Computational statistics for biological models

Adeline Leclercq Samson (Grenoble, France)

Joint work with

F. Comte (MAP5, Paris), S. Ditlevsen (Copenhagen),
J. Léon (Caracas), C. Prieur (LJK, Grenoble),
M. Tamborrino (Linz), M. Thieullen (LPMA, Paris), I. Tubikanec (Linz)

LABORATOIRE JEAN KUNTZMANN
MATHÉMATIQUES APPLIQUÉES - INFORMATIQUE

UNIVERSITÉ Grenoble Alpes

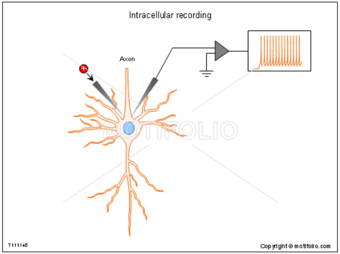
Intracellular recording data of one single neuron

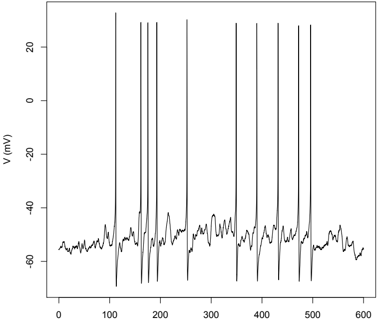
Data

- Membrane potential: difference in voltage between the interior and exterior of the cell
- High frequency records available ($\Delta = 0.1$ ms)

Objectives

- Prediction of spike emission
- Neuronal modeling with stochastic models
- Estimation/identification

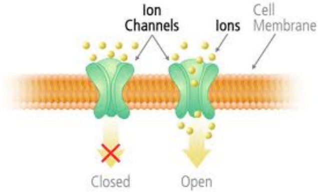




Hypoelliptic model for intracellular neuronal data

Deterministic Morris-Lecar neuronal model

- Calcium, potassium, leakage ionic currents
- g_{Ca}, g_K, g_L maximal conductances
- V_{Ca}, V_K, V_L reversal potential
- I input current
- C_t proportion of opened potassium channels
- Functions α and β: opening and closing rates



$$\frac{dV_t}{dt} = -g_{Ca} m_\infty(V_t)(V_t - V_{C_a}) - g_K C_t(V_t - V_K) - g_L(V_t - V_L) + I$$
$$\frac{dC_t}{dt} = \alpha(V_t)(1 - C_t) - \beta(V_t)C_t$$

Stochastic Morris-Lecar model

- N potassium gates
- $C_N(t)$ proportion of open gates among N gates at time t
- stochastic opening and closing at random times

$$\mathrm{C}_{\text{closed}} \underset{\beta\ (\mathrm{V})}{\overset{\alpha\ (\mathrm{V})}{\rightleftarrows}} \mathrm{C}_{\text{open}}$$

Between jumps of C_N, the trajectory of the continuous component V_t follows

$$\frac{dV_t}{dt} = -g_{Ca} m_\infty(V_t)(V_t - V_{C_a}) - g_K\, C_N(t)\,(V_t - V_K) - g_L(V_t - V_L) + I$$

$\Rightarrow$ **Piecewise Deterministic Markov Process**

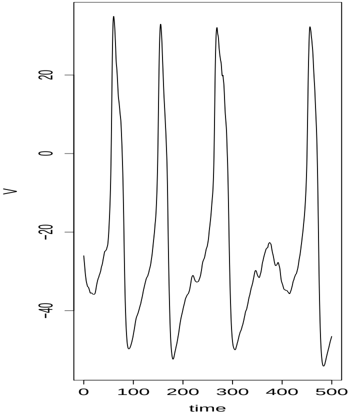
- **Diffusion approximation** [Wainrib, Thieullen, Pakdaman, EJP 2012]
  - $(V_t, C_N(t))$ is approximated by

$$\begin{aligned} dV_t &= \left(-g_{Ca} m_\infty(V_t)(V_t - V_{C_a}) - g_K C_t (V_t - V_K) - g_L (V_t - V_L) + I\right) dt \\ dC_t &= \left(\alpha(V_t)(1 - C_t) - \beta(V_t) C_t\right) dt + \sigma(V_t, C_t) dB_t \end{aligned}$$

- Diffusion of dimension 2 driven by only one Brownian motion

Hypoellitic diffusion

Hypoelliptic Morris-Lecar model



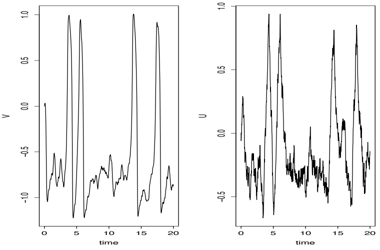
Morris-Lecar is highly non-linear $\Rightarrow$ Difficult to study

Fitzhugh-Nagumo model: a simplest model !

