Mapping the Hodkgin-Huxley Neuron to an Escape Rate Neuron

Masterarbeit der Philosophisch-naturwissenschaftlichen Fakultät der Universität Bern

vorgelegt von

Roman Caspar

2013

Leiter der Arbeit
Prof. Dr. Walter Senn

Acknowledgments

I would like to express my gratefulness to my adviser *Prof. Dr. phil. nat. Walter Senn* for his willingness to start the first time experiment of a mathematical master thesis at the department of physiology of the university of Bern. I really enjoyed the insight in a topic of applied mathematics in research.

Many thanks also to *Dr. phil. nat. Robert Urbanczik* who contributed a lot to this thesis. Again and again he found some time to discuss problems and give ideas for solving them.

Moreover I would like to thank all the members of the department of physiology for letting me work in their premises, especially *Andrea Sieber* who shared her bureau with me during the last months.

Finally I am very grateful for all the people around me, who motivated and encouraged me to keep on working, even if I have been stuck at certain points.

Abstract

The frequently used neural model of Hodgkin and Huxley (HH) reproduces the dynamics in a neuron very well. However, it is not practicable to describe reinforcement learning of a neuron, of neural networks respectively. Since it is possible to emulate the spiking behavior of the HH-model under noisy input by the simpler escape noise model, the latter is commonly used to model reinforcement learning.
Thus, the aim of this thesis is to formally map the model of HH to an escape rate neuron. The attempt presented in this work uses a discrete time approximation of the noisy membrane potential as the sum of a deterministic potential given by the HH-model and a noise term, which is shown to reasonably be described by an Ornstein-Uhlenbeck (OU) process. It turns out that the desired mapping is not at all feasible for a gaussian OU process, but works at least for subthreshold stimulus if it is replaced by an Ornstein-Uhlenbeck-Lévy process, namely an OU process with stable increments with stability-index $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$. Thus some important results about stable distributions and the OUL process are presented and proven to allow a mapping from the HH model to an escape rate model in the case of subthreshold stimulus.

Contents

1 Introduction

The mystery of the human brain has driven research for centuries. Why is it able to memorize words, pictures and even very complex and abstract connections? These kind of questions have motivated natural scientists as well as philosophers to develop theories and perform experimental research which lead to a far deeper insight into the mechanism of the brain. However, there are still a lot of questions that are not yet answered. Especially the questions about memorizing and learning are still in research and there are quite different concurrent approaches to solve these problems.
During the 20th century natural scientists began to develop mathematical models for nerve cells (neurons) and networks, based on the growing knowledge of the biological structure, to eventually understand how people or other creatures can process information, how they learn and, finally, how they can save this knowledge for future tasks.

1.1 Physiological Background

Like other cells, neurons are surrounded by a membrane which separates the interior from the extracellular space. Because of this membrane, it is possible to have a different ion concentration inside the neuron than in the surrounding liquid. This difference in concentration generates an electrical potential V which drives the neural dynamics.
A neuron can be devided roughly into three functionally distinct parts, the dendrites, the soma and the axon (see figure 1). The dendritic tree can be seen as the signal receiving unit, the soma as the processing unit and the axon as the transmitting unit. Most of the input from other neurons is received via synapses at the dendritic tree. In these synapses the incoming electric signals, viz. spikes, trigger a complex biochemical process which generates either an excitatory or an inhibitory postsynaptic potential (PSP). However, the strength of the PSP triggered by a certain input signal depends heavily on the physical and neurochemical characteristics of each synapse. The fact that it is the synapses where the brain is most flexible, is the reason why they are believed to be the location of learning and memory [12]. The impact of the synapses on the strength of the PSP is interpreted as an adjustable weight w.
At the end of the dendritic tree, all the PSP's add up in the soma generating a change in the membrane potential. If the total input produces a membrane potential crossing a certain neural threshold, the soma fires a spike that is transmitted through the axon, which itself is connected dendrites of other neurons.

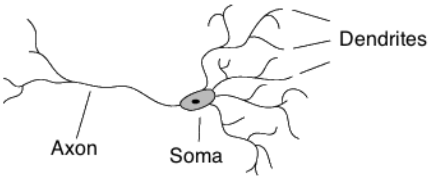

Figure 1: Schematic Neuron [28]

These observations motivate modeling the potential caused by incoming currents I_i by adding up the PSP's weighted by corresponding w_i's and letting the neuron spike if a threshold ϑ is reached.

1.2 Detailed vs. Formal Spiking Neural Models

The different attempts in modeling single neurons can be divided into two categories. The detailed neural models aim to describe the chemical processes in the cell, including the biophysics of ion-channels, dendritic trees, synaptic interaction and so on. A very popular example is the model proposed by Hodgkin and Huxley in 1952 which is presented in more detail in section 3. An argument for studying such complex models is the idea, that one has to consider all the anatomical and physiological details to eventually understand the nervous system. The problem of this concept is that such models are difficult to construct an often even more difficult to analyze. Typically, such models lead to a large number of nonlinear differential equations which make the model very expensive in finding numerical solutions. Already the model from Hodgkin and Huxley contains four nonlinear differential equations, even though it completely ignores the influence of the spacial structure.

The formal spiking neural models aim to reduce the complexity of the description of a neuron, while keeping the spiking nature and the essential elements of the behavior. The motivation for such models is that they allow studying more easily computational and functional principles of neural system such as the reinforcement learning from section 4. These simplified models cannot describe the entire range of functions of a neuron, so they usually focus on an essential feature and describe it as a function of what the modeler considers to be crucial. E.g. the escape noise model from section 2 describes the spiking as a function of the instantaneous membrane potential (and its slope) in order to perform reinforcement learning [9].

1.3 Aim of the Thesis

The models and learning processes treated in this thesis will be confined to a single neuron. As indicated above, the detailed neural model of Hodgkin and Huxley has the poor property, that it causes too much effort to be analyzed as to be feasible to describe reinforcement learning. However, simpler models such as the escape noise model can reproduce their spiking nature quite well if the input is noisy, which means that besides the desired input, there is always some stochastic input interpreted as background noise and implemented by a stochastic process ξ_t. For this reason the escape noise model is usually used as a simplification if applied for reinforcement learning. Thus the aim of this thesis is to give a formal reason for this simplification, by formally mapping the neural model from Hodgkin and Huxley to an escape noise model. Section 3.3 and all of section 7 contain the research parts of the thesis, while the other sections contain definitions and results from existing literature relevant for the research parts.

1.4 Idea

A previous attempt to perform the mapping from HH model to an escape rate model failed, because it assumed a simple voltage threshold [18, 19]. The idea of the mapping presented in this thesis is slightly different, since it assumes a threshold condition sensitive to the voltage V as well as its slope $\dot{V}$. This additional sensitivity on the slope is motivated by a reduction of the HH model shown in section 3.3.

Now the noisy potential $\tilde{V}$ and its slope $\dot{\tilde{V}}$ are expressed as the sum of their deterministic values V, $\dot{V}$ respectively, and a stochastic component, modeled by a discrete time approximation of a stochastic process ξ_t, namely an Ornstein-Uhlenbeck process, that vanishes in the limit $\Delta \to 0$. Due to the sensitivity to $\dot{V}$, the model can have an escape rate in the limit $\Delta \to 0$ even though the model gets deterministic. It will be shown that this turns the reduced HH model into an escape noise model in the limit $\Delta \to 0$ for subthreshold stimulus.

2 Escape Noise Neuron with Reset

The basic idea of the escape noise model is to split the input of a neuron into a deterministic input current, which may be seen as the input from a certain subset of neurons, and a stochastic input current, interpreted as background noise from the activity of the other neurons. This section shall give a brief insight into such models and describe why they are useful in terms of studying learning processes.

2.1 Definition

The escape noise neuron is a threshold model and thus spikes if the membrane potential reaches ϑ. The spikes are not modeled in detail but only described by a stereotype action potential. The stochastic input current yields the possibility that the neuron spikes even though the deterministic membrane potential has not yet reached the threshold. Since the exact value of the noise component can not be predicted, this gives a spiking probability, also called escape rate, depending on the instantaneous membrane potential of the neuron.

Definition 2.1. For $t > 0$, let V_t be the potential, I(t) the incoming current, τ_m a time constant. The *escape noise neuron with reset* is defined by

i) $V_0 = 0$

ii) $V_{\hat{t}} = 0$ if $\hat{t}$ spike time

iii) $\dot{V}_t = -\frac{V_t}{\tau_m} + I(t)$

and has an instantaneous escape rate $\varphi(V_t|\hat{t})$, in the following referred to as firing rate, where $\hat{t}$ is the time of the last spike, viz. reset.

Example. A common escape rate $\varphi(V_t|\hat{t})$ appears in the *Arrhenius model*

$$\varphi(V_t|\hat{t}) \sim \exp\left\{-\frac{(1-V_{t|\hat{t}})^2}{\sigma^2}\right\},$$

where the escape rate depends only on the distance between the threshold and the membrane potential.

Even though most authors use firing rates based on a pure voltage threshold, this concept is questionable. As discussed in section 3.2.2, the model from Hodgkin and Huxley indicates a dependence not only on the voltage V_t, but also on the slope $\dot{V}_t$. Thus, a more sophisticated version of the Arrhenius model could have a firing rate

$$\varphi(V_t, \dot{V}_t|\hat{t}) = w \exp\left\{-\frac{(1-(V_{t|\hat{t}}+\dot{V}_{t|\hat{t}}))^2}{\sigma^2}\right\}.$$

However, to keep notations simple, the slope will not be mentioned where there is no explicit reference to it.

2.2 Properties

By construction, an escape noise neuron has an instantaneous escape rate, that is independent of the history of the membrane potential. Therefore, the spike process generated by this neuron is an inhomogeneous Poisson process.

2.2.1 Spiking as a Poisson Process

Definition 2.2. A *stochastic process* is a collection $\{X_t; t \in T\}$ of random variables X_t, defined on the same probability space.

Definition 2.3. [4] A *counting process* $\{C_t; t \geq 0\}$ is a stochastic process that satisfies
i) $C_t \geq 0$
ii) C_t is an integer
iii) C_t is non-decreasing
for all $t \geq 0$.

According to these preliminary definitions, an inhomogeneous Poisson process can be defined as

Definition 2.4. [4] An *inhomogeneous Poisson process* is a continuous time counting process $\{N_t; t \geq 0\}$ with the properties
i) $N_0 = 0$
ii) N_t increases by jumps of size 1 almost sure
iii) for any $t, s \geq 0$, the increment $N_{t+s} - N_t$ is independent of the history $\{N_u; u \leq t\}$

Proposition 2.5. *[5] The number of events of an inhomogeneous Poisson process in an interval* $(a, b]$ *has a Poisson distribution with parameter*

$$\lambda_{a,b} = \int_a^b \lambda(t)dt,$$

where $\lambda(t)$ *is the time dependent poisson rate.*

The Poisson process generated by the escape noise neuron counts the number N of spikes and has the Poisson rate $\varphi(V_t)$.

2.2.2 Interspike Interval Distribution

It is common to characterize noisy neuron models by their interspike interval (ISI) distribution $P_I(t|\hat{t})$ depending on the incoming current I(t) and a given spike at $\hat{t}$.

Definition 2.6. Let $\hat{t}$ be the time of the last spike. Then the *interspike interval distribution* $P_I(t|\hat{t})$ of the escape noise neuron with reset is

$$P_I(t|\hat{t}) = \varphi(V_t|\hat{t}) \exp\left\{ -\int_{\hat{t}}^t \varphi(V_{t'}|\hat{t})dt' \right\}$$

Remark. The exponential term in this definition can be interpreted as the probability of the neuron staying quiescent from $\hat{t}$ to t. Thus $P_I(t|\hat{t})$ is the product of the firing rate at time t and the probability of not firing in $(\hat{t}, t)$.

3 Hodgkin-Huxley Model

Named after its developers Alan Lloyd Hodgkin an Andrew Fielding Huxley, who presented the model in 1952, the HH model describes the dynamics of a neuron. It is based on their research on the giant axon of the squid. They showed that the consistence of the membrane guarantees a nearly perfect electrical insulation. However, some specific proteins in the membrane act as ion gates and enable the ion concentration in the cell to change. The HH model now describes the voltage-dependent ion channels and the changes of the membrane potential caused by the currents through the channels and some incoming current $I(t)$.

3.1 Definition

Hodgkin and Huxley found in their experiments three different types of ion currents: an inward sodium (Na$^+$) current, an outward potassium (K$^+$) current and a leakage current (l) that consists mainly of Cl$^-$ ions. They describe the membrane as an electrical circuit, where the membrane potential is replaced by a capacitor C_m, and the ion channels by parallel resistances (see figure 2).

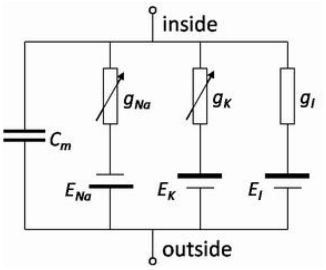

Figure 2: Diagram of the HH model [26]

If an external current I(t) is applied to the cell, the law of conservation of electric charge leads to:

$$I(t) = I_{C_m}(t) + I_{\text{Na}}(t) + I_K(t) + I_l(t) \tag{1}$$

The definition of the capacity $C_m = Q/V$ where Q is the charge and V the voltage across the capacitor yields $I_{C_m}(t) = dQ/dt = CdV/dt$ and thus

$$C_m \frac{dV}{dt} = -\sum_k I_k(t) + I(t) \qquad k = \text{Na,K,}l \tag{2}$$

The HH model describes the three ion channels not by resistances R_i but by conductances g_i, which is equivalent since $g_i = 1/R_i$. While the leakage channel has a constant conductance g_l, the other ion channels have a maximum conductance g_{Na} or g_{K}, respectively. However, only parts of the sodium an potassium channels are open at the same time. The ratio of open channels is by additional voltage dependent variables m,n and h. Hodgkin and Huxley formulated their model of the three components as follows:

$$\sum_k I_k(t) = g_{\text{Na}} m^3 h(V - E_{\text{Na}}) + g_{\text{K}} n^4 (V - E_{\text{K}}) + g_{\text{L}}(V - E_l) \tag{3}$$

The parameters E_{Na}, E_{K} and E_l are the equilibrium potentials, whose values, as well as those of the conductances, are basically results of experiments by Hodgkin and Huxley, even though some of them have been slightly corrected after further research.

