Statistical and numerical analysis of jump and diffusion models in biology

Anna Melnykova

Supervisors: Eva Löcherbach (Sorbonne Paris 1), Adeline Samson (Université Grenoble Alpes)

Paris Cergy Université, UMR-CNRS 8088
Université Grenoble Alpes, LJK UMR-CNRS 5224

8 December 2020

Plan of the presentation

1. **Introduction:** mathematical models of one and multiple neurons
2. **Simulation methods** for Stochastic Differential Equations
3. **Parametric inference** for SDE
4. **Non-asymptotic statistical test** of covariance matrix rank of SDE

One neuron: FitzHugh-Nagumo model

FitzHugh-Nagumo model (Fitzhugh, 1961, Nagumo et al., 1962)

$$\begin{cases} dV_t = \frac{1}{\varepsilon}\left(V_t - V_t^3 - U_t - s\right) dt \\ dU_t = (\gamma V_t - U_t + \beta)\, dt, \end{cases}$$

- V_t — the membrane potential, U_t — the input of ion channels.
- ε — scaling parameter, which regulates the time difference between the membrane voltage V_t, which changes at a faster scale than the recovery variable U_t,
- s — the stimuli current,
- γ and β — abstract constants determining the spiking or oscillatory behaviour of a neuron.

One neuron: FitzHugh-Nagumo model

FitzHugh-Nagumo model (Fitzhugh, 1961, Nagumo et al., 1962)

$$
\begin{cases}
dV_t = \frac{1}{\varepsilon}\left(V_t - V_t^3 - U_t - s\right) dt \quad +\sigma_1 \mathbf{dW_t^{(1)}} \\
dU_t = (\gamma V_t - U_t + \beta)\, dt \quad +\sigma_2 \mathbf{dW_t^{(2)}},
\end{cases}
$$

- V_t — the membrane potential, U_t — the input of ion channels.
- ε — scaling parameter, which regulates the time difference between the membrane voltage V_t, which changes at a faster scale than the recovery variable U_t,
- s — the stimuli current,
- γ and β — abstract constants determining the spiking or oscillatory behaviour of a neuron.
- σ_1 and σ_2 are diffusion coefficients, which can be null!

One neuron: FitzHugh-Nagumo model

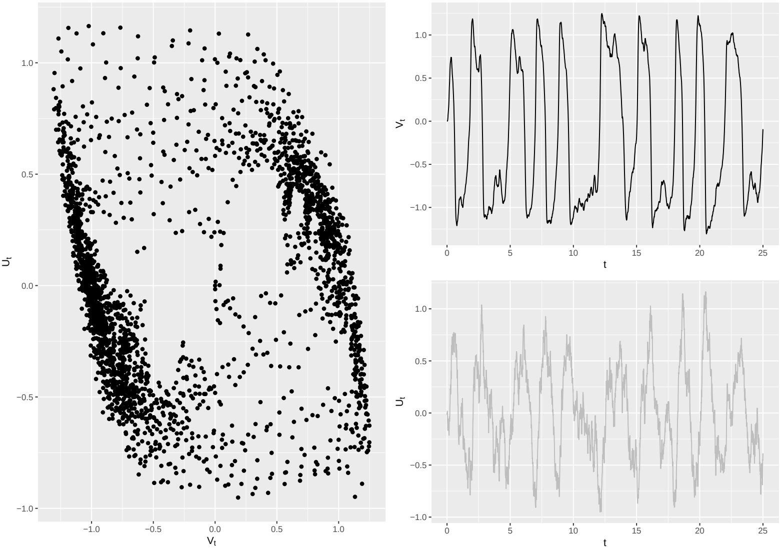


Figure: Neuronal activity simulated with stochastic FitzHugh-Nagumo model.

Population of neurons: point process

- **N neurons** structured in **K populations** (N_k neurons in each, $k = 1, \ldots, K$)
- **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** $(\lambda^{k,n}(t))_{t\geq 0}$,
- **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.
- $\mathbb{P}(Z_t^{k,n}$ has a jump in $(t, t+dt] | \mathcal{F}_t) = \lambda^{k,n}(t)dt$.

References

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

Population of neurons: point process

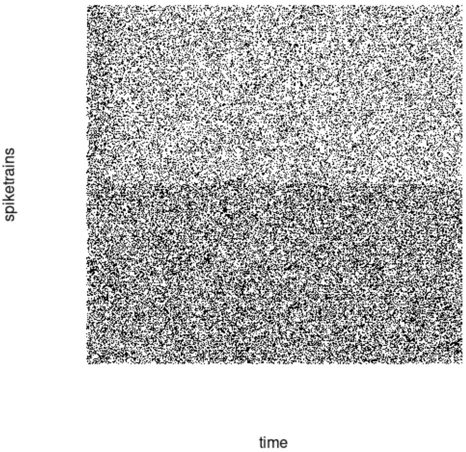


Figure: Simulation of the neuronal activity of a population of neurons as a point process

Some open questions

- How to **simulate these processes** as precise and efficient as possible?
- How to **estimate the parameters** of the models?
- How to **choose the right dimension of noise** for diffusion models?

I: Numerical simulation of stochastic diffusions[1].

[1]all the cited results in this part are from 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

Piece-wise deterministic Markov process

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

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

where

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

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

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

Piece-wise deterministic Markov process

$(\bar{X}_t)_{t\geq 0} = \{(\bar{X}_t^{k,j})_{t\geq 0},\ 1 \leq k \leq K, 1 \leq j \leq \eta_k + 1\}$ 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 $f_{k+1}(\bar{X}_{t-}^{k+1,1})$, $\bar{X}_0 = x_0 \in \mathbb{R}^\kappa$.

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 process by a diffusion!

