Parameter estimation techniques for hypoelliptic ergodic diffusions

Student: **Anna Melnykova**
Supervisor: Adeline Leclercq-Samson

M2SIAM, STAT

21 June 2017

UFR IM²AG
UNIVERSITÉ Grenoble Alpes

Motivation

UNIVERSITÉ Grenoble Alpes

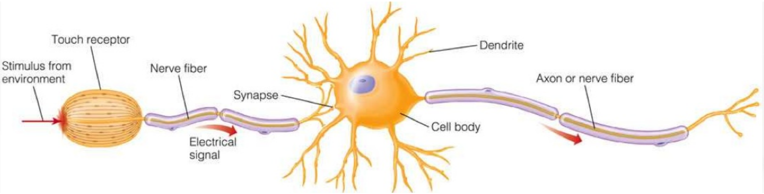

Figure 1: Touch sensor neuron

- **Membrane potential**: difference between the voltage in the interior and the exterior of the cell
- **Spikes**: stereotypic events, when the membrane potential becomes bigger than some threshold.

Introduction

Q: *How the neuronal activity can be described?*

A: Two approaches:

- Point process (Poisson, Hawkes process)
- Multidimensional stochastic diffusion

Introduction

Q: *How the neuronal activity can be described?*
A: Two approaches:

- Point process (Poisson, Hawkes process)
- Multidimensional stochastic diffusion

State-of-the-art

"All models are wrong but some are useful"
(Box and Draper, 1987, Gribbin, 2009; Paninski et al., 2009).

Example 1: Hodgkin-Huxley model

Experimentally it was shown that the behaviour of the neuron can be described by the following system of ODEs:

$$\begin{cases} I = C_m \frac{\mathrm{d}V_m}{\mathrm{d}t} + \bar{g}_K n^4 (V_m - V_K) + \bar{g}_{\mathsf{Na}} m^3 h (V_m - V_{Na}) + \bar{g}_l (V_m - V_l), \\ \frac{dn}{dt} = \alpha_n(V_m)(1-n) - \beta_n(V_m) n \\ \frac{dm}{dt} = \alpha_m(V_m)(1-m) - \beta_m(V_m) m \\ \frac{dh}{dt} = \alpha_h(V_m)(1-h) - \beta_h(V_m) h \end{cases}$$

where I — current per unit area, and α_i and β_i — rate functions for the i-th ion channel, $\bar{g}_n$ — maximal value of the conductance. n, m and $h \in (0,1)$ are associated with potassium channel activation, sodium channel activation, and sodium channel inactivation, respectively.

Examples

Example 2: FitzHugh-Nagumo

The behaviour of the neuron is defined through the solution of the system:

$$\begin{cases} dX_t = \frac{1}{\varepsilon}(X_t - Y_t^3 - Y_t - s)dt \\ dY_t = (\gamma X_t - Y_t + \beta)dt, \end{cases}$$

where the variable X_t represents the membrane potential of the neuron at time t, and Y_t is a recovery variable, which could represent channel kinetic.

No noise: **deterministic system**

Examples

UNIVERSITÉ Grenoble Alpes

Example 2: FitzHugh-Nagumo

The behaviour of the neuron is defined through the solution of the system:

$$\begin{cases} dX_t = \frac{1}{\varepsilon}(X_t - Y_t^3 - Y_t - s)dt + \sigma_1 dW_1 \\ dY_t = (\gamma X_t - Y_t + \beta)dt + \sigma_2 dW_2, \end{cases}$$

where the variable X_t represents the membrane potential of the neuron at time t, and Y_t is a recovery variable, which could represent channel kinetic.

Noise in both coordinates: **elliptic system**

Example 2: FitzHugh-Nagumo

The behaviour of the neuron is defined through the solution of the system:

$$\begin{cases} dX_t = \frac{1}{\varepsilon}(X_t - Y_t^3 - Y_t - s)dt \\ dY_t = (\gamma X_t - Y_t + \beta)dt + \sigma dW_t, \end{cases}$$

where the variable X_t represents the membrane potential of the neuron at time t, and Y_t is a recovery variable, which could represent channel kinetic.

