UNIVERSITÄT OSNABRÜCK

Adaptation of Neuronal Activation Functions to Arbitrary Distributions of In- and Output

April 20th, 2015

Author:

Johannes LEUGERING

Supervisors:

Prof. Dr. Gordon PIPA

MSc. Kristoffer APPEL

A thesis submitted in fulfillment
of the requirements for the degree of
Master of Science *(MSc.)*

in the

Institute of Cognitive Science, University of Osnabrück

Declaration

I hereby certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other university.

Signed: Osnabrück, April 20th 2015

<div style="text-align:right">

_____

Johannes Leugering

</div>

Abstract

This thesis investigates a probability theoretical approach to implementing *homeo-static intrinsic plasticity* in a *linear-non-linear-Poisson-spiking* (LNP) neuron model. A method is presented to derive a model neuron's activation function from an observed distribution of its membrane potential and a desired or observed distribution of activation or spiking output. For the special case of *exponential family* distributions of membrane potential and activation, some interesting properties and a simple, biologically plausible mechanism for homeostatic intrinsic plasticity are derived and analyzed. Links to *dynamical systems* and *generalized linear models* are established. Numerical simulations are used to validate the most interesting theoretical conclusions.

"Adapt or perish, now as ever, is Nature's inexorable imperative."

– H. G. WELLS
The Mind at the End of its Tether (1945)

Contents

List of Figures

Motivation and Introduction

A remarkable feature of biological systems is their ability to self-regulate, to sense and correct errors in their own behaviour, and thus maintain stable operation in changing environments, adapt to new environments and recover from damage. This key observation lead Ashby (1954) to develop his view of the brain in particular and cybernetics in general based on the concept of *adaptation* (or *homeostasis*), a property he considered fundamental for living organisms as well as many physical systems, which he defined as follows:

> [A] form of behaviour is **adaptive** if it maintains the essential variables [of the system] within physiological limits.[1]

He proposed that "unless the environment is wholly inactive, [the stabilizing effect of adaptation] is *necessary* for survival"[2] and concluded that "'adaptive' behaviour is equivalent to the behaviour of a stable system".[3] This however supposes the existence of adaptation mechanisms (active or passive) that are able to respond to and counter-act undesirable changes in the system – "*the constancy of some variables may involve the vigorous activity of others*".[4] From a modern perspective, the idea of perfect stability of a system might seem inadequate for describing systems subjected to as much noise as biological systems are, because it was formulated in a cybernetics framework that studies dynamical systems, for which stable and unstable (and thus *desirable* and *undesirable*) regimes are clear cut basins of

---

[1] Ashby 1954, p. 57.   [2] Ibid., p. 65.   [3] Ibid., p. 64.   [4] Ibid., p. 67.

attraction around well defined attractors. But the intriguing notion of a system defind on a state space with *desirable* and *undesirable* regimes can be generalized to a probability space with a density of *desirability*, which can be interpreted in a probabilistic framework as the prior probability of a "healthy" system to find itself at a certain point in the probability space. From this perspective, the system then needs to maintain a certain stable probability distribution over its states in response to its stochastic inputs in order to survive.

If cortical neurons are viewed as self-regulating systems in such a probabilistic framework, then their remarkable ability to maintain *desirable* levels of activation in spite of drastically changing distributions of synaptic inputs raises the question, what biological processes might underly this adaptation. Many candidate mechanisms for such homeostatic regulation have been suggested,[5] but their theoretical implications, minimal requirements and limits have not been conclusively dealt with.

In particular, considering mechanisms of *intrinsic plasticity* (i.e. mechanisms confined within the neuron), it's an open question how much and what kind of information about its surrounding is required for the neuron to be able to regulate its own activity. Some suggested mechanisms use simultaneous observations of the model neuron's in- and output and apply numerical methods to find parameters which minimize a loss function, that either measures the amount of information transmitted by the neuron,[6] or the deviation of the neuron's actual distribution of activity from a desired distribution.[7] While these approaches can be used to evaluate and optimize existing neuron models, they rely on information (tuples of membrane potential and simultaneous activation) that may be more specific than necessary[8] and algorithms that are hard to motivate biologically[9]. They fail to provide a satisfactory, abstract, minimal model that solves this problem in an

---

[5] Turrigiano and Nelson 2000.   [6] Joshi and Triesch 2009; Buesing and Maass 2010; Toyoizumi et al. 2005; Tishby, Pereira, and Bialek 2000; Klampfl, Legenstein, and Maass 2009; Bell and Sejnowski 1995; Stemmler and Koch 1999.   [7] Joshi and Triesch 2009; Savin, Joshi, and Triesch 2010; Triesch 2007.   [8] The desired result is the stable mapping from a probability distribution of the membrane potential to a probability distribution of its activation, thus in principle un-timed information about these two distributions might be sufficient.   [9] In several of the references, gradient descent rules are derived that result in fairly complex expressions, relating many variables and parameters in non-linear ways.

optimal way.

This thesis follows a different approach by presupposing a (time-varying) parametric distribution of membrane potential and activation and inferring a thusly parameterized adaptive optimal neuron model from these distributions. The result is a neuron model that exactly realizes the desired distribution of activation and thus serves as the upper limit of what intrinsic plasticity could achieve. The idea of such an optimal neuron model for some static probability distributions has been implicit in e.g. (Triesch 2007) as a test-case against which to compare the adapted neuron models and has been used to establish optimality of the encoding of some biological neurons.[10] Yet I am not aware of a fully developed theory that generalizes this concept to arbitrary distributions and allows the resulting optimal model to adapt to changing distributions.

This thesis attempts to provide such a framework in order to theoretically address two principal questions:

1. (How) can the operation of a neuron be derived from measurements of the statistical properties of its in- and output?

2. (How) can an intrinsic homeostatic mechanism be implemented that keeps the output statistics in a "desirable" regime?

To this end, chapter 2 introduces the basic linear-non-linear-Poisson (LNP) neuron model, some concepts from probability theory and how they can be used to derive an optimal adaptive neuron model. In chapter 3, this general definition of the adaptive neuron model is constrained to the class of exponential family distributions and a simple, yet interesting, example neuron model is derived, analyzed and discussed in more detail. Chapter 4 shows numerical results to support the main conclusions of chapters 2 and 3. Finally, chapter 5 presents a short summary of the thesis, relates it to other research and ends with suggested directions for further research.

---

[10] Laughlin 1981.

Theoretical Background

2.1 Linear-Nonlinear Point Process Neurons

This thesis focuses on neuron models of the *linear-non-linear-point-process* variety[1], where the neuron's spiking output is modeled as a stochastic point process driven by a time-varying *activation*. This activation at each point in time is modeled by a *linear transformation* of the neuron's inputs[2] into what is interpreted as the neuron's *membrane potential*, followed by a *non-linear transformation* by the neuron's *activation function*[3]. The binary spiking events drawn from a time-varying activation constitute a *Poisson process*, which is why these models are referred to as *linear-non-linear-Poisson-spiking neuron models* (*LNP models* for short) from here on. This class of neuron models can be derived as a statistical abstraction of more physiologically motivated dynamical systems models such as the commonly used *(Non-Linear-)Integrate-and-Fire* model.[4]

See figure 2.1.1 for an illustration of the different components of such a *LNP* model.

As discussed in the following, LNP models exhibit very desirable properties for

---

[1] See (Ostojic and Brunel 2011) for a motivation, derivation and comparison of such models with more biologically inspired spiking neuron models.   [2] The linear integration of inputs across different synapses as well as filtering in time with a filter that models the synapses' and membrane potential's temporal dynamics can be subsumed in this linear transformation.   [3] The activation function is also referred to as *transfer function* or *(static) non-linearity*   [4] Ostojic and Brunel 2011.
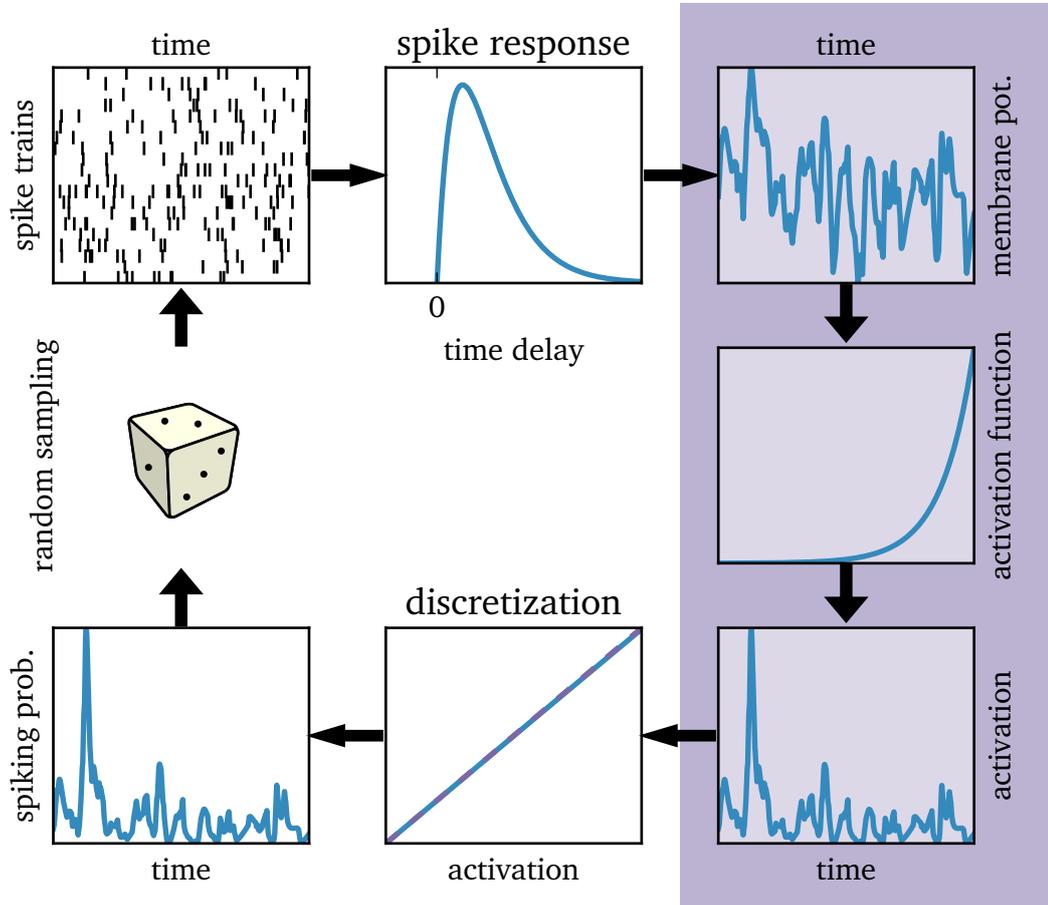
Figure 2.1.1: Clock-wise from the top left: Spike trains from presynaptic neurons are linearly combined and filtered to yield the neuron's membrane potential. A non-linear activation function maps the membrane potential to the neuron's instantaneous activation (or *rate*). In discrete time, the instantaneous rate is mapped to a spiking probability within one time-step. For very small time-steps, this mapping approaches a linear function. Finally, a new spike-train is drawn from the time-varying spiking probability distribution. Effects like refractoriness can be included by adding a virtual feedback-connection that uses a different spike response filter and contributes to the neuron's membrane potential. The theoretical results of this thesis mostly concern the mapping from membrane potentials to activations highlighted in the right column of the above figure.

statistical inference and have thus gained popularity in recent years, in particular in the form of *generalized linear models* (GLMs) which have been proposed for modeling neuronal spiking activity in a wide range of experimental setups.[5]

Mathematically, the LNP models used in this thesis can be summarized by equations 2.1.1 to 2.1.4.

$$X_j(t) = \sum_{i=1}^{T_j} \delta(t - \tau_{i,j}) \tag{2.1.1}$$

$$V(t) = \int_0^t \eta(\tau) \sum_j X_j(t - \tau) \omega_j \mathrm{d}\tau \tag{2.1.2}$$

$$\lambda(t) = f(V(t)) \tag{2.1.3}$$

$$P(\tau_{T_0+1,0} \in [t; t + \triangle t]) = 1 - \exp\left(-\int_t^{t+\triangle t} \lambda(\tau) \mathrm{d}\tau\right) \tag{2.1.4}$$

In equation 2.1.1, X_j denotes synapse j's spiking output over time, which is a sum of δ-pulses[6] centered at the times $\tau_{i,j}$ where synapse j fires. The total number of spikes emitted thus far by synapse j is denoted T_j. The pre-synaptic spike trains X_j are scaled by the corresponding synaptic weights ω_j, summed and filtered with the (causal) filter $\eta(t)$ to yield the function $V(t)$, interpreted to represent the neuron's membrane potential at time t. The non-linear activation function f in equation 2.1.3 maps this membrane potential to the neuron's instantaneous activation $\lambda(t)$. The neuron's own spiking output (represented by the index $j = 0$) is generated probabilistically by an inhomogeneous Poisson process with time-varying rate $\lambda(t)$. To this end, the probability of emitting at least one spike within the next time interval of length $\triangle t$ can be derived as a function of the total probability mass in that interval.

---

[5] Gerwinn, Macke, and Bethge 2010; Truccolo et al. 2005; Kass and Ventura 2001; Haslinger et al. 2012; Ostojic and Brunel 2011; Paninski 2004.   [6] Strictly speaking, this makes $X_j(t)$ a *distribution* rather than a function. However, this terminological inconsistency is resolved by the convolution in equation 2.1.2 with a filter $\eta(t)$, which results in a proper function $V(t)$ nevertheless.

In discrete time with T time-steps of size $\triangle t$, this can be rewritten as:

$$X \in \{0,1\}^{T \times N} \tag{2.1.5}$$

$$X_{i,j} = 1 \Leftrightarrow \exists k : \tau_{k,j} \in [i\triangle t; (i+1)\triangle t] \tag{2.1.6}$$

$$V = \eta \star X \omega \tag{2.1.7}$$

$$\lambda_i = f(V_i) \tag{2.1.8}$$

$$P_{i+1} = 1 - \exp\left(-\triangle t \cdot \lambda_i\right) \tag{2.1.9}$$

Here, X can be defined as a binary matrix with columns representing pre-synaptic spike trains. If $\triangle t$ is chosen sufficiently small, each time-interval can contain at most 1 spike due to refractoriness of the neuron, thus $X_{i,j} = 1$ represents a spike at synapse j and time-step i, whereas $X_{i,j} = 0$ represents no spike. V, η, λ and P are column vectors corresponding to the likewise named quantities for continuous time. The "$\star$" in equation 2.1.7 symbolizes discrete convolution. The probability of the neuron to spike within time-step $i+1$ is then given by P_{i+1}, which is calculated under the assumption of a constant rate λ_i throughout the ith time-interval of length $\triangle t$.

In the scope of this thesis, such models are used for one compelling reason:

> [They offer] a practical, broadly applicable solution to the computa-
> tional problem of fitting potentially complex point process models for
> neural spike trains by maximum likelihood [estimation].[7]

As discussed in section 3.3, the fact that *maximum likelihood estimation* is a suitable method for fitting a LNP model's parameters allows for the derivation of an adaptive neuron model that achieves *homeostasis* by continuously estimating the parameters of its input distribution.

However, a principal difference in the usage of LNP models, in particular GLMs, in thesis thesis compared to how they are also used in the analysis of neural data should be pointed out: While they can be used just as generic machine-learning tools suitable for "the analysis of the simultaneous effects of extrinsic covariates" on "concurrent ensemble spiking activity",[8] they are here viewed as statistical simplifications of the actual mechanism underlying spike generation in neurons.

---

[7] Truccolo et al. 2005.   [8] Ibid.

Thus less emphasis is placed on the computational power of such models, rather than on their mathematical simplicity and utility as models of biological neurons.

2.2 Mapping Membrane Potentials to Activations

In the previous section, the mapping $f : V \to \lambda$ in equations 2.1.3 and 2.1.8 is left unspecified. Many activation functions have been proposed in different contexts: from step functions[9] or entire logical predicates[10] in early perceptrons, via sine,[11] sigmoidal, conic parabolic and radial basis functions used in artificial neural networks,[12] to yet more general abstract classes[13] and biologically inspired functions.[14] Often, the choice of an activation function is based on its mathematical convenience[15] and performance for computation[16] or on simplifying assumptions about the operation of the neuron and its parameters are then fitted to neural data (see e.g.[17]).

In contrast to that, this section describes how the statistics of the neuron's input (membrane potential) and output (activation) alone (almost) uniquely determine the neuron's activation function, and how this insight can be used to construct an adaptive neuron model.

2.2.1 Mappings from Arbitrary In- to Output Distributions

A neuron's membrane potential, resulting from a bombardment of incoming spikes, can be modeled as a stochastic process such as the commonly used Ornstein-Uhlenbeck process.[18] This implies that for any point in time, the neuron's membrane potential can be expressed as a random variable, the distribution of which we refer to as the *membrane potential* or *input distribution*. The neuron's activation, a deterministic transformation of its stochastic input, is thus a random variable, as well, the distribution of which we refer to as the *activation* or *output distribution*. See figure 2.2.1 for an illustration of how different commonly used activation

[9] Rosenblatt 1958. [10] Minsky and A. Papert 1969; McCulloch and Pitts 1943. [11] Sopena, Romero, and Alquezar 1999. [12] Karlik and Olgac 2010. [13] DasGupta and Schnitger 1993. [14] Toyoizumi et al. 2005. [15] Sopena, Romero, and Alquezar 1999; Paninski 2004. [16] Karlik and Olgac 2010. [17] Toyoizumi et al. 2005. [18] Shimokawa and Shinomoto 2009.

functions can map the same normally distributed membrane potential to very different distributions of activation.

Given an input and an output distribution, this section explains how the neuron's non-linear activation function can be determined. The motivation for this approach, rather than using one of the many proposed activation functions and fitting them to neural data, is three-fold: Since these distributions only need to be observed, this might allow estimation of activation functions in vivo and could thus supplant or augment similar techniques like *white noise analysis*,[19] which require actively injecting stochastic inputs into the neuron. Furthermore, since no precise pairings $(V(t), \lambda(t) = f(V(t)))$ of input and corresponding output need to be known, recordings from different neurons could be efficiently pooled to estimate the neurons' input and output distributions and thus determine their expected activation functions. This also alleviates the problem that neural processing might not be instantaneous, because in- and output distributions are independent of any temporal structures.

We formalize this idea in theorem 5 after stating a few preliminary definitions and theorems. The proofs of all theorems presented in this section and lemmas required in the process can be found in Appendix A.

Definition 1 (Space of Cumulative Distribution Functions)**.** We define Π to be the space of cumulative distribution functions (CDFs), i.e. the space of non-decreasing, right-continuous functions $F : \mathbb{R} \to (0,1)$ with limits $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

To allow general statements about continuous as well as non-continuous random variables, a *generalized inverse*[20] needs to be defined that extends the concept of a function inverse to cumulative distribution functions, which in general are not invertible in the strict sense. Using specific properties of the space Π, this allows defining a quantile function for arbitrary random variables which can be considered an inverse function in a weaker sense.

---

[19] Chichilnisky 2001.   [20] Embrechts and Hofert 2013.
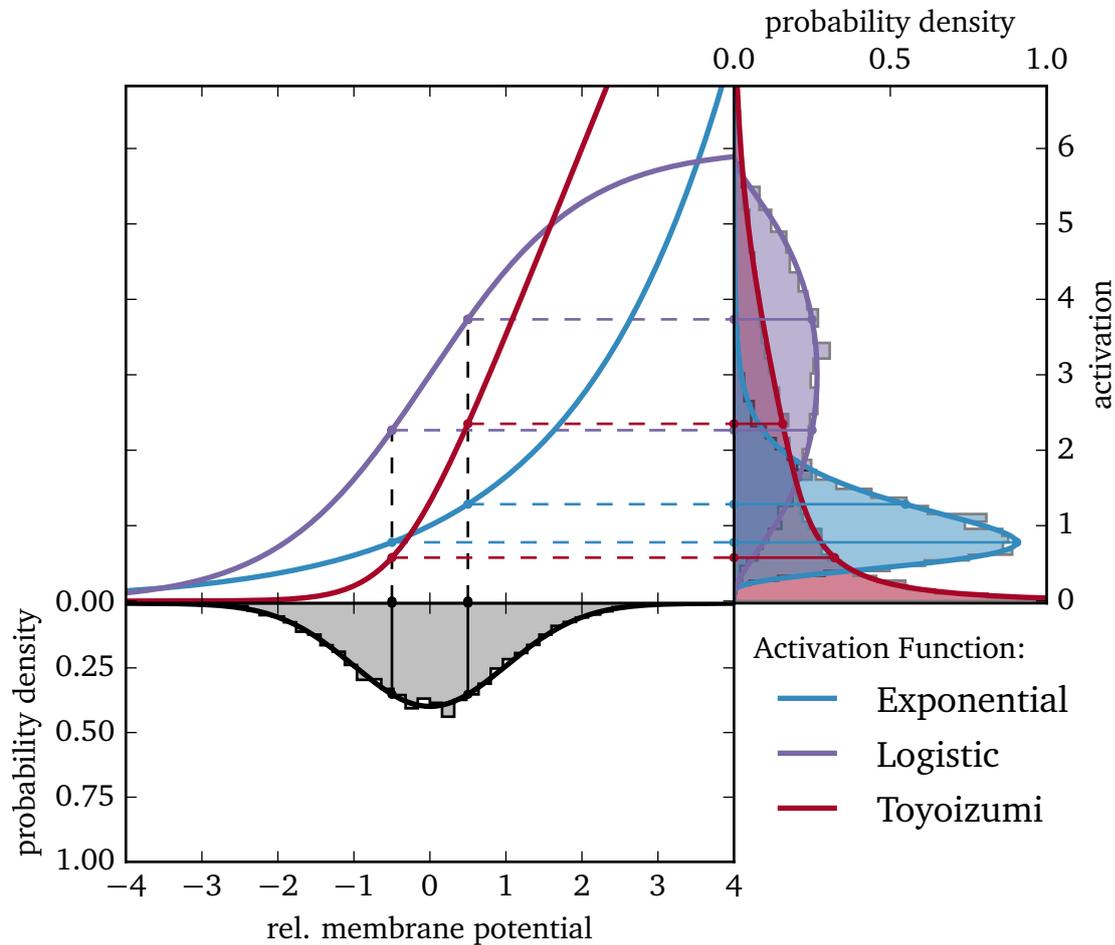
Figure 2.2.1: A random membrane potential with normal distribution is mapped onto an activation variable via 3 different activation functions. The resulting distributions of activation are qualitatively different. Solid lines show theoretically derived distributions on top of histograms for 5000 random samples and the respective function values of the activation functions. The activation functions used are a simple exponential function as used later in this thesis, the logistic function commonly used in artificial neural networks and the neuron model used by Toyoizumi et al. (2005), all using arbitrarily chosen parameters to illustrate the variability of resulting activation distributions.

Definition 2 (Generalized Inverse)**.** Let $f \in \Pi$ be a cumulative distribution function. We define the *generalized inverse* of f by the function

$$g : (0, 1) \to \mathbb{R}, y \mapsto \min\{x \in \mathbb{R} | f(x) \geq y\}$$

The minimum in the above expression is well defined, since f is by definition right-continuous and thus $f(\inf\{x \in \mathbb{R} | f(x) \geq y\}) = f(y) \Rightarrow \inf\{x \in \mathbb{R} | f(x) \geq y\} \in \{x \in \mathbb{R} | f(x) \geq y\}$, i.e. $\min\{x \in \mathbb{R} | f(x) \geq y\} = \inf\{x \in \mathbb{R} | f(x) \geq y\}$.

