Statistical testing of the covariance matrix rank in multidimensional neuronal models: non-asymptotic case

Anna Melnykova

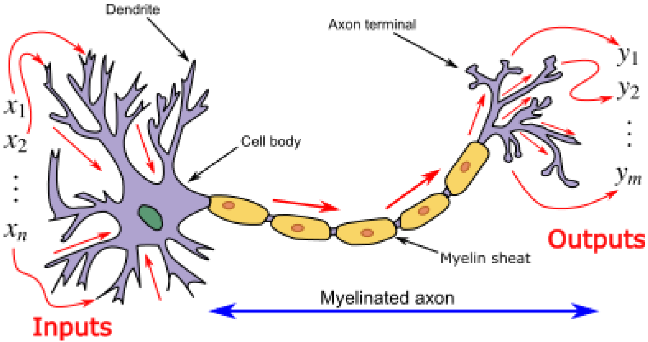
joint work with: Patricia Reynaud-Bouret, Adeline Samson

Institut National Polytechnique de Grenoble

Séminaire d'equipe Statistique et Optimisation, IMT Toulouse

30 March 2021

Motivation



Dendrite
Axon terminal
Cell body
Myelin sheat
Myelinated axon
Inputs
Outputs
x_1
x_2
x_n
y_1
y_2
y_m

Example 1: Hodgkin-Huxley model

Conductance-based model of **action potential in neurons**:

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

- I — membrane potential
- n, m, h — quantities between 0 and 1 that are associated with potassium channel activation, sodium channel activation, and sodium channel inactivation.

References: Hodgkin and Huxley (1952) — *1963 Nobel Prize in Physiology or Medicine*,

Example 2: FitzHugh-Nagumo model

The behaviour of the neuron is defined through:

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

- X_t — membrane potential
- Y_t — recovery variable
- s — magnitude of the stimulus current
-

References: Fitzhugh (1961), Nagumo et al. (1962)

Example 2: FitzHugh-Nagumo model

The behaviour of the neuron is defined through:

$$\begin{cases} dX_t = \frac{1}{\varepsilon}(X_t - X_t^3 - Y_t - s)dt + \sigma_1 dW_t^1 \\ dY_t = (\gamma X_t - Y_t + \beta)dt + \sigma_2 dW_t^2 \end{cases}$$

- X_t — membrane potential
- Y_t — recovery variable
- s — magnitude of the stimulus current
- $\sigma_1, \sigma_2 \geq 0$ — diffusion coefficients (possibly null)

References: Fitzhugh (1961), Nagumo et al. (1962)

FitzHugh-Nagumo model

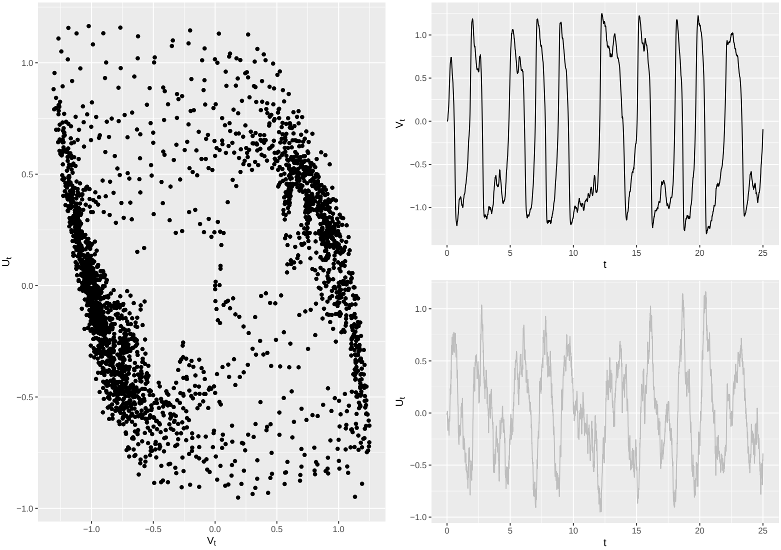


Figure: Neuronal activity simulated with stochastic FitzHugh-Nagumo model.

Where to put noise?

