

BACHELOR

The Limit of Learning

A Mean-Field Approach to Hebbian Learning for Integrate-and-Fire Neuron Models

de Leeuw, Cor D.

Award date:
2023

[Link to publication](#)

Disclaimer

This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain



Department of Mathematics and Computer Science
Applied Analysis Group, CASA

The Limit of Learning

*A Mean-Field Approach to Hebbian
Learning for Integrate-and-Fire Neuron
Models*

C.D. de Leeuw

Supervisors:
dr. O.T.C. Tse
dr. J. Hoeksema

Bachelor's Thesis

Eindhoven, July 2023

Abstract

In this thesis, a model for learning neuronal populations in a spatial setting is constructed, based on a general integrate-and-fire framework. The focus is on deriving mean-field results that predict neuronal activity and synaptic weight changes in response to each other and external input. Next to this, synchronization of neuron potentials is also investigated. We find that this tends to occur in excitatory populations. We present a learning rule based on spike-timing-dependent plasticity (STDP) to model Hebbian learning. In the mean-field limit we show that such a learning rule can coincide with a rate-based learning rule. The derived mean-field equations predict, under appropriate assumptions, the relation between external (sensory) input, activity, and network structure in the form of a system of partial differential equations. Numerical simulations using the Julia language are used to illustrate concepts and verify some of the results. These verifications indicate that the mean-field approximation is valid for at least fully connected homogeneous inhibitory populations.

Contents

Contents	v
List of Figures	vii
1 Introduction	1
2 Neuron models	3
2.1 Biological background	3
2.2 Biophysical models	4
2.3 Integrate-and-fire models	5
2.3.1 Leaky integrate-and-fire	5
2.3.2 Exponential integrate-and-fire	6
2.3.3 Simulation	8
2.4 General integrate-and-fire model	8
2.4.1 Formulation	8
2.4.2 Local solution existence	9
2.4.3 Global solution existence without resets	11
2.4.4 Solution construction with resets	12
2.4.5 Repetitive firing	13
3 Populations of neurons	17
3.1 Formulation of the coupled system	17
3.2 Solution construction	19
3.3 Homogeneous populations with full connectivity	22
3.3.1 Synchronization	22
3.3.2 Mean-field limit	28
3.3.3 Stationary solution	31
3.4 Spatial populations	33
3.4.1 Formulation	33
3.4.2 Mean-field limit	34
3.4.3 Stationary solution	42
4 Learning populations	43
4.1 Hebbian learning	43
4.2 Learning rule formulation	44
4.3 Mean-field limit	45
4.4 Short-memory limit	49
5 Conclusions	53
5.1 Summary	53
5.2 Discussion	53
5.3 Future research	54
Bibliography	55
A Background material	57
B Supporting lemmas	59

List of Figures

2.1	Schematic of a neuron with its incoming dendrites, central soma, and outgoing axon. Adapted from [14] and licensed under © CC BY-SA 3.0.	4
2.2	Potential evolution of the leaky integrate-and-fire model. Nondimensionalized parameters as in Table 2.2, with $I_{\text{ext}} \equiv 0.65$	6
2.3	Potential evolution of the exponential integrate-and-fire model. Nondimensionalized parameters as in Table 2.1, with $I_{\text{ext}} \equiv 0.2$	7
3.1	Variance of the potentials and activity histogram over time for the leaky model, with a homogeneous population of $N = 1000$ neurons. The constant external input is $I_0 = 0.65$ and the neurons are fully connected with $w_0 = 0.1$. Initial potentials were chosen uniformly at random, on different intervals.	27
3.2	Variance of the potentials and activity histogram over time for the exponential model, with a homogeneous population of $N = 1000$ neurons. The constant external input is $I_0 = 0.2$ and the neurons are fully connected with $w_0 = 0.1$. Initial potentials were chosen uniformly at random on $[0, 0.85]$	28
3.3	Activity histograms over time for the exponential model, with a homogeneous population of $N = 1000$ neurons. The constant external input is $I_0 = 0.2$ and the neurons are fully connected with $w_0 = 0.1$. Initial potentials were chosen uniformly at random on $[0, 0.85]$	29
3.4	Simulation results for the leaky model with a homogeneous population of $N = 1000$ neurons. The constant external input is $I_0 = 0.8$ and the neurons are fully connected with $w_0 = -0.1$. Initial potentials were chosen uniformly at random on $[0, 0.85]$	32
3.5	Snapshots of the state of a homogeneous population of $N = 100$ leaky neurons, which are distributed uniformly at random over $\Omega = [0, 1]^2$. A color scaling visualizes the current potential of the neurons, and firing neurons are marked with larger red dots. The constant external input is $I_0 = 0.65$ and the neurons are connected with Gaussian weights according to (3.34), with $w_0 = 0.1$ and $\sigma = 0.2$. Initial potentials were chosen uniformly at random on $[0, 0.85]$	34

Chapter 1

Introduction

Learning is arguably the most valuable cognitive skill humans and other mammals possess. The ability to adapt behavior based on past experiences within an organism's lifespan, instead of over many generations due to natural selection, is a significant survival advantage.

But the importance of learning is not limited to the biological domain. The rampant rise of artificial intelligence over the past few months has shown that emulating biological learning is possible, at least to some degree.

The mechanics behind learning are analogous in both cases. In organisms, the organ responsible for learning is the brain. It consists of a system of interconnected cells, called neurons. Machine learning is accomplished with artificial neural networks, which are modeled on their biological counterparts.

The fact that a group of (artificial) cells, which are relatively simple in and of themselves, can collectively achieve something as complex as learning is fascinating. Given that learning is also foundational for humans and enables tremendous advancements in AI, studying the process by which neur(on)al¹ networks learn is of great interest.

This thesis focuses on biological neurons, which we will model using the integrate-and-fire model. The benefit of this model is that it approximates empirical observations well, whilst still being simple enough to enable theoretical analysis. After establishing this model for individual neurons, the objective will be to model a system, usually called a population, of interacting neurons.

It can however be difficult to analyze a population of neurons due to the large number of neurons involved: there are over 10,000 neurons per cubic millimeter of brain tissue [8, p. 3]. This is why we will employ a mean-field approximation, which will enable us to effectively average the interactions of a neuron with all other neurons.

To model the learning ability of a neuronal population, we will primarily consider Hebbian theory. The tenet of this theory is that a connection between two neurons is strengthened if these neurons are often active together. This associative principle is believed to be at the basis of both memory and learning, although reward-based learning due to certain neurotransmitters also plays a role.

We will derive mean-field partial differential equations that predict, under appropriate assumptions, the relation between external (sensory) input, activity, and network structure. Numerical simulations will be used to verify these results in a few simplified settings.

¹"Neuronal" generally refers to biological neurons, whereas "neural" pertains to the artificial variants.

Chapter 2

Neuron models

To eventually make the leap to interacting and learning populations of neurons, we first have to understand the behavior of individual neurons. Before diving into the mathematical models used to model neurons, this chapter will give a short and simplified overview of the biological foundation upon which these models are built, following Gerstner et al. [8, pp. 3–7].

Once this background has been established we will settle on a certain type of neuron model to use, namely the integrate-and-fire model. We close the chapter by deriving properties of a general version of this model.

2.1 Biological background

A neuron is a type of cell that enables the functioning of the nervous system. It typically consists of three main components: incoming dendrites, the soma, and the axon, illustrated in Figure 2.1. The dendrites transfer electrical signals coming from other neurons into the soma. The soma combines all of this input, and if it collectively causes the soma's electric (membrane) potential to exceed a certain threshold an output signal is emitted through the axon. The axon branches out to other neurons, propagating the signal. The point where the axon connects to the dendrite of another neuron is called a synapse. Schematically, this process is illustrated in Figure 2.1.

It turns out that these electrical signals between the neurons consist of very short¹ bursts of electric current that cross the synapses, which are called spikes. A neuron that emits such a spike is said to fire. In the context of a firing neuron and a neuron on the receiving end of its spike, we speak of presynaptic and postsynaptic neurons respectively.

Recapitulating, the basic mechanic is as follows: each neuron receives input from other neurons, and if this input sums up to be large enough the neuron sends out input itself. However, experiments with neurons have revealed two complicating aspects of neuronal behavior: refractoriness and adaptation. Both of these aspects demonstrate that neurons can “remember” their past stimulus exposure to some degree.

If a neuron has recently fired there is a period during which it can not fire again, even when it receives large input. This period is called the absolute refractory period. It is followed by a phase of relative refractoriness, during which excitation is difficult but not impossible.

Adaptation occurs when a neuron starts being exposed to (near) constant stimulus, strong enough to trigger spikes. Due to the aforementioned refractory period, the frequency of these spikes is bounded. One might expect the neuron to commence firing at a fixed frequency when the exposure starts, but this is not the case. Instead, most neurons start firing at a frequency that decays to a lower frequency over time. This can be interpreted as the neuron adapting to the stimulus, becoming less sensitive to it over time.

To explain this diverse behavior, various models of neuronal dynamics exist. These range from the more accurate and complex biophysical models to the simpler integrate-and-fire models. We will discuss these in the coming sections.

¹The relatively short duration of spikes will be important because it justifies the approximation of spikes as instantaneous events as done by integrate-and-fire models (see Section 2.3).

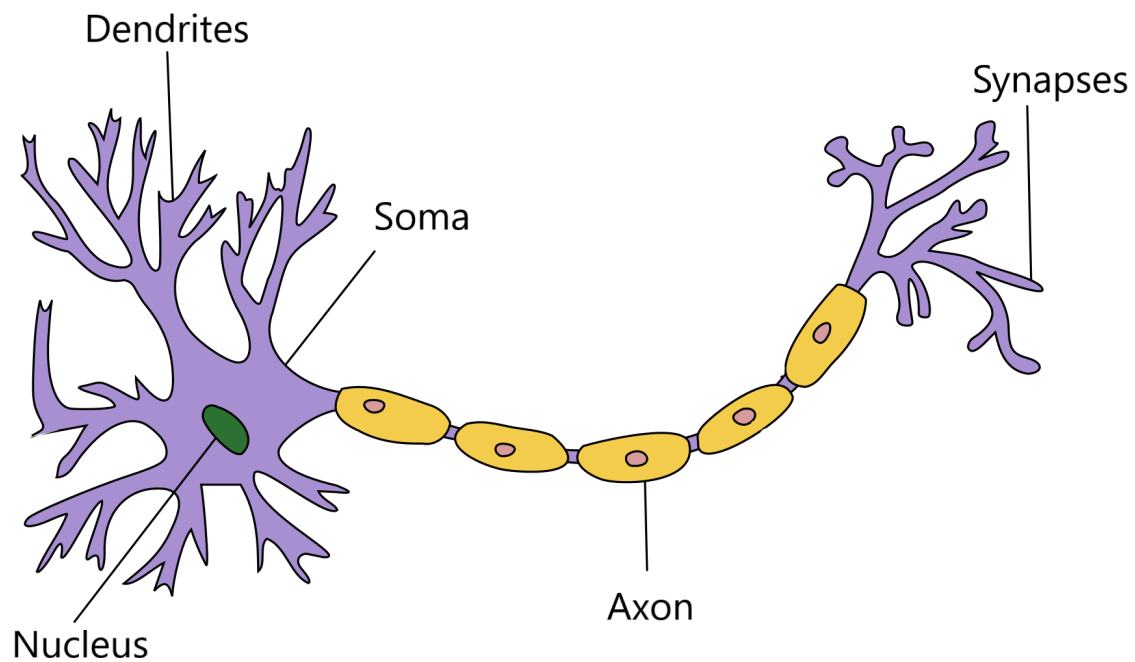


Figure 2.1: Schematic of a neuron with its incoming dendrites, central soma, and outgoing axon. Adapted from [14] and licensed under © CC BY-SA 3.0.

2.2 Biophysical models

Because neurons are ultimately cells, a large class of neuron models predicts spiking behavior by describing the biological processes that cause the build-up and discharge of potential in the soma. Generally, this is done by modeling various ion channels in the cell membrane, through which current can flow in and out of the cell. These channels open and close (partially) based on the current potential in the cell, which is modeled using differential equations in time. One of the oldest and most prominent models of this kind is the Hodgkin-Huxley model, which has 3 ion channels [8, ch. 2].

The variety in the behavior of these ion channels makes it possible to explain a large portion of observed neuronal behavior. For instance, the Hodgkin-Huxley model has a sodium ion channel that rapidly opens when the cell potential rises above some threshold, letting in even more current through the cell membrane. However, the same channel closes again when the potential is large, but this happens on a slightly longer timescale. The effect is a spike. After the spike the channel is closed off entirely, causing the post-spike potential to be decreased below the equilibrium (resting) potential. This causes a phase of (relative) refractoriness.

Similarly, by adding an ion channel that closes at an even slower speed when the potential is large, adaptive behavior can be modeled.² One can imagine that adding more ion channels with different properties can lead to behavior as intricate as desired. There are even spatial versions of these models that take into account the differences in potential along the dendrites and axons [8, ch. 3], which can be even more accurate.

However, the cost of this accuracy is a complicated system of coupled differential equations, which is hard to untangle and analyze analytically. This is why we will focus on a different, simpler kind of model in this thesis.

²Adaptation is, in essence, refractoriness on a longer timescale.

2.3 Integrate-and-fire models

Integrate-and-fire models are essentially simplified versions of biophysical models. The primary ingredient is still the time-dependent potential, denoted as $u(t)$, within the cell membrane, which evolves according to a differential equation. The main simplification is that spikes happen instantaneously once the potential exceeds a certain threshold: no time elapses during the transmission of current through the synapses and the closing of the ion channels to return the potential to the post-spike potential. This means that the mechanism that decreases the potential after a spike does not need to be incorporated into the model.

In this section, we introduce two prominent models of this type: the leaky integrate-and-fire model and the exponential integrate-and-fire model. We will use these two models as examples throughout this thesis.

2.3.1 Leaky integrate-and-fire

The simplest kind of integrate-and-fire model is the leaky integrate-and-fire model [8, sec. 1.3]. Its differential equation consists of a driving force that exponentially decays the current potential $u(t)$ back to some resting potential $V_{\text{rest}} \in \mathbb{R}$, and an arriving external input current $I_{\text{ext}}(t)$ containing for instance presynaptic spikes. The effect of this input current depends on the resistance of the cell membrane $R > 0$, and the speed of the entire process is controlled by a time constant $\tau > 0$.

The potential evolution thus satisfies the differential equation

$$\tau \frac{d}{dt} u = -(u - V_{\text{rest}}) + R I_{\text{ext}}(t). \quad (2.1)$$

This equation alone would suggest that u could grow arbitrarily large for large input I_{ext} , but as we know from Section 2.1 there should be a firing threshold $V_F > V_{\text{rest}}$. When u exceeds this threshold V_F the neuron fires, after which u is *immediately* reset to some reset potential $V_R < V_{\text{rest}}$. In combination with an initial condition $u(t_0) = u_0$, this completely describes the evolution of the potential u . A typical solution can be seen in Figure 2.2, where I_{ext} is constant.

The fact that the post-spike potential is reset to a value $V_R < V_{\text{rest}}$ causes a phase of relative refractoriness. This is because the required potential for causing a spike at that moment is $V_F - V_R$, which is larger than in the equilibrium state when it is $V_F - V_{\text{rest}}$.

Remark 2.1. Note that (2.1) may not be a classical differential equation. The input I_{ext} can contain presynaptic spikes from other neurons, which happen instantaneously (just like the reset from V_F to V_R). This means that $I_{\text{ext}}(t)$ can not be a continuous function, but instead contains infinitesimally short “pulses”. To represent this mathematically we will use Dirac measures, which we will formalize in Section 2.4.

Remark 2.2. Suppose that at time t either a presynaptic spike arrives or the neuron fires. In both cases the potential u changes immediately (discontinuously) at time t , but whether $u(t)$ is already affected by this jump is up to preference. We will adhere to the convention that u is a càdlàg function which means that the jump already affects $u(t)$, see Definition 2.1 below.

Definition 2.1. A function $h : J \rightarrow X$ from $J \subset \mathbb{R}$ to some metric space (X, d) is called a càdlàg³ function if it is right-continuous and has left-limits everywhere. That is, for all $t^* \in J$,

$$\lim_{t \downarrow t^*} h(t) = h(t^*), \quad \text{and} \quad \lim_{t \uparrow t^*} h(t) \text{ exists.}$$

Related to this definition, we introduce some shortened notation for left- and right-limits.

Notation 2.2. Left- and right-limits are denoted with superscript plus and minus signs on the function argument (when they exist). For example, for $f : I \rightarrow \mathbb{R}$ some function with $I \subset \mathbb{R}$ some open interval and $t \in I$ we write

$$f(t^+) := \lim_{\tau \downarrow t} f(\tau),$$

provided that this limit exists.

³French: “continue à droite, limite à gauche”

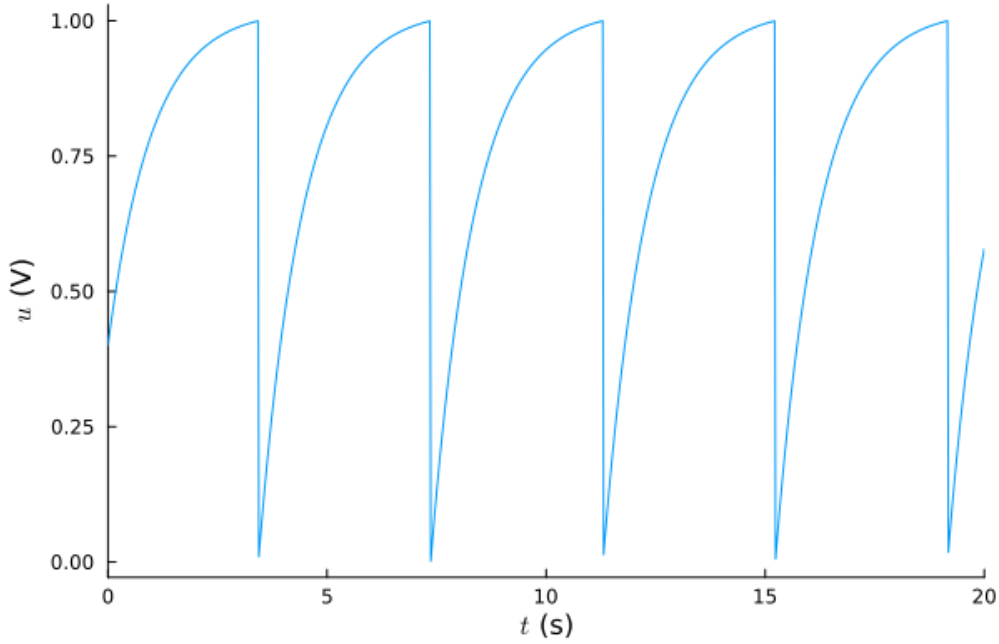


Figure 2.2: Potential evolution of the leaky integrate-and-fire model. Nondimensionalized parameters as in Table 2.2, with $I_{\text{ext}} \equiv 0.65$.

Nondimensionalization

To simplify the form of (2.1), we will nondimensionalize this equation to reduce the number of physical parameters. The involved dimensions are potential u [V], time t [s], current I_{ext} [A] and resistance R [$\Omega = VA^{-1}$]. We introduce the new parameters

$$\pi := \frac{t}{\tau}, \quad v := \frac{u - V_R}{V_F - V_R}, \quad \text{and } S(\pi) := \frac{RI_{\text{ext}}(\tau\pi)}{V_F - V_R},$$

which transforms the equation to

$$\frac{d}{d\pi}v = -(v - \frac{V_{\text{rest}} - V_R}{V_F - V_R}) + S(\pi),$$

with a reset from $v = \frac{V_F - V_R}{V_F - V_R} = 1$ to $v = \frac{V_R - V_R}{V_F - V_R} = 0$.

Relabeling the parameters and the constant $\frac{V_{\text{rest}} - V_R}{V_F - V_R}$ back to the old symbols (which is a slight abuse of notation) allows us to write,

$$\frac{d}{dt}u = -(u - V_{\text{rest}}) + I_{\text{ext}}(t), \quad (2.2)$$

with resets from 1 to 0 as the nondimensionalized leaky integrate-and-fire model.

2.3.2 Exponential integrate-and-fire

The exponential integrate-and-fire model is more detailed than the leaky variant [6]. In addition to the decay to the resting potential, this model also describes (part of) the blowup of the potential that defines a spike. This is modeled by an exponential term in the differential equation that can cause the potential to rapidly increase once it exceeds the so-called rheobase threshold $\vartheta_{\text{rh}} < V_F$, overpowering the decay to V_{rest} . The “sharpness” of this blowup is governed by a parameter $\Delta_T > 0$.

The differential equation corresponding to this is the following:

$$\tau \frac{d}{dt}u = -(u - V_{\text{rest}}) + \Delta_T \exp\left(\frac{u - \vartheta_{\text{rh}}}{\Delta_T}\right) + RI_{\text{ext}}(t). \quad (2.3)$$

with some initial condition $u(t_0) = u_0$ and a firing threshold $V_F \in \mathbb{R}$, after which a reset to $V_R < V_F$ takes place. Figure 2.3 shows a typical solution to this system, for constant external input. We used a

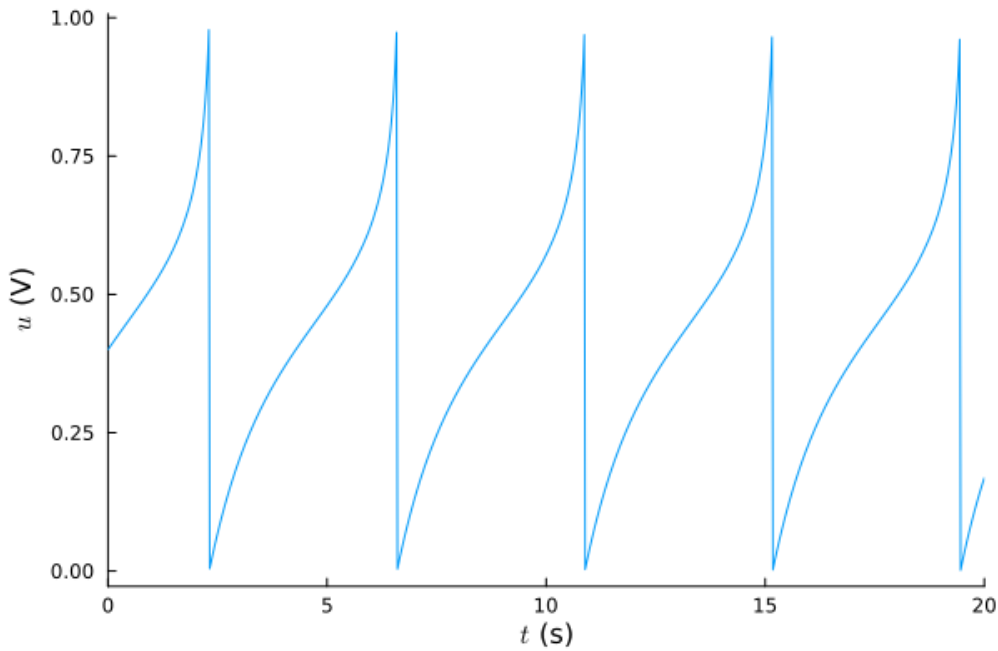


Figure 2.3: Potential evolution of the exponential integrate-and-fire model. Nondimensionalized parameters as in Table 2.1, with $I_{\text{ext}} \equiv 0.2$.

lower external input there than what was used for the leaky model to get approximately comparable firing frequencies.

Because the blowup caused by the exponential term is so rapid (as can be seen in Figure 2.3), the potential would go to infinity in finite time without the firing threshold. This means that the threshold V_F can be set equal to $+\infty$, or at least arbitrarily large, without significantly affecting the dynamics. For numerical purposes, this means that the exact value of V_F matters little when it is large enough.

This model, with fitted parameters, matches observed neuronal behavior much better than the leaky model [8, pp. 125–126]. The only facet of neuronal behavior that is not represented is adaptation. This could be added to increase the predictive performance of the model further [4], but because of the extra differential equations this requires we will not pursue such a model.

Remark 2.3. Often, an absolute refractory period of duration τ_{ref} is added to the exponential integrate-and-fire model. This means that when the neuron spikes, the integration restarts only after a time period of τ_{ref} . For simplicity, we opt not to add this to our version of the model. Relative refractoriness is still modeled because $V_R < V_{\text{rest}}$, as noted in Section 2.3.1.

Remark 2.4. When Δ_T is taken very close to zero, the behavior of the exponential integrate-and-fire model approaches a leaky integrate-and-fire model with a firing threshold of ϑ_{rh} . This is because, when Δ_T is small, the blowup happens very quickly after u exceeds ϑ_{rh} , causing V_F to be reached almost instantaneously.

Nondimensionalization

We will now nondimensionalize (2.3), just as we did for (2.1). Note that the argument of the exponential must be dimensionless, so $[\vartheta_{\text{rh}}] = [\Delta_T] = V$. A similar argument as given for the leaky model, where now also Δ_T and ϑ_{rh} are scaled by $V_F - V_R$ and the latter is shifted, yields⁴

$$\frac{d}{dt}u = -(u - V_{\text{rest}}) + \Delta_T \exp\left(\frac{u - \vartheta_{\text{rh}}}{\Delta_T}\right) + I_{\text{ext}}(t), \quad (2.4)$$

with resets from 1 to 0 as the nondimensionalized exponential integrate-and-fire model.

Remark 2.5. This transformation is only possible when $V_F < +\infty$. This is one of the reasons for keeping V_R and V_F in the general integrate-and-fire model we define in Section 2.4, instead of replacing them with 0 and 1 respectively.

⁴Here we again use the same symbols as for the dimensional parameters by abuse of notation.

2.3.3 Simulation

To illustrate the behavior of integrate-and-fire models, a simulation was created with the Julia programming language using the `DifferentialEquations.jl` library [12]. We used the default settings of this library, which automatically chooses an ODE solver based on the given problem. The source code can be accessed on GitHub ([link](#)). Throughout this thesis, we will use this simulation to visualize and verify some of our models and results, as done in Figure 2.2 and Figure 2.3.

To perform these simulations, a choice had to be made for the parameters in (2.1) and (2.3), and for V_R and V_F . Because this thesis focuses on studying the influence of the neuronal network structure and external input, all of these parameters, except for I_{ext} , were assigned fixed values.

To choose realistic values, we used parameters that were fitted to the biophysical Wang-Buzsáki model,⁵ as done in [6]. These can be seen in Table 2.1. The parameters τ and R are not in this table; as they pertain to the speed of the process and the external input respectively we set them equal to 1.

Parameter	Dimensional	Nondimensionalized
V_F	-50 mV^*	1
ϑ_{rh}	-59.9 mV	0.45
V_{rest}	-65 mV	0.17
V_R	-68 mV	0
Δ_T	3.48 mV	0.19

Table 2.1: Exponential integrate-and-fire parameters, based on the Wang-Buzsáki model.

* Fourcaud-Trocmé et al. [6] use $V_F = -30 \text{ mV}$, but this was lowered to improve the stability of the numerical integration.

We used the same reset and resting potential for the leaky model as for the exponential model. Based on Remark 2.4, we chose to take ϑ_{rh} from Table 2.1 as the firing threshold of this model. This results in Table 2.2.

Parameter	Dimensionalized	Nondimensionalized
V_F	-59.9 mV	1
V_{rest}	-65 mV	0.37
V_R	-68 mV	0

Table 2.2: Leaky integrate-and-fire parameters.

2.4 General integrate-and-fire model

In this section, we will formulate the integrate-and-fire model in full generality as an integral initial value problem, and prove results on the existence of solutions. At the end, we will also introduce the concept of repetitive firing and the gain function.

2.4.1 Formulation

The setting is as follows: the (membrane) potential $u : I \rightarrow U$ takes values in the potential space $U := (-\infty, V_F) \subset \mathbb{R}$ on some time interval $I \subset \mathbb{R}$, where $V_F \in \mathbb{R} \cup \{+\infty\}$ is the firing threshold. The evolution of u is governed by the following components:

1. A driving force that governs the potential in the absence of resets and pulse inputs, represented by a function $f : I \times U \rightarrow \mathbb{R}$. We will assume that f is continuous in its first argument (time) and either continuously differentiable or Lipschitz continuous in its second argument (potential). This is needed to ensure the existence of unique solutions to the differential equation. Note that f already contains any continuous external input $I_{\text{ext}}(t)$.
2. A set of arriving pulses independent of the potential, possibly from other neurons. We will represent these by $N \in \mathbb{N}$ arbitrary probability measures $\mu_1, \dots, \mu_N \in \mathcal{P}(\mathbb{R})$ on the time space with corresponding weights $\omega_1, \dots, \omega_N \in \mathbb{R}$. Note that in all applications these measures will simply be (sums of) Dirac measures.

⁵The Wang-Buzsáki model is a variant of the standard Hodgkin-Huxley model.

