
Università degli studi di Padova

Dipartimento di Matematica "Tullio Levi-Civita"

**Corso di Laurea Magistrale in Matematica**


# Emergence of collective periodic behaviour in a multi-population system of interacting neurons


Relatore:
**Prof. Marco Formentin**

Laureanda:
**Ilaria Di Pietro**
**Mat.2044682**



21 Luglio 2023

Anno Accademico 2022-2023

***Alla Nonna Marì,***

*che, con il suo costante stupore per la mia passione per la Matematica,*
*mi ha resa fiera ed orgogliosa di intraprendere questa strada.*

# Contents

# Introduction

The analysis of complex systems usually involves the study of huge families of interacting components: the aim is understanding the internal mechanisms that regulate such networks, modelling complexity and investigating eventual self-organization.

In this thesis, the interacting components we shall consider are the neurons in the brain; more specifically, we look at a multi-population system of interacting neurons and study their behaviour, with the purpose of answering the following question: *can a system exhibit a collective periodic behaviour even though the single units have no natural tendency to behave periodically?*
In the biological context, the answer is positive: indeed, it is a matter of fact that biological rhythms are ubiquitous in living beings, since the emergence of recurrent dynamical patterns is observed at all scales. We can just think of the 24-hours rhythm for the day/night regulation, of the circadian rhythm, or of the limb movements, which are essentially made of flexions and extensions. Moreover, spreading the discussion outside the neuronal context, we know rhythmic patterns and self-organized collective behaviours are exhibited also on larger levels: to have in mind some examples, we can think of the predator-prey equilibria, or of applauding audiences, and so on and so forth. Some of these phenomena are well-understood from a theoretical standpoint, others need further investigations.
In the present thesis we aim at modelling periodic behaviours of brain neural activity. We consider a multi-population network of neurons (each class representing, for instance, a different functional area of the brain), and we model it using a particular type of counting processes, a multi-class system of non-linear multivariate *Hawkes processes*, which possesses the right features to represent the spiking activity of a neuron. Within this framework, the activity of each neuron is represented via counting the number of successive spikes it emits in a certain interval of time. The specific nature of the Hawkes processes allows providing a good model of this synaptic integration phenomenon, since it takes into account all the proper features of the network, such as the spiking activity of the single neuron, the interaction (in terms of excitation/inhibition) between the different classes, the delay of transmission, the past history of the neuron and the noise of the network.

Equipped with this mathematical structure, and following the work carried out by Susanne Ditlevsen and Eva Löcherbach in their paper [17], throughout this thesis we study such a model with the aim of detecting the emergence of a periodic collective dynamics, i.e. self-sustained oscillations: not only we do this from a theoretical point of view, but also we provide evidence for such rhythms through numerical simulations.

In particular, the thesis is structured as follows:

- **In Chapter 1** we provide the reader a biological overview: we open a brief parenthesis about the general functioning of the nervous system and the spiking activity of neurons, trying to understand what are the usual mathematical means for modelling such ensembles. In addition, we deepen the theme of rhythmic collective behaviours emerging in living and complex systems, which leads to the discussion on oscillations for multi-class systems of interacting neurons which is the object of the current work.

- **In Chapter 2** we equip the reader with the mathematical tools needed throughout the thesis. Each section will be also supplied of the necessary references to eventually deepen the topics at stake and to figure out in which part of the thesis such tools are used.

- **In Chapter 3** we actually get into the main topic of the thesis.
  First of all, we set the mathematical framework and probabilistic setting by expliciting the model via multi-class systems of non-linear multivariate Hawkes processes, working in a mean-field framework. After that, having at disposal the multivariate process representing the spiking activity of the ensemble of neurons, we study its large population limit, i.e. we consider a situation in which the total number of neurons tends to infinity.
  In particular, we show that the system can be approximated by a system of inhomogeneous and independent Poisson Processes, hence arriving to the first relevant result: in Theorem 3.6 we prove *Propagation of Chaos*, discovering that, in the large population limit, within the same class, neurons converge in law to i.i.d. copies of the same limit law.
  In addition to that, we investigate the relation between the large time behaviour of the limit system and the large time behaviour of the finite size system: to this purpose, and considering both time and the total number of neurons tending to infinity, we provide -in Theorem 3.11- a Central Limit Theorem.