Table 1: Empirical parameters of the HH model [9]

x	E_x (mV)	g_x (mS/cm^2)
Na	115	120
K	-12	36
L	10.6	0.3

The gating varbiables m,n and h are defined by the following differential equations

$$\begin{aligned} \dot{m} &= \alpha_m(V)(1-m) - \beta_m(V)m \\ \dot{n} &= \alpha_n(V)(1-n) - \beta_n(V)n \\ \dot{h} &= \alpha_h(V)(1-h) - \beta_h(V)h \end{aligned} \tag{4}$$

where the voltage dependent functions α_k and β_k are empirical functions, adjusted by Hodgkin and Huxley to fit the experimental data as good as possible (see table 2).

Table 2: Empirical functions depending on V (mV) [9]

x	$\alpha_x(V)$	$\beta_x(V)$
n	$\frac{(0.01V-0.1)}{1-\exp\{1-0.1V\}}$	$\frac{0.125}{\exp\{V/80\}}$
m	$\frac{(0.1V-2.5)}{1-\exp\{2.5-0.1V\}}$	$\frac{4}{\exp\{V/18\}}$
h	$\frac{(0.07)}{\exp\{0.05V\}}$	$\frac{1}{1+\exp\{3-0.1V\}}$

Definition 3.1. [9] The *Hodgkin-Huxley model* describes the membrane potential of a neuron through the equations (2)-(4) with the empirically fitted values from table 1 and functions from table 2.

3.2 Properties

3.2.1 Dynamics

Obviously, the dynamics of the model is given by the behaviour of m,n and h. Figure 3 shows that m and n are increasing as V increases, while h is decreasing. In the figure $m_\infty(V), n_\infty(V), h_\infty(V)$ are the asymptotic values of m,n and h for a fixed potential V as $t \to \infty$.

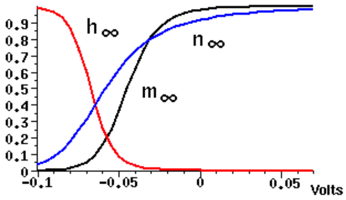

Figure 3: Voltage dependent gating variables [27]

Interesting is the fact, that these changes follow different time scales. This is obvious if the equations (4) are rewritten as

$$\begin{aligned} \dot{m} &= \frac{m_\infty(V) - m}{\tau_m(V)} \\ \dot{n} &= \frac{n_\infty(V) - n}{\tau_n(V)} \\ \dot{h} &= \frac{h_\infty(V) - h}{\tau_h(V)} \end{aligned} \tag{5}$$

where $m_\infty(V), n_\infty(V), h_\infty(V)$ are the asymptotic values of m,n and h for a fixed potential V as $t \to \infty$. The time constants $\tau_{m,n,h}$ indicate, how fast the gating variables react on a change of the potential.

$$m_\infty(V), n_\infty(V), h_\infty(V) = \frac{\alpha_{m,n,h}(V)}{\alpha_{m,n,h}(V) + \beta_{m,n,h}(V)} \tag{6}$$

$$\tau_{m,n,h} = \frac{1}{\alpha_{m,n,h}(V) + \beta_{m,n,h}(V)} \tag{7}$$

Figure 4 shows, that τ_m is always much smaller than τ_n and τ_h.

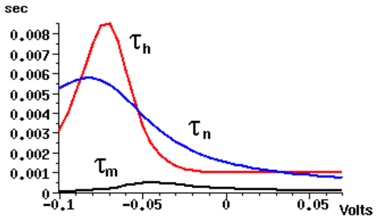

Figure 4: Time constants of the gating variables [27]

Thus, if an incoming current is rising V, m opens the sodium channels quickly and the I_{Na} current increases, whereas the variable h will close the channels and stop the I_{Na} with a certain delay. On a similar slow time scale, the variable n activates the potassium current, which lowers the potential. The effect of these two currents is a short action potential (spike) due to the fast activation of I_{Na}, followed by a negative overshoot due to the slow timescale of n.

3.2.2 Threshold Behavior

Not every incoming current generates a spike. $I(t)$ has to rise V sufficiently for m to allow a strong sodium current, and at the same time it has to rise V fast enough, such that h is not yet closing the channels and n does not activate the voltage lowering current I_K. Thus the HH neuron spikes only, if V and $\dot{V}$ are higher than some critical value, what motivates to describe the HH model as a threshold model [9].

3.3 Reduction

Even though it is very close to the biological processes in the soma, the HH model is not really feasible for describing reinforcement learning, since it is very laborious. In this context, the substantial interest is to describe the spiking behavior of the neuron (see section 4). This section motivates several reductions that should provide the main characteristics of the HH model in the spiking behavior.

3.3.1 Reduction of the Gating Variables

As seen in section 3.1, there are three types of ion gates defining the dynamics of the HH model. A first step of simplification is to eliminate the leakage current, which has a rather small influence, since the conductance g_l is relatively small (see table 1). Secondly the observations of section 3.2 show that, since I_{Na} is the only inward current, the sodium channel is responsible for generating a spike. Thus, even though eliminating the potassium current I_{K} will heavily change the after spike behavior of the model, the threshold property from section 3.2.2 survives due to the different dynamics of the gating variables m and h.

Over all, this first reduction leads to a simpler version of equations (2) and (3):

$$C_m \frac{dV}{dt} = -I_{\text{Na}}(t) + I(t) \tag{8}$$

where

$$I_{\text{Na}}(t) = g_{\text{Na}} m^3 h (V - E_{\text{Na}}) \tag{9}$$

3.3.2 Simplification of the Gating Variables

This reduced model still contains two differential equations for the variables m and h. While they describe the dynamics of these variables very precise, they prevent an efficient handling of the model. The following approximations yield a much easier version of equation (9):

The dynamics of the gating variables m and h indicate, that an inactivation by h can only happen, if V changes on a slow time scale compared to τ_m. Thus assume $m(V_t) \approx m_\infty(V_t)$, $h(V_t) \approx h_\infty(V_t)$ and let V_t change to $V_{t+\Delta t} = V + \Delta V$ in time $\Delta t \approx \tau_m$, with ΔV small to yield the desired slow changing potential. Then $m(V_{t+\Delta t})$ can be approximated by

$$m(V_{t+\Delta t}) \approx m_\infty(V_{t+\Delta t}) \tag{10}$$

since m changes on a similar time scale as V. In contrast, since $\Delta t \ll \tau_h$, h reacts a lot slower on changes in V and thus approximately stays at $h(V_t)$:

$$h(V_{t+\Delta t}) \approx h(V_t) \tag{11}$$

Thus by assumption

$$\begin{aligned} h(V_{t+\Delta t}) &\approx h_\infty(V_t) \\ &= h_\infty(V_{t+\Delta t} - (V_{t+\Delta t} - V_t)) \\ &= h_\infty(V_{t+\Delta t} - \tau_m \frac{V_{t+\Delta t} - V_t}{\tau_m}) \end{aligned}$$

Now $\Delta V / \tau_m \approx \Delta V / \Delta t =: \dot{V}_{t+\Delta t}$ yields

$$h(V_{t+\Delta t}) \approx h_\infty(V_{t+\Delta t} - \tau_m \dot{V}_{t+\Delta t}) \tag{12}$$

Thus the following reformulation of (9) is a feasible approximation for a potential V, that changes slowly compared to τ_m:

$$I_{\text{Na}}(t) = g_{\text{Na}} m_\infty^3(V) h_\infty(V - \tau_m \dot{V})(V - E_{\text{Na}}) \tag{13}$$

3.3.3 Threshold Behavior of the Reduction

As in the original model (see section 3.2.2), the reduction (13) emits a spike if I_{Na} is strong, which happens if V is high enough to increase m and at the same time $\dot{V}$ is high enough to ensure h not to close the channels. To get a simple threshold condition, $I_{\text{Na}}(V, \dot{V})$ is approximated linearly in V and $\dot{V}$ around $(V_0, \dot{V}_0)$:

$$I_{\text{Na}}(V, \dot{V}) \approx c + \frac{\partial I_{\text{Na}}}{\partial V}(V - V_0) + \frac{\partial I_{\text{Na}}}{\partial \dot{V}}(\dot{V} - \dot{V}_0) \tag{14}$$

Thus the threshold condition $I_{\text{Na}}(V, \dot{V}) > \theta$ yields

$$\frac{\partial I_{\text{Na}}}{\partial V}(V - V_0) + \frac{\partial I_{\text{Na}}}{\partial \dot{V}}(\dot{V} - \dot{V}_0) > \theta - c \tag{15}$$

which is basically equivalent to $aV + b\dot{V} > \theta$. Thus it is straight forward to replace it by

$$V + \dot{V} > 1 \tag{16}$$

Since the after-spike course of the membrane potential is again stereotype, it is replaced by a reset as in the escape rate model.

3.4 Reduced HH as Dynamic Integrate&Fire Neuron

Summarysing the steps from the previous section, the reduced model that will be treated in the following is defined by:

Definition 3.2. Let $I(t)$ be the incoming current, $I_{\text{Na}}(t)$ the inward sodium current and $m_\infty(V)$, $h_\infty(V)$ the gating variables defined by (6).
The *dynamic I&F model* (dI&F) is defined through the threshold condition

$$V_t + \dot{V}_t > 1$$

eliciting a spike when $V_t + \dot{V}_t$ crosses the threshold 1 from below. After a spike at time $\hat{t}$ the potential as well as its derivative are reset to 0:

$$V_{\hat{t}} = \dot{V}_{\hat{t}} = 0$$

V_t and $\dot{V}_t$ are given by the differential equation

$$C_m \frac{dV_t}{dt} = -I_{\text{Na}}(t) + I(t)$$

where $I_{\text{Na}}(t) = g_{\text{Na}} m_\infty^3(V_t) h_\infty(V_t - \tau_m \dot{V}_t)(V_t - E_{\text{Na}})$.

As mentioned in section 1.3 the escape noise neuron is only used if the input is noisy. Thus, the deterministic dI&F model will be analyzed for stochastic input in section 7.

4 Reinforcement Learning

This section gives a short introduction into the main ideas of the well studied concept of reinforcement learning and shows the advantage of an instantaneous escape rate model over the HH-model.

4.1 Definition

Empirical research lead to the conclusion, that the information in neuronal signals is neither represented by the individual spiking patterns nor by the spike timing. In fact experiments gave evidence for the spike firing rate to be the information carrier. Thus the considered learning rules and neuron models are defined by incoming and emitted firing rates.
Threshold neurons can be seen as decision units who control their firing rate y based on the incoming firing rates x_i of the connected neurons. Reinforcement learning now assumes that the neuron gets a feedback on its response y, that measures the quality of y compared to a target response $\hat{y}$ in terms of a reward function $R(y,x)$.

Definition 4.1. Let $X \subset \mathbb{R}_+^n$ be the set of possible of incoming patterns of firing rates, $Y \subset \mathbb{R}_+$ the set of response firing rates and $\hat{y} \in \mathbb{R}_+$ the target firing rate.
Then $R : Y \times X \to \mathbb{R}$ such that for $x \in X$ fixed

i) $R(\hat{y},x) \geq R(y,x)\ \forall y \in Y$

ii) $R(y,x) = R(\hat{y},x) \iff y = \hat{y}$

is a *reward function*.

Synaptic weights are adapted to generate neuronal responses such that the reward increases, is maximized respectively, meaning the neurons response reaches the target $\hat{y}$.
The process of generating a response y to an input x depends on synaptic weights w_i (see section 1.1), which allow the neuron to control the influence of an input x_i, and a threshold condition

$$\sum_n x_i w_i = x \cdot w > \theta$$

Thus learning, viz. increasing the reward, is a matter of adapting the weights w properly. Since the response y is only stochastically depending on the input x, the expected reward

$$E[R] = \sum_{x,y} R(y,x) P_w(y,x) = \sum_{x,y} R(y,x) P(x) P_w(y|x) \tag{17}$$

should be maximized.

4.1.1 Gradient Descent

To make sure that a change of w actually causes an improvement, the learning rule will be defined using method of gradient descent.

Definition 4.2. Let $F(w) : \mathbb{R}^n \to \mathbb{R}$ be $C^1(\mathbb{R}^n)$, with a local minimum at $\hat{w} \in \mathbb{R}^n$, $w_0 \in \mathbb{R}^n$ a guess for the minimum and $\lambda_i > 0$.
Iterating the rule

$$w := w - \lambda_i \nabla F(w)$$

defines the method of *gradient descent*.

Remark. The method of gradient descent acts as a method for finding a local maximum of a function by changing the direction of the defining rule to

$$w := w + \lambda_i \nabla F(w),$$

which will be denoted as learning rule $\Delta w \propto \nabla F(w)$.

The function to be maximized is $F(w) = E[R]$, given by (17), which yields

$$\begin{aligned}
\Delta w &\propto \frac{\partial}{\partial w} \sum_{x,y} R(y,x) P(x) P_w(y|x) \\
&= \sum_{x,y} R(y,x) P(x) \frac{\partial}{\partial w} P_w(y|x) \\
&= \sum_{x,y} R(y,x) P(x) P_w(y|x) \frac{1}{P_w(y|x)} \frac{\partial}{\partial w} P_w(y|x) \\
&= \sum_{x,y} R(y,x) P_w(y,x) \frac{\partial}{\partial w} \log P_w(y|x). \qquad (18)
\end{aligned}$$

The learning rule (18) is not very practicable, since evaluating the sum might be very expensive. Moreover, reinforcement learning as described above requires adaption of the weights after every incoming pattern, whereas (18) uses all possible pairs (y,x) in every single adaption.