3. A reset to $V_R < V_F$ whenever the firing threshold V_F is reached. The set of firing times is implicitly defined as follows

$$\mathfrak{T} := \{\tau \in I \mid u(\tau^-) + \sum_{i=1}^N \omega_i \mu_i(\{\tau\}) \geq V_F\}, \quad (2.5)$$

where we need to use $u(\tau^-)$ and take into account spikes arriving at time τ separately since we formulate the system in a way that makes u càdlàg. Namely, if $\tau \in \mathfrak{T}$ then by construction $u(\tau) = V_R$. We will revisit this set \mathfrak{T} in Section 2.4.4.

4. An initial condition $u_0 \in U$ at time $t_0 \in I$.

All in all, we formulate the following formal and implicit system for u :

$$\begin{cases} du &= f(t, u) dt + \sum_{\tau \in \mathfrak{T}} (V_R - u(\tau^-) - \sum_{i=1}^N \omega_i \mu_i(\{\tau\})) \delta_\tau(dt) + \sum_{i=1}^N \omega_i \mu_i(dt), & t \in I, \\ u(t_0) &= u_0. \end{cases} \quad (2.6)$$

In the expression for du , the first term is the contribution of the driving force f , the second term causes the reset to V_R , and the third term contains the arriving pulses. The reason for the subtraction of the singleton measures in the second term is that pulses arriving at a firing time should be able to cause the firing (see (2.5)), but they should not affect the post-spike potential V_R .⁶ As they are not included in $u(\tau^-)$, they have to be subtracted separately.

We are going to make this system more precise by rewriting it as an integral initial value problem, but to do so we first introduce some notation.

Notation 2.3. The Lebesgue measure $\mathcal{L} : \bar{\mathcal{B}}_{\mathbb{R}^d}^{\mathcal{L}} \rightarrow [0, +\infty]$, where $\bar{\mathcal{B}}_{\mathbb{R}^d}^{\mathcal{L}}$ is the completion of the Borel σ -algebra w.r.t. the Lebesgue measure, will be denoted by dx instead of $\mathcal{L}(dx)$ when confusion is unlikely. Here x is the variable from the measured space \mathbb{R}^d , which is often the integration variable.

Notation 2.4. For a measure space $(\mathbb{R}, \mathcal{F}, \mu)$ with $\mathcal{F} \supseteq \mathcal{B}_{\mathbb{R}}$ (so that half-open intervals are measurable), an interval $J \subset \mathbb{R}$, a μ -integrable function $f : J \rightarrow \mathbb{R}$ and constants $t_0, t_1 \in J$ with $t_1 \geq t_0$, we denote

$$\int_{t_0}^{t_1} f d\mu := \int_{(t_0, t_1]} f d\mu,$$

and for $t_1 < t_0$ we define

$$\int_{t_0}^{t_1} f d\mu := - \int_{t_1}^{t_0} f d\mu.$$

This convention ensures that $\int_{t_0}^{t_0} f d\mu = 0$ and $\int_{t_0}^s f d\mu + \int_s^{t_1} f d\mu = \int_{t_0}^{t_1} f d\mu$. The choice to include the upper bound of the integration domain makes it so that functions of the type $t \mapsto \int_{t_0}^t f d\mu$ are càdlàg (as a consequence of the Dominated Convergence Theorem A.4).

With these conventions in mind, we reformulate (2.6) as

$$u(t) = u_0 + \int_{t_0}^t \left(f(\tau, u(\tau)) d\tau + \sum_{\sigma \in \mathfrak{T}} (V_R - u(\sigma^-) - \sum_{i=1}^N \omega_i \mu_i(\{\sigma\})) \delta_\sigma(d\tau) + \sum_{i=1}^N \omega_i \mu_i(d\tau) \right), \quad t \in I. \quad (2.7)$$

Observe that this equation is still implicit because \mathfrak{T} depends on u . Therefore, the existence of (unique) solutions is not obvious. We will address this in the next three sections: first, we will prove the existence of (1) local and (2) global càdlàg solutions if we disregard the reset terms, and (3) then we will show how the resets can be incorporated by “gluing” together solutions without resets.

2.4.2 Local solution existence

When looking for a local solution to (2.7), we can disregard the reset term by assuming that the arriving spikes locally do not elevate the potential to V_F . In this case the existence of a unique local càdlàg solution can be shown, which is the topic of the following theorem.

⁶Allowing this could let u exceed V_F .

Remark 2.6. In the theorem below, the choice to assume continuous differentiability instead of Lipschitz continuity of f in its second argument is made to accommodate the exponential integrate-and-fire model, whose driving force f is not globally Lipschitz continuous in case $V_F = +\infty$.

Theorem 2.1 (Picard-Lindelöf with measures). *Let $(X, \|\cdot\|)$ be a real finite-dimensional Banach space with $U \subset X$ and $I \subset \mathbb{R}$ both open and nonempty. Let $f \in C^{0,1}(I \times U, X)$, meaning that f is continuous and continuously differentiable in its second argument. Also let $\mu_1, \dots, \mu_N \in \mathcal{P}(\mathbb{R})$, $N \in \mathbb{N}$ be Borel probability measures with associated weights $\omega_1, \dots, \omega_N \in X$. Suppose that for $(t_0, u_0) \in I \times U$ there exists a closed interval $\tilde{J} \subset I$ containing t_0 in its interior for which it holds that $\bar{B}(u_0, \sum_{i=1}^N \|\omega_i\| \mu_i(\tilde{J}) + 1) \subset U$, then there exists an open interval $J \subset \tilde{J}$, $t_0 \in J$ such that the (integral) initial value problem*

$$u(t) = u_0 + \int_{t_0}^t \left(f(\tau, u(\tau)) d\tau + \sum_{i=1}^N \omega_i \mu_i(d\tau) \right), \quad t \in J, \quad (2.8)$$

has a unique solution $u \in D_b(J, X)$, the space of bounded càdlàg functions from J to X . Here (2.8) should be interpreted with Notation 2.4 in mind.

Proof. The proof is inspired by the proof of the Picard-Lindelöf Theorem A.1.

Define $K := \sum_{i=1}^N \|\omega_i\| \mu_i(\tilde{J}) + 1$. We start by defining the interval J , using the assumptions on the function f .

By assumption, f is continuously differentiable in its second argument. Denote this derivative by $D_2f : I \times U \rightarrow \text{BLin}(X)$, which is thus continuous. Here $\text{BLin}(X)$ is the space of bounded linear maps from X to X . Because $\tilde{J} \times \bar{B}(u_0, K)$ is closed and bounded, it is compact by finite-dimensionality of the Banach space X . Thus, we can define

$$L := \max_{(t,x) \in \tilde{J} \times \bar{B}(u_0, K)} \|D_2f(t, x)\|_{\text{BLin}(X)}. \quad (2.9)$$

Moreover, by continuity of f , there exists an $M > 0$ such that

$$\|f(t, x)\| \leq M \text{ for } (t, x) \in \tilde{J} \times \bar{B}(u_0, K). \quad (2.10)$$

Now choose $R > 0$ such that $MR < 1$, $LR < 1$ and $J := (t_0 - R, t_0 + R) \subset \tilde{J}$. On this interval J we will prove the existence of a unique solution to (2.8).

Define the set $D := \bar{B}_{D_b(J, X)}(\bar{u}_0, K)$, where $\bar{u}_0 : J \rightarrow X : t \mapsto u_0$ is a constant function. Here $(D_b(J, X), \|\cdot\|_\infty)$ is the Banach space of bounded càdlàg functions from J to X with supremum norm (this space is complete by Lemma B.2). The supremum norm is defined as

$$\|u\|_\infty := \sup_{t \in J} \|u(t)\|.$$

Also define the map $\mathcal{T} : D \rightarrow D_b(J, X) : u \mapsto \mathcal{T}u$ by

$$(\mathcal{T}u)(t) := u_0 + \int_{t_0}^t \left(f(\tau, u(\tau)) d\tau + \sum_{i=1}^N \omega_i \mu_i(d\tau) \right), \quad t \in J. \quad (2.11)$$

This map is well-defined, since for $u \in D$ we have $u(\tau) \in U \forall \tau \in J$ (because $\bar{B}(u_0, K) \subset U$). The resulting function $\mathcal{T}u : J \rightarrow X$ is furthermore bounded, because (2.10) combined with $t_0, t \in J \subset \tilde{J}$ guarantees boundedness of the first term of the integrand whereas the finiteness of the measures μ_1, \dots, μ_N does so for the second term. By Lemma B.1, $\mathcal{T}u$ is also càdlàg. Therefore \mathcal{T} indeed maps from D into $D_b(J, X)$.

Note that if $u^* \in D$ is a fixed point of \mathcal{T} , i.e., $\mathcal{T}u^* = u^*$, then u^* satisfies (2.8). We will now apply the Banach Fixed Point Theorem (Theorem A.2) to show that \mathcal{T} has a unique fixed point. We do this by checking the 3 conditions of this theorem:

1. $D_b(J, X)$ is complete by Lemma B.2, so as D is a closed subset it is complete as well.
2. Let $v \in D$. Then we have for all $t \in J$ that

$$\begin{aligned}
 \|(\mathcal{T}v)(t) - u_0\| &= \left\| \int_{t_0}^t \left(f(\tau, v(\tau)) d\tau + \sum_{i=1}^N \omega_i \mu_i(d\tau) \right) \right\|, \\
 &= \left\| \int_{t_0}^t f(\tau, v(\tau)) d\tau + \operatorname{sgn}(t - t_0) \sum_{i=1}^N \omega_i \mu_i((t, t_0] \cup (t_0, t]) \right\|, \\
 &\leq \int_{t_0}^t \|f(\tau, v(\tau))\| d\tau + \sum_{i=1}^N \|\omega_i\| \mu_i(\tilde{J}), \\
 &\stackrel{(2.10)}{\leq} MR + K - 1 < K.
 \end{aligned}$$

The final inequality holds by our choice of R . Therefore, $\|\mathcal{T}v - \bar{x}_0\|_\infty < K$, which implies that $\mathcal{T} : D \rightarrow D$.

3. Finally, for any $v, w \in D$ and $t \in J$, it follows from the mean-value inequality that

$$\begin{aligned}
 \|(\mathcal{T}v)(t) - (\mathcal{T}w)(t)\| &= \left\| \int_{t_0}^t (f(\tau, v(\tau)) - f(\tau, w(\tau))) d\tau \right\|, \\
 &\leq R \cdot \sup_{\tau \in J} \|f(\tau, v(\tau)) - f(\tau, w(\tau))\|, \\
 &\stackrel{(MVI)}{\leq} R \cdot \sup_{(\tau, s) \in J \times (0, 1)} \left\{ \|D_2 f(\tau, (1-s)v(\tau) + sw(\tau))\|_{\text{BLin}(X)} \cdot \|v(\tau) - w(\tau)\| \right\}, \\
 &\stackrel{(2.9)}{\leq} R \cdot L \|v - w\|_\infty.
 \end{aligned}$$

Since the above holds for all $t \in J$, also

$$\|\mathcal{T}v - \mathcal{T}w\|_\infty \leq \underbrace{LR}_{< 1} \|v - w\|_\infty,$$

which means that \mathcal{T} is a contraction.

From the Banach Fixed Point Theorem, we may now conclude that \mathcal{T} has a unique fixed point $u^* \in D$. This function u^* is a unique bounded càdlàg solution to (2.8) on J . \blacksquare

2.4.3 Global solution existence without resets

In this section, we will demonstrate that the unique local solution to (2.7) from the previous section is global under the additional assumption of Lipschitz continuity of f in its second argument, where we still disregard the resets. Because we can no longer ensure that u stays away from V_F , we have to set $V_F = +\infty$ (so $U = \mathbb{R}$) to prevent resets from being necessary.

Remark 2.7. Due to the assumption of Lipschitz continuity, this result does not apply to the exponential integrate-and-fire model. This makes sense, as under that model V_F is reachable in finite time which prevents globality of the solution.

For the proof of Theorem 2.2, we first introduce the exponential norm for bounded functions.

Definition 2.5. Let $(X, \|\cdot\|)$ be a Banach space and consider the space of bounded functions $B([t_0, s], X)$ (where $t_0, s \in \mathbb{R}$ with $t_0 < s$). Then the exponential norm $\|\cdot\|_\lambda : B([t_0, s], X) \rightarrow \mathbb{R}_+$ with parameter $\lambda > 0$ is defined as

$$\|u\|_\lambda := \sup_{t \in [t_0, s]} e^{-\lambda(t-t_0)} \|u(t)\|.$$

Theorem 2.2. *The setting is mostly the same as in Theorem 2.1. Instead of assuming that f is continuously differentiable in its second argument we now assume that it is Lipschitz continuous in its second argument, with a Lipschitz constant $L(t) \geq 0$ that depends continuously on t . Additionally, assume that $I = \mathbb{R}$ and $U = X$ (the latter means that the requirement on $\sum_{i=1}^N \|\omega_i\| \mu_i$ we had in Theorem 2.1 is always satisfied). Then (2.8) has a unique solution in $D_b([t_0, +\infty), X)$.*

Proof. The proof is analogous to the proof of Theorem 2.1. This time we take $D := D_b([t_0, s], X)$ for some fixed $s > t_0$, and again define the map $\mathcal{T} : D \rightarrow D$ in the same way as in (2.11).

To apply the Banach Fixed Point Theorem, we use the exponential norm defined in Definition 2.5 with

$$\lambda := \sup_{t \in [t_0, s]} L(t).$$

Thus, we show the existence of a unique solution in $(D, \|\cdot\|_\lambda)$.

The first condition then holds in the same way as in (1) because of Lemma B.3, and (2) holds trivially this time due to $U = X$. For the third and final condition, consider any $v, w \in D$. Then

$$\begin{aligned} \|\mathcal{T}v - \mathcal{T}w\|_\lambda &= \sup_{t \in [t_0, s]} e^{-\lambda(t-t_0)} \|(\mathcal{T}v)(t) - (\mathcal{T}w)(t)\|, \\ &= \sup_{t \in [t_0, s]} e^{-\lambda(t-t_0)} \left\| \int_{t_0}^t (f(\tau, v(\tau)) - f(\tau, w(\tau))) d\tau \right\|, \\ &\leq \sup_{t \in [t_0, s]} e^{-\lambda(t-t_0)} \int_{t_0}^t L(\tau) \|v(\tau) - w(\tau)\| d\tau, \\ &= \sup_{t \in [t_0, s]} \int_{t_0}^t L(\tau) e^{-\lambda(t-\tau)} e^{-\lambda(\tau-t_0)} \|v(\tau) - w(\tau)\| d\tau, \\ &\leq \lambda \sup_{t \in [t_0, s]} \int_{t_0}^t e^{-\lambda(t-\tau)} \|v - w\|_\lambda d\tau, \\ &= \lambda \sup_{t \in [t_0, s]} \frac{1}{\lambda} (1 - e^{-\lambda(t-t_0)}) \|v - w\|_\lambda, \\ &= \underbrace{(1 - e^{-\lambda(s-t_0)})}_{<1} \|v - w\|_\lambda. \end{aligned} \tag{2.12}$$

This shows that \mathcal{T} is a contraction on D , so just as in Theorem 2.1 we conclude that it has a unique fixed point $u^* \in D$. This function u^* is a unique bounded càdlàg solution to (2.8) on $[t_0, s]$. Because $s > t_0$ is arbitrary, it follows that there must exist a unique solution in $D_b([t_0, +\infty), X)$. ■

Remark 2.8. Theorem 2.2 states that (sub)linear growth of f implies global solution existence, which can also be proved using Gronwall's inequality.

2.4.4 Solution construction with resets

Having established solution existence without resets, we now derive an explicit solution of the implicit system (2.7). To do this, we start from the following preliminary system without resets:

$$\begin{cases} d\tilde{u} &= f(t, \tilde{u}) dt + \sum_{i=1}^N \omega_i \mu_i(dt), & t \in I, \tilde{u} \in U, \\ \tilde{u}(\tilde{t}_0) &= \tilde{u}_0. \end{cases} \tag{2.13}$$

By Theorem 2.1 this system has a unique càdlàg solution on some open⁷ interval $J(\tilde{t}_0, \tilde{u}_0) \subset I$, which we will denote by $\tilde{u}(\cdot, \tilde{t}_0, \tilde{u}_0)$. Note that we explicitly incorporate the initial conditions $\tilde{t}_0 \in I$ and $\tilde{u}_0 \in U$. This will be important since we are going to construct the solution to (2.7) as a sum of instances of this function with different initial conditions.

We start by constructing the set of firing times $\mathfrak{T} \subset I$ of the neuron. Define the potential evolution before the first (possible) spike as

$$\tilde{u}^{(1)}(t) := \tilde{u}(t, t_0, u_0), \quad t \in J(t_0, u_0),$$

which will be the first part of the potential trajectory. In accordance with (2.5), we define the first firing time as

$$r_1 := \inf\{\tau \in J(t_0, u_0) \mid \tau > t_0 \wedge \tilde{u}^{(1)}(\tau^-) + \sum_{i=1}^N \omega_i \mu_i(\{\tau\}) \geq V_F\}.$$

⁷Or half-open in case $\tilde{t}_0 \in \partial I$

Often r_1 will coincide with $\sup J(t_0, u_0)$, because f is not defined at or above V_F .⁸ We only consider $\tau > t_0$ because of Remark 2.9 below. Note that when the above subset of $J(t_0, u_0)$ is empty, we have $r_1 = +\infty$. In that case the neuron never fires, and also no resets take place. We henceforth assume that there is at least one spike, so $r_1 \in J(t_0, u_0) \subset I$.

Recursively, we then define for $k \in \mathbb{N}$,

$$\tilde{u}^{(k+1)}(t) := \tilde{u}(t, r_k, V_R), \quad t \in J(r_k, V_R),$$

and

$$r_{k+1} := \inf\{\tau \in J(r_k, V_R) \mid \tau > r_k \wedge \tilde{u}^{(k+1)}(\tau^-) + \sum_{i=1}^N \omega_i \mu_i(\{\tau\}) \geq V_F\}. \quad (2.14)$$

In this way, we have constructed all firing times r_k . Thus

$$\mathfrak{T} = \{r_k \mid k \in \mathbb{N}\} \setminus \{+\infty\}$$

is the set of all firing times of the neuron. Here we exclude $+\infty$ for the technical reason that $r_k = +\infty$ when the infimum in (2.14) is empty.

Finally, we can explicitly express the potential evolution u for $t > t_0$ by gluing together all interspike solutions $u^{(k)}$. Namely,

$$u(t) := \sum_{k \in \mathbb{N}} \tilde{u}^{(k)}(t) \mathbb{1}_{[r_{k-1}, r_k)}(t), \quad t \in I \cap [t_0, \sup_{k \in \mathbb{N}} r_k),$$

where by convention we set $r_0 := t_0$. By construction, this solution satisfies (2.7).

We intersect I with $[t_0, \sup_{k \in \mathbb{N}} r_k)$ here because we could theoretically have $\sup_{k \in \mathbb{N}} r_k < \sup I$. However, we will see in Lemma 3.1 that this will not happen for $V_F < +\infty$, making this intersection unnecessary.

Remark 2.9. The construction above is only (uniquely) possible in the forward time-direction. Namely, if one were to traverse the potential trajectory backwards, then upon approaching V_R from above it would in general be impossible to know if V_R was reached from below or due to a reset from V_F . Therefore it only makes sense to consider (2.7) on intervals of the form $I = [t_0, +\infty)$.

2.4.5 Repetitive firing

Now that we have constructed solutions to the general integrate-and-fire model, it is time to explore some of their properties. Namely, we will investigate the repetitive firing of a neuron when exposed to constant input, which can be observed in Figure 2.2 and Figure 2.3. In particular, the frequency of this firing will be of interest.

Throughout this section we make the simplifying assumptions that there is no time-dependent input, meaning that driving force is homogeneous ($f(t, \cdot) = f(\cdot)$), and that no measures (external spikes) contribute ($\omega_i = 0$).

Lemma 2.3. *Assume the driving force is bounded below by some constant $c > 0$ on $[V_R, V_F)$. Then the neuron will enter a state of periodic firing with some period T for any initial condition in $[V_R, V_F)$. Furthermore, there exists a strictly increasing (so order-preserving) diffeomorphism $\nu : [0, T) \rightarrow [V_R, V_F)$ satisfying $\frac{d\nu}{dt} = f(\nu)$. When given a time since the last spike this function yields the current potential of the neuron.*

Proof. We define $\nu(t) := \tilde{u}(t, 0, V_R)$ for $t \in J(0, V_R)$ (recall (2.13)). Note that f is bounded below by $c > 0$ on $[V_R, V_F)$, so ν is strictly increasing. The fact that ν reaches V_F follows by the Mean-Value Theorem. Namely, for all $t \in J(0, V_R)$ with $t > 0$ we have

$$\nu(t) - V_R = \nu(t) - \nu(t_0) = f(\nu(s))(t - t_0) \geq c(t - t_0)$$

for some $s \in (0, t)$. As $c > 0$, this implies that for $t \geq \frac{V_F - V_R}{c}$ we would have $\nu(t) \geq V_F$ if $t \in J(0, V_R)$. Since ν 's range is $U = (-\infty, V_F)$, it follows that ν must have reached V_F from below at a time in the interval $(0, \frac{V_F - V_R}{c})$. It follows that we can define

$$T := \inf\{t \in J(0, V_R) \mid \nu(t^-) = V_F\} = \sup J(0, V_R).$$

Because ν is continuous and strictly increasing on $[0, T)$ with $\nu(0) = V_R$, it is a bijection between $[0, T)$ and $[V_R, V_F)$. \blacksquare

⁸To be precise, we have $\sup J(t_0, u_0) \leq r_1$. This inequality is strict if and only if \tilde{u} diverges to $-\infty$ in finite time.

Example 2.1. For the (nondimensionalized) leaky integrate-and-fire model under constant input $I_{\text{ext}} \equiv I_0$ it is easy to compute the function ν . Namely, we can analytically solve (2.2) with $I_{\text{ext}}(t) = I_0$ and $u(0) = V_R = 0$ to obtain

$$\nu(\tau) = \tilde{u}(\tau, 0, V_R) = V_{\text{rest}} + I_0 - (V_{\text{rest}} + I_0)e^{-\tau},$$

which will reach $V_F = 1$ provided $I_0 > V_F - V_{\text{rest}} = 1 - V_{\text{rest}}$.

The lemma above states that for sufficiently strong external stimulus (which ensures that f that is bounded below) the neuron will continuously fire with some fixed frequency. A natural question would be what this frequency is for a given input, which is precisely the purpose of the so-called gain function.

Definition 2.6 (Gain function). Suppose the driving force is of the form $f(t, u) = f(u) + I_0$, where $I_0 \in \mathbb{R}$ is a constant external input current. The gain function $g : \mathbb{R} \rightarrow \mathbb{R}_+$ relates this input current to the firing rate it elicits in a single neuron for an arbitrary initial condition in $[V_R, V_F)$. Possible one-time spikes due to the initial condition are ignored here. Thus, the firing frequency of a neuron exposed to a constant current I_0 is given by $g(I_0)$.

Gain functions are often studied experimentally by exposing an *in vitro* neuron to a fixed current. The empirical gain function this yields can be used to fit the parameters of a neuron model. To this end, it is useful that we can derive an analytical expression for the gain function of a neuron model, as done in the lemma below.

Remark 2.10. Gain functions can be classified into two types, namely continuous and discontinuous ones. This distinction is based on whether repetitive firing picks up gradually as the external current increases, or whether there is a jump from zero to a nonzero firing rate [for details see 8, sec. 4.4].

Lemma 2.4. The gain function $g : \mathbb{R} \rightarrow \mathbb{R}_+$, as defined in Definition 2.6, is given by

$$g(I_0) = \begin{cases} 0 & \text{if } I_0 \leq -\inf_{u \in [V_R, V_F)} f(u), \\ \left(\int_{V_R}^{V_F} \frac{1}{f(u) + I_0} du \right)^{-1} & \text{if } I_0 > -\inf_{u \in [V_R, V_F)} f(u). \end{cases} \quad (2.15)$$

Proof. First, suppose $I_0 \leq -\inf_{u \in [V_R, V_F)} f(u)$. Then we have $\inf_{u \in [V_R, V_F)} f(u) + I_0 \leq 0$. The potential of a neuron evolves according to $\frac{du}{dt} = f(u) + I_0$, so as f is continuous the uniqueness of the solution implies that $\tilde{u}(\cdot, 0, V_R)$ never attains the value V_F . Depending on u_0 there could still be one spike, but no repetitive spikes take place. This means that $g(I_0) = 0$.

Now consider the case $I_0 > -\inf_{u \in [V_R, V_F)} f(u)$. Then $\frac{du}{dt} = f(u) + I_0$ is bounded below on $[V_R, V_F)$ by a positive constant, so we are in the realm of Lemma 2.3. This lemma implies that periodic repetitive firing takes place. Hence $g(I_0) = \frac{1}{T} > 0$, where T is the period (as defined in the lemma).

We will now use the differentiable bijection $\nu : [0, T) \rightarrow [V_R, V_F)$ to determine T . Its inverse $\nu^{-1} : [V_R, V_F) \rightarrow [0, T)$; $u \mapsto \nu^{-1}(u)$ exists, and is differentiable with

$$\frac{d\nu^{-1}}{du}(u) = \left(\frac{d\nu}{dt}(\nu^{-1}(u)) \right)^{-1} = (f(\nu(\nu^{-1}(u))) + I_0)^{-1} = \frac{1}{f(u) + I_0}.$$

It now follows that

$$T = \nu^{-1}(V_F) - \nu^{-1}(V_R) = \int_{V_R}^{V_F} \frac{d\nu^{-1}}{du}(u) du = \int_{V_R}^{V_F} \frac{1}{f(u) + I_0} du.$$

Thus we conclude that

$$g(I_0) = \frac{1}{T} = \left(\int_{V_R}^{V_F} \frac{1}{f(u) + I_0} du \right)^{-1}. \quad \blacksquare$$

Example 2.2. We again consider the leaky model, for which the computations are easy. For this model we have $f(u) = -(u - V_{\text{rest}})$ on U , so clearly $-\inf_{u \in [V_R, V_F]} f(u) = V_F - V_{\text{rest}}$. For $I_0 > V_F - V_{\text{rest}}$ we find

$$\left(\int_{V_R}^{V_F} \frac{1}{f(u) + I_0} du \right)^{-1} = \left(\int_{V_R}^{V_F} \frac{1}{V_{\text{rest}} + I_0 - u} du \right)^{-1} = \left([-\ln(V_{\text{rest}} + I_0 - u)]_{u=V_R}^{u=V_F} \right)^{-1} = 1/\ln \left(\frac{V_{\text{rest}} + I_0 - V_R}{V_{\text{rest}} + I_0 - V_F} \right),$$

and thus

$$g(I_0) = \begin{cases} 0 & \text{if } I_0 \leq V_F - V_{\text{rest}}, \\ 1/\ln \left(\frac{V_{\text{rest}} + I_0 - V_R}{V_{\text{rest}} + I_0 - V_F} \right) & \text{if } I_0 > V_F - V_{\text{rest}}. \end{cases}$$

For the nondimensionalized model, this simplifies to

$$g(I_0) = \begin{cases} 0 & \text{if } I_0 \leq 1 - V_{\text{rest}}, \\ 1/\ln \left(\frac{V_{\text{rest}} + I_0}{V_{\text{rest}} + I_0 - 1} \right) & \text{if } I_0 > 1 - V_{\text{rest}}. \end{cases}$$

If we substitute the standard parameter $V_{\text{rest}} = 0.37$ and $I_0 = 0.65$ into this we find a predicted firing frequency of

$$g(0.65) = 1/\ln \left(\frac{0.37 + 0.65}{0.37 + 0.65 - 1} \right) \approx 0.25 \text{ Hz},$$

which corresponds approximately with the frequency observed in Figure 2.2.

In Section 3.3.3 we will see that there is an interesting link between the gain function of an individual neuron and the stationary activity of a population of neurons. These populations will be the topic of the next chapter.

Chapter 3

Populations of neurons

In this chapter, we will make the jump from individual neurons to populations of $N \in \mathbb{N}$ interacting neurons. With a population we mean a group of similar neurons that interact (predominantly) with each other. This notion is based on the existence of structures in the cortex, such as columns, layers, and neuronal assemblies, that consist of neurons with similar properties [8, pp. 293–297]. Although the model will be more general, for all applications and examples we will restrict to homogeneous populations. In such a population all neurons are identical in terms of parameters, but not necessarily in terms of external input.

As explained in Section 2.1, the interaction between the neurons of a population takes place through the synapses between them. Whenever a presynaptic neuron fires, the spike travels through the axon and traverses the outgoing synapses to postsynaptic neurons, causing a change in the latter’s membrane potentials. This change can be both an increase or a decrease in the potential, which determines whether we call the corresponding synapse excitatory or inhibitory respectively.