The generalized inverse G of the cumulative distribution function F of the random variable X is also called the probability distribution's *quantile function*, because for continuous random variables it allows defining the p-quantile, i.e. the interval $(-\infty, G(p)] = (-\infty, x] \subset \mathbb{R}$ for which $p = P(X \in (-\infty, x]) = F(x) = F(G(p))$. This equivalence also illustrates that for continuous random variables, the generalized inverse is identical to F^{-1}, the standard inverse[21].

Using these definitions, two well known theorems from probability theory, the *Probability Integral Transform* and the *Inverse Probability Integral Transform*, can be introduced.[22] They state that:

Theorem 3 (Probability Integral Transform)**.** *Let X_1 be a continuous random variable with cumulative distribution function $F_1 \in \Pi$. Then the random variable $U := F_1(X_1)$ has uniform distribution on the open interval $(0, 1)$.*

Theorem 4 (Inverse Probability Integral Transform)**.** *Let U be a random variable with uniform distribution on the open interval $(0, 1)$ and let $F_2 \in \Pi$ be a cumulative distribution function. Let G_2 be the generalized inverse of F_2 as defined in definition 2. Then the random variable $X_2 := G_2(U)$ has the cumulative distribution function F_2.*

---

[21] For non-continuous random variables, the CDF can be locally constant and may thus not be invertible in the traditional sense.   [22] Angus 1994.

With the two theorems stated above, the main corollary of this section can be constructed:

Theorem 5 (Transfer Function Theorem). *Let X_1 be a continuous random variable with cumulative distribution function $F_1 \in \Pi$ and let $F_2 \in \Pi$ be a cumulative distribution function. Let G_2 be the generalized inverse of F_2 as defined in definition 2.*

Then the random variable $X_2 := \tau(X_1)$ with $\tau := G_2 \circ F_1$ has the cumulative distribution function F_2. The function τ is called transfer function from X_1 to X_2 *and is non-decreasing and left-continuous.*

This theorem has the strong implication, that for any continuous random variable, a well defined function can be found that maps it onto a random variable with arbitrarily defined probability distribution. Since this theorem is used below to derive a neuron's activation function given its in- and desired output distribution, the question arises, whether the function τ defined in theorem 5 is the only sensible choice or just one among many. The following theorem addresses this question.

Theorem 6 (Uniqueness of the Transfer Function). *Let X_1, F_1, F_2 and G_2 be as defined in theorem 5. Let $\tilde{\tau} : \mathbb{R} \to \mathbb{R}$ be a non-decreasing, left-continuous function such that the random variable $\tilde{\tau}(X_1)$ has the cumulative distribution function F_2. Then $\tilde{\tau} = \tau = G_2 \circ F_1$.*

Thus τ is in fact the only reasonable choice![23] We have thus established a method to uniquely recover from just the cumulative distribution functions of two random variables X_1 and $X_2 := f(X_1)$, where f is a non-decreasing left-continuous

---

[23] Monotonicity is a reasonable requirement for neuronal activation functions. The constraint of left-continuity could be dropped, allowing the function values of τ to be chosen more freely on the set of discontinuities of τ. Relinquishing left-continuity for this seems unjustified though, because any transfer function would then still have to agree with τ *almost everywhere*, thus gaining little additional freedom at the expense of mathematical simplicity.

function, the transformation f between the two. Since this was derived in a very general way, allowing any continuous distribution for X_1 and any (continuous, discrete or neither) probability distribution for X_2, the following sections present a few examples to illustrate the utility of this result.

For proofs of the theorems above, see Appendix A.

2.3 Intrinsic Plasticity

Theorem 5 from the previous section provides a method to determine, given two random variables X_1 and X_2 with cumulative distribution functions F_1 and F_2, respectively, a function τ that maps the random variable X_1 onto a random variable with the same distribution as X_2. This function τ is defined by just the composition of the (generalized) inverse of F_2 with F_1 and can thus be calculated as soon as F_1 and F_2 are known. In the context of LNP-neurons, this can be used to define the activation function f of the neuron after determining the distribution of its membrane potential (F_1) and its activation (F_2). This could provide an improved way of determining neuronal activation functions from biological measurements: It is non-parametric in so far, as the activation function is parameterized only by parameters of the respective distributions F_1 and F_2, i.e. if F_1 and F_2 are fixed, the activation function has no more free parameters. It should prove simpler than methods that rely on simultaneous recordings from a neuron's dendrite and axon: Since only the *distribution* of membrane potential and activation need to be known, rather than tuples of membrane potential and activation *at fixed points in time*, the recordings need not be simultaneous and delay-effects induced by the biological processing in the neuron can be neglected. Furthermore, no regression-like techniques need to be used to fit a deterministic activation function through noisy data points of membrane potential and activation, since the stochasticity of both variables is adequately represented in their probability distributions, which then uniquely determine the activation function. Available literature can be surveyed for suggestions of membrane potential distributions as well as distributions of firing rates, and the implied activation function mapping one onto the other can be compared to contemporary neuron models (see section 3.3 for an extended

example).

In addition to that, since the activation function is parameterized by the parameters of the in- and output distribution, any change in a parameter of the input distribution can be compensated by also changing the corresponding parameter of the activation function:

Suppose the neuron's membrane potential V is a random variable with cumulative distribution function $F_1(V; \theta_1)$ with the parameter vector θ_1 and its activation λ is a random variable with cumulative distribution function $F_2(\lambda; \theta_2)$ with the parameter vector θ_2 and the generalized inverse G_2. If the neuron's activation function is defined as $f(V; \theta_1, \theta_2) := G_2(F_1(V; \theta_1); \theta_2)$, then it is fully parameterized by θ_1 and θ_2. Suppose now a change occurs in the probability distribution of V and θ_1 is updated to $\tilde{\theta}_1$. Then the random variable $f(V; \tilde{\theta}_1, \theta_2)$ again has the cumulative distribution function $F_2(\lambda; \theta_2)$. Thus, in order to maintain the distribution F_2 for its activation, the neuron has to adapt the parameters θ_1 of its activation function to the new parameters $\tilde{\theta}_1$. Since these parameters coincide with the parameters of the input distribution itself, maintaining a constant output distribution becomes a question of maintaining close approximations of the parameters of the neuron's input distribution.

By keeping the neuron's own output in a desirable regime (one could view the enforced output distribution as providing a measure of desirability of the potential outputs), this method implements the basic idea of *homeostatic intrinsic plasticity* as outlined in chapter 1. For the convenient special case of *exponential family* input distributions, these ideas are used in chapter 3 to formalize very simple homeostatic neuron models which are later numerically evaluated in chapter 4.

A second kind of neuronal plasticity could be characterized by changing the parameter vector θ_2, which allows the neuron to change its output distribution – a behavior that could be triggered by external effects such as LFPs or neuro-modulators and could be used to selectively (in)activate neurons or change their mode of operation (see chapter 5 for a few suggestions for how this concept could be utilized).

2.4   Stochastic Spike Generation

In sections [2.2](#) and [2.3](#), a method is presented to define the adaptive activation function f which has been left unspecified in equation [2.1.3](#) for continuous time and [2.1.8](#) for discretized time. This completes the specification of the deterministic part of the general LNP model used in this thesis, which leaves only the stochastically spiking point process to be discussed in this section.

2.4.1   Inverse Sampling

If a neuron's spiking output is modeled as a Poisson process with time-varying *rate* $\lambda(t)$, the probability p of the neuron to spike within a time-interval of fixed length $\triangle t$ at time t is given by equation [2.1.4](#) as $p = 1 - \exp(-\int_{t}^{t+\triangle t} \lambda(\tau)\mathrm{d}\tau)$[24]. Assuming that λ changes on a time-scale much slower than $\triangle t$, the integral can be approximated by $\int_{0}^{\triangle t} \lambda(t+\tau)\mathrm{d}\tau \approx \int_{0}^{\triangle t} \lambda(t)\mathrm{d}\tau = \lambda(t)\triangle t$. With this simplification, the spiking probability becomes $p = 1 - \exp(-\triangle t\lambda(0))$ as used in equation [2.1.9](#) for discrete time.

Given this probability, the neurons spiking output for this one time interval of length $\triangle t$ can be represented as a Bernoulli random variable, taking on a value of 1 (which corresponds to a spiking event within said interval) or 0 (which corresponds to no spike) with probabilities p and $1 - p$, respectively. A random variable like this can easily be generated form a uniform random variable U (*noise*) by using the inverse probability integral transform (see theorem [4](#)) derived in section [2.2](#):

The CDF F_{Bern} and its generalized inverse G_{Bern} of the Bernoulli distribution

---

[24] Temporal structure of spiking output such as refractoriness is disregarded here.

with parameter p are defined as:

$$F_{\text{Bern}}(x|p) = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 - p & \text{for } x \in [0, 1) \\ 1 & \text{for } x \geq 1 \end{cases} \tag{2.4.1}$$

$$G_{\text{Bern}}(x|p) = \min\{s \in \mathbb{R} | F_{\text{Bern}}(s|p) \geq X\} \tag{2.4.2}$$

$$= \begin{cases} 0 & \text{for } y \in (0, 1 - p) \\ 1 & \text{for } y \in [1 - p, 1) \end{cases} \tag{2.4.3}$$

Theorem 4 implies that X defined as follows has the CDF F_{Bern} with parameter p:

$$X := G_{\text{Bern}}(U|p) \tag{2.4.4}$$

$$= G_{\text{Bern}}(U|1 - \exp(-\triangle t \lambda(t))) \tag{2.4.5}$$

$$= \begin{cases} 0 & \text{for } y \in (0, \exp(-\triangle t \lambda(t))) \\ 1 & \text{for } y \in [\exp(-\triangle t \lambda(t)), 1) \end{cases} \tag{2.4.6}$$

However, this expression depends on the time-varying activation $\lambda(t) = f(V(t))$ which involves the neuron's activation function f that is defined in terms of the generalized inverse of the neurons activation distribution and might thus be difficult to compute. An alternative representation of X, which can be derived by equivalently transforming the condition $y \in (0, \exp(-\triangle t \lambda(t)))$ in equation 2.4.6, overcomes this problem:

17

$$y \in (0, \exp(-\triangle t \lambda(t))) \tag{2.4.7}$$

$$\Leftrightarrow \quad \frac{-\log(y)}{\triangle t} \in [\lambda(t), \infty) \tag{2.4.8}$$

$$\Leftrightarrow \quad \frac{-\log(y)}{\triangle t} \in [G_2(F_1(V(t))), \infty) \tag{2.4.9}$$

$$\Leftrightarrow \quad \frac{-\log(y)}{\triangle t} \in [\min\{s \in \mathbb{R} | F_2(s) \geq F_1(V(t))\}, \infty) \tag{2.4.10}$$

$$\Leftrightarrow \quad \frac{-\log(y)}{\triangle t} \in \{s \in \mathbb{R} | F_2(s) \geq F_1(V(t))\} \tag{2.4.11}$$

$$\Leftrightarrow \quad F_1(V(t)) \leq F_2\left(\frac{-\log(y)}{\triangle t}\right) \tag{2.4.12}$$

Substituting the expression in equation 2.4.12 back into equation 2.4.6 for the equivalent expression in equation 2.4.7 yields an alternative definition of the spiking output X that can be used to sample spikes from the membrane potential without the need to calculate or estimate the (generalized) inverse of the neuron's distribution of activations:

$$X := \left(F_1(V(t)) \leq F_2\left(\frac{-\log(y)}{\triangle t}\right)\right) \tag{2.4.13}$$

If F_1 is continuous, this can be further transformed:

$$X := \left(V(t) \leq F_1^{-1}\left(F_2\left(\frac{-\log(y)}{\triangle t}\right)\right)\right) \tag{2.4.14}$$

This demonstrates the equivalence of choosing an arbitrary activation function and subsequently sampling spikes to using a simple step-function with (suitably distributed) random thresholds. Such random thresholds have been used in conjunction with integrate-and-fire based neuron models,[25] for which random thresholds can be introduced very naturally and temporal dynamics can more easily be incorporated than in LNP models (see section 2.4.4).

Since for equation 2.4.13 only the cumulative distribution functions need to be

---

[25] Lazar, Pnevmatikakis, and Zhou 2010; Gestri, Mastebroek, and Zaagman 1980; Reich et al. 1997.

known, it might prove simpler to implement than equations 2.4.6 and 2.4.14 for commonly used distribution, which are often defined by their CDFs but need not have simple closed form quantile functions (i.e. generalized inverses).

This completes the specification of the basic LNP model as outlined in equations 2.1.1 to 2.1.9. The following section discusses, how this LNP models can be extended to cope with violations of the assumption of Poisson-spiking, such as temporal structure in the spike generation process resulting from refractoriness.

2.4.2 Incorporating refractoriness into the Neuron Model

A limitation of the LNP model described thus far is, that it assumes strictly Poisson spike generation, i.e. by a stochastic process that is memory-less. This however neglects the ample evidence that effects such as neuronal refractoriness are not just slight deviations from an ideal process due to physical limitations of the neuron but may in fact happen on a time-scale that is functionally relevant.

For example, an early study of locust neurons[26] quantifies the effect of relative changes in firing thresholds of the observed neurons as a function of time since the preceding spike. A sharp increase directly after the spike can be observed, followed by a roughly exponential decay back to the original threshold. See figure 2.4.1 for a comparison of the original data with a fitted exponential decay model. The observed decay time constants of $\approx$ 15ms are considerably longer than what could be expressed by absolute refractory periods alone.

The study concludes:

> The threshold recovery after impulse propagation in locust peripheral motor axons to flight muscles is described. The refractoriness is large enough to have functional significance for at least 10 msec, and may be detected after 50-100 msec. The refractoriness to successive impulses accumulates.[27]

Further biological studies indicate that refractoriness of neurons might play an important role in explaining the temporal precision of firing in biological neurons which simple LNP neuron models cannot reproduce.[28]

---

[26] Wilson 1964.   [27] Ibid.   [28] Berry and Meister 1998.
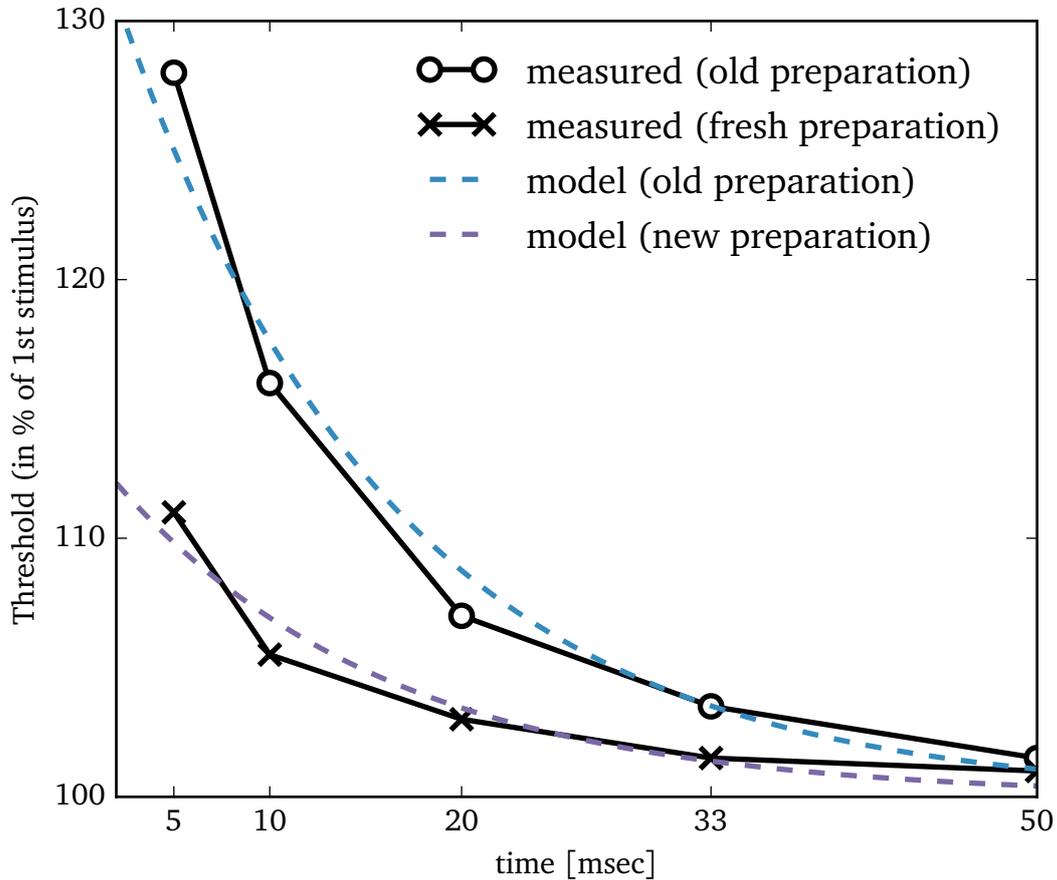
Figure 2.4.1: After inducing a spike in locust axons, relative changes in the threshold are measured as a function of time. Solid lines show data extracted from figure 1 of the original study by Wilson (1964) for two conditions (recorded in a fresh and an old preparation). Dashed lines show exponential functions fitted through the data-points using a generalized linear model with a log-link function (thus assuming Poisson-noise in the measurements). Although the fit is not very precise, it shows a roughly exponential decay towards 100% with a fixed time constant of 14.27ms (during fitting, both exponentials share the same time-constant and vary only in scaling).

SECTION 4. Stochastic Spike Generation

Due to these insights, a simple extension to the LNP framework has been suggested, where spike generation is modeled as an *inhomogeneous Markov Interval Process* (IMI-process),[29] rather than a Poisson-process. To this end, the neuron's activation variable $\lambda(t)$ is multiplied by a *recovery variable* $R(t - \tau_*)$ which depends on the time since the neuron's last spike τ_* and converges to 1 as $(t - \tau_*) \to \infty$. Refractoriness accumulated due to multiple previous spikes could similarly be included by multiplying with the product of the effects from previous spikes $\prod_{i \in \mathbb{N}} R(t - \tau_i)$ instead.

In terms of our LNP neuron model, this implies that equation 2.1.3 is replaced by equation 2.4.15:

$$\lambda(t) = f(V(t)) \cdot \prod_{i=1} R(t - \tau_{i,0}) \tag{2.4.15}$$

Here, R is the *recovery kernel* that models the recovery after each spike, the $\tau_{i,0}$ are spike times of the neuron itself and X_0 is its previous spike train.

However, this causes an obvious problem as it changes the distribution of the neuron's activation λ. In order to stay consistent with the interpretation that the activation function f is chosen to map the membrane potential to an activation with a given distribution, the effect of modulating the activation with a recovery variable needs to be compensated for.

One method, to maintain at least the expected value of the distribution of activation, could be to compensate the overall decrease in firing probability due to refractoriness by up-scaling the activation of the neuron accordingly. Although this fails to capture the different effects that refractoriness has for different levels of activation (due to temporal correlations, a neuron in a highly activated state has more likely spiked recently and is thus on average subject to stronger refractory effects), this might provide a reasonable approximation – in particular when firing rates are low on average. To this end, the expected value of the (cumulated) recovery variable, $\mathbb{E}_\delta[R(\delta)]$ needs to be calculated from the distribution of inter-spike-intervals (ISIs) δ. This could be done analytically for a given probability distribution of δ or numerically from a set of observed spike times.

A different approach is to use the fact that the activation function f is non-

---

[29] Berry and Meister 1998; Kass and Ventura 2001.

decreasing and non-negative, typically with $\lim_{v \to -\infty} f(v) = 0$, and introduce the recovery variable as an additive term of the activation function's input. Such feed-back of a neuron's output to its own input has also been used with dynamical systems models to model spike responses and after-effects.[30] After each of the neuron's spikes the *effective* input $\tilde{V}(t) = V(t) + \sum_i R(t - \tau_i)$ can be pulled towards large negative values, resulting in an activation $\tilde{\lambda}(t) = f(V(t) + \sum_i R(t - \tau_i))$ close to zero. While this approach is here not yet as well motivated as the previous suggestion, it offers a significant improvement: The influence of refractoriness, which is now incorporated as an additional summand in the neurons effective membrane potential $\tilde{V}$ like an additional inhibitory feed-back connection, is now compensated for by the intrinsic plasticity mechanism of the neuron itself.

For the special case of exponential family neuron models, these two approaches are shown to be compatible in chapter 3, resulting in a well motivated and convenient neuron model that incorporates auto-history dependence.

2.4.3 Fitting the Activation Function from an ISI distribution

The (extended) LNP model allows the extraction of a neuron's activation function from an observed distribution of its membrane potential and activation over time. However, the activation variable itself, introduced as a hidden variable underlying the spike generation process, cannot be observed itself in biological neurons. This section addresses that problem and offers a crude numerical way of fitting parameters of the activation distribution to raw spiking data, in case the activation variable itself is not observable.

This problem is typically avoided when fitting rate coding neuron models by averaging their spiking responses to identical stimuli over multiple trials into a so called *peri-stimulus time histogram* (PSTH), and subsequently smoothing the PSTH using techniques such as *generalized linear models* (GLMs; see section 3.3.5), *kernel density estimation*, *filtering*, *spline* fitting or *gaussian process regression*[31] to get a slow varying estimate of the neuron's activation. These mechanisms however mostly rely on further assumptions about e.g. the smoothness of the neuron's

---

[30] Keat et al. 2001; Kistler, Gerstner, and Hemmen 1997; Gerstner 1995.   [31] Behseta et al. 2007; Cunningham et al. 2008.

activation, which is reasonable as long as the stimulus injected into the neuron can be controlled. For passive in-vivo recordings, a PSTH cannot be created and the smoothness of the neuron's activation over time might not be known. Some sophisticated numerical methods to address the problem have been suggested,[32] but here, as a proof of concept, only a basic brute-force *maximum likelihood estimation* (MLE) approach to recovering the distribution of activation from a distribution of ISIs is presented.

The neuron's distribution of ISIs τ depends on the distribution of the neuron's activation λ via the spike generation mechanism described above and can be analytically derived once the activation is known. The distribution of λ is parameterized by an unknown parameter vector θ[33] that should be recovered. Using the Bayes theorem, the posterior distribution $P(\theta|\tau)$ can be derived from known quantities:

$$P(\theta|\tau) \propto P(\tau|\theta) = \int_0^\infty P(\tau, \lambda|\theta)\mathrm{d}\lambda = \int_0^\infty P(\tau|\lambda)P(\lambda|\theta)\mathrm{d}\lambda \qquad (2.4.16)$$

Here, $P(\tau|\lambda)$ represents the distribution of ISIs given a rate λ and $P(\lambda|\theta)$ is the distribution of λ with unknown parameters θ. Marginalizing out the activation λ leaves as the right hand side only the likelihood $P(\tau|\theta)$, proportional to the posterior $P(\theta|\tau)$ which can be maximized w.r.t. the parameters θ, thus giving a ML estimate $\hat{\theta}$. Maximizing this integral expression may prove difficult, but section 4.3 presents an example to demonstrate that this brute-force approach is in principle capable of recovering parameters of a simulated neuron's distribution of activation.

2.4.4 Relation to Integrate-and-Fire-type Models

The LNP model described so far, if auto-history effects are not included, is a memory-less point process. All the temporal dynamics concerning the integration of inputs that is characteristic for biological neurons has so far been modeled by a single filter $\eta(t)$ in equation 2.1.2. If η is chosen to include the synaptic response

---

[32] Cunningham et al. 2008.   [33] The basic shape of the distribution of λ is assumed to be known. Otherwise, parametric methods of approximating arbitrary distributions functions could be used here.

to a spike (transmitter release, re-uptake etc.) as well as the response of the membrane potential to the presence of neurotransmitters (leaky integration), the filtering can be split into two stages: given two filters η_S and η_M, that model the *impulse response* of the synapse and the membrane potential, respectively, then due to associativity of the convolution operation, the filter $\eta(t) := \int_0^t \eta_M(t-\tau)\eta_S(\tau)\mathrm{d}\tau$ models the combined response as it would result from filtering a spike-train with η_S to determine neurotransmitter concentration and then filtering it with η_M to model the leaky integration of the resulting input in the membrane potential.