Noise in only one coordinate: **hypoelliptic system**

Model

We consider the following SDE:

$$\begin{cases} dX_t = a_1(X_t, Y_t; \theta)dt \\ dY_t = a_2(X_t, Y_t; \theta)dt + \sigma dW_t, \end{cases} \tag{1}$$

or, in vector notations:

$$dZ_t = A(Z_t; \theta)dt + \tilde{\sigma} dW_t, \tag{2}$$

where $Z_t = (X_t, Y_t)$, $A(Z_t; \theta) = (a_1(X_t, Y_t; \theta),\ a_2(X_t, Y_t; \theta))^T$ is a drift, and $\tilde{\sigma} = (0, \sigma)^T$ is a variance term.

Model

We are working under the following assumptions:

A1 $\forall (x, y) \in \mathbb{R}^2 : \quad \partial_y a_1 \neq 0$

A2 *Lipschitz and linear growth conditions.*

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

A4 Both functions $a_1(x, y; \theta)$ and $a_2(x, y; \theta)$ are identifiable, that is $a_i(x, y; \theta) = a_i(x, y; \theta_0) \Leftrightarrow \theta = \theta_0,\ i = 1, 2.$

We assume that both variables are discretely observed at equally spaced periods of time on some finite time interval $[0, T]$, with a vector of observations being $Z_i = (X_i, Y_i)^T$, where Z_i is a value of the process at the time $i\Delta$, $i \in 0 \dots N$.

Objective

The **main objective** of this work is to propose a method of estimation of the unknown parameters from discretely observed data.

Objective

The **main objective** of this work is to propose a method of estimation of the unknown parameters from discretely observed data.

Classical approach:

1. Construct the discrete model of the continuous process.
2. Estimate the parameters using *Maximum Likelihood Estimator* based on the likelihood of the discretized model (or the *pseudo-likelihood* of the continuous model).

Difficulties

UNIVERSITÉ Grenoble Alpes

Euler-Maruyama scheme

Given system (2), its solution is approximated by:

$$Z_{i+1} = \bar{A}_{1.0}(Z_i;\theta) + \bar{B}_{1.0}(Z_i,\sigma)\Xi_i \tag{3}$$

where $\Xi = (\xi_1, \xi_2)^T$ is a two-dimensional standard Gaussian noise, drift term is approximated by

$$\bar{A}_{1.0}(Z_i;\theta) = \begin{pmatrix} X_i + \Delta a_1(X_i, Y_i;\theta) \\ Y_i + \Delta a_2(X_i, Y_i;\theta) \end{pmatrix}$$

and the variance, respectively, by $\bar{B}_{1.0}$ given by:

$$\bar{B}_{1.0}(Z_i;\sigma) = \begin{pmatrix} 0 & 0 \\ 0 & \sigma\sqrt{\Delta} \end{pmatrix}$$

Related works

We begin with the class of stochastic Damping Hamiltonian systems. They are defined as the solution of the following SDE:

$$\begin{cases} dX_t = Y_t dt \\ dY_t = a_2(X_t, Y_t; \theta)dt + \sigma dW_t. \end{cases} \tag{4}$$

Examples:

Stochastic Van der Pol oscillator, Harmonic oscillator etc.

Related works

First possible solution [Samson, Thieullen 2012]:

- Use Euler scheme, but estimate the parameters using only one coordinate.

Related works

First possible solution [Samson, Thieullen 2012]:

- Use Euler scheme, but estimate the parameters using only one coordinate.

The following contrast function is introduced:

$$\mathcal{L}_E = \sum_{i=0}^{N-1} \left(\frac{(Y_{i+1} - Y_i - \Delta a_2(X_i, Y_i; \theta))^2}{\Delta \sigma^2} + \log \sigma^2 \right).$$

Then the optimal value of parameters can be found as:

$$(\hat{\theta}_E, \hat{\sigma}_E^2) = \arg\min_{\theta, \sigma^2} \mathcal{L}_E(\theta, \sigma^2; Z_{0:N}).$$

First possible solution [Samson, Thieullen 2012]:

- Use Euler scheme, but estimate the parameters using only one coordinate.