In Section 3.1 we will incorporate these interaction mechanics rigorously into the single-neuron model formulated in Section 2.4.1, after which we will construct a solution analogous to Section 2.4.4. Then we will explore two cases of this model. First, a simplified version where the population is homogeneous with the same external input and full connectivity (all weights between the neurons are the same), for which we will study possible synchronization of the neurons. Afterwards, we consider a more general version of the model with a spatial component, in which the external input and weights depend on the (relative) positions of the neurons. For both of these models, we are ultimately interested in the behavior of the network when N grows large. We will approximate this behavior by taking the so-called mean-field limit, and explore stationary solutions of the resulting mean-field equations.

3.1 Formulation of the coupled system

The potentials of the different neurons at time $t \in I$ will be denoted by $u_1(t), \dots, u_N(t)$. Here $I \subset \mathbb{R}$ is still the time interval on which we consider the system. Because of Remark 2.9 we restrict to time intervals of the form $I = [t_0, t_e]$ or $I = [t_0, +\infty)$.

The potential u_i of each neuron $i \in \{1, \dots, N\}$ evolves according to the general integrate-and-fire model (2.3). The driving force $f_i : I \times U \rightarrow \mathbb{R}$ is now neuron-dependent because different neurons might have different parameters and/or external input.

The external pulses ω_k in (2.3) are replaced with Dirac measures located at the firing times of the other neurons. These model the spike arrivals from other neurons in the population. To describe whether and to what degree a synapse is excitatory or inhibitory, we introduce the synaptic weights $w_{ij} \in \mathbb{R}$ from presynaptic neuron j and postsynaptic neuron i .¹ For instance, if $w_{ij} = 1$ then the potential of neuron i is increased by 1 if neuron j fires. If $w_{ij} = 0$ there is no synapse from neuron j to neuron i . A neuron is not connected to itself, so the values w_{ii} are irrelevant.

Remark 3.1. We could have also preserved separate external pulses ω_k , which could model spikes from neurons of a different population, but for simplicity we choose not to do so. This means that all external input is contained in the driving forces f_i , and thus assumed to be continuous.

¹This ordering of the indices might seem unnatural, but this order is conventional.

Remark 3.2. According to Dale’s principle, a neuron is either excitatory or inhibitory. This would mean that for a fixed presynaptic neuron j the weights $\{w_{ij}\}_{i \in \{1, \dots, N\}}$ should all be of the same sign (with zero allowed). Thus, our model is usually more general than required, although there are neurons for which Dale’s principle does not hold [11, p. 10].

Based on (2.6) and the fact that values w_{ii} should be ignored, an obvious formulation of the system would be

$$\begin{cases} du_i &= f_i(t, u_i) dt + \sum_{\tau \in \mathfrak{T}_i} (V_R - u_i(\tau^-)) \delta_\tau(dt) + \sum_{j=1}^N \sum_{\tau \in \mathfrak{T}_j \setminus \mathfrak{T}_i} w_{ij} \delta_\tau(dt), & i \in \{1, \dots, N\}, t \in I, \\ u_i(t_0) &= v_i, & i \in \{1, \dots, N\}, \end{cases} \quad (3.1)$$

where \mathfrak{T}_j is the set of firing times for neuron j defined in (2.5), and $v_i \in U$ are initial potentials for all neurons at t_0 . Here the third term’s inner summation excludes \mathfrak{T}_i , which is equivalent to the subtraction of the sum of singleton measures in (2.6).

However, this exclusion of \mathfrak{T}_i has an undesirable property now that the pulses are no longer external. When the spike of a neuron j at time $\tau \in \mathfrak{T}_j$ causes the potential of another neuron i to also (immediately) exceed V_F , the resulting spike of neuron i does not affect the potential of neuron j in any way. Namely, we will simply have $u_j(\tau) = u_i(\tau) = V_R$. To see why this is problematic from a modeling perspective, we have to go back to the premise of our model.

The central assumption in the general integrate-and-fire model is that spikes occur and traverse synapses instantaneously. However, in reality or in biophysical models, these events take place over a short but non-infinitesimal time window. Consider again a pair of mutually connected neurons i and j , with the synapse from j to i being excitatory (so $w_{ij} > 0$). Neuron j fires and its spike subsequently causes the postsynaptic neuron i to fire as well. Then neuron j ’s potential has already been reset, or is in the process of resetting, when neuron i ’s spike arrives at neuron j . This means that neuron j ’s potential is affected by the spike from neuron i .²

As mentioned above, in our integrate-and-fire model both of these spikes would happen at the same moment in time $\tau \in \mathfrak{T}_i \cap \mathfrak{T}_j$ because $u_i(\tau^-) + w_{ij} \geq V_F$, even though in “reality” neuron j would have fired before neuron i . To reflect this ordering, we will reset neuron j to $V_R + w_{ji}$ and neuron i to V_R .

It could still be that two neurons i and j exceed V_F at the same time without one causing the other: either when both neurons continuously approach V_F simultaneously due to f_i and f_j , or if a presynaptic spike from some third neuron k puts both potentials above V_F .

In the first case, there is no way to establish any kind of order, so the effects of both spikes have to be processed at the same time. We choose not to let the neurons affect each other’s post-spike potential in this case.

In the second case it could be that (w.l.o.g.) u_j exceeds V_F by more than u_i , i.e., $u_j(\tau^-) + w_{jk} > u_i(\tau^-) + w_{ik}$. In that case, we assume that in the process we are modeling neuron j fired before neuron i , so we choose to process neuron j ’s spike first. This assumption is based on the idea that exceeding V_F by a larger amount causes the neuron to spike faster, which is generally consistent with biophysical models.

Remark 3.3. If we have $w_{ij} < 0$ in the example above, i.e., the synapse from neuron j to neuron i is inhibitory, this modeling choice means that neuron j ’s spike could prevent neuron i from firing even though neuron i ’s potential would have exceeded V_F due to neuron k ’s spike.

Remark 3.4. Most of these considerations about neurons immediately causing other neurons to spike are not relevant for the exponential integrate-and-fire model with $V_F = +\infty$. Under that model, a neuron j that fires at time $\tau \in \mathfrak{T}_j$ can never cause another neuron i to fire immediately at time τ as well because $u_i(\tau^-) + w_{ij} < +\infty = V_F$. When V_F is finite but large this immediate spiking is also unlikely to happen if w_{ij} is relatively small.

With these modeling choices in mind, we adapt (3.1) to

$$\begin{cases} du_i &= f_i(t, u_i) dt + \sum_{\tau \in \mathfrak{T}_i} (V_R - u_i(\tau^-)) \delta_\tau(dt) + \sum_{\tau \in \mathfrak{T}} \sum_{j \in \bigcup_{n=\ell_i(\tau)+1}^N \mathfrak{N}^n(\tau)} w_{ij} \delta_\tau(dt), & i \in \{1, \dots, N\}, t \in I, \\ u_i(t_0) &= v_i, & i \in \{1, \dots, N\}. \end{cases}$$

²Note that possible (absolute) refractory behavior is also relevant here, but we neglect this.

Here $\ell_i(\tau)$ and $\mathfrak{N}^n(\tau)$ are not yet defined, which we will do in the next section where it is more convenient. For now, we state that they provide an ordering of the neurons that fire at time τ in accordance with the choices made above.

3.2 Solution construction

The process behind this solution construction is the same as in Section 2.4.4. As in that section, we first formulate a preliminary system without resets or presynaptic spikes. We now do so for every $i \in \{1, \dots, N\}$:

$$\begin{cases} \frac{d}{dt} \tilde{u}_i &= f_i(t, \tilde{u}_i), & t \in I, \tilde{u}_i \in U, \\ \tilde{u}_i(t_0) &= \tilde{u}_0. \end{cases} \quad (3.2)$$

We again explicitly incorporate the initial conditions by writing $\tilde{u}_i(\cdot, \tilde{t}_0, \tilde{u}_0)$ to denote the solution of this system. Let $J_i(\tilde{t}_0, \tilde{u}_0) \subset I$ be the maximal existence interval of this solution.

Now we define the potential evolutions before the first spike as

$$\tilde{u}_i^{(1)}(t) := \tilde{u}_i(t, t_0, v_i), \quad t \in J_i(t_0, v_i), \quad i \in \{1, \dots, N\}. \quad (3.3)$$

Before the first firing time we, by definition, do not have any spikes. Thus the first firing time is given by

$$r_1 := \inf \left\{ \tau \in \bigcap_{i=1}^N J_i(t_0, v_i) \mid \exists i \in \{1, \dots, N\} : \tilde{u}_i^{(1)}(\tau^-) \geq V_F \right\}.$$

To determine which neurons fire at time r_1 , we need to take into account that presynaptic spikes can trigger postsynaptic spikes immediately. We have to be careful here because the order in which spikes and resets are processed can influence the behavior of the model as explained in Section 3.1.

We define $\mathfrak{N}(\tau)$ to be the set of neurons that fire at time $\tau \in I$ (which could be empty). To determine $\mathfrak{N}(r_1)$ we have to perform a recursive construction. Since this construction will be the same for all firing times r_k , we do it in general already. Let

$$\mathfrak{N}^1(r_k) := \left(\arg \max_{i \in \{1, \dots, N\}} \tilde{u}_i^{(k)}(r_k^-) \right) \cap \{i \in \{1, \dots, N\} \mid \tilde{u}_i^{(k)}(r_k^-) \geq V_F\}, \quad (3.4)$$

which we know is non-empty by definition of r_k . Here we are selecting the neurons whose potential is largest and above V_F , in accordance with Section 3.1. Unless two or more neurons have exactly the same potential, this set will contain just one neuron. Now define recursively, for $l \in \mathbb{N}$,

$$\begin{aligned} \mathfrak{N}^{l+1}(r_k) := & \left(\arg \max_{i \in \{1, \dots, N\} \setminus \bigcup_{n=1}^l \mathfrak{N}^n(r_k)} \tilde{u}_i^{(k)}(r_k) + \sum_{j \in \bigcup_{n=1}^l \mathfrak{N}^n(r_k)} w_{ij} \right) \\ & \cap \{i \in \{1, \dots, N\} \mid \tilde{u}_i^{(k)}(r_k) + \sum_{j \in \bigcup_{n=1}^l \mathfrak{N}^n(r_k)} w_{ij} \geq V_F\}, \end{aligned} \quad (3.5)$$

where by construction $\mathfrak{N}^{l+1}(r_k)$ contains the neurons that fire at time r_k after l previous “batches” of neurons that fired (at time r_k).

Note that the sequence of sets $\{\mathfrak{N}^l(r_k)\}_{l \in \mathbb{N}}$ is disjoint, and that for $l \in \mathbb{N}$,

$$\mathfrak{N}^l(r_k) = \emptyset \implies \forall m > l : \mathfrak{N}^m(r_k) = \mathfrak{N}^l(r_k) = \emptyset.$$

Since always $\mathfrak{N}^l(r_k) \subset \{1, \dots, N\}$, it follows that for all $l > N$ we have $\mathfrak{N}^l(r_k) = \emptyset$. This corresponds with the fact that at most all neurons $1, \dots, N$ neurons can fire at time r_k . Hence the set of neurons that fire at time r_k , which is $\mathfrak{N}(r_k)$ as defined above, satisfies

$$\mathfrak{N}(r_k) = \bigcup_{n=1}^N \mathfrak{N}^n(r_k). \quad (3.6)$$

We require a way to tell to which firing batch \mathfrak{N}^l a given neuron at a given firing time belongs. To this end, we define for a neuron $i \in \{1, \dots, N\}$ at one of its firing times $\tau \in \mathfrak{T}_i$ the integer $\ell_i(\tau) \in \{1, \dots, N\}$ to be the unique batch index such that $i \in \mathfrak{N}_i^{\ell_i(\tau)}$. It will be convenient for (3.11) below to also define $\ell_i(\tau) = 0$ if $\tau \notin \mathfrak{T}_i$.

Inductively we now define the other firing times and for $k \in \mathbb{N}$,

$$\tilde{u}_i^{(k+1)}(t) := \begin{cases} \tilde{u}_i(t, r_k, \tilde{u}_i^{(k)}(r_k) + \sum_{j \in \mathfrak{N}(r_k)} w_{ij}), & t \in J_i(r_k, \tilde{u}_i^{(k)}(r_k) + \sum_{j \in \mathfrak{N}(r_k)} w_{ij}), \quad \text{if } i \notin \mathfrak{N}(r_k), \\ \tilde{u}_i(t, r_k, V_R + \sum_{j \in \bigcup_{n=\ell_i(\tau)+1}^N \mathfrak{N}^n(r_k)} w_{ij}), & t \in J_i(r_k, V_R), \quad \text{if } i \in \mathfrak{N}(r_k), \end{cases} \quad (3.7)$$

and

$$r_{k+1} := \inf \left\{ \tau \in \bigcap_{i=1}^N J_i(r_k, \tilde{u}_i^{(k+1)}(r_k)), \tau > r_k \mid \exists i \in \{1, \dots, N\} : \tilde{u}_i^{(k+1)}(\tau^-) \geq V_F \right\}. \quad (3.8)$$

For (3.7) to be well-defined we have to avoid immediate refiring of a neuron, which could occur if the weights were large. To this end, we introduce Assumption 3.1 below, which ensures for (3.7) that in its second case the initial potential (third argument of \tilde{u}_i) is still less than V_F , so in U .

Assumption 3.1. For each neuron i we have

$$\sum_{j=1}^N \max\{0, w_{ij}\} < V_F - V_R.$$

With this, we have inductively constructed all firing times r_k and their associated sets of firing neurons. Thus,

$$\mathfrak{T} = \{r_k \mid k \in \mathbb{N}\} \setminus \{+\infty\} \quad (3.9)$$

is the set of all firing times.³

Finally, the potential evolution u_i for every neuron $i \in \{1, \dots, N\}$ is given by patching together all interspike solutions $u_i^{(k)}$. Namely,

$$u_i(t) := \sum_{k \in \mathbb{N}} \tilde{u}_i^{(k)}(t) \mathbb{1}_{[r_{k-1}, r_k)}(t), \quad t \in I \cap [t_0, \sup_{k \in \mathbb{N}} r_k), \quad (3.10)$$

where again by convention we set $r_0 := t_0$. Note that the difference between $\sup \mathfrak{T}$ and $\sup_{k \in \mathbb{N}} r_k$ is relevant here because the latter can be $+\infty$ when $\sup I < +\infty$ whereas the former can not. This ensures that u_i also defined on $(\sup \mathfrak{T}, \sup I]$ when $\sup I < +\infty$.

We intersect I with $[t_0, \sup_{k \in \mathbb{N}} r_k)$ here because we could theoretically have $\sup_{k \in \mathbb{N}} r_k < \sup I$. However, we will see in Lemma 3.1 that this will not happen for $V_F < +\infty$ due to the continuity of $\{f_i\}_{i \in \{1, \dots, N\}}$, making this intersection unnecessary.

By construction, u satisfies the following (formal) differential equation:

$$\begin{cases} du_i = f_i(t, u_i) dt + \sum_{\tau \in \mathfrak{T}_i} (V_R - u_i(\tau^-)) \delta_\tau(dt) + \sum_{\tau \in \mathfrak{T}} \sum_{j \in \bigcup_{n=\ell_i(\tau)+1}^N \mathfrak{N}^n(\tau)} w_{ij} \delta_\tau(dt), & i \in \{1, \dots, N\}, t \in I, \\ u_i(t_0) = v_i, & i \in \{1, \dots, N\}, \end{cases} \quad (3.11)$$

which we already stated in the previous section. Here the sets \mathfrak{T}_i depend implicitly on u . Because of the following lemma, in most cases the constructed solution will hold on all of I instead of only on $I \cap [t_0, \sup_{k \in \mathbb{N}} r_k)$.

Lemma 3.1. Suppose $I = [t_0, t_e]$, $V_F < +\infty$, and let Assumption 3.1 hold. Then there is a uniform lower bound $\Delta t_{\min} > 0$ for the time between two spikes of the same neuron, i.e.,

$$\forall i \in \{1, \dots, N\} : \inf_{\substack{\tau_1, \tau_2 \in \mathfrak{T}_i \\ \tau_1 \neq \tau_2}} |\tau_2 - \tau_1| \geq \Delta t_{\min}. \quad (3.12)$$

In particular, this implies that we have $\sup_{k \in \mathbb{N}} r_k \geq \sup I$. This means that the solution is global on I .

³We subtract $+\infty$ for the same reason as in Section 2.4.4.

Proof. As N is finite we can define, for all $u \in U$,

$$F(u) := \max_{i \in \{1, \dots, N\}} \max_{t \in I} f_i(t, u),$$

which is the maximum potential growth rate at u over all neurons and all of I . Here the inner maximum is well-defined as I is compact and f_i is continuous. Note furthermore that since f_i is continuous, for fixed $u \in U$ the map $t \mapsto f_i(t, u)$ on I is uniformly continuous as I is compact. This implies that the map $u \mapsto \max_{t \in I} f_i(t, u)$ is continuous on U for all $i \in \{1, \dots, N\}$, so F is also continuous.

Now choose some neuron $i \in \{1, \dots, N\}$ and one of its firing times $\tau_1 \in \mathfrak{T}_i$, and suppose there is a next firing time $\tau_2 = \min \mathfrak{T}_i \cap (\tau_1, +\infty)$. We can assume without loss of generality that all other neurons $\{1, \dots, N\} \setminus \{i\}$ fire at most once in $[\tau_1, \tau_2)$, because otherwise we would choose a different neuron i with firing times $\hat{\tau}_1, \hat{\tau}_2 \in [\tau_1, \tau_2)$ such that $\hat{\tau}_2 - \hat{\tau}_1 < \tau_2 - \tau_1$.

It suffices to construct $\Delta t_{\min} > 0$ in such a way that $\tau_2 - \tau_1 \geq \Delta t_{\min}$.

In the absence of arriving spikes from the other neurons, we would have $u_i(\tau_1) = V_R$, and τ_2 would depend only on the time it takes f_i to increase u_i by $V_F - V_R$. By an analogous argument as given in the proof of Lemma 2.4, we would have

$$\tau_2 - \tau_1 \geq \int_{V_R}^{V_F} \frac{1}{F(u)} du.$$

Here the fact that $\mathfrak{T} \neq \emptyset$ implies that F is bounded below by some positive constant (similar to the condition required in Lemma 2.3), so this integral is positive and finite.

However, the time between spikes can of course be reduced due to spike arrivals from excitatory synapses. This is where Assumption 3.1 comes in, which states that

$$\Delta w := \sum_{j=1}^N \max\{0, w_{ij}\} < V_F - V_R.$$

On $[\tau_1, \tau_2)$, u_i can thus be instantaneously incremented at most $N - 1$ times by in total at most Δw , as we assumed that each other neuron fires no more than once on that interval. As these increments can come at any moment in the trajectory of neuron i , we instead have

$$\tau_2 - \tau_1 \geq \inf \left\{ \sum_{k=1}^n \int_{q_{2k-1}}^{q_{2k}} \frac{1}{F(u)} du \mid n \in \mathbb{N}, V_R = q_1 \leq \dots \leq q_{2n} = V_F, \sum_{k=1}^{n-1} (q_{2k+1} - q_{2k}) \leq \Delta w \right\}. \quad (3.13)$$

Here q_{2k} and q_{2k+1} , $k = 1, \dots, n-1$, are respectively the values of the potential before and after the k th instantaneous increment of u_i .⁴ Note that $q_{2k+1} - q_{2k}$ is thus the size of the k th increment, of which the total has to be less than Δw . Also observe that on the intervals (q_{2k-1}, q_{2k}) the potential is continuous, which explains why the integral of $1/F$ contributes to the inter-spike time there, and that not all of these intervals are empty since

$$\sum_{k=1}^n (q_{2k} - q_{2k-1}) = q_{2n} - q_1 - \sum_{k=1}^{n-1} (q_{2k+1} - q_{2k}) \geq V_F - V_R - \Delta w > 0. \quad (3.14)$$

We could have imposed $n \leq N - 1$ in (3.13), because there are at most $N - 1$ increments, but for what follows this will not be necessary.

Let $V_\Delta \in \mathbb{R}$ be such that $V_R + \Delta w < V_\Delta < V_F$. Then F attains a maximum on $[V_R, V_\Delta]$, because F is continuous. Let $M := \max_{u \in [V_R, V_\Delta]} F(u)$. Now the idea is that at least a distance of $V_\Delta - V_R - \Delta w$ has to be traversed continuously (without jumps) during which F is bounded by M .

Consider any $n \in \mathbb{N}$ with q_1, \dots, q_{2n} as in (3.13), and define $m := \max\{k \in \{1, \dots, n-1\} \mid q_{2k} \leq V_\Delta\}$.

⁴For $k \neq 1, n$, we have for some firing time $\tau \in (\tau_1, \tau_2) \cap \mathfrak{T}$ that $q_{2k+1} = u_i(\tau^-)$ and $q_{2k} = u_i(\tau)$ (not necessarily for $k = 1, n$ because the first and last increment could be special due to happening at V_R and V_F).

Then by (3.14) we get

$$\begin{aligned}
 \sum_{k=1}^m (q_{2k} - q_{2k-1}) &\geq V_F - V_R - \Delta w - \sum_{k=m+1}^n (q_{2k} - q_{2k-1}), \\
 &\geq V_F - V_R - \Delta w - \sum_{k=2m+2}^{2n} (q_k - q_{k-1}), \\
 &= V_F - V_R - \Delta w - (V_F - q_{2m+1}), \\
 &\geq V_F - V_R - \Delta w - (V_F - V_\Delta), \\
 &= V_\Delta - V_R - \Delta w > 0.
 \end{aligned} \tag{3.15}$$

Now the argument is straightforward, because we have for all $k = 1, \dots, m$,

$$\int_{q_{2k-1}}^{q_{2k}} \frac{1}{F(u)} du \geq (q_{2k} - q_{2k-1})/M,$$

so using (3.15) we find that

$$\sum_{k=1}^m \int_{q_{2k-1}}^{q_{2k}} \frac{1}{F(u)} du \geq \sum_{k=1}^m (q_{2k} - q_{2k-1})/M \stackrel{(3.15)}{\geq} (V_\Delta - V_R - \Delta w)/M.$$

Thus, we have

$$\tau_2 - \tau_1 \geq \Delta t_{\min}$$

for

$$\Delta t_{\min} := (V_\Delta - V_R - \Delta w) / \left(\max_{u \in [V_R, V_\Delta]} F(u) \right) > 0$$

where we are free to choose $V_\Delta \in (V_R + \Delta w, V_F)$ (which could be used to maximize this lower bound). As this choice of Δt_{\min} is independent of the chosen neuron i and the spike times τ_1, τ_2 , we conclude that it satisfies (3.12).

We now show that the existence of this minimal inter-spike time implies that $\sup_{k \in \mathbb{N}} r_k \geq \sup I$. We argue by contradiction. Suppose that $\sup_{k \in \mathbb{N}} r_k < \sup I$. From (3.9) it is clear that this implies that $\#\mathfrak{T} = +\infty$. Consider (3.11) on the compact interval $\hat{I} := [t_0, \sup_{k \in \mathbb{N}} r_k]$, which contains all firing times in \mathfrak{T} . Let $L := \sup_{k \in \mathbb{N}} r_k - t_0$ be the length of this interval. Then we know that each individual neuron fired at most $\lceil L/\Delta t_{\min} \rceil$ times on \hat{I} , so

$$\#\mathfrak{T} \leq N \lceil L/\Delta t_{\min} \rceil < +\infty.$$

This contradicts $\#\mathfrak{T} = +\infty$, so $\sup_{k \in \mathbb{N}} r_k \geq \sup I$. ■

3.3 Homogeneous populations with full connectivity

As a first example of the system defined in Section 3.1, we will consider the simplified case where all neurons have the same driving force $f_i = f$ and are fully connected with equal synaptic weights $w_{ij} = \frac{w_0}{N}$, for some fixed parameter $w_0 \in \mathbb{R}$. The reason we divide by the number of neurons N here is to enable the mean-field limit analysis we will perform in Section 3.3.2, where we will take N to ∞ . But first, we analyze possible synchronization between the neurons under this connectivity scheme.

Remark 3.5. To satisfy Assumption 3.1 we require $w_0 < V_F - V_R$.

3.3.1 Synchronization

In excitatory networks, where $w_0 > 0$, the potentials of the neurons tend to synchronize over time when the driving force is decreasing in u . This behavior is interesting in and of itself, but it will also be relevant for the mean-field limit in the next section because it violates some of the absolute continuity assumptions we make there. Analyzing when synchronization takes place thus provides us with necessary conditions for the validity of some of the results we will derive.

The following proposition does not provide a full characterization of long-term synchronizing behavior, but it does establish that the potentials get closer together after one firing cycle if they start reasonably close together but are still separated enough so as to fire at separate times. It also gives a lower bound for the reduction in their spread after this firing cycle.

The proof is rather technical, but the basic idea is that because we consider a decreasing driving force f the effect of arriving pulses is greater on neurons close to V_F than on neurons close to V_R . This causes the “pull” of the neurons that fire first on the neurons behind them to be stronger than the “push” the neurons that fire last exert on the first neurons.

Remark 3.6. The assumption that all potentials are sufficiently well-separated in the proposition below is stronger than what is required. Specifically, it suffices for the N th neuron to fire separately (in time) from the $(N - 1)$ th neuron. It is however difficult to ensure this elegantly with an assumption on the initial conditions.

Proposition 3.2. *Assume that $w_0 > 0$, and that the driving force is homogeneous (meaning $f(t, \cdot) = f(\cdot)$), decreasing and bounded below by some positive constant on U ; this means that Lemma 2.3 applies (we will use the function ν defined there). Suppose that the initial potentials $u_i(t_0) = v_i$ satisfy $V_R < v_N < v_{N-1} < \dots < v_1 < V_F$. Moreover, we assume*

$$R_{\text{pre}} := \nu^{-1}(v_1) - \nu^{-1}(v_N) < \frac{T}{2}, \quad (3.16)$$

where T is the period of Lemma 2.3, and

$$\forall i \in \{1, \dots, N - 1\} : \nu^{-1}(v_i) - \nu^{-1}(v_{i+1}) > T - \nu^{-1}(V_F - \frac{w_0}{N}). \quad (3.17)$$

Then all neurons will have spiked once at firing time r_N , and at that point the spread of the potentials will have decreased, with

$$R_{\text{post}} := \nu^{-1}(u_1(r_N)) - \nu^{-1}(u_N(r_N)) < R_{\text{pre}} - (N - 1) \int_0^{\frac{w_0}{N}} \left(\frac{1}{f(u + T - R_{\text{pre}})} - \frac{1}{f(u + R_{\text{pre}})} \right) du < R_{\text{pre}}.$$

Proof. By Lemma 2.3 it follows that each neuron on its own would enter a mode of repetitive firing. The difference with the setting of Chapter 2 is that there are now also inter-neuron spikes, but as all weights $w_{ij} = \frac{w_0}{N}$ are positive it is clear that these can only accelerate the firing of a neuron, not prevent it.

We will use the bijection ν from Lemma 2.3 to map between potential-space $[V_R, V_F)$ and time-space $[0, T)$, where time-space refers to the position in time w.r.t. the periodic firing of an isolated neuron under the same conditions. To keep track of the position in time-space we define

$$\tau_i(t) := \nu^{-1}(u_i(t)), \quad i \in \{1, \dots, N\}.$$

A neuron i fires when its potential u_i reaches V_F , but because ν is an order-preserving bijection we can equivalently say (or even define) that a neuron fires when τ_i reaches T . The advantage of working in time-space is that the distance between any pair of neurons can change only at firing times. This is straightforward: suppose $s > t_0$ with $s \notin \mathfrak{T}$. Then for any $i \in \{1, \dots, N\}$ we have that u_i satisfies the ODE in (3.2) at s . Recall also the derivative of ν^{-1} as computed in the proof of Lemma 2.3, which is

$$\frac{d\nu^{-1}}{du}(u) = \frac{1}{f(u)}.$$

Using this, we obtain

$$\left. \frac{d}{dt} \tau_i(t) \right|_{t=s} = \left. \frac{d}{dt} \nu^{-1}(u_i(t)) \right|_{t=s} = \frac{d\nu^{-1}}{du}(u_i(s)) \frac{d}{dt} u_i(s) = \frac{1}{f(u_i(s))} f(u_i(s)) = 1.$$

So between firing times τ_i changes at a constant rate of 1. This also means that the distance between any pair of neurons in time-space remains the same. To study synchronization, we thus only have to look at what happens at the firing times.

