Memory and hypoellipticity in neuronal models

S. Ditlevsen  R. Höpfner  E. Löcherbach  M. Thieullen

Banff, 2017
What this talk is about:

- What is the effect of memory in probabilistic models for neurons?
- Study this in two classical models: the stochastic Hodgkin-Huxley model and a model of interacting Hawkes processes in high dimension.
- Why does memory lead to hypoellipticity?
- What are the probabilistic and statistical consequences of this fact?
Joint work with M. Thieullen and R. Höpfner.

Classically: HH-model = 4-dimensional deterministic model

\[(V_t, n_t, m_t, h_t), t \geq 0\]

for the membrane potential \( V \) of a single neuron, together with **gating variables** \( n, m, h \) that account for the **ion currents** of \( K^+ \) and \( Na^+ \) ions:

\[
\begin{align*}
    dV_t &= l_t dt - F(V_t, n_t, m_t, h_t) dt \\
    dn_t &= \left[ \alpha_n(V_t)(1 - n_t) - \beta_n(V_t)n_t \right] dt \\
    dm_t &= \left[ \alpha_m(V_t)(1 - m_t) - \beta_m(V_t)m_t \right] dt \\
    dh_t &= \left[ \alpha_h(V_t)(1 - h_t) - \beta_h(V_t)h_t \right] dt,
\end{align*}
\]

where \( t \to l_t \) is some deterministic signal, and where

\[
F(v, n, m, h) = g_K n^4 (v - E_K) + g_{Na} m^3 h (v - E_{Na}) + g_L (v - E_L).
\]
The model - continued

• All conductances $g_K$, $g_{Na}$, $g_L$ and equilibrium potentials $E_K$, ... are explicitly known.

Also the voltage dependent activation and inactivation functions $\alpha, \beta$:

$$\alpha_n(v) = \frac{0.1 - 0.01v}{\exp(1 - 0.1v) - 1}, \ldots$$

not bounded, but analytic functions.
The model - continued

- All conductances $g_K, g_{Na}, g_L$ and equilibrium potentials $E_K, \ldots$ are explicitly known.
Also the voltage dependent activation and inactivation functions $\alpha, \beta$:

$$\alpha_n(v) = \frac{0.1 - 0.01v}{\exp(1 - 0.1v) - 1}, \ldots$$

not bounded, but analytic functions.

• Let us call the equations for the gating variables $n, m, h$ memory equations.

• Indeed, for a fixed voltage trajectory $(V_t)_{t \geq 0}$, the equations for $n, m, h$ are linear, and we obtain e.g.

$$n_t = n_0 e^{-\int_0^t (\alpha_n + \beta_n)(V_s)ds} + \int_0^t \alpha_n(V_u) e^{-\int_u^t (\alpha_n + \beta_n)(V_r)dr} du,$$

and similar formulas for the other gating variables...
**Figure**: Deterministic HH with constant input $S = 15$ starting in a numerical approximation to its equilibrium point.
Add noise to the system!

Indeed: the neuron does certainly not receive the original deterministic signal:
It receives input from its dendritic system.

This system has a large number of synapses: they register spike trains from a huge number (about 10000!) of other neurons.
Höpfner 2007: used kernel estimators to estimate drift and diffusion coefficient of the membrane potential process - away from spike times ⇒ In some neurons, the input is well modeled by

\[ d\xi_t = \nu (S(t) - \xi_t) \, dt + \gamma \, dW_t \]

Ornstein-Uhlenbeck process of mean-reverting type, where

\[ t \rightarrow S(t) \]

is some deterministic stimulus processed by the network. In other neurons, a CIR-process is a good model for the stochastic input.
Stochastic Hodgkin-Huxley model with input $t \rightarrow S(t)$ given by

\[
\begin{align*}
    dV_t &= d\xi_t - F(V_t, n_t, m_t, h_t)dt \\
    dn_t &= [\alpha_n(V_t)(1 - n_t) - \beta_n(V_t)n_t]dt \\
    dm_t &= [\alpha_m(V_t)(1 - m_t) - \beta_m(V_t)m_t]dt \\
    dh_t &= [\alpha_h(V_t)(1 - h_t) - \beta_h(V_t)h_t]dt \\
    d\xi_t &= \nu(S(t) - \xi_t)dt + \gamma dW_t.
\end{align*}
\]