Main challenges:

- Highly non-linear systems
- Computational cost
- Measurements inaccuracy

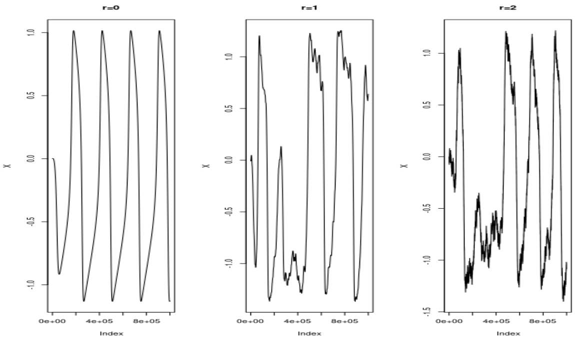


Figure: Membrane potential simulated wiht a FitzHugh-Nagumo model: deterministic, hypoelliptic, elliptic system

References: see Tuckwell (2005) for a general overview of neuronal models

Formalization

Given:

Discrete observations of the d-dimensional process X **with a fixed time step Δ**

$$dX_t = b_t dt + \sigma_t dW_t,\ t \in [0, T], \tag{1}$$

$b_t \in \mathbb{R}^d, \sigma_t \in \mathbb{R}^{d\times q}$.

Goal:

Propose a test

$$H_0 : rank(\Sigma) = r$$
$$H_1 : rank(\Sigma) \leq r,$$

where $\Sigma = \sigma_t \sigma_t^T$. If σ_t is not constant, we search $\sup_t \text{rank}(\Sigma)$ instead.

Ellipticity vs Hypoellipticity

Definition (Ellipticity)

We say that the system (1) is *elliptic*, if its covariance matrix is of full rank ($rank(\Sigma) = d$).

Definition (Hypoellipticity)

We say that the system (1) is *hypoelliptic*, if its covariance matrix is **not** of full rank ($rank(\Sigma) < d$), **but** the process X has a smooth transition density. It can be verified by Hörmander condition.

Remark

In this talk, when referring to "hypoelliptic" systems we will mean systems where some coordinates are not perturbed by the Brownian motion or some diffusion coefficients are negligibly small with respect to a given discretization step Δ.

What do we want to study?

Given a d-dimensional process X $i = 1, \ldots, N$, the main statistics is defined as

$$S = \frac{1}{n} \sum_{i=1}^{n_{inc}} \det \Phi_i^2,$$

where:

$$\Phi_i := \begin{pmatrix} \frac{X^{(1)}_{id+1} - X^{(1)}_{id}}{\sqrt{\Delta}} & \frac{X^{(1)}_{id+2} - X^{(1)}_{id+1}}{\sqrt{\Delta}} & \ldots & \frac{X^{(1)}_{id+d} - X^{(1)}_{id+d-1}}{\sqrt{\Delta}} \\ \frac{X^{(2)}_{id+1} - X^{(2)}_{id}}{\sqrt{\Delta}} & \frac{X^{(2)}_{id+2} - X^{(2)}_{id+1}}{\sqrt{\Delta}} & \ldots & \frac{X^{(2)}_{id+2d} - X^{(2)}_{id+d-1}}{\sqrt{\Delta}} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{X^{(d)}_{id+1} - X^{(d)}_{id}}{\sqrt{\Delta}} & \frac{X^{(d)}_{id+2} - X^{(d)}_{id+1}}{\sqrt{\Delta}} & \ldots & \frac{X^{(d)}_{id+d} - X^{(d)}_{id+d-1}}{\sqrt{\Delta}} \end{pmatrix}^2 \quad (2)$$

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

Why this statistics?

Main statistics is defined as

$$S = \frac{1}{n} \sum_{i=1}^{n_{inc}} \det \Phi_i^2.$$

- Under certain assumptions, the entries are independent. It is simpler to derive the probabilistic properties of the process.
- It can be profitable to study the distribution of rows.
- It has the following important property:

$$\det \Phi_i^2 \approx O(\Delta^{d-r_0}),$$

where r_0 is the rank of Σ.