4.1.2 Stochastic Gradient Descent

The solution solving the problem of the gradient descent described above is called stochastic gradient descent:

Definition 4.3. Let $F(w) = \sum_i F_i(w)$ and therefore $\nabla F(w) = \sum_i \nabla F_i(w)$.
Adapting w by evaluating the gradient at a single pattern (y,x)

$$w := w + \lambda_i \nabla F_i(w),$$

is the method of *stochastic gradient descent*.

Under some assumptions to the function F, the stochastic gradient descent guarantees the convergence of w.

Theorem 4.4. *[2] Let $F(w)$ have the following properties:*

i) $F(w) \in C^3(\mathbb{R}^n)$ and bounded from below

ii) The learning rates λ_i satisfy: $\sum_i \lambda_i = \infty$ and $\sum_i \lambda_i^2 < \infty$

iii) $E_x(\|(x,w)\|^k) \leq A_k + B_k\|w\|^k$ for $A_k,\ B_k > 0$, $k = 2,3,4$

iv) for some $D > 0$: $\inf_{w^2 \geq D} w\nabla F(w) > 0$

Then the stochastic gradient descent guarantees:

1. *The weights w_t are confined in a bounded region of $\mathbb{R}^n$ with probability 1.*
2. *$F(w_t)$ converges almost surely*

$$F(w_t) \xrightarrow[t\to\infty]{a.s.} F_\infty$$

3. *$\nabla F(w_t)$ converges almost surely to 0*

$$\nabla F(w_t) \xrightarrow[t\to\infty]{a.s.} 0$$

Remark. Since the function $F(w)$ may have several local extremal points (and even asymptotic plateaus), the stochastic gradient descent only guarantees the convergence to an extremal point, which may be different from the global extremal point. At least, the possibility of diverging on an asymptotic plateau is excluded by the confinement result.

Thus the learning rule is defined by (18) according to the stochastic gradient descent:

Definition 4.5. The reinforcement *learning rule* for adaption of the weights w of a neuron is

$$\Delta w \propto R(y,x)\frac{\partial}{\partial w}\log P_w(y|x)$$

4.1.3 Reinforcement Learning Process

After this preliminary definitions, reinforcement learning is defined as follows:

Definition 4.6. [10] A *reinforcement learning* process consists of

i) set of inputs $X \subset \mathbb{R}^n_+$

ii) set of responses $Y \subset \mathbb{R}_+$

iii) an initial weight vector $w \in \mathbb{R}^n$

iv) set of probability distributions $P_w(y, x)$ over X

v) reward function $R(y, x) : Y \times X \to \mathbb{R}$

vi) learning rule $\Delta w \propto R(y, x)\frac{\partial}{\partial w} \log P_w(y|x)$.

The system is learning by sampling $x \in X$ and adjusting the weights by Δw, depending on the response y . Theorem 4.4. shows the convergence under some conditions for $F(w) = E[R(y, x)]$.

4.2 Advantage of Escape Noise over Hodgkin-Huxley

The motivation to map the reduced version of the Hodgkin-Huxley model to an escape noise neuron is the need of an explicit formula for the conditional probability $P_w(y|x)$, in order to study the learning process of a single neuron as well as of a neuronal network. While this is possible in the case of an escape noise neuron, due to the property that the intervals between two spikes are independent and distributed according to definition 2.6, it is in general impossible for any integrate-and-fire models, such as the HH-model, since they heavily depend on the history of the membrane potential V_t.
For an escape noise neuron, the conditional probability $P_w(y, x)$ can be defined as follows:

Definition 4.7. A *spike train* x_i is a set of spiking times $t^1 < t^2 < \cdots < t^n$

$$x_i := \{t^1_{x_i}, t^2_{x_i}, \ldots, t^n_{x_i}\}.$$

The potential generated by an incoming spike train x is the sum of the postsynaptic potentials of the spikes:

$$\Delta V^x_t = w_i x^\varepsilon_i(t) \tag{19}$$

$$= w_i \sum_{t^j_{x_i} \in x_i} PSP(t - t^j_{x_i}) \tag{20}$$

and thus, the membrane potential at time t is

$$V_t = \sum_i w_i \sum_{t^j_{x_i} \in x_i} PSP(t - t^j_{x_i}) \tag{21}$$

where in the first sum i runs over all connected neurons. Note that V_t depends only on the incoming spike trains $\{x_i\}$ and the weights w.

Proposition 4.8. *Let $X = \{x_i\}$ be given a pattern of incoming spike trains. Then the probability of a spike train y of an escape noise neuron in an interval [0,T] is*

$$P(y|X) = \exp\left\{-\int_0^T \varphi(V_t)dt\right\} \prod_{t^{sp} \in y} \varphi(V_{t^{sp}})$$

where V_t is given by (21).

Proof. The spike train y can be interpreted as a sequence of ISI $I_1, I_2, \dots, I_n$ and a quiescent interval [t^n,T], with $I_j = t^j - t^{j-1}$ and $t^0 = 0$. As seen before, these ISI's are independent and it follows:

$$\begin{aligned}
P(y|X) &= \exp\left\{-\int_{t^n}^{T} \varphi(V_t)dt\right\} \prod_{j=1}^{n} P(t^j|t^{j-1}) \\
&= \exp\left\{-\int_{t^n}^{T} \varphi(V_t)dt\right\} \prod_{j=1}^{n} \varphi(V_{t^j}) \exp\left\{-\int_{t^{j-1}}^{t^j} \varphi(V_t)dt\right\} \\
&= \exp\left\{-\int_{t^n}^{T} \varphi(V_t)dt\right\} \exp\left\{-\sum_{j=1}^{n} \int_{t^{j-1}}^{t^j} \varphi(V_t)dt\right\} \prod_{j=1}^{n} \varphi(V_{t^j}) \\
&= \exp\left\{-\int_{0}^{T} \varphi(V_t)dt\right\} \prod_{j=1}^{n} \varphi(V_{t^j})
\end{aligned}$$

□

5 Ornstein-Uhlenbeck Processes

Many processes in physics, biology and financial mathematics are usually described by an Ornstein-Uhlenbeck (OU) process. While for a long time such phenomena have been assumed to result from gaussian distributed processes, there has been some recent research that comes to the conclusion that some applications have more likely a heavy tailed distribution, particularly a stable distribution (see section 5.3.1) [3]. In fact section 7 shows that in order to map the HH-model to an escape noise model, it is necessary to assume heavy tailed noise. Therefore main results for the standard case of the gaussian OU process as well as for the OU process for arbitrary stable distributions are presented in this section.

5.1 Notations and Preliminary Definitions

Definition 5.1. A stochastic Process $(X_t, t \geq 0)$ has *independent increments* if for any sequence $(t_i)_{i=1}^n$ with $0 \leq t_1 < t_2 < \cdots < t_{n-1} < t_n < \infty$

$$X_{t_2} - X_{t_1},\ X_{t_3} - X_{t_2},\ \dots\ ,\ X_{t_n} - X_{t_{n-1}}$$

are independent.

Definition 5.2. A stochastic Process $(X_t, t \geq 0)$ has *stationary increments* if for any $\tau > 0$ the distribution of $X_{t+\tau} - X_t$ is independent of t.

5.2 Ornstein-Uhlenbeck Process

The OU process can be motivated as a process describing the impact of the environment to the velocity of a Brownian particle. The random impacts of neighbouring particles is described by a Wiener process.

Definition 5.3. [23] A *standard Wiener process* W_t satisfies:
- i) $W_0 = 0$
- ii) W_t as independent increments
- iii) $W_t - W_s \sim \mathcal{N}(0, t-s)$ for $0 \leq s \leq t$

According to this definition, a time discrete Wiener process can be written as

$$(\Delta W_t) = W_{t+\tau} - W_t = \sqrt{\tau}\eta_t \tag{22}$$

where $\eta_t \sim \mathcal{N}(0,1)$.

An OU process as indicated above is therefore defined as

Definition 5.4. [5] Let $\beta, \sigma, \tau > 0$. A time discrete *OU process* ξ_t with zero mean solves

$$(\Delta \xi_t) = -\beta \xi_t \tau + \sigma(\Delta W_t)$$

and therefore satisfies

$$\xi_{t+\tau} - \xi_t = -\beta \xi_t \tau + \sigma \sqrt{\tau}\eta_t$$

The OU process ξ_t has again a gaussian distribution

Proposition 5.5. *Let ξ_t be an OU process as defined above and $\xi_0 = 0$. Then*

$$\xi_t \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})\right)$$

and ξ_t has the asymptotic stationary distribution $N(0, \frac{\sigma^2}{2\beta})$.

Remark. The proof of proposition 5.5 is analog to the proof of proposition 5.22 and thus is omitted.

5.3 Lévy Process

Definition 5.6. [17] A stochastic process X_t with $X_0 = 0$ a.s. is called *Lévy process*, if
- i) X_t has independent increments
- ii) X_t has stationary increments
- iii) X_t is continuous in probability: $\forall s, \epsilon > 0 \lim\limits_{s \to 0} P(|X_{t+s} - X_t| > \epsilon) = 0$

The class of Lévy processes contains a wide range of stochastic processes. Obviously the Wiener process is a Lévy process. However, since there is no condition on the distribution of the increments of a Lévy process, the increments can be changed to have distributions other than the gaussian, namely stable distributions, defining again a Lévy process, though with heavily different properties.

5.3.1 Definition of Stable Distributions

An important property of gaussian random variables is that their sum is again gaussian. Since stochastic processes ξ_t like the Wiener and the OU process are basically a sum of the increments, distributions without this property would in general make it impossible to find the distribution of ξ_t. In the following, the special case of degenerate distributions is excluded, since most statements about stable distribution do not hold in this case. Moreover, there is no use for a degenerate distribution in terms of noise.

Definition 5.7. [16] Let $X, X_1, \dots, X_n$ i.i.d. The distribution of X is called *stable*, if
$\forall n \in \mathbb{N}\ \exists c_n > 0, d_n$ s.t.

$$X_1 + \cdots + X_n \overset{d}{=} c_n X + d_n$$

The symbol $\overset{d}{=}$ means equality in distribution.
A distribution is *strictly stable* if this holds for $d_n = 0\ \forall n \in \mathbb{N}$.

Remark. Obviously, the gaussian distribution is stable and even strictly stable if it has zero mean. The only other stable distribution on $\mathbb{R}$, that can be written in a closed form, is the Cauchy distribution.

As indicated in the remark, most stable distributions can not be written in a closed form. However, they can be parametrized in a rather simple way by its characteristic function $\phi(k) = E[\exp\{ikx\}]$. The function $\phi(k)$ completely determines the distribution of X.

Proposition 5.8. *A random variable X is stable if and only if $X \overset{d}{=} \gamma Z + \delta$, where $\gamma > 0$, $\delta \in \mathbb{R}$ and Z has the characteristic function*

$$\phi(k) = \begin{cases} \exp\left\{-|k|^\alpha[1 - \text{sign}(k) i\beta \tan(\frac{\pi\alpha}{2})]\right\} & \alpha \neq 1 \\ \exp\left\{-|k|[1 + \text{sign}(k) i\beta \frac{2}{\pi} \log(|k|)]\right\} & \alpha = 1 \end{cases}$$

for $\alpha \in (0, 2]$ and $\beta \in [-1, 1]$.

Note that in the case of a symmetric distribution around zero $\beta = 0$ and $\delta = 0$. This gives the simpler form

$$\phi(k) = \exp\{-\gamma^\alpha |k|^\alpha\} \tag{23}$$

for the characteristic function of γZ. Since the noise term of the stochastic HH neuron (see section 7.1) will always be assumed symmetric and centered at zero, only the simpler case (23) will be treated to avoid irrelevant technical difficulties.

Definition 5.9. Let X be a stable random variable, with a symmetric distribution centered at zero. Then X is $S(\alpha, \gamma)$ if it as the characteristic function

$$\phi(k) = \exp\{-\gamma^\alpha |k|^\alpha\} \tag{24}$$

with $\alpha \in (0, 2]$ and $\gamma > 0$.

Thus, there are the two parameters α and γ left, that define the characteristic function. While γ is a scale parameter similar (but not equal) to the standard deviation in the gaussian case, α is the *index of stability*.

Example. The characteristic function of a gaussian distribution $N(0, \sigma^2)$ is

$$\phi(k) = \exp\left\{-\frac{1}{2}\sigma^2 k^2\right\}.$$

Thus, a random variable $X \sim S(2, \gamma)$ has a gaussian distribution $X \sim N(0, 2\gamma^2)$.

5.3.2 Properties of Stable Distributions

Even though there are no explicit formulas for most stable distributions, it is possible to state theoretical properties.

Theorem 5.10. *[16] All $S(\alpha, \gamma)$ distributions are continuous with a density that is $C^\infty(\mathbb{R})$.*

Unlike the gaussian distribution, stable distribution for $\alpha \neq 2$ are heavy tailed, meaning that the probability $P(X > x)$ decays not exponentially but only by a power-law as $x \to \infty$.

Theorem 5.11. *[16] Let $X \sim S(\alpha, \gamma)$ with $\alpha \in (0, 2)$ and $\gamma > 0$.*
Then as $x \to \infty$,

$$P(X > x) \sim c_\alpha \gamma^\alpha \frac{1}{x^\alpha} \tag{25}$$

$$f(x|\alpha, \gamma) \sim \alpha c_\alpha \gamma^\alpha \frac{1}{x^{1+\alpha}} \tag{26}$$

where $f(x|\alpha, \gamma)$ denotes the density and $c_\alpha = \sin(\frac{\pi\alpha}{2})\Gamma(\alpha)/\pi$. Since $S(\alpha, \gamma)$ distributions are symmetric, the analog properties hold for $P(X < -x)$ and $f(-x|\alpha, \gamma)$ as $x \to -\infty$.

Note that here $h(x) \sim g(x)$ as $x \to a$ means $\lim_{x \to a} h(x)/g(x) = 1$. Due to this slow decay, the existence of the mean and the variance depends on the value of α.