5–dimensional diffusion driven by 1–dimensional Brownian motion present in $\xi$ and in $V$. $S(t)$ supposed to be $T$–periodic signal, for some fixed periodicity $T$. 
The stochastic Hodgkin-Huxley model

Hörmander Condition

A system of interacting neurons

Erlang kernels and hypoellipticity

HH with signal and noise: $V^{\text{out}}$ function of $t$

HH with signal and noise: $n$ (violet), $m$ (blue), $h$ (grey) functions of $t$

Driving noisy input: $V^{\text{in}}$ function of $t$

The following parameters were used for $V^{\text{in}}$: period = 28, amplitude = 9, sigma = 0.5, tau = 0.75, $K = 15$
What we did with Reinhard and Michèle in a series of papers:

- We proved **limit theorems** for a large class of functionals of the process – under a certain condition of **periodic recurrence**.

- Analysis of **spiking patterns** in the neuron is then possible via SLLN. In particular: **Glivenko-Cantelli theorem for the empirical ISI distributions**.

- Tools: **Periodic ergodicity** induced by $T$—periodicity of the signal encoded in the drift, Nummelin- (or Doeblin-) type minorization condition based on existence of transition densities which are locally smooth.

This led us immediately to the notion of **Hypoellipticity**!
Hypoellipticity of the stochastic Hodgkin-Huxley model:

\[
\begin{align*}
    dV_t &= d\xi_t - F(V_t, n_t, m_t, h_t)\,dt \\
    dn_t &= \left[ \alpha_n(V_t)(1 - n_t) - \beta_n(V_t)\,n_t \right] \,dt \\
    \cdots \\
    d\xi_t &= \nu (S(t) - \xi_t) \,dt + \gamma\sqrt{\tau} \,dW_t.
\end{align*}
\]

Noise only present in first and fifth variable - this is a highly degenerate five-dimensional diffusion.
Hypoellipticity of the stochastic Hodgkin-Huxley model:

\[
\begin{align*}
    dV_t &= d\xi_t - F(V_t, n_t, m_t, h_t)dt \\
    dn_t &= \left[ \alpha_n(V_t)(1 - n_t) - \beta_n(V_t) n_t \right] dt \\
    \cdots \\
    d\xi_t &= \nu (S(t) - \xi_t) dt + \gamma \sqrt{\tau} dW_t
\end{align*}
\]

Noise only present in first and fifth variable - this is a **highly degenerate** five-dimensional diffusion. **What does Hypoelliptic mean?**

S. Ditlevsen, R. Höpfner, E. Löcherbach, M. Thieullen
• **Degenerate diffusion** are diffusions in dimension $m$ (here, 5), driven by lower-dimensional $n$–dimensional Brownian motion (here, $n = 1$).

• Question: does the $n$–dimensional noise generate **density** of the process in the whole $m$–dimensional state space?

If so, then we call the process **hypoelliptic**. (This is not exactly the definition of a **hypoelliptic diffusion generator**, but it is the basic idea.)

• What can be the problem in this degenerate case?
A toy model

To fix ideas, consider a two-dimensional toy model with noise only in one component:

\[ dX_t = dW_t, X_0 = x, \]
\[ dY_t = f(X_t)dt, Y_0 = y, \]

f smooth.
A toy model

To fix ideas, consider a two-dimensional toy model with noise only in one component:

\[ dX_t = dW_t, \quad X_0 = x, \]
\[ dY_t = f(X_t)dt, \quad Y_0 = y, \]

\( f \) smooth. What is the problem here?