Determinant expansion (Jacod and Podolskij, 2013)

If we use a very rough approximation (for example, Euler-Maruyama), we can write the matrix Φ_i as follows:

$$\Phi_i \approx A_i + \sqrt{\Delta} B_i,$$

where A_i is a matrix constituted of the increments of **the diffusion term**, and B_i is a matrix of increments of **the drift term**. Then,

$$\det\left(A_i + \sqrt{\Delta} B_i\right) = \det A_i + \sqrt{\Delta}\gamma_{d-1}(A_i, B_i) + \cdots + (\sqrt{\Delta})^d \det B_i,$$

where $\gamma_k(A_i, B_i)$ denotes a sum of determinants of all possible matrices created from k columns of A_i and $d - k$ columns of B_i (without permutation).

Determinant expansion: deterministic 3d-example

$$\det(E + hD) =$$

$$\det\begin{pmatrix} e_{11}+hd_{11} & e_{12}+hd_{12} & e_{13}+hd_{13} \\ e_{21}+hd_{21} & e_{22}+hd_{22} & e_{23}+hd_{23} \\ e_{31}+hd_{31} & e_{32}+hd_{32} & e_{33}+hd_{33} \end{pmatrix} = \det\begin{pmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{pmatrix} +$$

$$\overbrace{h\det\begin{pmatrix} e_{11} & e_{12} & d_{13} \\ e_{21} & e_{22} & d_{23} \\ e_{31} & e_{32} & d_{33} \end{pmatrix} + h\det\begin{pmatrix} e_{11} & d_{12} & e_{13} \\ e_{21} & d_{22} & e_{23} \\ e_{31} & d_{32} & e_{33} \end{pmatrix} + h\det\begin{pmatrix} d_{11} & e_{12} & e_{13} \\ d_{21} & e_{22} & e_{23} \\ d_{31} & e_{32} & e_{33} \end{pmatrix}}^{h\gamma_2(E,D)} +$$

$$\overbrace{h^2\det\begin{pmatrix} e_{11} & d_{12} & d_{13} \\ e_{21} & d_{22} & d_{23} \\ e_{31} & d_{32} & d_{33} \end{pmatrix} + h^2\det\begin{pmatrix} d_{11} & d_{12} & e_{13} \\ d_{21} & d_{22} & e_{23} \\ d_{31} & d_{32} & e_{33} \end{pmatrix} + h^2\det\begin{pmatrix} d_{11} & e_{12} & d_{13} \\ d_{21} & e_{22} & d_{23} \\ d_{31} & e_{32} & d_{33} \end{pmatrix}}^{h^2\gamma_1(E,D)} +$$

$$h^3\det\begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{pmatrix}$$

Determinant expansion (Jacod and Podolskij, 2013)

What this result means:

- $rank(\Sigma) = r_0$ means that $d - r_0$ variables are not perturbed by noise. In this case, the matrix A_i will also have a rank r_0.
- Asymptotically, when $\Delta \to 0$, the structure of drift plays no role in the accuracy of the test. Only increments of the Brownian motion matter.

Naive approach: Jacod et al. (2008)

- Compute matrices of increments (6)
- Define a converging sequence $\{\rho_N\}$ of "thresholds", s.t.:

$$\rho_N \to 0,\ \rho_N\sqrt{N} \to \infty$$

- Define by $\mathcal{A}_r$ class of all subsets of $\{1, \ldots d\}$ with r elements. Define $\det_K \Sigma$ the determinant of $r \times r$ submatrix of $\Sigma^{kl},\ k, l \in K$. Finally, define:

$$\det(r; \Sigma) = \sum_{K \in \mathcal{A}_r} \det_K \Sigma \tag{3}$$

- $\widehat{R}_{T,\Delta} = \inf\left\{r \in \{0, \ldots, d-1\} : \frac{1}{\Delta(r+1)!} \sum_{i=1}^{N-r+1} \det(r+1; \Phi_i) < \rho_N T\right\}$

Remark

Note that $\det(1; \Sigma) = Tr(\Sigma)$, *and* $\det(d; \Sigma) = \det(\Sigma)$.

Remarks to the method of Jacod et al. (2008)

- **Main idea of the method:** we check if the determinant of the submatrix of a given rank is "small enough"
- In practice, **the method is extremely costly and difficult to apply**: to compute the estimator we need to compute all possible minors of dimension from 1 to r and then compare them to a manually tuned threshold.
- The last problem can be omitted if we compare the determinant of a matrix to some threshold directly (since we know that it must is of order Δ^{d-r_0}).
- Disregarding of a threshold, the method will perform poorly in a non-asymptotic setting.