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

$$dX_t = (AX_t + B(X_t))dt + \frac{1}{\sqrt{N}}\sigma(X_t)dW_t, \quad X_0 = x_0, \tag{4}$$

- $X = (X^1, X^2)^T$ 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 $(-\nu_k, \ldots, -\nu_k)$ and upper-diagonal equal to $(1, \ldots, 1)$,
- $B(X) = (B^1(X^2), B^2(X^1))^T$, where $B^1(X^2) = (0, \ldots, 0, c_1 f_2(X^{2,1}))$ and $B^2(X^1) = (0, \ldots, 0, c_2 f_1(X^{1,1}))$.
- $\sigma(X) = \left(\sigma^1(X^2), \sigma^2(X^1)\right)^T$, where σ^1 and σ^2 read as

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

Cascade SDE: hypoellipticity

Note that the diffusion matrix has zeroes almost everywhere. But in fact, the Brownian motion enters in every variable of the system!

We can write the solution of the system 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 zeroes almost everywhere. But in fact, the Brownian motion enters in every variable of the system!

We can write the solution of the system 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:

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

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

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)

Cascade diffusion: splitting approach

Main difficulties:

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

Main idea of splitting:

Instead of approximating the solution of the ”complicated” system, we can split the system in ”simpler” subsystems and solve them explicitly.

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 diffusion: splitting approach

Main idea of splitting:

Instead of approximating the solution of the "complicated" system, we can split the system in "simpler" subsystems and solve them explicitly.

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

Thus, we 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.

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

$$\tilde{X}_{t_{i+1}} = \left(\psi_{\Delta}^{[1]} \circ \psi_{\Delta}^{[2]}\right)\left(\tilde{X}_{t_i}\right) = e^{A\Delta}\left(\tilde{X}_{t_i} + \Delta B(\tilde{X}_{t_i}) + \frac{\sqrt{\Delta}}{\sqrt{N}}\sigma(\tilde{X}_{t_i})w_i\right), \tag{5}$$

where $(w_i)_{i=1,\dots,i_{\max}}$ i.i.d. $\mathcal{N}(0,1)$. Note that the approximated solution (5) 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:

$$\tilde{X}_{t_{i+1}} = \left(\psi_{\Delta}^{[1]} \circ \psi_{\Delta}^{[2]}\right)\left(\tilde{X}_{t_i}\right) = e^{A\Delta}\left(\tilde{X}_{t_i} + \Delta B(\tilde{X}_{t_i}) + \frac{\sqrt{\Delta}}{\sqrt{N}}\sigma(\tilde{X}_{t_i})w_i\right), \tag{5}$$

where $(w_i)_{i=1,\ldots,i_{\max}}$ 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, $i = 1, \ldots, i_{max}$, where $\|\cdot\|$ denotes the Euclidean norm.

Cascade diffusion: splitting approach

Important remark

Splitting scheme does not only converge, but also preserves **first** and **second moment bounds**, and **ergodicity**. The Euler-Maruyama scheme converges, but not necessarily preserves the properties of the underlying diffusion (Mattingly et al., 2002)!

Cascade diffusion: splitting approach

Important remark

Splitting scheme does not only converge, but also preserves **first** and **second moment bounds**, and **ergodicity**. The Euler-Maruyama scheme converges, but not necessarily preserves the properties of the underlying diffusion (Mattingly et al., 2002)!

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* $|g| \leq \tilde{G}^m$,

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

- **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 $(w_i)_{i=1,\ldots,\eta^*+1}$ such that

$$y = \underbrace{(\psi_\Delta[w_{\eta^*+1}] \circ \cdots \circ \psi_\Delta[w_1])}_{\eta^*+1}(x), \tag{6}$$

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

Splitting approach. Numerical performance.

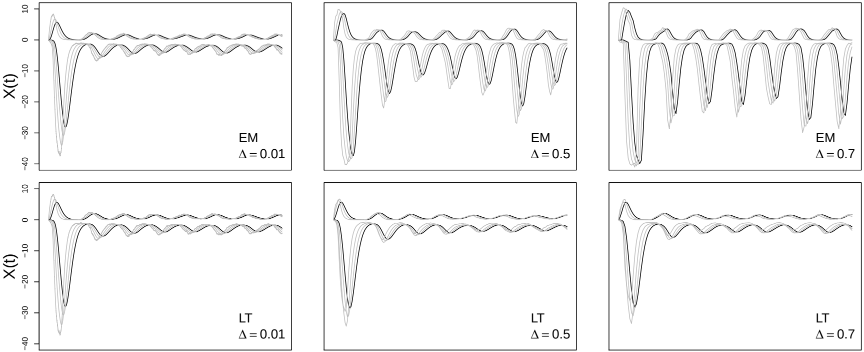


Figure: Trajectories of the diffusion, simulated with the Euler-Maruyama scheme and the Lie-Trotter splitting

Splitting approach. Numerical performance.

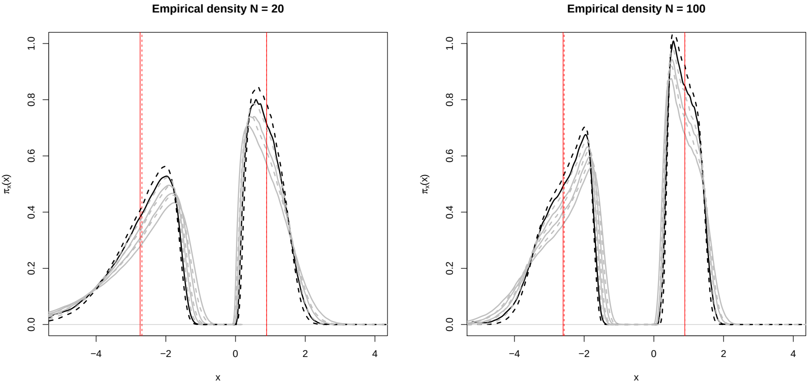


Figure: Empirical density of the diffusion vs empirical density of the PDMP (Markovian cascade)

Splitting approach. Computational time.