[Lindner et al 1999, Gerstner and Kistler, 2002, Lindner et al 2004, Berglund and Gentz, 2006]

$$
\begin{aligned}
dV_t &= \frac{1}{\varepsilon}(V_t - V_t^3 - C_t - s)dt, \\
dC_t &= (\gamma V_t - C_t + \beta)\, dt + \tilde{\sigma} dB_t,
\end{aligned}
$$

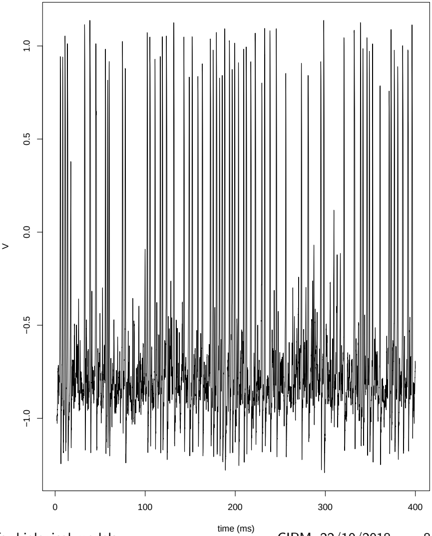
- V_t membrane potential of a single neuron
- C_t recovery variable / channel kinetics
- ε time scale separation
- s stimulus input, β position of the fixed point, γ duration of excitation
- B_t Brownian motion, $\tilde{\sigma}$ diffusion coefficient



Objectives of the talk

[Leòn and Samson, 2018]

1. Probabilistic properties of the system
   - Hypoellipticity
   - Stationary distribution
   - β-mixing
2. Parameter estimation
   - Why it's difficult for hypoelliptic SDE
   - Problem of partial observations
3. Estimation by pseudo-likelihood
   - Particle filter
   - Stochastic Approximation EM
4. Estimation by ABC
   - Numerical splitting scheme
   - Choice of the summaries

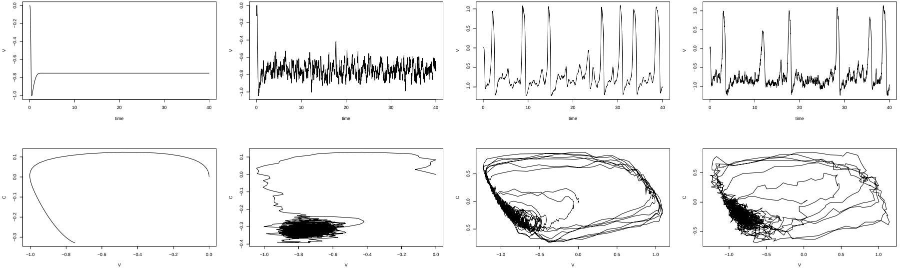


1. Probabilistic properties

[Leòn and Samson, 2018]

Hypoellipticity of the system

- Condition: drift of the first coordinate depends on C
- Noise of the second coordinate propagates to the first one



No noise — Noise on V — Noise on C — Noise on V, C

$\Rightarrow$ Hypoellipticity has consequences on the generation of spikes

Other probabilist properties

A difficult task

- Main results assume a non null noise
- A class of well studied hypoelliptic systems is

$$\begin{aligned} dV_t &= U_t dt, \\ dU_t &= -(c(V_t)\,U_t + \partial_v P(V_t))dt + \sigma dB_t, \end{aligned}$$

with $P(v)$ a potential, $c(v)$ a damping force.

  - Stochastic Damping Hamiltonian system [Wu 2001]
  - Langevin Equation [Wu 2001]
  - Hypocoercif model [Villani, 2009]

Good news!

- We enter the previous class by setting $U_t = \frac{1}{\varepsilon}(V_t - V_t^3 - C_t - s)$:

$$\begin{aligned} dV_t &= U_t dt, \\ dU_t &= \frac{1}{\varepsilon}\left(U_t(1-\varepsilon-3V_t^2) - V_t(\gamma-1) - V_t^3 - (s+\beta)\right) dt - \frac{\tilde{\sigma}}{\varepsilon} dB_t, \end{aligned}$$

Stationary distribution

- Existence of Lyapounov function $\Psi(v, u) = e^{F(v,u) - \inf_{\mathbb{R}^2} F}$ with explicit F
- Existence and uniqueness of the stationary density [Wu, 2001]

β-mixing

- Process (V_t, U_t) is β-mixing [Wu, 2001]

Same properties for process (V_t, C_t)

2. Estimation of parameters

$$
\begin{aligned}
dV_t &= \frac{1}{\varepsilon}(V_t - V_t^3 - C_t - s)dt, \\
dC_t &= (\gamma V_t - C_t + \beta)\, dt + \tilde{\sigma}\, dB_t,
\end{aligned}
$$