The crucial observation of the proof is that because f is decreasing, the increase $\tau_i(r) - \tau_i(r^-)$ caused by a presynaptic spike at some firing time r of strength w is larger when $\tau_i(r^-)$ is larger. This can be seen from the following explicit expression for this increase. Define for $\tau \in [0, T)$ and $w > 0$ with $\nu(\tau) + w < V_F$,

$$\begin{aligned} \Delta\tau[\tau, w] &:= \nu^{-1}(\nu(\tau) + w) - \tau, \\ &= \nu^{-1}(\nu(\tau)) + \int_{\nu(\tau)}^{\nu(\tau)+w} \frac{d\nu^{-1}}{du}(u) du - \tau, \\ &= \int_{\nu(\tau)}^{\nu(\tau)+w} \frac{1}{f(u)} du, \\ &= \int_0^w \frac{1}{f(u + \nu(\tau))} du. \end{aligned} \tag{3.18}$$

Because f is decreasing and ν is increasing, $\Delta[\cdot, w]$ is increasing. So when $\tau_i(r^-)$ is larger, it receives a larger boost (for fixed w). Because we only have to study what happens at the firing times, we define for $k \in \mathbb{N}$

$$\Delta r_k := r_k - r_{k-1},$$

where by convention $r_0 = t_0$. This is the increase in all τ 's between the $(k-1)$ th and k th firing time. It follows from this that in general

$$\tau_i(r_k) = \begin{cases} \Delta\tau[0, \frac{w_0}{N} \sum_{n=l+1}^N \#\mathfrak{N}^n(r_k)] & \text{if } i \in \mathfrak{N}^l(r_k), \\ \tau_i(r_{k-1}) + \Delta r_k + \Delta\tau[\tau_i(r_{k-1}) + \Delta r_k, \frac{w_0}{N} \#\mathfrak{N}(r_k)] & \text{if } i \notin \mathfrak{N}(r_k). \end{cases} \tag{3.19}$$

Because in our special case the second argument of $\Delta\tau$ will turn out to always be $\frac{w_0}{N}$, we will omit it. We claim that all neurons fire at separate moments in time during the first cycle, which means that

$$\forall k \in \{1, \dots, N\} : \mathfrak{N}(r_k) = \{k\}. \tag{3.20}$$

We will show this by induction using the assumption (3.17). To show that $\mathfrak{N}(r_1) = \{1\}$, we first observe that since $v_1 > v_2 > \dots > v_N$ (and the ordering of the potentials is preserved until r_1 because of local uniqueness), we certainly have $\mathfrak{N}^1(r_1) = \{1\}$ by (3.4). Note that this also means that

$$\Delta r_1 = r_1 - t_0 = T - \tau_1(t_0).$$

Subsequently, by (3.5),

$$\mathfrak{N}^2(r_1) = \arg \max_{i \in \{2, \dots, N\}} (u_i(r_1^-) + \frac{w_0}{N}) \cap \{i \in \{1, \dots, N\} \mid u_i(r_1^-) + \frac{w_0}{N} \geq V_F\}.$$

However, from (3.17) it follows that

$$\begin{aligned} u_2(r_k^-) + \frac{w_0}{N} &= \nu(\tau_2(r_k^-)) + \frac{w_0}{N}, \\ &= \nu(\tau_2(t_0) + \Delta r_1) + \frac{w_0}{N}, \\ &\stackrel{(3.17)}{<} \nu(\tau_1(t_0) + \nu^{-1}(V_F - \frac{w_0}{N}) - T + \Delta r_1) + \frac{w_0}{N}, \\ &= \nu(T + \nu^{-1}(V_F - \frac{w_0}{N}) - T) + \frac{w_0}{N}, \\ &= V_F - \frac{w_0}{N} + \frac{w_0}{N}, \\ &= V_F. \end{aligned}$$

And since $u_2(r_k^-) > u_3(r_k^-) > \dots > u_N(r_k^-)$ we find that $\mathfrak{N}^2(r_1) = \emptyset$, which means that indeed

$$\mathfrak{N}(r_k) = \{1\}$$

by (3.6). Now suppose (3.20) holds for all $k \leq m$ for some $m \in \{1, \dots, N-2\}$. Then

$$\tau_i(r_m) = \begin{cases} r_m - r_i + \sum_{n=i+1}^m \Delta\tau[\tau_i(r_{n-1}) + \Delta r_n] & \text{if } i \in \{1, \dots, m\}, \\ \tau_i(t_0) + r_m - t_0 + \sum_{n=1}^m \Delta\tau[\tau_i(r_{n-1}) + \Delta r_n] & \text{if } i \in \{m+1, \dots, N\}. \end{cases} \tag{3.21}$$

For firing time r_{m+1} we then find that $\tau_{m+1}(r_{m+1}^-) = \tau_i(r_m) + \Delta r_{m+1}$ is largest, since $\Delta\tau[\cdot]$ is increasing (and nonnegative) and $\tau_{m+1}(t_0) = v_{m+1} > v_{m+2} > \dots > v_N$. As $\tau_{m+1}(r_{m+1}^-) = \nu^{-1}(u_{m+1}(r_{m+1}^-))$ where ν is an order-preserving bijection, we must have $\mathfrak{N}^1(r_{m+1}) = \{m+1\}$.

Now, since $m + 1 < N$ we have that the next-largest potential is $u_{m+2}(r_{m+1}^-)$, so by a similar argument as given above, using (3.17), we get $\mathfrak{N}^2(r_{m+1}) = \emptyset$ meaning $\mathfrak{N}(r_{m+1}) = \{m + 1\}$.

By induction it now follows that (3.20) holds at least for all $k \in \{1, \dots, N - 1\}$. Furthermore, the first part of the argument also applies when $m = N - 1$, so $N \in \mathfrak{N}^1(r_N)$. We will show that that (3.20) also holds for $k = N$ in a moment.

Just before the N th firing time u_N is coming up on V_F (because $N \in \mathfrak{N}^1(r_N)$), so

$$\tau_N(r_N^-) = r_N - t_0 + \tau_N(t_0) + \sum_{n=1}^{N-1} \Delta\tau[\tau_N(r_{n-1}) + \Delta r_n] = T.$$

Therefore

$$r_N - t_0 = T - \tau_N(t_0) - \sum_{n=1}^{N-1} \Delta\tau[\tau_N(r_{n-1}) + \Delta r_n].$$

For the other neurons, we have a similar expression as in (3.21), namely for $i \in \{1, \dots, N - 1\}$,

$$\tau_i(r_N^-) = r_N - r_i + \sum_{n=i+1}^{N-1} \Delta\tau[\tau_i(r_{n-1}) + \Delta r_n].$$

In particular,

$$\begin{aligned} \tau_1(r_N^-) &= r_N - r_1 + \sum_{n=2}^{N-1} \Delta\tau[\tau_1(r_{n-1}) + \Delta r_n], \\ &= T - \tau_N(t_0) - \sum_{n=1}^{N-1} \Delta\tau[\tau_N(r_{n-1}) + \Delta r_n] - (T - \tau_1(t_0)) + \sum_{n=2}^{N-1} \Delta\tau[\tau_1(r_{n-1}) + \Delta r_n], \\ &= R_{\text{pre}} - \sum_{n=2}^{N-1} (\Delta\tau[\tau_N(r_{n-1}) + \Delta r_n] - \Delta\tau[\tau_1(r_{n-1}) + \Delta r_n]) - \Delta\tau[\tau_N(t_0) + \Delta r_1], \\ &< R_{\text{pre}} \stackrel{(3.16)}{<} \frac{T}{2}. \end{aligned} \tag{3.22}$$

Here we used that $\tau_N(r_{n-1}) > \tau_1(r_{n-1})$ for all $n \in \{2, \dots, N - 1\}$. This means that the firing of neuron N does not trigger neuron 1, since combining (3.16) and (3.17) clearly yields

$$\begin{aligned} (N - 1)(T - \nu^{-1}(V_F - \frac{w_0}{N})) &< \frac{T}{2}, \\ \frac{T}{2} &< (1 - \frac{1}{2(N - 1)})T < \nu^{-1}(V_F - \frac{w_0}{N}), \end{aligned}$$

and thus

$$u_1(r_N^-) + \frac{w_0}{N} = \nu(\tau_1(r_N^-)) + \frac{w_0}{N} < \nu(\frac{T}{2}) + \frac{w_0}{N} < \nu(\nu^{-1}(V_F - \frac{w_0}{N})) + \frac{w_0}{N} = V_F$$

Because $\tau_1(r_N^-)$ is the largest among the neurons that have already fired, we can now conclude that $\mathfrak{N}^2(r_N) = \emptyset$ so that (3.20) holds also for $k = N$ (as foreshadowed). This means that we can finally write

$$\begin{aligned} \tau_1(r_N) &= R_{\text{pre}} - \sum_{n=2}^{N-1} (\Delta\tau[\tau_N(r_{n-1}) + \Delta r_n] - \Delta\tau[\tau_1(r_{n-1}) + \Delta r_n]) \\ &\quad - (\Delta\tau[\tau_N(t_0) + \Delta r_1] - \Delta\tau[\tau_1(r_{N-1}) + \Delta r_N]). \end{aligned}$$

Note that for all $n \in \{2, \dots, N - 1\}$ we have

$$\tau_N(r_{n-1}) + \Delta r_n = \tau_N(r_n^-) > \tau_N(r_1^-) = \tau_N(t_0) + \Delta r_1,$$

and for all $n \in \{2, \dots, N\}$,

$$\tau_1(r_{n-1}) + \Delta r_n = \tau_1(r_n^-) < \tau_1(r_{N-1}^-) \stackrel{(3.22)}{<} R_{\text{pre}}.$$

Therefore

$$\begin{aligned}
 R_{\text{post}} &= \tau_1(r_N) - \tau_N(r_N), \\
 &= \tau_1(r_N), \\
 &< R_{\text{pre}} - (N-2)(\Delta\tau[\tau_N(t_0) + \Delta r_1] - \Delta\tau[R_{\text{pre}}]) - (\Delta\tau[\tau_N(t_0) + \Delta r_1] - \Delta\tau[R_{\text{pre}}]), \\
 &= R_{\text{pre}} - (N-1)(\Delta\tau[\tau_N(t_0) + \Delta r_1] - \Delta\tau[R_{\text{pre}}]), \\
 &= R_{\text{pre}} - (N-1)(\Delta\tau[\tau_N(t_0) + T - \tau_1(t_0)] - \Delta\tau[R_{\text{pre}}]), \\
 &= R_{\text{pre}} - (N-1)(\Delta\tau[T - R_{\text{pre}}] - \Delta\tau[R_{\text{pre}}]).
 \end{aligned}$$

Because of (3.16), we know that

$$T - R_{\text{pre}} > T - \frac{T}{2} = \frac{T}{2} > R_{\text{pre}},$$

which means that

$$\Delta\tau[T - R_{\text{pre}}] - \Delta\tau[R_{\text{pre}}] = \int_0^{\frac{w_0}{N}} \left(\frac{1}{f(u + T - R_{\text{pre}})} - \frac{1}{f(u + R_{\text{pre}})} \right) du > 0.$$

So we indeed have

$$R_{\text{post}} < R_{\text{pre}} - (N-1) \int_0^{\frac{w_0}{N}} \left(\frac{1}{f(u + T - R_{\text{pre}})} - \frac{1}{f(u + R_{\text{pre}})} \right) du < R_{\text{pre}}. \quad \blacksquare$$

The proof of the proposition above requires two rather technical assumptions in the form of (3.16) and (3.17), and also only claims that progress towards synchronization takes place during one firing cycle. A stronger result about long-term synchronization to some minimal distance between the potentials is probably possible. Especially (3.20) could likely be weakened, as noted in Remark 3.6. However, the technical assumptions do play a role in preventing certain pathological initial conditions from which synchronization will not take place at all.

For instance, suppose $N = 2$ with initial conditions such that $\tau_1(r_1^-) = T$, which always holds, and $\tau_2(r_1^-) = \frac{T}{2} - \varepsilon$ for some $0 < \varepsilon < \frac{T}{2}$ (note that this violates (3.16)) and weight parameter $w_0 > 0$ such that $\Delta\tau[\frac{T}{2} - \varepsilon, \frac{w_0}{2}] = 2\varepsilon$. Then we have $\tau_2(r_1) = \frac{T}{2} + \varepsilon$, so at the second firing time $r_2 = r_1 + T - \tau_2(r_1) = r_1 + \frac{T}{2} - \varepsilon$ we have $\tau_1(r_2^-) = \frac{T}{2} - \varepsilon$ and $\tau_2(r_2^-) = T$. We are back in the same situation as at r_1 (except with the roles of neuron 1 and 2 reversed), so this system will continue indefinitely without synchronizing.

Based on numerical results, examples like this one where the time-space distance of the neurons remains constant are likely unstable, meaning that synchronization would eventually occur for most initial conditions. This is also observed in the example below, in Figure 3.1b. Still, these pathological cases illustrate that a claim about global synchronization is not as straightforward as it might seem.

Example 3.1. Proposition 3.2 applies to the (nondimensionalized) leaky integrate-and-fire model under sufficient constant input $I_{\text{ext}} \equiv I_0$, because the driving force $f(u) = -(u - V_R) + I_0$ of that model is decreasing. Specifically, the proposition applies when $I_0 > V_F - V_{\text{rest}} = 1 - V_{\text{rest}}$ as then f is bounded below by a positive constant on U .

We will again consider our standard parameters $V_{\text{rest}} = 0.37$ and $I_0 = 0.65$. From Example 2.1 and Example 2.2 we find that in that case

$$\nu^{-1}(u) = \ln \frac{V_{\text{rest}} + I_0}{V_{\text{rest}} + I_0 - u} = \ln \frac{1.02}{1.02 - u},$$

and

$$T = 1/g(I_0) = \nu^{-1}(V_F) = \ln \frac{1.02}{1.02 - 1} \approx 3.93.$$

To satisfy (3.16), we thus require

$$\ln \frac{1.02 - v_N}{1.02 - v_1} = \nu^{-1}(v_1) - \nu^{-1}(v_N) < \frac{T}{2} \approx 1.97. \quad (3.23)$$

In Figure 3.1 we see simulation results for a case where the initial potentials are uniformly distributed on $[0, 0.85]$ and a case where they are distributed on $[0, 0.99] = U$. Note that the first case satisfies (3.23), because

$$\ln \frac{1.02 - 0}{1.02 - 0.85} \approx 1.79 < \frac{T}{2},$$

whereas the second case does not since

$$\ln \frac{1.02 - 0}{1.02 - 0.99} \approx 3.52 > \frac{T}{2}.$$

For the first case, we observe in Figure 3.1a that the variance of the potentials decreases rapidly to a minimum value. The activity histogram, which shows when spikes took place, shows that the firings of the neurons synchronize (almost) immediately.

Figure 3.1b corresponding to the second case is markedly different. The variance eventually decays to a low value, but this takes multiple firing cycles to achieve. If we look closely at the activity graph we see that between $t = 10$ s and $t = 40$ s most neurons fire around the same time, but a small contingent of neurons fires before the main group. This is likely an example of a pathological case similar to the one discussed above.

Remark 3.7. The spikes in the variance graph in Figure 3.1b (and the single variance spike in Figure 3.1a) are due to some neurons already being past V_F while some are still before it, causing their inter-distances to increase. The variance does not take into account the torus-like nature of U .

Remark 3.8. The reason that the variance does not converge to 0 but instead stays slightly above it is due to how we have chosen to handle simultaneous spikes. Namely, the post-spike potential of the first neuron that crosses V_F is affected by the pulses of the neurons that cross V_F afterwards. This prevents the neurons from attaining exactly the same potential.

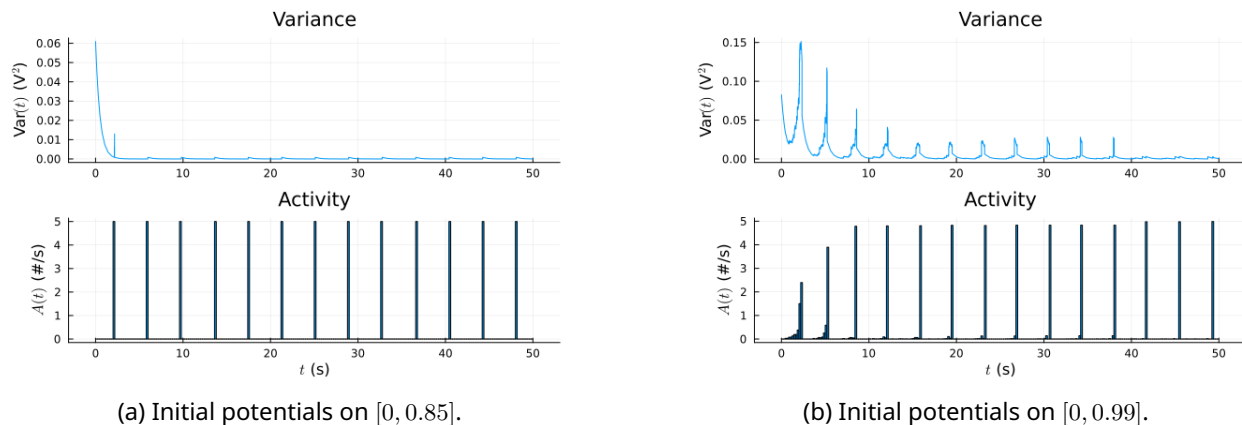


Figure 3.1: Variance of the potentials and activity histogram over time for the leaky model, with a homogeneous population of $N = 1000$ neurons. The constant external input is $I_0 = 0.65$ and the neurons are fully connected with $w_0 = 0.1$. Initial potentials were chosen uniformly at random, on different intervals.

Example 3.2. In contrast to the leaky integrate-and-fire model, the driving force of the exponential integrate-and-fire model is not decreasing in u (see (2.3) or (2.4)). This means that we do not expect synchronization on the basis of Proposition 3.2.

Indeed, simulating the same system as in Example 3.1 under the exponential model (with the parameters adapted following Section 2.3.3) demonstrates that no synchronization occurs, or at least not on the same timescale. This can be seen in Figure 3.2). However, some periodic pattern does seem to emerge in the activity.

In Figure 3.3 we can see the activity on a longer timescale for two different bin sizes. While there is no synchronization, a kind of regularity is undoubtedly present. Based on this pattern, we conjecture that various subgroups of neurons do undergo a form of synchronization, leading to the visually distinct periodic bin patterns we observe. Further research would be necessary to confirm this.

Remark 3.9. If the sharpness parameter Δ_T of this exponential model is decreased, we observe synchronizing behavior similar to Example 3.1. This is because then the exponential integrate-and-fire model approaches the leaky model, as noted in Remark 2.4.

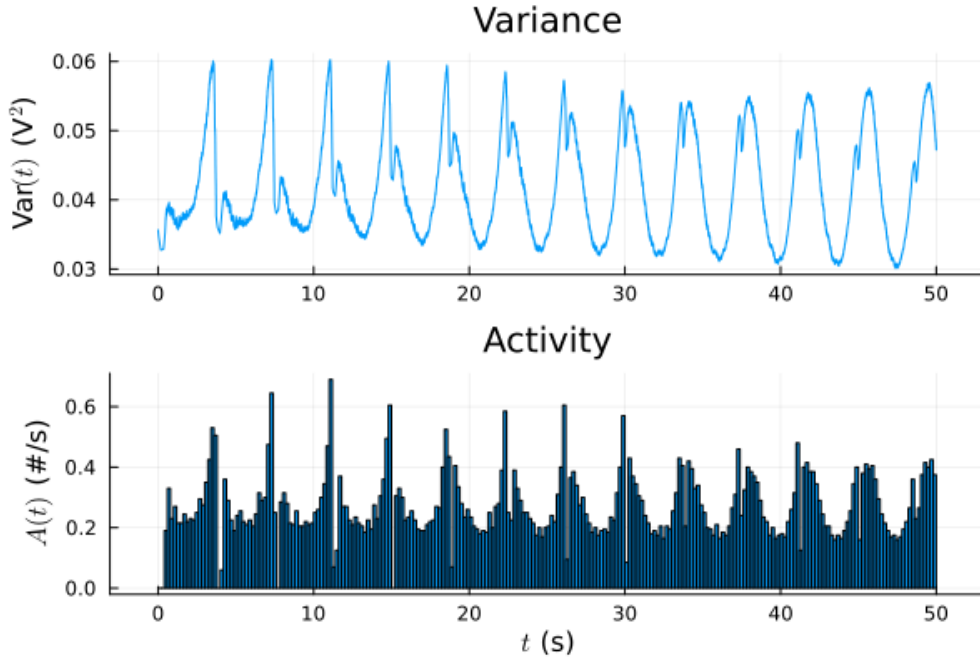


Figure 3.2: Variance of the potentials and activity histogram over time for the exponential model, with a homogeneous population of $N = 1000$ neurons. The constant external input is $I_0 = 0.2$ and the neurons are fully connected with $w_0 = 0.1$. Initial potentials were chosen uniformly at random on $[0, 0.85]$.

3.3.2 Mean-field limit

In this section, we will determine the mean-field limit of a population of neurons that are fully connected. This entails taking the limit $N \rightarrow \infty$ to say something about the behavior of the system when the number of neurons becomes very large. Under certain regularity assumptions, we will derive equations for the evolution of the distribution of the neuron potentials and the population activity in this limit sense.

From now on, we only consider our model on time intervals of the form $I = [t_0, t_e]$, $t_e > t_0$, which ensures finiteness of \mathfrak{T} by Lemma 3.1. Among other benefits, this makes the empirical population activity measure we will define below finite.

Before being able to perform a mean-field analysis we first introduce two relevant measures, starting with the population activity measure. As the name suggests, this measure measures how active the neuronal population is during a certain time period. In this context, activity refers to the number of spikes that occur at or around a given time t . We will stick to a straightforward definition in the form of an empirical measure.

Definition 3.2 (Empirical population activity measure). We define the empirical population activity measure $\mathcal{A}_N : \mathcal{B}_I \rightarrow [0, +\infty)$ as

$$\mathcal{A}_N(dt) := \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{T}_i} \delta_\tau(dt),$$

or equivalently, for $S \in \mathcal{B}_I$,

$$\mathcal{A}_N(S) := \frac{1}{N} \sum_{i=1}^N \#(\mathfrak{T}_i \cap S) = \frac{1}{N} \sum_{\tau \in \mathfrak{T} \cap S} \#\mathfrak{N}(\tau).$$

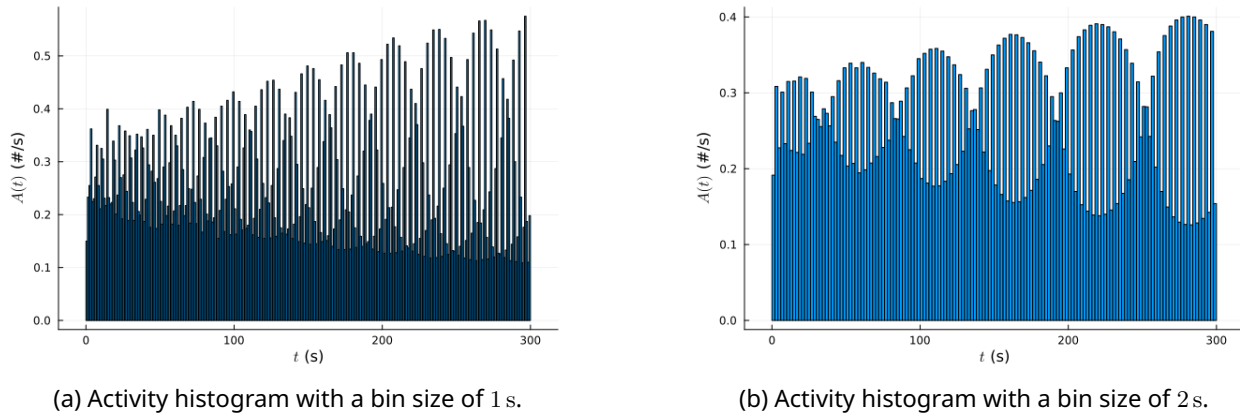


Figure 3.3: Activity histograms over time for the exponential model, with a homogeneous population of $N = 1000$ neurons. The constant external input is $I_0 = 0.2$ and the neurons are fully connected with $w_0 = 0.1$. Initial potentials were chosen uniformly at random on $[0, 0.85]$.

This empirical measure counts the number of neuronal firings in some time period S and scales it by the number of neurons. This scaling is necessary because the idea behind the mean-field limit is that for large N this empirical measure narrowly converges to some finite measure \mathcal{A} . Proving this convergence is outside of the scope of this thesis, so we make the following assumption:

Assumption 3.3. There exists a finite measure $\mathcal{A} : \mathcal{B}_I \rightarrow [0, +\infty)$ to which \mathcal{A}_N converges narrowly. That is, for all test functions $\varphi \in C_b(I)$ we have

$$\lim_{N \rightarrow \infty} \int_I \varphi(\tau) \mathcal{A}_N(d\tau) = \int_I \varphi(\tau) \mathcal{A}(d\tau).$$

We call \mathcal{A} the mean-field population activity measure. We also make a uniformity assumption on this convergence of \mathcal{A}_N . It states that the measure that has Radon-Nikodym derivative $\frac{\#\mathfrak{N}(\tau)}{N}$ w.r.t. \mathcal{A}_N evaluated on I converges to zero, i.e.,

$$\lim_{N \rightarrow \infty} \int_I \frac{\#\mathfrak{N}(\tau)}{N} \mathcal{A}_N(d\tau) = \lim_{N \rightarrow \infty} \int_I \mathcal{A}_N(\{\tau\}) \mathcal{A}_N(d\tau) = 0. \quad (3.24)$$

Remark 3.10. The last part of Assumption 3.3 could for instance be satisfied when the number of firing neurons at one time point grows at most sublinearly as N increases, uniformly over I . That is,

$$\limsup_{N \rightarrow \infty} \sup_{t \in I} \frac{\#\mathfrak{N}(t)}{N} = \limsup_{N \rightarrow \infty} \sup_{t \in I} \mathcal{A}_N(\{t\}) = 0.$$

This is because

$$\int_I \frac{\#\mathfrak{N}(\tau)}{N} \mathcal{A}_N(d\tau) \leq \sup_{t \in I} \frac{\#\mathfrak{N}(t)}{N} \mathcal{A}_N(I),$$

where $\mathcal{A}_N(I)$ converges to $\mathcal{A}(I) < +\infty$ by narrow convergence.

The second measure we are interested in is the potential measure. At every $t \in I$, this measure is the law of the distribution of the potentials in U . Thus, it contains important information on the state of the system: for instance how many neurons are close to the firing threshold. As for the population activity, we will have an empirical version of this measure for which we hope that a limiting measure exists.

Definition 3.4 (Empirical potential measure). The empirical potential measure for $N \in \mathbb{N}$ neurons $\rho_N : I \times \mathcal{B}_U \rightarrow [0, 1]$ is defined as

$$\rho_N(t, du) := \frac{1}{N} \sum_{i=1}^N \delta_{u_i(t)}(du).$$

We assume that this empirical measure converges uniformly narrowly to a mean-field limit version of the measure. The uniformity of this convergence will be needed, because this measure will be integrated against \mathcal{A}_N over time. We will also need some form of uniformity for the convergence of \mathcal{A}_N itself, but this will be part of the conditions of the theorem below.

Assumption 3.5. There exists a time-dependent probability measure $\rho : I \times \mathcal{B}_U \rightarrow [0, 1]$ such that, for all test functions $\varphi \in C_b(I \times U)$ and for all $t \in I$, we have the following uniform narrow convergence

$$\lim_{N \rightarrow \infty} \sup_{t \in I} \left| \int_U \varphi(u) \rho_N(t, du) - \int_U \varphi(u) \rho(t, du) \right| = 0.$$

We call ρ the mean-field potential measure. We also assume that for any $\varphi \in C_b(I \times U)$ the map

$$\tau \mapsto \int_U \varphi(\tau, u) \rho(\tau, du), \quad \tau \in I,$$

is continuous. Note that this implies that this map is also bounded since $I = [t_0, t_e]$ is compact.

Note that this assumption also implies pointwise narrow convergence of ρ_N to ρ , as in for fixed t . Here we are thus also implicitly assuming that the initial potentials $u_i(t_0)$ are distributed with law $\rho(t_0, du)$ as N increases, although they need not be random variables.

Remark 3.11. We will mostly use test functions that are smooth and compactly supported, so for our purposes the assumptions of (uniform) narrow convergence on ρ_N and \mathcal{A}_N are stronger than strictly necessary. However, these stronger assumptions are the most convenient. Weaker versions of the mean-field equations we will derive might be possible under different topologies, e.g., weak, weak*, or vague.

Under these assumptions, we can derive a mean-field limit equation, at least when the firing threshold is finite. Under additional absolute continuity assumptions, we can derive a partial differential equation (PDE) from this equation.

Theorem 3.3. *Assume that the firing threshold V_F is finite. As stated above, we consider a time interval of the form $I = [t_0, t_e]$.*

(i.) *Under Assumption 3.3 and Assumption 3.5, the mean-field potential measure ρ and the mean-field activity measure \mathcal{A} are related by the mean-field equation*

$$\begin{aligned} & \int_U (\varphi(t_e, u) \rho(t_e, du) - \varphi(t_0, u) \rho(t_0, du)) \\ &= \int_I \int_U [\partial_t \varphi(\tau, u) + \partial_u \varphi(\tau, u) f(\tau, u)] \rho(\tau, du) d\tau + \int_I (\varphi(\tau, V_R) - \varphi(\tau, V_F)) \mathcal{A}(d\tau) \\ &+ \int_I \int_U w_0 \partial_u \varphi(\tau, u) \rho(\tau, du) \mathcal{A}(d\tau), \end{aligned} \quad (3.25)$$

for any smooth and compactly supported test function $\varphi \in C_c^\infty(I \times \bar{U})$.