The following contrast function is introduced:

$$\mathcal{L}_E = \sum_{i=0}^{N-1} \left(\frac{(Y_{i+1} - Y_i - \Delta a_2(X_i, Y_i; \theta))^2}{\Delta \sigma^2} + \log \sigma^2 \right).$$

Then the optimal value of parameters can be found as:

$$(\hat{\theta}_E, \hat{\sigma}^2_E) = \arg\min_{\theta, \sigma^2} \mathcal{L}_E(\theta, \sigma^2; Z_{0:N}).$$

Euler contrast is **consistent** and **asymptotically normal** under conditions $N\Delta \to \infty$ and $N\Delta^2 \to 0$.

Alternative solution [Pokern et. al 2007]:
Approximate the drift term as in Euler scheme ($\bar{A}_P \equiv \bar{A}_{1.0}$), but introduce noise in the first coordinate with higher order, with variance and covariance matrix given respectively by:

$$\bar{B}_P = \sigma \begin{pmatrix} \frac{1}{12}\Delta^{\frac{3}{2}} & \frac{1}{2}\Delta^{\frac{3}{2}} \\ 0 & \sqrt{\Delta} \end{pmatrix} \quad \bar{B}_P \bar{B}_P^T = \sigma^2 \begin{pmatrix} \frac{\Delta^3}{3} & \frac{\Delta^2}{2} \\ \frac{\Delta^2}{2} & \Delta \end{pmatrix}$$

Alternative solution [Pokern et. al 2007]:
Approximate the drift term as in Euler scheme ($\bar{A}_P \equiv \bar{A}_{1.0}$), but introduce noise in the first coordinate with higher order, with variance and covariance matrix given respectively by:

$$\bar{B}_P = \sigma \begin{pmatrix} \frac{1}{12}\Delta^{\frac{3}{2}} & \frac{1}{2}\Delta^{\frac{3}{2}} \\ 0 & \sqrt{\Delta} \end{pmatrix} \quad \bar{B}_P \bar{B}_P^T = \sigma^2 \begin{pmatrix} \frac{\Delta^3}{3} & \frac{\Delta^2}{2} \\ \frac{\Delta^2}{2} & \Delta \end{pmatrix}$$

Covariance matrix is no longer singular! We can estimate the parameters in the following way:

$$\hat{\theta}_P = \arg\min_{\theta} \frac{1}{2(N-1)} \sum_{i=0}^{N-1} \|\bar{B}_P^{-1}(Z_i - \bar{A}_P(Z_i;\theta))\|^2,$$

$$\hat{\sigma}_P^2 = \frac{1}{2(N-1)} \sum_{i=0}^{N-1} \|\bar{B}_P^{-1}(Z_i - \bar{A}_P(Z_i;\theta))\|^2.$$

Related works

Drawbacks of the described schemes:

- Order of the approximation is not high enough to ensure accurate estimation

Related works

Drawbacks of the described schemes:

- Order of the approximation is not high enough to ensure accurate estimation
- Applicable only to a very limited number of models (with $a_1(X_t, Y_t; \theta) \equiv Y_t$)...

Related works

Drawbacks of the described schemes:

- Order of the approximation is not high enough to ensure accurate estimation
- Applicable only to a very limited number of models (with $a_1(X_t, Y_t; \theta) \equiv Y_t$)...
- ... and sometimes we need to transform the data (bring system (2) to (4)).

Related works

For systems (2) the only available solution is **1.5 strong order scheme** [Samson and Ditlevsen, work in progress] .

For systems (2) the only available solution is **1.5 strong order scheme** [Samson and Ditlevsen, work in progress] .

