Diffusion approximation of multi-class Hawkes processes: theoretical and numerical analysis

Anna Melnykova

joint work with J. Chevallier (Université Grenoble Alpes) and I. Tubikanec (JKU Linz, Austria)

Institut Nationale Polytechnique de Grenoble, LJK UMR-CNRS 5224

Séminaire d'équipe ProbaStat du Laboratoire Paul Painlevé,
Lille
9 March 2021

Figure: **Julien Chevallier** (MCF at Université Grenoble Alpes) and **Irene Tubikanec** (PhD student in JKU Linz, Austria)

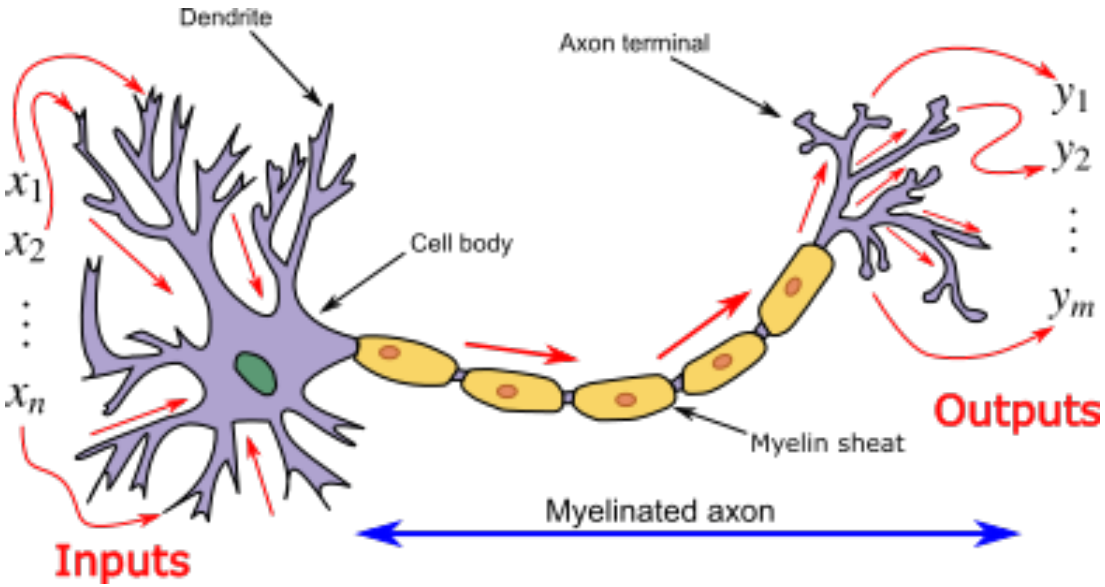
Source:

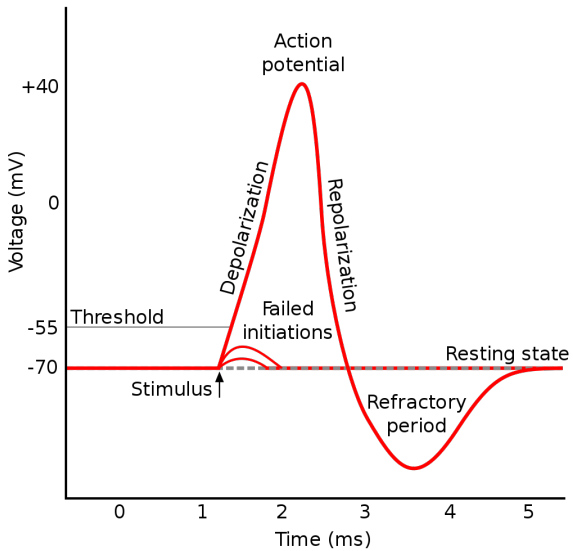
J. Chevallier, A. Melnykova, I. Tubikanec *"Diffusion approximation of multi-class Hawkes Processes: theoretical and numerical analysis"*, to appear in *Advances in Applied Probability*, https://doi.org/10.14760/OWP-2020-09

Plan of the presentation

1. Motivation: biological neurons
2. Hawkes process with Erlang kernel and its diffusion approximation
3. Simulation of the diffusion
4. Simulation of the Piece-Wise Deterministic Markov process
5. Conclusion

Motivation: biological neurons





How to model activity of the biological neurons?

One can consider the membrane potential of **a single cell** as a continuous stochastic process (*diffusion-type models*).

- **Examples:** stochastic Hodgkin-Huxley, Morris-Lecar, FitzHugh-Nagumo models.
- **Advantages:** very precise, can take into account a lot of factors
- **Disadvantage:** complicated to analyze when the number of neurons the large. (Extremely) Computationally expensive to model the networks.

How to model activity of the biological neurons?

If we are interested only in spiking activity of a **network of neurons**, we need the model which would:

- keep track of spiking events in a binary form (*counting process*)
- be able to model the refractory period (i.e., *to have a memory*)
- be able to model the interactions between the populations of neurons (neuronal groups with different functions), i.e., to be *a multi-class process*.

How to model activity of the biological neurons?

If we are interested only in spiking activity of a **network of neurons**, we need the model which would:

- keep track of spiking events in a binary form (*counting process*)
- be able to model the refractory period (i.e., *to have a memory*)
- be able to model the interactions between the populations of neurons (neuronal groups with different functions), i.e., to be *a multi-class process*.

A natural suitable candidate is **Hawkes process**, which is a point process with conditional intensity given by

$$\lambda(t) = \mu(t) + \int_0^t h(t-s)ds.$$

Population of neurons: Hawkes process

- **N neurons** structured in **K populations** (N_k neurons in each, $k = 1, \dots, K$)
- **Spike trains** $\{(Z_t^{k,n})_{t\geq 0},\ 1 \leq k \leq K,\ 1 \leq n \leq N_k\}$ are characterized by the **intensity processes** $(\lambda^{k,n}(t))_{t\geq 0}$,

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

- **Mean-field framework:** intensity processes $\lambda^{k,n}(t)$ are 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.

References

Delarue and Menozzi (2010), Ditlevsen and Löcherbach (2017)

Population of neurons: Hawkes process

- **N neurons** structured in **K populations** (N_k neurons in each, $k = 1, \dots, K$)
- **Spike trains** $\{(Z_t^{k,n})_{t\geq 0},\ 1 \leq k \leq K,\ 1 \leq n \leq N_k\}$ are characterized by the **intensity processes** $(\lambda^{k,n}(t))_{t\geq 0}$,

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

- **Mean-field framework:** intensity processes $\lambda^{k,n}(t)$ are 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.

References

Delarue and Menozzi (2010), Ditlevsen and Löcherbach (2017)

Population of neurons: Hawkes process

- **N neurons** structured in **K populations** (N_k neurons in each, $k = 1, \dots, K$)
- **Spike trains** $\{(Z_t^{k,n})_{t\geq 0},\ 1 \leq k \leq K,\ 1 \leq n \leq N_k\}$ are characterized by the **intensity processes** $(\lambda^{k,n}(t))_{t\geq 0}$,

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

- **Mean-field framework:** intensity processes $\lambda^{k,n}(t)$ are 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.

References

Delarue and Menozzi (2010), Ditlevsen and Löcherbach (2017)

Hawkes process with Erlang kernels

Consider the following process (Ditlevsen and Löcherbach, 2017):

$$\bar{X}_t^{k,1} = \frac{1}{N_{k+1}} \sum_{1 \le m \le N_{k+1}} \int_{(0,t)} h_{kk+1}(t-s) dZ_s^{k+1,m}, \tag{2}$$

where

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

- $c_k = \pm 1$ is a "sign" of the population (-1 for inhibitory and +1 for excitatory)
- η_k — an integer parameter, determining the memory order for the interaction function
- $\frac{\eta_{k+1}}{\nu_{k+1}}$ — the maximum delay of the influence. Larger ratio assigns more importance to the "old" events. If η_{k+1} and ν_{k+1} tend to ∞, while the ratio is constant, then only one specific moment of time is important.