Random perturbation approach

Main reference: Jacod and Podolskij (2013)

Given a d-dimensional diffusion process X, consider 2 new processes:

$$\tilde{X}_t^{(k)} = X_t + \sqrt{k\Delta}\tilde{\sigma}\tilde{W}_t,$$

where $k = 1, 2$, and $\tilde{\sigma}$ is such that $\tilde{\sigma}\tilde{\sigma}^T =: \tilde{\Sigma}$ is a non-random matrix of full rank.

$$S^k = 2d\Delta \sum_{i=0}^{N-1} \det \begin{pmatrix} \frac{\tilde{X}^{1,(k)}_{2id+k} - \tilde{X}^{1,(k)}_{2id}}{\sqrt{k\Delta}} & \frac{\tilde{X}^{1,(k)}_{2id+2k} - \tilde{X}^{1,(k)}_{2id+k}}{\sqrt{k\Delta}} & \dots & \frac{\tilde{X}^{1,(k)}_{2id+2d} - \tilde{X}^{1,(k)}_{2id+kd-k}}{\sqrt{k\Delta}} \\ \frac{\tilde{X}^{2,(k)}_{2id+k} - \tilde{X}^{2,(k)}_{2id}}{\sqrt{k\Delta}} & \frac{\tilde{X}^{2,(k)}_{2id+2k} - \tilde{X}^{2,(k)}_{2id+k}}{\sqrt{k\Delta}} & \dots & \frac{\tilde{X}^{2,(k)}_{2id+2d} - \tilde{X}^{2,(k)}_{2id+kd-k}}{\sqrt{k\Delta}} \\ \dots & \dots & \dots & \dots \\ \frac{\tilde{X}^{d,(k)}_{2id+k} - \tilde{X}^{d,(k)}_{2id}}{\sqrt{k\Delta}} & \frac{\tilde{X}^{d,(k)}_{2id+2k} - \tilde{X}^{d,(k)}_{2id+k}}{\sqrt{k\Delta}} & \dots & \frac{\tilde{X}^{d,(k)}_{2id+2d} - \tilde{X}^{d,(k)}_{2id+kd-k}}{\sqrt{k\Delta}} \end{pmatrix}^2$$

Random perturbation approach

Define:

$$\hat{R}_{T,\Delta} = d - \frac{\log \frac{S^2}{S^1}}{\log 2}$$

$$V_{T,\Delta} := Var\left[\widehat{R}_{T,\Delta}\right] = \frac{\left(\frac{E[S_T^1]}{E[S_T^2]}\right)^2 Var[S_T^2] - 2\frac{E[S_T^1]}{E[S_T^2]} Cov[S_T^1 S_T^2] + Var[S_T^1]}{(E[S_T^1]\log 2)^2}.$$

Jacod and Podolskij (2013)

$$\frac{\widehat{R}_{T,\Delta} - r_0}{\sqrt{\Delta V_{T,\Delta}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \text{ as } \Delta \to 0$$

How does it work?

Main idea: compute how "different" are matrices $\Phi_i^{(1)}$ and $\Phi_i^{(2)}$ (matrices of increments constructed with a different discretization step).

Remember that:

$$\frac{\det(E + 2hD)}{\det(E + hD)} \approx 2^{d-r},$$

for any matrices E and D.

Because:

$$\det(E + hD) = \det E + h\gamma_{d-1}(E, D) + \cdots + h^d \det D,$$

where γ_r, $r = 1, \ldots, d$ stands for a sum of determinants of all possible matrices, whose r columns are equal to r columns of a matrix E (with the same indexes), and the remaining $d - r$ — to the corresponding columns of a matrix D.