Observations

- Data: discrete observations $V_{0:n} = (V_0, \ldots, V_n)$ at times $t_0 = 0 < t_1 = \Delta < \ldots < t_n = n\Delta$
- Hidden coordinate (C_t)

Difficult because

- Hypoellipticity
- No explicit transition density of the SDE
- Hidden coordinate C

Ideal case of complete observations and noise on both coordinates $X_t = (V_t, C_t)$:

$$dX_t = b_\mu(X_t)dt + \Sigma dB_t, \quad \Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$$

Discretization with **Euler-Maruyama** of $X_{i+1} = (V_{i+1}, U_{i+1})$:

$$X_{i+1} = X_i + \Delta b_\mu(X_i) + \sqrt{\Delta}\Sigma\eta_i, \quad \eta_i \sim_{iid} \mathcal{N}(0, I)$$

Minimum contrast estimator [Genon-Catolot, Jacod, 1993; Kessler 1996]
Set $\Gamma = \Sigma'\Sigma$.

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \left(\sum_{i=1}^{n-1} (X_{i+1} - X_i - \Delta b_\mu(X_i))' \, \Gamma^{-1} \, (X_{i+1} - X_i - \Delta b_\mu(X_i)) + \sum_{i=1}^{n-1} \log\det\Gamma \right)$$

- $\hat{\mu}$, $\hat{\Gamma}$ asymptotically normal

What about hypoelliptic SDE ?

Impossible to apply because

$$\Gamma = \begin{pmatrix} 0 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \quad \text{not invertible}$$

Previous contrast

$$\sum_{i=1}^{n-1} (X_{i+1} - X_i - \Delta b_\mu(X_i))' \, \Gamma^{-1} \, (X_{i+1} - X_i - \Delta b_\mu(X_i)) + \sum_{i=1}^{n-1} \log \det \Gamma$$

is not defined...

Idea: change of variable

- **Assume ε known**, and change the system with $U_t = \frac{1}{\varepsilon}(V_t - V_t^3 - C_t - s)$
- New system:

$$\begin{aligned} dV_t &= U_t dt, \\ dU_t &= \frac{1}{\varepsilon}\left(U_t(1-\varepsilon-3V_t^2) - V_t(\gamma-1) - V_t^3 - (s+\beta)\right) dt - \frac{\tilde{\sigma}}{\varepsilon} dB_t, \end{aligned}$$

Class of Hamiltonian (hypoelliptic) SDE

$$\begin{aligned} dV_t &= U_t dt \\ dU_t &= b_\mu(V_t, U_t)dt + \sigma_2 dB_t \end{aligned}$$

Estimation

- Martingale estimating functions [Ditlevsen, Sorensen, 2004]
- Gibbs sampler [Pokern et al, 2010]
- Euler contrast [Gloter 2006, Samson, Thieullen, 2012];
- Higher order contrast [Ditlevsen, Samson, 2018]

Partial observations

- U_t **not observed** but can be replaced by

$$\bar{V}_i := \frac{V_{i+1} - V_i}{\Delta} = \frac{\int_{i\Delta}^{(i+1)\Delta} U_s ds}{\Delta} \approx U_{i\Delta}$$

- Contrast function on the second equation with plug-in $\bar{V}$
  - $\mu = (\beta, \gamma, s)$

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \left(\sum_{i=1}^{n-1} \frac{\left(\overline{V}_{i+1} - \overline{V}_i - \Delta b_\mu(V_{i-1}, \overline{V}_{i-1})\right)^2}{\Delta\sigma^2} + \sum_{i=1}^{n-1} \log\sigma^2 \right)$$

  - $\hat{\mu}$ is unbiased, asymptotically normal
  - $\hat{\sigma}$ is biased (because $\bar{V}_i$ is not Markovian)

Partial observations

- U_t **not observed** but can be replaced by

$$\bar{V}_i := \frac{V_{i+1} - V_i}{\Delta} = \frac{\int_{i\Delta}^{(i+1)\Delta} U_s ds}{\Delta} \approx U_{i\Delta}$$

- Contrast function on the second equation with plug-in $\bar{V}$
  - $\mu = (\beta, \gamma, s)$

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left(\frac{3}{2} \sum_{i=1}^{n-2} \frac{\left(\overline{V}_{i+1} - \overline{V}_i - \Delta b_\mu(V_{i-1}, \overline{V}_{i-1})\right)^2}{\Delta \sigma_2^2} + \sum_{i=1}^{n-2} \log \sigma_2^2 \right)$$

  - $\hat{\mu}$ is unbiased, asymptotically normal
  - $\hat{\sigma}$ is unbiased, asymptotically normal [Gloter 2006]