As shown below, an appropriate choice of η_M allows our model to capture the same dynamic effects as the commonly used (nonlinear) *integrate-and-fire* (IF, also called *integrate-to-threshold*) neuron model. This provides an intuitive way to relate the LNP model defined above to the more biologically inspired integrate-and-fire model with random thresholds:[34]

Suppose the filter η is the convolution of two components, a truly synaptic filter η_S and an exponential membrane potential filter η_M defined as follows:

$$\eta_M(t) := \begin{cases} 0 & \text{if } t < 0 \\ \alpha_M \exp\left(-\alpha_M t\right) & \text{otherwise} \end{cases} \tag{2.4.17}$$

Here $\frac{1}{\alpha_M}$ is the membrane time-constant of the neuron. A new variable $I(t)$ is introduced to capture the cumulated synaptic responses resulting from pre-synaptic spikes.

If we define the set $T := \{(j,i) \in \mathbb{N}^2 | \tau_{j,i} < t\}$ of all the indices of pre-synaptic spikes fired before t, then the filter $\eta(t)$, the membrane potential $V(t)$ from

---

[34] Gestri, Mastebroek, and Zaagman 1980; Reich et al. 1997.

equation 2.1.2 and the post-synaptic input $I(t)$ can be written as:

$$\eta(t) = \int_0^t \eta_M(t - \tau)\eta_S(\tau)\mathrm{d}\tau \tag{2.4.18}$$

$$I(t) = \sum_{(j,i)\in T} \omega_j \eta_S(t - \tau_{j,i}) \tag{2.4.19}$$

$$V(t) = \sum_{(j,i)\in T} \omega_j \eta(t - \tau_{j,i}) \tag{2.4.20}$$

Using differentiation under the integral sign to calculate the derivative with respect to time, the membrane potential $V(t)$ can be expressed as a simple differential equation of first order:

$$\frac{\mathrm{d}\eta(t)}{\mathrm{d}t} = \eta_S(t)\eta_M(0) + \int_0^t \eta_M'(t - \tau)\eta_S(\tau)\mathrm{d}\tau \tag{2.4.21}$$

$$= \alpha_M \eta_S(t) - \alpha_M \int_0^t \eta_M(t - \tau)\eta_S(\tau)\mathrm{d}\tau \tag{2.4.22}$$

$$= \alpha_M \left(\eta_S(t) - \eta(t)\right) \tag{2.4.23}$$

$$\frac{1}{\alpha_M}\frac{\mathrm{d}V(t)}{\mathrm{d}t} = \frac{1}{\alpha_M}\sum_{(j,i)\in T} \omega_j \frac{\mathrm{d}}{\mathrm{d}t}\eta(t - \tau_{j,i}) \tag{2.4.24}$$

$$= \sum_{(j,i)\in T} \omega_j \eta_S(t - \tau_{j,i}) - \sum_{(j,i)\in T} \omega_j \eta(t - \tau_{j,i}) \tag{2.4.25}$$

$$= -V(t) + I(t) \tag{2.4.26}$$

Equation 2.4.26 is the well known expression of the integrate-and-fire neuron model[35] with resting potential $V_0 = 0$, driven by the synaptic input variable $I(t)$. Thus given an appropriately chosen filter η, the membrane potential of the LNP model presented here is equivalent to that of an integrate and fire neuron.

Assuming a continuous probability distribution with CDF F_1 for the membrane potential V, equation 2.4.14 can thus be used to generate spikes according to some target distribution with CDF F_2, resulting in the simple IF model with random

---

[35] Burkitt 2006a; Burkitt 2006b.

thresholds described by equations 2.4.27 to 2.4.28:

$$\tau_M \frac{\mathrm{d}}{\mathrm{d}t} V(t) = -V(t) + I(t) \tag{2.4.27}$$

$$\text{if } V(t) \geq \theta : \text{spike}, \quad \theta \leftarrow F^{-1}(F_2(\frac{-\log(U)}{\triangle t})), \quad t \leftarrow t + \triangle t \tag{2.4.28}$$

Here, for consistency with the notation used elsewhere, the membrane time constant is called $\tau_M := \frac{1}{\alpha_M}$. Line 2.4.28 is meant to summarize that when the membrane potential crosses the threshold, a spike is fired, a new threshold is drawn from a transformation of the uniform random variable U and simulation is paused for a time interval of length $\triangle t$, the absolute refractory period.

Using the intrinsic plasticity mechanism outlined in section 2.3, this allows for the construction of an adaptive dynamical systems model that maintains a fixed distribution of activations! See section 3.3.3 for an example. A numerical simulation of such a neuron model can be found in section 4.2

Exponential Family Neuron Models

In chapter 2, a general statistical approach is presented to determine a neuron's activation function given it's observed or desired membrane potential and activation distribution. This chapter refines these results for a particularly convenient class of probability distributions, members of an *exponential family* (EF), for which adaptive neuron models can be derived very naturally. A particular example of a neuron model that maps normally distributed membrane potentials to log-normally distributed activations is discussed in detail.

3.1   Exponential Families

A so called *exponential family* (or *class*) *of distributions* can be defined in various ways, using different *base measures*, *non-canonical* forms or factorizing terms in different ways.[1] Here, densities are always assumed to be in their *canonical* form, simplifying the definition to a form consistent with, but more informal than Canu and Smola (2006)[2] as well as Nielsen and Garcia (2009):

---

[1] Charnes, Frome, and Yu 1976; Canu and Smola 2006; Geyer 1990; Nielsen and Garcia 2009; Dobson 2002; McCullagh and Nelder 1989.   [2] Canu and Smola (2006) measure w.r.t. to the distribution's base measure and can thus drop it from the definition of the density.

Definition 7 (Exponential Families). A family of probability distributions with probability density (or mass) functions $f(x|\theta)$ parameterized by the (bounded) vector θ, is called an *exponential family* (in *canonical form*), if and only if the densities can be expressed as:

$$f(x|\theta) = \exp(\langle \phi(x), \theta \rangle - g(\theta) + h(x))$$

Here $\langle \cdot, \cdot \rangle$ denotes the standard scalar product, $\phi(x) : \Omega \to \mathbb{R}^N$ is referred to as the *sufficient statistics* of the distribution, $\theta \in \mathbb{R}^N$ is the *natural parameter* of the distribution, $g(\theta) : \mathbb{R}^N \to \mathbb{R}$ is the *log-partition function* and $h(x) : \mathbb{R} \to \mathbb{R}$ is the distribution's *base measure*. The scalar $\eta := \langle \phi(x), \theta \rangle$ is referred to as the *linear predictor* of the distribution.

Remark. Here, only the standard scalar product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^N$ of the finite dimensional vector of sufficient statistics with the vector of natural parameters is considered, but in principle the concept of linearity of the predictor η can be extended to bilinear *kernels* $h(\cdot, \cdot)$ on *Reproducing Kernel Hilbert Spaces*[3] for even greater generality.

Remark. In order for the function f in definition 7 to be a proper probability density, it must be normalized to 1. Thus $g(\theta) = \log \left(\int_\Omega \exp(\langle \phi(x), \theta \rangle + h(x)) dx \right)$, which implies that g must be convex.[4]

The class of such distribution contains a surprising number of commonly used probability distributions,[5] as a partial list by Nielsen and Garcia (2009) shows:

> Gaussian or normal (generic, isotropic Gaussian, diagonal Gaussian, rectified Gaussian or Wald distributions, log-normal), Poisson, Bernoulli, binomial, multinomial (trinomial, Hardy-Weinberg distribution), Laplacian, Gamma (including the chi-squared), Beta, exponential, Wishart, Dirichlet, Rayleigh, probability simplex, negative binomial distribution, Weibull, Fisher-von Mises, Pareto distributions, skew logistic, hyperbolic secant, negative binomial, etc.[6]

---

[3] Canu and Smola 2006.   [4] Ibid.   [5] Brown 1986; McCullagh and Nelder 1989; Dobson 2002; Canu and Smola 2006.   [6] Nielsen and Garcia 2009.

SECTION 1. Exponential Families

A very convenient property of exponential family distributions is the fact, that maximum likelihood estimation of their parameters given a dataset is possible in a simple closed form given only the empirical mean of the sufficient statistics on the dataset. This relies on the fact that the log-likelihood function is concave[7] and the log-partition function g convex.[8]

Suppose a dataset $X := \{x_1, \ldots, x_M\}$ of i.d.d. values from an exponential family distribution with density f is given, and the ML estimator of the distribution's natural parameters θ is to be determined. The joint likelihood $\mathcal{L}$ of the parameters θ given the dataset X is then proportional to the expression in equation 3.1.4:

$$\mathcal{L}(\theta|X) = \mathbb{P}(X|\theta) = \prod_{i=1}^{M} f(x_i|\theta) \tag{3.1.1}$$

$$= \prod_{i=1}^{M} \exp\left(\langle \phi(x_i), \theta \rangle - g(\theta) + h(x_i)\right) \tag{3.1.2}$$

$$= \exp\left(\left\langle \sum_{i=1}^{M} \phi(x_i), \theta \right\rangle - M g(\theta) + \left(\sum_{i=1}^{M} h(x_i)\right)\right) \tag{3.1.3}$$

$$\propto \exp\left(\left\langle \sum_{i=1}^{M} \phi(x_i), \theta \right\rangle - M g(\theta)\right) \tag{3.1.4}$$

Using the standard ML estimation procedure of setting the derivative of the joint log-likelihood with respect to θ to 0 and solving for θ, we determine:

$$\frac{\mathrm{d}\log\mathcal{L}(\theta|X)}{\mathrm{d}\theta_j} = \frac{\mathrm{d}}{\mathrm{d}\theta_j}\left(\left\langle \sum_{i=1}^{M} \phi(x_i), \theta \right\rangle - M g(\theta)\right) \tag{3.1.5}$$

$$= \sum_{i=1}^{M} \phi_j(x_i) - M \frac{\mathrm{d}}{\mathrm{d}\theta_j} g(\theta) \overset{!}{=} 0 \tag{3.1.6}$$

$$\Longleftrightarrow \frac{\mathrm{d}}{\mathrm{d}\hat{\theta}_j} g(\hat{\theta}) = \frac{1}{M} \sum_{i=1}^{M} \phi_j(x_i) = \mathbb{E}_X[\phi_j(X)] \tag{3.1.7}$$

$$\Longleftrightarrow \nabla g(\hat{\theta}) = \mathbb{E}_X[\phi(X)] \tag{3.1.8}$$

Solving equation for $\hat{\theta}$ yields the maximum likelihood estimator of the parameters for the given dataset X. The fact that the ML estimate depends only on the

[7] Geyer 1990.   [8] Canu and Smola 2006.

(empirical) expectation $\mathbb{E}_X[\phi(X)]$ is the reason for calling ϕ the sufficient statistics of the distribution: the expected value of the sufficient statistics fully characterizes the corresponding probability distribution.

3.2 Adaptive Mapping between Exponential Families

3.2.1 Online Parameter Estimates with Conjugate Priors

If not a whole dataset X is observed simultaneously but instead continuously sampled one value at a time (and $\mathbb{E}_X[\phi(X)]$ can thus not be calculated at once), *Bayesian sequential updates* can be used to derive a running ML estimate of the parameters given the data points seen so far. To this end, we use the definition of the *conjugate prior* by Hogg, McKean, and Craig (2005):

Definition 8 (Conjugate Priors)**.** A class of prior pdfs for the family of distributions with pdfs $f(x|\theta), \theta \in \Omega$ is said to define a *conjugate family of distributions* if the posterior pdf of the parameter is in the same family of distributions as the prior.

In other words, the conjugate prior of an exponential family with a PDF as in definition 7 takes the same form as the joint likelihood for a dataset consisting of M_0 *virtual* observations $Y := \{y_1, \ldots, y_{M_0}\}$ with $\mu_0 := \frac{1}{M_0} \sum_{i=1}^{M_0} \phi(y_i)$ (cf. Nielsen and Garcia (2009)):

$$p(\theta|Y) = p(\theta|\mu_0) \propto \exp\left(\left\langle \sum_{i=1}^{M_0} \phi(y_i), \theta \right\rangle - M_0 g(\theta) \right) \tag{3.2.1}$$

$$= \exp(\langle M_0 \mu_0, \theta \rangle - M_0 g(\theta)) \tag{3.2.2}$$

Using this prior, that corresponds to M_0 virtual observations of the sufficient statistics with the sample mean μ_0, the posterior probability of θ after seeing a single data point x is proportional to the expression in equation 3.2.6:

$$f(\theta|x, \mu_0) \propto f(x|\theta)p(\theta|\mu_0) \tag{3.2.3}$$

$$\propto \exp\left(\langle \phi(x), \theta \rangle - g(\theta)\right) \exp\left(\langle M_0\mu_0, \theta \rangle - M_0 g(\theta)\right) \tag{3.2.4}$$

$$= \exp\left(\langle \phi(x) + M_0\mu_0, \theta \rangle - (1 + M_0)g(\theta)\right) \tag{3.2.5}$$

$$= \exp\left(\langle M_1\mu_1, \theta \rangle - M_1 g(\theta)\right) \tag{3.2.6}$$

$$\text{where}: \quad M_1 := M_0 + 1; \quad y_{M_1} := x; \quad \mu_1 := \frac{1}{M_1}\sum_{i=1}^{M_1}\phi(y_i) \tag{3.2.7}$$

The posterior has the same form as the expression in 3.2.6 and the prior in 3.2.2, which is thus shown to be a conjugate prior of the distribution.

Since prior and posterior come from the same family of distributions, only with different parameters, it is possible to use the resulting posterior as a prior when adding a new data point, and to continue doing so inductively. This allows us to define the posterior after observing the data points $X_n := \{x_i \in X | 1 \leq i \leq n\}$ recursively for $n \geq 1$:

$$f(\theta|X_n, \mu_n) \propto \exp\left(\langle M_n\mu_n, \theta \rangle - M_n g(\theta)\right) \tag{3.2.8}$$

$$\text{where}: \quad M_n := M_{n-1} + 1; \quad \mu_n := \frac{1}{M_n}\phi(x) + (1 - \frac{1}{M_n})\mu_{n-1} \tag{3.2.9}$$

For $M_0 = 0$, the posterior $f(\theta|X_n, \mu_n)$ is identical to the joint likelihood of X_n as defined in equation 3.1.4 and μ_n corresponds to the sample mean of the sufficient statistics on the sample X_n. As to be expected, estimates μ_n converge to $\mathbb{E}[\phi(X)]$ as the number of samples n goes to infinity due to the *law of large numbers,*[9] since with an increasing n, the influence of the prior (or any individual data point) decays with $\mathcal{O}(n^{-1})$.

Since the ML estimator for the parameter vector θ is implicitly defined by equation 3.1.8 as the vector $\hat{\theta}$ for which $\nabla g(\hat{\theta}) = \mathbb{E}_X[\phi(X)]$ and $\mu_n = \mathbb{E}_{X_n}[\phi(X)] \approx \mathbb{E}_X[\phi(X)]$, a running maximum likelihood estimate $\hat{\theta}_n$ can be defined as the solution $\hat{\theta}_n$ of $\nabla g(\hat{\theta}_n) = \mu_n$. If the (component-wise) inverse of the gradient of g exists and is called γ, then this can be expressed as $\hat{\theta}_n = \gamma(\mu_n)$.

---

[9] Hogg, McKean, and Craig 2005.

If we want the estimation to not converge to some fixed value, but instead want all previously seen data (regardless of how much that might have been) to serve as a prior with a fixed weight, then the diminishing weight $\frac{1}{M_n}$ in line 3.2.9 can be replaced by the constant weight α, which yields $\tilde{\mu}_n := \alpha\phi(x) + (1-\alpha)\mu_{n-1}$. If this exponentially weighted running estimate of the sufficient statistics is used to calculate a "running ML estimate" $\hat{\theta}(t)$, then this corresponds, in continuous time, to the following differential equation with time-constant $\frac{1}{\alpha}$:

$$\frac{1}{\alpha}\frac{\mathrm{d}\tilde{\mu}(t)}{\mathrm{d}t} = -\tilde{\mu}(t) + \phi(x(t)) \tag{3.2.10}$$

$$\hat{\theta}(t) = \gamma(\tilde{\mu}(t)) \tag{3.2.11}$$

3.2.2 Sensitivity to Parameter Estimates

We now assume that a neuron, as defined in chapter 2, has a membrane potential that is distributed according to some continuous member distribution of an exponential family and a continuous desired distribution of activation. Neither of this appears to be a strong concession. Since the true parameters underlying the membrane potential distribution are unknown a priori and the running ML estimate defined above only approximates the true maximum likelihood estimators, a first interesting thing to analyze is how closely the neuron's distribution of activation matches the desired distribution.

Since both membrane potential distribution and desired activation distribution are assumed to be continuous, the corresponding cumulative distribution functions $F_1(V|\theta_1)$ and $F_2(\lambda|\theta_2)$ are strictly monotonically increasing, thus the neuron's activation function f as well as its inverse g are well defined and monotonically increasing:

$$f(v;\theta_1,\theta_2) = (F_2^{-1} \circ F_1)(v;\theta_1,\theta_2) \tag{3.2.12}$$

$$g(\lambda;\theta_1,\theta_2) = f^{-1}(\lambda;\theta_1,\theta_2) = (F_1^{-1} \circ F_2)(\lambda;\theta_1,\theta_2). \tag{3.2.13}$$

Let the neuron's true distribution of membrane potentials V be of the exponen-

tial family form:

$$f_1(v|\theta_1) = \exp(\langle \phi_1(v), \theta_1 \rangle - g_1(\theta_1) + h_1(v)) \tag{3.2.14}$$

If the maximum likelihood estimates for the parameters are given by $\hat{\theta}_1$, then the neuron's actually resulting distribution of activation can be determined:

$$\hat{f}_2(\lambda|\hat{\theta}_1, \theta_2) = f_1(g(\lambda; \hat{\theta}_1, \theta_2)|\theta_1) \left| \frac{\mathrm{d}g(\lambda; \hat{\theta}_1, \theta_2)}{\mathrm{d}\lambda} \right| \tag{3.2.15}$$

$$= f_1(g(\lambda; \hat{\theta}_1, \theta_2)|\theta_1) \frac{\mathrm{d}F_1^{-1}(F_2(\lambda|\theta_2)|\hat{\theta}_1)}{\mathrm{d}\lambda} \tag{3.2.16}$$

$$= \frac{f_1(g(\lambda; \hat{\theta}_1, \theta_2)|\theta_1)}{f_1\left(F_1^{-1}(F_2(\lambda|\theta_2)|\hat{\theta}_1)|\hat{\theta}_1\right)} \frac{\mathrm{d}F_2(\lambda|\theta_2)}{\mathrm{d}\lambda} \tag{3.2.17}$$

$$= \frac{f_1(g(\lambda; \hat{\theta}_1, \theta_2)|\theta_1)}{f_1(g(\lambda; \hat{\theta}_1, \theta_2)|\hat{\theta}_1)} f_2(\lambda|\theta_2) \tag{3.2.18}$$

The expression in equation 3.2.18 relates the actual output distribution $\hat{f}_2(\lambda|\hat{\theta}_1, \theta_2)$ to the desired output distribution $f_2(\lambda|\theta_2)$. It can be used to calculate the *relative entropy* (also referred to as *Kullback-Leibler Divergence* or *Distance*) $D(\hat{f}_2||f_2)$, "a measure of the distance between two distributions"[10] or rather of the "inefficiency of assuming that the distribution is $[f_2]$ when the true distribution is $[\hat{f}_2]$":[11]

$$D(\hat{f}_2||f_2) = \int_0^\infty \hat{f}_2(\lambda|\hat{\theta}_1, \theta_2) \log \frac{\hat{f}_2(\lambda|\hat{\theta}_1, \theta_2)}{f_2(\lambda|\theta_2)} \mathrm{d}\lambda \tag{3.2.19}$$

$$= \int_0^\infty \hat{f}_2(\lambda|\hat{\theta}_1, \theta_2) \log \frac{f_1(g(\lambda; \hat{\theta}_1, \theta_2)|\theta_1)}{f_1(g(\lambda; \hat{\theta}_1, \theta_2)|\hat{\theta}_1)} \mathrm{d}\lambda \tag{3.2.20}$$

$$= \int_0^\infty \hat{f}_2(\lambda|\hat{\theta}_1, \theta_2) \left(\langle \phi_1(g(\lambda; \hat{\theta}_1, \theta_2)), \theta_1 - \hat{\theta}_1 \rangle - g_1(\theta_1) + g_1(\hat{\theta}_1) \right) \mathrm{d}\lambda \tag{3.2.21}$$

---

[10] Cover and Thomas (2005). The KLD is however not a metric, as they are careful to point out: *"relative entropy is always nonnegative and is zero if and only if $[\hat{f}_2 = f_2]$. However, it is not a true distance between distributions since it is not symmetric and does not satisfy the triangle inequality. Nonetheless, it is often useful to think of relative entropy as a 'distance' between distributions."*
[11] Ibid.

Substituting $v = g(\lambda; \hat{\theta}_1, \theta_2); \frac{\mathrm{d}v}{\mathrm{d}\lambda} = \frac{f_2(\lambda|\theta_2)}{f_1(v|\hat{\theta}_1)}$ back into equation 3.2.21 yields:

$$D(\hat{f}_2||f_2) = \int_{-\infty}^{\infty} f_1(v|\theta_1)\langle\phi_1(v), \theta_1 - \hat{\theta}_1\rangle\mathrm{d}v - g_1(\theta_1) + g_1(\hat{\theta}_1) \tag{3.2.22}$$

$$= \langle\mathbb{E}[\phi_1(v)], \theta_1 - \hat{\theta}_1\rangle - g_1(\theta_1) + g_1(\hat{\theta}_1) \tag{3.2.23}$$

$$= |\langle\nabla g_1(\hat{\theta}_1), \theta_1 - \hat{\theta}_1\rangle - (g_1(\theta_1) - g_1(\hat{\theta}_1))| \tag{3.2.24}$$

$$\leq ||\nabla g_1(\hat{\theta}_1)|| \cdot ||\theta_1 - \hat{\theta}_1|| + K||\theta_1 - \hat{\theta}_1|| \tag{3.2.25}$$

$$\leq 2K||\theta_1 - \hat{\theta}_1|| \tag{3.2.26}$$

This follows from the fact that g is Lipschitz continuous on a compact set including $\hat{\theta}_1$ and θ_1.[12] Thus also $D(\hat{f}_2||f_2)$ is Lipschitz continuous, which implies that it converges at least as fast to 0 as the difference $||\hat{\theta}_1 - \theta_1||$. In other terms, a good approximation of the parameters θ_1 of the input distribution results in a low KL divergence, and a parameter estimation technique that converges in $\mathcal{O}(n^{-1})$ implies that the KLD also converges with $\mathcal{O}(n^{-1})$.

This shows that as long as the parameter estimates are close to the real parameters of the membrane potential distribution, the distribution of the neuron's activation is also close to the desired distribution.

3.3 Normal-Log-Normal Neuron Model

In this section, a plastic neuron model is finally derived for a specific membrane potential and an activation distribution. It serves as a simple example of the methods outline above, rather than a suggestion of a neuron model, but it exhibits a number of interesting properties worth investigating.