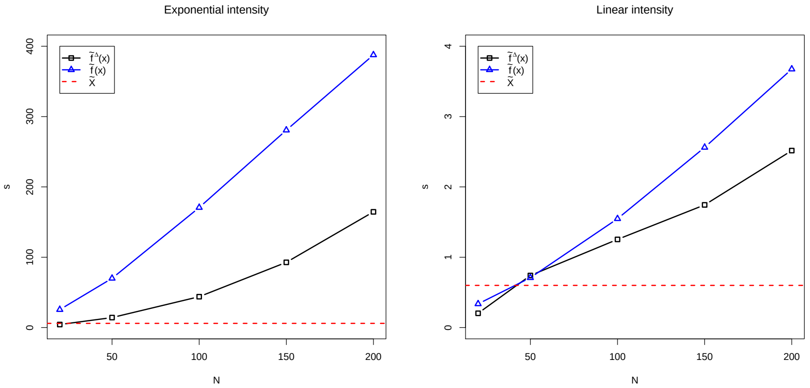


Figure: Execution time for exponential and linear intensity functions: PDMP, simulated with 2 methods (thinning procedure with local $\tilde{f}^{\Delta}(x)$ and global $\tilde{f}(x)$ intensity bounds), and the stochastic diffusion $\tilde{X}$.

Conclusion of the Chapter

- **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)?

II. Parametric inference in hypoelliptic diffusions [2].

[2] all the cited results are from A. Melnykova *"Parametric inference for hypoelliptic ergodic diffusions with full observations"* Statistical Inference for Stochastic Processes, 23 (2020), 595-635, DOI:https://doi.org/10.1007/s11203-020-09222-4

Model and assumptions

$$\begin{cases} dV_t = a_1(V_t, U_t; \theta^{(1)})dt \\ dU_t = a_2(V_t, U_t; \theta^{(2)})dt + b(V_t, U_t; \sigma)dW_t, \end{cases} \tag{7}$$

- $X_t := \ (V_t, U_t)^T \in \mathbb{R} \times \mathbb{R}$,
- $A(X_t; \theta) := \ (a_1(V_t, U_t; \theta^{(1)}), \ a_2(V_t, U_t; \theta^{(2)}))^T$ — drift term,
- $B(X_t; \sigma) := \ \begin{pmatrix} 0 & 0 \\ 0 & b(V_t, U_t; \sigma) \end{pmatrix}$ — **degenerate** diffusion coefficient,
- dW_t is a standard Brownian motion,
- $(\theta^{(1)}, \theta^{(2)}, \sigma) \in \Theta_1 \times \Theta_2 \times \Xi$ — vector of the unknown parameters.

Goal:

Estimation of $(\theta^{(1)}, \theta^{(2)}, \sigma)$ from $(V_{t_i}, U_{t_i})^T, \ i \in 1, \ldots n$ on time interval $[0, T], T = n\Delta$. $\mathbf{T} \to \infty, \Delta \to \mathbf{0}$

Main assumptions

A1 $a_1(v, u; \theta^{(1)}))$ and $a_2(v, u; \theta^{(2)}))$ have bounded partial derivatives of every order, uniformly in θ. Furthermore $\partial_u a_1 \neq 0 \quad \forall (v, u) \in \mathbb{R}^2$ **(ensures hypoellipticity)**.

A2 Global Lipschitz and linear growth conditions **(ensures existence of a unique strong solution)**.

A3 Process X_t is **ergodic** and there exists a unique invariant probability measure ν_0 with finite moments of any order.

A4 $a_1(X_t; \theta^{(1)})$, $a_2(X_t; \theta^{(2)})$ and $b(X_t; \sigma)$ are **identifiable**, that is $a_k(X_t; \theta^{(k)}) = a_k(X_t; \theta_0^{(k)}) \Leftrightarrow \theta^{(k)} = \theta_0^{(k)}$. Moreover, $b(X_t; \sigma) > 0$, $\partial_\sigma b(X_t; \sigma) \neq 0 \quad \forall t$.

A5° The following holds:

$$a_1(x; \theta^{(1)}) = f(u) + (\theta^{(1)})^T g(v),$$

where $g(x)$ is a vector-valued function of the same dimension as the vector $\theta^{(1)}$, $f(x)$ is a continuous function.

Difficulties:

- Degenerate diffusion coefficient $\longrightarrow$ non-invertible covariance matrix of the approximated transition density
- Each coordinate has a variance of different order $\longrightarrow$ difficult to analyze

Solution:

- Use a high-order scheme to "catch" the propagated noise in all coordinates
- Build a contrast estimator based on the approximated density

Contrast estimators in literature

- **Elliptic systems**: Kessler (1997), Genon-Catalot et al. (2000), Gloter (2006), Gloter and Sørensen (2009). General reference: Kutoyants (2013)
- **Hypoelliptic systems with $\mathbf{a_1}(\mathbf{V_t}, \mathbf{U_t}; \theta^{(1)}) \equiv \mathbf{U_t}$**: Ozaki (1989), Samson and Thieullen (2012), León et al. (2018)
- **General hypoelliptic systems**: Ditlevsen and Samson (2017)

Discretization: Local linearization scheme

Consider (7):

$$dX_s = A(X_s;\theta)ds + B(X_s;\sigma)d\tilde{W}_s, \quad s \in (\tau, \tau+\Delta].$$

Discretization: Local linearization scheme

On each small time interval of size Δ we approximate (7) by the linear equation:

$$dX_s \approx \left(A(X_\tau;\theta) + J(X_\tau;\theta)(X_s - X_\tau) + \frac{1}{2} b^2(X_\tau;\sigma)\partial^2_{uu} A(X_\tau;\theta)(s-\tau) \right) ds + B(X_\tau;\sigma)\, d\tilde{W}_s, \quad s \in (\tau, \tau+\Delta],$$

where $J(X_\tau;\theta)$ is a Jacobian of vector A.