Piece-wise deterministic Markov process

Along with the **spiking rate functions** f_k the process $\bar{X}$ identifies PDMP $(Z_t^{k,n})_{t\geq 0}$ as follows:

$$\mathbb{P}(Z_t^{k,n} \text{ has a jump in } (t, t+dt]|\mathcal{F}_t) = f_k\left(\bar{X}_t^{k,1}\right) dt$$

$(\bar{X}_t)_{t\geq 0} = \{(\bar{X}_t^{k,j})_{t\geq 0},\ 1 \leq k \leq K, 1 \leq j \leq \eta_k + 1\}$ is a **Piece-Wise Deterministic Markov process**, which solves the following system of dimension $\kappa = \sum_{k=1}^{K}(\eta_k + 1)$:

$$\begin{cases} d\bar{X}_t^{k,j} = \left[-\nu_k \bar{X}_t^{k,j} + \bar{X}_t^{k,j+1}\right] dt, \text{ for } j = 1, \dots, \eta_k, \\ d\bar{X}_t^{k,\eta_k+1} = -\nu_k \bar{X}_t^{k,\eta_k+1} dt + c_k d\bar{Z}_t^{k+1}, \end{cases} \tag{3}$$

where $\bar{Z}_t^{k+1} = \frac{1}{N_{k+1}} \sum_{n=1}^{N_{k+1}} Z_t^{k+1,n}$, each $Z_t^{k+1,n}$ jumping at rate $f_{k+1}(\bar{X}_{t-}^{k+1,1})$, $\bar{X}_0 = x_0 \in \mathbb{R}^\kappa$.

Piece-wise deterministic Markov process

When the number of neurons is large and $\frac{N_k}{N} = p_k \quad \forall k$, the following approximation holds (Ditlevsen and Löcherbach, 2017):

$$d\bar{Z}_t^{k+1} \approx \frac{f_{k+1}\left(\bar{X}_{t-}^{k+1,1}\right)}{p_{k+1}}dt + \sqrt{\frac{f_{k+1}\left(\bar{X}_{t-}^{k+1,1}\right)}{p_{k+1}}}dW_t$$

In other words, we can approximate the PDMP (3) by a diffusion, which solves the following **Stochastic Differenetial Equation** of the same dimension as PDMP:

$$\begin{cases} d\bar{X}_t^{k,j} = \left[-\nu_k \bar{X}_t^{k,j} + \bar{X}_t^{k,j+1}\right] dt, \text{ for } j = 1, \dots, \eta_k, \\ d\bar{X}_t^{k,\eta_k+1} = \left(-\nu_k \bar{X}_t^{k,\eta_k+1} + c_k \frac{f_{k+1}\left(\bar{X}_{t-}^{k+1,1}\right)}{p_{k+1}}\right) dt + c_k \sqrt{\frac{f_{k+1}\left(\bar{X}_{t-}^{k+1,1}\right)}{p_{k+1}}} dW_t^k. \end{cases} \tag{4}$$

Cascade SDE for 2 populations ($K = 2$)

$$dX_t = (AX_t + B(X_t))dt + \frac{1}{\sqrt{N}}\sigma(X_t)dW_t, \quad X_0 = x_0, \tag{5}$$

- $X = (X^1, X^2)^T$ is composed of two $\eta_k + 1$ dimensional vectors, $k = 1, 2$.
- $W = (W^1, W^2)^T$ is a 2-dimensional Brownian motion
- $A = \begin{pmatrix} A_{\nu_1} & \mathbb{O}_{(\eta_1+1)\times(\eta_2+1)} \\ \mathbb{O}_{(\eta_2+1)\times(\eta_1+1)} & A_{\nu_2} \end{pmatrix}$, where A_{ν_k} is a tri-diagonal matrix with lower-diagonal equal to 0_{η_k}, diagonal equal to $(-\nu_k, \dots, -\nu_k)$ and upper-diagonal equal to $(1, \dots, 1)$,
- $B(X) = (B^1(X^2), B^2(X^1))^T$, where $B^1(X^2) = (0, \dots, 0, c_1 f_2(X^{2,1}))$ and $B^2(X^1) = (0, \dots, 0, c_2 f_1(X^{1,1}))$.
- $\sigma(X) = \left(\sigma^1(X^2), \sigma^2(X^1)\right)^T$, where σ^1 and σ^2 read as

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

Population of neurons: Hawkes process

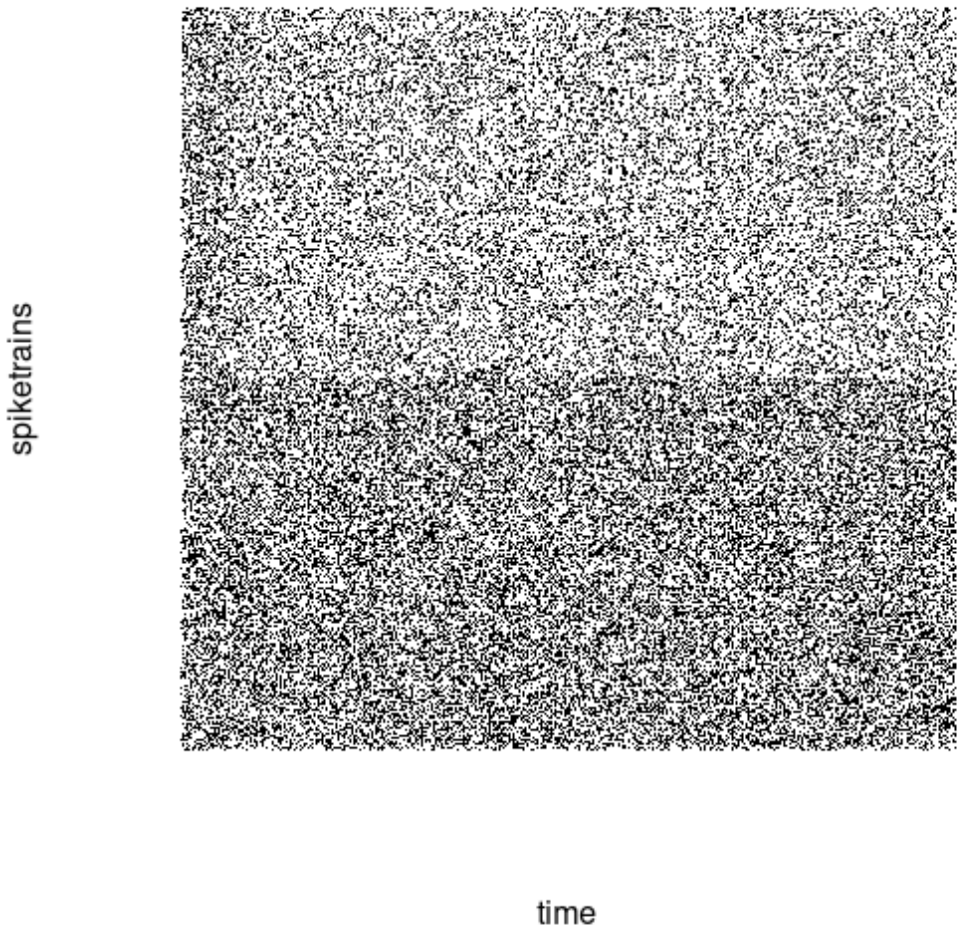


Figure: Simulation of the $Z^{k,n}$ (spiking times of a population of neurons)

Population of neurons: SDE

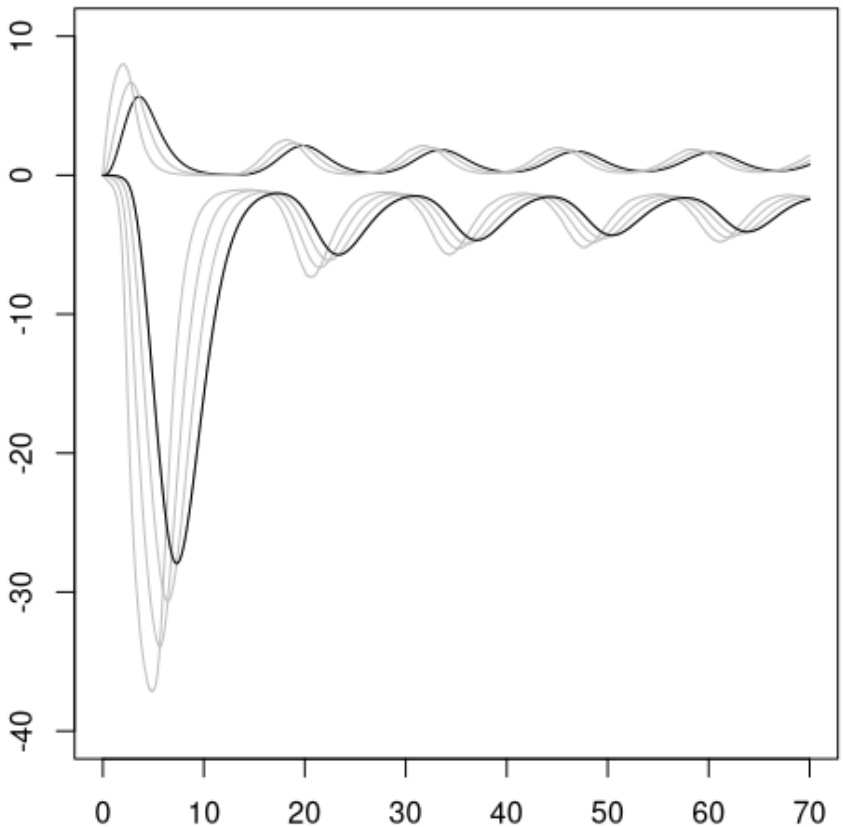


Figure: Simulation of X (the integrated intensity for 2 populations, approximated with a SDE)

Weak and strong error bounds

Theorem (Ditlevsen and Löcherbach (2017))

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 $t \geq 0$,

$$\sup_{x \in \mathbb{R}^k} |\mathbb{E}_x \varphi(\bar{X}_t) - \mathbb{E}_x \varphi(X_t)| \leq Ct \frac{\|\varphi\|_{4,\infty}}{N^2}, \tag{6}$$

where $\mathbb{E}_x$ denotes the conditional expectation given that $\bar{X}_0 = X_0 = x$.