Proposition 5.12. *[16] Let $X \sim S(\alpha, \gamma)$ and $\alpha \in (0, 2)$.*
Then $E[|X|^p] < \infty$ if $p \in (0, \alpha)$ and $E[|X|^p] = \infty$ if $p \geq \alpha$

Thus the mean and the variance are given by

Lemma 5.13. *Let $X \sim S(\alpha, \gamma)$. Then*

$$E[X] = \begin{cases} 0 & \alpha \in (1, 2] \\ \textit{not defined} & \alpha \in (0, 1] \end{cases} \qquad Var(X) = \begin{cases} 2\gamma^2 & \alpha = 2 \\ \infty & \alpha \in (0, 2) \end{cases}$$

The distribution of sums or weighted sum of stable random variables are given in an intuitive way, given by proposition 5.14.

Proposition 5.14. *[16] The $S(\alpha, \gamma)$ distribution has the following properties*

a) If $X \sim S(\alpha, \gamma)$, then for $a \neq 0$

$$aX \sim S(\alpha, |a|\gamma)$$

b) The characteristic functions, densities and distribution functions are jointly continuous in x, k respectively, and both parameters (α, γ).

c) If $X_1 \sim S(\alpha, \gamma_1)$ and $X_2 \sim S(\alpha, \gamma_2)$ are independent, then

$$X_1 + X_2 \sim S(\alpha, \gamma)$$

where $\gamma^\alpha = \gamma_1^\alpha + \gamma_2^\alpha$.

From these properties follows immediately

Lemma 5.15. *Let $X_i \sim S(\alpha, \gamma_i)$. Then*

$$X_1 + \cdots + X_n \sim S(\alpha, (\gamma_1^\alpha + \cdots + \gamma_n^\alpha)^{\frac{1}{\alpha}})$$

Since the classical central theorem only holds for $\alpha = 2$, otherwise the variance is infinite by Proposition 5.13 and the conditions of the theorem are not satisfied, this very simple form of a sum of i.i.d. stable random variables motivates the question wether there is a similar property in the case $\alpha < 2$.

Theorem 5.16. Classical central limit theorem
Let $X_1, X_2, \ldots$ be i.i.d. with $E[X] = 0$ and Var(X) $= \sigma^2 < \infty$. Then

$$\frac{1}{\sigma\sqrt{n}}(X_1 + \cdots + X_n) \xrightarrow{d} Z \sim \text{N(0,1) as } n \to \infty$$

In fact, the stability property allows to state a generalized version of the central limit theorem that holds for all $\alpha \in (0, 2]$. Again, the distributions are assumed to be symmetric and centered at zero.

Theorem 5.17. [16] Generalized central limit theorem
Let $\alpha \in (0, 2]$. A random variable Z is stable if and only if there is a sequence $X_1, X_2, \ldots$ i.i.d. and constants $a_n > 0$ such that

$$a_n(X_1 + \cdots + X_n) \xrightarrow{d} Z \text{ as } n \to \infty$$

The generalized central theorem states, that the only possible distribution being the asymptotic distribution of a scaled sum $X_1 + \cdots + X_n$ of i.i.d. random variables is a stable distribution.

Definition 5.18. A random variable X is in the *domain of attraction* of Z if and only if there are constants $a_n > 0$ such that

$$a_n(X_1 + \cdots + X_n) \xrightarrow{d} Z,$$

where $X_1, X_2, \ldots$ are i.i.d. copies of X.

By theorem 5.17 the only distributions having a domain of attraction are stable distributions. The following proposition follows from theorems 5.11 and 5.17, and gives conditions under which the asymptotic distribution is a standardized stable distribution, analog to the statement in the classical central limit theorem.

Proposition 5.19. *[16] Let $X_1, X_2, \dots$ be i.i.d. with a symmetric distribution centered at zero. Let the tail probabilities satisfy $x^\alpha P(X > x) \to c$ as $x \to \infty$, for $\alpha \in (1, 2)$ and $c > 0$.*
Then

$$a_n(X_1 + \cdots + X_n) \xrightarrow{d} Z \sim S(\alpha, 1) \text{ as } n \to \infty$$

if $a_n = \frac{1}{n^{\frac{1}{\alpha}}}(\Gamma(\alpha)\sin(\frac{\pi\alpha}{2})/c\pi)$.

This shows, that a distribution with heavy tails has to be scaled by $a_n \sim n^{-\frac{1}{\alpha}}$ in order to get convergence to a stable distribution. Note that for $\alpha = 2$, this is again the scaling of the classical theorem 5.16.

5.3.3 Lévy Flight

For increments η_t with a stable distribution of index $0 < \alpha < 2$, the corresponding process L_t^α, often called Lévy flight, has the following properties:

Proposition 5.20. *Let η_t be i.i.d. $S(\alpha, 1)$ with $\alpha \in (0, 2)$, and let $L_t^\alpha = \sum_{t=0}^{t} \eta_t$. Then:*
i) L_t^α is a Lévy process
ii) $L_t^\alpha - L_s^\alpha \sim (t-s)^{\frac{1}{\alpha}} S(\alpha, 1)$ for t > s.

Proof. w.l.g. s=0:
i) by the definition of L_t^α
ii) by lemma 5.15 and since $\gamma_t = 1 \; \forall t$:

$$\gamma^\alpha = \sum_{t=0}^{t} \gamma_t^\alpha = t$$

Thus

$$L_t^\alpha \sim S(\alpha, t^{\frac{1}{\alpha}}) = t^{\frac{1}{\alpha}} S(\alpha, 1)$$

□

The last property indicates that a Lévy flight is always scaled by $t^{\frac{1}{\alpha}}$. For $\alpha = 2$ this shows the scaling of the Wiener process.

5.4 Ornstein-Uhlenbeck-Lévy Process

As will be shown in section 7.2, the OU process from definition 5.4 does not allow a proper scaling to yield a firing rate in the sense of the escape noise model. However, it turns out that this is indeed only possible for stable distributions with index $\alpha \in (1, 2)$, whereas a mapping to an escape noise model can only be achieved if $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$. Proposition 5.22 shows that an OU process ξ_t driven by a Lévy flight instead of a Wiener process, has again a stable distribution and is therefore a useful modification of the standard case. The definition of the Ornstein-Uhlenbeck-Lévy (OUL) process is analog to the OU process:

Definition 5.21. Let $\beta, \sigma, \tau > 0$. A time discrete *Ornstein-Uhlenbeck-Lévy process* ξ_t centered at zero solves

$$(\Delta\xi_t) = -\beta\xi_t\tau + \gamma(\Delta L_t^\alpha)$$

and therefore satisfies

$$\xi_{t+\tau} - \xi_t = -\beta\xi_t\tau + \gamma\tau^{\frac{1}{\alpha}}\eta_t$$

for $\eta_t \sim S(\alpha, 1)$.

The OUL process ξ_t has again a stable distribution of index α:

Proposition 5.22. *Let ξ_t be an OUL process as defined above and $\xi_0 = 0$. Then*

$$\xi_t \sim S\left(\alpha, \frac{\gamma}{(\alpha\beta)^{\frac{1}{\alpha}}}(1 - e^{-\alpha\beta t})^{\frac{1}{\alpha}}\right)$$

and ξ_t has the asymptotic stationary distribution $S(\alpha, \gamma/(\alpha\beta)^{\frac{1}{\alpha}})$.

Remark. This result is a generalization of proposition 5.5 for the standard OU process. For $\alpha = 2$:

$$\tilde{\sigma}^2 = 2\tilde{\gamma}^2 = \frac{2\gamma^2}{2\beta}(1 - e^{-2\beta t}) = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})$$

Proof. Consider the homogeneous case of definition 5.21: $d\xi_t^H = -\beta\xi_t^H dt$, which has the solution

$$\xi_t^H = C_1 e^{-\beta t}$$

By the method of variation of constants obtain

$$C_1'(t) = \gamma e^{\beta t} dL_t^\alpha \implies C_1(t) = \gamma \int_0^t e^{\beta s} dL_s^\alpha + C_2$$

and thus

$$\xi_t = C_2 e^{-\beta t} + \gamma e^{-\beta t} \int_0^t e^{\beta s} dL_s^\alpha$$

Then, the initial condition $\xi_0 = 0$ yields

$$\xi_t = \gamma e^{-\beta t} \int_0^t e^{\beta s} dL_s^\alpha$$

The latter can be expressed as the limit of a discrete expression:

$$\xi_t = \lim_{n\to\infty} \sum_{i=0}^{n-1} \gamma e^{-\beta(t-\tau_i)} [L_{\tau_{i+1}}^\alpha - L_{\tau_i}^\alpha]$$

where $0 = \tau_0 < \tau_1 < \cdots < \tau_n = t$ is a partition of $[0,t]$. Denote $\tau_{i+1} - \tau_i$ by $(\Delta\tau_i)$. By proposition 5.20 each summand is distributed as $\gamma e^{-\beta(t-\tau_i)} (\Delta\tau_i)^{\frac{1}{\alpha}} S(\alpha, 1)$ or equivalently $S(\alpha, \gamma e^{-\beta(t-\tau_i)} (\Delta\tau_i)^{\frac{1}{\alpha}})$. Then by Lemma 5.15

$$\begin{aligned}
\xi_t &\sim \lim_{n\to\infty} S\left(\alpha, \left(\sum_{i=0}^{n-1} \gamma^\alpha e^{-\alpha\beta(t-\tau_i)} (\Delta\tau_i)\right)^{\frac{1}{\alpha}}\right) \\
&= S\left(\alpha, \gamma\left(\int_0^t e^{-\alpha\beta(t-s)} ds\right)^{\frac{1}{\alpha}}\right) \\
&= S\left(\alpha, \frac{\gamma}{(\alpha\beta)^{\frac{1}{\alpha}}} (1 - e^{-\alpha\beta t})^{\frac{1}{\alpha}}\right)
\end{aligned}$$

and the statement follows. Note that the second equation holds, since by proposition 5.14 the distribution is jointly continuous in both parameters. ☐

5.5 OU vs. OUL

Even though the framework of the OUL process is very similar to the OU process, the properties differ quite heavily. From the properties of the distributions $S(\alpha, \gamma)$ follows that ξ_t has infinite variance for all $\alpha < 2$ and not even a mean for $\alpha < 1$. Whereas in the gaussian case $\alpha = 2$ the sample paths are almost sure continuous, the sample paths of an OUL process are not continuous for $\alpha < 2$. Due to the heavy tails of the increment distribution for $\alpha < 2$, the sample paths of an OUL process are of jump-type as illustrated by figure 6 [7].

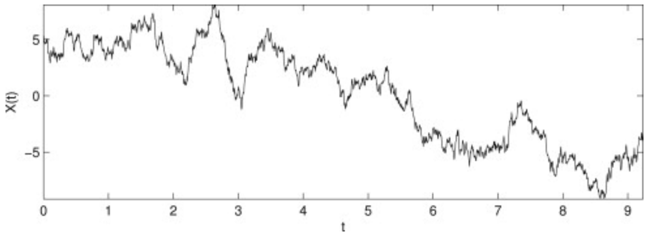

Figure 5: Sample path of an OUL process with $\alpha = 2$ [11]

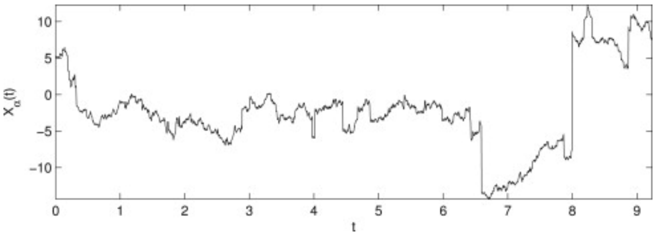

Figure 6: Sample path of an OUL process with $\alpha = 1.5$ [11]

6 Stochastic Version of the dI&F Neuron

As already mentioned in section 2, in vivo experiments give reasons to divide the presynaptic neurons into two categories. The first category contains the neurons who's spikes are actually carrying information relevant for the observed neuron, for the learning process respectively. The membrane potential caused by this neurons will be referred to as the deterministic part given by the differential equation from the dynamic I&F (dI&F) neuron from definition 3.2. The second category is interpreted as background noise, coming from the activity of neurons that are not directly involved in the considered learning process an thus will be described by a stochastic process, reasonably chosen to be an OU process as shown in this section. The introduction of the stochastic potential in 6.2 leads to a dI&F neuron with stochastic input, which is mapped to an escape noise neuron in section 7.

6.1 Noise as Ornstein-Uhlenbeck Process

The large number N of presynaptic neurons causing the background noise is assumed to emit spikes and initiate postsynaptic potentials at random times and to be stochastically independent. The PSP's are double exponentials of the form

$$PSP(t|t_i) = V_i(1 - e^{\frac{t-t_i}{\tau_r}})e^{-\frac{t-t_i}{\tau_d}} \tag{27}$$

where $\tau_r \ll \tau_d$, such that the raise time is very small compared to the decay time, and the amplitudes V_i are assumed to be gaussian distributed with zero mean. Then the following theorem shows that summing up all the PSP's leads to an OU process in the limits $\tau_r \to 0$ and N $\to \infty$.

Theorem 6.1. *[21] Let N presynaptic neurons initiate PSP's at random times t_i given by a Poisson process and the PSP's evolve as in (27) in the limit $\tau_r \to 0$. Let $V_{max} \sim N(0, \sigma^2)$. Then*

$$\xi_t := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} PSP(t|t_i)$$

converges in distribution to an OU process as $N \to \infty$.

Figure 7 shows two PSP's of the form (27) - PSP$^+$ (excitatory) initiated at time t_i and PSP$^-$ (inhibitory) initiated at time t_j. The solid line is the potential resulting by summing PSP$^+$ and PSP$^-$.