How does it work: toy 1d example

Take the process:

$$dX_t = adt + \sigma dW_t$$

Add a random perturbation:

$$\tilde{X}_t^{(1)} = adt + \sigma dW_t + \sqrt{\Delta}\tilde{\sigma}\tilde{W}_t,$$
$$\tilde{X}_t^{(2)} = adt + \sigma dW_t + \sqrt{2\Delta}\tilde{\sigma}\tilde{W}_t$$

Using the first-order approximation, compute:

$$\mathbb{E}\left[\left(\frac{\tilde{X}_{i+1}^{(k)} - \tilde{X}_i^{(k)}}{\sqrt{k\Delta}}\right)^2\right] = \sigma^2 + k\Delta a + k\Delta\tilde{\sigma} =: s_i^k$$

Notice that

$$\frac{S^2}{S^1} \underset{\Delta\to 0}{\longrightarrow} \begin{cases} 1 & \text{if } \sigma \neq 0 \\ 2 & \text{if } \sigma = 0 \end{cases}$$

How does it work: toy 1d example

Take the process:

$$dX_t = adt + \sigma dW_t$$

Add a random perturbation:

$$\tilde{X}_t^{(1)} = adt + \sigma dW_t + \sqrt{\Delta}\tilde{\sigma}\tilde{W}_t,$$
$$\tilde{X}_t^{(2)} = adt + \sigma dW_t + \sqrt{2\Delta}\tilde{\sigma}\tilde{W}_t$$

Using the first-order approximation, compute:

$$\mathbb{E}\left[\left(\frac{\tilde{X}_{i+1}^{(k)} - \tilde{X}_i^{(k)}}{\sqrt{k\Delta}}\right)^2\right] = \sigma^2 + k\Delta a + k\Delta\tilde{\sigma} =: s_i^k$$

Notice that

$$\frac{S^2}{S^1} \xrightarrow[\Delta\to 0]{} \begin{cases} 1 & \text{if } \sigma \neq 0 \\ 2 & \text{if } \sigma = 0 \end{cases} \quad \Rightarrow \quad 1 - \frac{\log \frac{S^2}{S^1}}{\log 2} \xrightarrow[\Delta\to 0]{} \begin{cases} 1 & \text{if } \sigma \neq 0 \\ 0 & \text{if } \sigma = 0 \end{cases}$$

What happens if Δ is fixed?

Take the process:

$$dX_t = bdt + \sigma dW_t$$

Fix: $\Delta = 0.01$, $\sigma = 0.05$, $b = 1$, $\tilde{\sigma} = 0.01$

Add the random perturbation:

$$\tilde{X}_t^{(1)} = bdt + \sigma dW_t + \sqrt{\Delta}\tilde{\sigma}\tilde{W}_t,$$
$$\tilde{X}_t^{(2)} = bdt + \sigma dW_t + \sqrt{2\Delta}\tilde{\sigma}\tilde{W}_t$$

Notice that

$$\mathbb{E}\left[\frac{S^2}{S^1}\right] \approx 1.8 \quad \Rightarrow \quad \mathbb{E}\left[\hat{R}_{T,\Delta}\right] \approx 0.15$$

What happens if Δ is fixed?

Conclusions for the methods from Jacod et al. (2008) and Jacod and Podolskij (2013):

- Method of Jacod et al. (2008) is powerful, but requires a careful *tuning of thresholds.* In addition, it is difficult to quantify a probability of wrongly rejecting ellipticity assumption.
- Method of Jacod and Podolskij (2013) is easy to interpret (two-sided test), but requires *tuning of the perturbation rate* (not relevant when Δ is negligible)
- Both methods perform rather poor when Δ is large.

What happens if Δ is fixed?

Ongoing work

A. M., Adeline Samson, Patricia Reynaud-Bouret

Question 1: What can we actually infer in a non-asymptotic setting?
What do we do: study the distribution of the statistics S in *a non-asymptotic setting.*

Question 2: How can we ameliorate the performance of the test in a non-asymptotic setting?
What do we do: We center the increments around their mean value. Numerically, it implies that the method needs to be coupled with the estimator of the drift.

Structure of the project

Ongoing work

A. M., Adeline Samson, Patricia Reynaud-Bouret

- **First,** we evaluate the distribution of S when it can be done explicitly: in 1- and 2-dimensional case.
- **Second**, we demonstrate how the test works in practice.
- **Finally,** we consider a $d-$dimensional case and evaluate the distribution of the statistics S with concentration inequalities.

1d toy model

Consider a one-dimensional process with constant drift and diffusion coefficients:

$$dX_t = b dt + \sigma dW_t$$

Our aim is to construct the following test:

$$H_0 : \sigma^2 = \delta$$
$$H_1 : \sigma^2 \leq \delta,$$

where δ is a pre-chosen parameter.