Why it doesn't work assuming ε unknown

What can not be applied

- Change of variable
- Euler contrast

Alternatives

- **Pseudo-likelihood**
  - Higher order discrete scheme that propagates the noise to the first coordinate
  - Contrast for complete observations as in the "ideal" case of non null
  - Particle filter for the partial observations
  - Stochastic Approximation EM algorithm coupled to Particle filter

- **Approximate Bayesian Computation (ABC)**
  - Likelihood free method
  - Numerical scheme to simulate from the hypoelliptic SDE
  - Higher order discrete scheme or Splitting scheme

3. Pseudo-likelihood and particle filter approach

$$
\begin{aligned}
dV_t &= b_1(V_t, C_t, \mu)dt \\
dC_t &= b_2(V_t, C_t, \mu)dt + \sigma_2 dB_{2t}
\end{aligned}
$$

- Previous trick where C_i is replaced by $\frac{V_{i+1}-V_i}{\Delta}$ not available
- We want to filter C_i and compute $\pi_{n,\theta} f = \mathbb{E}\left(f(C_n) | V_{0:n}; \theta\right)$
  - Kalman filter when SDE is linear and Gaussian
  - **Particle filter/Sequential Monte Carlo** (SMC)
    [Del Moral et al, 2001; Doucet et al, 2001; Chopin, 2004; ...]

Particle filter/Sequential Monte Carlo

- Iterative algorithm
- Simulation of K particles $C^k_{0:n}$ and computation of weights $w_n(C^k_{0:n})$
- Empirical measure $\Psi^K_{n;\theta} = \sum_{k=1}^K w_n(C^k_{0:n})\mathbf{1}_{C^k_{0:n}}$

At time $j = 1, \dots, n$, $\forall\, k = 1, \dots, K$:

1. simulation of $C_j^{(k)} \sim q(\cdot|V_j, C_{j-1}^{(k)};\theta)$
2. calculation of weights

$$w\left(C_{0:j}^{(k)}\right) = \frac{p(V_{0:j}, C_{0:j}^{(k)}|\theta)}{p(V_{0:j-1}, C_{0:j-1}^{(k)}\theta)q(C_j^{(k)}|V_j, C_{j-1}^{(k)};\theta)}$$

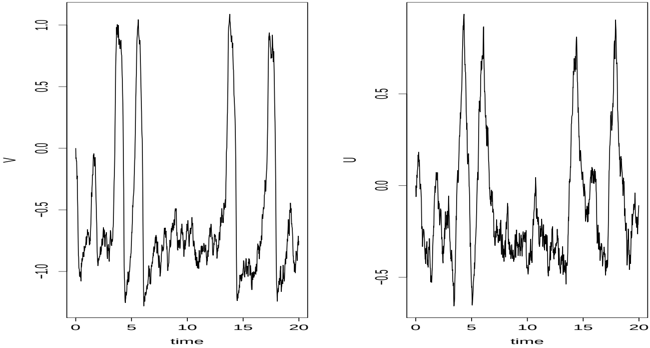
Particle filter for hypoelliptic SDE

- **Degenerate Hidden Markov Model**:
  - (V_i, C_i) Markovian but not (C_i)
  - Set $X_i = (V_i, C_i)$, with Markov kernel $p(dV_i, dC_i | V_{i-1}, C_{i-1})$
  - $V_i = X_i^{(1)}$ with transition kernel $\mathbb{1}_{\{V = X^{(1)}\}}$ (zero almost everywhere)
- **Discretization of the SDE** [Ditlevsen, Samson, 2018]
  - High order Taylor-Ito development
  - Propagation of the noise to the first coordinate

$$\tilde{\Gamma} = \sigma^2 \begin{pmatrix} \frac{\Delta^3}{3} & \frac{\Delta^2}{2} \\ \frac{\Delta^2}{2} & \Delta \end{pmatrix}$$

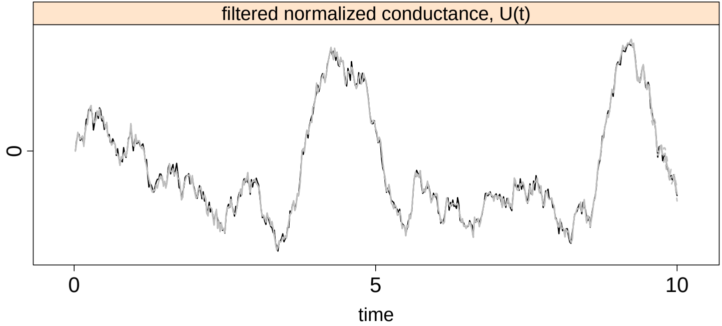
FitzHugh-Nagumo model

$$
\begin{aligned}
dV_t &= \frac{1}{\varepsilon}(V_t - V_t^3 - C_t - s)dt \\
dC_t &= (\gamma V_t - C_t + \beta)\, dt + \sigma_2 dB_{2t}
\end{aligned}
$$