(ii.) *Suppose further that ρ and \mathcal{A} are absolutely continuous w.r.t. the Lebesgue measure, i.e., $\rho(t, du) = p(t, u) du$ and $\mathcal{A}(dt) = A(t) dt$. For the potential density $p : I \times U \rightarrow \mathbb{R}_+$ we assume that it is differentiable everywhere except at $u = V_R$, but still càdlàg in u .⁵ We also assume that $p(t, V_F^-)$ and $f(t, V_F^-)$ exist for all $t \in I$. For the population activity density $A : I \rightarrow \mathbb{R}_+$ we assume that it is continuous. Then the population activity density is given by*

$$A(t) = \frac{f(t, V_F^-) p(t, V_F^-)}{1 - w_0 p(t, V_F^-)} = \frac{f(t, V_R) [p(t, V_R) - p(t, V_R^-)]}{1 - w_0 (p(t, V_R) - p(t, V_R^-))}, \quad t \in I, \quad (3.26)$$

and p satisfies the following transport Cauchy problem:

$$\begin{cases} \partial_t p(t, u) + \partial_u J(t, u) = 0, & (t, u) \in I \times (U \setminus \{V_R\}), \\ J(t, V_R) - J(t, V_R^-) = J(t, V_F^-), & t \in I, \end{cases} \quad (3.27)$$

where the flux $J : I \times U \rightarrow \mathbb{R}$ is defined as

$$J(t, u) := [f(t, u) + w_0 A(t)] p(t, u), \quad (t, u) \in I \times U.$$

Proof. Because this result is a special case of Theorem 3.4, the proof is postponed. ■

⁵The values of $p(\cdot, V_R)$ do not matter, as it is a density of an absolutely continuous measure. Thus we can choose p to be right-continuous here.

In the expression for $A(t)$, which is (3.26), we see that when $p(t, V_F^-)$ is close to $\frac{1}{w_0}$ it is possible for the activity to blow up. This can only happen when $w_0 > 0$, which might correspond with the fact that synchronization can take place for some models in that case (see Section 3.3.1). This syncing would probably violate some of the assumptions we made to derive the second part of Theorem 3.3 over time, for instance the absolute continuity of \mathcal{A} . Thus, the second part of Theorem 3.3 likely does not hold for the leaky model with $w_0 > 0$, or at least not globally on I .

Remark 3.12. Observe that (3.26), (3.27) and the definition of the flux J imply

$$A(t) = J(t, V_F^-) = J(t, V_R) - J(t, V_R^-), \quad t \in I,$$

which is in fact the result from which (3.26) is derived (see (3.48) in the proof of Theorem 3.4).

3.3.3 Stationary solution

After determining the Cauchy problem governing the behavior of the system in the limit of a large number of particles, the natural next step would be to solve this equation. However, even though the PDE in (3.27) looks like a simple first-order linear transport system, it is significantly more complicated due to the nonlinear dependence of $A(t)$ on $p(t, V_F^-)$. We therefore only determine a stationary solution of (3.27), meaning that we assume $p(t, \cdot) = p(\cdot)$ and $f(t, \cdot) = f(\cdot)$ for all $t \in I$, i.e., no time-dependence. The system then simplifies to

$$\begin{cases} \partial_u J(u) = 0, & u \in U \setminus \{V_R\}, \\ J(V_R) - J(V_R^-) = J(V_F^-), \end{cases} \quad (3.28)$$

with

$$J(u) = [f(u) + w_0 A] p(u), \quad u \in U, \quad (3.29)$$

where

$$A = \frac{f(V_F^-) p(V_F^-)}{1 - w_0 p(V_F^-)} = \frac{f(V_R) [p(V_R) - p(V_R^-)]}{1 - w_0 (p(V_R) - p(V_R^-))},$$

is the now constant population activity. The first equation in (3.28) yields

$$J(u) = \begin{cases} J_1 & \text{if } u < V_R, \\ J_2 & \text{if } u \geq V_R, \end{cases}$$

for some constants $J_1, J_2 \in \mathbb{R}$, where we had to take the possible discontinuity at $u = V_R$ into account and the fact that p , and thus also J , is imposed to be càdlàg. The second equation in (3.28) now yields that $J_2 - J_1 = J_2$, so $J_1 = 0$. Furthermore, we know from Remark 3.12 that $J_2 = J(V_F^-) = A$. It follows from (3.29) that

$$p(u) = \frac{J(u)}{f(u) + w_0 A} = \begin{cases} 0 & \text{if } u < V_R, \\ \frac{A}{f(u) + w_0 A} & \text{if } u \geq V_R. \end{cases} \quad (3.30)$$

Because p is a probability density, we must have

$$\int_U p(u) du = 1,$$

so the constant A is constrained by the equation

$$\int_{V_R}^{V_F} \frac{A}{f(u) + w_0 A} du = 1. \quad (3.31)$$

Example 3.3. In case of the nondimensionalized leaky integrate-and-fire model from (2.2) with constant external input I_0 , we have $f(u) = -(u - V_{\text{rest}}) + I_0$ and $V_R = 0, V_F = 1$. Then (3.31) becomes

$$1 = \int_{V_R}^{V_F} \frac{A}{f(u) + w_0 A} du = A \int_0^1 \frac{1}{V_{\text{rest}} + I_0 + w_0 A - u} du = A \ln \left(\frac{V_{\text{rest}} + I_0 + w_0 A}{V_{\text{rest}} + I_0 + w_0 A - 1} \right), \quad (3.32)$$

provided $V_{\text{rest}} + I_0 + w_0 A \notin [0, 1]$. Now, note that for any activity to be expected ($A > 0$) we require $V_{\text{rest}} + I_0 > 1$, else (2.2) will become negative before u reaches $V_F = 1$.

We are primarily interested in the case $w_0 < 0$, because as noted at the end of the previous section the absolutely continuous version of the mean-field limit is unlikely to be valid for $w_0 > 0$. For $w_0 < 0$, the map

$$A \mapsto \frac{V_{\text{rest}} + I_0 + w_0 A}{V_{\text{rest}} + I_0 - 1 + w_0 A}$$

has a vertical asymptote at $A = \frac{1 - V_{\text{rest}} - I_0}{w_0} > 0$, and is increasing for $A \in [0, \frac{1 - V_{\text{rest}} - I_0}{w_0})$. From this, we deduce that the map

$$A \mapsto \ln \left(\frac{V_{\text{rest}} + I_0 + w_0 A}{V_{\text{rest}} + I_0 - 1 + w_0 A} \right)$$

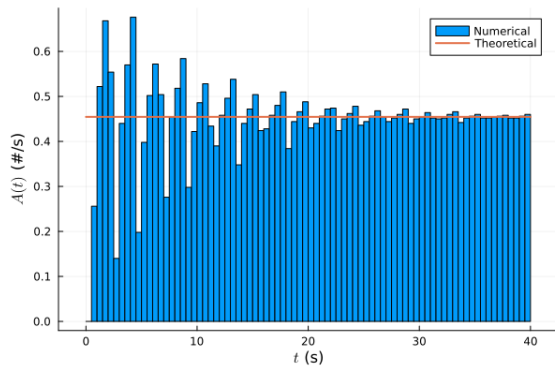
is continuous and increasing on $[0, \frac{1 - V_{\text{rest}} - I_0}{w_0})$, zero at $A = 0$ and diverging to $+\infty$ for $A \uparrow \frac{1 - V_{\text{rest}} - I_0}{w_0}$. It follows that (3.32) has a unique solution $A \in [0, \frac{1 - V_{\text{rest}} - I_0}{w_0})$, which can be determined numerically. Note that this A satisfies $V_{\text{rest}} + I_0 + w_0 A > V_{\text{rest}} + I_0 + 1 - V_{\text{rest}} - I_0 = 1$ which is sufficient for the validity of (3.32).

Taking the standard parameter $V_R = 0.37$ as in Table 2.2 and $I_0 = 0.8$ combined with $w_0 = -0.1$, numerically solving (3.32) yields $A \approx 0.454$ which is indeed less than $\frac{1 - V_{\text{rest}} - I_0}{w_0} = 1.7$. Thus we get from (3.30) that

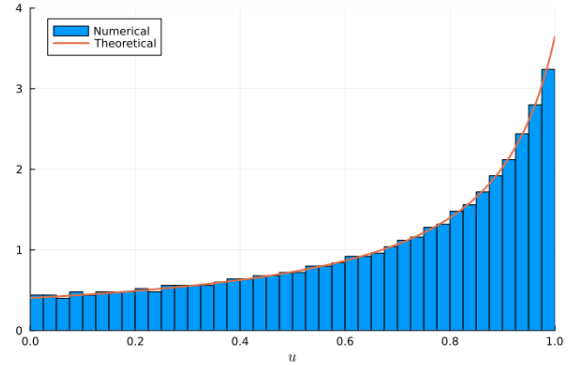
$$p(u) = \begin{cases} 0 & \text{if } u < V_R, \\ \frac{A}{f(u) + w_0 A} & \text{if } u \geq V_R, \end{cases} \approx \begin{cases} 0 & \text{if } u < 0, \\ \frac{0.454}{1.125 - u} & \text{if } u \geq 0, \end{cases} \quad (3.33)$$

In Figure 3.4 we can see the numerical results for these parameters. In Figure 3.4a we see that the activity seems to be converging to the theoretical stationary value $A \approx 0.454$. Furthermore, in Figure 3.4b we see that at the end of the simulation, the numerical potential density agrees closely with the function p from (3.33).

Different values of $w_0 < 0$ and I_0 were simulated, for which the results also appeared to match the theoretical results. This suggests that for $w_0 < 0$ the mean-field equations of Theorem 3.3 not only hold, but also that the system converges to a stationary solution of those equations.



(a) Population activity over time, visualized using a histogram of neuronal firings, versus the theoretical stationary value $A \approx 0.454$.



(b) Potential density histogram at $t_e = 40$ s versus the theoretical stationary density p according to (3.33).

Figure 3.4: Simulation results for the leaky model with a homogeneous population of $N = 1000$ neurons. The constant external input is $I_0 = 0.8$ and the neurons are fully connected with $w_0 = -0.1$. Initial potentials were chosen uniformly at random on $[0, 0.85]$.

The expression in (3.32) might look familiar, as it is quite similar to the gain function found for the leaky model in Example 2.2. This is no coincidence, there is a close relation between the stationary solution in (3.30) and the gain function corresponding to a single neuron (recall Definition 2.6). From Lemma 2.4 we know that the gain function is of the form

$$g(I_0) = \begin{cases} 0 & \text{if } I_0 \leq - \inf_{u \in [V_R, V_F]} f(u), \\ \left(\int_{V_R}^{V_F} \frac{1}{f(u) + I_0} du \right)^{-1} & \text{if } I_0 > - \inf_{u \in [V_R, V_F]} f(u). \end{cases}$$

where $I_0 \in \mathbb{R}$ is constant external input.

Now, observe that

$$g(w_0 A) = A$$

is equivalent to the implicit equation (3.31) for the activity, provided I_0 is large enough. This is no coincidence; since each neuron is identical and receives the same input (both external and from all other neurons) the firing rate of one neuron given by $g(I_0 + w_0 A)$ is equal to the average firing rate A . This relation holds more generally, for instance also in models with noise [see 8, sec. 12.4.2].

3.4 Spatial populations

The assumption of a fully homogeneous population of neurons with synapses of equal weight between all of them is an oversimplification. In this section, we will consider a more realistic version of the model developed in Section 3.1. We will assign to each neuron i a coordinate x_i that determines its parameters and external input. The synaptic connection strength to and from a neuron j will now depend on x_i and x_j .

The coordinates do not necessarily refer (only) to a neuron's physical position in the brain. Namely, x_i could just as well contain information about the parameters of its driving force, or the kind of sensory input neuron i responds to. An example of the latter is the response of neurons in the visual cortex to images in different orientations, which can be used to group the neurons into columns with shared properties [8, pp. 293–294]. However, for simplicity we will use spatial terminology such as “position” and “location”.

3.4.1 Formulation

Adding the spatial component does not require altering the model mechanics as defined in Section 3.1. Rather, we only have to add coordinates to each neuron which determine the weights w_{ij} and the driving force f_i . Let $\Omega \subset \mathbb{R}^d$ be a compact d -dimensional space. As stated above, to every neuron $i \in \{1, \dots, N\}$ we now associate not only a potential $u_i : I \rightarrow U$ but also a coordinate $x_i \in \Omega$.

This location plays a role in determining the strength of the synaptic connections between the neurons, the neuron's characteristics, and the strength of the external input it receives. The last two are encapsulated in the driving force $f_i(t, u) := f(t, x_i, u)$. The function $f : I \times \Omega \times U \rightarrow \mathbb{R}$ is assumed to be continuous, also in its second argument.

The synaptic weight between postsynaptic neuron i and presynaptic neuron j is given by

$$w_{ij} = \frac{1}{N} w(x_i, x_j),$$

where the weight function $w : \Omega^2 \rightarrow [-w_0, w_0]$ represents a certain connectivity pattern. We require this function to be continuous. Here $w_0 \in \mathbb{R}_+$ is a fixed parameter. Note that in contrast to Section 3.3, this parameter is now always nonnegative. We still scale by a factor of $\frac{1}{N}$ to facilitate the mean-field limit we will take in the next section.

Observe that choosing $f(t, x, u) := f(t, u)$ and $w(x, y) := w_0$ recovers the full connectivity model from Section 3.3. Thus, this spatial model is a generalization.

Remark 3.13. For simulation purposes, we will always take $\Omega = [0, 1]^2$. This choice for two dimensions is based on the fact that the cerebral cortex is approximately two-dimensional [7, p. 315].

Remark 3.14. We will not pay much attention to how the neurons are assigned a location in Ω as N increases. This could be either a deterministic or a stochastic process, although we will not consider the x_i 's to be random variables to avoid having to distinguish between sure and almost sure convergence. In our simulation, we will distribute them uniformly at random over Ω , which should at least be more realistic than a fixed grid structure.

Implicitly however, we are assuming that this process is somewhat uniform due to the convergence assumptions we will make on the potential and activity measures in the next section. In Chapter 4 we will need this explicitly, where for Definition 4.1 we will also assume that no pair of neurons shares the same coordinate.

Example 3.4. Consider a homogeneous population of leaky neurons with constant external input. Instead of full connectivity, we now study a Gaussian weight function $w : \Omega^2 \rightarrow [-w_0, w_0]$ of the form

$$w(x, y) := w_0 e^{-\frac{\|x-y\|^2}{2\sigma^2}}, \quad (3.34)$$

where $\sigma > 0$ is a parameter determining how “wide” the function is. Note that this weight function only depends on the distance between the neurons. Thus, it is symmetric.

Because $w_0 > 0$, all weights are positive. Also, all neurons are connected, albeit with negligible synaptic strength if $\|x_i - x_j\|$ is large enough. Thus, the circumstances are somewhat similar to those in Example 3.1, leading us to expect synchronization.

In Figure 3.5 we indeed see this occurring. At $t = 10.30$ we observe that the neurons have already mostly synchronized in local clusters, before reaching global synchronization at $t = 25.55$ where they fire all at the same time.

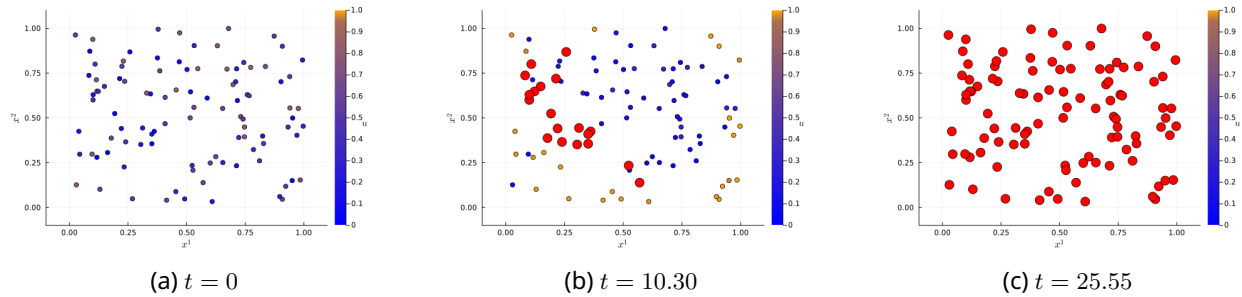


Figure 3.5: Snapshots of the state of a homogeneous population of $N = 100$ leaky neurons, which are distributed uniformly at random over $\Omega = [0, 1]^2$. A color scaling visualizes the current potential of the neurons, and firing neurons are marked with larger red dots. The constant external input is $I_0 = 0.65$ and the neurons are connected with Gaussian weights according to (3.34), with $w_0 = 0.1$ and $\sigma = 0.2$. Initial potentials were chosen uniformly at random on $[0, 0.85]$.

Remark 3.15. If we reduce σ in Example 3.4, we observe that local synchronization still takes place whereas global synchronization does not. There should be a critical value σ_c above which global synchronization occurs, and below which it no longer does. This value likely depends on the number of neurons and their spatial distribution, as these determine the maximum distance between any two neurons.

3.4.2 Mean-field limit

We are again going to examine the mean-field limit of this model, which will yield a more general result than we found in Section 3.3.2. To do so, we first reintroduce the population activity and potential measures which will now also measure space, and their convergence assumptions. This will be entirely analogous to Section 3.3.2, with the assumptions we make here being stronger versions of the ones we already made.

Definition 3.6 (Empirical spatiotemporal population activity measure). The spatiotemporal population activity measure $\mathcal{A}_N : \mathcal{B}_I \otimes \mathcal{B}_\Omega \rightarrow [0, +\infty)$ is defined as

$$\mathcal{A}_N(dt dx) := \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{I}_i} \delta_{(\tau, x_i)}(dt dx) = \frac{1}{N} \sum_{\tau \in \mathfrak{I}} \sum_{i \in \mathfrak{N}(\tau)} \delta_{(\tau, x_i)}(dt dx).$$

Definition 3.7 (Empirical spatiotemporal potential measure). The spatiotemporal potential measure $\rho_N : I \times \mathcal{B}_\Omega \otimes \mathcal{B}_U \rightarrow [0, 1]$ is defined as

$$\rho_N(t, dx du) := \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, u_i(t))}(dx du).$$

Remark 3.16. The notation for the spatiotemporal population activity and potential measures is the same as for the population activity measure from Definition 3.2 and the potential measure from Definition 3.4. This should cause no confusion, as these measures can be distinguished by their arguments.

Assumption 3.8. There exists a time-dependent probability measure $\rho : I \times \mathcal{B}_\Omega \otimes \mathcal{B}_U \rightarrow [0, 1]$ such that we have the following uniform narrow convergence, for all test functions $\varphi \in C_b(I \times \Omega \times U)$,

$$\lim_{N \rightarrow \infty} \sup_{t \in I} \left| \int_\Omega \int_U \varphi(t, x, u) \rho_N(t, dxdu) - \int_\Omega \int_U \varphi(t, x, u) \rho(t, dxdu) \right| = 0.$$

We call ρ the mean-field potential measure. We also assume that for any $\varphi \in C_b(I \times \Omega \times U)$ the map

$$\tau \mapsto \int_\Omega \int_U \varphi(\tau, x, u) \rho(\tau, dxdu), \quad \tau \in I,$$

is continuous. Note that this implies that this map is also bounded since $I = [t_0, t_e]$ is compact.

Assumption 3.9. There exists a finite measure $\mathcal{A} : \mathcal{B}_I \rightarrow [0, +\infty)$ such that \mathcal{A}_N converges narrowly to \mathcal{A} . This means that for all test functions $\varphi \in C_b(I \times \Omega)$ we have

$$\lim_{N \rightarrow \infty} \int_I \int_\Omega \varphi(\tau, x) \mathcal{A}_N(d\tau dx) = \int_I \int_\Omega \varphi(\tau, x) \mathcal{A}(d\tau dx).$$

We call \mathcal{A} the mean-field population activity measure. We again also make the following uniformity assumption on this convergence of \mathcal{A}_N :

$$\lim_{N \rightarrow \infty} \int_I \int_\Omega \frac{\#\mathfrak{N}(\tau)}{N} \mathcal{A}_N(d\tau dx) = \lim_{N \rightarrow \infty} \int_I \int_\Omega \mathcal{A}_N(\{\tau\} \times \Omega) \mathcal{A}_N(d\tau dx) = 0. \quad (3.35)$$

Having reintroduced the relevant measures and assumptions, we can now state the mean-field result for the spatial model. Note that this result is a generalization of Theorem 3.3, so by proving it we will prove that theorem as well.

Theorem 3.4. Assume that the firing threshold V_F is finite.

(i.) Under Assumption 3.9 and Assumption 3.8, the mean-field potential measure ρ and the mean-field activity measure \mathcal{A} are related by the mean-field equation

$$\begin{aligned} & \int_\Omega \int_U (\varphi(t_e, x, u) \rho(t_e, dxdu) - \varphi(t_0, x, u) \rho(t_0, dxdu)) \\ &= \int_I \int_\Omega \int_U [\partial_t \varphi(\tau, x, u) + \partial_u \varphi(\tau, x, u) f(\tau, x, u)] \rho(\tau, dxdu) d\tau + \int_I \int_\Omega (\varphi(\tau, x, V_R) - \varphi(\tau, x, V_F)) \mathcal{A}(d\tau dx) \\ &+ \int_I \int_{\Omega^2} \int_U w(y, x) \partial_u \varphi(\tau, y, u) \rho(\tau, dydu) \mathcal{A}(d\tau dx), \end{aligned} \quad (3.36)$$

for any smooth and compactly supported test function $\varphi \in C_c^\infty(I \times \Omega \times \bar{U})$.

(ii.) Suppose further that ρ and \mathcal{A} are absolutely continuous w.r.t. the Lebesgue measure, i.e., $\rho(t, dxdu) = p(t, x, u) dxdu$ and $\mathcal{A}(dtdx) = A(t, x) dtdx$. For the potential density $p : I \times \Omega \times U \rightarrow \mathbb{R}_+$ we assume that it is differentiable everywhere except at $u = V_R$, but still càdlàg in u and continuous in time at $u = V_R$. We also assume that $p(t, x, V_F^-)$ and $f(t, x, V_F^-)$ exist for all $t \in I$, $x \in \Omega$. For the population activity density $A : I \times \Omega \rightarrow \mathbb{R}_+$ we assume that it is continuous. Then the population activity density is a solution of

$$A(t, x) = \left[f(t, x, V_F^-) + \int_\Omega w(x, y) A(t, y) dy \right] p(t, x, V_F^-), \quad (t, x) \in I \times \Omega,$$

and p satisfies the following transport Cauchy problem:

$$\begin{cases} \partial_t p(t, x, u) + \partial_u J(t, x, u) = 0, & (t, x, u) \in I \times \Omega \times (U \setminus \{V_R\}), \\ J(t, x, V_R) - J(t, x, V_F^-) = J(t, x, V_F^-), & (t, x) \in I \times \Omega, \end{cases} \quad (3.37)$$

where the flux $J : I \times \Omega \times U \rightarrow \mathbb{R}$ is defined as

$$J(t, x, u) := \left[f(t, x, u) + \int_\Omega w(x, y) A(t, y) dy \right] p(t, x, u), \quad (t, x, u) \in I \times \Omega \times U.$$

Notation 3.10. When integrating over Ω twice we often abuse notation by abbreviating this with an integral over Ω^2 , even when the integration is performed against different measures. For an example, see (3.41) in the proof below. This is just for notational conciseness; the integrals should be interpreted as being separate.

In a similar vein, we do not explicitly denote it when we integrate w.r.t. product measures, but instead simply juxtapose the component measures. Note that Fubini's Theorem will usually apply in these cases, because we will often integrate continuous and compactly supported functions against finite measures, so this should not cause confusion.

Proof. We consider the system (3.11), with $N \in \mathbb{N}$ fixed for now. Consider the set of firing times $\{r_1, \dots, r_M\} = \mathfrak{T}$, which is finite as a consequence of Lemma 3.1. For notational convenience also define $r_0 := t_0$ and $r_{M+1} := t_e$.

Now observe that on all time intervals (r_k, r_{k+1}) , $k \in \{0, \dots, M\}$ no neurons fire, meaning that $u_i(\tau)$ is differentiable with derivative $f(\tau, x_i, u_i(\tau))$. Using this we obtain for any neuron $i \in \{1, \dots, N\}$ and any test function $\varphi \in C_c^\infty(I \times \Omega \times \bar{U})$,

$$\begin{aligned}
 & \varphi(t_e, x_i, u_i(t_e)) - \varphi(t_0, x_i, u_i(t_0)) \\
 &= \sum_{k=0}^M (\varphi(r_{k+1}, x_i, u_i(r_{k+1})) - \varphi(r_k, x_i, u_i(r_k))), \\
 &= \sum_{k=0}^M (\varphi(r_{k+1}, x_i, u_i(r_{k+1}^-)) - \varphi(r_k, x_i, u_i(r_k)) + \varphi(r_{k+1}, x_i, u_i(r_{k+1})) - \varphi(r_{k+1}, x_i, u_i(r_{k+1}^-))), \\
 &= \sum_{k=0}^M (\varphi(r_{k+1}, x_i, u_i(r_{k+1}^-)) - \varphi(r_k, x_i, u_i(r_k))) + \sum_{k=0}^M (\varphi(r_{k+1}, x_i, u_i(r_{k+1})) - \varphi(r_{k+1}, x_i, u_i(r_{k+1}^-))), \\
 &\stackrel{(*)}{=} \sum_{k=0}^M \int_{r_k}^{r_{k+1}} \frac{d}{d\tau} \varphi(\tau, x_i, u_i(\tau)) d\tau + \sum_{\tau \in \mathfrak{T}} (\varphi(\tau, x_i, u_i(\tau)) - \varphi(\tau, x_i, u_i(\tau^-))), \\
 &\stackrel{(**)}{=} \int_I [\partial_t \varphi(\tau, x_i, u_i(\tau)) + \partial_u \varphi(\tau, x_i, u_i(\tau)) f(\tau, x_i, u_i(\tau))] d\tau \\
 &\quad + \sum_{\tau \in \mathfrak{T}_i} (\varphi(\tau, x_i, u_i(\tau)) - \varphi(\tau, x_i, u_i(\tau^-))) + \sum_{\tau \in \mathfrak{T} \setminus \mathfrak{T}_i} (\varphi(\tau, x_i, u_i(\tau)) - \varphi(\tau, x_i, u_i(\tau^-))), \\
 &= \int_{t_0}^{t_e} [\partial_t \varphi(\tau, x_i, u_i(\tau)) + \partial_u \varphi(\tau, x_i, u_i(\tau)) f(\tau, x_i, u_i(\tau))] d\tau + \sum_{\tau \in \mathfrak{T}_i} (\varphi(\tau, x_i, V_R + \Delta_i^{\text{post}}(\tau)) - \varphi(\tau, x_i, V_F - \Delta_i^{\text{pre}}(\tau))) \\
 &\quad + \sum_{\tau \in \mathfrak{T} \setminus \mathfrak{T}_i} \left[\varphi \left(\tau, x_i, u_i(\tau^-) + \sum_{j \in \mathfrak{N}(\tau)} w_{ij} \right) - \varphi(\tau, x_i, u_i(\tau^-)) \right].
 \end{aligned} \tag{3.38}$$

In step (*) we used the fact that φ is smooth and u_i is càdlàg. Note that this equality also holds when $r_M = r_{M+1} = t_e$, because then the final terms in the two summations of the previous line cancel. Afterwards in step (**), we used that $\mathfrak{T} = \{r_1, \dots, r_M\}$ is a finite set, and thus a (Lebesgue) null set, to combine the $M + 1$ integrals in the first term into one integral.

We introduced two new symbols in (3.38), which are

$$\Delta_i^{\text{post}}(\tau) := \sum_{j \in \bigcup_{n=\ell_i(\tau)+1}^N \mathfrak{N}^n(\tau)} w_{ij}, \quad \text{and} \quad 0 \leq \Delta_i^{\text{pre}}(\tau) \leq \sum_{j \in \bigcup_{n=1}^{\ell_i(\tau)-1} \mathfrak{N}^n(\tau)} w_{ij},$$

where $\Delta_i^{\text{pre}}(\tau)$ is such that $u_i(\tau^-) + \Delta_i^{\text{pre}}(\tau) = V_F$. These are error terms after and before the spike, respectively. Note that $\Delta_i^{\text{post}}(\tau)$ can be negative.