1d toy model

The main statistics of the test is given by

$$S = \frac{1}{n}\sum_{i=1}^{n-1}\left(X_{(i+1)\Delta} - X_{i\Delta}\right)^2 = \frac{\sigma^2}{n}\sum_{i=1}^{n-1}\left(\eta_i + \Delta\frac{b}{\sigma}\right)^2,$$

where η_i are i.i.d. distributed standard normal variables. Then,

$$S \sim \frac{\sigma^2}{n}\chi_n^2(\lambda),$$

where $\chi_n^2(\lambda)$ is a chi-squared distributed random variable with a non-centrality parameter λ, defined as follows:

$$\lambda(\sigma) = \frac{\Delta^2 b^2}{n\sigma^2}.$$

1d toy model

Now let us define a α-quantile under H_0. Note that

$$\mathbb{P}\left(S \leq \varepsilon\right) = 1 - Q_{n/2}\left(\sqrt{\lambda(\delta)}, \sqrt{\frac{\varepsilon}{\delta}}\right),$$

where $Q_m(a, b)$ is a Markum Q-function, defined as:

$$Q_m(a, b) = \int_b^{\infty} x \left(\frac{x}{a}\right)^{m-1} \exp\left(-\frac{x^2 + a^2}{2}\right) I_{m-1}(ax)dx, \tag{4}$$

where I_{m-1} is a modified Bessel function of the first kind.
Then, H_0 hypothesis is rejected if $S \geq z_\alpha$, where z_α is such that $Q_{n/2}\left(\sqrt{\lambda(\delta)}, \sqrt{\frac{z_\alpha}{\delta}}\right) = \alpha$.

Some remarks

- **One-dimensional test has a purely mathematical interest**: it shows how S behaves when it is constructed on $1-$ dimensional process. If we observe a discrete trajectory of a continuous process, it is straigtforward to see if it is deterministic or not.
- The non-centrality parameter is of order $O(\Delta^2)$. It means that as $\Delta \to 0$, the law of S transforms in a **standard chi-squared law**. The same effect is attained if we center the statistics S.
- Note that we do not do any approximation for the tail probabilities of S, **quantiles are explicit**. They can be evaluated in most of the statistical packages.

2-dimensional case (drift known)

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

$$dX_t = b_t dt + \sigma dW_t, \tag{5}$$

- $b_t = (b_t^1, b_t^2)^T$ is a drift vector
- $\sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$,
- W is a 2-dimensional Brownian motion.

Non-asymptotic setting

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

Outline of the results for 2-dimensional test

- First, we modify the statistics first proposed by Jacod et al. (2008) to obtain a good **separation rate.**
- Second, we study the distribution of the new modified statistics.
- Then, we explain under which condition the Type I and Type II error can be controlled (i.e., **wrongly rejecting**, or **wrongly failing to reject** the ellipticity assumption)
- Finally, we do a numerical study and explain why the previous methods do not work very well for a fixed Δ.

First step: centering the statistics

We assume the drift b_t to be known. Then, we can write the "centered matrix":

$$\dot{\Phi}_i = \begin{pmatrix} \frac{X^{(1)}_{(2i+2)\Delta} - X^{(1)}_{(2i+1)\Delta} - \int_{(2i+1)\Delta}^{(2i+2)\Delta} b^{(1)}_t \, dt}{\sqrt{\Delta}} & \frac{X^{(1)}_{(2i+1)\Delta} - X^{(1)}_{2i\Delta} - \int_{(2i+1)\Delta}^{(2i+2)\Delta} b^{(1)}_t \, dt}{\sqrt{\Delta}} \\ \frac{X^{(2)}_{(2i+2)\Delta} - X^{(2)}_{(2i+1)\Delta} - \int_{(2i+1)\Delta}^{(2i+2)\Delta} b^{(2)}_t \, dt}{\sqrt{\Delta}} & \frac{X^{(2)}_{(2i+1)\Delta} - X^{(2)}_{2i\Delta} - \int_{(2i+1)\Delta}^{(2i+2)\Delta} b^{(2)}_t \, dt}{\sqrt{\Delta}} \end{pmatrix}$$

What has changed: Now, only the "power" of noise determines the order of statistics!