Local Linearization

Ozaki (1989), Jimenez and Carbonell (2015)

Discretization: Local linearization scheme

On each small time interval of size Δ we approximate (7) by the linear equation:

$$dX_s \approx \left(A(X_\tau;\theta) + J(X_\tau;\theta)(X_s - X_\tau) + \frac{1}{2} b^2(X_\tau;\sigma)\partial^2_{uu} A(X_\tau;\theta)(s-\tau) \right) ds + B(X_\tau;\sigma) d\tilde{W}_s, \quad s \in (\tau, \tau+\Delta],$$

where $J(X_\tau;\theta)$ is a Jacobian of vector A. Note that X_s is a Gaussian process, with expectation and variance given by:

$$\mathbb{E}[X_t | X_\tau] = X_\tau + \int_\tau^t e^{J(X_\tau;\theta)(\tau+\Delta-s)} \cdot \left(A(X_\tau;\theta) - J(X_\tau;\theta)X_\tau + \frac{1}{2} b^2(X_\tau;\sigma)\partial^2_{uu} A(X_\tau;\theta)(s-\tau) \right) ds$$

$$\Sigma(X_t;\theta,\sigma^2) = \mathbb{E}\left[\left(\int_\tau^t e^{J_\tau (t-s)} B(X_\tau;\sigma) d\tilde{W}_s \right) \left(\int_\tau^t e^{J_\tau (t-s)} B(X_\tau;\sigma) d\tilde{W}_s \right)^T \right]. \quad (8)$$

Discretization: Local linearization scheme

Then, we can recursively approximate the solution of (7) at time $i\Delta$:

$$\tilde{X}_{t_{i+1}} = \bar{A}(\tilde{X}_{t_i}; \theta, \sigma) + \bar{B}(\tilde{X}_{t_i}; \theta, \sigma)\Xi_i,$$

- Ξ_i — standard Gaussian 2-dimensional random vector
- $\bar{A}$ is a discrete approximation of $\mathbb{E}[X_t | X_\tau]$.
- $\bar{B}$ — any matrix s. t. $\bar{B}\bar{B}^T = \Sigma(X_t; \theta, \sigma^2)$.

Proposition

The second-order Taylor approximation of matrix $\Sigma(\tilde{X}; \theta, \sigma^2)$ has the following form:

$$b^2(\tilde{X}_\tau; \sigma) \begin{pmatrix} (\partial_u a_1)^2 \frac{\Delta^3}{3} & (\partial_u a_1)\frac{\Delta^2}{2} + (\partial_u a_1)(\partial_u a_2)\frac{\Delta^3}{3} \\ (\partial_u a_1)\frac{\Delta^2}{2} + (\partial_u a_1)(\partial_u a_2)\frac{\Delta^3}{3} & \Delta + (\partial_u a_2)\frac{\Delta^2}{2} + (\partial_u a_2)^2\frac{\Delta^3}{3} \end{pmatrix} + O(\Delta^4),$$

where the derivatives are computed at time τ.

Contrast estimator

The contrast function is defined as follows:

$$\mathcal{L}(\theta, \sigma^2; \tilde{X}_{t_0:t_n}) = \frac{1}{2}\sum_{i=0}^{n-1}(\tilde{X}_{t_{i+1}} - \bar{A}(\tilde{X}_{t_i};\theta))^T \Sigma_\Delta^{-1}(\tilde{X}_{t_{i+1}} - \bar{A}(\tilde{X}_{t_i};\theta)) + \sum_{i=0}^{n-1} \log\det(\Sigma_\Delta).$$

The estimator is then:

$$(\hat{\theta}^{(1)}_{n,\Delta}, \hat{\theta}^{(2)}_{n,\Delta}, \hat{\sigma}^2_{n,\Delta}) = \underset{\theta^{(1)},\theta^{(2)},\sigma^2}{\arg\min}\ \mathcal{L}(\theta^{(1)}, \theta^{(2)}, \sigma^2; \tilde{X}_{t_0:t_n})$$

Contrast estimator

The contrast function is defined as follows:

$$\mathcal{L}(\theta, \sigma^2; \tilde{X}_{t_0:t_n}) = \frac{1}{2}\sum_{i=0}^{n-1}(\tilde{X}_{t_{i+1}} - \bar{A}(\tilde{X}_{t_i};\theta))^T \Sigma_\Delta^{-1}(\tilde{X}_{t_{i+1}} - \bar{A}(\tilde{X}_{t_i};\theta)) + \sum_{i=0}^{n-1}\log\det(\Sigma_\Delta).$$

The estimator is then:

$$(\hat{\theta}^{(1)}_{n,\Delta}, \hat{\theta}^{(2)}_{n,\Delta}, \hat{\sigma}^2_{n,\Delta}) = \underset{\theta^{(1)},\theta^{(2)},\sigma^2}{\arg\min}\ \mathcal{L}(\theta^{(1)}, \theta^{(2)}, \sigma^2; \tilde{X}_{t_0:t_n})$$

Analogous approach

Contrasts can also be built on marginal densities of each coordinate as in Ditlevsen and Samson (2017)!

Theorem (Consistency and asymptotic normality of $\hat{\theta}^{(1)}_{n,\Delta}$.)