Drift is approximated by:

$$\bar{A}_{1.5}(Z_i;\theta) = \begin{pmatrix} X_i + \Delta a_1(X_i, Y_i; \theta_1) + \frac{\Delta^2}{2}\left[\frac{\partial a_1}{\partial X_i} a_1 + \frac{\partial a_1}{\partial Y_i} a_2\right] \\ Y_i + \Delta a_2(X_i, Y_i; \theta_2) + \frac{\Delta^2}{2}\left[\frac{\partial a_2}{\partial X_i} a_1 + \frac{\partial a_2}{\partial Y_i} a_2\right] \end{pmatrix}$$

and the covariance matrix is given by:

$$\Sigma_{1.5}(Z_i;\theta,\sigma) = \sigma^2 \begin{pmatrix} (\partial_{Y_i} a_1)^2 \frac{\Delta^3}{3} & \partial_{Y_i} a_1 \frac{\Delta^2}{2} \\ \partial_{Y_i} a_1 \frac{\Delta^2}{2} & \Delta \end{pmatrix}$$

Related works

In this work were studied the following estimators:

$$\mathbb{C}_{\theta_1} = \frac{3}{\Delta^3}\sum_{i=0}^{N-1}\frac{\left(X_{i+1}-\bar{A}_{(1),1.5}(Z_i;\theta_1,\theta_2)\right)^2}{\left(\partial_{Y_i}a_1\right)^2\sigma^2} + (N-1)\log((\partial_{Y_i}a_1)^2\,\sigma^2)$$

$$\mathbb{C}_{\theta_2} = \sum_{i=0}^{N-1}\left[\frac{\left(Y_{i+1}-\bar{A}_{(2),1.5}(Z_i;\theta_1,\theta_2)\right)^2}{\Delta\sigma^2} + \log\sigma^2\right],$$

where $\bar{A}_{(1),1.5}$ and $\bar{A}_{(2),1.5}$ denote first and the second coordinate in the vector $\bar{A}_{1.5}$ respectively.

Related works

It was proven that the estimator is consistent under the following conditions:

1. If $\Delta \to 0,\ N\Delta \to \infty$, then

$$\hat{\theta}_{1,N} \xrightarrow{P} \theta_{1,0}$$

2. If $\Delta \to 0,\ N\Delta \to \infty$, then

$$(\hat{\theta}_{2,N}, \hat{\sigma}_N^2) \xrightarrow{P} (\theta_{2,0},\ \sigma_0^2),$$

where $\theta_{1,0}$ and $\theta_{2,0}$ are real values of the parameters.

What we propose?

Our idea (inspired by work [Ozaki 1985]):

- Approximate the non-linear system by a piece-wise linear SDE, for which the close-form solution is known

What we propose?

Our idea (inspired by work [Ozaki 1985]):

- Approximate the non-linear system by a piece-wise linear SDE, for which the close-form solution is known
- Discretize this solution on each small interval with desired accuracy

What we propose?

Our idea (inspired by work [Ozaki 1985]):

- Approximate the non-linear system by a piece-wise linear SDE, for which the close-form solution is known
- Discretize this solution on each small interval with desired accuracy
- Construct the estimator using the discretized process

Discretization scheme

Main assumption: $\frac{\partial A(Z_t;\theta)}{\partial Z_t} \triangleq J_i \approx const$, $\forall t \in [i\Delta, (i+1)\Delta)$.

Discretization scheme

UNIVERSITÉ Grenoble Alpes

Main assumption: $\frac{\partial A(Z_t;\theta)}{\partial Z_t} \triangleq J_i \approx const$, $\forall t \in [i\Delta, (i+1)\Delta)$.
Then the solution on each interval is given by:

$$Z_i e^{J_i \Delta} + \int_{i\Delta}^{(i+1)\Delta} e^{J_i((i+1)\Delta - s)} \tilde{\sigma} dW_s \tag{5}$$

Discretization scheme

Main assumption: $\frac{\partial A(Z_t;\theta)}{\partial Z_t} \triangleq J_i \approx const$, $\forall t \in [i\Delta, (i+1)\Delta)$.