- **In Chapter 4** we focus on the limit system, leading finally towards the emergence of oscillations. We consider a situation in which the

different classes of neurons interact following a Monotone Cyclic Feedback System *(MFC)*, in which the choice of the memory kernels in form of an Erlang distribution is essential. In this context, and by means of dynamical systems theory, we get to Theorem 4.2, in which, under suitable assumptions, it is proven that the limit system possesses attracting and non-constant periodic orbits, i.e., it presents oscillations.

- **In Chapter 5** we deepen the study of the model analyzing several aspects.
  In the first part of the chapter, we exploit the specific Erlang structure of the memory kernels to associate to the original finite size system represented by the multivariate Hawkes process a Piecewise Deterministic Markov process *(PDMP)* which entirely determines the dynamics of the original process.
  After that, with the aim of relating the behaviour of the limit system with the one of the finite size system, we consider the newly built *(PDMP)* and we construct an Approximating Diffusion Process *(ADP)*. Having at disposal these two new processes, in Theorem 5.1 we discover that, as the total number of neurons tends to infinity, the *(PDMP)* - whose dynamics entirely determines the one of the original finite size process- is well approximated by the diffusion and that both processes converge to the limit process.

  After that, we deepen the study of the diffusion process in the particular case of just two interacting populations. The aim is, having in mind the oscillatory behaviour of the limit system *(MFC)*, seeing how the *(ADP)* imitates those oscillations. Towards a Lyapunov argument, we show that once the diffusion enters the basin of attraction of the periodic orbit of the limit system, it keeps entering a compact set- which contains such a basin of attraction- infinitely often, almost surely. This means that, for $N$ -total number of neurons- large enough, the approximating diffusion presents *the same type of oscillations* as the limit system.

- **In Chapter 6** to end up, we provide simulations to highlight the theoretical results achieved throughout the whole thesis: essentially, we first simulate the limit system *(MFC)* and the diffusion approximation *(ADP)* for a fixed number of neurons, to compare their oscillatory behaviours. Secondly, we simulate *(ADP)* in a larger time horizon with respect to the first set and gradually increasing the total number of neurons, to make evidence of the fact *(ADP)* follows the same trend of the limiting system.

# Chapter 1

# Biological overview

The human brain is an unbelievably complex system, spanning several spatial scales of organizations, from microcircuits to whole-brain networks. In order to have in mind -at least from an elementary point of view- which are the internal mechanisms that undergo the functioning of such an impressively structured system, we proceed by giving a brief overview about the nervous system and the processes that drive the information transmission thanks to the electrical impulses fired by neurons.

After that, we give a small outline concerning how mathematical modelling of neuronal networks works; to conclude, we focus on rhythmic behaviours and the emerging of oscillating attitudes in biological systems: this will lead us to the main aim of the current thesis, that is investigating the emergence of oscillatory behaviour in a multi-population system of interacting neurons.

The reader can hence consider this chapter as an introduction in order to have in mind the biological context in which the current thesis is set.

## 1.1 Nervous system and spiking neurons

The nervous system, which coordinates actions and sensory information by transmitting signals to and from different parts of its body, in vertebrates consists of two main parts, the *Central Nervous system* and the *Peripheral Nervous system*, which, respectively, involve the brain and spinal cord on the one side and the nerves -that connect the Central Nervous system to every other part of the body- on the other side.

When considering the Central Nervous system, the brain and the spinal cord play a central role: they are vitally important elements on for the living being, being encased in the bone for protection and located in the dorsal body cavity. The brain, situated in the cranial vault, is for sure the main pillar of the whole system; each brain hemisphere has four sections, called lobes: frontal, parietal, temporal and occipital, and each lobe controls

specific functions. Hence, when modelling the functioning of the brain, it is natural to consider different hierarchical groups or areas in it, and let them interact.