When $\Delta \to 0$, $n\Delta \to \infty$ and $n\Delta^2 \to 0$, the following holds:

$$\hat{\theta}^{(1)}_{n,\Delta} \xrightarrow{\mathbb{P}_0} \theta^{(1)}_0,$$

$$\sqrt{\frac{n}{\Delta}}\left(\hat{\theta}^{(1)}_{n,\Delta} - \theta^{(1)}_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ 3\left(\int \frac{(\partial_{\theta^{(1)}} a_1)(\partial_{\theta^{(1)}} a_1)^T}{b^2(x;\sigma)(\partial_y a_1)}\nu_0(dx)\right)^{-1}\right.$$
$$\left.\int \frac{b^2(x;\sigma_0)}{b^4(x;\sigma)}(\partial_{\theta^{(1)}} a_1)(\partial_{\theta^{(1)}} a_1)^T\left(1+\frac{1}{(\partial_u a_1)^2}\right)\nu_0(dx)\right),$$

where $\partial_u a_1$ is a simplified notation for $\partial_u a_1(x;\theta^{(1)}_0)$

Theorem (Consistency and asymptotic normality of $\hat{\theta}^{(2)}_{n,\Delta}$ and $\hat{\sigma}_{n,\Delta}$.)

When $\Delta \to 0$, $n\Delta \to \infty$ and $n\Delta^2 \to 0$ the following holds:

$$\hat{\theta}^{(2)}_{n,\Delta} \xrightarrow{\mathbb{P}_0} \theta^{(2)}_0, \quad \hat{\sigma}_{n,\Delta} \xrightarrow{\mathbb{P}_0} \sigma_0$$

and

$$\sqrt{n\Delta}(\hat{\theta}^{(2)}_{n,\Delta} - \theta^{(2)}_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \left(\int \frac{(\partial_{\theta^{(2)}} a_2(x; \theta^{(2)}_0))(\partial_{\theta^{(2)}} a_2(x; \theta^{(2)}_0))^T}{b^2(x, \sigma)} \nu_0(dx)\right)^{-1}\right)$$

$$\sqrt{n}(\hat{\sigma}_{n,\Delta} - \sigma_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, 2\left(\int \frac{(\partial_\sigma b(x, \sigma_0))(\partial_\sigma b(x, \sigma_0))^T}{b^2(x, \sigma_0)} \nu_0(dx)\right)^{-1}\right).$$

Consistency and asymptotic normality results

Main idea of the proof:

- Find a sequence of estimators $\hat{\theta}^{(1)}_{n,\Delta_n}$ such that the sequence $(\hat{\theta}^{(1)}_{n,\Delta_n} - \theta^{(1)}_0)$ is tight (as in Gloter and Sørensen (2009))
- Use the tightness in combination with the rate of convergence obtained for $\hat{\theta}^{(1)}_{n,\Delta}$ to prove the convergence of $\hat{\theta}^{(2)}_{n,\Delta}$ and $\hat{\sigma}_{n,\Delta}$

Why (A5) is needed?

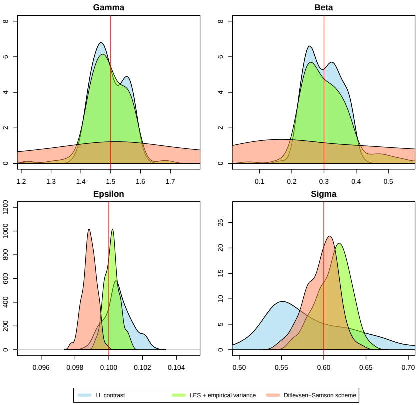
- When $\theta^{(1)}$ is included in the derivative $\partial_y a_1$, this parameter is present both in the drift and the variance term. It causes troubles, because drift and variance converge at different rates.
- Linear shape of a_1 with respect to $\theta^{(1)}$ allows to fully use the speed of convergence of $\hat{\theta}^{(1)}_{n,\Delta_n}$ when studying the convergence of the other parameters.

Numerical performance of the contrast estimator

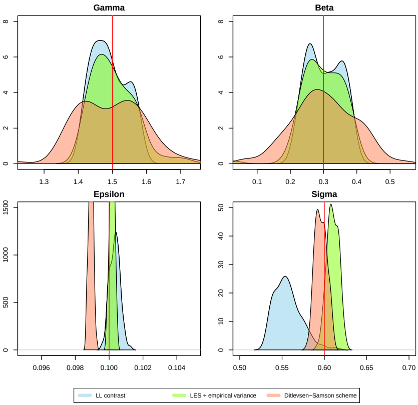
FitzHugh-Nagumo model

$$\begin{cases} dV_t = \frac{1}{\varepsilon}(V_t - V_t^3 - U_t - s)dt \\ dU_t = (\gamma V_t - U_t + \beta)dt + \sigma dW_t, \end{cases}$$

- Parameters to estimate are: $\gamma, \beta, \varepsilon, \sigma^2$.
- Data is simulated with $\Delta = 10^{-4}$, $T = 50$.
- Trajectories are subsampled, and the horizon of observations is cut (for more realistic conditions).
- Linearized constrast estimator (LC) is compared to the alternative methods: conditional least square estimator (LSE) (Melnykova, 2020), and 1.5 order scheme (Ditlevsen and Samson, 2017).



(a) $T = 5$



(b) $T = 50$

Figure: Estimation density for the LL contrast (blue), the LSE (red) and 1.5 scheme (green) estimators for the excitatory set. $\Delta = 0.01$

Conclusion of the Chapter

- Parameters can be estimated through the **minimization of a simple unique criteria**.
- **Works in practice even for a much larger class** of models. For example, FitzHugh-Nagumo satisfies only (A3)-(A4).
- Drawback: to be applied to real data, the method must be able to work with **missing observations** (when only V_t is available).
- Another drawback: when we apply the method to a 2-dimensional model, **we implicitly assume that $\boldsymbol{\sigma_1 \equiv 0}$**, and are estimating σ_2 only.

III: What is the dimension of the noise? [3]

[3] Joint work with Adeline Samson and Patricia Reynaud-Bouret.