First order approximation gives (for small \( t \)):

\[
\begin{pmatrix}
X_t \\
Y_t
\end{pmatrix}
\approx
\begin{pmatrix}
W_t \\
0
\end{pmatrix}
+
\begin{pmatrix}
x \\
y + f(x)t
\end{pmatrix}.
\]

Brownian motion only in one coordinate: this vector does not possess a two-dimensional Lebesgue density!
So: need a second order approximation. Use Itô:

\[ f(X_t) = f(x) + \int_0^t f'(X_s)dW_s + \frac{1}{2} \int_0^t f''(X_s)ds. \]
So: need a second order approximation. Use Itô:

\[ f(X_t) = f(x) + \int_0^t f'(X_s) dW_s + \frac{1}{2} \int_0^t f''(X_s) ds. \]

Plug this in: \( Y_t = y + \int_0^t f(X_s) ds : \)

\[
Y_t = y + f(x) t + f'(x) \int_0^t \int_0^s dW_u \, ds + \text{remainder} \\
= y + f(x) t + f'(x) \int_0^t (t - u) dW_u + \text{remainder}, \quad t << 1.
\]
Conclusion

\[
\begin{pmatrix}
X_t \\
Y_t
\end{pmatrix} = \begin{pmatrix}
W_t \\
f'(x) \int_0^t (t - u) dW_u
\end{pmatrix} + \text{remainder}.
\]

↑ Gaussian vector: has a two-dimensional density iff \( f'(x) \neq 0 \).
The stochastic Hodgkin-Huxley model

Hörmander Condition

A system of interacting neurons
Erlang kernels and hypoellipticity

Conclusion

\[
\begin{pmatrix}
X_t \\
Y_t
\end{pmatrix} = \begin{pmatrix}
W_t \\
f'(x) \int_0^t (t - u) dW_u
\end{pmatrix} + \text{remainder}.
\]

↑ Gaussian vector: has a two-dimensional density iff \( f'(x) \neq 0 \).

Points \( x \) with \( f'(x) \neq 0 \) are points where the (weak) Hörmander condition holds. If \( f'(x) = 0 \), iterate the above argument.

Sufficient: \( \exists n: f^{(n)}(x) \neq 0 \).
The stochastic Hodgkin-Huxley model

Hörmander Condition

A system of interacting neurons

Erlang kernels and hypoellipticity

Conclusion

\[
\begin{pmatrix}
X_t \\
Y_t
\end{pmatrix} = \left( \begin{pmatrix}
W_t \\
f'(x) \int_0^t (t-u) dW_u
\end{pmatrix} \right) + \text{remainder.}
\]

↑ Gaussian vector: has a two-dimensional density iff \( f'(x) \neq 0 \).

Points \( x \) with \( f'(x) \neq 0 \) are points where the (weak) Hörmander condition holds. If \( f'(x) = 0 \), iterate the above argument. Sufficient: \( \exists n: f^{(n)}(x) \neq 0 \).

The weak Hörmander condition is a condition which generalizes the well-known ellipticity condition.
The stochastic Hodgkin-Huxley model

Hörmander Condition

A system of interacting neurons

Erlang kernels and hypoellipticity

Conclusion

\[
\begin{pmatrix}
X_t \\
Y_t
\end{pmatrix}
= \begin{pmatrix}
W_t \\
 f'(x) \int_0^t (t - u)dW_u
\end{pmatrix} + \text{remainder}.
\]

↑ Gaussian vector: has a two-dimensional density iff \( f'(x) \neq 0 \).

Points \( x \) with \( f'(x) \neq 0 \) are points where the (weak) Hörmander condition holds. If \( f'(x) = 0 \), iterate the above argument.

Sufficient: \( \exists n : f^{(n)}(x) \neq 0 \).

The weak Hörmander condition is a condition which generalizes the well-known ellipticity condition. And it implies the hypo-ellipticity of the process.
A first comment on the structure of the covariance matrix

The Gaussian vector

\[
\begin{pmatrix}
W_t \\
f'(x) \int_0^t (t - u) dW_u
\end{pmatrix}
\]

has covariance matrix

\[
\begin{pmatrix}
t & f'(x) \frac{1}{2} t^2 \\
f'(x) \frac{1}{2} t^2 & (f'(x))^2 \frac{2}{3} t^3
\end{pmatrix}
\]

- To keep in mind: for small \( t \), this is of order \( t^{3/2} \).
- The geometrical structure has changed, and the coordinates of the process do not travel at the same speed.
- The more iterations one needs to span the whole space, the worse this speed is.
Weak Hörmander condition for stochastic HH model

The memory equations for the gating variables $n, m, h$ feel the noise only through the voltage variable $V_t$. 
Weak Hörmander condition for stochastic HH model

The memory equations for the gating variables $n, m, h$ feel the noise only through the voltage variable $V_t$. So we have to consider the derivatives with respect to $v$ of $\alpha_n(v)$ etc....