Theorem (Chevallier, Melnykova, Tubikanec, 2020)

There exists a constant $C > 0$ such that, for all $T > 0$,

$$\sup_{t \leq T} \|\bar{X}_t - X_t\|_\infty \leq \Theta_N e^{CT} \frac{\log(N)}{N} \text{ a.s.},$$

where Θ_N is a r.v. with exponential moments whose distribution does not depend on N.

Idea of the proof for the strong error bound

- ”Decouple” the populations, consider them as two independent processes
- Rescale the time for the Hawkes process, present it as a transformation a standard Poisson process
- Show that a standard Brownian motion and a standard Poisson process can be constructed on the same probability space
- Evaluate the difference between the processes (Ethier and Kurtz, 2009):

$$\sup_{t\geq 0} \frac{|\Pi_t - t - W_t|}{\log(2 \vee t)} \leq \Xi < \infty \text{ a.s.},$$

where Ξ is a r.v. with exponential moments.

Strong error bound: (not an) illustration

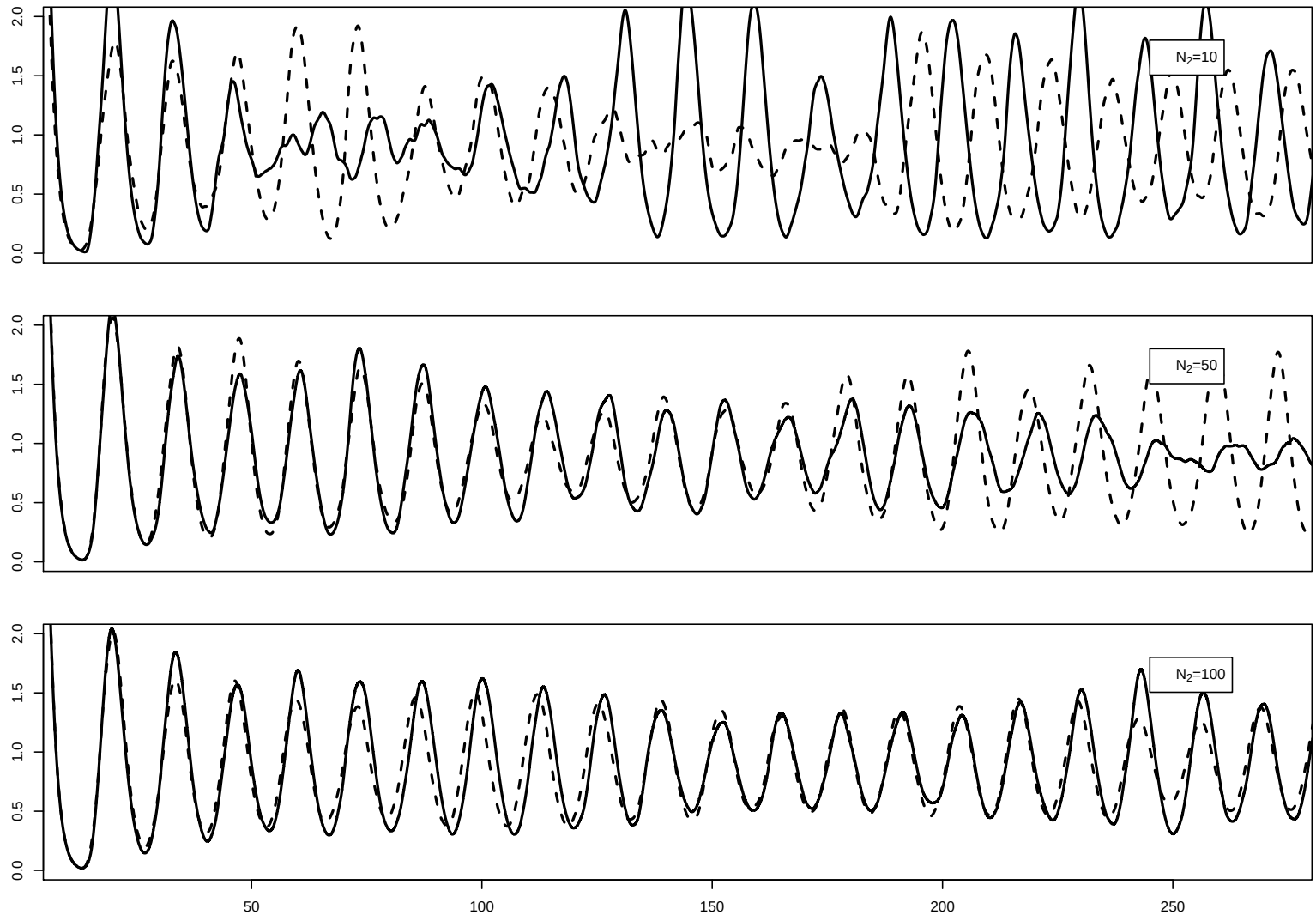


Figure: Solid line: PDMP, dashed line: diffusion

Few remarks on weak and strong convergence

- The obtained results can be used to prove that both PDMP and stochastic diffusion converge to PDE as $N \to \infty$
- However, diffusion approximation becomes worse as time increases (at **exponential rate** for strong error).
- Unfortunately, the strong error bound cannot be illustrated numerically: the coupling between a specific Poisson process with a specific Brownian motion is not explicit.

Why it is good to work with the stochastic diffusion?

It is easier to study the behaviour of the system:

- Its stationary regime, dependence on parameters, bifurcation points etc. (Ditlevsen and Löcherbach, 2017)
- Its long-time behaviour, large deviations (Löcherbach, 2019)
- Its moments, ergodicity (Chevallier, *Melnykova*, Tubikanec, 2020)
- **It is much faster to simulate!** (Chevallier, *Melnykova*, Tubikanec, 2020)

Main difficulties:

- System is high-dimensional, and the dimension depends on the parameter η.
- Non-linear drift.
- Degenerate diffusion coefficient (*hypoelliptic* system).

Classical solution

- Split the observation interval $[0, T]$ in small intervals of size Δ
- Taking x_{t_i} as an initial value, use Itô-Taylor expansion to approximate $x_{t_i+\Delta} =: x_{t_{i+1}}$
- If in Itô-Taylor expansion we take the terms up to order Δ, we obtain the classical *Euler-Maruyama scheme*.

Drawbacks

- Does not always converge for oscillating systems
- In case of convergence, does not necessarily preserves the ergodic property and the moments

Main difficulties:

- System is high-dimensional, and the dimension depends on the parameter η.
- Non-linear drift.
- Degenerate diffusion coefficient (*hypoelliptic* system).

Main idea of splitting:

Instead of approximating the solution of the ”complicated” system, the SDE is splitted in ”simpler” subsystems, which can be explicitly solved.

Splitting for SDEs

Mattingly et al. (2002), Shardlow (2003), Leimkuhler and Matthews (2015), Ableidinger and Buckwar (2016), Ableidinger et al. (2017), Buckwar et al. (2019)

Cascade SDE: hypoellipticity

Note that the diffusion matrix has only two non-zero components. But in fact, the Brownian motion enters in every variable of the system!

We can write the solution of the system (4) as follows:

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

Cascade SDE: hypoellipticity

Note that the diffusion matrix has only two non-zero components. But in fact, the Brownian motion enters in every variable of the system!