3.3.1 Normally Distributed Inputs

If the membrane potential $V(t)$ of a neuron is modeled using an Ornstein-Uhlenbeck process[13] (a *noisy relaxation process*) resulting from the "relaxing" effect of a leaky

---

[12] Roberts and Varberg 1974. [13] Shimokawa and Shinomoto (2009) provide as parameters for such an OU process to reproduce biological findings the means $\mu_\lambda = 50$Hz and $\mu_\kappa = 1.0$, standard deviations $\sigma_\lambda = 25$Hz and $\sigma_\kappa = 1.0$ and the joint timescale $\tau = 600$ms.

membrane and the "noisy" bombardment with ePSPs, its distribution at a fixed point t in time is Gaussian with mean $\mu(t)$ and variance $\sigma^2(t)$.[14]

This choice of membrane potential distribution appears sensible from both a statistical as well as an implementation perspective: Under the simplifying assumption that the membrane potential resembles a linear combination of a large number of independent, identically distributed post-synaptic potentials, the central limit theorem states that the membrane potential should roughly resemble a normal distribution. As the maximum entropy distribution given a fixed mean and variance and real support, the normal distribution also makes the weakest assumptions about the distribution of membrane potentials given its first two moments, mean and variance.[15]

For a mean μ_1 and a standard deviation σ_1, the PDF f_1 and CDF F_1 of the neuron's membrane potential are given as follows:[16]

$$f_1(v|\mu_1,\sigma_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\left(\frac{v-\mu_1}{\sqrt{2}\sigma_1}\right)^2\right) \tag{3.3.1}$$

$$F_1(v|\mu_1,\sigma_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \int_{-\infty}^{v} \exp\left(-\left(\frac{x-\mu_1}{\sqrt{2}\sigma_1}\right)^2\right) \mathrm{d}x \tag{3.3.2}$$

$$= \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{v-\mu_1}{\sqrt{2}\sigma_1}\right) \tag{3.3.3}$$

$$\text{where } \mathrm{erf}(v) := \frac{2}{\sqrt{\pi}} \int_0^v \exp(-x^2)\mathrm{d}x \tag{3.3.4}$$

When re-parameterizing with the new parameters $\theta_1 \in \mathbb{R}^+$ and $\theta_2 \in \mathbb{R}^- \setminus \{0\}$, the PDF takes its canonical exponential family form:

---

[14] Rudolph and Destexhe 2005.   [15] Cover and Thomas 2005.   [16] Hogg, McKean, and Craig 2005.

$$\text{set } \begin{pmatrix} \mu_1 \\ \sigma_1^2 \end{pmatrix} \rightarrow \begin{pmatrix} -\frac{\theta_1}{2\theta_2} \\ -\frac{1}{2\theta_2} \end{pmatrix} \tag{3.3.5}$$

$$f_1(v|\theta_1, \theta_2) = \sqrt{-\frac{\theta_2}{\pi}} \exp\left((v + \frac{\theta_1}{2\theta_2})^2 \theta_2 \right) \tag{3.3.6}$$

$$= \exp\left(v^2\theta_2 + v\theta_1 + \frac{\theta_1^2}{4\theta_2} + \frac{1}{2}\log(-2\theta_2) - \frac{1}{2}\log(2\pi) \right) \tag{3.3.7}$$

$$= \exp\left(\langle \phi(v), \theta \rangle - g(\theta) + h(v) \right) \tag{3.3.8}$$

with $\theta := \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$, $\phi(v) := \begin{pmatrix} v \\ v^2 \end{pmatrix}$, $g(\theta) := -\frac{\theta_1^2}{4\theta_2} - \frac{1}{2}\log(-2\theta_2)$ and $h(v) := -\frac{1}{2}\log(2\pi)$.

The gradient of the log-partition function g as well as its component-wise inverse, which is needed for maximum likelihood estimation, can be calculated easily:

$$\nabla g(\theta) = \begin{pmatrix} \frac{dg(\theta)}{d\theta_1} \\ \frac{dg(\theta)}{d\theta_2} \end{pmatrix} = \begin{pmatrix} -\frac{\theta_1}{2\theta_2} \\ \left(\frac{\theta_1}{2\theta_2}\right)^2 - \frac{1}{2\theta_2} \end{pmatrix} \tag{3.3.9}$$

$$\gamma(\Theta) := (\nabla g)^{-1}\left(\begin{pmatrix} \Theta_1 \\ \Theta_2 \end{pmatrix} \right) = \begin{pmatrix} \frac{\Theta_1}{\Theta_2 - \Theta_1^2} \\ \frac{-1}{2(\Theta_2 - \Theta_1^2)} \end{pmatrix} \tag{3.3.10}$$

The continuous time running ML estimate of the parameters of the membrane potential distribution from equations 3.2.10 and 3.2.11 can thus be written as a system of two differential equations:

$$\frac{1}{\alpha}\frac{d\tilde{\mu}_1(t)}{dt} = -\tilde{\mu}_1(t) + \phi_1(v(t)) = -\tilde{\mu}_1(t) + v(t) \tag{3.3.11}$$

$$\frac{1}{\alpha}\frac{d\tilde{\mu}_2(t)}{dt} = -\tilde{\mu}_2(t) + \phi_2(v(t)) = -\tilde{\mu}_2(t) + v(t)^2 \tag{3.3.12}$$

$$\hat{\theta}_1(t) = \gamma_1(\tilde{\mu}(t)) = \frac{\tilde{\mu}_1(t)}{\tilde{\mu}_2(t) - \tilde{\mu}_1(t)^2} \tag{3.3.13}$$

$$\hat{\theta}_2(t) = \gamma_2(\tilde{\mu}(t)) = \frac{-1}{2(\tilde{\mu}_2(t) - \tilde{\mu}_1(t)^2)} \tag{3.3.14}$$

Reversing the parameter transformation results in estimators $\hat{\mu}$ and $\hat{\sigma}$ for the parameters of the distribution[17]:

$$\hat{\mu}(t) = -\frac{\hat{\theta}_1}{2\hat{\theta}_2} = \tilde{\mu}_1(t) \tag{3.3.15}$$

$$\hat{\sigma}^2(t) = -\frac{1}{2\hat{\theta}_2} = \tilde{\mu}_2 - \tilde{\mu}_1(t)^2 \tag{3.3.16}$$

This result is hardly surprising, as it resembles just the running estimate of mean and variance, but it validates the approach chosen so far.

While the first moment corresponds to a term just like the neuron's leaky membrane potential, only with a slower time-constant, and could thus easily be implemented in a biological system, an estimate of the second moment requires a dependency on the square of the membrane potential and thus some other underlying mechanism. Triesch (2007) suggests that "[the] required estimate of the second moment of the neuron's firing rate may be implemented by an agent A that binds two Ca^{2+} ions: $A + 2Ca^{2+} \rightarrow A'$. The concentration of A' could then approximate the square of the current firing rate".

3.3.2 Log-Normally Distributed Firing Rates

Based on a review of a large number of biological studies, Buzsáki and Mizuseki (2014) suggest that log-normal distributions of firing rates can be found on different scales throughout the brain. They attribute this to the fact that "biological mechanisms possess emergent and collective properties as a result of many interactive processes, and multiplication of a large number of variables, each of which is positive, gives rise to lognormal distributions".[18] This effect – an intuitive multiplicative analog of the additive *central limit theorem* – can be observed in many other sufficiently complex systems, as well.[19]

Following their suggestion, the neuron's desired output distribution is here defined to be a log-normal distribution (i.e. a distribution of a random variable λ,

---

[17] The notation here is a bit unfortunate, because $\tilde{\mu}$ refers to the running estimate of the expected value of the sufficient statistics of the membrane potential distribution, whereas $\hat{\mu}$ refers to an estimator of its mean. [18] Buzsáki and Mizuseki 2014. [19] Limpert, Stahel, and Abbt 2001.

the logarithm of which is normally distributed) with parameters μ_2 and σ_2 and the cumulative distribution function

$$F_2(\lambda|\mu_2, \sigma_2) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\log(\lambda) - \mu_2}{\sqrt{2}\sigma_2}\right) \qquad (3.3.17)$$

Solving for y gives the quantile function G_2 of the log-normal distribution[20]:

$$G_2^{-1}(x) = \exp\left(\mu_2 + \sigma_2\sqrt{2}\mathrm{erf}^{-1}(2x - 1)\right) \qquad (3.3.18)$$

The same model is briefly discussed and analyzed by Roxin et al. (2011) to model the population firing rate of neurons with exponential activation functions and subsequently compared to different models. Their conclusion is, however, that besides strong evidence to the contrary at least "for homogeneous neuronal populations, a lognormal distribution of firing rates is, in general, not to be expected". Thus finding the right distribution of activation to use remains an interesting question; here the log-normal distribution is used to provide a proof of concept because of the previously cited evidence and mathematical convenience, but the results could easily transfer to other similar distribution.

3.3.3 Adaptive Normal-Log-Normal Neuron Model

Using Theorem 5, the random variable $\lambda := f(V; m_1, s_1, \mu_2, \sigma_2)$ has the desired log-normal distribution with cumulative distribution function F_2 if and only if $f(v; m_1, s_1, \mu_2, \sigma_2) = G_2(F_1(v; m_1, s_1); \mu_2, \sigma_2)$, where $m_1 = \mu_1$, $s_1 = \sigma_1$, μ_2 and σ_2 are the parameters of F_1 and F_2, respectively. It exhibits intrinsic plasticity in the sense of section 2.3, if the parameters m_1 and s_1 are set to the running estimates $\hat{\mu}_1$ and $\hat{\sigma}_1$ as defined in equations 3.3.15 and 3.3.16.

The dynamic membrane potential and running parameter estimates are as

---

[20] Because F_2 is continuous and strictly increasing, it is invertible with a (generalized) inverse G_2

defined above and repeated here for convenience:

$$\tau_M \frac{\mathrm{d}}{\mathrm{d}t} V(t) = -V(t) + I(t)$$

$$\frac{1}{\alpha} \frac{\mathrm{d}\tilde{\mu}_1(t)}{\mathrm{d}t} = -\tilde{\mu}_1(t) + V(t)$$

$$\frac{1}{\alpha} \frac{\mathrm{d}\tilde{\mu}_2(t)}{\mathrm{d}t} = -\tilde{\mu}_2(t) + V(t)^2$$

The activation function and its inverse can be derived as follows:

$$f(V; m_1, s_1, \mu_2, \sigma_2) = \exp\left(\frac{\sigma_2}{s_1}(V(t) - m_1) + \mu_2 \right) \tag{3.3.19}$$

$$f^{-1}(\lambda; m_1, s_1, \mu_2, \sigma_2) = (\log(\lambda) - \mu_2)\frac{s_1}{\sigma_2} + m_1 \tag{3.3.20}$$

This also determines the neuron's activation $\lambda(t)$ given $V(t)$:

$$\lambda(t) = f(V(t); \tilde{\mu}_1, \sqrt{\tilde{\mu}_2 - \tilde{\mu}_1(t)^2}, \mu_2, \sigma_2) \tag{3.3.21}$$

$$= \exp\left(\frac{\sigma_2}{\sqrt{\tilde{\mu}_2(t) - \tilde{\mu}_1(t)^2}}(V(t) - \tilde{\mu}_1) + \mu_2 \right) \tag{3.3.22}$$

Equivalently, this defines the random threshold of a dynamical systems model representing the same distribution of activation:

$$\theta = f^{-1}(\frac{-\log(U)}{\Delta t}; \tilde{\mu}_1, \sqrt{\tilde{\mu}_2 - \tilde{\mu}_1(t)^2}, \mu_2, \sigma_2) \tag{3.3.23}$$

$$= (\log(\frac{-\log(U)}{\Delta t}) - \mu_2)\frac{\sqrt{\tilde{\mu}_2 - \tilde{\mu}_1(t)^2}}{\sigma_2} + \tilde{\mu}_1 \tag{3.3.24}$$

This can be used to implement a homeostatic integrate-and-fire neuron model with random thresholds, the distribution of which is modulated via intrinsic plasticity, that achieves a fixed firing rate distribution.

In principle, once the neuron's activation distribution and the distribution of the number and weight of synaptic connections is known, a neuron's membrane potential distribution could be inferred, instead of being presupposed as in the previous sections. Unfortunately, there is no closed form expression for the dis-

tribution of a linear combination of i.i.d. log-normal random variables. However, two different approximations are possible: As argued before, if the number of synaptic connections approaches infinity, the membrane potential distribution, a linear combination of i.i.d. random variables, must converge to a normal distribution due to the central limit theorem (assuming the vector of synaptic weights is bounded). For a smaller number of synapses, the probability distribution resulting from a linear combination of i.i.d. activities can be approximated using e.g. an Edgeworth series, because the first 4 cumulants of the log-normal distribution are finite.[21] Here, we assume the numbers of synaptic connections and external noise to be sufficiently large to make the membrane potential normally distributed, without this being inconsistent with the assumption of log-normal distributions of activation.

3.3.4 Incorporating Refractoriness and Spike-Responses

As motivated in section 2.4.2, the spiking probability of the neuron can be modulated by auto-history effects as expressed in the equation $\tilde{\lambda}(t) = f(V(t) + \sum_i R(t - \tau_i))$, where R is a function of the time since a previous spike i. Using an exponential activation function, this can be motivated more rigorously: Suppose that each spike the neuron fires triggers a cascade of biological processes that could each interrupt the generation of subsequent spikes. If, after a spike at time τ_i, the triggered processes have the time-varying probability $\tilde{R}(t - \tau_i)$ of preventing the firing of a spike at time t and the effects of different spikes accumulate and are assumed to be independent of each other and the neuron's current activation, then $\tilde{\lambda}(t)$, the instantaneous spiking probability with refractoriness, is given by the joint probability $\lambda(t) \prod_i \tilde{R}(t - \tau_i)$, where $\lambda(t)$ denotes the activation in the absence of refractory effects. This can be simplified due to the fact that $\lambda(t) = f(V(t)) = \exp(c_1 V(t) + c_2)$ for some constants c_1 and c_2, because using the logarithm, the product can be absorbed into the exponential as a sum, resulting in $\tilde{\lambda}(t) = \exp(c_1 V(t) + c_2 + \log(\prod_i \tilde{R}(t - \tau_i))) = f(V(t) + \frac{1}{c_1} \sum_i \log \tilde{R}(t - \tau_i))$. We now call $R := \frac{1}{c_1} \log \tilde{R}$ the *refractory function* and get a model similar to the *spike-response model*.[22]

---

[21] McCullagh and Nelder 1989.   [22] Gerstner 1995.

Using the neuron's own spiking output $X(t) = \sum_i \delta(t - \tau_i)$, the sum $\sum_i R(t - \tau_i)$) can be rewritten as the convolution $V_{\mathrm{ref}}(t) := \int_0^t R(t - \tau)X(\tau)\mathrm{d}\tau$, which has the same form as the post-synaptic potential induced by a single synapse, only using the refractory function R instead of a synaptic filter η. This allows us to treat auto-history effects such as refractoriness or more general spike responses as an additional internal input to the virtual membrane potential of the neuron. One side effect of this is, that the changes of the spike distribution resulting from inclusion of auto-history effects are compensated for by the homeostatic adaptation process within the neuron and should thus be alleviated.

When using an integrate-and-fire type neuron model with random thresholds, the negative auto-history can equivalently be subtracted from the random threshold, rather than used as an additive component of the virtual membrane potential $\tilde{V}$[23]. This is sometimes used to account for refractoriness and *spike rate adaptation*.[24] The refractoriness of the neuron is then modeled by a time-varying component of the threshold that can, for sufficiently simple refractory functions, be expressed as a one or two-dimensional system of differential equations.

3.3.5 Relations to Generalized Linear Models

Using exponential family distributions to model the neuron's activation puts this thesis into the context of the widely used *generalized linear models*,[25] so a few interesting connections are listed here.

The activation function in equation 3.3.22 can be viewed in a quite different way, as well. If we interpret the neuron as a system that receives a linear combination of pre-synaptic inputs and transforms it in a non-linear way, resulting in a spiking probability from which spikes are randomly drawn, then the neuron corresponds to a *generalized linear model*: If $f(V) = \exp(\eta(t)) = \lambda(t) = \mathbb{E}[X(t)]$ is the expected value of the instantaneous random spiking output $X(t)$ at time t and $\eta(t) = \frac{\sigma_2}{s_1}V(t) + (\mu_2 - \frac{\sigma_2 m_1}{s_1}) = \sum_j X_j \beta_j$ is a linear function of the neu-

---

[23] Note however, that doing so means that refractoriness is added at the last stage of the model and thus cannot be compensated for by intrinsic plasticity. This in turn means, that the neuron's distribution of spikes can be systematically different from the distribution resulting from the desired distribution of activation in the absence of refractory effects!   [24] Dayan and Abbott 2001, ch. 5.
[25] McCullagh and Nelder 1989.

ron's synaptic inputs X_j, then $\eta(t) = \log(\mathbb{E}[X(t)])$ and the neuron resembles a generalized linear model with *log-link* function.[26] If we further assume that $X(t)$ at any point t is the expected number of spikes within a short time interval around t and follows a Bernoulli, binomial or Poisson distribution, then this becomes a GLM with *conjugate* log-link function. Finding the most likely parameters β for which the random observations are distributed around the mean $\mathbb{E}[X(t)] = \exp(\eta(t)) = \exp(\sum_j X_j \beta_j)$ then reduces to the problem of approximately solving the linear system $\sum_j X_j \beta_j = \log(\mathbb{E}[X])$ with respect to an appropriately chosen error measure.[27] Such log-link models are commonly used for *Poisson regression analysis*, where multiplicative effects of several processes are to be modeled,[28] and can be applied to a wide range of systems.[29]

If the model is expressed in discrete time[30], as in equations 2.1.6 to 2.1.9, the spiking variable X becomes a Bernoulli random variable, and the probability $\rho = \mathbb{E}[X]$ to spike within one time-interval is related to the activation λ via $\lambda = -\log(1-\rho)$. Concatenated with the log-link function, this results in the model $\eta(t) = \log(\lambda(t)) = \log(-\log(1-\rho)) = \log(-\log(\mathbb{E}[X]))$, i.e. a GLM with the so called *complementary log-log function* link that is often used, as here, to model binary data.[31] Curiously, this relates the model presented here in discrete time as well as in continuous time to another commonly used neural activation function, the *logistic function*, by the fact that "[for] small values of π, the complementary log-log function is close to the logistic, both being close to $\log(\pi)$".[32]

Within the GLM framework, the task of modeling a spiking point-process can be related to tools such as so called *survival analysis*, which is commonly used to analyze point processes like a patient's or machine's probability of "survival" over a period of time (hence the name), with the neuron's activation $\lambda(t)$ over time representing the so called *hazard function*.[33]

---

[26] McCullagh and Nelder 1989, ch. 2. [27] Charnes, Frome, and Yu 1976. [28] McCullagh and Nelder 1989, ch. 6. [29] Limpert, Stahel, and Abbt 2001. [30] Assume without loss of generality that the timescale is chosen such that $\triangle t = 1$. [31] McCullagh and Nelder 1989, ch. 4. [32] Ibid., ch. 4. [33] Ibid., ch. 13.

3.3.6 Implications for Neural Computations

Exponential activation functions have equivalent approximating power to e.g. the commonly used standard sigmoid and radial basis functions[34] and are thus reasonable candidates from a computational point of view. Furthermore, they fall into the category of activation functions for which the joint likelihood is convex[35] and maximum likelihood estimation (and thus the intrinsic plasticity mechanism discussed here) converges to the distribution's true parameters.

Determining the activation function of a neuron from its membrane potential and activation distribution can be very helpful in analyzing the "calculus" used by neurons for computation. The assumption of fixed threshold units, for example, on which the idea of a logical calculus of neural computation[36] is based, has testable implications for the shape of the resulting distribution of firing rates in a network. Conversely, measuring the distributions in a network and inferring the *effective* activation function (without having to disrupt in vivo processing, clamp membrane potentials or otherwise distort the neuron's environment) allows building a calculus based on the inferred activation function. Similarly to how a logical calculus of neuronal activity can be built atop the assumption of fixed threshold neurons, a multiplicative "calculus" can be derived from the assumption of exponential activation functions:

Given that sensory neurons encode their sensory modality logarithmically, rather than linearly[37], the activation of a neuron receiving multiple sensory inputs $\gamma_i = \log(\beta_i)$ is given by $\lambda = \exp(\sum_i \gamma_i \omega_i) = \prod_i \exp(\log(\beta_i)\omega_i) = \prod_i \beta_i^{\omega_i}$, which corresponds to a multiplication of the measured sensory modalities β_i. This method of calculating products is referred to as the *log-exp transform* by Koch and Poggio (1992) and could provide a simple mechanism for *coincidence detection*.

The hypothesis that neurons can multiply inputs is supported by a host of biological evidence[38] and could be attributed to very different potential mechanisms.[39] Using passive observations of an in vivo system however, as might be possible

---

[34] DasGupta and Schnitger 1993.   [35] Paninski 2004.   [36] McCulloch and Pitts 1943.   [37] This is the well known Weber-Fechner law, that states in German: "Die Grösse der Empfindung (γ) steht im Verhältnisse nicht zu der absoluten Grösse des Reizes (β), sondern zu dem Logarithmus der Grösse des Reizes" (Fechner 1860, p. 13).   [38] Koch and Poggio 1992; Koch and Segev 2000; Koch 1997; Gabbiani et al. 2002.   [39] Koch and Poggio 1992.

using the methods discussed here, to capture the effective activation function in vivo might help answer these questions.

3.3.7   Implications for Synaptic Plasticity

So far, Gaussian membrane potential distributions were discussed. This section addresses the question, what happens when the membrane potential of the same neuron model with exponential activation function takes on a different distribution and is subject to simple Hebbian synaptic plasticity and synaptic scaling.

   The approach and conclusions here are similar to those of Triesch (2007), who used an information theoretical approach to tune parameters of a fixed activation function to enforce an exponential distribution of outputs. Thus naturally, the conclusions would be very similar if an exponential, instead of a log-normal, activation distribution was chosen here. But despite the different resulting distribution of activation, some general conclusions about the interaction of the homeostatic mechanisms of intrinsic plasticity and synaptic plasticity can be equivalently derived from both approaches.

   Dropping the homeostatic plasticity equations, the activation function of the neuron given by equation 3.3.22 and the linear combination of synaptic inputs can be summarized in the following equations for discrete time:

$$V_t = \langle x_t, \omega \rangle \tag{3.3.25}$$

$$\lambda_t = \exp\left(c_1 \cdot V_t + c_2\right) \tag{3.3.26}$$

where x_t is the vector of (filtered) pre-synaptic activities[40], ω is the vector of synaptic weights through which the inputs are projected, V_t is the neuron's membrane potential, c_1 and c_2 are constants and λ_t is the neuron's activation, each at discrete time-step t.

   As shown in the following, from this very simple model and a few basic assumptions, a Hebbian learning rule allowing single neurons to extract independent components from mixed signals can be derived as the solution to the interesting

---

[40] Due to linearity of the convolution operator, filtering the pre-synaptic inputs before calculating the linear combination is the same as filtering the linear combination as written in equation 3.3.22

optimization problem of *blind source separation*.[41]

In the following, the vector ω is assumed to have fixed length $\|\omega\|_2 = 1$, a constraint that can be attributed to fast acting mechanisms of *synaptic scaling*.[42]

Independent component analysis (ICA) can be loosely defined as the process of de-mixing non-Gaussian signals by finding a basis of the input space for which the Fourier-coefficients of the input vectors are "maximally non-Gaussian". It is built on the observation that "a sum of two independent random variables usually has a distribution that is closer to gaussian than any of the two original random variables",[43] and thus "[m]aximizing the nongaussianity"[44] of the recovered *components* should result in the original unmixed, *independent components*.