- $b_n, b_m, b_h$ : drift terms of $n, m, h$.

- Important quantity:

$$D(v, n, m, h) := \det \begin{pmatrix} \frac{\partial^2 b_n}{\partial v^2} & \frac{\partial^3 b_n}{\partial v^3} & \frac{\partial^4 b_n}{\partial v^4} \\ \frac{\partial^2 b_m}{\partial v^2} & \frac{\partial^3 b_m}{\partial v^3} & \frac{\partial^4 b_m}{\partial v^4} \\ \frac{\partial^2 b_h}{\partial v^2} & \frac{\partial^3 b_h}{\partial v^3} & \frac{\partial^4 b_h}{\partial v^4} \end{pmatrix} (v, n, m, h).$$

**Proposition**

The weak Hörmander condition holds for all points $(v, n, m, h, \zeta)$ such that $D(v, n, m, h) \neq 0$. 
Weak Hörmander condition for stochastic HH model

The memory equations for the gating variables $n, m, h$ feel the noise only through the voltage variable $V_t$. So we have to consider the derivatives with respect to $v$ of $\alpha_n(v)$ etc....

- $b_n, b_m, b_h$: drift terms of $n, m, h$.

- Important quantity:

$$D(v, n, m, h) := \det \begin{pmatrix} \frac{\partial^2 b_n}{\partial v^2} & \frac{\partial^3 b_n}{\partial v^3} & \frac{\partial^4 b_n}{\partial v^4} \\ \frac{\partial^2 b_m}{\partial v^2} & \frac{\partial^3 b_m}{\partial v^3} & \frac{\partial^4 b_m}{\partial v^4} \\ \frac{\partial^2 b_h}{\partial v^2} & \frac{\partial^3 b_h}{\partial v^3} & \frac{\partial^4 b_h}{\partial v^4} \end{pmatrix} (v, n, m, h).$$

Proposition

The weak Hörmander condition holds for all points $(v, n, m, h, \zeta)$ such that $D(v, n, m, h) \neq 0$. This is an open set of full Lebesgue measure.
• We then checked that the above determinant is non-zero at certain **attainable points** of the stochastic HH model (in the sense of deterministic control).

• Using arguments from control theory (in particular, Arnold and Kliemann 1987), and from Nagano 1966, Susmann 1973 (the weak Hörmander condition is transported by control paths, due to the analyticity of the coefficients of the system)

**Theorem (HLT 2016, to appear in Esaim P&S)**

*For the stochastic Hodgkin-Huxley system driven by an Ornstein-Uhlenbeck type process or a CIR type process, the weak Hörmander condition holds everywhere, and thus the process is hypoelliptic.*
Summary

• Hypoellipticity of stochastic HH-model is induced by the absence of (Brownian) noise in the *gating* (or: *memory*) variables $n$, $m$, $h$.

• Even if we may assume these gating variables to be subject to noise as well, the *channel noise* is most likely not to be of the same type.

• Dealing with hypoelliptic rather than with elliptic diffusions has drastical consequences for any probabilistic study. Most important: the *control theory* is not at all established and has to be done “by hand”.
Second example

Not any longer a model for a single neuron, but a system of interacting neurons.

Joint work with Susanne Ditlevsen (SPA, to appear in 2017)
We consider a large system of interacting Hawkes processes, describing each one neuron. That is, we describe the successive appearences of spikes of a given neuron.

This system is made of $n$ populations, $n$ is fixed.