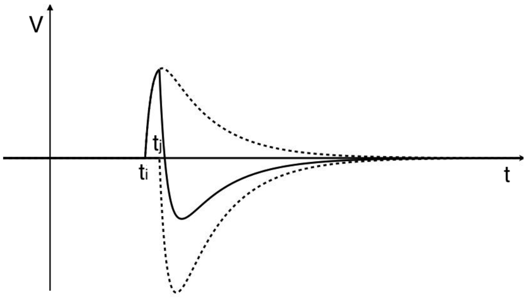

Figure 7: PSP$^+$ and PSP$^-$ (both dashed) and the resulting Potential

The resulting potential obviously converges to a continuous but non-differentiable function in the limit $\tau_r \to 0$, matching the corresponding properties of the OU process. Thus the noise term ξ_t will be modeled by an OU process in the gaussian case and analog by an OUL process if the amplitudes are assumed to have a stable distribution of index $\alpha < 2$.
Note that the gaussian as well as the heavy-tailed case will be called OU process where there is no explicit reference to their specific properties.

6.2 Stochastic Potential

Section 6.1 motivates a model which describes the total membrane potential $\tilde{V}_t$ as the sum of the deterministic potential V_t and a stochastic noise ξ_t given by an OU process in discrete time. Since the membrane potential of a HH neuron is deterministic, the potential of the model should get deterministic in the limit $\Delta \to 0$. Thus the noise ξ_t is scaled by a function $a(\Delta)$ such that $\lim_{\Delta \to 0} a(\Delta) = 0$:

Definition 6.2. Let V_t be the potential given by definition 3.2, ξ_t a time discrete OU process and $a(\Delta)$ s.t. $\lim_{\Delta \to 0} a(\Delta) = 0$. The *stochastic potential* of the dI&F neuron is then defined by

$$\tilde{V}_t := V_t + a(\Delta)\xi_t.$$

Obviously $\lim_{\Delta \to 0} \tilde{V}_t = V_t$. Condition 1 in section 7 gives a condition on $a(\Delta)$ that guarantees a stochastic spiking behavior of the dI&F neuron even in the limit $\Delta \to 0$.

7 dI&F Neuron with Stochastic Input

As motivated in the previous section, the membrane potential $\tilde{V}_t$ can be split into a deterministic part V_t and a stochastic part ξ_t, where the latter can be expressed by a time-discrete OU process, later formally converted into a continuous-time process by taking the limit $\Delta \to 0$. In section 7.2 the common attempt of gaussian noise is discussed, which turns out to not even allow to define a firing rate, and in section 7.3 it is shown, that the same setting for Lévy-stable noise of index $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$ allows a mapping to an escape noise neuron if the stimulus is assumed to be subthreshold.

7.1 Setting

Since the dI&F threshold-model depends on the slope $\dot{\tilde{V}}_t$ of the potential as well as on $\tilde{V}_t$, define the derivative of the stochastic potential from definition 6.2 as

$$\dot{\tilde{V}}_t = \dot{V}_t + a(\Delta)\dot{\xi}_t \tag{28}$$

where $\dot{\xi}_t$ is given by definiton 7.1.

Definition 7.1. Let ξ_t be a time discrete OU process with time step Δ. The derivative of ξ_t is defined as

$$\dot{\xi}_t := \frac{\xi_{t+k\Delta} - \xi_t}{k\Delta}$$

for a $k \in \mathbb{N}$.

Replacing V_t and $\dot{V}_t$ by their stochastic versions $\tilde{V}_t$ and $\dot{\tilde{V}}_t$ in definition 3.2 yields the dI&F neuron with stochastic input.

7.1.1 Firing Rate

Goal is to find a scaling $a(\Delta)$ such that the noise term disappears in the limit $\Delta \to 0$. However, the convergence of $a(\Delta)$ to zero has to be of a specific rate to guarantee the existence of a finite firing rate.

Definition 7.2. Let $\tilde{V}_t$ be the stochastic potential of the dI&F neuron. The probability $P(\tilde{V}_t + \dot{\tilde{V}}_t > 1)$ of a spike in $[t, t+\Delta]$ is

$$P_t(\Delta) = P\left(\xi_t + \dot{\xi}_t > \frac{1 - (V_t + \dot{V}_t)}{a(\Delta)}\right)$$

Definition 7.3. Let $P_t(\Delta)$ be the probability of having a spike in $[t, t+\Delta]$.

$$\rho(V_t, \dot{V}_t) := \frac{P_t(\Delta)}{\Delta}$$

is the *firing rate* of the discrete neuron model at time t.

To be able to map the dI&F neuron with stochastic input to the escape noise model, it is necessary that $\rho(V_t, \dot{V}_t)$ stays finite and non-zero even in the limit $\Delta \to 0$, which is achieved if $P_t(\Delta) \propto \Delta$. Therefore the first condition is

Condition 1. The scaling $a(\Delta)$ satisfies $\lim\limits_{\Delta \to 0} a(\Delta) = 0$ and yields $P_t(\Delta) \propto \Delta$.

Remark. In order to prevent $P_t(\Delta) \to 1$ as $\Delta \to 0$, which would obviously contradict $P_t\Delta) \propto \Delta$, the stimulus needs to be assumed subthreshold. Thus, the following theorems contain the condition $V_t + \dot{V}_t \leq 1 - \varepsilon$ for some $\varepsilon > 0$.

7.1.2 ISI Distribution

If the firing rate $\rho(V_t, \dot{V}_t)$ exists, viz. condition 1 is satisfied, the dI&F neuron with stochastic input should also allow to get an expression for the distribution of the interspike intervals as in definition 2.6.

Definition 7.4. Let $\rho(V_t, \dot{V}_t)$ be the firing rate at time t and $P_I(t|\hat{t})$ the ISI distribution for $t > \hat{t}$, where $\hat{t}$ is the time of the last spike. The ISI distribution is of *escape noise type* if it is of the form

$$P_I(t|\hat{t}) = \rho(V_t, \dot{V}_t) \exp\left\{ -\int_{\hat{t}}^{t} \rho(V_s, \dot{V}_s) ds \right\}$$

Therefore a second condition on the dI&F neuron with stochastic input can be formulated as follows

Condition 2. The ISI distribution is of escape noise type in the limit $\Delta \to 0$

7.1.3 Continuity and Differentiability

Reinforcement learning from definition 4.6 requires $P_w(y|x)$ to be continuous in w and $\frac{\partial}{\partial w} P_w(y|x)$ to be finite. Since only V_t and $\dot{V}_t$ depend on w, this means that $P_w(y|x)$ has to be continuous in V_t and $\dot{V}_t$ and moreover the derivatives of $P_w(y|x)$ with respect to V_t, $\dot{V}_t$ respectively, have to be finite. Moreover, the derivatives should not vanish in the limit $\Delta \to 0$ in order to allow reinforcement learning. Proposition 4.8 and the ISI distribution of definition 7.4 imply that these conditions are satisfied for $P_w(y|x)$ if they hold for $\rho(V_t, \dot{V}_t)$. Thus the third condition is

Condition 3. The firing rate $\rho(V_t, \dot{V}_t)$ is continuous in V_t and $\dot{V}_t$ and its first order derivatives $\frac{\partial}{\partial V_t} \rho(V_t, \dot{V}_t)$ and $\frac{\partial}{\partial \dot{V}_t} \rho(V_t, \dot{V}_t)$ have to exist and be finite and non-zero in the limit $\Delta \to 0$.

7.2 dI&F Neuron with Gaussian Noise

It seems most natural to describe noise in a biological phenomenon by a gaussian distribution. Thus by definition 5.4 the noise term ξ_t is given by

$$\xi_{t+\Delta} = (1-\beta\Delta)\xi_t + \sigma\sqrt{\Delta}\eta_t \tag{29}$$

where $\eta_t \sim N(0,1)$. However, it turns out that a membrane potential with gaussian noise does not allow a firing rate independent of Δ. Thus the stochasticity of the dI&F neuron disappears in the limit $\Delta \to 0$.

Theorem 7.5. *The dI&F neuron with stochastic input can not be mapped to an escape rate neuron if the noise is gaussian.*

For the proof of theorem 7.5, the propositions 7.6 and 7.7 lead to an expression for $\rho(V_t, \dot{V}_t)$ depending on $a(\Delta)$, which does not allow a proper scaling according to condition 1.

Proposition 7.6. *Let $\eta_t \sim N(0,1)$, ξ_t as in definition 5.4 and $\dot{\xi}_t$ as in definition 7.1. Then $\xi_t + \dot{\xi}_t$ has a gaussian distribution with zero mean and variance $\frac{\sigma^2}{\Delta}$ as $\Delta \to 0$. Thus the variance tends to infinity as $\frac{1}{\Delta}$.*

Proof. To rewrite $\dot{\xi}_t$, note that $\xi_{t+k\Delta}$ can be written as

$$\begin{aligned}\xi_{t+k\Delta} &= (1-\beta\Delta)^k\xi_t + \sigma\sum_{i=0}^{k-1}(1-\beta\Delta)^i\sqrt{\Delta}\eta_{t+i\Delta}\\ &\approx \xi_t - k\beta\Delta\xi_t + \sigma\sqrt{\Delta}\sum_{i=0}^{k-1}\eta_{t+i\Delta}\end{aligned}$$

where the second equation skips all terms with coefficients of order $o(\Delta)$.

Thus as $\Delta \to 0$

$$\begin{aligned}\xi_t + \dot{\xi}_t &\approx \xi_t + \frac{-k\beta\xi_t\Delta + \sigma\sqrt{\Delta}\sum_{i=0}^{k-1}\eta_{t+i\Delta}}{k\Delta}\\ &= (1-\beta)\xi_t + \frac{\sigma}{k\sqrt{\Delta}}\sum_{i=0}^{k-1}\eta_{t+i\Delta}\\ &\approx \frac{\sigma}{k\sqrt{\Delta}}\sum_{i=0}^{k-1}\eta_{t+i\Delta}.\end{aligned}$$

Since η_t are i.i.d. N(0,1), their sum is again gaussian with mean zero and variance

$$Var\left(\sum_{i=0}^{k-1}\eta_{t+i\Delta}\right) = \sum_{i=0}^{k-1}Var(\eta_{t+i\Delta}) = k$$

Thus by the approximation above

$$\mathrm{Var}(\xi_t + \dot{\xi}_t) = \frac{\sigma^2}{\Delta}$$

□

Proposition 7.7 now gives an explicit formula for the firing rate $\rho(V_t, \dot{V}_t)$.

Proposition 7.7. *Let* $\tilde{V}_t$ *and* $\dot{\tilde{V}}_t$ *be as in section 7.1, and* $\xi_t + \dot{\xi}_t$ *as in proposition 7.6. Then*

$$\rho(V_t, \dot{V}_t) = \frac{1-(V_t+\dot{V}_t)}{\sigma 2\pi} \exp\left\{ -\frac{(1-(V_t+\dot{V}_t))^2}{2\sigma^2}\frac{\Delta}{a^2(\Delta)} \right\} \frac{a(\Delta) - 2\Delta a'(\Delta)}{2\sqrt{\Delta}a^2(\Delta)}$$

Proof. Denoting $f(x) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}x^2\right\}$ the density of the standard gaussian distribution, definition 7.2 yields

$$P_t(\Delta) = \int_{g(\Delta)}^{\infty} f(x)dx$$

where $g(\Delta) = \frac{1-(V_t+\dot{V}_t)}{\sigma}\frac{\sqrt{\Delta}}{a(\Delta)}$. By definition 7.3 $P_t(\Delta) = \rho(V_t, \dot{V}_t)\Delta$ and thus

$$\begin{aligned}\rho(V_t, \dot{V}_t) &= \frac{d}{d\Delta}P_t(\Delta) \\ &= -f(g(\Delta))\frac{d}{d\Delta}g(\Delta) \\ &= -f(g(\Delta))\frac{1-(V_t+\dot{V}_t)}{\sigma}\frac{\frac{1}{2\sqrt{\Delta}}a(\Delta) - \sqrt{\Delta}a'(\Delta)}{a^2(\Delta)}\end{aligned}$$

and the statement follows. □

To satisfy condition 1, the scaling $a(\Delta)$ should be chosen such that $\rho(V_t, \dot{V}_t)$ in proposition 7.7 is independent of Δ and non-zero.

Proof of theorem 7.5. By proposition 7.7 the firing rate $\rho(V_t, \dot{V}_t)$ is independent of Δ if and only if $a(\Delta) \propto \sqrt{\Delta}$. However, this yields $\rho(V_t, \dot{V}_t) = 0$ since $P_t(\Delta)$ is independent of Δ as well and thus condition 1 can not be satisfied. □

7.3 dI&F Neuron with Heavy-Tailed Noise

As seen in the previous section, modeling the noise by a standard OU process fails in satisfying condition 1. This rises the question wether there are other distributions for η_t which allow to avoid the problem of the gaussian case. Since the noise should still follow an OU process, the only reasonable candidates are the stable distributions described in section 5.3.1. Then by definition 5.21 the OUL process satisfies

$$\xi_{t+\Delta} = (1-\beta\Delta)\xi_t + \gamma\Delta^{\frac{1}{\alpha}}\eta_t \tag{30}$$

with $\eta_t \sim S(\alpha, 1)$ and $\alpha \in (0, 2]$.

7.3.1 Scaling

To perform an approximation analog to the one in proposition 7.6 it is necessary to restrict the stability index to $\alpha > 1$. Since the gaussian case has been excluded as well, only $\alpha \in (1,2)$ will be considered in the following.

Theorem 7.8. *Let $\tilde{V}_t$ and $\dot{\tilde{V}}_t$ be the dI&F neuron with stochastic input, $V_t + \dot{V}_t \leq 1 - \varepsilon$ for a $\varepsilon > 0$ and $\eta_t \sim S(\alpha, 1)$ with index $\alpha \in (1,2)$. Then as $\Delta \to 0$*

$$\rho(V_t, \dot{V}_t) \sim \frac{c_\alpha}{k^{\alpha-1}} \left(\frac{\gamma C}{1 - (V_t + \dot{V}_t)} \right)^\alpha$$

if and only if $a(\Delta) = C\Delta$, $C > 0$.