Then the solution on each interval is given by:

$$Z_i e^{J_i \Delta} + \int_{i\Delta}^{(i+1)\Delta} e^{J_i((i+1)\Delta - s)} \tilde{\sigma} dW_s \tag{5}$$

Then $\mathbb{E}[Z_{i+1}|Z_i] = Z_i e^{J_i \Delta}$ and the covariance matrix is given by:

$$\Sigma_{i+1} = \mathbb{E}\left[\left(\int_{i\Delta}^{(i+1)\Delta} e^{J_i((i+1)\Delta - s)} \tilde{\sigma} dW_s\right)\left(\int_{i\Delta}^{(i+1)\Delta} e^{J_i((i+1)\Delta - s)} \tilde{\sigma} dW_s\right)^T\right] \tag{6}$$

Discretization scheme

.... which approximately corresponds to:

$$Z_i + \Delta A(Z_i;\theta) + \frac{\Delta^2}{2} J_i A(Z_i;\theta) + \mathcal{O}(\Delta^3) \triangleq \bar{A}(Z_i;\theta)$$

and

$$\Sigma_\Delta(Z_i;\theta,\sigma) \triangleq \sigma^2 \begin{pmatrix} (\partial_{Y_i} a_1)^2 \frac{\Delta^3}{3} & \partial_{Y_i} a_1 \frac{\Delta^2}{2} \\ \partial_{Y_i} a_1 \frac{\Delta^2}{2} & \Delta \end{pmatrix}$$

respectively.

Discretization scheme

Strong sides of our scheme:

+ Covariance matrix is invertible —> now we can apply the same approach as for elliptic equations.
+ Due to higher-order terms it is more stable than the Euler scheme.

Discretization scheme

Strong sides of our scheme:

+ Covariance matrix is invertible —> now we can apply the same approach as for elliptic equations.
+ Due to higher-order terms it is more stable than the Euler scheme.

Weak points:

- Drift and variance terms now depend on the same set of parameters.
- Variance of each coordinate is of different order —> hard to study.
- Covariance matrix is highly ill-conditioned $\det(\Sigma_\Delta) \approx \mathcal{O}(\Delta^4)$.

Estimator

Then we introduce a contrast function, which we can intuitively define as -2 times pseudo-likelihood:

$$\mathcal{L}(\theta, \sigma^2; Z_{0:N}) =$$

$$= \sum_{i=0}^{N-1} (Z_{i+1} - \bar{A}(Z_i; \theta))^T \Sigma_\Delta^{-1}(Z_i; \theta, \sigma^2)(Z_{i+1} - \bar{A}(Z_i; \theta))$$

$$+ \sum_{i=0}^{N-1} \log\det(\Sigma_\Delta(Z_i; \theta, \sigma^2)).$$

Estimator

Then we introduce a contrast function, which we can intuitively define as -2 times pseudo-likelihood:

$$\mathcal{L}(\theta, \sigma^2; Z_{0:N}) =$$
$$= \sum_{i=0}^{N-1} (Z_{i+1} - \bar{A}(Z_i;\theta))^T \Sigma_\Delta^{-1}(Z_i;\theta,\sigma^2)(Z_{i+1} - \bar{A}(Z_i;\theta))$$
$$+ \sum_{i=0}^{N-1} \log\det(\Sigma_\Delta(Z_i;\theta,\sigma^2)).$$

ATTENTION: it is biased because of the covariance matrix! σ^2 cannot be estimated correctly.

Estimator

We prove that the correct contrast function is defined by:

$$\mathcal{L}(\theta, \sigma^2; Z_{0:N}) =$$
$$= \mathbf{\frac{1}{2}} \sum_{i=0}^{N-1} (Z_{i+1} - \bar{A}(Z_i;\theta))^T \Sigma_\Delta^{-1}(Z_i;\theta,\sigma^2)(Z_{i+1} - \bar{A}(Z_i;\theta))$$
$$+ \sum_{i=0}^{N-1} \log\det(\Sigma_\Delta(Z_i;\theta,\sigma^2)).$$

Estimator

We prove that the correct contrast function is defined by:

$$\mathcal{L}(\theta, \sigma^2; Z_{0:N}) =$$
$$= \frac{1}{2} \sum_{i=0}^{N-1} (Z_{i+1} - \bar{A}(Z_i;\theta))^T \Sigma_\Delta^{-1}(Z_i;\theta,\sigma^2)(Z_{i+1} - \bar{A}(Z_i;\theta))$$
$$+ \sum_{i=0}^{N-1} \log\det(\Sigma_\Delta(Z_i;\theta,\sigma^2)).$$

Then the optimal values of parameters may be estimated as:

$$(\hat{\theta}, \hat{\sigma}^2) = \underset{\theta,\sigma^2}{\arg\min}\, \mathcal{L}(\theta,\sigma^2; Z_{0:N}).$$

Estimator

Let us denote the parameters in the **first coordinate** by φ, and in the **second** — by ψ.