Each population $k$ consists of $N_k$ neurons described by their counting processes

$$Z_{k,i}(t), 1 \leq i \leq N_k.$$ 

Within a population, all neurons behave in the same way. This is a mean-field assumption.
- Intensity of $i$–th neuron belonging to population $k$:

$$\lambda_{k,i}(t) = f_k \left( \frac{1}{N_l} \sum_{l=1}^{n} \sum_{1 \leq j \leq N_l} \int_{[0,t]} h_{kl}(t-s) dZ_{l,j}(s) \right).$$

- $f_k =$ jump rate function of population $k$; *supposed to be Lipschitz continuous*.

- $h_{kl}$ measures the influence of a typical neuron of population $l$ on a typical neuron of population $k$; *supposed to be in $L^2_{loc}(\mathbb{R}_+, \mathbb{R})$*. 
• Intensity of $i$–th neuron belonging to population $k$:

$$\lambda_{k,i}(t) = f_k \left( \sum_{l=1}^{n} \frac{1}{N_l} \sum_{1 \leq j \leq N_l} \int_{]0,t[} h_{kl}(t - s) dZ_{l,j}(s) \right).$$

• $f_k = \text{jump rate function of population } k; \text{ supposed to be Lipschitz continuous.}$

• $h_{kl}$ measures the influence of a typical neuron of population $l$ on a typical neuron of population $k; \text{ supposed to be in } L^2_{loc}(\mathbb{R}_+, \mathbb{R}).$

• We are in a mean field frame: population $l$ influences population $k$ only through its empirical measure.
In the $N \to \infty$ limit, a typical neuron of pop $k$ spikes at rate $\lambda^k_t$:

$$\lambda^k_t = f_k \left( \sum_{l=1}^{n} \int_0^t h_{kl}(t-u)\lambda^l_u du \right), 1 \leq k \leq n.$$  

**Theorem (with S. Ditlevsen, SPA 2017)**

*For memory kernels $h_{kl}$ which are Erlang kernels and for a specific interaction graph structure, the limit system has oscillatory behavior.*

This oscillatory behavior is mainly a consequence of the memory in the system (and of course also of the structure of the graph of interactions).
Hawkes memory

- Hawkes processes are truly infinite memory processes - the intensity depends on the whole history.
- Suppose $n = 1$ (only one population) and stochastic intensity (for fixed $N$, thus before passing to the limit)

$$\lambda(t) = f \left( \int_0^t h(t - s) d\bar{Z}_N(s) \right) =: f(X(t)), \quad \bar{Z}_N(s) = \frac{1}{N} \sum_{i=1}^N Z_i(s),$$

and $h$ is an Erlang kernel:

$$h(t) = c \frac{t^m}{m!} e^{-\nu t}, \quad \nu > 0, \ m \in \mathbb{N}_0, \ c \in \mathbb{R}.$$  

That means: The delay of influence of the past is distributed. It takes its maximum at about $m/\nu$ time units back in the past.
Developing the memory - continued

- $\lambda(t) = f(X(t))$. We show that $X(t)$ is Markov and almost a degenerate diffusion process.

- Suppose e.g. $h(t) = cte^{-\nu t}$ (memory length $m = 1$) and define
  
  $$h_1(t) := ce^{-\nu t}.$$  

  Then
  
  $$X(t) = \int_0^t h(t - s)d\tilde{Z}_N(s), \quad Y(t) = \int_0^t h_1(t - s)d\tilde{Z}_N(s)$$  

  is a two dimensional Markov process with
  
  $$dX_t = -\nu X_t + dY_t, \quad dY_t = -\nu Y_t dt + c \, d\tilde{Z}_N(t).$$
Developing the memory - continued

- $\lambda(t) = f(X(t))$. We show that $X(t)$ is **Markov** and almost a **degenerate** diffusion process.