A major benefit of this approach when compared to *principal component analysis* (PCA) is the fact that e.g. even a mixture of uncorrelated signals with identity covariance matrix can be de-mixed using ICA as shown in the following, whereas PCA can not. This qualitative difference comes from the fact that PCA, operating on the covariance matrix alone, can utilize at most the second centralized moment, the covariance, of the mixtures, whereas ICA, a non-linear method, depends on higher order moments, as well. Such dependencies on higher order moments – excess kurtosis (defined via fourth and second centralized moment) in particular – are often used as measures of a distribution's non-gaussianity for ICA,[45] with higher values of high order moments intuitively implying "heavier tails than Gaussian".

The exponential activation function of our model neurons acts as such a non-linear transformation of the neuron's membrane potential, that introduces a monotonic dependency of the neuron's mean activation on higher order moments of its membrane potential, i.e. its expected activation is the higher, the more "non-gaussian" the membrane potential distribution becomes for a fixed mean and variance (see the appendix section B.1 for a derivation that shows this dependency on higher order moments). Thus choosing synaptic weights to maximize the neuron's expected activation under the constraint $\|\omega\| = 1$ simultaneously maximizes the non-gaussianity of the resulting membrane potential, which in turn implies that an independent component is extracted from the mixture. This is analytically shown in detail for mixtures of i.i.d. Laplace random variables in appendix section

---

[41] Hyvärinen and Oja 2000.   [42] Turrigiano and Nelson 2000.   [43] Hyvärinen and Oja 2000.
[44] Ibid.   [45] Ibid.

B.2, but generalizes to other distributions as well.

In order to maximize the neuron's expected output with respect to the vector of synaptic weights ω, a simple *stochastic gradient descent* mechanism can be derived. The objective is to to find, given an empirical distribution of V:

$$\hat{\omega} := \arg\max_{w} \mathbb{E}[\exp(c_1 V + c_2)] \tag{3.3.27}$$

Taking at each time step t the gradient of the activation resulting from a single random sample x_t, rather than the gradient of the expected value, an iterative, *stochastic gradient ascent* algorithm can be formulated to maximize this expression[46]:

$$\omega_{t+1} = \omega_t + \eta \nabla_{\omega_t} \exp(c_1 \langle x_t, \omega_t \rangle + c_2) \tag{3.3.28}$$

$$= \omega_t + \eta c_1 x_t \underbrace{\exp(c_1 \langle x_t, \omega_t \rangle + c_2)}_{\lambda_t} \tag{3.3.29}$$

$$= \omega_t + \eta_0 x_t \lambda_t \tag{3.3.30}$$

Under the constraint on the norm of the weight vector $||\omega|| = 1$, this algorithm will rotate ω in the input space into the direction for which the resulting membrane potential distribution is the "least gaussian", thus extracting an independent component. As can be seen from equation 3.3.30, this update mechanism takes the form of the most basic Hebbian learning rule of synaptic plasticity – the product of simultaneous *pre-synaptic activities* (x_t) times *post-synaptic activity* (λ_t) with *learning rate* η_0. This implies that simple Hebbian learning with (rate-coding) exponential neurons and synaptic scaling is sufficient to realize ICA where linear methods such as PCA fail. A numerical proof of concept of this conclusion is given in chapter 4.

Contrary to e.g. Savin, Joshi, and Triesch (2010), *intrinsic plasticity* in this setup does not appear necessary for ICA, but it could play the supporting role of

---

[46] This is a technique referred to as *stochastic gradient ascent* – the stochasticity is due to the fact that the gradient is not averaged but sampled – for which according to Bottou (2010) "almost sure convergence under mild conditions" can be shown. These conditions include those used for ordinary gradient ascent, such as sufficiently small step sizes and "well behaved" functions to maximize.

realizing a zero mean unit variance distribution of the data as chosen here and often enforced during preprocessing by *data whitening*.[47]

---

[47] Hyvärinen and Oja 2000.

Simulation-Based Numerical Results

All programs used in this chapter are written in the Python programming language using the SciPy library stack.[1] Numerical solutions to differential equations are calculated using discrete time steps of length $\triangle t = 1$ms of simulated time. The codes used to generate the shown figures are included in Appendix C.

4.1 Mappings for a Wide Range of Distributions

In this section, an exponential family neuron model is implemented and simulated, in order to numerically test the hypothesized ability of the neuron to adaptively map different, time-varying membrane potential distributions to fixed distributions of activation.

Since the focus in this section is on the mapping between membrane potential and activation, the spike generation, refractoriness and integration of synaptic inputs in the membrane potential are not modeled here. The model used in this section is given by the following equations, that are repeated from earlier chapters:

$$\mu_t = \alpha \phi(V_t) + (1 - \alpha)\mu_{t-1}$$
$$\hat{\theta}_t = \gamma(\mu_t)$$
$$\lambda_t = G_2(F_1(V; \hat{\theta}_t), \psi)$$

---

[1] Oliphant 2007; Rossum 2012; Jones et al. 2007.

Here, V_t is the random membrane potential at time t, drawn from a distribution with CDF $F_1(V; \theta_t)$. The input distribution is a continuous exponential family distribution with sufficient statistics ϕ, log-partition function g, some base measure and the (time-varying) natural parameters θ_t. The running parameter estimate $\hat{\theta}_t$ is derived from a running estimate μ_t of the sufficient statistics ϕ via $\gamma = (\nabla g)^{-1}$, the inverse gradient of the log-partition function g. In each of the experimental setups, the activation function is chosen corresponding to the respective distributions of membrane potential and activation, where the parameters of the output distribution are given and the parameters of the input distribution are estimated via the neuron's intrinsic plasticity mechanism.

In the first setup (see figure 4.1.1), the neuron is presented with a normal distribution of membrane potentials and a desired log-normal distribution of activation. The parameters of the membrane potential distribution change at some points in time, and the adaptation of the neuron's estimate of the membrane potential distribution towards the true distribution as well as the adapation of its true activation distribution towards the desired distribution are illustrated. The expression for the Kullback-Leibler Distance (KLD) derived in 3.2.2 is used to quantify the statement made there, that the true activation distribution "converges"[2] in KLD towards the desired distribution as the parameter estimates of the membrane potential distribution converge to the true parameters. The results support the hypothesis and show that the exponential example neuron model discussed at length in section 3.3 realizes the hypothesized homeostatic mapping from a normal to a log-normal distribution.

In a similar setup (see figure 4.1.2), multiple different exponential family distributions are used for the membrane potential as well as the activation in order to demonstrate that the results shown for the normal-log-normal model generalize to other exponential family distributions, as well. Cumulative histograms of random samples of the neuron's activation are compared to the desired CDF using the *Kolmogorov-Smirnov Distance* or *test*, a widely used measure for how well an empirical distribution approximates an analytically given distribution,[3] in order to

---

[2] Since KLD is not a metric, rather than convergence in the strict sense this means that the KLD converges to 0.   [3] Shimokawa and Shinomoto 2009; Haslinger et al. 2012; Brown et al. 2002; Hromádka, DeWeese, and Zador 2008; Truccolo et al. 2005.

Figure 4.1.1: See page 53. The model neuron receives membrane potentials drawn from one of three different normal distributions for a period of 10000ms of simulated time each (color coded throughout the figure). The top row shows the resulting membrane potential over time. Due to the neuron's intrinsic plasticity mechanism, the membrane potential distribution estimated by the neuron slowly converges to the true distribution. This is shown in the second row, where for each of the three periods the true membrane potential distribution is shown dashed in black, and estimated distributions from different points in time during that period are shown as solid lines (with early estimates in lighter shades). Towards the end, the estimated and true distribution are virtually indistinguishable. The resulting activation of the neuron is shown as a function of time in the middle row. The true activation distribution of the neuron at each point in time can be calculated using the true and estimated membrane potential distribution. As the parameter estimates converge to the true parameters, the activation distribution should converge to the desired distribution. This can be seen in row 4, where for the three periods the desired activation distribution is plotted in dashed black (identical across all periods) on top of the true activation distribution at different points in time during the corresponding period. Again, the match towards the end of each period is very good. To quantify this, the bottom row shows the Kullback-Leibler Divergence between the true and the desired activation distribution over time. It is apparent, that changing the input distribution results in a sudden mismatch of these distributions, but over time the divergence slowly converges to 0, thus supporting the analyical results from section 3.2.2.

verify the analytical predictions. Similarly to figure 4.1.1, where the KLD between the analytically derived actual and desired activation distribution is shown to increase and then decay to zero at the points in time where the parameters of the membrane potential distribution change, the KSD, measuring the mismatch between empirical and desired distribution, increases at the same points and decays back to low values over time. The results demonstrate the flexibility and power of the class of adaptive exponential family neuron model defined in chapter 3.

Figure 4.1.2: See pages . In a very similar setup to figure 4.1.1, a model neuron receives membrane potentials drawn from an exponential family distribution parameterized by one of three sets of parameters for time-periods of 15000ms each. The procedure is repeated for 3 different families of distributions: Laplace, normal and gamma distributions, resulting in 3 full page figures. The true membrane potential distributions (dashed black), as well as the estimate at the end of each period (solid lines, color coded), are shown in the top left of each figure. The sampled potentials are shown in the top right as a function of time. For 5 different desired distributions of activation (dashed black, all exponential family members), a histogram of the resulting activations in the last 5000ms of each period are shown in the left colum, showing a good match in all cases. The right column shows the mismatch (Kolmogorov-Smirnov Distance) between the desired activation distribution and the histogram of the actual activations within a sliding window of 200ms.

probabilities /rel. frequencies

sampled activations

In: Laplace

activation

KSD over time

Out: Weibull

Out: Beta

Out: Log-Normal

Out: Log-Logistic

Out: Gamma

activation

time [s]
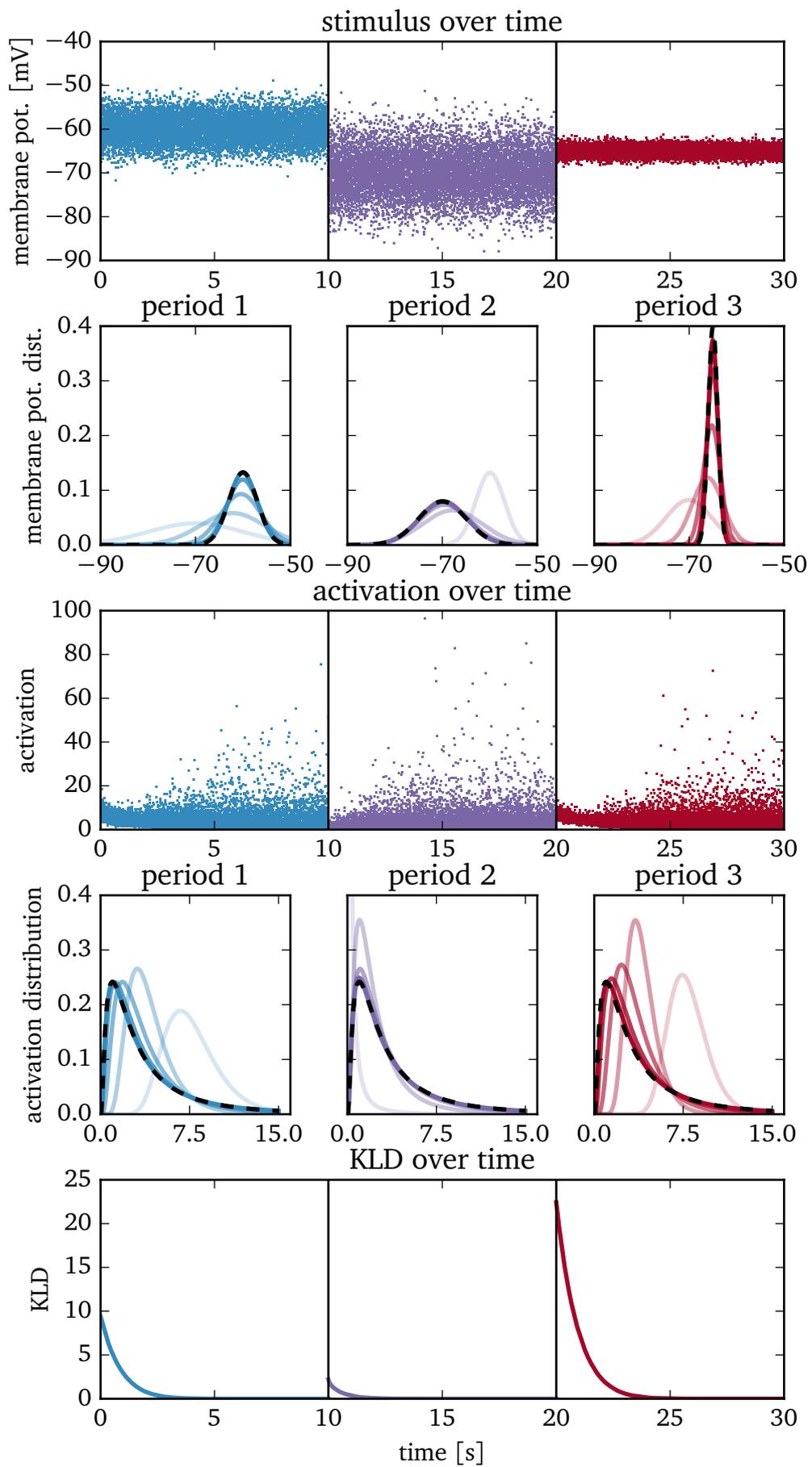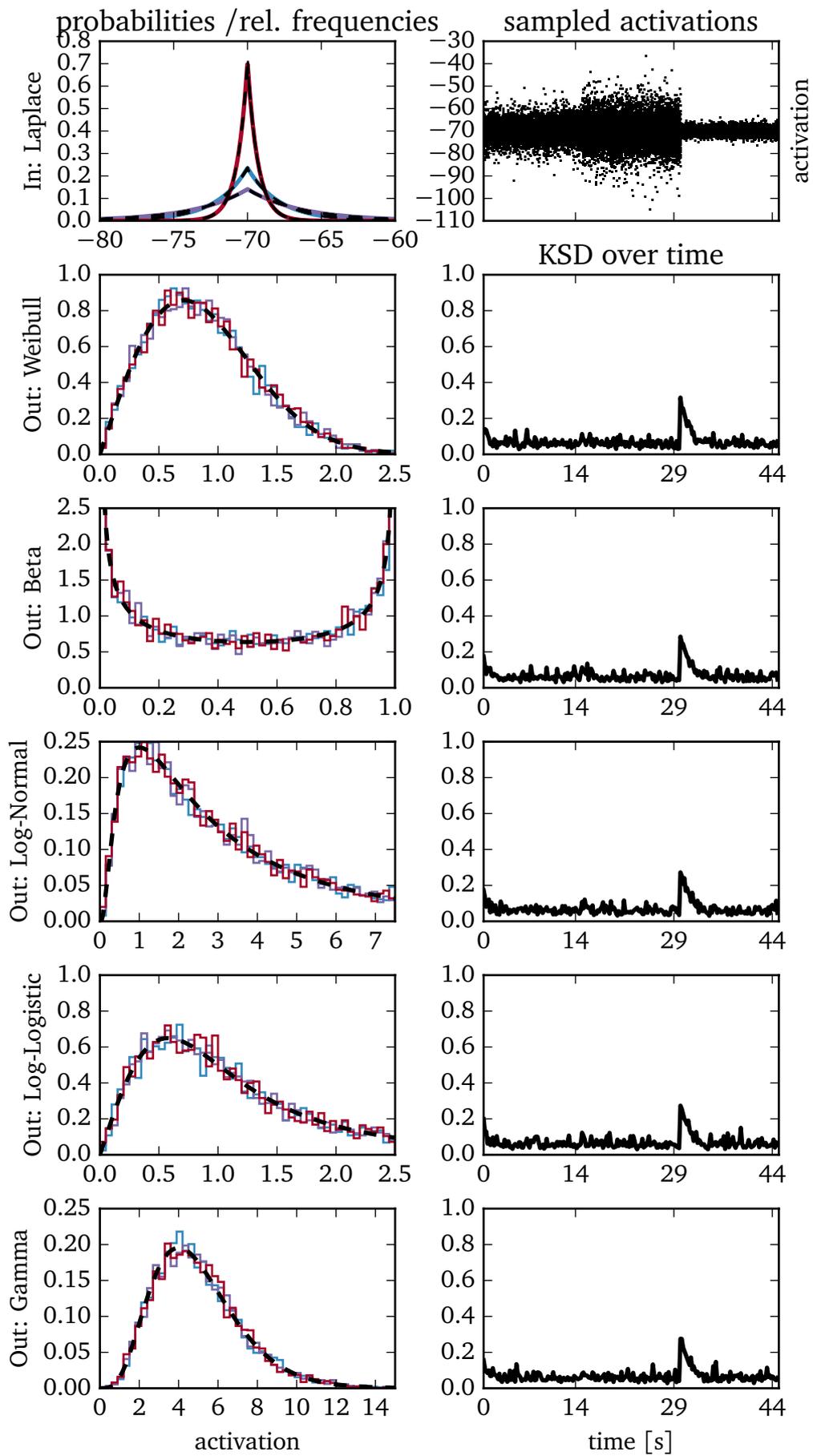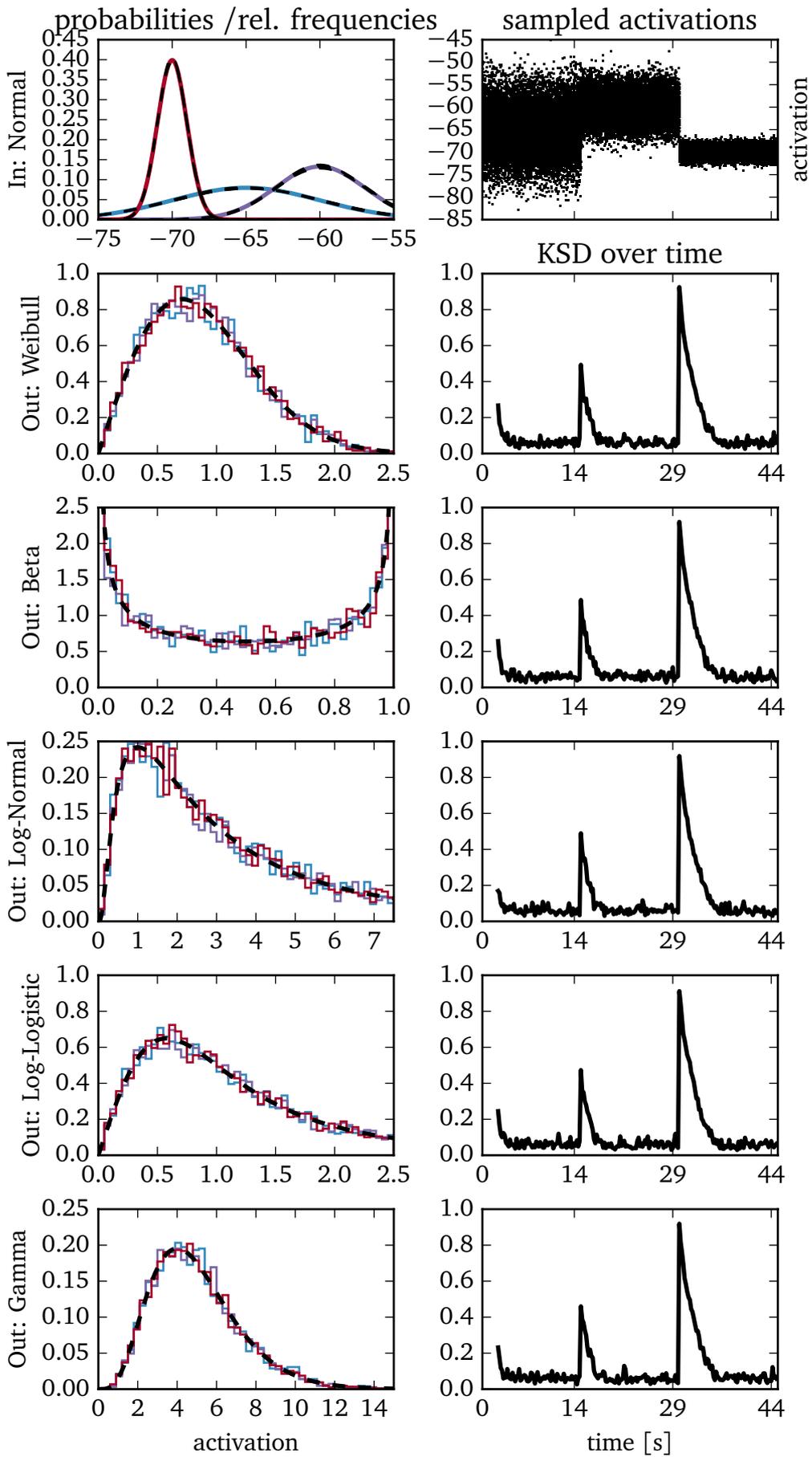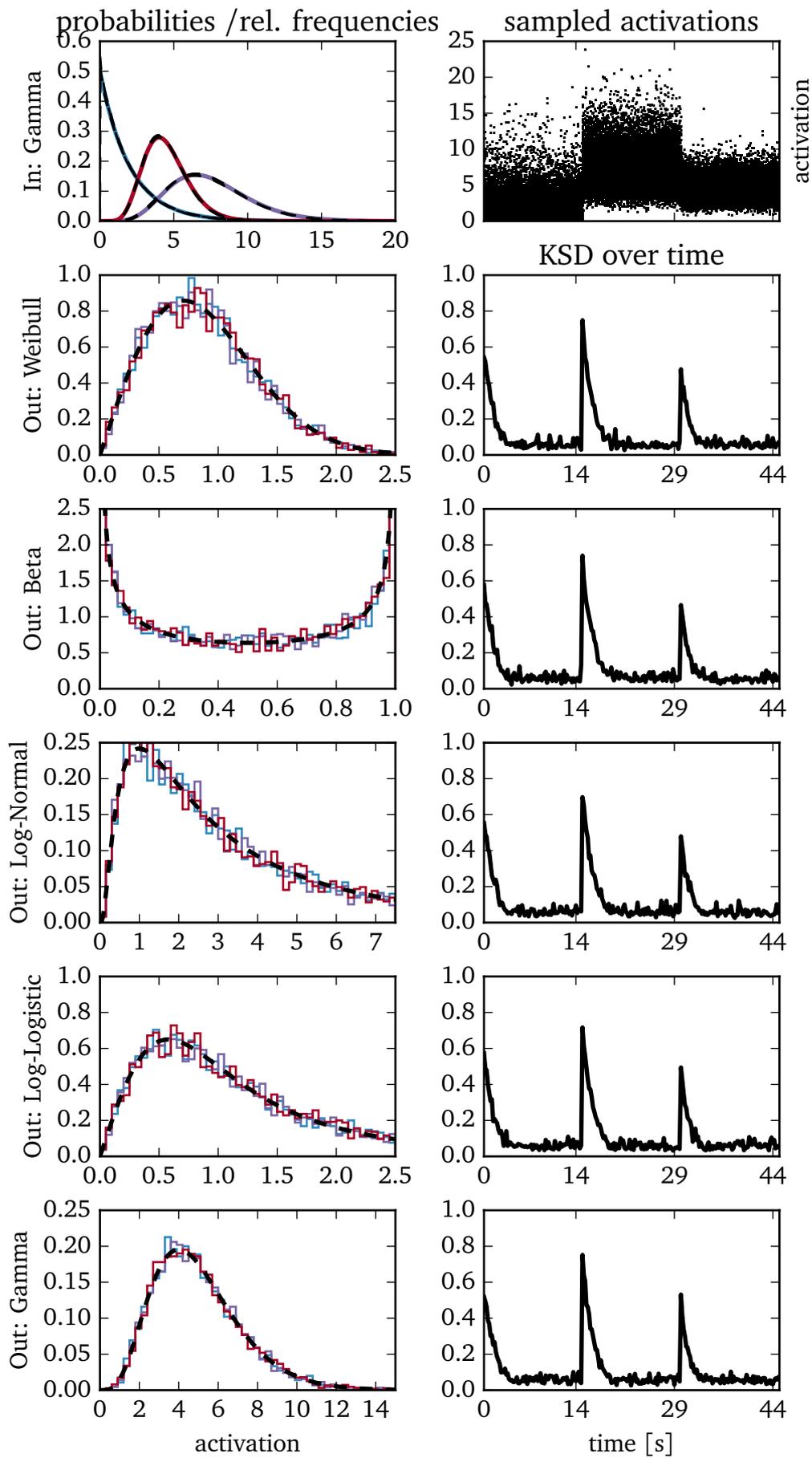
4.2   Dynamical Systems Model

A link between the LNP model as defined in this thesis and dynamical systems models is established in section 3.3.3. Here, the proposed neuron model is implemented as a system of first order differential equations and numerically solved to provide a proof of concept.