Theorem

Under assumptions (A1)-(A4) and $\Delta_N \to 0$ and $N\Delta_N \to \infty$ the following holds:

$$\hat{\varphi}_{N,\Delta_N} \xrightarrow{\mathbb{P}_\theta} \varphi_0, \ \textit{given } \psi_0$$
$$\hat{\psi}_{N,\Delta_N} \xrightarrow{\mathbb{P}_\theta} \psi_0, \ \textit{given } \varphi_0$$
$$\hat{\sigma}^2_{N,\Delta_N} \xrightarrow{\mathbb{P}_\theta} \sigma^2_0, \ \textit{given } \varphi_0$$

Estimator

Issues with the proof:

1. It was hard to consider the whole criteria
2. For φ:

$$\frac{\Delta}{N}\left[\mathcal{L}_{N,\Delta}(\varphi,\psi_0,\sigma^2;Z_{0:N})-\mathcal{L}_{N,\Delta}(\varphi_0,\psi_0,\sigma^2;Z_{0:N})\right]$$

3. For ψ:

$$\frac{1}{N\Delta}\left[\mathcal{L}_{N,\Delta}(\varphi_0,\psi,\sigma^2;Z_{0:N})-\mathcal{L}_{N,\Delta}(\varphi_0,\psi_0,\sigma^2;Z_{0:N})\right]$$

Estimator

Strong sides of the estimator:

+ We do not need to transform the system
+ It is possible to estimate all the parameters of the system...

Estimator

Strong sides of the estimator:

+ We do not need to transform the system
+ It is possible to estimate all the parameters of the system...

Weak points:

- ...theoretically.

Simulation study

Step-by-step:

1. Generate 100 trajectories with Local Linearization scheme for each set of parameters with $\Delta = 0.001$ and $N = 1000000$.
2. Subsample data with bigger time step (specified in the table).
3. Use **Conjugate Gradient Descent** method in order to minimize the contrast.
4. Take the median of the estimated values as a result.

All experiments are conducted in **R**.

Simulation study: FitzHugh-Nagumo

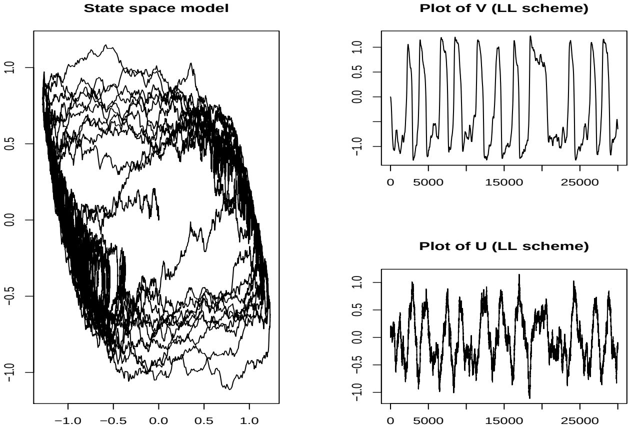

Figure 2: Excitatory: $\sigma = 0.6,\ \gamma = 1.5,\ \beta = 0.3,\ \varepsilon = 0.1,\ s = 0.01$

Simulation study: FitzHugh-Nagumo

UNIVERSITÉ Grenoble Alpes

N	1000	1000	10000	10000
Δ	0.1	0.01	0.01	0.001
$\beta_0 = 0.3$	0.216	0.361	0.305	0.462
	0.007	0.033	0.002	0.071
$\gamma_0 = 1.5$	1.196	1.613	1.526	1.610
	0.092	0.031	0.003	0.050
$\sigma_0 = 0.6$	0.628	0.666	0.606	0.623
	0.001	0.006	0.000	0.001
$\varepsilon_0 = 0.1$	603.78	1334.45	1530.76	1705.80