- Suppose e.g. $h(t) = cte^{-\nu t}$ (memory length $m = 1$) and define

  \[
  h_1(t) := ce^{-\nu t}.
  \]

  Then

  \[
  X(t) = \int_{0}^{t} h(t - s) d\tilde{Z}_N(s), \quad Y(t) = \int_{0}^{t} h_1(t - s) d\tilde{Z}_N(s)
  \]

  is a two dimensional Markov process with

  \[
  dX_t = -\nu X_t + dY_t, \quad dY_t = -\nu Y_t dt + c \, d\tilde{Z}_N(t).
  \]

- For general Erlang kernels $h(t) = c \frac{t^m}{m!} e^{-\nu t}$ we obtain an $m + 1$-dimensional process...
Diffusion approximation

We approximate the small jumps by a Brownian motion. In this way each neuron jumps at rate \( f(\tilde{X}(t)) \) where

\[
\begin{align*}
    d\tilde{X}(t) &= -\nu \tilde{X}(t)dt + \tilde{Y}_t dt \\
    d\tilde{Y}(t) &= -\nu \tilde{Y}(t)dt + cf(\tilde{X}(t))dt + \frac{c}{\sqrt{N}} \sqrt{f(\tilde{X}(t))dB_t}
\end{align*}
\]

- Can be extended to the general case of \( n \) populations and of general Erlang memory kernels.
- We obtain a diffusion of high dimension driven by only few (actually, \( n \)) Brownian motions.
• The **degeneracy** of this diffusion is once more due to the memory variables of the system.

• Due to the **cascade structure** of the drift - coming from the development of the memory - it is easy to show that the diffusion satisfies the weak Hörmander condition.

• This cascade structure (a coordinate does only depend on itself and the following coordinate) is reminiscent of Delarue-Menozzi (2010) (density estimates for hypo-elliptic diffusions describing chains of reactions).
On the covariance matrix

• In this specific system, if the Erlang kernel is of order $m$, then we have $m + 1$ diffusion coordinates.

• The last is driven by Brownian motion: it travels at speed $t^{1/2}$.

• The before last one travels at speed $t^{3/2}$.

• And so on: the first at speed $t^{(2m+1)/2}$.

⇒ the process possesses a Lebesgue density, but its “geometry” is “flattened” compared to the elliptic case.

⇒ Moreover, it is not a priori clear that this density is positive – or more precisely, where exactly it is positive: Control problem!
In the above cascade situation, the cost functional $V_t(x, y)$ of the process (in the sense of deterministic control: cost of steering the process from $x$ to $y$ within a time interval $[0, t]$) behaves as

$$V_t(x, y) \sim t|T_t^{-1}(\varphi_t(x) - y)|^2,$$

where

$$T_t = \text{diag}(t^{m+1}, t^m, \ldots, t)$$

and where $\varphi_t(x)$ is the associated deterministic flow with zero noise.
Coming back to our system studied with Susanne:

- Using a convenient Lyapunov-function and the control theorem (and ideas inspired by the work we did with Michèle Thieullen and Reinhard Höpfner on the stochastic Hodgkin-Huxley system)

\[ \Rightarrow \] Under certain conditions on the structure of the graph of interactions:

\[ \exists \] attainable point (which is the unstable equilibrium of the limit dynamical system).

\[ \Rightarrow \] diffusion is recurrent in the sense of Harris, with unique invariant probability measure $\mu^N$. 
• Limit system is attracted to non constant periodic orbit (limit cycle)...

• Presence of noise: diffusion may escape from a tube around the limit cycle - after longer and longer periods $\Rightarrow$ natural to study the large deviations of the system. But:

As a consequence of the hypoellipticity, classical results on control theory needed as main ingredient for sample path large deviations in the small noise regime, à la Freidlin-Wentzell, are not granted and have to be checked “by hand”.
Theorem (Lö 2016)

\[
\mu^N(D) \sim Ce^{-N[\inf_{x\in D} W(x)]},
\]

where (in case of a periodic orbit \(\Gamma\) and unstable equilibrium \(x^*\)), the cost function is given by

\[
W(x) = V(\Gamma, x) \land [V(\Gamma, x^*) + V(x^*, x)].
\]

Here, \(V(x, y)\) is the cost of steering the process from \(x\) to \(y\) in any time.
Final comments

- Refractory period and age dependence?
- What about Hypoellipticity in the original Hawkes frame, without diffusion approximation (should be okay, Vandermonde determinant)


Thank you for your attention.