2-dimensional case (drift known)

The density of the distribution of $\det \dot{\Phi}_i^2$ can be written explicitly.

Proposition

Denote $\dot{s}_i := \det \dot{\Phi}_i^2$. *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 variables with k and $k-1$ degrees of freedom. The same paper provides an analogous result for two chi-squared variables with non-centrality parameter.

Second step: transforming the statistics

Denote $\dot{S} = \frac{1}{n}\sum_{i=1}^{n} \dot{s}$. Recall the result about the determinant expansion from Jacod and Podolskij (2013): when the diffusion matrix is of full rank, then $\forall i$

$$\dot{S} = O(1).$$

Then, if the system is elliptic,

$$\frac{\log \dot{S}}{\log \Delta} = 0.$$

So, the **hypothesis of ellipticity** ($\sigma_1^2\sigma_2^2 = \delta$) will be rejected if $\frac{\log \dot{S}}{\log \Delta}$ is "large enough".

2-dimensional case (drift known)

Proposition

Under H_0 the following bound holds:

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

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:

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

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

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

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

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

Proposition

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

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

How to implement it in practice?

Of course, in practice the drift is not known. However, we can estimate it:

- *Non-parametrically* (for example, with kernel estimates)
- *Parametrically* (Bayesian statistic, contrast estimators etc).

How do we proceed:

- We estimate the parameters of the system with least-square estimates (Kutoyants, 2013)
- We plug-in the estimators in the system and substract the higher-order term in drift approximation
- We build $\dot{S}$ and evaluate it under H_0 and H_1 hypothesis.

Performance on FitzHugh-Nagumo model

Parameters are the following:

- **Elliptic case:** $\varepsilon = 0.1, \beta = 0.3, \gamma = 1.5, s = 0.01, \sigma_1 = 0.1, \sigma_2 = 1$.
- **Hypoelliptic case:** $\varepsilon = 0.1, \beta = 0.3, \gamma = 1.5, s = 0.01, \sigma_1 = 0, \sigma_2 = 1$.

Procedure:

1. Generate 1000 trajectories with $\Delta = 1e - 5$, $T = 10$, using 1.5 strong order scheme (see Kloeden et al. (2003))
2. Subsample the data with a bigger Δ *(fixing $n = 1000$).*
3. Compute test statistics for elliptic and hypoelliptic data
4. Report the results

"Old" setting, based on (Jacod et al., 2008)

Test

We are interested in the following hypotheses:

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

where δ is some chosen "sensitivity" threshold.

Statistics

Consider vectors of increments:

$$\Phi_i = \begin{pmatrix} \frac{X^{(1)}_{(2i+2)\Delta} - X^{(1)}_{(2i+1)\Delta}}{\sqrt{\Delta}} & \frac{X^{(1)}_{(2i+1)\Delta} - X^{(1)}_{2i\Delta}}{\sqrt{\Delta}} \\ \frac{X^{(2)}_{(2i+2)\Delta} - X^{(2)}_{(2i+1)\Delta}}{\sqrt{\Delta}} & \frac{X^{(2)}_{(2i+1)\Delta} - X^{(2)}_{2i\Delta}}{\sqrt{\Delta}} \end{pmatrix}$$

Denote $s_i = \det \Phi_i^2$ and $S = \sum_{i=1}^n s_i$.

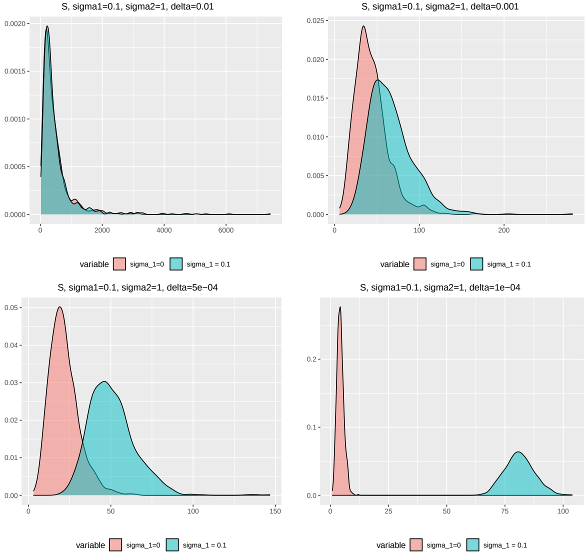


Figure: Distribution of the test statistics S (uncentered) for 1000 trajectories, sampled with different step sizes. Elliptic case (blue color) corresponds to the case $\sigma_1 = 0.1, \sigma_2 = 1$, hypoelliptic case (rose color) — to $\sigma_1 = 0, \sigma_2 = 1$.