Table 3: LL scheme: excitatory behaviour

First line: *median*, ***second line***: *standard deviation*

Simulation study: FitzHugh-Nagumo

N	1000	1000	10000	10000
Δ	0.1	0.01	0.01	0.001
$\beta_0 = 0.3$	0.230	0.329	0.308	0.452
	0.008	0.051	0.004	0.089
$\gamma_0 = 1.5$	1.216	1.569	1.514	1.707
	0.084	0.057	0.006	0.015
$\sigma_0 = 0.6$	0.614	0.598	0.600	0.600
	0.000	0.000	0.000	0.000

Table 4: LL scheme, ε is fixed: excitatory behaviour

***First line**: median, **second line**: standard deviation*

Simulation study: FitzHugh-Nagumo

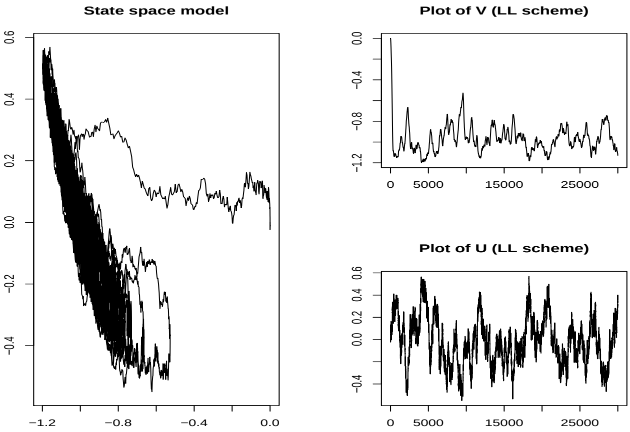

Figure 5: Oscillatory: $\sigma = 0.4,\ \gamma = 1.2,\ \beta = 1.3,\ \varepsilon = 0.1,\ s = 0.01$

Simulation study: FitzHugh-Nagumo

N	1000	1000	10000	10000
Δ	0.1	0.01	0.01	0.001
$\beta_0 = 1.3$	1.687	2.015	1.762	1.856
	0.168	0.543	0.213	0.309
$\gamma_0 = 1.2$	1.582	1.882	1.641	1.652
	0.148	0.490	0.195	0.204
$\sigma_0 = 0.4$	0.407	0.415	0.405	0.407
	0.000	0.000	0.000	0.000
$\varepsilon_0 = 0.1$	424.22	867.31	862.54	905.12

Table 6: LL scheme: oscillatory behaviour

First line: *median*, ***second line***: *standard deviation*

Simulation study: FitzHugh-Nagumo

UNIVERSITÉ Grenoble Alpes

N	1000	1000	10000	10000
Δ	0.1	0.01	0.01	0.001
$\beta_0 = 1.3$	1.481	1.674	1.638	1.655
	0.097	0.189	0.167	0.197
$\gamma_0 = 1.2$	1.385	1.568	1.527	1.566
	0.094	0.180	0.158	0.201
$\sigma_0 = 0.4$	0.400	0.401	0.400	0.400
	0.000	0.000	0.000	0.000

Table 7: LL scheme: inhibitory behaviour

First line: *median*, ***second line***: *standard deviation*

Simulation study

Main problems:

- Straightforward two-dimensional implementation did not work for small Δ due to the almost singular covariance matrix.
- ε is not estimated!
- Sensitive to noise (especially to the ratio $\frac{\sigma}{\varepsilon}$).
- Scheme is not stable: θ_{init} must be close to θ_0, many outliers.

Simulation study

UNIVERSITÉ Grenoble Alpes

Main problems:

- Straightforward two-dimensional implementation did not work for small Δ due to the almost singular covariance matrix.
- ε is not estimated!
- Sensitive to noise (especially to the ratio $\frac{\sigma}{\varepsilon}$).
- Scheme is not stable: θ_{init} must be close to θ_0, many outliers.

What is positive:

- We could estimate the parameters of the rough variables without fixing anything
- It works even for relatively big time step and short observation interval

Simulation study

UNIVERSITÉ Grenoble Alpes

Example 3: Introduction

In the large network of neurons spike occurrences can be described by a **multivariate Hawkes process** ($\approx$ "Poisson process with a memory"). In the next example we consider the stochastic approximation of the multivariate Hawkes process. Result is due to [Ditlevsen, Löcherbach 2015].