Filtering with the hypoelliptic neuronal FitzHugh-Nagumo

Parameter θ fixed at the true value



Estimation by Expectation-Maximization (EM) algorithm

- Likelihood non explicit, even with the discretisation scheme

$$p_\Delta(V_{0:n};\theta) = \int \prod_{i=1}^{n} p_\Delta(V_i, C_i | V_{i-1}, C_{i-1}; \theta) dC_{0:n}$$

- Incomplete data model
  - Observed data $(V_{0:n})$
  - Complete data $(V_{0:n}, C_{0:n})$

- EM algorithm [Dempster, Laird, Rubin, 1977], iteration m
  - *E Step*: computation of $Q_{m+1}(\theta) = \mathbb{E}_\Delta\left[\log p_\Delta(V_{0:n}, C_{0:n}; \theta) \mid V_{0:n}, \widehat{\theta}_m\right]$
  - *M Step*: update $\widehat{\theta}_{m+1} = \arg\max_\theta Q_{m+1}(\theta)$

- Convergence results [Wu, 1983]

Stochastic Approximation (SAEM) algorithm [Delyon, Lavielle, Moulines, 1999, Ditlevsen, Samson, 2014]

- *E Step*
  - *S Step*: simulation of $C_{0:n}^{(m)}$ under $p_\Delta(C_{0:n}|V_{0:n};\widehat{\theta}_m)$ with particle filter
  - *SA Step*: stochastic approximation of Q_{m+1}

$$Q_{m+1}(\theta) = (1-\alpha_m)Q_m(\theta) + \alpha_m \log p_\Delta(V_{0:n}, C_{0:n}^{(m)};\theta)$$

- *M Step*: $\widehat{\theta}_{m+1} = \arg\max_\theta Q_{m+1}(\theta)$

Convergence [Ditlevsen, Samson, 2014, 2018]

Assumptions

1. $\sum_m \alpha_m = \infty$, $\sum_m \alpha_m^2 < \infty$.
2. Number of particles $K(m) = \log(m^{1+\delta})$

$$\widehat{\theta}_m \xrightarrow[m\to\infty]{a.s.} \text{(local) max of likelihood } p_\Delta(V_{0:n};\theta)$$

Tool: convergence of Robbins-Monroe scheme and inequality deviation for the particle filter

Some estimation results obtained from 100 simulated data sets

		ε fixed	ε fixed	ε estimated
	True	New contrast	Euler Contrast	New contrast
ε	0.100	–	–	0.105 (0.010)
γ	1.500	1.523 (0.130)	1.499 (0.196)	1.592 (0.160)
β	0.800	0.821 (0.110)	0.779 (0.107)	0.866 (0.130)
σ	0.300	0.293 (0.008)	0.381 (0.038)	0.306 (0.020)

4. Approximate Bayesian Computation (ABC)

Posterior distribution

$$\underbrace{\pi(\theta|V)}_{\textit{posterior}} \propto \underbrace{\pi(V|\theta)}_{\substack{\textit{likelihood}\\(\textit{intractable})}} \underbrace{\pi(\theta)}_{\textit{prior}}$$

Approximate Bayesian Computation

1. Simulate $\theta^k \sim \pi(\theta)$ for $k = 1, \ldots, K$
2. Generate pseudo-data V_θ from the SDE model for each θ^k
3. Introduce summaries $s(V)$ and $s(V_\theta)$ of the data
4. Approximate the posterior

$$\pi(\theta|V) \approx \pi_{d,\eta,s}(\theta|V) = \pi(\ \theta \mid d(s(V), s(V_\theta)) < \eta\ )$$

for a 'small' η

Difficulties with hypoelliptic SDE

$$\begin{aligned} dV_t &= b_1(V_t, C_t, \mu)dt \\ dC_t &= b_2(V_t, C_t, \mu)dt + \sigma_2 dB_{2t} \end{aligned}$$

- Transition density unknown
  - No explicit scheme to simulate a solution
- Discretization schemes
  - Euler scheme does not preserve the structure of the noise
  - Higher order schemes do not preserve the structure of the invariant measure
- Choice of the summaries for temporal series

Numerical splitting scheme

[Samson, Tamborrino, Tubikanec, work in progress]

The FHN equation can be splitted into two subsystems.