The model is defined according to the following system of equations, all of which are repetitions from previous sections:

$$I(t) \sim \text{Normal} \tag{4.2.1}$$

$$U(t) \sim \text{Uniform} \tag{4.2.2}$$

$$\tau_M \frac{\mathrm{d}V}{\mathrm{d}t} = -V + I(t) \tag{4.2.3}$$

$$\tau_A \frac{\mathrm{d}\tilde{\mu}_1}{\mathrm{d}t} = -\tilde{\mu}_1 + V \tag{4.2.4}$$

$$\tau_A \frac{\mathrm{d}\tilde{\mu}_2}{\mathrm{d}t} = -\tilde{\mu}_2 + V^2 \tag{4.2.5}$$

$$\theta = (\log(\frac{-\log(U)}{\triangle t}) - \mu_2)\frac{\sqrt{\tilde{\mu}_2 - \tilde{\mu}_1(t)^2}}{\sigma_2} + \tilde{\mu}_1 \tag{4.2.6}$$

$$X(t) = (V(t) > \theta(t)) \tag{4.2.7}$$

Here, I and U are piece-wise constant random functions of time. They are step-functions with steps of length $\triangle t$ and represent the neurons synaptic input and a source of randomness for the threshold generation. Mean and variance of I are changed after 20 and 40 seconds of simulated time.

As before, V is the neurons membrane potential, which is modeled to decay back to its resting value (the expected value $\mathbb{E}[I] \in \{-65\text{mV}, -60\text{mV}, -55\text{mV}\}$) with time-constant τ_M, chosen to be very small (2ms). The adaptation variables $\tilde{\mu}_1$ and $\tilde{\mu}_2$ have the much slower time-constant τ_A (set here to the still very small value of 1s). The threshold θ is given by the inverse of the activation function of the LNP model, parameterized by the adaptation variables as well as the parameters of the desired output distribution. Spikes are sampled whenever the membrane potential exceeds the threshold. For the purpose of better visualization, refractory

effects of the model are neglected, time constants are chosen artificially small and the desired activation distribution is chosen to enforce high firing rates. See figure 4.2.1 for a summary of the results.

It is visually apparent, that the neuron model succeeds in estimating its own input distribution, as to be expected considering that the model is equivalent to the model analyzed in the previous section. The homeostatic nature of the model can be seen in the fact that systematic changes in membrane potentials are followed by compensatory systematic changes in the neuron's threshold, thus maintaining the overall firing probability.

4.3   Model Fitting from Spikes

Section 2.4.3 discusses the problem that neuronal activation can typically not be directly observed and must thus be reconstructed from spiking activity. A simple theoretical solution to this problem is given by equation 2.4.16 that gives the likelihood function of the parameters η given a set of observed *inter-spike intervals* (ISIs) τ as

$$P(\eta|\tau) \propto \int_0^\infty P(\tau|\lambda)P(\lambda|\eta)\mathrm{d}\lambda \tag{4.3.1}$$

For the example neuron model as defined in section 3.3 under the assumption of slow changing activation $\lambda(t) \approx \lambda_0$ for some time interval around t, the probability distribution over ISIs given the activation is

$$P(\tau|\lambda) = \int_0^\tau \lambda(t)\mathrm{d}t \exp\left(-\int_0^\tau \lambda(t)\mathrm{d}t\right) \approx \tau\lambda_0\exp(-\tau\lambda_0) \tag{4.3.2}$$

The distribution of activation is here assumed to be the log-normal distribution with PDF $P(\lambda_0|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma\lambda_0}\exp\left(-\left(\frac{\log(\lambda_0)-\mu}{\sqrt{2}\sigma}\right)^2\right)$. The resulting joint log-likelihood for a set of independelty observed ISIs is then given by the sum of the log-likelihoods of the individual ISIs τ, which is given by

$$\log\mathscr{L}(\mu,\sigma|\tau) = \log\left(\int_0^\infty \underbrace{\exp(-\left(\frac{\log(\lambda_0)-\mu}{\sqrt{2}\sigma}\right)^2 - \tau\lambda_0)\mathrm{d}\lambda_0}_{=:K(\lambda_0,\tau,\mu,\sigma)}\right) + \log\left(\frac{\tau}{\sqrt{2\pi}\sigma}\right) \tag{4.3.3}$$

Figure 4.2.1: The model as defined by equations 4.2.1 to 4.2.7 is numerically solved using *Heun's method*. It receives a piecewise constant (for time intervals of $\triangle t = 1$ms each) randomly drawn input from a normal distribution with mean and variance that are changed after 20 and 40 seconds of simulated time. Blue dots show the neuron's membrane potential V, purple dots its threshold θ and vertical bars the times of spikes (their vertical position is centered at the neuron's membrane potential at the time of the spike). The gray area shows ± 1 standard deviation around the mean (solid red line) of the neuron's estimated Gaussian input distribution. For comparison, dashed lines show the corresponding confidence interval of the true input distribution.

The derivative with respect to a parameter $\theta \in \{\mu, \sigma\}$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}\theta}K(\lambda_0, \tau, \mu, \sigma) = K(\lambda_0, \tau, \mu, \sigma)\frac{\mu - \log(\lambda_0)}{\sigma}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\log(\lambda_0) - \mu}{\sigma}\right) \tag{4.3.4}$$

$$\frac{\mathrm{d}}{\mathrm{d}\mu}K(\lambda_0, \tau, \mu, \sigma) = K(\lambda_0, \tau, \mu, \sigma)\frac{-\mu + \log(\lambda_0)}{\sigma^2} \tag{4.3.5}$$

$$\frac{\mathrm{d}}{\mathrm{d}\sigma}K(\lambda_0, \tau, \mu, \sigma) = K(\lambda_0, \tau, \mu, \sigma)\frac{(-\mu + \log(\lambda_0))^2}{\sigma^3} \tag{4.3.6}$$

$$\frac{\mathrm{d}\mathscr{L}(\mu, \sigma|\tau)}{\mathrm{d}\theta} = \frac{\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}\theta}K(\lambda_0, \tau, \mu, \sigma)\mathrm{d}\lambda_0}{\int_0^\infty K(\lambda_0, \tau, \mu, \sigma)\mathrm{d}\lambda_0} - \frac{\mathrm{d}}{\mathrm{d}\theta}\log(\sigma) \tag{4.3.7}$$

$$\frac{\mathrm{d}\mathscr{L}(\mu, \sigma|\tau)}{\mathrm{d}\mu} = \frac{\int_0^\infty K(\lambda_0, \tau, \mu, \sigma)\frac{-\mu + \log(\lambda_0)}{\sigma^2}\mathrm{d}\lambda_0}{\int_0^\infty K(\lambda_0, \tau, \mu, \sigma)\mathrm{d}\lambda_0} \tag{4.3.8}$$

$$\frac{\mathrm{d}\mathscr{L}(\mu, \sigma|\tau)}{\mathrm{d}\sigma} = \frac{\int_0^\infty K(\lambda_0, \tau, \mu, \sigma)\frac{(-\mu + \log(\lambda_0))^2}{\sigma^3}\mathrm{d}\lambda_0}{\int_0^\infty K(\lambda_0, \tau, \mu, \sigma)\mathrm{d}\lambda_0} - \frac{1}{\sigma} \tag{4.3.9}$$

The joint log-likelihood and its gradient can be numerically evaluated for a given dataset of spikes and parameters μ and σ. This makes it possible to use an iterative algorithm to numerically find the parameters for which the likelihood is largest.

For 4 different arbitrarily chosen sets of parameters, a total of 1000 ISIs is sampled from the theoretical ISI distributions resulting from homogeneous Poisson processes with rates λ randomly drawn from the corresponding log-normal distribution[4]. The joint log-likelihood given the set of ISIs is plotted as a contour plot for illustration purposes, evaluated on a two-dimensional grid of $25 \times 25 = 625$ parameter combinations of μ and σ, covering the search space $(\mu, \sigma) \in [-2, 2] \times [0.01, 2.0]$[5]. An optimal parameter vector is found using the *Newton conjugate gradient ascent* algorithm, for which the expressions given above are numerically evaluated.

Much more sophisticated methods could be used,[6] but this basic method provides an intuitive proof of the concept, that it is possible in principle to infer the

---

[4] In between each two spikes, the firing rate is assumed to be constant, which is a strong simplification   [5] Points are logarithmically spaced in the σ-direction, due to its divisive role, and linearly in the μ direction.   [6] Cunningham et al. 2008.

distribution of activation from spiking activity alone, without the need to know or control the neuron's membrane potential or make assumptions about the timescale of changes in the activation. Figure 4.3.1 shows the log-likelihood surface on the search space in a color coded contour plot (red indicating high, blue indicating low likelihood), as well as the parameters used to generate the ISIs and the reconstructed estimates. Figure 4.3.2 shows the corresponding ISI distributions resulting from the different parameterizations of the distribution of activation. It seems apparent, that different parameter choices of the distribution of activation are reflected in the empirical distribution of ISIs and that it might thus be possible to recover them.

4.4   Symbiosis with Synaptic Plasticity

Section 3.3 concludes that for a neuron with exponential activation function, corresponding to the assumption of normally distributed membrane potentials and log-normally distributed activations with the same parameters μ and σ, Hebbian synaptic plasticity acting on a normalized weight vector maximizes the neuron's expected output and thus implements independent component analysis. These predictions are tested in this section by numerical simulations with a neuron model as defined in section 4.1, using an exponential activation function.

To illustrate this, a toy problem is constructed (see figure 4.4.1 for a summary): Suppose a neuron in the visual system receives synaptic inputs from a 3×3 grid of sensory neurons, coding for e.g. light intensity. For simplicity, the response of these sensors is assumed to be a value in $\mathbb{R}$ and the potential inputs to this *receptive field* are thus the elements of $\mathbb{R}^9$. Any such input can, by definition, be represented as a linear combination of any basis of this space and can thus be viewed as the *linear mixture* of a set of 9 arbitrary orthogonal vectors, or *independent components*. Since any input can be decomposed with respect to any basis, the task of recovering the originally used orthogonal basis is an underdetermined problem, but it can be solved probabilistically.[7]

---

[7]   The basis is chosen to be orthogonal, because this implies that the fourier coefficients of an input with respect to this basis can be calculated for each of the basis functions independently.
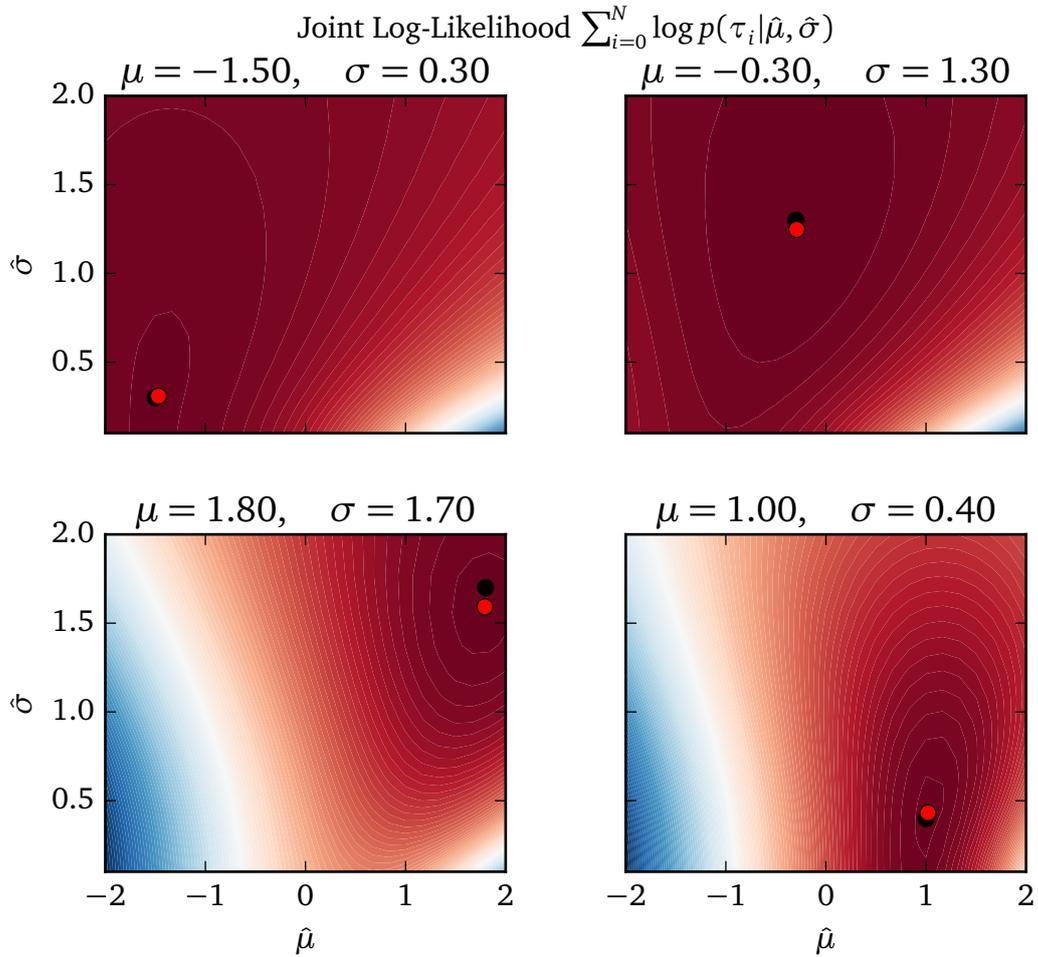
Figure 4.3.1: The subplots show a contour plot of the estimated log-likelihood function for the parameters μ and σ of a log-normal distribution of activation estimated from a total of 1000 simulated spikes per condition. The real parameters (marked by black dots) could be closely recovered by the *Newton-CG method* (red dots show ML estimates).
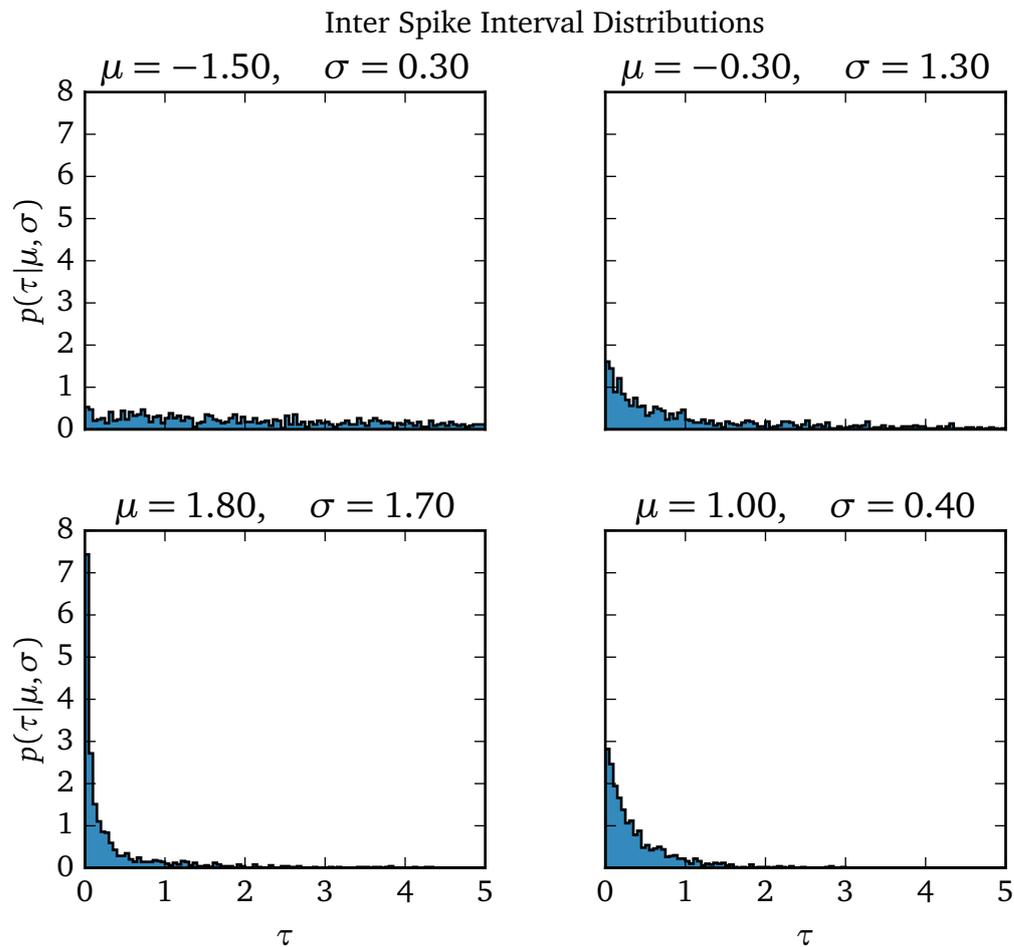
Figure 4.3.2: The subplots of this figure show histograms of the inter-spike intervals created from the log-normal activation distributions corresponding to the subplots in figure 4.3.1. The ISI distributions are qualitatively different, thus allowing inference about the underlying parameters used to generate them.

Figure 4.4.1: See page 65. A random set of orthogonal basis vectors of $\mathbb{R}^9$ is used to represent different features in a 2D, 3×3 pixel visual receptive field, shown in the left column of the figure. The features are repeatedly linearly combined using random weights drawn from a Laplace distribution to yield random visual inputs as shown in the second column. These inputs are projected through different vectors of synaptic weights (using the vector product in $\mathbb{R}^9$) to multiple model neurons. The resulting scalar membrane potentials are transformed by an exponential activation function (fourth column) into the neurons' activations (rightmost column). Over the course of 10^7 samples, the neurons' weights, updated with simple hebbian plasticity, converge to the vectors shown in the middle column. They resemble the basis functions from which the inputs are generated, thus showing that the independent components are succesfully recovered. For the purpose of illustration, the neurons are manually decorrelated via lateral inhibition and thus forced to discover different independent components.

If the coefficients of the linear combination of independent components are randomly drawn from a non-gaussian distribution (e.g. Laplace, as here), the results from section 3.3.7 suggest that any resulting mixtures are at least as "gaussian" as the unmixed independent components. Maximizing non-gaussianity, as measured by the neuron's expected activation due to its non-linear activation function should thus result in demixing of the signals, i.e. the recovery of independent components by the neurons' weight vectors. Maximization of the expected outputs is done via the simple hebbian learning rule that modulates each neuron i's weight vector $\omega_{i,t}$ given by the equation $\omega_{i,t} = \omega_{i,t} + \eta_0 x_t \lambda_{i,t}$ based on the neuron's activations $\lambda_{i,t}$ and the input vector x_t at time t.

The random mixing coefficients are drawn from a centered Laplace distribution, thus centering the multivariate random input vectors. Intrinsic plasticity is not included in the model, but under the assumption that it operates on a time-scale much slower than synaptic plasticity, it could be used to account for this centering of the inputs in a more biologically plausible way.

The simulation results support the hypothesis, that a simple model neuron with an exponential activation function and Hebbian synaptic plasticity acting on fixed-length weight vectors can implement independent component analysis.

features mixtures weights act. func. activation
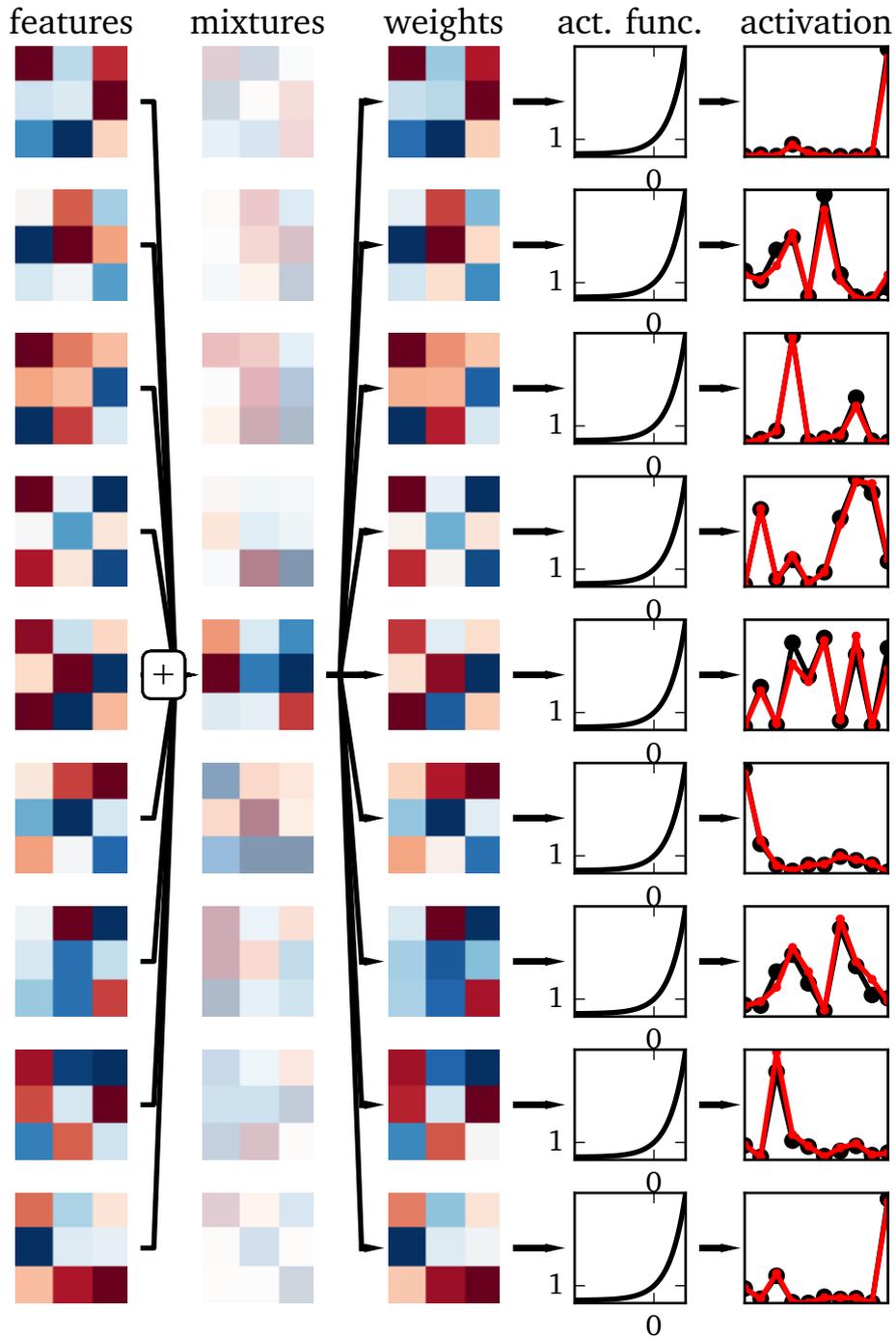
Discussion

5.1 Summary

In this thesis, a simple insight from probability theory, the *probability integral transform theorem* and its inverse, was used to address two principal questions about modeling the behavior of neurons:

1. (How) can the operation of a neuron be derived from measurements of the statistical properties of its in- and output?