Problems with the test

There are two main problems with the tests:

- **First,** statistics coincide when Δ is not small enough. Reason: drift has an order $\sqrt{\Delta}$, which is equal to σ_1 in the elliptic case for $\Delta = 0.01$.
- **Second**, it seems that an "optimal" threshold is linked to the step size. It is difficult to set.

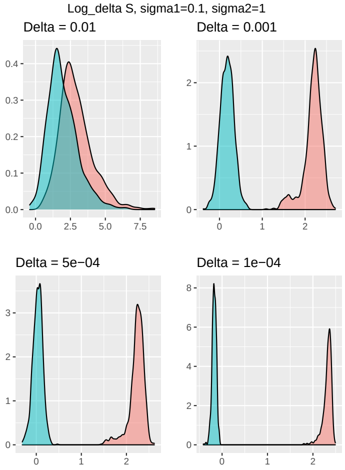


Figure: Distribution of the test statistics $\dot{S}$ for 1000 trajectories, sampled with different step sizes. Elliptic case (blue color) corresponds to the case $\sigma_1 = 0.1, \sigma_2 = 1$, hypoelliptic case (rose color) – to $\sigma_1 = 0, \sigma_2 = 1$.

What happens in a d-dimensional case?

In the last part of the presentation, we want to go back to the ”classical” setting adopted by Jacod et al. (2008), Jacod and Podolskij (2013), for the following reasons:

- It is interesting **how S behaves in a high-dimensional non-asymptotic setting**.
- Results, studying the distribution of a determinant for matrices with non-centered, not-unit variance, not identically distribed normal entries are not available.
- It is interesting (though challenging) to give the conditions when the asymptotic setting fails to separate two hypotheses.

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:

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

Main statistics of the test

Consider the matrix:

$$\Phi_i := \begin{pmatrix} \frac{X^{(1)}_{id+1} - X^{(1)}_{id}}{\sqrt{\Delta}} & \frac{X^{(1)}_{id+2} - X^{(1)}_{id+1}}{\sqrt{\Delta}} & \dots & \frac{X^{(1)}_{id+d} - X^{(1)}_{id+d-1}}{\sqrt{\Delta}} \\ \frac{X^{(2)}_{id+1} - X^{(2)}_{id}}{\sqrt{\Delta}} & \frac{X^{(2)}_{id+2} - X^{(2)}_{id+1}}{\sqrt{\Delta}} & \dots & \frac{X^{(2)}_{id+2d} - X^{(2)}_{id+d-1}}{\sqrt{\Delta}} \\ \dots & \dots & \dots & \dots \\ \frac{X^{(d)}_{id+1} - X^{(d)}_{id}}{\sqrt{\Delta}} & \frac{X^{(d)}_{id+2} - X^{(d)}_{id+1}}{\sqrt{\Delta}} & \dots & \frac{X^{(d)}_{id+d} - X^{(d)}_{id+d-1}}{\sqrt{\Delta}} \end{pmatrix}^2, \tag{6}$$

$i = 1, \dots, n$, 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:

$$\Omega_c^+ := \left\{ \left\| \xi_i^j \right\|^2 \leq \mathbb{E}\left[\left\| \xi_i^j \right\|^2 \right] + c_i^j \quad \forall j \in 1, \dots d,\ i = 1, \dots, n \right\}, \qquad (7)$$

  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:

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

Lemma

In Ω_c^+ the following inequality holds:

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

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.

Literature I

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

A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane currents and its application to conduction and excitation in nerve. *Journal of Physiology-London*, 117(4):500–544, 1952.

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.

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`.

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

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

Henry C Tuckwell. *Introduction to theoretical neurobiology: volume 2, nonlinear and stochastic theories*, volume 8. Cambridge University Press, 2005.

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!