1. **Subsystem a:** Linear SDE

$$d\begin{pmatrix} V(t) \\ C(t) \end{pmatrix} = \begin{pmatrix} -\frac{1}{\epsilon}C(t) \\ \gamma V(t) - C(t) \end{pmatrix} dt \; + \; \begin{pmatrix} 0 \\ \sigma \end{pmatrix} dW(t)$$

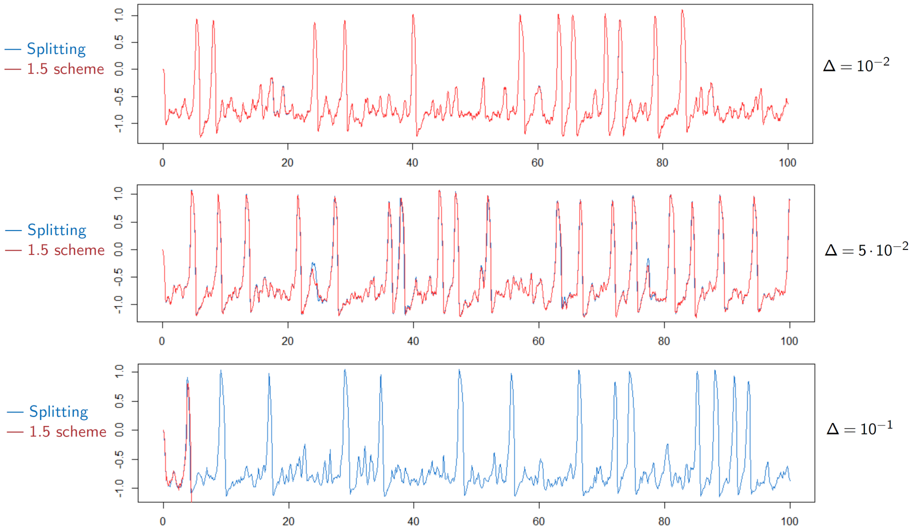
2. **Subsystem b:** Non-linear (decoupled) ODE

$$d\begin{pmatrix} V(t) \\ C(t) \end{pmatrix} = \begin{pmatrix} \frac{1}{\epsilon}(V(t) - V^3(t)) \\ \beta \end{pmatrix} dt$$

Numerical splitting scheme with time step Δ

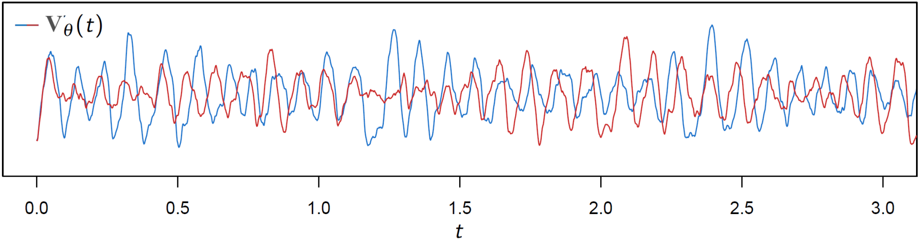
$$\hat{\mathbf{X}} = X^b_{\Delta/2} \circ X^a_{\Delta} \circ X^b_{\Delta/2}$$

Comparison of Splitting and Order 1.5 Strong Taylor Scheme



Choice of the summaries in ABC

How to account for the variability in the data for identical θ?



$\Longrightarrow$ **Transform the data from time to frequency domain**

Spectral Density (Periodogram)

- Stationary stochastic process: $\mathbf{V}_\theta = (V_\theta(t))_{t\geq 0}$
- Autocovariance function: $Cov(V_\theta(t), V_\theta(s)) = r_\theta(\tau = t - s)$

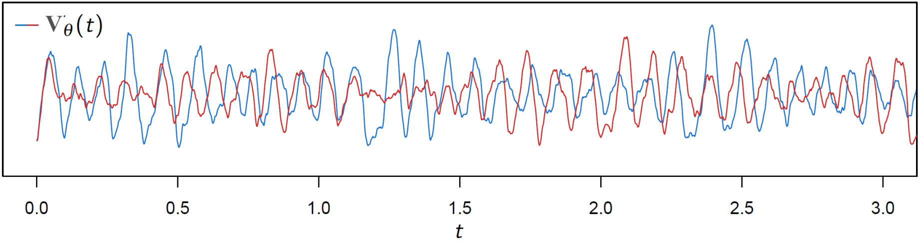
$$S_{\mathbf{V}_\theta}(\omega = 2\pi f) = \int_{-\infty}^{\infty} r_\theta(\tau) e^{-i\omega\tau} \, d\tau, \quad \omega \in [-\pi, \pi]$$

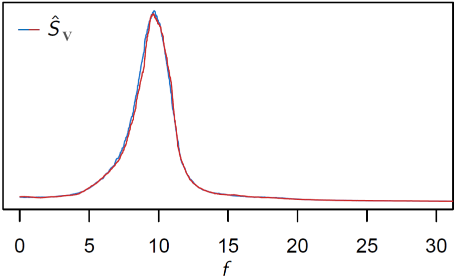
Definition (Periodogram)

$$s(V) = \hat{S}_V(\omega) = \frac{1}{n} \left| \sum_{j=1}^{n} V_j e^{-i\omega j} \right|^2$$

