k(t+1) = \psi_\Delta[\xi_{t+1}](x(t))_k$, and thus

$$y = x(\bar{\eta}+1) = \underbrace{\left(\psi_\Delta[\xi_{\bar{\eta}+1}] \circ \cdots \circ \psi_\Delta[\xi_1]\right)}_{\eta^*+1}(x),$$

which proves the result.

References

[1] ABI JABER, E., LARSSON, M. AND PULIDO, S. (2019). Affine Volterra processes. *Ann. Appl. Prob.* **29**, 3155–3200.

[2] ABLEIDINGER, M. AND BUCKWAR, E. (2016). Splitting integrators for the stochastic Landau–Lifshitz equation. *SIAM J. Sci. Comput.* **38**, A1788–A1806.

[3] ABLEIDINGER, M., BUCKWAR, E. AND HINTERLEITNER, H. (2017). A stochastic version of the Jansen and Rit neural mass model: analysis and numerics. *J. Math. Neurosci.* **7**, 1–35.

[4] BLANES, S., CASAS, F. AND MURUA, A. (2009). Splitting and composition methods in the numerical integration of differential equations. *Bol. Soc. Esp. Mat. Apl.* **45**, 89–145.

[5] BRÉHIER, C.-E. AND GOUDENÈGE, L. (2019). Analysis of some splitting schemes for the stochastic Allen–Cahn equation. *Discrete Continuous Dynam. Systems B* **24**, 4169–4190.

[6] BRÉMAUD, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer, Berlin, Heidelberg.

[7] BUCKWAR, E., TAMBORRINO, M. AND TUBIKANEC, I. (2020). Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs. *Statist. Comput.* **30**, 627–648.

[8] CATANI, M., JONES, D. K., DONATO, R. AND FFYTCHE, D. H. (2003). Occipito-temporal connections in the human brain. *Brain* **126**, 2093–2107.

[9] CHEN, C.-T. (1998). *Linear System Theory and Design*. Oxford University Press.

[10] CHEVALLIER, J., CÁCERES, M. J., DOUMIC, M. AND REYNAUD-BOURET, P. (2015). Microscopic approach of a time elapsed neural model. *Math. Models Meth. Appl. Sci.* **25**, 2669–2719.

[11] CHORNOBOY, E. S., SCHRAMM, L. P. AND KARR, A. F. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybernet.* **59**, 265–275.

[12] DASSIOS, A. AND ZHAO, H. (2013). Exact simulation of Hawkes process with exponentially decaying intensity. *Electron. Commun. Prob.* **18**, 1–13.

[13] DELARUE, F. AND MENOZZI, S. (2010). Density estimates for a random noise propagating through a chain of differential equations. *J. Funct. Anal* **259**, 1577–1630.

[14] DELATTRE, S., FOURNIER, N. AND HOFFMANN, M. (2016). Hawkes processes on large networks. *Ann. Appl. Prob.* **26**, 216–261.

[15] DITLEVSEN, S. AND LÖCHERBACH, E. (2017). Multi-class oscillating systems of interacting neurons. *Stoch. Process. Appl.* 127, 1840–1869.

[16] DUARTE, A., LÖCHERBACH, E. AND OST, G. (2019). Stability, convergence to equilibrium and simulation of non-linear Hawkes processes with memory kernels given by the sum of Erlang kernels. *ESAIM Prob. Statist.* **23**, 770–796.

[17] ETHIER, S. N. AND KURTZ, T. G. (2009). *Markov Processes: Characterization and Convergence*. John Wiley, New York.

[18] FISCHER, M. AND NAPPO, G. (2009). On the moments of the modulus of continuity of Itô processes. *Stoch. Anal. Appl.* **28**, 103–122.

[19] HAIRER, E., LUBICH, C. AND WANNER, G. (2006). *Geometric Numerical Integration*. Springer, Heidelberg.