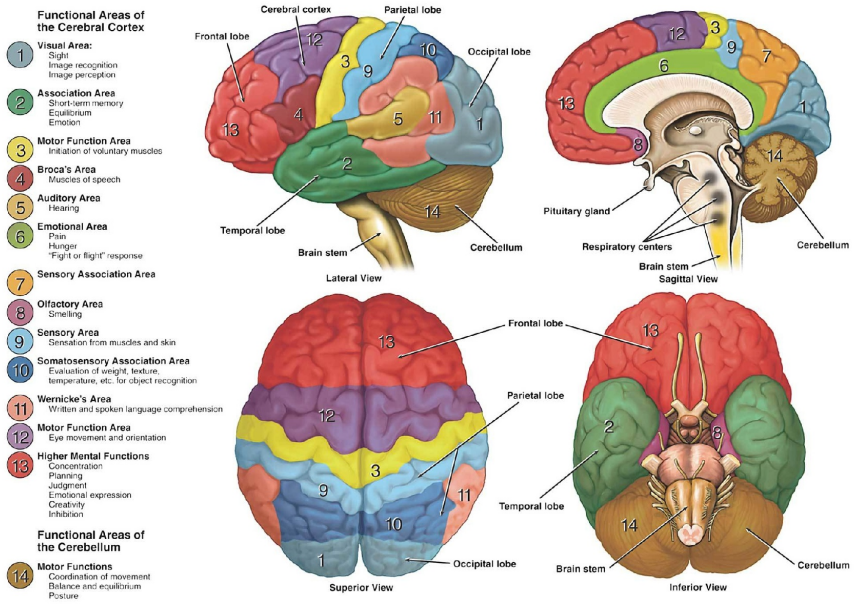



Figure 1.1: Functional areas of the brain. Image credits: https://www.dana.org/article/neuroanatomy-the-basics/

The real motors of the brain are, of course, *neurons*: they consist of highly specialized cells for receiving, processing and transmitting information to other neurons or effector cells via electrical and chemical signals; being electrically excitable, they work through firing electric signals called *action potentials* or *spikes*- which are localised electrochemical phenomena consisting of fast trans-membrane currents of $Na^+$ and $K^+$ ions- across the neuronal network which constitutes the whole brain.

A single neuron is made of a cell body and of "extensions", which are an *axon* on one side and the *dendrites* on the other; the whole structure is surrounded by protective membrane, called *plasma membrane*. The cell body contains the nucleus, and it often has a spherical or pyramidal shape, while the dendrites are branched fibres, with a granular appearance and irregular contours. The axon, instead, is a fibre of variable length with a sharp and regular outline. The axons of human nerve cells are wrapped by *Schwann cells* that form the insulating myelin sheath, which is interrupted at so-called *nodes of Ranvier*.

*How can neurons interact via electrochemical impulses?* First of all, we must point out that the plasma membrane of a neuron is polarised, i.e., it has a

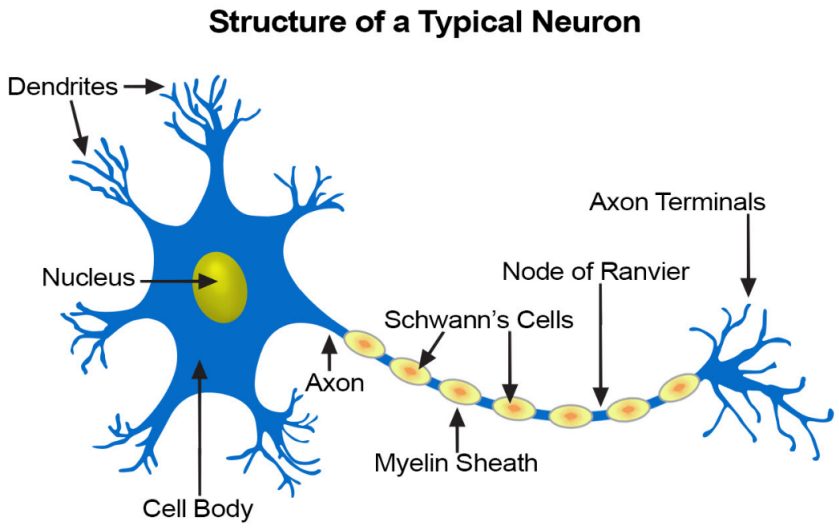



Figure 1.2: What a typical neuron looks like. Image credits: https://training.seer.cancer.gov/anatomy/nervous/tissue.html

difference in electrical charge between the inside and outside of the cell. This difference is due to the concentration of positive ions (mainly sodium ions $Na^+$), which is greater on the outside than on the inside: such an asymmetry produces an electrical potential difference called *resting potential*, which is maintained by the action of a membrane protein - the sodium-potassium pump - which transports $Na^+$ ions from the inside to the outside of the cell and $K^+$ potassium ions from the outside to the inside.