2. (How) can an intrinsic homeostatic mechanism be implemented that keeps the output statistics in a "desirable" regime?

The first question was answered by defining a *linear-non-linear* neuron model that is uniquely defined by and able to reproduce given in- and output statistics. The method is fairly general and works for arbitrary distributions of activation and continuous distributions of membrane potential. Some possible extensions and alterations to the model such as *Poisson spike generation*, inclusion of *auto-history* effects and a reformulation as a *dynamical systems* model with random thresholds were presented.

An approach was briefly outline how such a model could be derived when only statistics of the spiking output, not the continuous activation (i.e. rate), are available.

The theoretical results were then enlarged upon for the special case of *exponential family* distributions of membrane potential and activation. For this class, the second principal question was addressed using the concept of *Bayesian sequential updates* with *conjugate priors*. The *sensitivity* of the neuron's distribution of activation to changes in the membrane potential distribution and the *convergence* to the desired activation distribution were analyzed.

To illustrate these concepts, the simple adaptive neuron model resulting from a normal membrane potential and a log-normal activation distribution was derived explicitly and an expression for the threshold of an equivalent dynamical systems model was presented. The resulting exponential activation function was used to illustrate, how in such a model auto-history effects could be included as "virtual" synaptic inputs, thus simplifying the model.

A re-interpretation of the resulting model as a *generalized linear model* with *log-link* function was briefly discussed, which might allow some transfer of knowledge and vocabulary from these well studied methods. The perspective of the neuron as a GLM was then used to motivate that a neuron with exponential activation function implements a *multiplication calculus* in some ways similar to the logical calculus traditionally suggested in early perceptron studies.

Interactions of the example neuron model with hebbian *synaptic plasticity* were then analyzed in a simple example problem for which the emergence of *independent component analysis* could be proven.

Finally, numerical simulations were used to illustrate and experimentally verify the most interesting aspects of the theoretically derived results.

5.2   Scientific Context

Homeostatic plasticity has been suggested as a candidate mechanism to explain a host of biological phenomena, in particular *gain modulation* or self regulation of the activity in neural circuits.[1] Specifically *intrinsic plasticity* and its interaction with synaptic plasticity and synaptic scaling have been studied to explain homeostatis in biological neurons,[2] but with an emphasis on underlying biological mechanisms and the resulting complexity. Similar questions have been addressed from a

---

[1] Turrigiano and Nelson 2000.   [2] Ibid.

more abstract perspective by Triesch (2007), for example, but with a focus on finding good parameter values for predefined activation functions which result in a good yet imperfect fit. [3] The approach here differs in so far, as the activation function itself is derived from the supposed distributions of membrane potential and activation, thus reducing the problem of finding optimal parameters for the activation function to the problem of finding parameters for the corresponding distributions, which can be solved by the neuron itself by estimating the sufficient statistics of its membrane potential distribution.

Since this approach requires only passive observations of membrane potential and activation, it can also be used by researchers to "adapt" their neuron model to in vivo data, thus allowing the model to approximate the neuron's natural behavior without the need to emulated the neuron's environment.

5.3   Outlook

The results of this thesis suggest several directions in which further research should be conducted:

5.3.1   Adaptive Distribution of Activation

Here, only adaptation of the neuron with respect to its membrane potential was discussed, while the output distribution was assumed to be fixed. In a framework of homeostatic plasticity however, it is also reasonable to assume that the neuron should maintain a stable distribution of membrane potentials by modulating its activation such that some feedback mechanism stabilizes its input – much akin to an *operational amplifier* in electronics. Research in this direction would require some further insights into biological candidates for such feedback mechanisms.

A different proposition made by Triesch (ibid.) is to study how external modulators might influence the neuron's desired distribution of activation. Modulating the parameters of a neuron's activation distribution could allow the brain to selectively

---

[3] Some of the ideas presented in this thesis are implied there, albeit not fully fleshed out, which I only discovered late in the progress of writing my thesis. Where appropriate, I retrospectively attributed or referenced ideas already present there.

(in)activate neurons or fine-tune them to optimize the operation of neural circuits. In addition to homeostasis based on the individual neuron's activity, this could allow homeostasis with respect to network activity or other global criteria.

5.3.2   Link to Dynamical Systems Models

The connection to dynamical systems models should be further investigated, since the perspective of LNP neuron models can contribute tools for statistical analysis, whereas the dynamical systems perspective offers a more natural and biologically plausible way of expressing the resulting adaptive process. The connection between integrate-and-fire neuron models with stochastic thresholds and LNP neuron models appears to provide a good basis for that.

5.3.3   Exploring Different Models

Of course, only a narrow section of possible neuron models was presented here. While the focus on exponential family probability distributions was motivated both from a theoretical perspective as well as from some biological evidence, more complex models mapping from or to e.g. mixture models could be studied. This could be used to derive models for neurons with an observed multi-modal activation, e.g. neurons that express so called *up-* and *down states* or neurons that receive inputs from a multi-stable network.

Finally, due to the generality of the model discussed here, the same tools could be used to implement homeostatic mechanisms in domains other than neuroscience, such as cybernetics, machine learning of finance, where models need to be able to cope with and adapt to systematic changes in incoming data.

Appendices

Lemmas and Proofs

Lemma 9. *The generalized inverse g of a cumulative distribution function f is a non-decreasing left-continuous function with the limits $\lim_{x \to 0} g(x) = -\infty$ and $\lim_{x \to 1} g(x) = \infty$.*

Proof. Theses proofs can be found in Embrechts and Hofert (2013, prop. 2.3). We only prove here that the generalized inverse is non-decreasing:

Let f and g be as defined in lemma 9. The function g is non-decreasing, because for all $y_0 \leq y_1 \in (0,1)$ and for all $x \in \mathbb{R}$ we have $f(x) \geq y_1 \Rightarrow f(x) \geq y_0$ and thus $g(y_0) = \min\{x \in \mathbb{R}|f(x) \geq y_0\} \leq \min\{x \in \mathbb{R}|f(x) \geq y_1\} = g(y_1)$. $\qquad\square$

Lemma 10. *Let $g : (0,1) \to \mathbb{R}$ be a non-decreasing, left-continuous function and*

$$f : \mathbb{R} \to (0,1), x \mapsto \max\{s \in (0,1)|g(s) \leq x\}$$

Then g is the generalized inverse of f. The function $f \in \Pi$ is a CDF.

Proof. Let f and g be as defined in lemma 10 and let $y \in (0,1)$ be arbitrarily chosen.

The following two results hold because g is non-decreasing and left-continuous:

Let $t = g(y)$: Then $f(t) = \max\{s \in (0,1)|g(s) \leq t = g(y)\} \geq y$.

Let $t < g(y)$: Then $f(t) = \max\{s \in (0,1)|g(s) \leq t < g(y)\} < y$.

This implies that $g(y) = \min\{t \in \mathbb{R} | f(t) \geq y\}$ which in turn proves that g is the generalized inverse of f.

The proof the limits and right-continuity of f as well as the proof that f is non-decreasing are analogous to the proof of the corresponding properties of the generalized inverse in lemma 9 and are not shown here. $\qquad\square$

Lemma 11. *Let* $f \in \Pi$*, then for each point* $y \in f(\mathbb{R})$*, the fiber* $f^{-1}(\{y\})$ *is a half-closed or closed interval* $[x_{min}, x_{max})$ *or* $[x_{min}, x_{max}]$*, respectively.*

Proof. The fiber is empty by definition if and only if $y \notin f(\mathbb{R})$. It is a singleton set $\{x\}$ if and only if there exists a unique element $x \in \mathbb{R}$ with $f(x) = y$, which corresponds to the trivial interval $[x, x]$.

Now assume that the fiber $f^{-1}(\{y\})$ is neither empty nor a singleton set. The fiber must then be a convex set, as follows from the assumption that f is non-decreasing: Let $x_0 \leq x_1 \leq x_2 \in \mathbb{R}$ be arbitrary points with $x_0, x_2 \in f^{-1}(\{y\})$. It follows that $y = f(x_0) \leq f(x_1) \leq f(x_2) = y$ and thus $f(x_1) = y \Rightarrow x_1 \in f^{-1}(\{y\})$. Since we consider only subsets of $\mathbb{R}$, this implies that the fiber is an interval.

The fiber is closed at the infimum, as follows from the right-continuity of f: Let $(x_i)_{i \in \mathbb{N}}$ be a sequence in $f^{-1}(\{y\})$ that converges to $x_{\inf} := \inf f^{-1}(\{y\})$. Then $(f(x_i))_{i \in \mathbb{N}}$ is just the constant sequence $(y)_{\mathbb{N}}$ and thus converges to y. Since f is right-continuous we have $f(x_{\inf}) = \lim_{i \to \infty} f(x_i) = y$ and thus $x_{\inf} \in f^{-1}(\{y\})$, the fiber is thus closed at the infimum. $\qquad\square$

Lemma 12. *Let* X *be a random variable with uniform distribution, let* $F_1 \in \Pi$ *be a CDF and let* G_1 *be the generalized inverse of* F_1*. Let* $G_2 : (0, 1) \to \mathbb{R}$ *be a non-decreasing, left-continuous function and let* $Y := G_2(X)$ *be a random variable with CDF* F_1*. Then* $G_1 = G_2$*.*

Proof. Let X, F_1, G_1, G_2 and Y be defined as in lemma 12.

Using lemma 10 we construct the function $F_2 : x \rightarrowtail \max\{s \in (0, 1) | G_2(s) \leq x\}$. Since G_2 is the generalized inverse of $F_2 \in \Pi$, theorem 4 implies that the random variable $Y = G_2(X)$ must have the CDF F_2. Because the random variable Y has by definition the CDF F_1, this implies that $F_1 = F_2$.

Since G_1 and G_2 are uniquely defined as the generalized inverse of F_1 and $F_2 = F_1$, respectively, this implies that $G_1 = G_2$. $\qquad\square$

APPENDIX A. LEMMAS AND PROOFS

Proof of theorem 3. Let X_1 be a continuous random variable with cumulative distribution function F_1, let F_U denote the CDF of $U := F_1(X_1)$ and let $x \in \mathbb{R}$ be arbitrarily chosen.

Since F_1 is a non-decreasing function, $X \leq x \Rightarrow F_1(X) \leq F_1(x)$, thus also:

$$X \leq x \vee F_1(X) = F_1(x) \Rightarrow F_1(X) \leq F_1(x)$$

The converse also holds: Since F_1 is a non-decreasing function, $x \leq X \Rightarrow F_1(x) \leq F_1(X)$ and thus by contradiction $F_1(x) \not\leq F_1(X) \Rightarrow x \not\leq X$, i.e. $F_1(X) < F_1(x) \Rightarrow X < x \Rightarrow X \leq x$. Thus:

$$F_1(X) \leq F_1(x) \Rightarrow X \leq x \vee F_1(X) = F_1(x)$$

Combining these results and factorizing the right hand side yields:

$$F_1(X) \leq F_1(x) \Longleftrightarrow (X \leq x) \vee (X > x \wedge F_1(X) = F_1(x))$$

Since the right hand side of this expression is a disjunction of mutually exclusive propositions, we can write for the probability :

$$\begin{aligned}
(F_U \circ F_1)(x) &= P(U \leq F_1(x)) = P(F_1(X) \leq F_1(x)) \\
&= P((X \leq x) \vee (X > x \wedge F_1(X) = F_1(x))) \\
&= P(X \leq x) + P(X > x \wedge F_1(X) = F_1(x)) \\
&= F_1(x) + P(X \in (x, b) := (x, \infty) \cap F_1^{-1}(\{x\})) \\
&= F_1(x) + \lim_{t \uparrow b} F_1(t) - F_1(x) \\
&= F_1(x) + F_1(x) - F_1(x) = F_1(x)
\end{aligned}$$

Lemma 11 was used above where the fiber of x is treated as an interval. Since $x \in \mathbb{R}$ was arbitrarily chosen and $F_1(\mathbb{R}) = (0, 1)$ due to the continuity of F_1, this implies that $F_{U|F_1(\mathbb{R})} = F_{U|(0,1)} = \text{id}$, thus U has a uniform distribution on $(0, 1)$. $\square$

Proof of theorem 4. Let U, X_2, F_2 and G_2 be as defined as in theorem 4, $x \in \mathbb{R}$ arbitrary, let F_{X_2} and F_U denote the CDFs of X_2 and U, respectively, and let $M_U := \{s \in \mathbb{R} : F_2(s) \geq U\}$. The fact that F_2 is non-decreasing implies that if

75

$G_2(U) = \min M_U \leq x$, then $F_2(G_2(U)) = F_2(\min M_U) = U \leq F_2(x)$. If $F_2(x) \geq U$, then $x \in M_U$ and thus $G_2(U) = \min M_U \leq x$. Thus $G_2(U) \leq x \Leftrightarrow U \leq F_2(x)$. From this, we can conclude:

$$F_{X_2}(x) = P(X_2 \leq x) = P(G_2(U) \leq x) = P(U \leq F_2(x)) = \underbrace{F_U}_{\text{id}}(F_2(x)) = F_2(x)$$

Since $x \in \mathbb{R}$ was arbitrarily chosen, this implies that F_2 is the cumulative distribution function of X_2. $\qquad\square$

Proof of theorem 5. Let X_1, F_1, F_2, G_2 and τ as defined in theorem 5. The random variable $U := F_1(X_1)$ then has a uniform distribution due to theorem 3. Consequently, the random variable $X_2 := G_2(U) = G_2(F_1(X_1)) = \tau(X_1)$ has the cumulative distribution function F_2 according to theorem 4. The function τ is non-decreasing and left-continuous, because it is a composition of non-decreasing functions, one of which (G_2) is left-continuous and the other (F_1) is continuous. $\quad\square$

Proof of theorem 6. Let $F_1, F_2, G_2, \tilde{\tau}$ and τ be as defined in theorem 6. Suppose now that $\tilde{\tau} \neq \tau$ and let U be a random variable with uniform distribution on the open interval $(0, 1)$. Then, since F_1 is continuous CDF and thus increasing and invertible, the random variable $X_1 := F_1^{-1}(U)$ is well defined and has CDF F_1. Thus both random variables $\tilde{\tau}(X_1) = (\tilde{\tau} \circ F_1^{-1})(U)$ and $\tau(X_1) = (\tau \circ F_1^{-1})(U) = (G_2 \circ F_1 \circ F_1^{-1})(U) = G_2(U)$ have the cumulative distribution function F_2. But since F_2^{-1} is invertible, $\tilde{\tau} \neq \tau$ implies that also $\tilde{\tau} \circ F_1^{-1} \neq \tau \circ F_1^{-1} = G_2$. This is contradicted by the lemma 12, which states that because $\tilde{\tau} \circ F_1^{-1}$ is non-decreasing left-continuous function with $(\tilde{\tau} \circ F_1^{-1})(U) \sim F_2$ and G_2 is the generalized inverse of F_2 that $\tilde{\tau} \circ F_1^{-1} = G_2$.

It thus follows by contradiction that $\tilde{\tau} = \tau = G_2 \circ F_1$. $\qquad\square$

Independent Component Analysis

B.1 Dependency on High Order Moments

The expected activation $\mathbb{E}[\lambda]$ of the exponential neuron model in response to a random membrane potential V can be expressed in terms of the *moment generating function M_V* of the distribution of V:

$$\mathbb{E}[\lambda] = \mathbb{E}[\exp(c_1 V + c_2)] \tag{B.1.1}$$

$$= \exp(c_2)\mathbb{E}[\exp(c_1 V)] \tag{B.1.2}$$

$$= \exp(c_2)M_V(c_1) \tag{B.1.3}$$

where c_1 and c_2 are constants in $\mathbb{R}$.

Using the series expansion of the exponential function, the moment generating

function $M_X(c_1)$ can be rewritten as

$$M_V(c_1) = \mathbb{E}[\exp(c_1 V)] \tag{B.1.4}$$

$$= \mathbb{E}\left[1 + c_1 V + \frac{c_1^2 V^2}{2!} + \frac{c_1^3 V^3}{3!} + \frac{c_1^4 V^4}{4!} + \sum_{n=5}^{\infty} \frac{c_1^n V^n}{n!}\right] \tag{B.1.5}$$

$$= 1 + c_1 \mathbb{E}[V] + \frac{c_1^2}{2!}\mathbb{E}[V^2] + \frac{c_1^3}{3!}\mathbb{E}[V^3] + \frac{c_1^4}{4!}\mathbb{E}[V^4] + \sum_{n=5}^{\infty} \frac{c_1^n}{n!}\mathbb{E}[V^n] \tag{B.1.6}$$

$$= 1 + c_1 m_1 + \frac{c_1^2}{2!}m_2 + \frac{c_1^3}{3!}m_3 + \frac{c_1^4}{4!}\underbrace{m_4}_{\text{kurtosis}} + \underbrace{\sum_{n=5}^{\infty} \frac{c_1^n}{n!}m_n}_{\text{higher order moments}} \tag{B.1.7}$$

where m_i is the i-th order raw moment of the distribution of V. Assuming for simplicity that the distribution of V is symmetrical, one can further assume without loss of generality that it is centered around 0 and scaled to unit variance (any such transformation can be reversed by appropriately choosing c_1 and c_2). The first 3 raw moments of any admissible distribution for V are then uniquely determined: For the first moment, the mean, $m_1 = 0$ by definition. The second raw moment is identical to the second centered moment, as the mean is 0, and thus determined by the variance $m_2 = 1$. The third (as well as any other moment of odd order) must be zero due to the symmetry of the distribution of V. While different such distributions are thus indistinguishable by their first three moments, they differ in the fourth and higher order moments.

As can be seen in equations B.1.3 and B.1.7, the expected activation of the neuron can thus be understood as measuring the non-gaussianity or *tail-weight* of the membrane potential distribution, favoring non-gaussian distributions with larger probability mass in the tails. For *heavy-tailed* distributions of V, equation B.1.7 even diverges to ∞.

B.2   De-Mixing Laplace Inputs

If a random variable X_i follows a Laplace distribution with zero mean and unit variance, its moment generating function is given as:

$$M_{X_i}(t) = \frac{1}{1 - \frac{1}{2}t^2} \quad \text{for } |t| < \sqrt{2} \tag{B.2.1}$$

If $X = \sum_{i=0}^{N} \alpha_i X_i$ is a mixture of $N \in \mathbb{N}$ independent random variables X_i for $i \in \mathbb{N}$ with respective weights α_i, the moment generating function of X is given as the product $M_X(t) = \prod_{i=1}^{N} M_{X_i}(\alpha_i t)$. For a mixture X of N iid. Laplace random variables with zero mean and unit variance and weights $\alpha_i \leq 1$, this together with equation B.1.3 implies that the expected output of the neuron in response to the input X is given by

$$\mathbb{E}[Y] = \exp(c_2) M_X(c_1)$$
$$= \frac{\exp(c_2)}{\prod_{i=1}^{N} \left(1 - \frac{1}{2}\alpha_i^2 c_1^2\right)} \quad \text{for } |c_1| < \sqrt{2} \tag{B.2.2}$$

To develop a first intuition, two simple extreme cases can now be distinguished while fixing the (Euclidean) length of the weight vector $||\alpha||_2 = || \left(\alpha_1 \quad \dots \quad \alpha_N \right)^T ||_2 = \sqrt{\sum_{i=1}^{N} \alpha_i^2} = 1$.

First, consider that all weights $\alpha_{i \neq j} = 0$ except $\alpha_j = 1$ for some $1 \leq j \leq N$ [1]. In that case equation B.2.2 simplifies to:

$$\mathbb{E}[Y_1] = \frac{\exp(c_2)}{1 - \frac{1}{2}c_1^2} \tag{B.2.3}$$

For the other extreme, consider that all weights are identical, i.e. $\alpha_i := \frac{1}{\sqrt{N}}$ for

---

[1] This maximizes the sup-norm of α with $||\alpha||_{\sup} = 1$.

all $i \in \{1, \ldots, N\}$ [2]. In this case equation B.2.2 simplifies to:

$$\mathbb{E}[Y_N] = \frac{\exp(c_2)}{\left(1 - \frac{1}{2}\frac{c_1^2}{N}\right)^N} \tag{B.2.4}$$

The expected output of the neuron in equation B.2.4 monotonically decreases with N, starting with the right-hand side of equation B.2.3 for the special case $N = 1$. The monotonicity can be shown by using the *inequality of arithmetic and geometric means* to show that the denominator is monotonically increasing with N:

$$\sqrt[N+1]{\left(1 - \frac{c_1^2}{2N}\right)^N} = \sqrt[N+1]{\prod_{i=1}^{N}\left(1 - \frac{c_1^2}{2N}\right) \cdot 1} \tag{B.2.5}$$

$$\leq \frac{1}{N+1}\left(\sum_{i=1}^{N}\left(1 - \frac{c_1^2}{2N}\right) + 1\right) \tag{B.2.6}$$

$$= \frac{1}{N+1}\left((N+1)\left(1 - \frac{c_1^2}{2N}\right) + \frac{c_1^2}{2N}\right) \tag{B.2.7}$$

$$= 1 - \frac{c_1^2}{2N} + \frac{c_1^2}{2N(N+1)} \tag{B.2.8}$$

$$= 1 - \frac{(N+1)c_1^2 - c_1^2}{2N(N+1)} \tag{B.2.9}$$

$$= 1 - \frac{c_1^2}{2(N+1)} \tag{B.2.10}$$

If we consider the limit as $N \to \infty$, the mixture becomes Gaussian (as implied by the central limit theorem) and the denominator in equation B.2.4 converges to an exponential function:

$$\lim_{N\to\infty} \mathbb{E}[Y_N] = \frac{\exp(c_2)}{\lim_{N\to\infty}\left(1 - \frac{1}{2}\frac{c_1^2}{N}\right)^N} \tag{B.2.11}$$

$$= \exp\left(c_2 + \frac{1}{2}c_1^2\right) \tag{B.2.12}$$

---

[2] This minimizes the sup-norm of α with $||\alpha||_{\text{sup}} = \frac{1}{\sqrt{N}}$.

Informally speaking, this shows that an equal mixture of an increasing number of Laplace random variables becomes more gaussian, thus decreasing the neuron's output (the neuron acts as a measure of the non-gaussianity).