Example 3: Introduction

In the large network of neurons spike occurrences can be described by a **multivariate Hawkes process** ($\approx$ "Poisson process with a memory"). In the next example we consider the stochastic approximation of the multivariate Hawkes process. Result is due to [Ditlevsen, Löcherbach 2015].

We deal with:

- 2 population of neurons: inhibitory and excitatory
- Only 1 neuron in each population

$\Longrightarrow$ 4-dimensional equation

Simulation study

Example 3: Stochastic Approximation of Hawkes process

Oscillatory behaviour of the system is described by:

$$\begin{cases} X_1 = (-\nu_1 X_1 + X_2)dt \\ X_2 = (-\nu_1 X_2 + c_1 f_2(X_3))dt + c_1\sqrt{2 f_2(X_3)}dW_1 \\ X_3 = (-\nu_2 X_3 + X_4)dt \\ X_4 = (-\nu_2 X_4 + c_2 f_1(X_1))dt + c_2\sqrt{2 f_1(X_1)}dW_2, \end{cases}$$

where

$$f_1(x) = \begin{cases} 10e^x & x < \log(20) \\ \frac{400}{1+400e^{-2x}} & x \geq \log(20) \end{cases};$$

$$f_2(x) = \begin{cases} e^x & x < \log(20) \\ \frac{40}{1+400e^{-2x}} & x \geq \log(20) \end{cases}.$$

Simulation study

Drift and the variance are driven by the same set of parameters.
Is it possible to estimate them using only the drift?

Drift and the variance are driven by the same set of parameters.
Is it possible to estimate them using only the drift?

Let us denote:

$$\mathcal{L}_p(\theta; Z_{0:N}) = \sum_{i=1}^{N-1} (Z_{i+1} - \bar{A}(Z_i;\theta))(Z_{i+1} - \bar{A}(Z_i;\theta))^T,$$

where $\bar{A}(Z_i;\theta)$ is an approximation of the drift.

Simulation study

N	1000	1000	10000	10000
Δ	0.1	0.01	0.01	0.001
$c_{1,0} = -1$	-1.046	-0.982	-0.998	-1.050
	0.008	0.004	0.005	0.006
$c_{2,0} = 1$	1.056	1.036	0.983	1.101
	0.018	0.032	0.007	0.040
$\nu_{1,0} = 0.8$	0.801	0.788	0.808	0.805
	0.001	0.000	0.000	0.000
$\nu_{2,0} = 1.2$	1.299	1.288	1.280	1.382
	0.008	0.005	0.007	0.014

Table 8: Stochastic approximation of the Hawkes process

First line*: mean,* ***second line****: standard deviation*

Prospective work

Q: *Is there anything we can use for the variance independent estimation?*

A: For linear homogeneous equations we can use Girsanov formula and construct a consistent estimator [Alain Le Breton and Musiela, 1985] without taking into account variance:

$$\hat{\mathcal{L}}_{MLE} = \left[\int_0^T dZ_t Z_t^T\right]\left[\int_0^T Z_t Z_t^T dt\right]^{-1}$$

Prospective work

Further...

When we work with **real data**, we don't know:

- how many variables are driven by noise,
- what are the functions included in the model,
- how many equations?

Prospective work

Further...

When we work with **real data**, we don't know:

- how many variables are driven by noise,
- what are the functions included in the model,
- how many equations?

To be done: Statistical tests

Prospective work

Further...

When we work with **real data**, we don't know:

- how many variables are driven by noise,
- what are the functions included in the model,
- how many equations?

To be done: Statistical tests

Moreover:

- Problem of noisy data
- Problem of **partial observations**

Prospective work

Further...

When we work with **real data**, we don't know:

- how many variables are driven by noise,
- what are the functions included in the model,
- how many equations?

To be done: Statistical tests

Moreover:

- Problem of noisy data
- Problem of **partial observations**

To be done: Filtering

Thank you for your attention!