We can write the solution of the system (4) as follows:

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

If we look closer at the highlighted part, we will note that it results in a non-zero vector. More precisely, the diffusion coefficient of $j-$th variable, describing $k-$th population, is given by:

$$\frac{c_k\sqrt{f_{k+1}(X^{k+1,1})}}{N\sqrt{p_{k+1}}}\frac{e^{-\nu_k t}\, t^{\eta_k+1-j}}{(\eta_k+1-j)!}dW_t, \quad j=1,\dots,\eta_k+1$$

This "propagation of chaos" is ensured by **hypoellipticity**.

Cascade diffusion: closer look on the structure

Let us note the following:

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

- Red part is linear
- Blue part contains a lot of zeroes (only 2 non-zero variables)

Cascade diffusion: splitting approach

We can split the equation in two subsystems:

$$dX_t^{[1]} = AX_t^{[1]}dt,$$
$$dX_t^{[2]} = B(X_t^{[2]})dt + \frac{1}{\sqrt{N}}\sigma(X_t^{[2]})dW_t.$$

The solutions (or "flows") of the respective subsystems are given as follows:

$$\psi_t^{[1]}(x) := e^{At}x,$$
$$\psi_t^{[2]}(x) := x + tB(x) + \frac{\sqrt{t}}{\sqrt{N}}\sigma(x)w,$$

where $w = (w^1, w^2)^T$ is a 2-dimensional standard normal vector.

Cascade diffusion: splitting approach

Lie-Trotter splitting (Mclachlan and Quispel, 2002) can be written as follows:

$$\tilde{X}_{t_{i+1}} = \left(\psi_{\Delta}^{[1]} \circ \psi_{\Delta}^{[2]}\right)(\tilde{X}_{t_i}) = e^{A\Delta}\left(\tilde{X}_{t_i} + \Delta B(\tilde{X}_{t_i}) + \frac{\sqrt{\Delta}}{\sqrt{N}}\sigma(\tilde{X}_{t_i})w_i\right), \tag{7}$$

where $(w_i)_{i=1,\dots,i_{\max}}$ i.i.d. $\mathcal{N}(0,1)$. Note that the approximated solution (7) is a discrete analogue of

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

Cascade diffusion: splitting approach

Lie-Trotter splitting (Mclachlan and Quispel, 2002) can be written as follows:

$$\tilde{X}_{t_{i+1}} = \left(\psi_{\Delta}^{[1]} \circ \psi_{\Delta}^{[2]}\right)(\tilde{X}_{t_i}) = e^{A\Delta}\left(\tilde{X}_{t_i} + \Delta B(\tilde{X}_{t_i}) + \frac{\sqrt{\Delta}}{\sqrt{N}}\sigma(\tilde{X}_{t_i})w_i\right), \tag{7}$$

where $(w_i)_{i=1,\dots,i_{\max}}$ i.i.d. $\mathcal{N}(0,1)$.

Theorem (Mean-square convergence of the splitting scheme)

$\tilde{X}_{t_i}$ 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, $i = 1, \dots, i_{max}$, where $\|\cdot\|$ denotes the Euclidean norm.

Some details of the proof

- We show that in our case Euler-Maruyama scheme coincides with the Milstein scheme, which is known to converge with mean-square order 1 (Kloeden et al., 2003)
- Then, we apply a triangle inequality to show that

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

- The result follows from

$$\|X_{t_i} - \tilde{X}_{t_i}\|_{L^2} = O(\Delta) + O(\Delta^{\frac{3}{2}})$$

Some remarks

- Splitting is *not unique*: other decompositions are possible.
- Lie-Trotter is not the only composition scheme. One could also use Strang splitting (Strang, 1968):

$$\tilde{X}_{t_{i+1}} = \left(\psi^{[1]}_{\Delta/2} \circ \psi^{[2]}_{\Delta} \circ \psi^{[1]}_{\Delta/2}\right)\left(\tilde{X}_{t_i}\right). \tag{8}$$

- In Lie-Trotter or Strang splitting one could use more than 2 steps. For ODEs it can improve the performance, but it does not always hold for SDEs.
- In this talk, **we focus on the Lie-Trotter splitting**: it is the most "intuitive" approach. In Chevallier et al. (2020) we also study the numerical performance for Strang splitting.

Mean-square order convergence

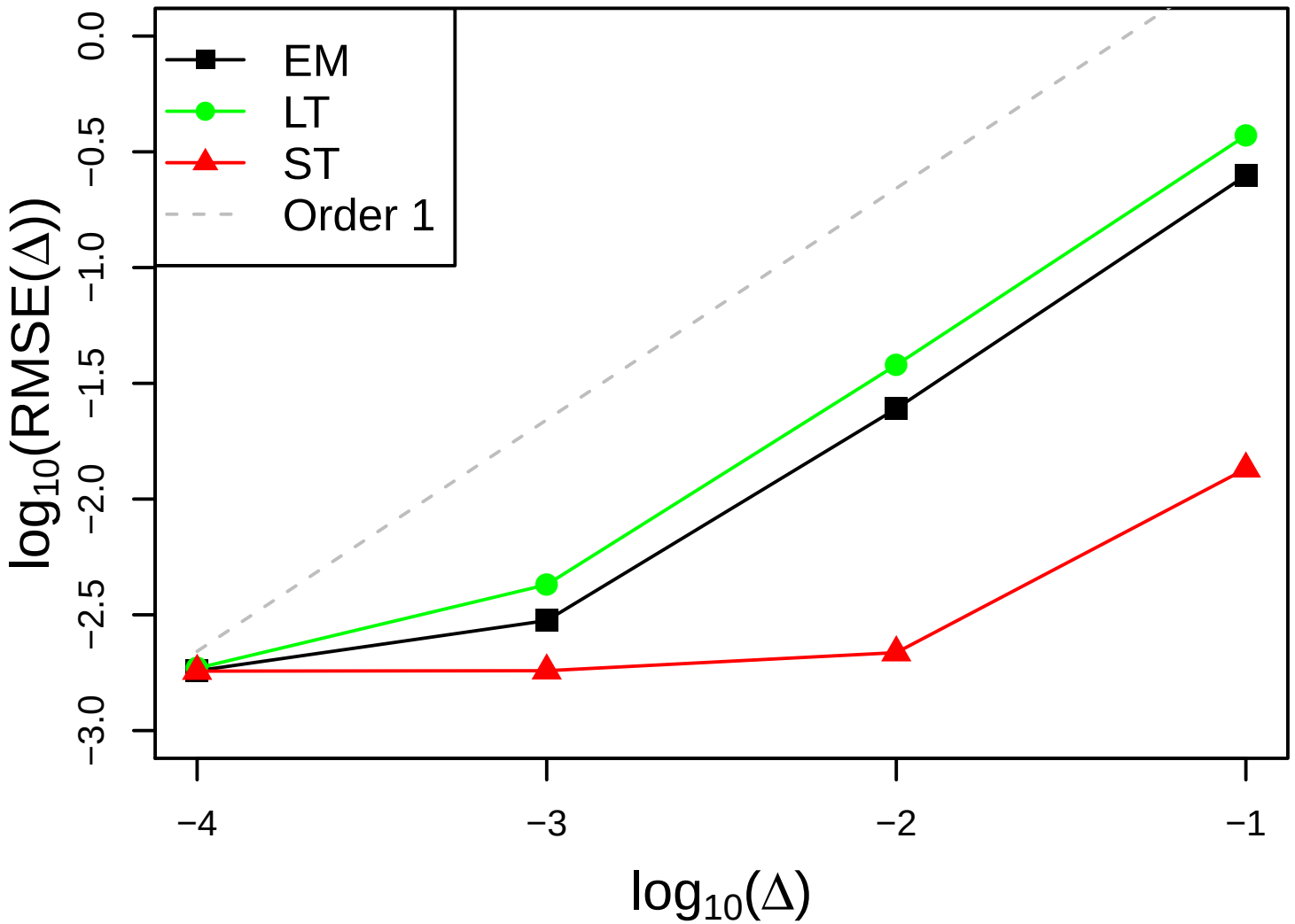


Figure: Mean-square order convergence: Lie-Trotter and Strang splitting compared to Euler-Maruyama

Splitting approach: theoretical properties

Our current goal is to show that the splitting scheme does not only converge, but also:

- preserves **first** and **second moment bounds**,
- has **ergodic** property.

Important remark

For systems with oscillating dynamics, the Euler-Maruyama scheme converges, but not necessarily preserves the properties of the underlying diffusion (Mattingly et al., 2002)!