Corollary 7.9. By theorem 7.8, condition 1 is satisfied for η_t with a stable distribution of index $\alpha \in (1,2)$ and a scaling by $a(\Delta) = C\Delta$.

For the proof of theorem 7.8 the following proposition gives an approximation of $\xi_t + \dot{\xi}_t$ as $\Delta \to 0$ and its distribution.

Proposition 7.10. *Let $\alpha \in (1,2)$, ξ_t be as in definition 5.21 and $\dot{\xi}_t$ as in definition 7.1. Then $\xi_t + \dot{\xi}_t$ has a stable distribution of index α and scaling parameter $\gamma(k\Delta)^{\frac{1-\alpha}{\alpha}}$ as $\Delta \to 0$. Thus the scaling parameter tends to infinity as $1/\Delta^{\frac{\alpha-1}{\alpha}}$.*

Proof. $\xi_{t+k\Delta}$ can be expressed in terms of ξ_t and the increments $\eta_t, \ldots, \eta_{t+k\Delta}$:

$$\begin{aligned} \xi_{t+k\Delta} &= (1 - \beta\Delta)^k \xi_t + \gamma \sum_{i=0}^{k-1} (1 - \beta\Delta)^i \Delta^{\frac{1}{\alpha}} \eta_{t+i\Delta} \\ &\approx \xi_t - k\beta\xi_t\Delta + \gamma\Delta^{\frac{1}{\alpha}} \sum_{i=0}^{k-1} \eta_{t+i\Delta} \end{aligned}$$

where the second equation skips all terms with coefficients of order $o(\Delta)$.

Thus as $\Delta \to 0$

$$\begin{aligned} \xi_t + \dot{\xi}_t &\approx \xi_t + \frac{-k\beta\xi_t\Delta + \gamma\Delta^{\frac{1}{\alpha}} \sum_{i=0}^{k-1} \eta_{t+i\Delta}}{k\Delta} \\ &= (1 - \beta)\xi_t + \frac{\gamma}{k}\Delta^{\frac{1-\alpha}{\alpha}} \sum_{i=0}^{k-1} \eta_{t+i\Delta} \\ &\approx \frac{\gamma}{k}\Delta^{\frac{1-\alpha}{\alpha}} \sum_{i=0}^{k-1} \eta_{t+i\Delta}. \end{aligned}$$

The approximation holds since $\alpha \in (1, 2)$. Then by proposition 5.14

$$\begin{aligned}\xi_t + \dot{\xi}_t &\sim S(\alpha, \frac{\gamma}{k} \Delta^{\frac{1-\alpha}{\alpha}} k^{\frac{1}{\alpha}}) \\ &= S(\alpha, \gamma(k\Delta)^{\frac{1-\alpha}{\alpha}})\end{aligned}$$

□

According to theorem 7.8, in contrast to the gaussian case, the choice of $a(\Delta) = C\Delta$ yields a finite firing rate $\rho(V_t, \dot{V}_t)$.

Proof of theorem 7.8. Note, that here again $f(\Delta) \sim g(\Delta) \iff \lim_{\Delta \to 0} f(\Delta)/g(\Delta) = 1$. By proposition 7.10: $\xi_t + \dot{\xi}_t \sim S\left(\alpha, \gamma(k\Delta)^{\frac{1-\alpha}{\alpha}}\right)$ as $\Delta \to 0$. The subthreshold condition yields $\frac{1-(V_t+\dot{V}_t)}{a(\Delta)} \to \infty$ as $\Delta \to 0$ and thus by theorem 5.11

$$\begin{aligned}P_t(\Delta) &= P\left(\dot{\xi}_t > \frac{1-(V_t+\dot{V}_t)}{a(\Delta)}\right) \\ &\sim c_\alpha(\gamma(k\Delta)^{\frac{1-\alpha}{\alpha}})^\alpha \left[\frac{1-(V_t+\dot{V}_t)}{a(\Delta)}\right]^{-\alpha} \\ &= \frac{c_\alpha \gamma^\alpha k^{1-\alpha}}{(1-(V_t+\dot{V}_t))^\alpha} \Delta^{1-\alpha} a^\alpha(\Delta)\end{aligned}$$

By choosing $a(\Delta) = C\Delta$ follows $P_t(\Delta) \propto \Delta$ and the proposition. □

7.3.2 ISI Distribution

By corollary 7.9, the dI&F neuron with stable noise of index $\alpha \in (1, 2)$ satisfying condition 1 is defined by

$$\tilde{V}_t = V_t + C\Delta\,\xi_t \tag{31}$$

$$\tilde{\dot{V}}_t = \dot{V}_t + C\Delta\,\dot{\xi}_t \tag{32}$$

and guarantees the existence of a Poisson-like firing rate. However, this does not yet proof an ISI distribution of escape noise type in the limit $\Delta \to 0$. Thus this section proofs that this can be achieved if $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$. Moreover, all the following statements assume $k = 1$ in definition 7.1.

Theorem 7.11. *Let $P_I(t|\hat{t})$ be the ISI of the dI&F neuron with stochastic input after a spike at $\hat{t}$, and $\rho(V_s, \dot{V}_s)$ the firing rate from definition 7.3. Let $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$ and assume $V_s + \dot{V}_s \leq 1 - \varepsilon$ for $s \in [\hat{t}, t]$ and some $\varepsilon > 0$. Then in the limit $\Delta \to 0$*

$$P_I(t|\hat{t}) = \rho(V_t, \dot{V}_t) \exp\left\{-\int_{\hat{t}}^{t} \rho(V_s, \dot{V}_s) ds\right\}$$

Corollary 7.12. Let $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$ and assume $V_s + \dot{V}_s \leq 1 - \varepsilon$ for $s \in [\hat{t}, t]$ and some $\varepsilon > 0$, then by theorem 7.11 condition 2 is satisfied.

The proof of theorem 7.11 follows from the following theorem which gives an expression for a quiescent interval $[t, \hat{t}]$

Theorem 7.13. *Let $\tilde{P}_I(t|\hat{t})$ be the probability of a quiescent interval $[\hat{t}, t]$, and $\rho(V_s, \dot{V}_s)$ the firing rate from definition 7.3. Let $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$ and assume $V_s + \dot{V}_s \leq 1 - \varepsilon$ for $s \in [\hat{t}, t]$ and some $\varepsilon > 0$. Then in the limit $\Delta \to 0$*

$$\tilde{P}_I(t|\hat{t}) = \exp\left\{ - \int_{\hat{t}}^{t} \rho(V_s, \dot{V}_s) ds \right\}$$

Proof of theorem 7.11. Note that $\int_{\hat{t}}^{t} P_I(s|\hat{t}) ds = 1 - \tilde{P}_I(t|\hat{t})$. Thus

$$\begin{aligned}
P_I(t|\hat{t}) &= \frac{\partial}{\partial t} \left(1 - \tilde{P}_I(t|\hat{t}) \right) \\
&= -\frac{\partial}{\partial t} \exp\left\{ - \int_{\hat{t}}^{t} \rho(V_s, \dot{V}_s) ds \right\} \\
&= \exp\left\{ - \int_{\hat{t}}^{t} \rho(V_s, \dot{V}_s) ds \right\} \frac{\partial}{\partial t} \int_{\hat{t}}^{t} \rho(V_s, \dot{V}_s) ds \\
&= \rho(V_t, \dot{V}_t) \exp\left\{ - \int_{\hat{t}}^{t} \rho(V_s, \dot{V}_s) ds \right\}
\end{aligned}$$

where the second equation follows from theorem 7.13. □

Thus the rest of this section yields the proof of theorem 7.13. Note that since $\hat{t}$ is arbitrary, $\hat{t} = 0$ is assumed and any reference to $\hat{t}$ is omitted to keep notations simple.

Definition 7.14. Define $x_t = \tilde{V}_t + \dot{\tilde{V}}_t$. Then the probability $\tilde{P}(t)$ of a *quiescent interval* $[0, t]$ can be defined as

$$\tilde{P}_I(t) := E_{\eta_{0,\ldots,t}}[\prod_{s \leq t} \theta(1 - x_s)]$$

where $\theta(x_t)$ is an indicator function such that $\theta(x_t - 1) = \begin{cases} 1 & x_t > 1 \\ 0 & x_t \leq 1 \end{cases}$.

Remark. Note that $E[\theta(1 - x_s)] = P(x_s \leq 1) = P(V_t + \dot{V}_t + C\Delta(\xi_t + \dot{\xi}_t) \leq 1)$.

Theorem 7.15 gives an estimation of the error done by replacing the indicator function of the last time step t by $(1 - P_t(\Delta))$. Note that $f(\Delta) = o(\Delta) \iff \lim_{\Delta \to 0} f(\Delta)/\Delta = 0$.

Proposition 7.15. *Let $\tilde{P}_I(t)$ be the probability of a quiescent interval $[0, t]$, and $P_s(\Delta)$ the probability of a spike at time s. Let $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$ and assume $V_s + \dot{V}_s \leq 1 - \varepsilon$ for $s \in [0, t]$ and some $\varepsilon > 0$. Then as $\Delta \to 0$*

$$\tilde{P}_I(t) = E_{\eta_{0,\ldots,t-\Delta}}[\prod_{s<t} \theta(1 - x_s)](1 - P_t(\Delta)) + o(\Delta)$$

Proof. By definition x_s is independent of η_t for $s < t$. Thus

$$\begin{aligned}\tilde{P}_I(t) &= E_{\eta_{0,\dots,t}}[\prod_{s\leq t} \theta(1-x_s)] \\ &= E_{\eta_{0,\dots,t-\Delta}}[\prod_{s<t} \theta(1-x_s) E_{\eta_t}[\theta(1-x_t)]]\end{aligned}$$

Note that $k = 1$ in definition 7.1 yields $\dot{\xi}_t \approx \gamma\Delta^{\frac{1-\alpha}{\alpha}}\eta_t$ as $\Delta \to 0$. Thus, consistent with the definition above, denote $x_t = V_t + \dot{V}_t + C\Delta\,\xi_t + \gamma C\Delta^{\frac{1}{\alpha}}\eta_t$. Then

$$E_{\eta_t}[\theta(1-x_t)] = \int_{-\infty}^{\infty} \theta(1-(V_t + \dot{V}_t + C\Delta\,\xi_t + \gamma C\Delta^{\frac{1}{\alpha}}\eta_t))d\mu(\eta_t)$$

and the θ-function yields $\eta_t < \frac{1-(V_t+\dot{V}_t)}{C\gamma\Delta^{\frac{1}{\alpha}}} - \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma}\xi_t$. Hence

$$\begin{aligned}E_{\eta_t}[\theta(1-x_t)] &= \int_{-\infty}^{\frac{1-(V_t+\dot{V}_t)}{C\gamma\Delta^{\frac{1}{\alpha}}} - \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma}\xi_t} d\mu(\eta_t) \\ &= \int_{-\infty}^{\frac{1-(V_t+\dot{V}_t)}{C\gamma\Delta^{\frac{1}{\alpha}}}} d\mu(\eta_t) - \int_{\frac{1-(V_t+\dot{V}_t)}{C\gamma\Delta^{\frac{1}{\alpha}}} - \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma}\xi_t}^{\frac{1-(V_t+\dot{V}_t)}{C\gamma\Delta^{\frac{1}{\alpha}}}} d\mu(\eta_t) \\ &= (1-P_t(\Delta)) - h(V_t + \dot{V}_t, \Delta, \xi_t)\end{aligned}$$

Thus

$$\tilde{P}_I(t) = E_{\eta_{0,\dots,t-\Delta}}[\prod_{s<t} \theta(1-x_s)](1-P_t(\Delta)) - E_{\eta_{0,\dots,t-\Delta}}[\prod_{s<t} \theta(1-x_s) h(V_t + \dot{V}_t, \Delta, \xi_t)]$$

where $h(V_t + \dot{V}_t, \Delta, \xi_t) := \int_{\frac{1-(V_t+\dot{V}_t)}{C\gamma\Delta^{\frac{1}{\alpha}}} - \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma}\xi_t}^{\frac{1-(V_t+\dot{V}_t)}{C\gamma\Delta^{\frac{1}{\alpha}}}} d\mu(\eta_t)$.
To show: $\mathrm{Err}_t(\Delta) := E_{\eta_{0,\dots,t-\Delta}}[\prod_{s<t} \theta(1-x_s) h(V_t + \dot{V}_t, \Delta, \xi_t)] = o(\Delta)$.
Note that $\prod_{s<t} \theta(1-x_s) \leq 1$ yields

$$|\mathrm{Err}_t(\Delta)| \leq E[|h(V_t + \dot{V}_t, \Delta, \xi_t)|]$$

Set $\tilde{\varepsilon} = \frac{\varepsilon}{2}$, denote $l_t := 1 - (V_t + \dot{V}_t)$ and write the mean above as $J(\Delta) + K(\Delta)$, where

$$\begin{aligned}J(\Delta) &:= E[\theta(l_t - C\Delta\xi_t - \tilde{\varepsilon})\,|h(1-l_t, \Delta, \xi_t)|] \\ K(\Delta) &:= E[\theta(\tilde{\varepsilon} + C\Delta\xi_t - l_t)\,|h(1-l_t, \Delta, \xi_t)|]\end{aligned}$$

Note that $J(\Delta)$ is the mean over $\xi_t \leq \frac{l_t - \tilde{\varepsilon}}{C\Delta}$ and thus $K(\Delta)$ the mean over $\xi_t > \frac{l_t - \tilde{\varepsilon}}{C\Delta}$. Then

$$|\mathrm{Err}_t(\Delta)| \leq J(\Delta) + K(\Delta)$$

and by the lemmas 7.16 and 7.17, $J(\Delta)$ as well as $K(\Delta)$ are $o(\Delta)$ and the proposition follows. $\square$

The following lemmas show that $J(\Delta)$ and $K(\Delta)$ are both $o(\Delta)$ and thus complete the proof of proposition 7.15.