2-dimensional case (drift known)

Consider a 2-dimensional process, defined by the solution of:

$$dX_t = Adt + BdW_t, \tag{9}$$

- $A = (a_1, a_2)^T$ is a drift vector (supposed to be known)
- $B = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$,
- W is a 2-dimensional Brownian motion.

Non-asymptotic setting

Observations of the process are available with a **fixed time step Δ**!

Test

We are interested in the following hypotheses:

$$H_0 : \sigma_1^2\sigma_2^2 = \delta$$
$$H_1 : \sigma_1^2\sigma_2^2 \geq \delta,$$

where δ is some chosen "sensitivity" threshold.

Statistics (Jacod et al., 2008, Jacod and Podolskij, 2013)

Consider vectors of increments:

$$\begin{pmatrix} \frac{X^{(1)}_{t_{2i+1}} - X^{(1)}_{t_{2i}}}{\sqrt{\Delta}} & \frac{X^{(1)}_{t_{2i+2}} - X^{(1)}_{t_{2i+1}}}{\sqrt{\Delta}} \\ \frac{X^{(2)}_{t_{2i+1}} - X^{(2)}_{t_{2i}}}{\sqrt{\Delta}} & \frac{X^{(2)}_{t_{2i+2}} - X^{(2)}_{t_{2i+1}}}{\sqrt{\Delta}} \end{pmatrix} =: mat\left(\xi_i^1, \xi_i^2\right)$$

Denote $s_i = \det mat\left(\xi_i^1, \xi_i^2\right)^2$.

2-dimensional case (drift known)

The distribution of s_i is difficult to write explicitly. But if we center the statistics, using the known drift, the cumulative distribution function is explicitly known!

Proposition

Denote $\dot{s}_i := \det\left(mat\left(\xi_i^1 - \sqrt{\Delta}A, \xi_i^2 - \sqrt{\Delta}A\right)\right)^2$. *The following holds for all i:*

$$\mathbb{P}\left(\dot{s}_i \leq x\right) = 1 - \left(\sqrt{\frac{x}{\sigma_1^2\sigma_2^2}} + 1\right) e^{-\sqrt{\frac{x}{\sigma_1^2\sigma_2^2}}}$$

NB: Based on the result of Wells et al. (1962) about the product of two chi-squared variable with k and $k-1$ degrees of freedom.

2-dimensional case (drift known)

Proposition

Under H_0 the following bound holds:

$$\mathbb{P}_0\left(\dot{S} \geq \delta\left(1 + L_W\left(-\frac{\alpha^{1/n}}{e}\right)\right)^2\right) \leq \alpha \quad \forall \alpha > 0,$$

where $\dot{S} = \frac{1}{n}\sum_{i=1}^{n} \dot{s}$ and L_W denotes Lambert W function.

Then the decision rule is the following:

$$H_0 \text{ is rejected if } \dot{S} \geq z_\alpha,$$

where $z_\alpha := \delta\left(1 + L_W\left(-\frac{\alpha^{1/n}}{e}\right)\right)^2$ plays the role of the quantile.

2-dimensional case (drift known)

Proposition (Type II risk)

For fixed levels of Type I and Type II risks $\alpha > 0$ and $\beta > 0$ respectively and if

$$\sigma_1^2\sigma_2^2 \geq \delta \left(\frac{1 + L_W\left(-\frac{\alpha^{1/n}}{e}\right)}{1 + L_W\left(-\frac{(1-\beta)^{1/n}}{e}\right)} \right)^2,$$

the following inequality holds: $\quad \mathbb{P}_1\left(\dot{S} \leq z_\alpha\right) \leq \beta.$

Proposition

For $\alpha, \beta \in (0,1)$, the following holds when $n \to \infty$:

$$\lim_{n\to\infty} \left(\frac{1 + L_W\left(-\frac{\alpha^{1/n}}{e}\right)}{1 + L_W\left(-\frac{(1-\beta)^{1/n}}{e}\right)} \right)^2 = \frac{\ln \alpha}{\ln(1-\beta)}.$$

Generalization to a d-dimensional case

Now, consider a d-dimensional process X, which solves

$$dX_t = A_t dt + B_t dW_t,$$

where A_t, B_t — unknown time-dependent d and $d \times d$-dimensional drift and diffusion coefficients respectively. Discrete observations of X are available with a *fixed* time step Δ.

We are interested in the following hypotheses:

$$H_0 : \sigma_1^2 \cdot \dots \cdot \sigma_d^2 = \delta$$
$$H_1 : \sigma_1^2 \cdot \dots \cdot \sigma_d^2 \geq \delta.$$

Test in the asymptotic case

Jacod et al. (2008), Jacod and Podolskij (2013)

Main statistics of the test

Consider the matrix:

$$\Xi_i := \begin{pmatrix} \frac{X^{(1)}_{id+1} - X^{(1)}_{id}}{\sqrt{\Delta}} & \frac{X^{(1)}_{id+2} - X^{(1)}_{id+1}}{\sqrt{\Delta}} & \dots & \frac{X^{(1)}_{id+d} - X^{(1)}_{id+d-1}}{\sqrt{\Delta}} \\ \frac{X^{(2)}_{id+1} - X^{(2)}_{id}}{\sqrt{\Delta}} & \frac{X^{(2)}_{id+2} - X^{(2)}_{id+1}}{\sqrt{\Delta}} & \dots & \frac{X^{(2)}_{id+2d} - X^{(2)}_{id+d-1}}{\sqrt{\Delta}} \\ \dots & \dots & \dots & \dots \\ \frac{X^{(d)}_{id+1} - X^{(d)}_{id}}{\sqrt{\Delta}} & \frac{X^{(d)}_{id+2} - X^{(d)}_{id+1}}{\sqrt{\Delta}} & \dots & \frac{X^{(d)}_{id+d} - X^{(d)}_{id+d-1}}{\sqrt{\Delta}} \end{pmatrix}^2, \tag{10}$$

$i = 1, \dots, n$, and we denote each vector-column in this matrix by ξ_i^j.