[20] HIGHAM, D. J. AND STRØMMEN MELBØ, A. H. (2004). Numerical simulation of a linear stochastic oscillator with additive noise. *Appl. Numer. Math.* **51**, 89–99.

[21] JOHNSON, D. H. (1996). Point process models of single-neuron discharges. *J. Comput. Neurosci.* **3**, 275–299.

[22] KLOEDEN, P. E., PLATEN, E. AND SCHURZ, H. (2003). *Numerical Solution of SDE Through Computer Experiments*. Springer, Berlin, Heidelberg.

[23] KURTZ, T. G. (1978). Strong approximation theorems for density dependent Markov chains. *Stoch. Process. Appl.* **6**, 223–240.

[24] LEIMKUHLER, B. AND MATTHEWS, C. (2015). *Molecular Dynamics: with Deterministic and Stochastic Numerical Methods*. Springer, Cham.

[25] LEIMKUHLER, B., MATTHEWS, C. AND STOLTZ, G. (2016). The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA J. Numer. Anal.* **36**, 16–79.

[26] LEWIS, P. A. W. AND SHEDLER, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Res. Logistics Quart.* **26**, 403–413.

[27] LÖCHERBACH, E. (2019). Large deviations for cascades of diffusions arising in oscillating systems of interacting Hawkes processes. *J. Theoret. Prob.* **32**, 131–162.

[28] MALLIAVIN, P. AND THALMAIER, A. (2006). *Stochastic Calculus of Variations in Mathematical Finance*. Springer, Berlin, Heidelberg.

[29] MAO, X. (2007). *Stochastic Differential Equations and Applications*. Woodhead Publishing, Oxford.

[30] MATTINGLY, J. C., STUART, A. M. AND HIGHAM, D. J. (2002). Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Process. Appl.* **101**, 185–232.

[31] MCLACHLAN, R. AND QUISPEL, G. (2002). Splitting methods. *Acta Numer.* **11**, 341–434.

[32] MILSTEIN, G. N. AND TRETYAKOV, M. V. (2003). Quasi-symplectic methods for Langevin-type equations. *IMA J. Numer. Anal.* **23**, 593–626.

[33] MILSTEIN, G. N. AND TRETYAKOV, M. V. (2004). *Stochastic Numerics for Mathematical Physics*. Springer, Berlin.

[34] MISAWA, T. (2001). A Lie algebraic approach to numerical integration of stochastic differential equations. *SIAM J. Sci. Comput.* **23**, 866–890.

[35] OGATA, Y. (1981). On Lewis simulation method for point processes. *IEEE Trans. Inf. Theory* **27**, 23–30.

[36] PERNICE, V., STAUDE, B., CARDANOBILE, S. AND ROTTER, S. (2011). How structure determines correlations in neuronal networks. *PLOS Comput. Biol.* **7**, 1–14.

[37] PETERSEN, W. P. (1998). A general implicit splitting for stabilizing numerical simulations of Itô stochastic differential equations. *SIAM J. Numer. Anal.* **35**, 1439–1451.

[38] RENNIE, B. C. AND DOBSON, A. J. (1969). On Stirling numbers of the second kind. *J. Combinatorial Theory* **7**, 116–121.

[39] REYNAUD-BOURET, P., RIVOIRARD, V., GRAMMONT, F. AND TULEAU-MALOT, C. (2014). Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *J. Math. Neurosci.* **4**, 3.

[40] SHARDLOW, T. (2003). Splitting for dissipative particle dynamics. *SIAM J. Sci. Comput.* **24**, 1267–1282.

[41] STRANG, G. (1968). On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.* **5**, 506–517.

[42] STROOCK, D. W. AND VARADHAN, S. R. S. (2007). *Multidimensional Diffusion Processes*. Springer, Berlin, Heidelberg.

[43] WOOD, D. C. (1992). The computation of polylogarithms. Tech. Rep. 15-92∗, University of Kent, Computing Laboratory.