Lemma 7.16. *Let* $l_t := 1 - (V_t + \dot{V}_t)$, $J(\Delta) = E[\theta(l_t - C\Delta\xi_t - \tilde{\varepsilon})\,|h(V_t + \dot{V}_t, \Delta, \xi_t)|]$ *and* $V_t + \dot{V}_t$ *be subthreshold as in the theorem. Then*

$$J(\Delta) = o(\Delta)$$

Proof. The θ-function yields that only $\xi_t \leq \frac{l_t - \tilde{\varepsilon}}{C\Delta}$ contributes to $J(\Delta)$. Note that the length of the interval of the integration is $\frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma}\xi_t$. Since $\int_a^b f(x)dx \leq (b-a)\max_{x\in(a,b)} f(x)$ if $f(x) \geq 0$, it follows that

$$J(\Delta) \leq E[|\frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma}\xi_t f_{\text{max}}|]$$

where f_{max} refers to the maximum of the density of η_t on the corresponding interval. Note that

$$f(x) \leq \begin{cases} f\left(\frac{l_t}{\gamma C \Delta^{1/\alpha}}\right) & \text{if } \xi_t \leq 0 \\ f\left(\frac{\tilde{\varepsilon}}{\gamma C \Delta^{1/\alpha}}\right) & \text{if } \xi_t > 0 \end{cases}$$

Thus for $\xi_t \leq 0$:

$$\begin{aligned} J(\Delta) &< \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma} f\left(\frac{l_t}{\gamma C \Delta^{1/\alpha}}\right) E[|\xi_t|] \\ &\sim \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma} \alpha c_\alpha \gamma^\alpha \frac{(\gamma C \Delta^{1/\alpha})^{\alpha+1}}{l_t^{\alpha+1}} E[|\xi_t|] \\ &= \alpha c_\alpha \gamma^{2\alpha} C^{\alpha+1} \frac{\Delta^2}{l_t^{\alpha+1}} E[|\xi_t|] \end{aligned}$$

By assumption $l_t \geq \varepsilon$. Thus

$$\begin{aligned} &\leq \frac{\alpha c_\alpha \gamma^{2\alpha} C^{\alpha+1}}{\varepsilon^{\alpha+1}} \Delta^2 E[|\xi_t|] \\ &= o(\Delta) \end{aligned}$$

since $E[|\xi_t|] < \infty$ by lemma 5.13.
For $\xi_t > 0$

$$\begin{aligned} J(\Delta) &< \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma} f\left(\frac{\tilde{\varepsilon}}{\gamma C \Delta^{1/\alpha}}\right) E[|\xi_t|] \\ &\sim \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma} \alpha c_\alpha \gamma^\alpha \frac{(\gamma C \Delta^{1/\alpha})^{\alpha+1}}{\tilde{\varepsilon}^{\alpha+1}} E[|\xi_t|] \\ &= \frac{\alpha c_\alpha \gamma^{2\alpha} C^{\alpha+1}}{\tilde{\varepsilon}^{\alpha+1}} \Delta^2 E[|\xi_t|] \\ &= o(\Delta) \end{aligned}$$

again since $E[|\xi_t|] < \infty$. $\square$

Lemma 7.17. *Let* $l_t := 1 - (V_t + \dot{V}_t)$, $K(\Delta) = E[\theta(\tilde{\varepsilon} + C\Delta\xi_t - l_t)\,|h(V_t + \dot{V}_t, \Delta, \xi_t)|]$ *and* $V_t + \dot{V}_t$ *be subthreshold as in the theorem. Then*

$$K(\Delta) = o(\Delta)$$

Proof. The θ-function yields that only $\xi_t > \frac{l_t - \tilde{\varepsilon}}{C\Delta}$ contributes to $K(\Delta)$. Again the length of the interval of the integration is $\frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma}\xi_t$. The same idea as in the proof of the previous lemma leeds to

$$K(\Delta) \leq E[\theta(\tilde{\varepsilon} + C\Delta\xi_t - l_t)|\frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma}\xi_t f_{\max}|]$$

where $f_{\max} < \infty$ since $f(x)$ is a density and C^∞ by theorem 5.10. Thus

$$\begin{aligned} K(\Delta) &\leq \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma} f_{\max} E[\theta(\tilde{\varepsilon} + C\Delta\xi_t - l_t)|\xi_t|] \\ &= \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma} f_{\max} \int_{\frac{l_t - \tilde{\varepsilon}}{C\Delta}}^{\infty} |\xi_t| d\mu(\xi_t) \end{aligned}$$

By propostion 5.22 $\xi_t \sim S(\alpha, \gamma/(\alpha\beta)^{\frac{1}{\alpha}})$. Thus as $\Delta \to 0$

$$\begin{aligned} K(\Delta) &\leq \frac{\Delta^{1-\frac{1}{\alpha}}}{\gamma} f_{\max} \int_{\frac{l_t - \tilde{\varepsilon}}{C\Delta}}^{\infty} \alpha c_\alpha \frac{\gamma^\alpha}{\alpha\beta} \frac{\xi_t}{\xi_t^{\alpha+1}} d\xi_t \\ &= \frac{c_\alpha \gamma^{\alpha-1}}{\beta} \Delta^{1-\frac{1}{\alpha}} f_{\max} \int_{\frac{l_t - \tilde{\varepsilon}}{C\Delta}}^{\infty} \frac{1}{\xi_t^\alpha} d\xi_t \\ &= \frac{c_\alpha \gamma^{\alpha-1}}{\beta} \Delta^{1-\frac{1}{\alpha}} f_{\max} \left(\frac{C\Delta}{l_t - \tilde{\varepsilon}}\right)^{\alpha-1} \\ &\leq \frac{c_\alpha (2\gamma C)^{\alpha-1}}{\beta \varepsilon^{\alpha-1}} f_{\max} \Delta^{\alpha - \frac{1}{\alpha}} \end{aligned}$$

since $l_t - \tilde{\varepsilon} \geq \varepsilon/2$. Note that $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$ yields $\alpha - \frac{1}{\alpha} > 1$ and thus

$$K(\Delta) = o(\Delta)$$

☐

Using theorem 7.15, the following theorem states that $\tilde{P}_I(t)$ is given by the product of the non-spike probabilities up to an error E(Δ) that vanishes in the limit $\Delta \to 0$.

Theorem 7.18. *Let* $\tilde{P}_I(t) = E_{\eta_{0,\ldots,t}}[\prod_{s\leq t} \theta(1 - x_s)]$ *be the probability of a quiescent interval* $[0, t]$, *and* $P_s(\Delta)$ *the probability of a spike at time s. Let* $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$ *and assume* $V_s + \dot{V}_s \leq 1 - \varepsilon$ *for* $s \in [0, t]$ *and some* $\varepsilon > 0$. *Then as* $\Delta \to 0$

$$\tilde{P}_I(t) = \prod_{s\leq t}(1 - P_s(\Delta)) + E(\Delta),$$

where $\lim_{\Delta\to 0} E(\Delta) = 0$.

Proof. By iterating proposition 7.15:

$$\tilde{P}_I(t) = \prod_{s\leq t}(1 - P_s(\Delta)) + \sum_{s=0}^{t/\Delta-1} Err_s(\Delta)$$

Hence

$$\tilde{P}_I(t) = \prod_{s\leq t}(1 - P_s(\Delta)) + \mathrm{E}(\Delta)$$

where $\mathrm{E}(\Delta) := \sum_{s=0}^{t/\Delta-1} Err_s(\Delta) = \frac{t}{\Delta}o(\Delta)$ and thus $\lim\limits_{\Delta\to 0} \mathrm{E}(\Delta) = 0$. $\square$

Now theorem 7.13 can be proven as follows.

Proof of theorem 7.13. By theorem 7.18 $\tilde{P}_I(t) = \prod_{s\leq t}(1 - P_s(\Delta)) + \mathrm{E}(\Delta)$. Define $\tilde{\mathrm{E}}(\Delta) := 1 + \frac{\mathrm{E}(\Delta)}{\prod_{s\leq t}(1-P_s(\Delta))}$. Then

$$\begin{aligned}
\log \tilde{P}_I(t) &= \log\left(\prod_{s\leq t}(1 - P_s(\Delta)) + \mathrm{E}(\Delta)\right) \\
&= \log\left(\tilde{\mathrm{E}}(\Delta)\prod_{s\leq t}(1 - P_s(\Delta))\right) \\
&= \sum_{s\leq t} \log\left(1 - \rho(V_s, \dot{V}_s)\Delta\right) + \log\left(\tilde{\mathrm{E}}(\Delta)\right)
\end{aligned}$$

By the taylor approximation of $\log(x)$ around $x = 1$:

$$\begin{aligned}
\log \tilde{P}_I(t) &= \sum_{s\leq t}\left[-\rho(V_s, \dot{V}_s)\Delta + o(\Delta)\right] + \log\left(\tilde{\mathrm{E}}(\Delta)\right) \\
&= -\sum_{s\leq t}\rho(V_s, \dot{V}_s)\Delta + \frac{t}{\Delta}o(\Delta) + \log\left(\tilde{\mathrm{E}}(\Delta)\right)
\end{aligned}$$

Thus

$$\log \tilde{P}_I(t) = -\int_0^t \rho(V_s, \dot{V}_s)ds$$

in the limit $\Delta \to 0$ if $\lim\limits_{\Delta\to 0} \log\left(\tilde{\mathrm{E}}(\Delta)\right) = 0$. Therefore, proving that $\lim\limits_{\Delta\to 0} \frac{\mathrm{E}(\Delta)}{\prod_{s\leq t}(1-P_s(\Delta))} = 0$ completes the proof. Note that $V_s + \dot{V}_s \leq 1 - \varepsilon$ yields $P_s(\Delta) < \frac{1}{2}$ and thus for $\Delta > 0$: $\prod_{s\leq t}(1 - P_s(\Delta)) > 0$. Moreover

$$\prod_{s\leq t}(1 - P_s(\Delta)) \geq (1 - \rho_{\max}\Delta)^{\frac{t}{\Delta}}$$

and thus, since $\rho_{\max} := c_\alpha \left(\frac{\gamma C}{\varepsilon}\right)^\alpha < \infty$, in the limit $\Delta \to 0$

$$\prod_{s \le t}(1 - P_s(\Delta)) \ge \exp\{-t\rho_{\max}\} > 0$$

which yields $\lim\limits_{\Delta\to 0} \frac{\mathrm{E}(\Delta)}{\prod_{s\le t}(1-P_s(\Delta))} \le \exp\{t\rho_{\max}\} \lim\limits_{\Delta\to 0} \mathrm{E}(\Delta) = 0.$ ☐

7.3.3 Continuity and Differentiability

Since $\rho(V_t, \dot{V}_t) = \frac{P_t(\Delta)}{\Delta}$ the continuity in V_t as well as in $\dot{V}_t$ follows from the continuity of $P_t(\Delta)$ (see theorem 5.10). Theorem 7.19 shows that the firing rate from theorem 7.8 has to desired properties the perform reinforcement learning.

Theorem 7.19. *Let $\rho(V_s, \dot{V}_s)$ be the firing rate from theorem 7.8, $\alpha \in (1, 2)$ and $V_s + \dot{V}_s \le 1 - \varepsilon$ for $s \in [\hat{t}, t]$ and some $\varepsilon > 0$.*
Then $\rho(V_s, \dot{V}_s)$ is strictly increasing and differentiable in both variables.

Thus condition 3 is satisfied for subthreshold stimulus as in the theorem:

Corollary 7.20. Let $\rho(V_s, \dot{V}_s)$ be the firing rate from theorem 7.8, $\alpha \in (1, 2)$ and $V_s + \dot{V}_s \le 1 - \varepsilon$ for $s \in [\hat{t}, t]$ and some $\varepsilon > 0$, then by theorem 7.19 condition 3 is satisfied.

Proof. Note that $\frac{\partial}{\partial V_s}\rho(V_s, \dot{V}_s) = \frac{\partial}{\partial V_s}\frac{P_s(\Delta)}{\Delta} = \frac{1}{\Delta}\frac{\partial}{\partial V_s}P_s(\Delta)$ and use again theorem 5.10 for the differentiability of $P_s(\Delta)$ which proofs the differentiability of $\rho(V_s, \dot{V}_s)$.

$$\begin{aligned}
\frac{\partial}{\partial V_s}P_s(\Delta) &= \frac{\partial}{\partial V_s}P\left(\xi_s + \dot{\xi}_s > \frac{1 - (V_s + \dot{V}_s)}{C\Delta}\right) \\
&= -f\left(\frac{1 - (V_s + \dot{V}_s)}{C\Delta}\right)\frac{\partial}{\partial V_s}\frac{1 - (V_s + \dot{V}_s)}{C\Delta} \\
&= \frac{1}{C\Delta}f\left(\frac{1 - (V_s + \dot{V}_s)}{C\Delta}\right)
\end{aligned}$$

By proposition 7.10 $\xi_s + \dot{\xi}_s \sim S(\alpha, \gamma\Delta^{\frac{1-\alpha}{\alpha}})$, where the tails of the probability density f are given by theorem 5.11, thus as $\Delta \to 0$

$$\begin{aligned}
\frac{\partial}{\partial V_s}P_s(\Delta) &\sim \frac{\alpha c_\alpha(\gamma\Delta^{\frac{1-\alpha}{\alpha}})^\alpha}{C\Delta}\left(\frac{C\Delta}{1 - (V_s + \dot{V}_s)}\right)^{\alpha+1} \\
&= \frac{\alpha c_\alpha(\gamma C)^\alpha}{(1 - (V_s + \dot{V}_s))^{\alpha+1}}\Delta
\end{aligned}$$

Thus

$$\frac{\partial}{\partial V_s}\rho(V_s, \dot{V}_s) = \frac{\alpha c_\alpha(\gamma C)^\alpha}{(1 - (V_s + \dot{V}_s))^{\alpha+1}}$$

in the limit $\Delta \to 0$. Hence $\rho(V_s, \dot{V}_s)$ is strictly increasing since $\frac{\partial}{\partial V_s}\rho(V_s, \dot{V}_s) > 0$. Note that the proof for $\frac{\partial}{\partial \dot{V}_s}\rho(V_s, \dot{V}_s)$ is analog and thus omitted. $\square$

Remark. From the proof of theorem 7.19 follows that the derivatives of $\rho(V_s, \dot{V}_s)$ are bounded by $\frac{\alpha c_\alpha (\gamma C)^\alpha}{\varepsilon^{\alpha+1}}$. Thus the subthreshold condition guarantees the finiteness of the reinforcement learning rule from definition 4.6, whereas the learning rule would tend to infinity in the limit $V_s + \dot{V}_s \to 1$.