Theorem (First moment bounds of the diffusion process)

The following bounds hold for the components of $\mathbb{E}[X_t]$*:*

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

where

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

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

Theorem (First moment bounds of the approximated process)

The following bounds hold for the components of $\mathbb{E}[\tilde{X}_{t_i}]$*:*

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

where

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

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

Corollary

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

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

where $\kappa^{k,j} := \eta_k + 1 - j$ *and* $Li_{-\kappa^{k,j}}\left(e^{-\nu_k \Delta}\right)$ *is a polylogarithm function.*
(ii) The following bounds hold for the components of $\mathbb{E}[\tilde{X}_{t_i}]$ *as* $i \to \infty$ *and* $\Delta \to 0$*:*

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

Remarks on the obtained results

- We assume that the *intensity functions f_k are strictly positive and bounded from below and above*. This condition is also necessary for the diffusion approximation to hold (Ditlevsen and Löcherbach, 2017).
- Practical meaning of the bounds: we can verify that the scheme has the same properties as the continuous process, even for a fixed Δ.
- In asymptotics $\Delta \to 0, n \to \infty$, bounds on the first and the second moments for continuous and approximated process coincide (Chevallier et al., 2020).

First and second moment bounds. Illustration.

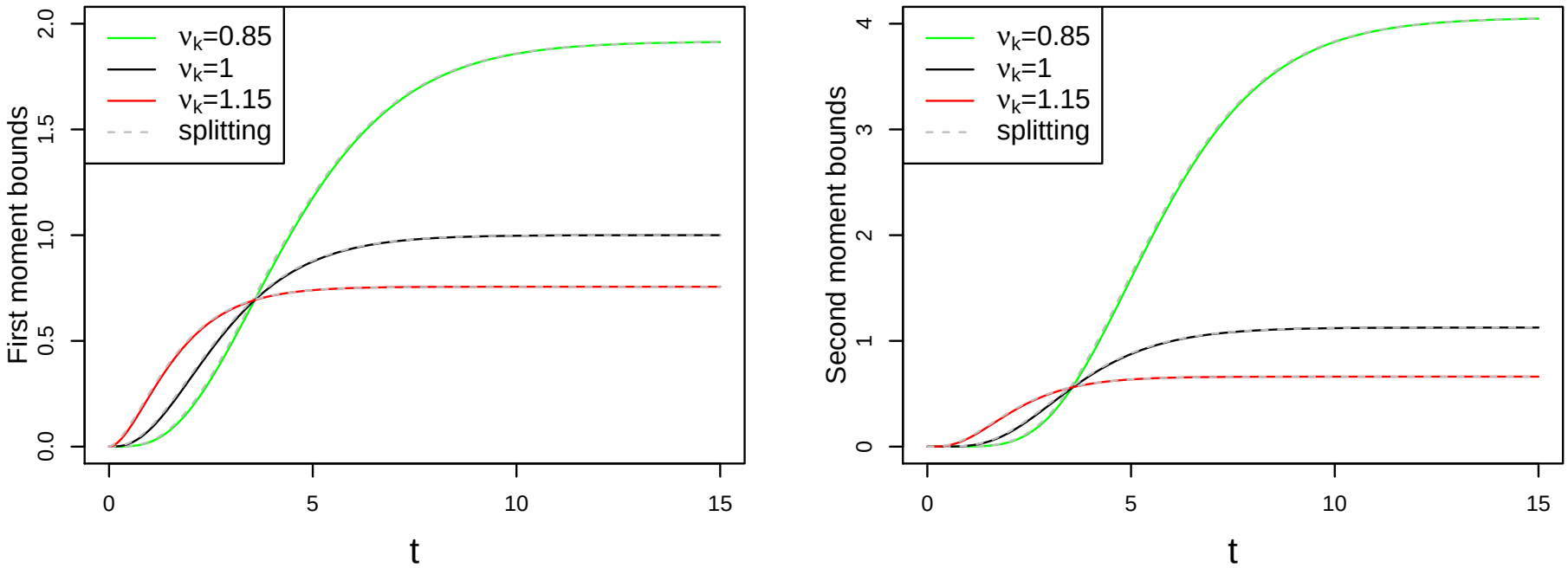


Figure: First (left panel) and second (right panel) moment bounds of the excitatory population $k = 2$ for different values of ν_2. The moment bounds for the diffusion are in solid lines and the moment bounds for the splitting scheme are in dashed lines. The bound of the rate function is fixed to $f_1^{\max} = 1$. The parameters are $\eta_2 = 3$, $N = 100$, $p_1 = 1/2$ and the time step $\Delta = 0.1$ is used.

Long-time behaviour: ergodicity

Theorem (Geometric ergodicity)

The process $(\tilde{X}_{t_i})_{i=0,\dots,i_{\max}}$ *has a unique invariant measure* π^{Δ} *on* $\mathbb{R}^{\kappa}$. *Denote by* G *a Lyapunov function of* $\tilde{X}$. *Then,* $\forall x_0, m \geq 1, \exists \tilde{C} = C(m,\Delta) > 0$ *and* $\tilde{\lambda} = \tilde{\lambda}(m,\Delta) > 0$ *such that, for all measurable functions* $g : \mathbb{R}^{\kappa} \to \mathbb{R}$ *such that* $|g| \leq \tilde{G}^m$,

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

Geometric ergodicity: proof

- **Lyapunov condition**: $\tilde{G}(x) = \sum_{k=1}^{2} \sum_{j=1}^{\eta_k+1} \frac{j}{\nu_k^{j-1}} \left|x^{k,j}\right|$, is a Lyapunov function for $\tilde{X}$, i.e., there exist constants $\alpha \in [0, 1)$ and $\beta \geq 0$, such that

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

- **Hypoellipticity**: granted by the convolution-type structure of the scheme
- **Irreducibility**: $\forall x, y \in \mathbb{R}^{\kappa}$ there exists some sequence of 2-dimensional vectors $(w_i)_{i=1,\dots,\eta^*+1}$ such that

$$y = \underbrace{(\psi_\Delta[w_{\eta^*+1}] \circ \cdots \circ \psi_\Delta[w_1])}_{\eta^*+1}(x), \tag{9}$$

where ψ_Δ denotes one step of the splitting scheme, $\eta^\star = \max(\eta_1, \eta_2)$.

Splitting approach. Numerical performance.

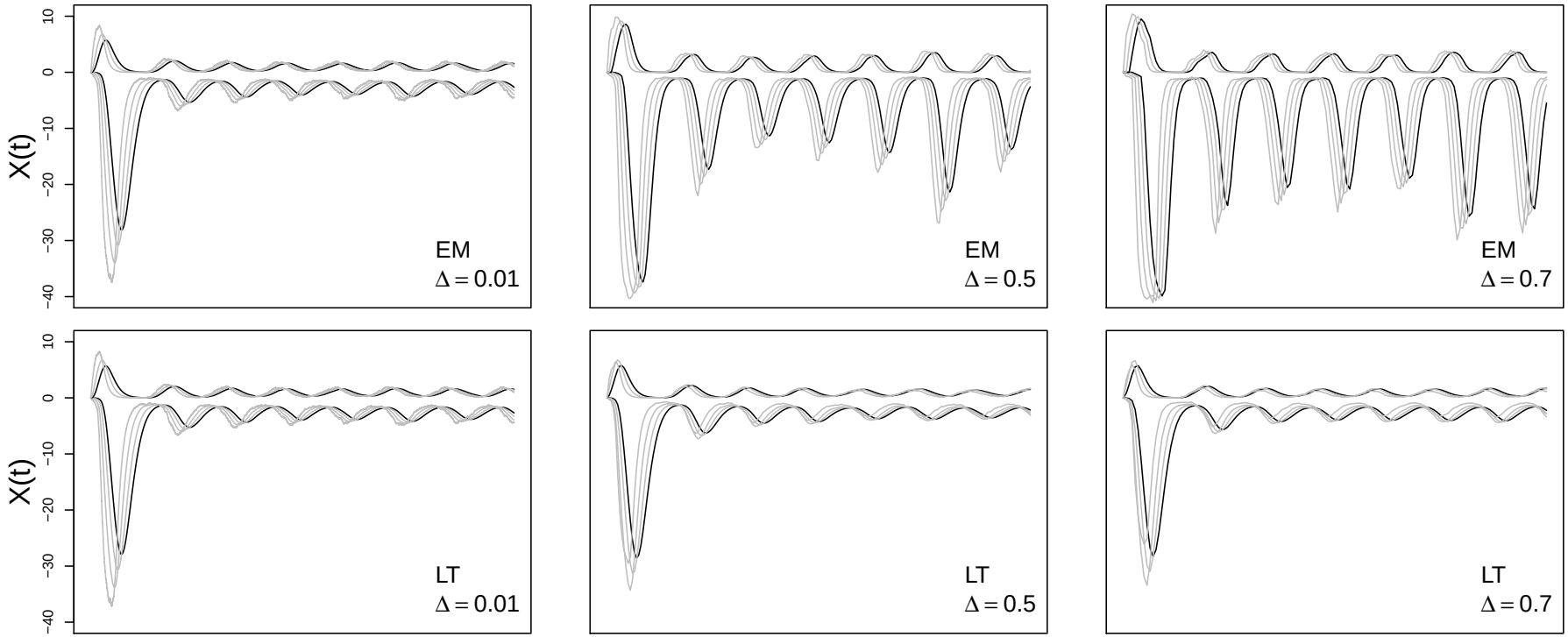


Figure: Trajectories of the diffusion, simulated with the Euler-Maruyama scheme and the Lie-Trotter splitting

Splitting approach. Empirical densities.