The derivation in (3.38) holds in general for the model formulated in Section 3.1. If we substitute our connectivity assumption $w_{ij} = \frac{1}{N}w(x_i, x_j)$, we get

$$\begin{aligned} & \varphi(t_e, x_i, u_i(t_e)) - \varphi(t_0, x_i, u_i(t_0)) \\ &= \int_I [\partial_t \varphi(\tau, x_i, u_i(\tau)) + \partial_u \varphi(\tau, x_i, u_i(\tau)) f(\tau, x_i, u_i(\tau))] d\tau + \sum_{\tau \in \mathfrak{I}_i} (\varphi(\tau, x_i, V_R + \Delta_i^{\text{post}}(\tau)) - \varphi(\tau, x_i, V_F - \Delta_i^{\text{pre}}(\tau))) \\ &+ \sum_{\tau \in \mathfrak{I} \setminus \mathfrak{I}_i} \left(\varphi \left(\tau, x_i, u_i(\tau^-) + \sum_{j \in \mathfrak{N}(\tau)} \frac{1}{N} w(x_i, x_j) \right) - \varphi(\tau, x_i, u_i(\tau^-)) \right). \end{aligned}$$

We now combine this expression with the empirical potential measure to obtain

$$\begin{aligned} & \int_{\Omega} \int_U (\varphi(t_e, x, u) \rho_N(t_e, dxdu) - \varphi(t_0, x, u) \rho_N(t_0, dxdu)) \\ &= \frac{1}{N} \sum_{i=1}^N (\varphi(t_e, x_i, u_i(t_e)) - \varphi(t_0, x_i, u_i(t_0))), \\ &= \frac{1}{N} \sum_{i=1}^N \left[\int_I [\partial_t \varphi(\tau, x_i, u_i(\tau)) + \partial_u \varphi(\tau, x_i, u_i(\tau)) f(\tau, x_i, u_i(\tau))] d\tau \right. \\ &+ \sum_{\tau \in \mathfrak{I}_i} (\varphi(\tau, x_i, V_R + \Delta_i^{\text{post}}(\tau)) - \varphi(\tau, x_i, V_F - \Delta_i^{\text{pre}}(\tau))) \\ &+ \left. \sum_{\tau \in \mathfrak{I} \setminus \mathfrak{I}_i} \left(\varphi \left(\tau, x_i, u_i(\tau^-) + \sum_{j \in \mathfrak{N}(\tau)} \frac{1}{N} w(x_i, x_j) \right) - \varphi(\tau, x_i, u_i(\tau^-)) \right) \right], \\ &= \underbrace{\int_I \int_{\Omega} \int_U [\partial_t \varphi(\tau, x, u) f(\tau, x, u) + \partial_u \varphi(\tau, x, u)] \rho_N(\tau, dxdu) d\tau}_{\text{(I)}} \\ &+ \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{I}_i} (\varphi(\tau, x_i, V_R + \Delta_i^{\text{post}}(\tau)) - \varphi(\tau, x_i, V_F - \Delta_i^{\text{pre}}(\tau))) \\ &+ \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{I} \setminus \mathfrak{I}_i} \left(\varphi \left(\tau, x_i, u_i(\tau^-) + \sum_{j \in \mathfrak{N}(\tau)} \frac{1}{N} w(x_i, x_j) \right) - \varphi(\tau, x_i, u_i(\tau^-)) \right). \end{aligned}$$

We will take the mean-field limit $N \rightarrow \infty$ for each of these three terms separately.

- I. This is the term corresponding to the driving force f . It has a straightforward limit using the narrow convergence of the empirical measure ρ_N to ρ , which is implied by Assumption 3.8. Namely,

$$\begin{aligned} & \lim_{N \rightarrow \infty} \int_I \int_{\Omega} \int_U [\partial_t \varphi(\tau, x, u) + \partial_u \varphi(\tau, x, u) f(\tau, x, u)] \rho_N(\tau, dxdu) d\tau \\ &= \int_I \int_{\Omega} \int_U [\partial_t \varphi(\tau, x, u) + \partial_u \varphi(\tau, x, u) f(\tau, x, u)] \rho(\tau, dxdu) d\tau. \end{aligned} \quad (3.39)$$

Note that f is not necessarily bounded. However, due to the multiplication with $\partial_u \varphi$, which has compact support, the product is a continuous and compactly supported function, and thus bounded. Thus, we can indeed apply Assumption 3.8.

- II. This second term arises due to the potential resets. We apply a Taylor expansion with Lagrange remainder (Theorem A.3), which is possible because $\varphi \in C_c^\infty(I \times \Omega \times \bar{U})$. This gives us

$$\begin{aligned} \varphi(\tau, x_i, V_F - \Delta_i^{\text{pre}}(\tau)) &= \varphi(\tau, x_i, V_F) - \Delta_i^{\text{pre}}(\tau) \partial_u \varphi(\tau, x_i, V_F - \xi_i^{\text{pre}}(\tau)), \\ \varphi(\tau, x_i, V_R + \Delta_i^{\text{post}}(\tau)) &= \varphi(\tau, x_i, V_R) + \Delta_i^{\text{post}}(\tau) \partial_u \varphi(\tau, x_i, V_R + \xi_i^{\text{post}}(\tau)), \end{aligned}$$

where $\xi_i^{\text{pre}}(\tau) \in [0, \Delta_i^{\text{pre}}(\tau)] \subset [0, w_0 \frac{\#\mathfrak{N}(\tau)-1}{N}]$ and similarly $|\xi_i^{\text{post}}(\tau)| \leq w_0 \frac{\#\mathfrak{N}(\tau)-1}{N}$.

Using this expansion we obtain

$$\begin{aligned}
 & \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{I}_i} (\varphi(\tau, x_i, V_R + \Delta_i^{\text{post}}(\tau)) - \varphi(\tau, x_i, V_F - \Delta_i^{\text{pre}}(\tau))) \\
 &= \frac{1}{N} \sum_{\tau \in \mathfrak{I}_i} (\varphi(\tau, x_i, V_R) - \varphi(\tau, x_i, V_F) + \Delta_i^{\text{post}}(\tau) \partial_u \varphi(\tau, x_i, V_R + \xi_i^{\text{post}}(\tau)) + \Delta_i^{\text{pre}}(\tau) \partial_u \varphi(\tau, x_i, V_F - \xi_i^{\text{pre}}(\tau))), \\
 &= \int_I \int_{\Omega} (\varphi(\tau, x, V_R) - \varphi(\tau, x, V_F)) \mathcal{A}_N(d\tau dx) \\
 &+ \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{I}_i} (\Delta_i^{\text{post}}(\tau) \partial_u \varphi(\tau, x_i, V_R + \xi_i^{\text{post}}(\tau)) + \Delta_i^{\text{pre}}(\tau) \partial_u \varphi(\tau, x_i, V_F - \xi_i^{\text{pre}}(\tau))).
 \end{aligned}$$

Because of the narrow convergence of \mathcal{A}_N to \mathcal{A} , the limit of the first term is clear. We claim that the second term vanishes in the limit of $N \rightarrow \infty$. For this, we require the convergence assumption we made on \mathcal{A}_N in (3.35). Also, we use that $\partial_u \varphi$ is bounded by virtue of being continuous and compactly supported, and both Δ_i -factors are bounded by $w_0 \frac{\#\mathfrak{N}(\tau)}{N}$. This yields

$$\begin{aligned}
 & \left| \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{I}_i} (\Delta_i^{\text{post}}(\tau) \partial_u \varphi(\tau, x_i, V_R + \xi_i^{\text{post}}(\tau)) + \Delta_i^{\text{pre}}(\tau) \partial_u \varphi(\tau, x_i, V_F - \xi_i^{\text{pre}}(\tau))) \right| \\
 & \leq \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{I}_i} 2w_0 \frac{\#\mathfrak{N}(\tau)}{N} \|\partial_u \varphi\|_{\infty}, \\
 & = 2w_0 \|\partial_u \varphi\|_{\infty} \int_I \int_{\Omega} \frac{\#\mathfrak{N}(\tau)}{N} \mathcal{A}_N(d\tau dx).
 \end{aligned} \tag{3.40}$$

This upper bound converges to zero because of (3.35) in Assumption 3.9. It follows that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{I}_i} (\Delta_i^{\text{post}}(\tau) \partial_u \varphi(\tau, x_i, V_R + \xi_i^{\text{post}}(\tau)) + \Delta_i^{\text{pre}}(\tau) \partial_u \varphi(\tau, x_i, V_F - \xi_i^{\text{pre}}(\tau))) = 0.$$

Thus,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{I}_i} (\varphi(\tau, x_i, V_R + \Delta_i^{\text{post}}(\tau)) - \varphi(\tau, x_i, V_F - \Delta_i^{\text{pre}}(\tau))) = \int_I \int_{\Omega} (\varphi(\tau, x, V_R) - \varphi(\tau, x, V_F)) \mathcal{A}(d\tau dx).$$

III. We consider the last term now, which stems from the spikes between the neurons. Using the fact that $\varphi \in C_c^\infty(I \times \Omega \times \bar{U})$, we again Taylor expand with Lagrange remainder. We do so around $u_i(\tau)$ instead of $u_i(\tau^-)$ to prevent possible issues with this left-limit. This yields

$$\begin{aligned}
 & \varphi \left(\tau, x_i, u_i(\tau^-) + \sum_{j \in \mathfrak{N}(\tau)} \frac{1}{N} w(x_i, x_j) \right) - \varphi(\tau, x_i, u_i(\tau^-)) \\
 &= \varphi(\tau, x_i, u_i(\tau)) - \varphi \left(\tau, x_i, u_i(\tau) - \sum_{j \in \mathfrak{N}(\tau)} \frac{1}{N} w(x_i, x_j) \right), \\
 &= \frac{1}{N} \left(\sum_{j \in \mathfrak{N}(\tau)} w(x_i, x_j) \partial_u \varphi(\tau, x_i, u_i(\tau)) - \underbrace{\frac{1}{2} \partial_u^2 \varphi(\tau, x_i, u_i(\tau))}_{=:\kappa_i(\tau)} \left(\sum_{j \in \mathfrak{N}(\tau)} w(x_i, x_j) \right)^2 \right), \\
 &= \frac{1}{N} \left(\sum_{j \in \mathfrak{N}(\tau)} w(x_i, x_j) \partial_u \varphi(\tau, x_i, u_i(\tau)) - \frac{1}{N} \kappa_i(\tau) \zeta_i(\tau)^2 \right),
 \end{aligned}$$

with $|\eta_i(\tau)| \leq \frac{1}{N} |\zeta_i(\tau)| \leq w_0 \frac{\#\mathfrak{N}(\tau)}{N}$. Note that $\kappa_i(\tau)$ is uniformly bounded for all i and τ , because $\varphi \in C_c^\infty(I \times \Omega \times \bar{U})$. Thus

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{T} \setminus \mathfrak{T}_i} \left(\varphi \left(\tau, x_i, u_i(\tau^-) + \sum_{j \in \mathfrak{N}(\tau)} \frac{1}{N} w(x_i, x_j) \right) - \varphi(\tau, x_i, u_i(\tau^-)) \right) \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{\tau \in \mathfrak{T} \setminus \mathfrak{T}_i} \left(\sum_{j \in \mathfrak{N}(\tau)} w(x_i, x_j) \partial_u \varphi(\tau, x_i, u_i(\tau)) - \frac{1}{N} \kappa_i(\tau) \zeta_i(\tau)^2 \right), \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{\tau \in \mathfrak{T}} \sum_{j \in \mathfrak{N}(\tau)} w(x_i, x_j) \partial_u \varphi(\tau, x_i, u_i(\tau)) - \frac{1}{N^2} \sum_{i=1}^N \sum_{\tau \in \mathfrak{T}_i} \zeta_i(\tau) \partial_u \varphi(\tau, x_i, u_i(\tau)) \\ &\quad - \frac{1}{N^2} \sum_{i=1}^N \sum_{\tau \in \mathfrak{T} \setminus \mathfrak{T}_i} \kappa_i(\tau) \frac{\zeta_i(\tau)^2}{N}. \end{aligned}$$

In this expression, the second and third terms vanish for $N \rightarrow \infty$. This can be shown with a similar argument to the one used for the reset term (II) in (3.40), using (3.35) combined with $|\zeta_i(\tau)| \leq w_0 \frac{\#\mathfrak{N}(\tau)}{N}$ and the fact that $\partial_u \varphi(\tau, x_i, u_i(\tau))$ and $\kappa_i(\tau)$ are bounded uniformly over i and τ .

Therefore, we only have to consider the first term. Taking note of Notation 3.10, we can rewrite this term in terms of empirical measures as

$$\begin{aligned} \frac{1}{N^2} \sum_{i=1}^N \sum_{\tau \in \mathfrak{T}} \sum_{j \in \mathfrak{N}(\tau)} w(x_i, x_j) \partial_u \varphi(\tau, x_i, u_i(\tau)) &= \frac{1}{N} \sum_{\tau \in \mathfrak{T}} \sum_{j \in \mathfrak{N}(\tau)} \int_{\Omega} \int_U w(y, x_j) \partial_u \varphi(\tau, y, u) \rho_N(\tau, dxdu), \\ &= \int_I \int_{\Omega^2} \int_U w(y, x) \partial_u \varphi(\tau, y, u) \rho_N(\tau, dydu) \mathcal{A}_N(d\tau dx), \\ &\rightarrow \int_I \int_{\Omega^2} \int_U w(y, x) \partial_u \varphi(\tau, y, u) \rho(\tau, dydu) \mathcal{A}(d\tau dx), \quad \text{as } N \rightarrow \infty. \end{aligned} \tag{3.41}$$

where the convergence follows from Lemma B.4, because of the uniform narrow convergence of the empirical potential according to Assumption 3.8, the continuity and boundedness of w , and narrow convergence of the population activity measure according to Assumption 3.9.

Combining the three terms from (I), (II) and (III), the following mean-field equation follows from the uniformly narrow convergence of ρ_N from Assumption 3.8

$$\begin{aligned} & \int_{\Omega} \int_U (\varphi(t_e, x, u) \rho(t_e, dxdu) - \varphi(t_0, x, u) \rho(t_0, dxdu)) \\ &= \lim_{N \rightarrow \infty} \left(\int_{\Omega} \int_U \varphi(t_e, x, u) \rho_N(t_e, dxdu) - \int_{\Omega} \int_U \varphi(t_0, x, u) \rho_N(t_0, dxdu) \right), \\ &= \int_I \int_{\Omega} \int_U [\partial_t \varphi(\tau, x, u) + \partial_u \varphi(\tau, x, u) f(\tau, x, u)] \rho(\tau, dxdu) d\tau + \int_I \int_{\Omega} (\varphi(\tau, x, V_R) - \varphi(\tau, x, V_F)) \mathcal{A}(d\tau dx) \\ &\quad + \int_I \int_{\Omega^2} \int_U w(y, x) \partial_u \varphi(\tau, y, u) \rho(\tau, dydu) \mathcal{A}(d\tau dx), \end{aligned} \tag{3.42}$$

for any $\varphi \in C_c^\infty(I \times \Omega \times \bar{U})$. This proves part (i.) of the theorem.

To prove part (ii.) we use the assumptions of absolute continuity w.r.t. the Lebesgue measure, viz. $\rho(\tau, dxdu) = p(\tau, x, u) dxdu$ and $\mathcal{A}(d\tau dx) = A(\tau, x) d\tau dx$. This reduces (3.42) to

$$\begin{aligned} & \int_{\Omega} \int_U (\varphi(t_e, x, u) p(t_e, x, u) - \varphi(t_0, x, u) p(t_0, x, u)) dudx \\ &= \int_I \int_{\Omega} \int_U [\partial_t \varphi(\tau, x, u) + \partial_u \varphi(\tau, x, u) f(\tau, x, u)] p(\tau, x, u) dudxd\tau + \int_I \int_{\Omega} (\varphi(\tau, x, V_R) - \varphi(\tau, x, V_F)) A(\tau, x) dx d\tau \\ &\quad + \int_I \int_{\Omega^2} \int_U w(y, x) \partial_u \varphi(\tau, y, u) p(\tau, y, u) A(\tau, x) dudydxd\tau. \end{aligned} \tag{3.43}$$

We now stipulate that φ is separable as $\varphi(\tau, x, u) = \psi(\tau)\phi(x, u)$, where $\psi \in C_c^\infty(I)$ and $\phi \in C_c^\infty(\Omega \times \bar{U})$. Substituting this into (3.43) yields

$$\begin{aligned} & \int_{\Omega} \int_U (\psi(t_e)\phi(x, u)p(t_e, x, u) - \psi(t_0)\phi(x, u)p(t_0, x, u)) \, dudx \\ &= \int_I \int_{\Omega} \int_U [\psi'(\tau)\phi(x, u) + \psi(\tau)\partial_u\phi(x, u)f(\tau, x, u)]p(\tau, x, u) \, dudxd\tau + \int_I \int_{\Omega} \psi(\tau) (\psi(x, V_R) - \psi(x, V_F)) A(\tau, x) \, dx d\tau \\ &+ \int_I \int_{\Omega^2} \int_U \psi(\tau)\partial_u\phi(y, u)w(y, x)p(\tau, y, u)A(\tau, x) \, dudydx d\tau. \end{aligned} \tag{3.44}$$

As we assumed that p is differentiable in its first argument, and since $I = [t_0, t_e]$, we can perform partial integration (p.i.) on the first term of the integrand of the first term in (3.44) as follows

$$\begin{aligned} \int_I \int_{\Omega} \int_U \psi'(\tau)\phi(x, u)p(\tau, x, u) \, dudxd\tau &\stackrel{\text{p.i.}}{=} \int_{\Omega} \int_U (\psi(t_e)\phi(x, u)p(t_e, x, u) - \psi(t_0)\phi(x, u)p(t_0, x, u)) \, dudx, \\ &- \int_I \int_{\Omega} \int_U \psi(\tau)\phi(x, u)\partial_t p(\tau, x, u) \, dudxd\tau. \end{aligned}$$

Here we observe that the first term is the same as the left-hand side of (3.44). Cancellation of these terms allows us to simplify (3.44) to

$$\begin{aligned} & \int_I \int_{\Omega} \int_U \psi(\tau)\phi(x, u)\partial_t p(\tau, x, u) \, dudxd\tau \\ &= \int_I \int_{\Omega} \int_U \psi(\tau)\partial_u\phi(x, u)f(\tau, x, u)p(\tau, x, u) \, dudxd\tau + \int_I \int_{\Omega} \psi(\tau) (\psi(x, V_R) - \psi(x, V_F)) A(\tau, x) \, dx d\tau \\ &+ \int_I \int_{\Omega^2} \int_U \psi(\tau)\partial_u\phi(y, u)w(y, x)p(\tau, y, u)A(\tau, x) \, dudydx d\tau. \end{aligned}$$

Note that in this expression we are integrating all terms over I and multiplying all integrands with $\psi \in C_c^\infty(I)$ which is arbitrary. Also, f , p , and A are all assumed to be continuous in time. This implies that we have

$$\begin{aligned} & \int_{\Omega} \int_U \phi(x, u)\partial_t p(\tau, x, u) \, dudx \\ &= \int_{\Omega} \int_U \partial_u\phi(x, u)f(\tau, x, u)p(\tau, x, u) \, dudx + \int_{\Omega} (\psi(x, V_R) - \psi(x, V_F)) A(\tau, x) \, dx \\ &+ \int_{\Omega^2} \int_U \partial_u\phi(y, u)w(y, x)p(\tau, y, u)A(\tau, x) \, dudydx, \end{aligned} \tag{3.45}$$

for all $\tau \in I$, which follows by concentrating ψ around a fixed time τ .

Because of the resets to V_R , we expect p to have a discontinuity at $u = V_R$. Thus, we distinguish between $p(\cdot, \cdot, V_R^-)$ and $p(\cdot, \cdot, V_R^+) = p(\cdot, \cdot, V_R)$ (p is assumed càdlàg). We again perform partial integration on (3.45) but now on the potential integral over U , where we take into account the aforementioned discontinuity at $u = V_R$. This yields for all $\tau \in I$ that

$$\begin{aligned}
 & \int_{\Omega} \int_U \phi(x, u) \partial_t p(\tau, x, u) \, dudx \\
 &= \int_{\Omega} \int_U \partial_u \phi(x, u) f(\tau, x, u) p(\tau, x, u) \, dudxd\tau + \int_{\Omega} (\phi(x, V_R) - \phi(x, V_F)) A(\tau, x) dx \\
 & \quad + \int_{\Omega^2} \int_U \partial_u \phi(x, u) w(x, y) p(\tau, x, u) A(\tau, y) \, dudxdy, \\
 &= \int_{\Omega} \left(\int_U \partial_u \phi(x, u) \left[f(\tau, x, u) + \int_{\Omega} w(x, y) A(\tau, y) \, dy \right] p(\tau, x, u) \, du + (\phi(x, V_R) - \phi(x, V_F)) A(\tau, x) \right) dx, \\
 &= \int_{\Omega} \left(\int_U \partial_u \phi(x, u) J(\tau, x, u) \, du + (\phi(x, V_R) - \phi(x, V_F)) A(\tau, x) \right) dx, \\
 &= \int_{\Omega} \left(\int_{-\infty}^{V_R} \partial_u \phi(x, u) J(\tau, x, u) \, du + \int_{V_R}^{V_F} \partial_u \phi(x, u) J(\tau, x, u) \, du + (\phi(x, V_R) - \phi(x, V_F)) A(\tau, x) \right) dx, \\
 &\stackrel{\text{p.i.}}{=} \int_{\Omega} \left([\phi(x, u) J(\tau, x, u)]_{u \rightarrow -\infty}^{u=V_R^-} + [\phi(x, u) J(\tau, x, u)]_{u=V_R^+}^{u=V_F^-} - \int_{-\infty}^{V_R} \phi(x, u) \partial_u J(\tau, x, u) \, du \right. \\
 & \quad \left. - \int_{V_R}^{V_F} \phi(x, u) \partial_u J(\tau, x, u) \, du + (\phi(x, V_R) - \phi(x, V_F)) A(\tau, x) \right) dx, \\
 &= \int_{\Omega} \left(\phi(x, V_F) [J(\tau, x, V_F^-) - A(\tau, x)] + \phi(x, V_R) [J(\tau, x, V_R^-) - J(\tau, x, V_R) + A(\tau, x)] \right. \\
 & \quad \left. - \int_U \phi(x, u) \partial_u J(\tau, x, u) \, du \right) dx.
 \end{aligned} \tag{3.46}$$

Note that possible issues with integrability in this derivation are avoided because our test function φ is compactly supported. This is also why the boundary condition at $u \rightarrow -\infty$ vanished after the partial integration.

As (3.46) holds for all $\phi \in C_c^\infty(\Omega \times \bar{U})$ and J , $\partial_u J$ and A are assumed to be continuous (except at $u = V_R$), it follows that

$$\forall t \in I, x \in \Omega, u \in U \setminus \{V_R\} : \partial_t p(t, x, u) = -\partial_u J(t, x, u). \tag{3.47}$$

Since (3.47) holds \mathcal{L} -a.e. on $I \times \Omega \times U$ the inner U -integrals cancel on both sides of (3.46), so it follows that for all $\phi \in C_c^\infty(\Omega \times \bar{U})$,

$$\int_{\Omega} (\phi(x, V_F) [J(\tau, x, V_F^-) - A(\tau, x)] + \phi(x, V_R) [J(\tau, x, V_R^-) - J(\tau, x, V_R) + A(\tau, x)]) \, dx = 0.$$

By choosing $\phi(\cdot, V_F)$ or $\phi(\cdot, V_R)$ to be arbitrarily concentrated around some $x \in \Omega$, we find that

$$\begin{aligned}
 J(t, x, V_F^-) &= A(t, x), & (t, x) \in I \times \Omega, \\
 J(t, x, V_R) - J(t, x, V_R^-) &= A(t, x), & (t, x) \in I \times \Omega.
 \end{aligned} \tag{3.48}$$

Writing this without the a priori unknown activity gives

$$J(t, x, V_R) - J(t, x, V_R^-) = J(t, x, V_F^-), \quad (t, x) \in I \times \Omega.$$

Also, the population activity density thus satisfies the equation

$$A(t, x) = J(t, x, V_F^-) = \left[f(t, x, V_F^-) + \int_{\Omega} w(x, y) A(t, y) \, dy \right] p(t, x, V_F^-), \quad (t, x) \in I \times \Omega.$$

Similarly, using the continuity of f ,

$$\begin{aligned}
 A(t, x) &= J(t, x, V_R) - J(t, x, V_R^-), \\
 &= \left[f(t, x, V_R) + \int_{\Omega} w(x, y) A(t, y) dy \right] p(t, x, V_R) - \left[f(t, x, V_R^-) + \int_{\Omega} w(x, y) A(t, y) dy \right] p(t, x, V_R^-), \\
 &= f(t, x, V_R) [p(t, x, V_R) - p(t, x, V_R^-)] + (p(t, x, V_R) - p(t, x, V_R^-)) \int_{\Omega} w(x, y) A(t, y) dy, \\
 &= \left[f(t, x, V_R) + \int_{\Omega} w(x, y) A(t, y) dy \right] (p(t, x, V_R) - p(t, x, V_R^-)).
 \end{aligned}$$

This shows that part (ii.) of the theorem also holds, concluding the proof. ■

3.4.3 Stationary solution

Entirely analogous to Section 3.3.3, we will determine the equations corresponding to a stationary solution of the mean-field equations (3.37). We thus assume $p(t, \cdot, \cdot) = p(\cdot, \cdot)$ and $f(t, \cdot, \cdot) = f(\cdot, \cdot)$ for all $t \in I$, i.e., no time-dependence. The system then simplifies to

$$\begin{cases} \partial_u J(x, u) = 0, & (x, u) \in \Omega \times (U \setminus \{V_R\}), \\ J(x, V_R) - J(x, V_R^-) = J(x, V_R^-), \end{cases}$$

with

$$J(x, u) = \left[f(x, u) + \int_{\Omega} w(x, y) A(t, y) dy \right] p(x, u), \quad (x, u) \in \Omega \times U.$$

By the same argument as in Section 3.3.3, we find that for all $x \in \Omega$,

$$p(x, u) = \frac{J(x, u)}{f(x, u) + \int_{\Omega} w(x, y) A(y) dy} = \begin{cases} 0 & \text{if } u < V_R, \\ \frac{A(x)}{f(x, u) + \int_{\Omega} w(x, y) A(y) dy} & \text{if } u \geq V_R. \end{cases}$$

where the function A is constrained by the equation

$$\int_{\Omega} \int_{V_R}^{V_F} \frac{A(x)}{f(x, u) + \int_{\Omega} w(x, y) A(y) dy} du dx = 1. \tag{3.49}$$

In contrast to Section 3.3.3 however, it is difficult to find solutions to (3.49), even for simple f and w .

Chapter 4

Learning populations

In the previous two chapters, we have developed a model of individual neurons that together form an interacting population. We will now add the missing piece: making this network learn. To achieve this, we are going to model synaptic plasticity. This refers to the process that changes the connectivity pattern between the neurons over time, by growing new synapses or adapting the strength of existing synapses. Such changes in the network, if persistent, can permanently alter the response of the population to external input.

The biological causes for synaptic plasticity are complex and full of exceptions, but there is a principle that, at least under simple conditions, is able to predict the type of changes that take place in the network of a neuronal population [8, p. 493]. This form of learning is called Hebbian learning, which we will explain in the first section of this chapter. Afterwards, we will incorporate a learning rule capable of emulating this into our model, of which we will again take the mean-field limit. We conclude by considering a few concrete examples of learning rules for our model, and how they relate to variants in the literature.

4.1 Hebbian learning

The essence of Hebbian learning, sometimes also called associative learning, is captured by the following pithy claim:

Neurons that fire together, wire together.

In slightly more words: when two neurons fire close in time (together), the synapse between them is strengthened (wired). This idea is due to Donald Hebb, who proposed it back in 1949. As noted by Gerstner et al. in [8, p. 492], the original postulate due to Hebb was more detailed than the summary above:

When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased. (Hebb [9])

Here Hebb distinguishes between the neuron that fires first and the neuron that fires second, with the second firing being (partially) caused by the first. This necessitates a temporal ordering between the spikes, giving rise to the concept of spike-timing-dependent plasticity (STDP). Indeed, experimental results suggest that whether the synaptic weight change is an increase or a decrease depends on the order of the pre- and postsynaptic spikes, even when they are very close in time [10]. Therefore, the time between the pre- and postsynaptic spikes will play a role in our model.

Remark 4.1. An alternative to STDP is the so-called rate-based approach. Under this model, each neuron has an associated (time-dependent) firing rate, with spikes arising from a Poisson process with that rate. The changes in synaptic weights are no longer dependent on individual spikes, but instead on these firing rates.