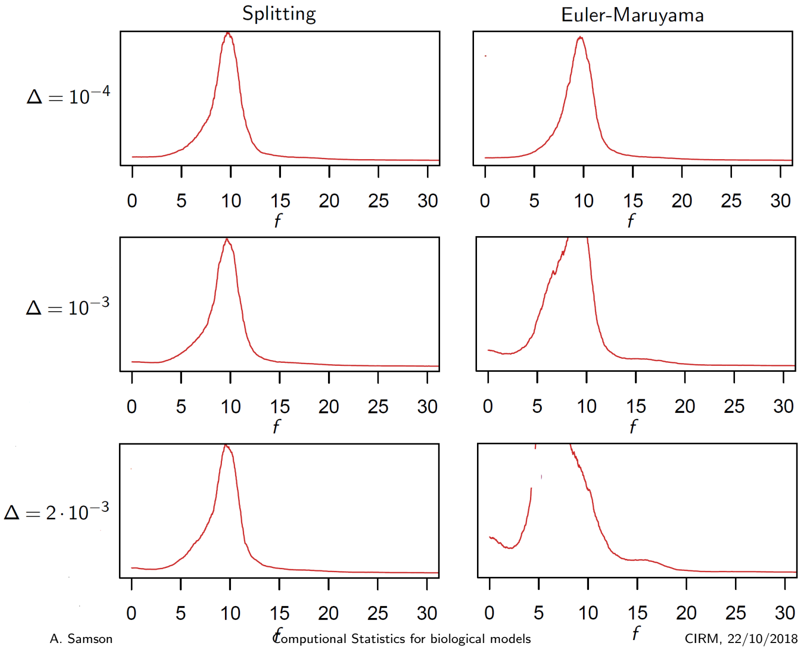
- Time domain: Discrete data $V = (V_1, ..., V_n)$
- Frequency domain: R-function `spectrum` (Fast Fourier Transform)

Simulated Data and Periodogram Estimates





The Splitting Scheme preserves the Periodogram



Structure-Preserving ABC-Algorithm

- Precompute the periodograms $s(V) = (\hat{S}_{V_1}, ..., \hat{S}_{V_M}) = (s_1, ..., s_M)$
- Choose prior distribution $\pi(\theta)$ and tolerance η

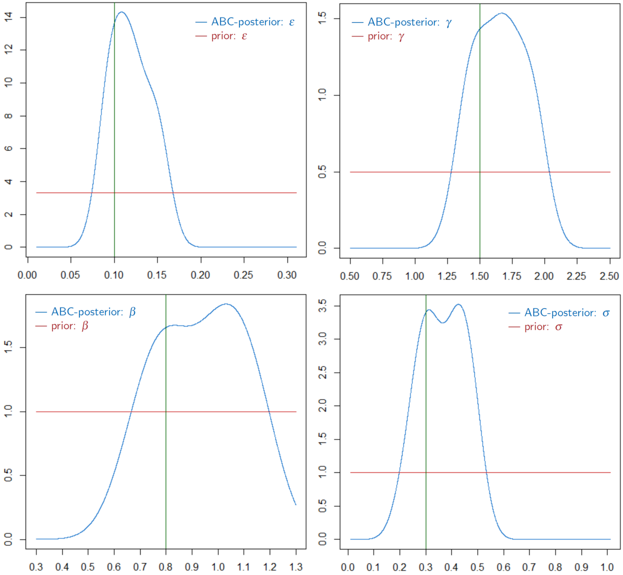
for k=1:N

- Draw θ^k from the prior $\pi(\theta)$
- Simulate new data $\hat{V}_{\theta^k}$ using the numerical **splitting** approach
- Compute the **periodogram** $s(\hat{V}_{\theta^k}) = \hat{s}_{\theta_i}$
- Calculate $D^k(s(V), \hat{s}_{\theta^k})$
- Store samples (D^k, θ^k)

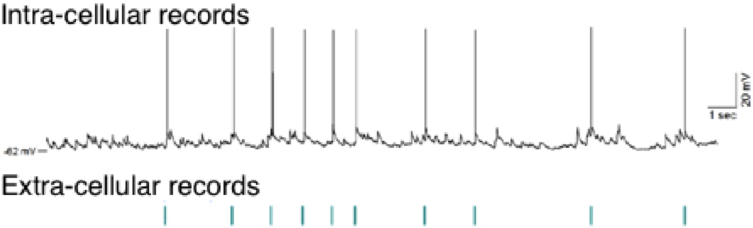
If $D^k < \eta$, keep θ^k as a sample of the posterior $\pi_{d,\eta,s,\hat{V}_\theta}(\theta|V)$