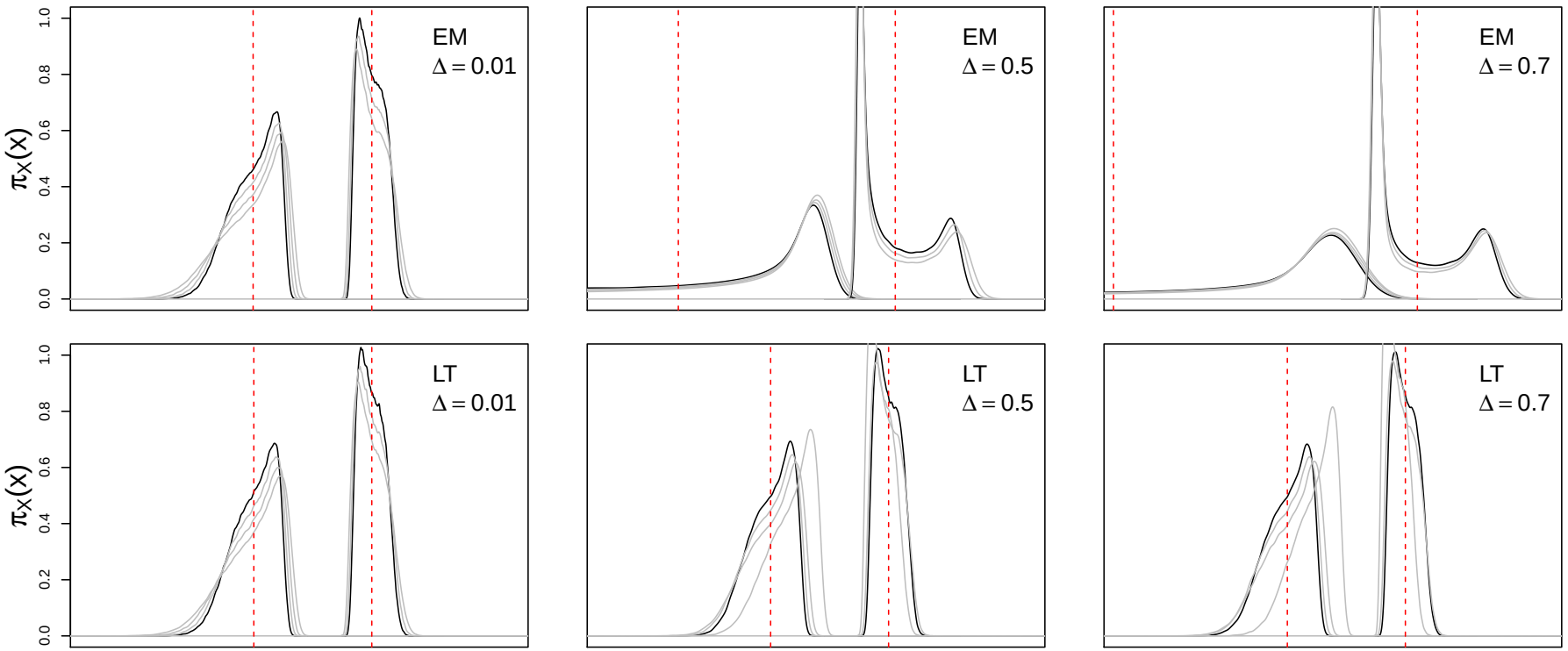


Figure: Empirical density of the diffusion, simulated with the Lie-Trotter splitting and the Euler-Maruyama scheme

We have proven (and demonstrated!) that the splitting scheme is a good approximation for the SDE.

We have proven (and demonstrated!) that the splitting scheme is a good approximation for the SDE.

What remains to show is that the approximation of the diffusion replicates well the dynamics of the PDMP.

Simulation of the PDMP: Thinning procedure

- **Main difficulty:** the intensity at the (future) spiking time t^* depends on the process up to time t^*
- **Acceptance-rejection procedure**:
  1. Find a "dominating" intensity Λ, which is strictly larger than the "true" intensity $\lambda(t)$.
  2. Simulate a "candidate" spiking time t^* as an exponential waiting time of a homogeneous Poisson process with intensity Λ
  3. Accept it as a "true" spiking time with probability $\frac{\lambda(t^*)}{\Lambda}$
- **Main reference:** Ogata (1981)

Simulation of the PDMP: Thinning procedure

- **Main difficulty:** the intensity at the (future) spiking time t^* depends on the process up to time t^*
- **Acceptance-rejection procedure**:
  1. Find a "dominating" intensity Λ, which is strictly larger than the "true" intensity $\lambda(t)$.
  2. Simulate a "candidate" spiking time t^* as an exponential waiting time of a homogeneous Poisson process with intensity Λ
  3. Accept it as a "true" spiking time with probability $\frac{\lambda(t^*)}{\Lambda}$
- **Main reference:** Ogata (1981)

Important note

The choice of Λ is of utmost importance: when the dominating intensity is not sharp enough, the algorithm will do a lot of rejects before accepting one point!