4.2 Learning rule formulation

To model STDP, the weights $w_{ij}(t)$ (which are now time-dependent) will be updated whenever a pre- or postsynaptic spike occurs at neuron j or i respectively. These weight changes will depend on the current weight, which of the two neurons fired, and the time elapsed since previous firings of the pre- or postsynaptic neuron. The sign and magnitude of an update will depend on the update functions $F_{\pm} : [-w_0, w_0] \times \mathbb{R}_+ \rightarrow \mathbb{R}$ and $G_{\pm} : [-w_0, w_0] \rightarrow \mathbb{R}$, inspired by [8, p. 500]. These functions are assumed to be continuous everywhere on their domains. The plus-variants are used for postsynaptic spikes, and the minus-variants are relevant in case of a presynaptic spike. As we are modeling STDP, the functions F_{\pm} will take into account all past pre- and postsynaptic spikes respectively. Their two arguments are

- the current normalized weight $w_N(t, x_i, x_j)$ of the synapse (defined in Definition 4.1 below); and
- the time elapsed since a past presynaptic spike in case of a current postsynaptic spike for F_+ , and vice versa for F_- .

The functions G_{\pm} will be applied only once for each post- and presynaptic spike respectively, with their argument being the current normalized weight.

Definition 4.1 (Normalized weight function). We define $w_N : I \times \Omega^2 \rightarrow [-w_0, w_0]$ for $t \in I$ and $x, y \in \Omega$ as

$$w_N(t, x, y) := \begin{cases} Nw_{ij}(t), & \text{if } \exists i, j \in \{1, \dots, N\} : (x, y) = (x_i, x_j), \\ 0, & \text{otherwise.} \end{cases}$$

This scaling by N is necessary, because in order to take the mean-field limit we still require that the true weights w_{ij} are proportional to $\frac{1}{N}$. Their limit for $N \rightarrow \infty$ would thus not be interesting to study.

Remark 4.2. In Definition 4.1 we are tacitly assuming that no two neurons share the same coordinate. This could be avoided by for instance averaging w_{ij} and w_{ik} (in case $x_j = x_k$), but we do not do this here for simplicity.

From a modeling perspective this is not very important, since two neurons would be identical and in the same location. Furthermore, if the coordinates are assigned stochastically from a continuous probability distribution, almost surely no coordinates would coincide (for finite N).

We will construct our learning rule in a way that makes the function w_N càdlàg in time. Note that w_N will only change at firing times of either the pre- or postsynaptic neuron. Namely, for fixed neurons $i, j \in \{1, \dots, N\}$, the effects of pre- and postsynaptic spikes on $w_N(\cdot, x_i, x_j)$ (and by extension on $w_{ij}(\cdot)$) are defined as follows:

- At $\tau \in \mathfrak{I}_i \setminus \mathfrak{I}_j$ a firing time of only the postsynaptic neuron i , we have

$$w_N(\tau, x_i, x_j) - w_N(\tau^-, x_i, x_j) := \sum_{\sigma \in \mathfrak{I}_j \cap (t_0, \tau]} F_+(w_N(\tau^-, x_i, x_j), \tau - \sigma) + G_+(w_N(\tau^-, x_i, x_j)), \quad (4.1)$$

- For a presynaptic firing time $\sigma \in \mathfrak{I}_j \setminus \mathfrak{I}_i$ the synaptic is updated following

$$w_N(\sigma, x_i, x_j) - w_N(\sigma^-, x_i, x_j) := \sum_{\tau \in \mathfrak{I}_i \cap (t_0, \sigma]} F_-(w_N(\sigma^-, x_i, x_j), \sigma - \tau) + G_-(w_N(\sigma^-, x_i, x_j)). \quad (4.2)$$

- When $\tau \in \mathfrak{I}_i \cap \mathfrak{I}_j$, we distinguish between the firing batches introduced in Section 3.2. The batch indices $\ell_i(\tau)$ and $\ell_j(\tau)$ provide information on which neuron fires “first”, even though both fire at the same moment in time in the model. A lower batch index means that the neuron fired earlier, so using this we can the update analogous to the previous two cases:

$$w_N(\tau, x_i, x_j) - w_N(\tau^-, x_i, x_j) := \begin{cases} \sum_{\sigma \in \mathfrak{I}_j \cap (t_0, \tau]} F_+(w_N(\tau^-, x_i, x_j), \tau - \sigma) + G_+(w_N(\tau^-, x_i, x_j)), & \text{if } \ell_i(\tau) > \ell_j(\tau), \\ \sum_{\sigma \in \mathfrak{I}_i \cap (t_0, \tau]} F_-(w_N(\tau^-, x_i, x_j), \tau - \sigma) + G_-(w_N(\tau^-, x_i, x_j)), & \text{if } \ell_i(\tau) < \ell_j(\tau), \\ 0, & \text{if } \ell_i(\tau) = \ell_j(\tau). \end{cases} \quad (4.3)$$

Remark 4.3. The choice we made for the last case in (4.3) matters relatively little from a modeling perspective because $\ell_i(\tau) = \ell_j(\tau)$ is rather unlikely to happen in non-pathological cases (the potentials of i and j would have to be identical). Here we chose to not change the weight in that case, because no temporal ordering can be assigned to the pre- and the postsynaptic spike. An alternative is applying both (4.1) and (4.2) simultaneously.

Remark 4.4. This class of learning rules could be extended by adding the inter-neuron distance $\|x_i - x_j\|$ as an argument to F_{\pm} and G_{\pm} . This could for instance be used to inhibit the growth of connections between neurons that are far apart in Ω .¹

The functions F_{\pm} and G_{\pm} can not be completely arbitrary for these updating rules to work. Namely, they must satisfy the following assumption:

Assumption 4.2. The functions F_{\pm} and G_{\pm} are such that none of (4.1), (4.2) nor (4.3) can cause w_N to exit its range $[-w_0, w_0]$, provided $w_N(t_0, x_i, x_j) \in [-w_0, w_0]$ for all $i, j \in \{1, \dots, N\}$.

This condition is mild for G_{\pm} because the current value of w_N is known, but for F_{\pm} one needs to be careful because these update functions are summed over sets of firing times (of a single neuron). It should still be possible to guarantee this assumption is met by requiring that F_{\pm} decay fast enough in time, using for instance the minimum time between firings of a single neuron from Lemma 3.1.

4.3 Mean-field limit

With this learning rule model under our belt, it is time to study its behavior in the limit of $N \rightarrow \infty$. That is, we will again take the mean-field limit. The empirical population activity and potential measures \mathcal{A}_N and ρ_N are again of interest, for which we still assume that Assumption 3.9 and Assumption 3.8 hold. As alluded to in Remark 3.14, we will now explicitly need notation for the distribution of the neuron coordinates. The associated measure is closely related to the potential measure.

Definition 4.3 (Empirical spatial measure). The empirical spatial measure $\mathfrak{X}_N : \mathcal{B}_{\Omega} \rightarrow [0, 1]$ is defined as

$$\mathfrak{X}_N(dx) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}(dx).$$

Note that for any $S \in \mathcal{B}_{\Omega}$ and all $t \in I$ we have $\mathfrak{X}_N(S) = \rho_N(t, S \times U)$.

With Assumption 3.8, we have already implicitly assumed that this empirical measure \mathfrak{X}_N converges narrowly to a probability measure $\mathfrak{X}(S) := \rho(t, S \times U)$ for arbitrary $t \in I$. The absolute continuity assumption made on this measure in Theorem 3.4 similarly implies absolute continuity of \mathfrak{X} , with density $\chi(x) := \int_U p(t, x, u) du$ for arbitrary $t \in I$.

Remark 4.5. If the coordinates x_i are sampled from a distribution with law \mathfrak{X} , then this (almost sure) narrow convergence of \mathfrak{X}_N would be a consequence of the Strong Law of Large Numbers. Thus, it is not dependent on Assumption 3.8 in that case.

Of course, the primary new object of interest is the limit of the weights themselves. To this end, we introduce the following empirical measure.

Definition 4.4 (Empirical weight measure). We define the (signed) empirical weight measure $\omega_N : I \times \mathcal{B}_{\Omega^2} \rightarrow [-w_0, w_0]$ as

$$\omega_N(t, dxdy) := \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N w_N(t, x_i, x_j) \delta_{(x_j, x_i)}(dxdy).$$

This measure computes the average normalized synaptic efficacy from dx to dy . Note the order of the coordinates in the subscript of the Dirac measure.

Similar to the empirical measures from Chapter 3, we require a convergence assumption on ω_N . In particular, we need narrow convergence. With the next assumption in mind, we will also immediately assume absolute continuity here.

¹If Ω (predominantly) corresponds to the spatial positions of the neuron, this would not necessarily be realistic. Some axons connect neurons in different brain areas or even travel all the way through the spinal cord [8, pp. 72–75]

Assumption 4.5. There exists a time-dependent signed measure $\omega : I \times \mathcal{B}_{\Omega^2} \rightarrow [-w_0, w_0]$ such that for each fixed $t \in I$, $\omega_N(t, dxdy)$ converges narrowly to $\omega(t, dxdy)$. We call ω the mean-field weight measure. This measure ω is absolutely continuous w.r.t. the Lebesgue measure, with a continuous Radon-Nikodym derivative $w : I \times \Omega^2 \rightarrow [-w_0, w_0]$ such that $\omega(t, dxdy) = w(t, y, x)dxdy$.² Furthermore, w is differentiable in time.

Next to ω_N , we also need the normalized weight function w_N itself to converge. Pointwise convergence on all of Ω^2 is not feasible, as w_N is only nonzero at neuron coordinates. However, if we restrict to the support of the product measure the empirical population activity measure and twice the empirical spatial measure from Definition 4.3 (which only contains neuron coordinates), there is hope for convergence.

Assumption 4.6. The normalized weight function w_N converges uniformly on $\text{supp } \mathcal{A}_N(dt)\mathfrak{X}_N(dx)\mathfrak{X}_N(dy)$ to the density $w : I \times \Omega^2 \rightarrow [-w_0, w_0]$ from Assumption 4.5. Here \mathcal{A}_N refers to the global population activity measure from Definition 3.2. That is,

$$\lim_{N \rightarrow \infty} \sup_{(t,x,y) \in \text{supp } \mathcal{A}_N \otimes \mathfrak{X}_N \otimes \mathfrak{X}_N} |w(t, y, x) - w_N(t^-, y, x)| = 0.$$

Note that $\mathcal{A}_N(dt dx) \ll \mathcal{A}_N(dt)\mathfrak{X}(dx)$, so this in particular implies uniform convergence of w_N on $\text{supp } \mathcal{A}_N(d\sigma dx)\mathcal{A}_N(dt dy)$ (for σ a dummy argument of w_N) and on $\text{supp } \mathfrak{X}_N(dx)\mathcal{A}_N(dt dy)$. These two properties will be used in the proof of Theorem 4.1.

The fact that we are assuming the limit of w_N to coincide with the density of ω might seem unreasonable, but given the fact that w_N is by Definition 4.4 the Radon-Nikodym derivative of ω_N w.r.t. to the product spatial measure, it is somewhat tenable.

We also require an additional assumption related to the population activity measure, which guarantees that for fixed i and j the sets $\mathfrak{X}_i \cap \{\tau \mid \ell_i(\tau) \leq \ell_j(\tau)\}$ and $\mathfrak{X}_j \cap \{\tau \mid \ell_i(\tau) \geq \ell_j(\tau)\}$ do not contribute to the limit. This is for instance achieved when their superset $\mathfrak{X}_i \cap \mathfrak{X}_j$ is empty or small for all i, j .

Assumption 4.7. We have

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{\substack{\tau \in \mathfrak{X}_i \\ \ell_i(\tau) \leq \ell_j(\tau)}} \left(\sum_{\sigma \in \mathfrak{X}_j \cap (t_0, \tau]} F_+(w_N(\tau^-, x_i, x_j), \tau - \sigma) + G_+(w_N(\tau^-, x_i, x_j)) \right) = 0,$$

and

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{\substack{\sigma \in \mathfrak{X}_j \\ \ell_i(\sigma) \geq \ell_j(\sigma)}} \left(\sum_{\tau \in \mathfrak{X}_i \cap (t_0, \sigma]} F_-(w_N(\sigma^-, x_i, x_j), \sigma - \tau) + G_-(w_N(\sigma^-, x_i, x_j)) \right) = 0.$$

With all of these definitions and assumptions in mind, we finally arrive at the following result for the mean-field weight function w :

Theorem 4.1. Under all assumptions from Theorem 3.4, plus Assumption 4.5, Assumption 4.6 and Assumption 4.7, item (ii.) of Theorem 3.4 still holds (after adding a time dependence to the function w there). Furthermore,

$$\begin{aligned} \partial_t w(t, y, x) &= A(t, y) \left(\int_{t_0}^t F_+(w(t, y, x), t - \sigma) A(\sigma, x) d\sigma + \chi(x) G_+(w(t, y, x)) \right) \\ &+ A(t, x) \left(\int_{t_0}^t F_-(w(t, y, x), t - \sigma) A(\sigma, y) d\sigma + \chi(y) G_-(w(t, y, x)) \right), \quad (t, x, y) \in I \times \Omega^2. \end{aligned} \quad (4.4)$$

²This function w should not be confused with the weight function w of the spatial model of Section 3.4, which was a fixed function independent of the model dynamics.

Proof. The fact that (ii.) still holds follows in the same way as in the proof of Theorem 3.4, using that still $|w_{ij}(t)| = \frac{1}{N} |w_N(t, x_i, x_j)| \leq \frac{w_0}{N}$ by Assumption 4.2. The only location where some care is required is in (3.41), which now requires Assumption 4.6 to hold. We will thus focus on proving (4.4). For any $\varphi \in C_c^\infty(I \times \Omega^2)$ we have

$$\begin{aligned}
 & \int_{\Omega^2} (\varphi(t_e, x, y) \omega_N(t_e, dx dy) - \varphi(t_0, x, y) \omega_N(t_0, dx dy)) \\
 &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (\varphi(t_e, x_j, x_i) w_N(t_e, x_i, x_j) - \varphi(t_0, x_j, x_i) w_N(t_0, x_i, x_j)), \\
 & \text{to which we can apply a similar argument as done in (3.38), yielding} \\
 &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\int_I \partial_t \varphi(\tau, x_j, x_i) w_N(\tau, x_i, x_j) d\tau + \sum_{\tau \in \mathfrak{I}} \varphi(\tau, x_j, x_i) (w_N(\tau, x_i, x_j) - w_N(\tau^-, x_i, x_j)) \right), \\
 &= \int_I \int_{\Omega^2} \partial_t \varphi(\tau, x, y) \omega_N(\tau, dx dy) d\tau + \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{\tau \in \mathfrak{I}} \varphi(\tau, x_j, x_i) (w_N(\tau, x_i, x_j) - w_N(\tau^-, x_i, x_j)), \\
 & \tag{4.5}
 \end{aligned}$$

As a consequence of Assumption 4.5, the limits for $N \rightarrow \infty$ of the left-hand side and the first term of the right-hand side are clear. For the second term we have, by Section 4.2,

$$\begin{aligned}
 & \sum_{\tau \in \mathfrak{I}} \varphi(\tau, x_j, x_i) (w_N(\tau, x_i, x_j) - w_N(\tau^-, x_i, x_j)) \\
 &= \sum_{\substack{\tau \in \mathfrak{I} \\ \ell_i(\tau) > \ell_j(\tau)}} \varphi(\tau, x_j, x_i) (w_N(\tau, x_i, x_j) - w_N(\tau^-, x_i, x_j)) + \sum_{\substack{\sigma \in \mathfrak{I} \\ \ell_i(\sigma) < \ell_j(\sigma)}} \varphi(\sigma, x_j, x_i) (w_N(\sigma, x_i, x_j) - w_N(\sigma^-, x_i, x_j)), \\
 &= \sum_{\substack{\tau \in \mathfrak{I}_i \\ \ell_i(\tau) > \ell_j(\tau)}} \varphi(\tau, x_j, x_i) (w_N(\tau, x_i, x_j) - w_N(\tau^-, x_i, x_j)) + \sum_{\substack{\sigma \in \mathfrak{I}_j \\ \ell_i(\sigma) < \ell_j(\sigma)}} \varphi(\sigma, x_j, x_i) (w_N(\sigma, x_i, x_j) - w_N(\sigma^-, x_i, x_j)), \\
 &= \sum_{\substack{\tau \in \mathfrak{I}_i \\ \ell_i(\tau) > \ell_j(\tau)}} \varphi(\tau, x_j, x_i) \left(\sum_{\sigma \in \mathfrak{I}_j \cap (t_0, \tau]} F_+(w_N(\tau^-, x_i, x_j), \tau - \sigma) + G_+(w_N(\tau^-, x_i, x_j)) \right) \\
 &+ \sum_{\substack{\sigma \in \mathfrak{I}_j \\ \ell_i(\sigma) < \ell_j(\sigma)}} \varphi(\sigma, x_j, x_i) \left(\sum_{\tau \in \mathfrak{I}_i \cap (t_0, \sigma]} F_-(w_N(\sigma^-, x_i, x_j), \sigma - \tau) + G_-(w_N(\sigma^-, x_i, x_j)) \right).
 \end{aligned}$$

Recall here that $\ell_i(\tau) = 0$ if and only if $\tau \notin \mathfrak{I}_i$, which enabled the second equality. We also used that in case $\ell_i(\tau) = \ell_j(\tau)$ there is no change in w_N at τ . Adding back the summations over i and j , we obtain

$$\begin{aligned}
 & \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{\tau \in \mathfrak{I}} \varphi(\tau, x_j, x_i) (w_N(\tau, x_i, x_j) - w_N(\tau^-, x_i, x_j)) \\
 &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left[\sum_{\substack{\tau \in \mathfrak{I}_i \\ \ell_i(\tau) > \ell_j(\tau)}} \varphi(\tau, x_j, x_i) \left(\sum_{\sigma \in \mathfrak{I}_j \cap (t_0, \tau]} F_+(w_N(\tau^-, x_i, x_j), \tau - \sigma) + G_+(w_N(\tau^-, x_i, x_j)) \right) \right. \\
 & \left. + \sum_{\substack{\sigma \in \mathfrak{I}_j \\ \ell_i(\sigma) < \ell_j(\sigma)}} \varphi(\sigma, x_j, x_i) \left(\sum_{\tau \in \mathfrak{I}_i \cap (t_0, \sigma]} F_-(w_N(\sigma^-, x_i, x_j), \sigma - \tau) + G_-(w_N(\sigma^-, x_i, x_j)) \right) \right].
 \end{aligned}$$

From the first term we can split off the following summation:

$$\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{\substack{\tau \in \mathfrak{I}_i \\ \ell_i(\tau) \leq \ell_j(\tau)}} \varphi(\tau, x_j, x_i) \left(\sum_{\sigma \in \mathfrak{I}_j \cap (t_0, \tau]} F_+(w_N(\tau^-, x_i, x_j), \tau - \sigma) + G_+(w_N(\tau^-, x_i, x_j)) \right),$$

which vanishes by Assumption 4.7, since φ is bounded. Thus we only have to consider

$$\begin{aligned} & \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left[\sum_{\tau \in \mathfrak{I}_i} \varphi(\tau, x_j, x_i) \left(\sum_{\sigma \in \mathfrak{I}_j \cap (t_0, \tau]} F_+(w_N(\tau^-, x_i, x_j), \tau - \sigma) + G_+(w_N(\tau^-, x_i, x_j)) \right) \right. \\ & \quad \left. + \sum_{\sigma \in \mathfrak{I}_j} \varphi(\sigma, x_j, x_i) \left(\sum_{\tau \in \mathfrak{I}_i \cap (t_0, \sigma]} F_-(w_N(\sigma^-, x_i, x_j), \sigma - \tau) + G_-(w_N(\sigma^-, x_i, x_j)) \right) \right], \\ &= \frac{1}{N} \sum_{i=1}^N \sum_{\tau \in \mathfrak{I}_i} \int_{t_0}^{\tau} \int_{\Omega} \varphi(\tau, x, x_i) F_+(w_N(\tau^-, x_i, x), \tau - \sigma) \mathcal{A}_N(d\sigma dx) \\ & \quad + \frac{1}{N} \sum_{j=1}^N \sum_{\sigma \in \mathfrak{I}_j} \int_{t_0}^{\sigma} \int_{\Omega} \varphi(\sigma, x_j, y) F_-(w_N(\sigma^-, y, x_j), \sigma - \tau) \mathcal{A}_N(d\tau dy) \\ & \quad + \frac{1}{N} \sum_{j=1}^N \int_I \int_{\Omega} \varphi(\tau, x_j, y) G_+(w_N(\tau^-, y, x_j)) \mathcal{A}_N(d\tau dy) + \frac{1}{N} \sum_{i=1}^N \int_I \int_{\Omega} \varphi(\sigma, x, x_i) G_-(w_N(\sigma^-, x_i, x)) \mathcal{A}_N(d\sigma dx), \\ &= \int_I \int_{t_0}^{\tau} \int_{\Omega^2} \varphi(\tau, x, y) F_+(w_N(\tau^-, y, x), \tau - \sigma) \mathcal{A}_N(d\sigma dx) \mathcal{A}_N(d\tau dy) \\ & \quad + \int_I \int_{t_0}^{\sigma} \int_{\Omega^2} \varphi(\sigma, x, y) F_-(w_N(\sigma^-, y, x), \sigma - \tau) \mathcal{A}_N(d\tau dy) \mathcal{A}_N(d\sigma dx) \\ & \quad + \int_I \int_{\Omega^2} \varphi(\tau, x, y) G_+(w_N(\tau^-, y, x)) \mathfrak{X}_N(dx) \mathcal{A}_N(d\tau dy) + \int_I \int_{\Omega^2} \varphi(\sigma, x, y) G_-(w_N(\sigma^-, y, x)) \mathfrak{X}_N(dy) \mathcal{A}_N(d\sigma dx), \\ &\rightarrow \int_I \int_{t_0}^{\tau} \int_{\Omega^2} \varphi(\tau, x, y) F_+(w(\tau, y, x), \tau - \sigma) \mathcal{A}(d\sigma dx) \mathcal{A}(d\tau dy) \\ & \quad + \int_I \int_{t_0}^{\sigma} \int_{\Omega^2} \varphi(\sigma, x, y) F_-(w(\sigma, y, x), \sigma - \tau) \mathcal{A}(d\tau dy) \mathcal{A}(d\sigma dx) \\ & \quad + \int_I \int_{\Omega^2} \varphi(\tau, x, y) G_+(w(\tau, y, x)) \mathfrak{X}(dx) \mathcal{A}(d\tau dy) + \int_I \int_{\Omega^2} \varphi(\sigma, x, y) G_-(w(\sigma, y, x)) \mathfrak{X}(dy) \mathcal{A}(d\sigma dx), \quad \text{as } N \rightarrow \infty. \end{aligned}$$

Here the final convergence follows from the narrow convergence of both \mathcal{A}_N and \mathfrak{X}_N , which by [5, p. 330] implies narrow convergence of the product measures we are integrating against here. It also relies on the continuity of F_{\pm} and G_{\pm} combined with Assumption 4.6 and Lemma B.4. Going back to (4.5), the narrow convergence of ω_N to ω now implies

$$\begin{aligned} & \int_{\Omega^2} (\varphi(t_e, x, y) \omega(t_e, dx dy) - \varphi(t_0, x, y) \omega(t_0, dx dy)) \\ &= \int_I \int_{\Omega^2} \partial_t \varphi(\tau, x, y) \omega(\tau, dx dy) d\tau + \int_I \int_{t_0}^{\tau} \int_{\Omega^2} \varphi(\tau, x, y) F_+(w(\tau, y, x), \tau - \sigma) \mathcal{A}(d\sigma dx) \mathcal{A}(d\tau dy) \\ & \quad + \int_I \int_{t_0}^{\sigma} \int_{\Omega^2} \varphi(\sigma, x, y) F_-(w(\sigma, y, x), \sigma - \tau) \mathcal{A}(d\tau dy) \mathcal{A}(d\sigma dx) \\ & \quad + \int_I \int_{\Omega^2} \varphi(\tau, x, y) G_+(w(\tau, y, x)) \mathfrak{X}(dx) \mathcal{A}(d\tau dy) + \int_I \int_{\Omega^2} \varphi(\sigma, x, y) G_-(w(\sigma, y, x)) \mathfrak{X}(dy) \mathcal{A}(d\sigma dx). \end{aligned}$$

Now we use that $\omega(t, dx dy)$ has Radon-Nikodym derivative $w(t, y, x)$ w.r.t. the Lebesgue measure, and that also \mathcal{A} and \mathfrak{X} are absolutely continuous to get

$$\begin{aligned}
 & \int_{\Omega^2} (\varphi(t_e, x, y)w(t_e, y, x) - \varphi(t_0, x, y)w(t_0, y, x)) dx dy \\
 &= \int_I \int_{\Omega^2} \partial_t \varphi(\tau, x, y)w(\tau, y, x) dx dy d\tau + \int_I \int_{t_0}^{\tau} \int_{\Omega^2} \varphi(\tau, x, y) F_+(w(\tau, y, x), \tau - \sigma) A(\sigma, x) A(\tau, y) d\sigma dx d\tau dy \\
 & \quad + \int_I \int_{t_0}^{\sigma} \int_{\Omega^2} \varphi(\sigma, x, y) F_-(w(\sigma, y, x), \sigma - \tau) A(\tau, y) A(\sigma, x) d\tau dy d\sigma dx \\
 & \quad + \int_I \int_{\Omega^2} \varphi(\tau, x, y) G_+(w(\tau, y, x)) \chi(x) A(\tau, y) dx d\tau dy + \int_I \int_{\Omega^2} \varphi(\sigma, x, y) G_-(w(\sigma, y, x)) \chi(y) A(\sigma, x) dy d\sigma dx, \\
 &= \int_I \int_{\Omega^2} \partial_t \varphi(\tau, x, y)w(\tau, y, x) dx dy d\tau \\
 & \quad + \int_I \int_{t_0}^{\tau} \int_{\Omega^2} \varphi(\tau, x, y) [F_+(w(\tau, y, x), \tau - \sigma) A(\sigma, x) A(\tau, y) + F_-(w(\tau, y, x), \tau - \sigma) A(\sigma, y) A(\tau, x)] dx dy d\sigma d\tau \\
 & \quad + \int_I \int_{\Omega^2} \varphi(\tau, x, y) [G_+(w(\tau, y, x)) \chi(x) A(\tau, y) + G_-(w(\tau, y, x)) \chi(y) A(\tau, x)] dx dy d\tau.
 \end{aligned}$$

Using the fact that w is assumed to be differentiable in time, a similar argument as in the proof of Theorem 3.4 with separation of the test function into a temporal and a spatial factor yields, for all $\phi \in C_c^\infty(\Omega^2)$ and $t \in I$,

$$\begin{aligned}
 & \int_{\Omega^2} \phi(x, y) \partial_t w(t, y, x) dx dy \\
 &= \int_{\Omega^2} \phi(x, y) \left(\int_{t_0}^t [F_+(w(t, y, x), t - \sigma) A(\sigma, x) A(t, y) + F_-(w(t, y, x), t - \sigma) A(\sigma, y) A(t, x)] d\sigma \right. \\
 & \quad \left. + G_+(w(t, y, x)) \chi(x) A(t, y) + G_-(w(t, y, x)) \chi(y) A(t, x) \right) dx dy,
 \end{aligned}$$

and thus pointwise

$$\begin{aligned}
 \partial_t w(t, y, x) &= A(t, y) \left(\int_{t_0}^t F_+(w(t, y, x), t - \sigma) A(\sigma, x) d\sigma + \chi(x) G_+(w(t, y, x)) \right) \\
 & \quad + A(t, x) \left(\int_{t_0}^t F_-(w(t, y, x), t - \sigma) A(\sigma, y) d\sigma + \chi(y) G_-(w(t, y, x)) \right),
 \end{aligned}$$

for all $(t, x, y) \in I \times \Omega^2$, which is identical to (4.4). \blacksquare

Remark 4.6. The two terms in (4.4) correspond closely to the definitions of the weight updates in (4.1) and (4.2). The multiplications with $A(t, y)$ and $A(t, x)$ respectively are a consequence of the fact that these updates are applied at postsynaptic and presynaptic firing times respectively.

4.4 Short-memory limit

In this final section, we will show how our mean-field equation for w can be simplified by limiting the time-dependence of our learning rule. In this context, we also provide some examples of concrete learning rules. This will also demonstrate a link between STDP-based rules and the rate-based rules mentioned in Remark 4.1.

We consider update functions $F_{\pm}(w, \Delta t)$ whose only time-dependence is a factor $e^{-\lambda \Delta t}$. For large enough $\lambda > 0$ such a rule should be able to satisfy Assumption 4.2 due to the exponential decay in time. If appropriately rescaled, we can reduce the ‘‘memory’’ of this rule by taking $\lambda \rightarrow \infty$. The effect this has on the mean-field equation for $\partial_t w$ is that the time-integral from t_0 to t disappears, leaving us with a form analogous to rate-based learning rules.

Corollary 4.1.1. *Suppose the spike-timing-dependent update functions are of the form*

$$F_{\pm}^{\lambda}(w, \Delta t) = \lambda \bar{F}_{\pm}(w) e^{-\lambda \Delta t}, \quad (4.6)$$

where $\lambda > 0$ with \bar{F}_{\pm} continuous functions.