The main statistics of the test is defined as follows:

$$S = \frac{1}{n} \sum_{i=1}^{n} \det \Xi_i^2$$

Main reference

Jacod et al. (2008), Jacod and Podolskij (2013)

Concentration inequalities for $S = \frac{1}{n}\sum_{i=1}^{n} \det \Xi_i^2$

Lemma (Sub-gaussian lower tail)

The following inequality holds:

$$\mathbb{P}\left(S - \mathbb{E}[S] \leq -\varepsilon\right) \leq \exp\left(\frac{-\varepsilon^2 n^2}{4\sum_{i=1}^{n}\mathbb{E}[\det \Xi_i^4]}\right).$$

Note: here it is difficult to evaluate $\mathbb{E}[\det \Xi_i^4]$!

Lemma (Upper tail)

The following bound holds:

$$\mathbb{P}\left(S - \mathbb{E}[S] \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2 n^2}{\sum_{i=1}^{n}\prod_{j=1}^{d}\left(\mathbf{tr}\left(\Sigma_i^j\right) + \|\mu_i^j\|^2 + c_i^j\right)}\right) + dne^{-c},$$

where c_i^j is a constant.

Upper tail of $S = \frac{1}{n}\sum_{i=1}^{n} \det \Xi_i^2$. Details of the proof.

- Probability space Ω is splitted in two subspaces: Ω_c^+ contains events in which $\det \Xi_i^2$ is bounded, and the other is equal to $\bar{\Omega}_c^+ = \Omega \setminus \Omega_c^+$.
- Define the set $\Omega_c^+ \subset \Omega$ as follows:

$$\Omega_c^+ := \left\{ \left\|\xi_i^j\right\|^2 \leq \mathbb{E}\left[\left\|\xi_i^j\right\|^2\right] + c_i^j \quad \forall j \in 1, \dots d,\ i = 1, \dots, n \right\}, \tag{11}$$

where $\|\cdot\|$ is the Euclidean norm,
- c_i^j is given as follows:

$$c_i^j = tr\left(\Sigma_i^j\right)\left(d + 2\sqrt{dc} + 2c - 1\right) + 2\left\|\mu_i^j\right\|^2 \sqrt{\frac{c}{d}},$$

and the constant c is independent of i and j.

Upper tail of $S = \frac{1}{n}\sum_{i=1}^{n} \det \Xi_i^2$. Details of the proof.

We evaluate the probability by the following two lemmas:

Lemma

The following holds:

$$\mathbb{P}(\bar{\Omega}_c^+) \le dne^{-c}$$

Lemma

In Ω_c^+ the following inequality holds:

$$\mathbb{P}\left(S - \mathbb{E}[S] \ge \varepsilon \middle| \Omega_c^+\right) \le \exp\left(-\frac{2\varepsilon^2 n^2}{\sum_{i=1}^{n}\prod_{j=1}^{d}\left(\mathbf{tr}\left(\Sigma_i^j\right) + \|\mu_i^j\|^2 + c_i^j\right)}\right).$$

Conclusions of the Chapter

- **In 2-dimensional case:** the law of S can be written explicitly.
  - For a practical application, a good estimator of drift is required.
- **In d-dimensional case:** the law of S cannot be written explicitly, but the lower and upper tail probabilities can be evaluated with concentration inequalities.
  - The lower bound is difficult to evaluate because it depends on the moments of $\det \Xi_i^2$.
  - The upper-bound is not sub-Gaussian and its sharpness decreases rapidly as d increases.
  - The constants are difficult to tune.

Global perspectives and further questions

- **Numerics**:
  - Is it possible to propose a "good" scheme preserving hypoellipticity for a general SDE with arbitrary number of noisy and non-noisy variables?
- **Statistics:**
  - Implement the testing procedure from the last Chapter and test it on neuronal models.
  - Is it possible to estimate the parameters of PDMP using contrast-type estimators? (*Ongoing work with Susanne Ditlevsen and Adeline Samson*)
  - Is it possible to estimate the parameters of PDMP using 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`.

S. Ditlevsen and A. Samson. Hypoelliptic diffusions: discretization, filtering and inference from complete and partial observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 07 2017. doi: 10.1111/rssb.12307.

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

Literature II

Richard Fitzhugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1(6):445–466, 1961.

V Genon-Catalot, T Jeantheau, and C Larédo. Stochastic volatility models as hidden markov models and statistical applications. bernoulli 6 1051–1079. *Mathematical Reviews (MathSciNet)*, 10:3318471, 2000.

Arnaud Gloter. Parameter estimation for a discretely observed integrated diffusion process. *Scandinavian Journal of Statistics*, 33(1):83–104, 2006.

Arnaud Gloter and Michael Sørensen. Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Processes and their Applications*, 119(3):679–699, 2009. URL `https://EconPapers.repec.org/RePEc:eee:spapps:v:119:y:2009:i:3:p:679-699`.

Jean Jacod and Mark Podolskij. A test for the rank of the volatility process: the random perturbation approach. *The Annals of Statistics*, 41(5):2391–2427, 2013.

Jean Jacod, Antoine Lejay, Denis Talay, et al. Estimation of the brownian dimension of a continuous itô process. *Bernoulli*, 14(2):469–498, 2008.

J.C. Jimenez and F. Carbonell. Convergence rate of weak local linearization schemes for stochastic differential equations with additive noise. *Journal of Computational and Applied Mathematics*, 279:106 – 122, 2015. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2014.10.021. URL `http://www.sciencedirect.com/science/article/pii/S0377042714004671`.