ABC for the FHN Model

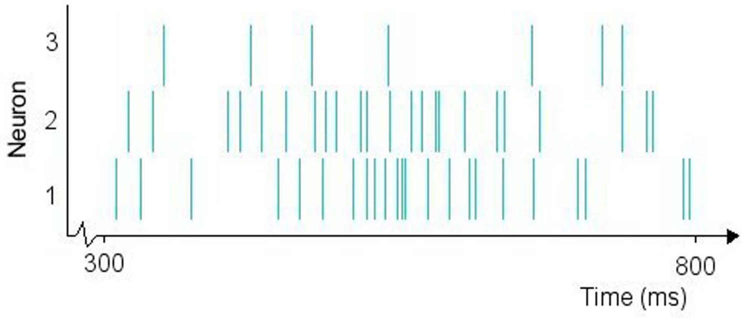
- *Estimated parameters*: $\theta = (\epsilon, \gamma, \beta, \sigma)$
- *True values*: $\epsilon = 0.1$, $\gamma = 1.5$, $\beta = 0.8$, $\sigma = 0.3$
- *Priors*: $\epsilon \sim U(0.01, 0.31)$, $\gamma \sim U(0.5, 2.5)$, $\beta \sim U(0.3, 1.3)$, $\sigma \sim U(0.01, 1.01)$
- *Observed data*: $M = 30$ paths of V for $\Delta = 7.5 \cdot 10^{-2}$ and $T = 10^4$.
- *Simulated data in ABC*: $N = 10^6$ paths of V with same Δ and T.
- *Kept samples fom the posterior*: 0.01^{th} percentile


ABC-posterior: ε
prior: ε
0.00 0.05 0.10 0.15 0.20 0.25 0.30
0 2 4 6 8 10 12 14
ABC-posterior: γ
prior: γ
0.50 0.75 1.00 1.25 1.50 1.75 2.00 2.25 2.50
0.0 0.5 1.0 1.5
ABC-posterior: β
prior: β
0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3
0.0 0.5 1.0 1.5
ABC-posterior: σ
prior: σ
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5

From intra to extracellular recording data



Extracellular recordings of several neurons



Objectives

- Understanding connexion between neurons
- Neuronal modeling with stochastic models
- Estimation/identification

Extracellular stochastic models

A Multi-timescale Adaptive Threshold model [Samson, Tamborrino, work in progress]

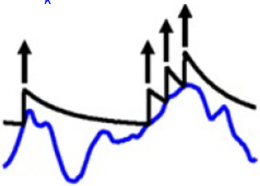
- Potential of a model neuron u_t, modeled by an Ornstein-Uhlenbeck process

$$du_t = \left(-\frac{u_t}{\tau} + \mu\right) dt + \sigma dW_t$$

  - τ membrane time constant
  - μ, σ positiv drift and diffusion coefficients

- Decaying threshold $\theta_u(t)$
  - If $u_t > \theta_u(t) \implies$ emits a spike at time t
  - Threshold linearly modulated by spikes

$$\theta_u(t) = \theta_\infty + \sum_k H_u(t - t_k), \quad H_u(t) = \sum_{j=1}^{L} \alpha_j \exp(-\lambda_j t)$$



Estimation for Multi-timescale Adaptive Threshold model

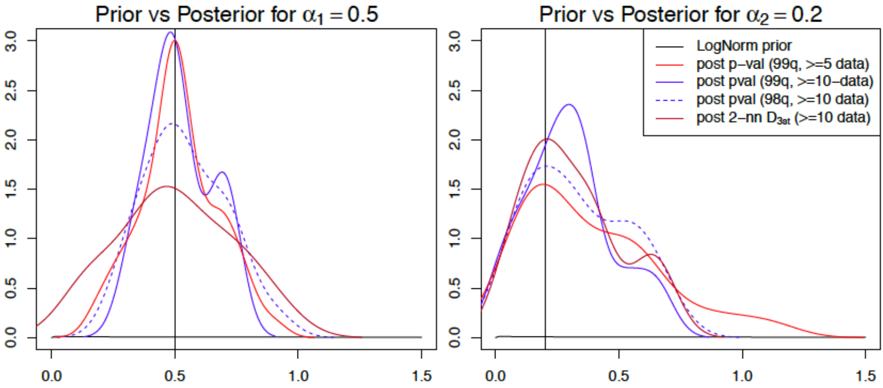
Difficulties

- Process is **not renewal**, consecutive inter-times intervals $t_k - t_{k-1}$ are **not iid**
- Potential u_t and threshold $\theta_u(t)$ are not observed
- Conditioned on hitting times, $\theta_u(t)$ is deterministic
- Not a hidden Markov model

ABC

- Choice of the summaries
- Similarity between two point processes
  - Periodogram
  - Kolmogorov Smirnov two sample hypothesis test
  - k-Nearest Neighbors

First results



Conclusion/Perspectives

- Hypoelliptic FHN system
  - Existence of stationary density
  - Parametric estimation
    - Particle filter
    - EM algorithm
    - Numerical splitting scheme: more complex models?
    - ABC and choice of the summaries
- Models for spike trains
  - Point processes, MAT models
  - Choice of the summaries??