Now consider more generally a mixture of N Laplace random variables X_i with potentially different weights $\alpha_i \in \mathbb{R}$:

$$X = \sum_{i=1}^{N} \alpha_i X_i \tag{B.2.13}$$

$$\mathbb{E}[Y_N] = \frac{\exp(c_2)}{\prod_{i=1}^{N}\left(1 - \frac{1}{2}\alpha_i^2 c_1^2\right)} \tag{B.2.14}$$

$$= \exp\left(c_2 - \log\left(\prod_{i=1}^{N}\left(1 - \frac{1}{2}\alpha_i^2 c_1^2\right)\right)\right) \tag{B.2.15}$$

$$= \exp\left(c_2 - \sum_{i=1}^{N}\log\left(1 - \frac{1}{2}\alpha_i^2 c_1^2\right)\right) \tag{B.2.16}$$

In order to maximize the output of the neuron under the constraint that the 2-norm of the weights $||\alpha||_2 = 1$, Lagrange multipliers can be used. The constraint optimization problem to be solved can be expressed as

$$\text{minimize} \quad \mathbb{E}[Y_N] = \exp\left(c_2 - \sum_{i=1}^{N}\log\left(1 - \frac{1}{2}\alpha_i^2 c_1^2\right)\right)$$

$$\text{subject to} \quad ||\alpha||_2 = 1$$

Two auxiliary function $g(\alpha) := ||\alpha||_2 - 1$ and $f(\alpha) := \mathbb{E}[Y_N]$ can be introduced to simplify the notation. If a solution exists at all, then there must be a $\lambda \in \mathbb{R}$ such that the weight vector α is a solution of the above optimization problem if and only if it satisfies two simultaneous conditions:

$$0 \overset{!}{=} g(\alpha) \tag{B.2.17}$$

$$0 \overset{!}{=} \nabla f(\alpha) - \lambda \nabla g(\alpha) \tag{B.2.18}$$

Deriving f component-wise yields ∇f:

$$\frac{\mathrm{d}f(\alpha)}{\mathrm{d}\alpha_j} = \frac{\mathrm{d}\mathbb{E}[Y_N]}{\mathrm{d}\alpha_j} = \frac{\mathrm{d}}{\mathrm{d}\alpha_j}\exp\left(c_2 - \sum_{i=1}^{N}\log\left(1 - \frac{1}{2}\alpha_i^2 c_1^2\right)\right) \tag{B.2.19}$$

$$= f(\alpha)\frac{\mathrm{d}}{\mathrm{d}\alpha_j}\left(c_2 - \sum_{i=1}^{N}\log\left(1 - \frac{1}{2}\alpha_i^2 c_1^2\right)\right) \tag{B.2.20}$$

$$= f(\alpha)\left(\frac{\frac{\mathrm{d}}{\mathrm{d}\alpha_j}\left(1 - \frac{1}{2}\alpha_j^2 c_1^2\right)}{1 - \frac{1}{2}\alpha_j^2 c_1^2}\right) \tag{B.2.21}$$

$$= f(\alpha)\left(\frac{\alpha_j c_1^2}{1 - \frac{1}{2}\alpha_j^2 c_1^2}\right) \tag{B.2.22}$$

$$\frac{\mathrm{d}g(\alpha)}{\mathrm{d}\alpha_j} = \frac{\mathrm{d}}{\mathrm{d}\alpha_j}(\|\alpha\|_2 - 1) = \frac{\mathrm{d}}{\mathrm{d}\alpha_j}(\sqrt{\sum_{i=1}^{N}\alpha_i^2} - 1) \tag{B.2.23}$$

$$= \frac{\frac{\mathrm{d}}{\mathrm{d}\alpha_j}\sum_{i=1}^{N}\alpha_i^2}{2\sqrt{\sum_{i=1}^{N}\alpha_i^2}} = \frac{2\alpha_j}{2} = \alpha_j \tag{B.2.24}$$

Plugging equations B.2.22 and B.2.24 into condition B.2.18 yields N simultaneous conditions for $j \in \{1, \dots, N\}$:

$$0 \stackrel{!}{=} f(\alpha)\left(\frac{\alpha_j c_1^2}{1 - \frac{1}{2}\alpha_j^2 c_1^2}\right) - \lambda\alpha_j \tag{B.2.25}$$

$$\Leftrightarrow 0 = \alpha_j\left(f(\alpha)c_1^2 - \lambda\left(1 - \frac{1}{2}\alpha_j^2 c_1^2\right)\right) \tag{B.2.26}$$

$$\Leftrightarrow \alpha_j = 0 \vee \alpha_j^2 = \underbrace{\frac{2}{c_1^2} - 2\frac{f(\alpha)}{\lambda}}_{\text{const.}} =: \gamma(\alpha) \tag{B.2.27}$$

Since for each $j \in \{1, \dots, N\}$ the weight $\alpha_j \in \{0, -\sqrt{\gamma(\alpha)}, \sqrt{\gamma(\alpha)}\}$ where $\gamma(\alpha)$ is independent of j, condition B.2.17 implies, that there must be $M \geq 1$ non-zero weights α_{i_1} to α_{i_M}, such that the 2-norm of the vector α is constrained to 1. Since all non-zero components of α must be equal to $-\sqrt{\gamma(\alpha)}$ or $\sqrt{\gamma(\alpha)}$, it follows

that $||\alpha||_2^2 = \sum_{j=1}^{N} \alpha_j^2 = \sum_{j=1}^{M} \alpha_{i_j}^2 = M\gamma(\alpha) = 1 \Rightarrow \alpha_{i_j} = \pm\sqrt{\gamma(\alpha)} = \pm\frac{1}{\sqrt{M}}$ for $j \in \{1, \ldots, M\}$. The weight vector α that yields the largest expected output of the neuron thus has, without loss of generality, the form:

$$\alpha = \Big(\underbrace{\pm\tfrac{1}{\sqrt{M}} \quad \cdots \quad \pm\tfrac{1}{\sqrt{M}}}_{M\,\text{times}} \quad \underbrace{0 \quad \cdots \quad 0}_{N-M\,\text{times}} \Big)^T \qquad\qquad \text{(B.2.28)}$$

Since the Laplace random variables are symmetric around 0, the sign of the weights has no impact on the resulting expected output of the neuron, which is then given by equation B.2.4 when replacing N by M. This in turn implies, that the neuron's expected output becomes largest when $M = 1$, i.e. when all but one weights are 0.

In other words, maximizing the neuron's expected output by adapting the weight vector α corresponds to finding a weight vector such that the neuron's input resembles a single one of the Laplace source random variables. If the pre-synaptic inputs of the neuron are themselves already linear mixtures of such Laplace sources, the neuron's input is still a linear mixture of the Laplace sources and it's output thus becomes maximal when the weight vector is chosen such that it extracts one of the Laplace sources from the mixed pre-synaptic inputs.

Simulation Codes

The code used to generate the figures presented in chapter 4 is provided here in digital form on the attached SD-card.

All programs are written in the Python programming language using the SciPy library stack.[1]
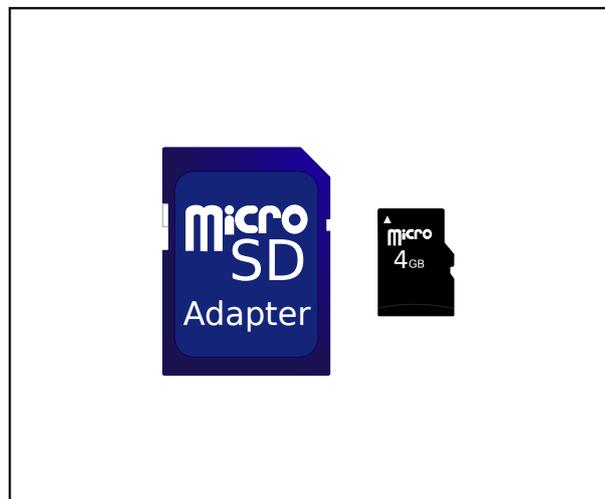
**Memory Disk containing
the code in digital format.**

---

[1] Oliphant 2007; Rossum 2012; Jones et al. 2007.

Bibliography

Angus, John E. (1994). "The Probability Integral Transform and Related Results". In: *SIAM Review* 36.4, pp. 652–654. DOI: 10.1137/1036146 (cit. on p. 12).

Ashby, William Ross (1954). *Design for a Brain*. 2nd ed. New York, USA: John Wiley & Sons, Inc., p. 286. DOI: 10.5962/bhl.title.6969 (cit. on p. 1).

Behseta, Sam, Robert E. Kass, David E. Moorman, and Carl R. Olson (2007). "Testing equality of several functions: Analysis of single-unit firing-rate curves across multiple experimental conditions". In: *Statistics in Medicine*. Vol. 26. 21, pp. 3958–3975. DOI: 10.1002/sim.2940 (cit. on p. 22).

Bell, A J and T J Sejnowski (1995). "An information-maximization approach to blind separation and blind deconvolution." In: *Neural computation* 7.6, pp. 1129–1159. DOI: 10.1162/neco.1995.7.6.1129 (cit. on p. 2).

Berry, M J and M Meister (1998). "Refractoriness and neural precision." In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 18.6, pp. 2200–2211 (cit. on pp. 19, 21).

Bottou, Léon (2010). "Large-Scale Machine Learning with Stochastic Gradient Descent". In: *Proceedings of COMPSTAT'2010*. Ed. by Yves Lechevallier and Gilbert Saporta. Heidelberg: Physica-Verlag HD, pp. 177–186. DOI: 10.1007/978-3-7908-2604-3_16 (cit. on p. 46).

Brown, Emery N, Riccardo Barbieri, Valérie Ventura, Robert E Kass, and Loren M Frank (2002). "The time-rescaling theorem and its application to neural spike train data analysis." In: *Neural computation* 14.2, pp. 325–346. DOI: 10.1162/08997660252741149 (cit. on p. 50).

Brown, Lawrence D. (1986). *Fundamentals of Statistical Exponential Families*. Ed. by Shanti S. Gupta. 9th ed. Vol. 9. Hayward, USA: Institude of Mathematical Statistics, pp. 1–279 (cit. on p. 28).

Buesing, Lars and Wolfgang Maass (2010). "A Spiking Neuron as Information Bottleneck". In: *Neural computation* 22.8, pp. 1961–1992. DOI: 10.1162/neco.2010.08-09-1084 (cit. on p. 2).

Burkitt, a. N. (2006a). "A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input". In: *Biological Cybernetics* 95.1, pp. 1–19. DOI: 10.1007/s00422-006-0068-6 (cit. on p. 25).

— (2006b). *A review of the integrate-and-fire neuron model: II. Inhomogeneous synaptic input and network properties*. DOI: 10.1007/s00422-006-0082-8 (cit. on p. 25).

Buzsáki, György and Kenji Mizuseki (2014). "The log-dynamic brain: how skewed distributions affect network operations." In: *Nature reviews. Neuroscience* 15.4, pp. 264–78. DOI: 10.1038/nrn3687 (cit. on p. 37).

Canu, Stéphane and Alex Smola (2006). "Kernel methods and the exponential family". In: *Neurocomputing* 69.7-9 SPEC. ISS. Pp. 714–720. DOI: 10.1016/j.neucom.2005.12.009 (cit. on pp. 27–29).

Charnes, a., E. L. Frome, and P. L. Yu (1976). "The Equivalence of Generalized Least Squares and Maximum Likelihood Estimates in the Exponential Family". In: *Journal of the American Statistical Association* 71.353, pp. 169–171. DOI: 10.1080/01621459.1976.10481508 (cit. on pp. 27, 42).

Chichilnisky, E J (2001). "A simple white noise analysis of neuronal light responses." In: *Network* 12.2, pp. 199–213. DOI: 10.1080/713663221 (cit. on p. 10).

Cover, Thomas M. and Joy A. Thomas (2005). *Elements of Information Theory*. Wiley Series in Telecommunications. New York, USA: John Wiley & Sons, Inc., pp. 1–748. DOI: 10.1002/047174882X. arXiv: ISBN0-471-06259-6 (cit. on pp. 33, 35).

Cunningham, John P, Byron M Yu, Krishna V Shenoy, and Maneesh Sahani (2008). "Inferring Neural Firing Rates from Spike Trains Using Gaussian Processes". In: *Advances in Neural Information Processing Systems 20*. Ed. by J C Platt, D Koller,

Y Singer, and S T Roweis. Curran Associates, Inc., pp. 329–336 (cit. on pp. 22, 23, 60).

DasGupta, B and Georg Schnitger (1993). "The power of approximating: a comparison of activation functions". In: *Advances in Neural Information Processing Systems 5*. Morgan Kaufmann, pp. 615–622 (cit. on pp. 9, 43).

Dayan, Peter and Laurence F. Abbott (2001). *Theoretical Neuroscience*. Cambridge: MIT Press (cit. on p. 41).

Dobson, Annette J (2002). *An introduction to generalized linear models*. Ed. by C. Chatfield and J. Zidek. Boca Raton, USA: Chapman and Hall/CRC, p. 225 (cit. on pp. 27, 28).

Embrechts, Paul and Marius Hofert (2013). "A note on generalized inverses". In: *Mathematical Methods of Operations Research* 77.3, pp. 423–432. DOI: 10.1007/s00186-013-0436-7 (cit. on pp. 10, 73).

Fechner, G. Th. (1860). *Elemente der Psychophysik*. 2nd ed. Leipzig: Breitkopf & Härtel (cit. on p. 43).

Gabbiani, Fabrizio, Holger G Krapp, Christof Koch, and Gilles Laurent (2002). "Multiplicative computation in a visual neuron sensitive to looming." In: *Nature* 420.6913, pp. 320–324. DOI: 10.1038/nature01190 (cit. on p. 43).

Gerstner, Wulfram (1995). "Time structure of the activity in neural network models". In: *Physical Review E* 51.1, pp. 738–758. DOI: 10.1103/PhysRevE.51.738 (cit. on pp. 22, 40).

Gerwinn, Sebastian, Jakob H Macke, and Matthias Bethge (2010). "Bayesian inference for generalized linear models for spiking neurons." In: *Frontiers in computational neuroscience* 4.May, p. 12. DOI: 10.3389/fncom.2010.00012 (cit. on p. 7).

Gestri, G., H. A. K. Mastebroek, and W. H. Zaagman (1980). "Stochastic constancy, variability and adaptation of spike generation: Performance of a giant neuron in the visual system of the fly". In: *Biological Cybernetics* 38.1, pp. 31–40. DOI: 10.1007/BF00337399 (cit. on pp. 18, 24).

Geyer, Cj (1990). "Likelihood and Exponential Families". Dissertation. University of Washington (cit. on pp. 27, 29).

Haslinger, Robert, Gordon Pipa, Bruss Lima, Wolf Singer, Emery N. Brown, and Sergio Neuenschwander (2012). "Context matters: The illusive simplicity of

macaque V1 receptive fields". In: *PLoS ONE* 7.7, pp. 1–17. DOI: 10.1371/journal.pone.0039699 (cit. on pp. 7, 50).

Hogg, Robert, Joeseph McKean, and Allen Craig (2005). *Introduction to Mathematical Statistics*. Ed. by George Lobell and Sally Yagan. 6th ed. Upper Saddle River: Pearson Education International (cit. on pp. 30, 31, 35).

Hromádka, Tomáš, Michael R. DeWeese, and Anthony M Zador (2008). "Sparse representation of sounds in the unanesthetized auditory cortex". In: *PLoS Biology* 6.1, pp. 124–137. DOI: 10.1371/journal.pbio.0060016 (cit. on p. 50).

Hyvärinen, A and E Oja (2000). "Independent component analysis: Algorithms and applications". In: *Neural Networks* 13.4-5, pp. 411–430. DOI: 10.1016/S0893-6080(00)00026-5 (cit. on pp. 45, 47).

Jones, Eric, Travis Oliphant, Pearu Pterson, and Others. *SciPy: Open Source Scientific Tools for Python* (cit. on pp. 49, 84).

Joshi, Prashant and Jochen Triesch (2009). "Rules for information maximization in spiking neurons using intrinsic plasticity". In: *Proceedings of the International Joint Conference on Neural Networks*, pp. 1456–1461. DOI: 10.1109/IJCNN.2009.5178625 (cit. on p. 2).

Karlik, Bekir and Av Olgac (2010). "Performance analysis of various activation functions in generalized MLP architectures of neural networks". In: *International Journal of Artificial Intelligence and Expert Systems (IJAE)* 1, pp. 111–122 (cit. on p. 9).

Kass, Robert E and Valérie Ventura (2001). "A spike-train probability model." In: *Neural computation* 13.8, pp. 1713–1720. DOI: 10.1162/08997660152469314 (cit. on pp. 7, 21).

Keat, Justin, Pamela Reinagel, R. Clay Reid, and Markus Meister (2001). "Predicting every spike: A model for the responses of visual neurons". In: *Neuron* 30.3, pp. 803–817. DOI: 10.1016/S0896-6273(01)00322-1 (cit. on p. 22).

Kistler, Werner M., Wulfram Gerstner, and J. Leo Van Hemmen (1997). "Reduction of the Hodgkin-Huxley Equations to a Single-Variable Threshold Model". In: *Neural Computation* 9.5, pp. 1015–1045. DOI: 10.1162/neco.1997.9.5.1015 (cit. on p. 22).

BIBLIOGRAPHY

Klampfl, Stefan, Robert Legenstein, and Wolfgang Maass (2009). "Spiking neurons can learn to solve information bottleneck problems and extract independent components." In: *Neural computation* 21.4, pp. 911–59. DOI: 10.1162/neco.2008.01-07-432 (cit. on p. 2).

Koch, C (1997). "Computation and the single neuron." In: *Nature* 385.6613, pp. 207–210. DOI: 10.1038/385207a0 (cit. on p. 43).

Koch, C and I Segev (2000). "The role of single neurons in information processing." In: *Nature neuroscience* 3, pp. 1171–1177. DOI: 10.1038/81444 (cit. on p. 43).

Koch, Christof and Tomaso Poggio (1992). "Multiplying with Synapses and Neurons". In: *Single Neuron Computation*. Ed. by Thomas McKenna, Joel Davis, and Steven F. Zornetzer. 1st ed. San Diego: Academic Press, Inc. Chap. 12, pp. 315–345 (cit. on p. 43).

Laughlin, Simon (1981). "A simple coding procedure enhances a neuron's information capacity." In: *Zeitschrift fur Naturforschung. Teil C: Biochemie, Biophysik, Biologie, Virologie* 36.9-10, pp. 910–912 (cit. on p. 3).

Lazar, Aurel a., Eftychios a. Pnevmatikakis, and Yiyin Zhou (2010). "Encoding natural scenes with neural circuits with random thresholds". In: *Vision Research* 50.22, pp. 2200–2212. DOI: 10.1016/j.visres.2010.03.015 (cit. on p. 18).

Limpert, Eckhard, Werner A. Stahel, and Markus Abbt (2001). "Log-normal Distributions across the Sciences: Keys and Clues". In: *BioScience* 51.5, pp. 341–352. DOI: 10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2 (cit. on pp. 37, 42).

McCullagh, Peter and John A Nelder (1989). *Generalized Linear Models*. Vol. 2. Chapman and Hall London (cit. on pp. 27, 28, 40–42).

McCulloch, Warren S. and Walter Pitts (1943). "A logical calculus of the ideas immanent in nervous activity". In: *The Bulletin of Mathematical Biophysics* 5.4, pp. 115–133. DOI: 10.1007/BF02478259 (cit. on pp. 9, 43).

Minsky, Marvin L. and Seymour A. Papert (1969). *Perceptrons*. MIT Press, p. 258 (cit. on p. 9).

Nielsen, Frank and Vincent Garcia (2009). "Statistical exponential families: A digest with flash cards". In: 2011, pp. 1–28. arXiv: 0911.4863 (cit. on pp. 27, 28, 30).

Oliphant, Travis E. (2007). "Python for scientific computing". In: *Computing in Science and Engineering* 9.3, pp. 10–20. DOI: 10.1109/MCSE.2007.58 (cit. on pp. 49, 84).

Ostojic, Srdjan and Nicolas Brunel (2011). "From spiking neuron models to linear-nonlinear models". In: *PLoS Computational Biology* 7.1. DOI: 10.1371/journal.pcbi.1001056 (cit. on pp. 5, 7).

Paninski, Liam (2004). "Maximum likelihood estimation of cascade point-process neural encoding models." In: *Network* 15.4, pp. 243–262. DOI: 10.1088/0954-898X/15/4/002 (cit. on pp. 7, 9, 43).

Reich, D S, J D Victor, B W Knight, T Ozaki, and E Kaplan (1997). "Response variability and timing precision of neuronal spike trains in vivo." In: *Journal of neurophysiology* 77.5, pp. 2836–2841 (cit. on pp. 18, 24).

Roberts, A. W. and D. E. Varberg (1974). "Another proof that convex functions are locally Lipschitz". In: *The American Mathematical Monthly* 81.9, pp. 1014–1016 (cit. on p. 34).

Rosenblatt, Frank (1958). "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6, pp. 386–408. DOI: 10.1037/h0042519 (cit. on p. 9).

Rossum, G. van (2012). *Python Programming Language and Interpreter* (cit. on pp. 49, 84).

Roxin, a., N. Brunel, D. Hansel, G. Mongillo, and C. van Vreeswijk (2011). "On the Distribution of Firing Rates in Networks of Cortical Neurons". In: *Journal of Neuroscience* 31.45, pp. 16217–16226. DOI: 10.1523/JNEUROSCI.1677-11.2011 (cit. on p. 38).

Rudolph, M and a Destexhe (2005). "An extended analytic expression for the membrane potential distribution of conductance-based synaptic noise." In: *Neural computation* 17.11, pp. 2301–2315. DOI: 10.1162/0899766054796932 (cit. on p. 35).

Savin, Cristina, Prashant Joshi, and Jochen Triesch (2010). "Independent component analysis in spiking neurons." In: *PLoS computational biology* 6.4. DOI: 10.1371/journal.pcbi.1000757 (cit. on pp. 2, 46).

Shimokawa, Takeaki and Shigeru Shinomoto (2009). "Estimating instantaneous irregularity of neuronal firing." In: *Neural computation* 21.7, pp. 1931–1951. DOI: 10.1162/neco.2009.08-08-841 (cit. on pp. 9, 34, 50).

Sopena, J.M., Enirque Romero, and Rene Alquezar (1999). "Neural networks with periodic and monotonic activation functions: a comparative study in classification problems". In: *9th International Conference on Artificial Neural Networks ICANN '99*, pp. 323–328. DOI: 10.1049/cp:19991129 (cit. on p. 9).

Stemmler, Martin and Christof Koch (1999). "How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate." In: *Nature neuroscience* 2.6, pp. 521–527. DOI: 10.1038/9173 (cit. on p. 2).

Tishby, Naftali, Fernando C. Pereira, and William Bialek (2000). "The information bottleneck method". In: *arXiv preprint physics*, pp. 1–11. DOI: 10.1108/eb040537. arXiv: 0004057 [physics] (cit. on p. 2).

Toyoizumi, Taro, Jean-pascal Pfister, Kazuyuki Aihara, and Wulfram Gerstner (2005). "Generalized Bienenstock-Cooper-Munro rule for spiking neurons that maximizes information transmission." In: *Proceedings of the National Academy of Sciences of the United States of America* 102.14, pp. 5239–5244. DOI: 10.1073/pnas.0500495102 (cit. on pp. 2, 9, 11).

Triesch, Jochen (2007). "Synergies between intrinsic and synaptic plasticity mechanisms." In: *Neural computation* 19.4, pp. 885–909. DOI: 10.1162/neco.2007.19.4.885 (cit. on pp. 2, 3, 37, 44, 69).

Truccolo, Wilson, Uri T Eden, Matthew R Fellows, John P Donoghue, and Emery N Brown (2005). "A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects." In: *Journal of neurophysiology* 93.2, pp. 1074–1089. DOI: 10.1152/jn.00697.2004 (cit. on pp. 7, 8, 50).

Turrigiano, Gina G. and Sacha B. Nelson (2000). "Hebb and homeostasis in neuronal plasticity". In: *Current Opinion in Neurobiology* 10.3, pp. 358–364. DOI: 10.1016/S0959-4388(00)00091-X (cit. on pp. 2, 45, 68).

BIBLIOGRAPHY

Wilson, Donald M (1964). "Relative Refractoriness and Patterned Discharge of Locust Flight Motor Neurons". In: *The Journal of Experimental biology* 41, pp. 191–205 (cit. on pp. 19, 20).