Simulation of the PDMP: Illustration

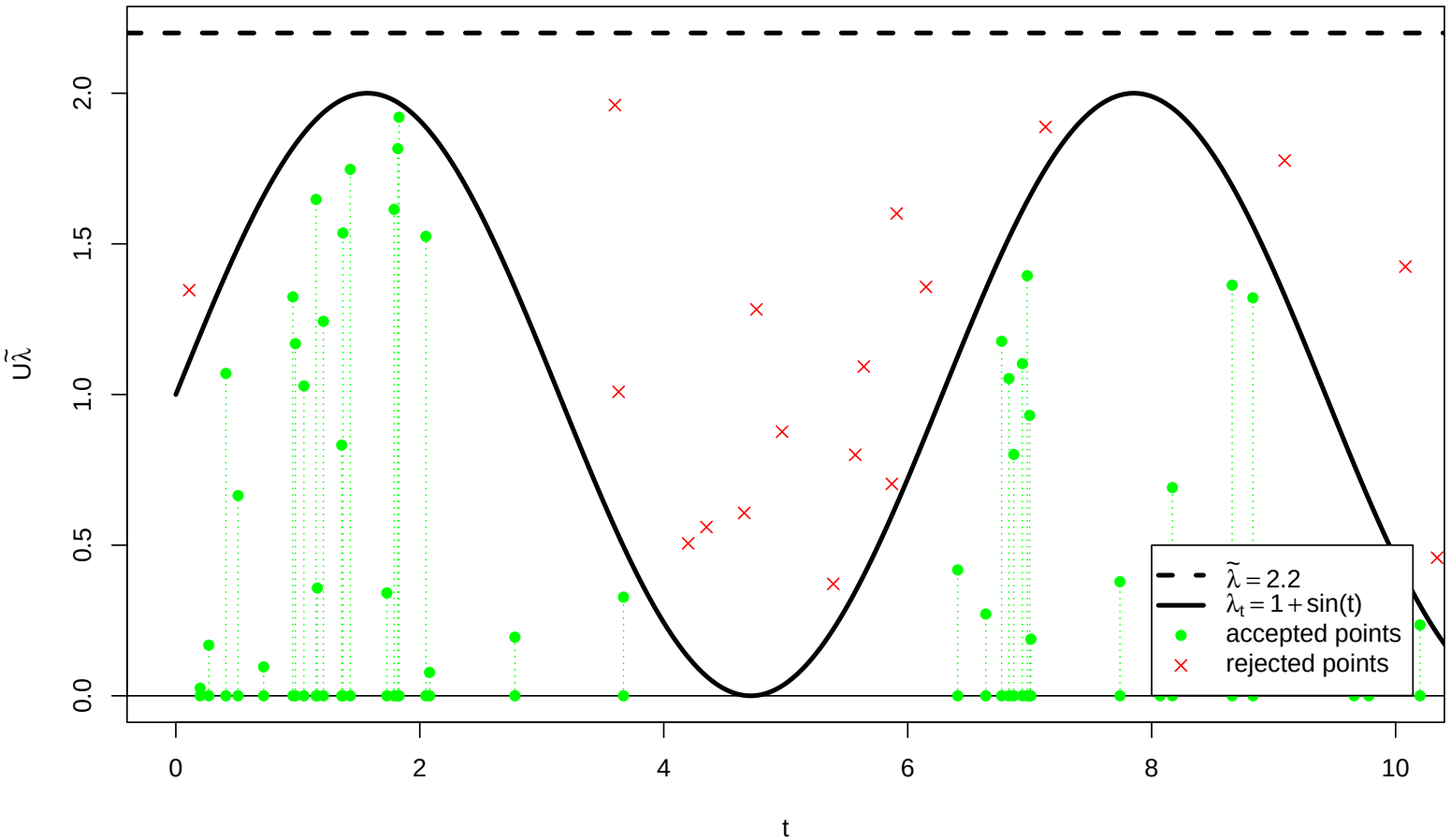


Figure: Thinning procedure for simulating a non-homogeneous Poisson process

Dynamics of PDMP

- When there is no spike, $\bar{X}$ evolves as the solution of an ODE, i.e. $\bar{X} = e^{At}x_0$
- Spike occurs with an intensity

$$\lambda_t^{k,n} = f_k\left(\left(e^{At}x_0\right)^{k,1}\right)$$

- f_k are non-decreasing, so we need to upper-bound $\left(e^{At}x_0\right)^{k,1}$.

Lemma (Global bound)

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

$$\Phi_k(x) \leq \tilde{\Phi}_k(x) = \max_{j=1,\dots,\eta_k+1}\left\{0, \frac{x^{k,j}}{\nu_k^{j-1}}\right\}. \tag{10}$$

Dynamics of PDMP

Lemma (Local bound)

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

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

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

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

Sharpness of bounds: illustration

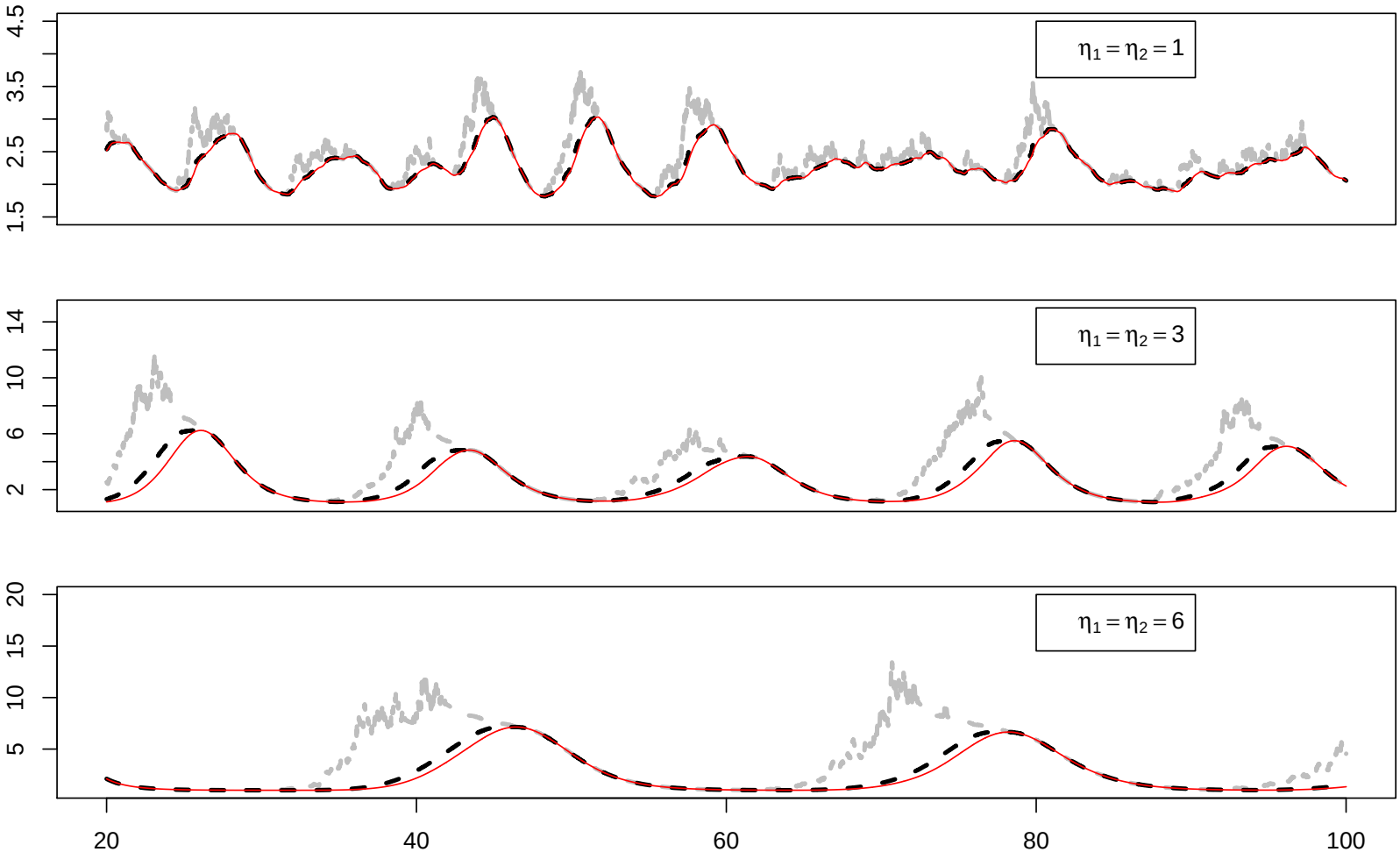


Figure: Gray line: global bound, dashed line: local bound, red line: true intensity.

Comparison: diffusion vs PDMP.

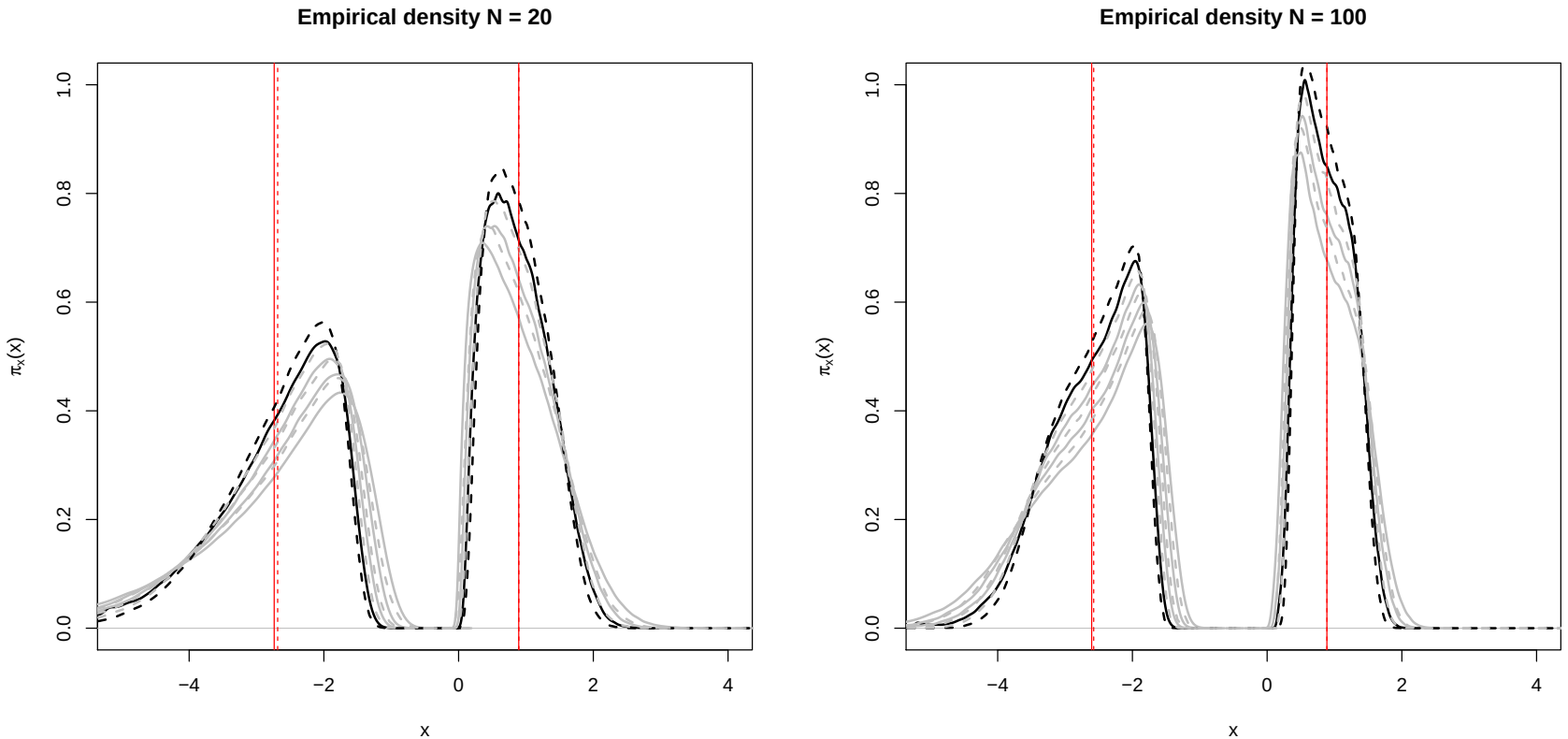


Figure: Empirical density of the diffusion vs empirical density of the PDMP (Markovian cascade)

Computational time.