Mathieu Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24(2):211–229, 1997.

Literature III

Yury A Kutoyants. *Statistical inference for ergodic diffusion processes.* Springer Science & Business Media, 2013.

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

José León, Luis Rodriguez, and Roberto Ruggiero. Consistency of a likelihood estimator for stochastic damping hamiltonian systems. totally observed data. 01 2018.

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.

Anna Melnykova. Parametric inference for hypoelliptic ergodic diffusion with full observations. *Statistical Inference for Stochastic Processes*, 20:595–635, July 2020. doi: 10.1007/s11203-020-09222-4.

Jinichi Nagumo, Suguru Arimoto, and Shuji Yoshizawa. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50(10):2061–2070, 1962.

Thoru Ozaki. Statistical identification of nonlinear random vibration systems. *Journal of Applied Mechanics*, 56:186–191, 1989.

Literature IV

Adeline Samson and Michèle Thieullen. Contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Processes and their Applications*, 122:2521–2552, 2012. doi: 10.1016/j.spa.2012.04.006. URL `https://hal.archives-ouvertes.fr/hal-00598553`.

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

WT Wells, RL Anderson, John W Cell, et al. The distribution of the product of two central or non-central chi-square variates. *The Annals of Mathematical Statistics*, 33(3):1016–1020, 1962.

Thank you for your attention!

Supplementary material

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 $t \geq 0$,

$$\sup_{x \in \mathbb{R}^k} |\mathbb{E}_x \varphi(\bar{X}_t) - \mathbb{E}_x \varphi(X_t)| \leq Ct \frac{\|\varphi\|_{4,\infty}}{N^2}, \tag{12}$$

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

Theorem (Chevallier et al. (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.

Splitting approach. Empirical densities.

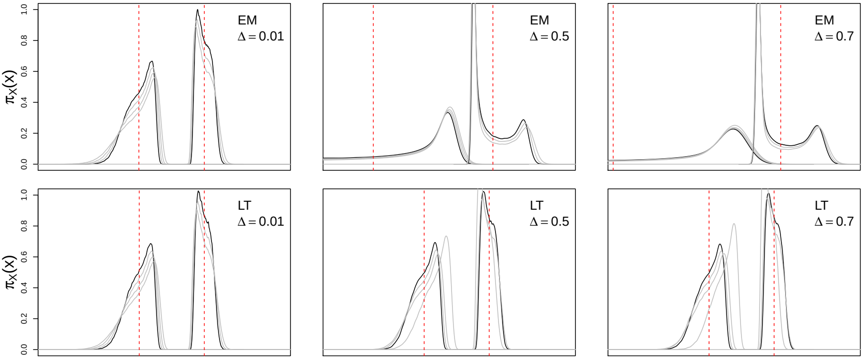


Figure: Empirical density of the diffusion, simulated with the Lie-Trotter splitting and the Euler-Maruyama scheme

Part II: How to prove hypoellipticity in 2d example

1. Rewrite (7) in Stratonovich form
2. Write the coefficients as two vector fields

$$A_0(v,u) = \begin{pmatrix} a_1(v,u;\theta^{(1)}) \\ a_2(v,u;\theta^{(2)}) - \frac{1}{2} b(v,u;\sigma)\partial_u b(v,u;\sigma) \end{pmatrix}$$
$$A_1(v,u) = \begin{pmatrix} 0 \\ b(v,u;\sigma) \end{pmatrix}.$$

3. Compute Lie bracket

$$[A_0, A_1] = \begin{pmatrix} \partial_y a_1(v,u;\theta^{(1)}) \\ \partial_u a_2(v,u;\theta^{(2)}) - \frac{1}{2}\partial_u b(v,u;\sigma)\partial^2_{uv} b(v,u;\sigma) \end{pmatrix}.$$

4. Hörmander condition is satisfied, as long as the first element of $[A_0, A_1]$ is non-zero (so, $[A_0, A_1]$ and $A_1(v,u)$ generate $\mathbb{R}^d$.

The drift approximation is following:

$$\begin{aligned}
\bar{A}_1(\tilde{X}_{t_i};\theta,\sigma) := {} & \tilde{V}_{t_i} + \Delta a_1(\tilde{X}_{t_i};\theta^{(1)}) + \\
& \frac{\Delta^2}{2}\left(\partial_x a_1(\tilde{X}_{t_i};\theta^{(1)}) + \partial_y a_1(\tilde{X}_{t_i};\theta^{(1)})\right) a_1(\tilde{X}_{t_i};\theta^{(1)}) + \frac{\Delta^2}{4} b^2(\tilde{X}_{t_i};\sigma)\partial_{vv} a_1(\tilde{X}_{t_i};\theta^{(1)}) \\
\bar{A}_2(\tilde{X}_{t_i};\theta,\sigma) := {} & \tilde{U}_{t_i} + \Delta a_2(\tilde{X}_{t_i};\theta^{(2)}) + \\
& \frac{\Delta^2}{2}\left(\partial_x a_2(\tilde{X}_{t_i};\theta^{(2)}) + \partial_y a_2(\tilde{X}_{t_i};\theta^{(2)})\right) a_2(\tilde{X}_{t_i};\theta^{(2)}) + \frac{\Delta^2}{4} b^2(\tilde{X}_{t_i};\sigma)\partial_{vv} a_2(\tilde{X}_{t_i};\theta^{(2)}).
\end{aligned}$$

Part III: Lambert W function

L_W is a multivariate function, with two real-valued branches giving the solution of the following equation:

$$ye^y = x, x \geq -\frac{1}{e}, \tag{13}$$

$$y = \begin{cases} (L_W)_{-1}(x) & \text{if } x \in [-\frac{1}{e}, 0), \\ (L_W)_0(x) & \text{if } x \geq 0. \end{cases}$$