7.3.4 Conclusion

The results from sections 7.3.1 -7.3.3 show, that the dI&F model with heavy-tailed noise turns into an escape noise neuron in the limit $\Delta \to 0$.

Theorem 7.21. *Let the potential of the dI&F neuron with stochastic input be given by $\tilde{V}_t$, $\dot{\tilde{V}}_t$ from section 7.1 with $a(\Delta) = C\Delta$ and $k = 1$ in definition 7.1. Let ξ_t be an OUL process as in definition 5.21 with $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$ and $V_s + \dot{V}_s \leq 1 - \varepsilon$ for $s \in [\hat{t}, t]$ and some $\varepsilon > 0$. Then the ISI distribution of the dI&F neuron with stochastic input is given by*

$$P_I(t|\hat{t}) = \rho(V_t, \dot{V}_t) \exp\left\{ -\int_{\hat{t}}^{t} \rho(V_s, \dot{V}_s) ds \right\}$$

in the limit $\Delta \to 0$. Moreover $P_I(t|\hat{t})$ is continuous and differentiable in w.

Thus, the dI&F neuron with stochastic input can be mapped to an escape noise model if the noise is chosen to have a stable distribution of index $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$ and the potential is assumed to be subthreshold such that $V_t + \dot{V}_t \leq 1 - \varepsilon$ for some $\varepsilon > 0$.

7.3.5 Further Reduction

For the classical escape noise neuron the firing rate is a function of the instantaneous voltage only, without dependence on its derivative. Once the HH model is reduced to an escape noise neuron with instantaneous rate $\rho(V_t, \dot{V}_t) = \tilde{\rho}(V_t + \dot{V}_t)$, one can reparametrize the voltage into an effective voltage of the form $\tilde{V}_t = V_t + \dot{V}_t$. Correspondingly, the effective postsynaptic potential triggered by a presynaptic spike will have the form $\widetilde{\mathrm{PSP}}(t) = \mathrm{PSP}(t) + \dot{\mathrm{PSP}}(t)$, such that $\tilde{V}_t$ is still the waited sum of the $\widetilde{\mathrm{PSP}}$'s analogous to (19). Note that for a PSP-shape representing a low-pass filtering of the form given in Fig. 7 (i.e. defining $\widetilde{\mathrm{PSP}}$ via $\tau\dot{\widetilde{\mathrm{PSP}}}(t) = -\widetilde{\mathrm{PSP}}(t) + \mathrm{PSP}(t)$), the effective $\widetilde{\mathrm{PSP}}$ will again show a biologically reasonable shape (starting at 0, quickly increasing, and decaying back to 0).

8 Summary

8.1 Preliminaries

The well known model of Hodgkin-Huxley (HH) has been simplified to the differential equation

$$C_m \frac{dV_t}{dt} = -I_{\mathrm{Na}}(t) + I(t) \tag{33}$$

where $I(t)$ is an external current and $I_{\mathrm{Na}}(t) = g_{\mathrm{Na}} m_\infty^3(V_t) h_\infty(V_t - \tau_m \dot{V}_t)(V_t - E_{\mathrm{Na}})$. The dynamics of this model indicates not only a sensitivity on the potential V_t but as well on its slope $\dot{V}_t$, which motivates the definition of the *dynamic I&F neuron*, given by the equations above and the threshold condition $V_t + \dot{V}_t > 1$ with a reset to $V_t = \dot{V}_t = 0$ after every spike.
The membrane potential caused by a noisy input current $I(t)$ is modeled by a discrete time approximation of time step Δ, given by

$$\tilde{V}_t = V_t + a(\Delta)\xi_t \tag{34}$$

$$\dot{\tilde{V}}_t = \dot{V}_t + a(\Delta)\dot{\xi}_t \tag{35}$$

where ξ_t is a discrete time Ornstein-Uhlenbeck (OU) process and $\dot{\xi}_t$ a formal derivative given by definition 7.1. Since the HH model is deterministic, the scaling $a(\Delta)$ is chosen such that $\lim\limits_{\Delta \to 0} \tilde{V}_t = V_t$ and $\lim\limits_{\Delta \to 0} \dot{\tilde{V}}_t = \dot{V}_t$.
The aim of the thesis was to map the HH model, simplified to this setting, to an escape noise model, viz. to yield an expression for the interspike interval (ISI) distribution of the form

$$P_I(t|\hat{t}) = \rho(V_t, \dot{V}_t) \exp\left\{ -\int_{\hat{t}}^{t} \rho(V_s, \dot{V}_s) ds \right\} \tag{36}$$

in the limit $\Delta \to 0$. This obviously requires the firing rate $\rho(V_t, \dot{V}_t)$ to survive and stay finite in the limit $\Delta \to 0$ which is equivalent to being independent of Δ. To be a feasible model for reinforcement learning, the firing rate $\rho(V_t, \dot{V}_t)$ is additionally required to be continuous and differentiable in both variables.

8.2 Results

The most intuitive way of modeling noise by a gaussian distributed input, which yields a gaussian distributed OU process ξ_t, turns out not to allow a mapping to an escape noise model. The reason for this being the simple fact, that there is no scaling $a(\Delta)$ yielding a firing rate $\rho(V_t, \dot{V}_t)$ independent of Δ.
The same setting has been analyzed for an input noise that is assumed to have a stable distribution of index $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$, where $\alpha = 2$ would be a gaussian distribution. The resulting OU process is called Ornstein-Uhlenbeck-Lévy (OUL) process and allows a mapping to an escape noise neuron, as shown in theorem 7.21:

Theorem. *Let the dI&F neuron with stochastic input be given by* $\tilde{V}_t$, $\dot{\tilde{V}}_t$ *as defined above with* $a(\Delta) = C\Delta$. *Let* ξ_t *be an OUL process as in definition 5.21 with* $\alpha \in (\frac{1+\sqrt{5}}{2}, 2)$ *and* $V_s + \dot{V}_s \leq 1 - \varepsilon$ *for* $s \in [\hat{t}, t]$ *and some* $\varepsilon > 0$. *Then the ISI distribution of the dI&F neuron with sochastic input is given by*

$$P_I(t|\hat{t}) = \rho(V_t, \dot{V}_t) \exp\left\{ - \int_{\hat{t}}^{t} \rho(V_s, \dot{V}_s) ds \right\}$$

in the limit $\Delta \to 0$. *Moreover* $P_I(t|\hat{t})$ *is continuous and differentiable in* w.

Hence, the mapping of the heavy-tailed dI&F neuron to an escape noise neuron works only for subthreshold stimulus such that $V_t + \dot{V}_t \leq 1 - \varepsilon$ for some $\varepsilon > 0$. In contrast, common escape noise models are well-defined for arbitrary stimulus.

8.3 Discussion

The rapid and variable spike initiation process has recently been debated in terms of the underlying biophysical process [14, 15, 25]. The ideas presented in this work offer a computational perspective to this phenomenon. It particularly explains why the sodium current I_{Na} is inactivating when the voltage depolarizes. In fact, the existence of the gating variable h in the HH model implies, that action potentials (spikes) can only be generated if there is a strong **and** fast depolarization. This not only makes the spikes more sparse, but – with the noise scenario presented here – also makes the spike generation be an instantaneous function of the voltage and its derivative. But being an instantaneous function, the synapses can easily calculate their contribution to the postsynaptic activity and, for instance, implement a learning rule which follows the gradient of the expected reward [24].

Introducing heavy-tailed noise instead of gaussian noise at least allows a mapping for subthreshold stimulus. This raises the question, whether noise caused by random activity of other neurons is necessarily gaussian, which could be studied further since heavy tailed weight distributions have been experimentally found in [13, 22] and where suggested to emerge from a spike-timing dependent synaptic plasticity [8].

Even though mapping the dI&F model to an escape noise model worked only for subthreshold stimulus, the results of this work give interesting insights about how attempts for a complete mapping could be designed. The sensitivity to the slope $\dot{V}_t$ offers the possibility to yield an instantaneous firing rate and should therefore be retained. Since the model has to be feasible for arbitrary stimulus, the input noise needs to be implemented in a different way, e.g. by modeling the ion channels by a Poisson process with a rate depending on V_t and $\dot{V}_t$.

The existence of a real mapping would be a computational link between the biologically reasonable model of HH and the escape noise model and thus an important motivation for working on escape noise models.

References

[1] Borak Szymon, Härdle Wolfgang, Weron Rafal; *Stable Distributions*; Humboldt-Universität zu Berlin; 2005

[2] Bottou Léon; *Online Learning and Stochastic Approximations*; AT&T Labs-Research, Red Bank; 2012

[3] Chechkin A.B., Metzler R., Klafter J., Gonchar V. Yu.; *Introduction to the Theory of Lévy Flights*;

[4] Çinlar Erhan; *Introduction to Stochastic Processes*; Prentice-Hall, Inc.; 1975

[5] Cox D.R., Miller H.D.; *The Theory of Stochastic Processes*; Chapman&Hall; 1965

[6] Dayan Peter, Abott L.F.; *Theoretical Neuroscience*; The MIT Press; 2001

[7] Garbaczewski Piotr, Olkiewicz Robert; *Ornstein-Uhlenbeck-Cauchy process*; AIP Journal of Mathematical Physics; 2000

[8] Gilson M., Fukai T.; *Stability versus neuronal specialization for STDP: long-tail weight distributions solve the dilemma*; PLoS ONE, Vol. 6; 2011

[9] Gerstener Wulfram, Kistler Werner; *Spiking Neuron Models*; Cambridge University Press; 2002

[10] Hu Junling, Wellman Michael P.; *Multiagent Reinforcement Learning: Theoretical Framework and Algorithm*; Artificial Intelligence Laboratory, University of Michigan;

[11] Janczura Joanna, Orzel Sebastian, Wylomanska Agnieszka; *Subordinated α-stable Ornstein-Uhlenbeck process as a tool for financial data description*; Elsevier Inc.; 2011

[12] Kandel Eric, Squire Larry; *Memory: From Mind to Molecules*;Roberts and Company Publishers; 2008

[13] Lefort S., Tomm C., Floyd Sarria J. C., Petersen, C. C.; *The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex*; Neuron, Vol. 61; January 2009

[14] McCormick D. A., Shu Y., Yu Y.; *Neurophysiology: Hodgkin and Huxley model–still standing?*; Nature, Vol. 445; January 2007

[15] Naundorf B., Wolf F., Volgushev M.; *Unique features of action potential initiation in cortical neurons*; Nature, Vol. 440; April 2006

[16] Nolan John P.; *Stable Distributions, Models for Heavy Tailed Data*; American University; Processed May 13, 2009

[17] Papapantoleon Antonis; *Stochastic Analysis for Jump Processes*; TU Berlin; 2012

[18] Plessner Hans E., Gerstner Wulfram; *Escape Rate Models for Noisy Integrate-and-Fire Neurons*; Neurocomputing, Elsevier, pages 219-224; June 2000

[19] Plesser Hans E., Gerstner Wulfram; *Noise in integrate-and-fire neurons: from stochastic input to escape rates* ; Neural Computation 12, pages 367-384; 2000

[20] Rauch Alexander, La Camera Giancarlo, Lüscher Hans-Rudolf, Senn Walter, Fusi Stefano; *Neocortical Pyramidal Cells Respond as Integrate-and-Fire Neuron to In Vivo-Like Input Currents; Institute of Physiology*; University of Bern; 2003

[21] Smith Philip L.; *From Poisson shot noise to the integrated Ornstein-Uhlenbeck process*; Elsevier Inc.; 2009

[22] Song S., Sjostrom P. J., Reigl M. , Nelson S. , Chklovskii D. B.; *Highly nonrandom features of synaptic connectivity in local cortical circuits*; PLoS Biol., Vol. 3; March 2005

[23] Tuckwell Henry C.; *Introduction to theoretical neurobiology*, Volume 2; Cambridge University Press; 2005

[24] Urbanczik R., Senn W.; *Reinforcement learning in populations of spiking neurons*; Nat. Neuroscience, Vol. 12; March 2009

[25] Yu Y.; Shu Y., McCormick D. A.; *Cortical action potential backpropagation explains spike threshold variability and rapid-onset kinetics*; J, Neuroscience., Vol. 28; July 2008

[26] http://www.zhang.pi.titech.ac.jp/en/?Project:Equivalent_Circuit_of_Neuron; December 11, 2012

[27] http://ecee.colorado.edu/∼ecen4831/HHsumWWW/HHsum.html, December 11, 2012

[28] http://www2.chemie.uni-erlangen.de/projects/vsc/chemoinformatik/erlangen/datenanalyse/nn_kneuron.html, Decemeber 11, 2012

Erklärung

gemäss Art. 28 Abs. 2 RSL 05

Name/Vorname: Caspar Roman

Matrikelnummer: 04-106-209

Studiengang: Mathematik

Bachelor ☐ Master ☒ Dissertation ☐

Titel der Arbeit: Mapping the Hodgkin-Huxley Neuron to an
Escape Noise Neuron

LeiterIn der Arbeit: Prof. Dr. Walter Senn

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetztes vom 5. September 1996 über die Universität zum Entzug des auf Grund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, 21.04.2013
Ort/Datum

Unterschrift