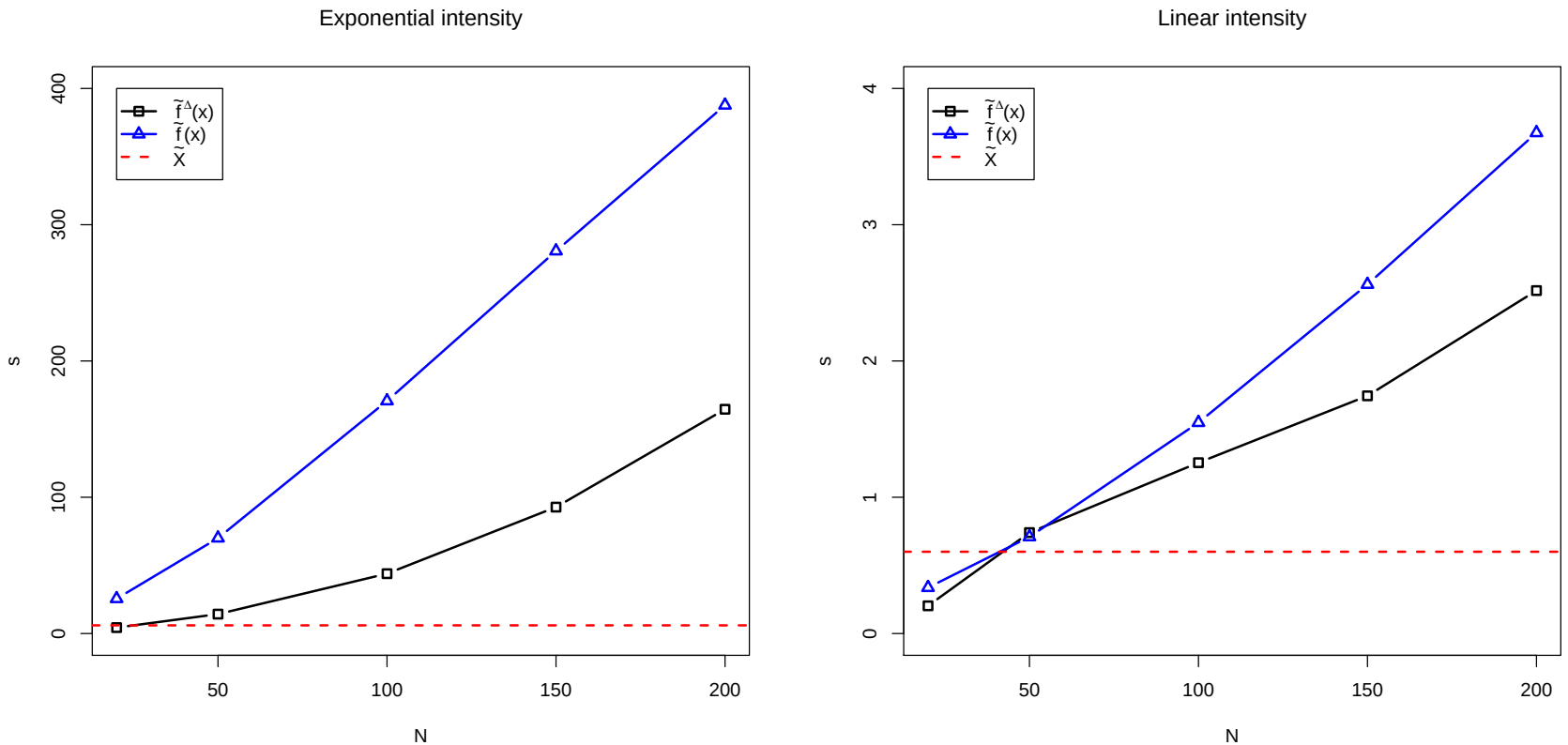


Figure: Execution time for exponential and linear intensity functions: PDMP, simulated with 2 methods (thinning procedure with local $\tilde{f}^{\Delta}(x)$ and global $\tilde{f}(x)$ intensity bounds), and the stochastic diffusion $\tilde{X}$.

Conclusion

- **Splitting scheme converges with the mean-square order** 1 (as Euler-Maruyama)...
- ... **it preserves first and the second moment**,
- ... **is ergodic**,
- ... as a result **it is stable even for large** Δ.
- As a consequence, **it allows to simulate the integrated intensity process**, describing the neuronal activity of a large network of neurons **at negligible computational cost**.
- **In perspective:** integration in simulation-based parametric inference procedures (for example, Approximate Bayesian Computation)?

Literature I

Markus Ableidinger and Evelyn Buckwar. Splitting integrators for the stochastic landau–lifshitz equation. *SIAM J. Sci. Comput.*, 38:A1788–A1806, 01 2016. doi: 10.1137/15M103529X.

Markus Ableidinger, Evelyn Buckwar, and Harald Hinterleitner. A stochastic version of the jansen and rit neural mass model: Analysis and numerics. *The Journal of Mathematical Neuroscience*, 7(1):1, August 2017. ISSN 2190-8567. doi: 10.1186/s13408-017-0046-4. URL `http://dx.doi.org/10.1186/s13408-017-0046-4`.

Evelyn Buckwar, Massimiliano Tamborrino, and Irene Tubikanec. Spectral density-based and measure-preserving abc for partially observed diffusion processes. an illustration on hamiltonian sdes. 2019.

Julien Chevallier, Anna Melnykova, and Irene Tubikanec. Theoretical analysis and simulation methods for Hawkes processes and their diffusion approximation. working paper or preprint, March 2020. URL `https://hal.archives-ouvertes.fr/hal-02513614`.

François Delarue and Stéphane Menozzi. Density Estimates for a Random Noise Propagating through a Chain of Differential Equations. *Journal of Functional Analysis*, 259:1577–1630, September 2010. URL `https://hal.archives-ouvertes.fr/hal-00436051`.

Susanne Ditlevsen and Eva Löcherbach. Multi-class oscillating systems of interacting neurons. *SPA*, 127: 1840–1869, 2017.

Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

Literature II

Peter Eris Kloeden, Eckhard Platen, and Henri Schurz. *Numerical solution of SDE through computer experiments*. Universitext. Springer, 2003. ISBN 3540570748,9783540570745. URL `http://gen.lib.rus.ec/book/index.php?md5=FED28E7C31D55C92DA839156DB965F99`.

B. Leimkuhler and C. Matthews. *Molecular dynamics: with deterministic and stochastic numerical methods*. Springer International Publ., Cham, 2015. ISBN 9783319163741; 3319163744.

Eva Löcherbach. Large deviations for cascades of diffusions arising in oscillating systems of interacting hawkes processes. *Journal of Theoretical Probability*, 32(1):131–162, 2019.

Jonathan C Mattingly, Andrew M Stuart, and Desmond J Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.

R. Mclachlan and G. Quispel. Splitting methods. *Acta Numer.*, 11:341–434, 01 2002. doi: 10.1017/S0962492902000053.

Yosihiko Ogata. On lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

T. Shardlow. Splitting for dissipative particle dynamics. *SIAM J. Sci. Comput.*, 24(4):1267–1282, 2003. doi: 10.1137/S1064827501392879.

G. Strang. On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.*, 5(3): 506–517, 1968. doi: 10.1137/0705041. URL `https://doi.org/10.1137/0705041`.

Thank you for your attention!