Let w_{λ} and A_{λ} be the λ -dependent weight density and population activity density. Assume that w_{λ} converges pointwise to a continuous function $\hat{w} : I \times \Omega^2$ for $\lambda \rightarrow \infty$, which is differentiable in time. Also suppose that $\partial_t w_{\lambda}$ converges pointwise to $\partial_t \hat{w}$. For A_{λ} we assume that it converges uniformly in time and pointwise in space to a continuous function $\hat{A} : I \times \Omega \rightarrow \mathbb{R}_+$ in the limit of $\lambda \rightarrow \infty$. Then the equation for $\partial_t \hat{w}$ becomes

$$\partial_t \hat{w}(t, y, x) = \hat{A}(t, y) \hat{A}(t, x) \bar{F}(\hat{w}(t, y, x)) + \hat{A}(t, y) \chi(x) G_+(\hat{w}(t, y, x)) + \hat{A}(t, x) \chi(y) G_-(\hat{w}(t, y, x)), \quad (4.7)$$

where $\bar{F} := \bar{F}_+ + \bar{F}_-$.

Proof. For fixed $\lambda > 0$ we know from Theorem 4.1 that the mean-field equation for $\partial_t w_{\lambda}$ is

$$\begin{aligned} \partial_t w_{\lambda}(t, y, x) &= A_{\lambda}(t, y) \left(\int_{t_0}^t \lambda \bar{F}_+(w_{\lambda}(t, y, x)) e^{-\lambda(t-\sigma)} A_{\lambda}(\sigma, x) d\sigma + \chi(x) G_+(w_{\lambda}(t, y, x)) \right) \\ &\quad + A_{\lambda}(t, x) \left(\int_{t_0}^t \lambda \bar{F}_-(w_{\lambda}(t, y, x)) e^{-\lambda(t-\sigma)} A_{\lambda}(\sigma, y) d\sigma + \chi(y) G_-(w_{\lambda}(t, y, x)) \right), \\ &= A_{\lambda}(t, y) \left(\bar{F}_+(w_{\lambda}(t, y, x)) \int_{t_0}^t \lambda e^{-\lambda(t-\sigma)} A_{\lambda}(\sigma, x) d\sigma + \chi(x) G_+(w_{\lambda}(t, y, x)) \right) \\ &\quad + A_{\lambda}(t, x) \left(\bar{F}_-(w_{\lambda}(t, y, x)) \int_{t_0}^t \lambda e^{-\lambda(t-\sigma)} A_{\lambda}(\sigma, y) d\sigma + \chi(y) G_-(w_{\lambda}(t, y, x)) \right). \end{aligned}$$

Note that for any family of continuous functions $\{g_{\lambda}\}_{\lambda>0} : \mathbb{R} \rightarrow \mathbb{R}$ that converges uniformly to some function $g : \mathbb{R} \rightarrow \mathbb{R}$, we have for any $t > t_0$ that

$$\lim_{\lambda \rightarrow \infty} \int_{t_0}^t \lambda g_{\lambda}(\sigma) e^{-\lambda(t-\sigma)} d\sigma = g(t).$$

Because of the continuity of \bar{F}_{\pm} and G_{\pm} , and our convergence assumptions on w_{λ} and A_{λ} , we thus obtain

$$\begin{aligned} \partial_t \hat{w}(t, y, x) &= \lim_{\lambda \rightarrow \infty} \partial_t w_{\lambda}(t, y, x), \\ &= \hat{A}(t, y) \left(\hat{A}(t, x) \bar{F}_+(\hat{w}(t, y, x)) + \chi(x) G_+(\hat{w}(t, y, x)) \right) \\ &\quad + \hat{A}(t, x) \left(\hat{A}(t, y) \bar{F}_-(\hat{w}(t, y, x)) + \chi(y) G_-(\hat{w}(t, y, x)) \right), \\ &= \hat{A}(t, y) \hat{A}(t, x) \bar{F}(\hat{w}(t, y, x)) \\ &\quad + \hat{A}(t, y) \chi(x) G_+(\hat{w}(t, y, x)) + \hat{A}(t, x) \chi(y) G_-(\hat{w}(t, y, x)). \end{aligned}$$

■

An interesting consequence of the reduction of the memory of the learning rule is that the individual update functions \bar{F}_+ and \bar{F}_- no longer matter. Only their sum \bar{F} appears in (4.7).

Remark 4.7. Note that we are taking the limit $\lambda \rightarrow \infty$ on the mean-field equation for w . The result could be different if we take this limit before taking the mean-field limit itself.

We are making a few strong convergence assumptions to do this, which could be weakened, because this corollary merely serves as an example of what is possible.

Remark 4.8. A simpler way to achieve the effect of Corollary 4.1.1 would be a learning rule with a time-window, e.g., $F_{\pm}(w, \Delta t) = \lambda \bar{F}_{\pm}(w) \mathbb{1}\{\Delta t < \frac{1}{\lambda}\}$. The reason the form used in Corollary 4.1.1 is preferred is that it is continuous in Δt .

Example 4.1. The simplest Hebbian STDP learning rule arises from

$$\bar{F}(w) := cw,$$

and $G_{\pm} \equiv 0$, where $c > 0$ is some constant. Then the short-memory limit according to Corollary 4.1.1 of the mean-field equation is

$$\partial_t \hat{w}(t, y, x) = c \hat{w}(t, y, x) \hat{A}(t, x) \hat{A}(t, y). \quad (4.8)$$

This equation shows why this is called a Hebbian learning rule. Namely, when both the pre- and postsynaptic neurons are active we should have $\hat{A}(t, x) \hat{A}(t, y) > 0$, which enables the synaptic weight between them to grow (or decrease if it starts out inhibitory). So indeed, “neurons that fire together, wire together”.

Remark 4.9. The simplest rate-based Hebbian rule mentioned in [8, p. 497] is

$$\frac{d}{dt} w_{ij} = c w_{ij} \nu_i \nu_j.$$

Note that this equation is entirely analogous to (4.8) if we interpret the short-memory, mean-field population activity density \hat{A} as a firing rate. In fact, all rate-based rules that have at most first-order factors ν_i and ν_j can be obtained from short-memory limits of mean-field limits of suitably chosen STDP rules.

A downside of the simple Hebbian rule covered in Example 4.1 is that its differential equation for w exhibits blow-up behavior, which would violate Assumption 4.2 and, more importantly, not lead to a stable connectivity pattern. This makes sense, as the only nonzero update function \bar{F} is unbounded on $[-w_0, w_0]$.

The next example contains a more realistic learning rule that takes into account that the synaptic weight between two neurons can not grow unbounded. This is done by limiting the synaptic weight by the activity of the presynaptic neuron.

Example 4.2. Consider the learning rule based on

$$\bar{F}(w) := c, \quad G_+(w) := -cw/\mathcal{L}(\Omega), \quad \text{and} \quad G_-(w) := 0,$$

where again $c > 0$ is a constant, and suppose the neurons are distributed uniformly over Ω such that $\chi(x) = \frac{1}{\mathcal{L}(\Omega)}$ for all $x \in \Omega$. Here $\mathcal{L}(\Omega)$ is the Lebesgue measure of Ω . The short-memory limit of this rule then becomes

$$\partial_t \hat{w}(t, y, x) = c \hat{A}(t, y) \left[\hat{A}(t, x) - \hat{w}(t, y, x) \right].$$

This equation has two stationary solutions for w , namely $\hat{A}(t, y) = 0$ and $\hat{w}(t, y, x) = \hat{A}(t, x)$. The interpretation of this is that a synaptic weight only updates when there is postsynaptic activity, in which case it starts to converge towards the presynaptic activity.

Thus, under this learning rule, it is theoretically possible for a stable connectivity pattern to arise, which would mean that the population learned. The weights could function as long-term storage of past presynaptic activity, which would only be erased when postsynaptic activity returned.

Note that again this equation has a rate-based counterpart, which is discussed in [8, pp. 506–507].

Chapter 5

Conclusions

5.1 Summary

Throughout Chapter 2, Chapter 3 and Chapter 4 we developed a model for learning neuronal populations, which uses the integrate-and-fire model for individual neurons. In Chapter 2, we started by demonstrating the existence of local solutions to the potential equation of an individual neuron, after which we constructed global solutions that take into account the resets inherent to the integrate-and-fire model. At the end of that chapter, we derived the gain function for a neuron undergoing repetitive firing.

We extended the model in Chapter 3, where we considered a population of interacting neurons. Here we encountered synchronization as a possible property of excitatory populations with a driving force that is decreasing in u , such as the leaky model. This was confirmed numerically. We went on to determine the mean-field limit for both homogeneous and spatial populations with a fixed connectivity pattern. This yielded a system of partial differential equations, for which we determined (implicit) stationary solutions. For fully connected homogeneous inhibitory populations, numerical simulations corroborated the validity of these results. Furthermore, we confirmed a link between these stationary solutions and the gain function of an individual neuron of the population.

Finally, in Chapter 4, we added a learning rule to our model that updates the synaptic weights between the neurons based on spike-timing-dependent plasticity (STDP), with the intent of modeling Hebbian learning. We again derived the mean-field limit, which resulted in an additional differential equation for the change in the synaptic weight function. We discovered that when the memory of an STDP rule is decreased by shortening its time-window, the mean-field equation for the synaptic weight function becomes analogous to the equation found in rate-based models of synaptic plasticity.

5.2 Discussion

In the construction of our population model, we had to make several modeling choices, many of which were based on the existing literature. One exception is the way of ordering simultaneous spikes, which we presented in Section 3.1. This method was designed to achieve consistency with both the exponential integrate-and-fire model and biophysical models. In the end, this modeling choice mattered little for the focus of this thesis, which was deriving mean-field limits, because we assumed that the number of simultaneous spikes was negligible. However, this framework could still be valuable for other applications.

To derive our mean-field results we had to make a significant number of assumptions, which are unlikely to hold under all possible conditions. Still, some of our mean-field-based predictions did match simulation outcomes, and the short-memory limit of the mean-field equation of our STDP rule agreed with the rate-based approach. This suggests that the derived equations do capture important aspects of the behavior of large neuronal populations, and are (partially) valid.

Provided that this is the case, the mean-field equations we obtained relate the external (sensory) input neurons receive, their activity, and the changes in the synaptic weights between them, at least for large populations. This could be useful in a biological context, but might also have applications in artificial neural networks that are very similar to biological neuronal systems.

5.3 Future research

There is an abundance of opportunities for further research on this topic. We present some of the most interesting options here, in no particular order:

- The most obvious direction for future research would be a translation of the various (convergence) assumptions into conditions on the model parameters. This would tell us to which kind of neuronal populations the mean-field results apply. Many assumptions could likely also be weakened, or maybe even removed entirely by generalizing the results. For instance, Assumption 4.7 is a prime candidate for refinement.
- A particularly desirable generalization of our mean-field findings would be a weak formulation of the system of partial differential equations we derived, in which the assumption of absolute continuity of the measures is dropped. This could potentially enable the inclusion of synchronized solutions in the form of Dirac measures.
- Another possible continuation would be solving the derived partial differential equations. As noted in Section 3.3.3, they are more complicated than they seem due to the nonlinear dependence of the activity on the potential density. In fact, for the spatial model it is not even clear when and if the implicit population activity equation in Theorem 3.4 has a (unique) solution.
- Related to the previous item, the stability of solutions to these equations would also be interesting. This applies in particular to the learning rule examples of Section 4.4, and other examples, because a stationary solution for w would suggest that the network learned something in the long term.
- It could also be interesting to increase the realism of the used neuron models, for instance by adding an absolute refractory period or incorporating adaptation in accordance with what we discussed in Section 2.1.
- More numerical verifications of the results would be desirable. For example, for the model with a learning rule it could be investigated whether or not a stationary weight pattern arises for certain external input.
- The model we considered in this thesis was entirely deterministic. In contrast, most of the literature includes a stochastic term in the differential equation for the membrane potential of the integrate-and-fire model (which is (2.6)). This makes sense from a modeling perspective since the input neurons receive from other parts of the brain is often noisy. Such a stochastic component could also be added to the update functions from Chapter 4 to see what the effect would be on the learning process.

Bibliography

- [1] Robert A. Adams and Christopher Essex. *Calculus: a complete course*. Ninth edition. Toronto: Pearson Canada Inc., 2017. 1085 pp. ISBN: 978-0-13-415436-7.
- [2] Praveen Agarwal, Mohamed Jleli, and Bessem Samet. "Banach Contraction Principle and Applications." In: *Fixed Point Theory in Metric Spaces: Recent Advances and Applications*. Ed. by Praveen Agarwal, Mohamed Jleli, and Bessem Samet. 1st ed. Singapore: Springer, 2018, pp. 1–23. ISBN: 978-9-811-32913-5. doi: 10.1007/978-981-13-2913-5_1.
- [3] Vladimir I. Bogachev. *Measure Theory*. Berlin, Heidelberg: Springer, 2007. 1101 pp. ISBN: 978-3-540-34514-5. doi: 10.1007/978-3-540-34514-5.
- [4] Romain Brette and Wulfram Gerstner. "Adaptive Exponential Integrate-and-Fire Model as an Effective Description of Neuronal Activity." In: *Journal of Neurophysiology* 94.5 (Nov. 2005), pp. 3637–3642. ISSN: 1522-1598. doi: 10.1152/jn.00686.2005.
- [5] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Vol. 38. Stochastic Modelling and Applied Probability. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. ISBN: 978-3-642-03311-7. doi: 10.1007/978-3-642-03311-7.
- [6] Nicolas Fourcaud-Trocmé et al. "How Spike Generation Mechanisms Determine the Neuronal Response to Fluctuating Inputs." In: *The Journal of Neuroscience* 23.37 (Dec. 17, 2003), pp. 11628–11640. ISSN: 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.23-37-11628.2003.
- [7] Wulfram Gerstner and Werner M. Kistler. "Mathematical formulations of Hebbian learning." In: *Biological Cybernetics* 87.5 (Dec. 1, 2002), pp. 404–415. ISSN: 03401200. doi: 10.1007/s00422-002-0353-y.
- [8] Wulfram Gerstner et al. *Neuronal dynamics: from single neurons to networks and models of cognition*. 1st ed. Cambridge, United Kingdom: Cambridge University Press, 2014. 577 pp. ISBN: 978-1-107-44761-5. doi: 10.1017/CB09781107447615.
- [9] D. O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. 1st ed. New York: Psychology Press, May 1, 2002. 378 pp. ISBN: 978-1-4106-1240-3. doi: 10.4324/9781410612403.
- [10] H. Markram et al. "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs." In: *Science (New York, N.Y.)* 275.5297 (Jan. 10, 1997), pp. 213–215. ISSN: 0036-8075. doi: 10.1126/science.275.5297.213.
- [11] Dennis R. Mosier. "CHAPTER 1 - Clinical Neuroscience." In: *Neurology Secrets*. Ed. by Loren A. Rolak. 5th ed. Philadelphia: Mosby, Jan. 1, 2010, pp. 7–17. ISBN: 978-0-323-05712-7. doi: 10.1016/B978-0-323-05712-7.00001-5.
- [12] Christopher Rackauckas and Qing Nie. "DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia." In: *Journal of Open Research Software* 5.1 (May 25, 2017), p. 15. ISSN: 2049-9647. doi: 10.5334/jors.151.
- [13] Thomas C. Sideris. *Ordinary differential equations and dynamical systems*. Atlantis Studies in Differential Equations 2214-6253. Paris: Atlantis Press, 2013. 225 pp. ISBN: 978-94-6239-021-8. doi: 10.2991/978-94-6239-021-8.
- [14] Quasar Jarosz at English Wikipedia. *Neuron-no labels2.png in Inkscape and hand-tuned to reduce filesize*. URL: https://commons.wikimedia.org/wiki/File:Neuron_Hand-tuned.svg. Aug. 11, 2009.

Appendix A

Background material

Theorem A.1 (Picard-Lindelöf). *Let $(X, |\cdot|)$ be a finite-dimensional Banach space with $U \subset X$ and $I \subset \mathbb{R}$ both open and nonempty. Suppose $f \in C^{0,1-}(I \times U, X)$. Then for every $(t_0, x_0) \in I \times U$ there exists an open interval $J \subset I$ such that $t_0 \in J$ and the initial value problem*

$$\begin{cases} \dot{x} &= f(t, x), \\ x(t_0) &= x_0, \end{cases} \quad (\text{A.1})$$

has a unique solution $u \in C(J, U)$.

Proof. See for instance [13, p. 23]. ■

Theorem A.2 (Banach Fixed Point Theorem). *Let (X, d) be a complete metric space, with a contraction $\mathcal{T} : X \rightarrow X$ with Lipschitz constant $K < 1$. Then \mathcal{T} has a unique fixed point in $x^* \in X$, and for every $x \in X$ the sequence $\{\mathcal{T}^n x\}_{n \in \mathbb{N}}$ converges to x^* .*

Proof. See for instance [2]. ■

Theorem A.3 (Taylor's Theorem with Lagrange remainder in one variable). *Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be $(n+1)$ -times differentiable on some open interval I , and let $a \in I$ be fixed. Then for all $x \in I$ there exists an s between a and x such that*

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + \frac{f^{(n+1)}(s)}{(n+1)!} (x-a)^{n+1}.$$

Proof. See for instance [1, p. 278]. ■

Theorem A.4 (Dominated Convergence Theorem). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space, and let $(f_n)_{n \in \mathbb{N}} : \mathcal{X} \rightarrow \mathbb{R}$ be a sequence of \mathcal{F} -measurable functions that converges pointwise to a function $f : \mathcal{X} \rightarrow \mathbb{R}$. Furthermore, assume that there exists a μ -integrable function $g : \mathcal{X} \rightarrow [0, \infty]$ such that $|f_n| \leq g$ μ -a.e. Then f is μ -integrable and*

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X}} f_n d\mu = \int_{\mathcal{X}} f d\mu.$$

Proof. See for instance [3, p. 130]. ■

Appendix B

Supporting lemmas

Lemma B.1. For fixed $u \in D$, the map $\mathcal{T}u$ defined in (2.11) is càdlàg.

Proof. Recall that

$$(\mathcal{T}u)(t) := x_0 + \int_{t_0}^t \left(f(\tau, u(\tau)) d\tau + \sum_{i=1}^N \omega_i \mu_i(d\tau) \right), \quad t \in J.$$

Let $t^* \in J$. By Definition 2.1 it suffices to show that $\lim_{t \downarrow t^*} (\mathcal{T}u)(t) = (\mathcal{T}u)(t^*)$ and that $\lim_{t \uparrow t^*} (\mathcal{T}u)(t)$ exists.

1. Since $\tau \mapsto f(\tau, u(\tau))$ is bounded on J , the Dominated Convergence Theorem A.4 implies that

$$\lim_{t \downarrow t^*} \int_{t_0}^t f(\tau, u(\tau)) d\tau = \int_{t_0}^{t^*} f(\tau, u(\tau)) d\tau.$$

Furthermore,

$$\lim_{t \downarrow t^*} \int_{t_0}^t \sum_{i=1}^N \omega_i \mu_i(d\tau) = \lim_{t \downarrow t^*} \sum_{i=1}^N \omega_i \mu_i((t_0, t]) = \sum_{i=1}^N \omega_i \mu_i((t_0, t^*]),$$

by Notation 2.4 and the fact that measures are continuous from above (and μ_1, \dots, μ_N are finite). Thus,

$$\begin{aligned} \lim_{t \downarrow t^*} (\mathcal{T}u)(t) &= \lim_{t \downarrow t^*} \left(x_0 + \int_{t_0}^t \left(f(\tau, u(\tau)) d\tau + \sum_{i=1}^N \omega_i \mu_i(d\tau) \right) \right), \\ &= x_0 + \int_{t_0}^{t^*} \left(f(\tau, u(\tau)) d\tau + \sum_{i=1}^N \omega_i \mu_i(d\tau) \right), \\ &= (\mathcal{T}u)(t^*). \end{aligned}$$

2. As $\tau \mapsto f(\tau, u(\tau))$ is bounded on J and $\mathcal{L}(\{t^*\}) = 0$, the Dominated Convergence Theorem A.4 implies that

$$\lim_{t \uparrow t^*} \int_{t_0}^t f(\tau, u(\tau)) d\tau = \int_{t_0}^{t^*} f(\tau, u(\tau)) d\tau.$$

Also,

$$\lim_{t \uparrow t^*} \int_{t_0}^t \sum_{i=1}^N \omega_i \mu_i(d\tau) = \lim_{t \uparrow t^*} \sum_{i=1}^N \omega_i \mu_i((t_0, t]) = \sum_{i=1}^N \omega_i \mu_i((t_0, t^*)),$$

by Notation 2.4 and the fact that measures are continuous from below. Since both of these limits exist, clearly $\lim_{t \uparrow t^*} (\mathcal{T}u)(t)$ also exists.

Since both properties hold, we conclude that $\mathcal{T}u$ is càdlàg. ■

Lemma B.2. Let $(X, \|\cdot\|)$ be a Banach space, and $J \subset \mathbb{R}$ non-empty and open. The Skorokhod space $D_b(J, X)$ of bounded càdlàg functions from J to X is complete w.r.t. the supremum norm $\|\cdot\|_\infty$.

Proof. Consider the space of bounded functions $B(J, X)$ which is complete w.r.t. the sup-norm. We will show that $D_b(J, X) \subset B(J, X)$ is closed, which immediately implies that it is complete.

To do so, let $\{h_n\}_{n \in \mathbb{N}} \subset D_b(J, X)$ be a sequence that converges to some $h \in B(J, X)$. It suffices to show that $h \in D_b(J, X)$.

Let $t^* \in J$. We will show that the two properties required for h to be càdlàg hold in t^* .

1. We claim that h is right-continuous in t^* . Proving this is very similar to showing that uniform convergence of a sequence of functions implies continuity of the limiting function, but we demonstrate it for completeness. We have to show the following:

$$\forall \varepsilon > 0 : \exists \delta > 0 : \forall t \in (t^*, t^* + \delta) : \|h(t) - h(t^*)\| < \frac{\varepsilon}{3}. \quad (\text{B.1})$$

Let $\varepsilon > 0$. Since $\lim_{n \rightarrow \infty} h_n = h$, we can choose $N \in \mathbb{N}$ such that $\|h - h_N\|_\infty < \varepsilon$. Now using that h_N is right-continuous in t^* (because $h_N \in D_b(J, X)$), we can choose $\delta > 0$ such that

$$\forall t \in (t^*, t^* + \delta) : \|h_N(t) - h_N(t^*)\| < \frac{\varepsilon}{3}.$$

Therefore we get for $t \in (t^*, t^* + \delta)$ that

$$\|h(t) - h(t^*)\| \leq \|h(t) - h_N(t)\| + \|h_N(t) - h_N(t^*)\| + \|h_N(t^*) - h(t^*)\| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Having shown (B.1), we conclude that h is right-continuous in t^* .

2. We claim that the one-sided limit $\lim_{t \uparrow t^*} h(t)$ exists. To show this, we first construct a candidate limit. Define the sequence $\{x_n\}_{n \in \mathbb{N}}$ by

$$x_n := \lim_{t \uparrow t^*} h_n(t), \quad n \in \mathbb{N}. \quad (\text{B.2})$$

This sequence is well-defined because h_n is càdlàg for all $n \in \mathbb{N}$. We claim that $\{x_n\}_{n \in \mathbb{N}}$ is a Cauchy sequence. To this end, we will show

$$\forall \varepsilon > 0 : \exists N \in \mathbb{N} : \forall m, n \geq N : \|x_m - x_n\| < \varepsilon.$$

Let $\varepsilon > 0$. As $\{h_n\}_{n \in \mathbb{N}}$ converges (to h), it is a Cauchy sequence. Choose $N \in \mathbb{N}$ such that

$$\forall m, n \geq N : \|h_m - h_n\|_\infty < \frac{\varepsilon}{3}.$$

Now let $m, n \geq N$. Because

$$\lim_{t \uparrow t^*} h_m(t) = x_m \quad \text{and} \quad \lim_{t \uparrow t^*} h_n(t) = x_n,$$

we can choose a $\hat{t} \in J$ with the property

$$\|x_m - h_m(\hat{t})\| < \frac{\varepsilon}{3} \quad \text{and} \quad \|x_n - h_n(\hat{t})\| < \frac{\varepsilon}{3}.$$

We now find that

$$\|x_m - x_n\| \leq \|x_m - h_m(\hat{t})\| + \|h_m(\hat{t}) - h_n(\hat{t})\| + \|x_n - h_n(\hat{t})\| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Therefore $\{x_n\}_{n \in \mathbb{N}}$ is a Cauchy sequence. Since $(X, \|\cdot\|)$ is complete, $\{x_n\}_{n \in \mathbb{N}}$ converges to some $x^* \in X$. We now claim that $\lim_{t \uparrow t^*} h(t) = x^*$. Thus, we show that

$$\forall \varepsilon > 0 : \exists \delta > 0 : \forall t \in (t^* - \delta, t^*) : \|h(t) - x^*\| < \varepsilon. \quad (\text{B.3})$$

Let $\varepsilon > 0$. Choose $N \in \mathbb{N}$ such that

$$\|h - h_N\|_\infty < \frac{\varepsilon}{3} \quad \text{and} \quad \|x^* - x_N\| < \frac{\varepsilon}{3},$$

by using that $\{h_n\}_{n \in \mathbb{N}}$ converges to h and $\{x_n\}_{n \in \mathbb{N}}$ converges to x^* . By (B.2) we can choose $\delta > 0$ such that

$$\forall t \in (t^* - \delta, t^*) : \|h_N(t) - x_N\| < \frac{\varepsilon}{3}.$$

It follows that for all $t \in (t^* - \delta, t^*)$,

$$\|h(t) - x^*\| \leq \|h - h_N\|_\infty + \|h_N(t) - x_N\| + \|x_N - x^*\| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Thus (B.3) holds, so $\lim_{t \uparrow t^*} h(t) = x^*$. In particular, this limit exists.

From both of these properties we conclude that h is càdlàg, i.e., $h \in D_b(J, X)$. This implies that $D_b(J, X) \subset B(J, X)$ is closed, and thus also complete. ■

Lemma B.3. *The exponential norm $\|\cdot\|_\lambda$ as defined in Definition 2.5 is equivalent to the standard supremum norm $\|\cdot\|_\infty$ on $B([t_0, s], X)$ for all $\lambda > 0$.*

Proof. For all $u \in B([t_0, s], X)$ we have

$$\|u\|_\lambda = \sup_{t \in [t_0, s]} e^{-\lambda(t-t_0)} \|u\| \leq \sup_{t \in [t_0, s]} \|u\| = \|u\|_\infty,$$

and

$$\|u\|_\infty = \sup_{t \in [t_0, s]} \|u\| = \sup_{t \in [t_0, s]} \left(e^{\lambda(t-t_0)} e^{-\lambda(t-t_0)} \|u\| \right) \leq e^{\lambda(s-t_0)} \cdot \sup_{t \in [t_0, s]} e^{-\lambda(t-t_0)} \|u\| = e^{\lambda(s-t_0)} \|u\|_\lambda,$$

where $e^{\lambda(s-t_0)}$ is a positive constant.

Thus,

$$\|u\|_\lambda \leq \|u\|_\infty \leq e^{\lambda(s-t_0)} \|u\|_\lambda,$$

meaning that the two norms are indeed equivalent on $B([t_0, s], X)$. ■

Lemma B.4. *Let X be a finite-dimensional Banach space with Borel σ -algebra \mathcal{F} . Suppose a sequence of finite measures $(\mu_n)_{n \in \mathbb{N}} : \mathcal{F} \rightarrow [0, \infty)$ converges narrowly to some finite measure $\mu : \mathcal{F} \rightarrow [0, \infty)$. Moreover, let $(f_n)_{n \in \mathbb{N}} : X \rightarrow \mathbb{R}$ be a sequence of \mathcal{F} -measurable functions that converges uniformly on $\text{supp } \mu_n$ to some (\mathcal{F} -measurable) function $f \in C_b(X)$, i.e.,*

$$\lim_{n \rightarrow \infty} \sup_{x \in \text{supp } \mu_n} |f(x) - f_n(x)| = 0. \quad (\text{B.4})$$

Then we have

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu_n = \int_X f d\mu.$$

Proof. We have

$$\begin{aligned} \left| \int_X f d\mu - \int_X f_n d\mu_n \right| &= \left| \int_X (f - f_n) d\mu_n - \int_X f (d\mu_n - d\mu) \right|, \\ &\leq \left| \int_X (f - f_n) d\mu_n \right| + \left| \int_X f (d\mu - d\mu_n) \right|, \\ &\leq \sup_{x \in \text{supp } \mu_n} |f(x) - f_n(x)| \mu_n(X) + \left| \int_X f (d\mu - d\mu_n) \right|, \\ &\rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Here the convergence of the first term follows by (B.4) and the fact that $\mu_n(X)$ converges to $\mu(X) < \infty$ by narrow convergence (follows by taking the unit test function). The second term vanishes by the narrow convergence of μ to μ_n and the fact that $f \in C_b(X)$. Thus indeed

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu_n = \int_X f d\mu. \quad \blacksquare$$