Entering more in detail, $K^+$ ions can pass freely through membrane proteins and tend to balance their concentration, moving from the inside to the outside, while $Na^+$ ions would also distribute themselves evenly by passing through other proteins, but this is not the case because when the neuron is at rest these channels are closed. In this way, the resting potential is kept constant at around $-70$ mV.
Hence, the nerve impulse is generated by a sudden change in this potential difference, and it is always followed by the restoration of normal conditions. In fact, if the neuron is stimulated, due to the opening of certain sodium channel proteins, the membrane potential can rise from about $-70$ mV to about $-50$ mV, and such a value is called the *threshold potential*.
When this threshold value is reached, many sodium channels open and many $Na^+$ ions pass from outside to inside the cell. As the concentration of positive charges inside increases, the potential is suddenly reversed and reaches a value of $+35$ mV, that is the *action potential*. This whole sequence of phenomena is called *membrane depolarisation*.

A few moments later, the *membrane repolarisation* phenomenon takes place: the sodium channel proteins close again, while those for potassium - which were closed in the meantime - reopen and, the sodium-potassium-pump restores the resting condition.

Once that the nerve impulse is generated, it is transmitted along the membrane of nerve cell axons. The nerve impulse proceeds by "jumping" from one node of Ranvier to another and thus advances much faster than it would if it were to travel along the entire axon. In order to have this transportation of the impulse, depolarisation must be transmitted from the area where it occurred to the immediately adjacent tract: the propagation of the stimulus only occurs in one direction due to the fact that, in the tract affected by the action potential, the sodium-potassium pump is working to restore conditions to rest.

An important fact to be highlighted is that the spiking activity is always followed by a refractory period, that turns out in the so called *delay*, in which the same neuron can't emit another firing; during this time, which lasts approximately 2 milliseconds, the membrane cannot receive any stimulus and this period of refractoriness prevents the transmission of depolarisation in the same direction from which it came, effectively preventing the transmission of the impulse from being blocked. More specifically, it happens that the concentration of potassium ions in the interstitial fluid outside the cell is, for a brief instant, higher than under normal conditions at rest; this causes a hyper-polarisation of the membrane which persists until the value of $-70$ mV is restored.

Once understood how the transmission process works, we can now point out how the transmission continues, in order to reach the target cells. Indeed, the situation is the following: sensory neurons, which acquire the sensory stimulus, and association neurons, which perform the function of processing and integration, pass the impulse on to other nerve cells, while motor neurons, on the other hand, pass the stimulus to muscle cells.

At this point the synapses play the crucial role in the transmission of the impulses: indeed, a nerve impulse is transferred from one cell to another and to the muscle cells thanks to synapses. They indeed consists of the contact points between two neurons or between a neuron and a muscle cell, and they transmit the nerve impulse to the next cell, entering in the game in the moment at which the impulse reaches the end of the axon.
Notice that the synapses can be either *chemical* or *electrical*. When considering electrical synapses, the action potential passes from one cell to the next in a mechanism that is very similar to the propagation of the impulse within the cell itself, and in this case the transmission of the nerve impulse can take place in both directions. On the other side, chemical synapses are found where impulse processing is more varied and complex; the trans-

mission of the impulse between one neuron and the next is here mediated by certain chemicals, contained in vesicles, called *neurotransmitters*. Notice that chemical synapses have a small space between the pre-synaptic neuron and the post-synaptic neuron, and it is exactly this separation that prevents the direct transfer of the electrical impulse.

For what concerns the basics needed in order to have the right context in which the current thesis is submerged, the above biological information seem to be enough. We are now ready to open a brief parenthesis about modelling neuronal networks, before passing to the more compelling discussion about rhythmic behaviours in living beings.

## 1.2 Modelling ensembles of interacting neurons

Once understood how the single firing activity of a neuron works, it is useful to focus for a while on the so-called *biological neuron models*, usually denoted as a *spiking neuron models*: they simply consist of a formal mathematical description of the properties of neurons generating their action potentials or spikes; building such models seems to be necessary to study and try to understand the internal mechanisms that undergo the complex network of neuronal connections.

Recall for a while the context we work in: a neuronal network is composed of interacting units -neurons- combined in such a way that the spiking intensity of a single neuron depends not only on the past history of the neuron itself, but also on the action of other neurons in the network. The interaction between neurons, as already pointed out, happens thanks to the synapses: the spike of a pre-synaptic neuron leads to an increase (*excitatory* synapse), or a decrease (*inhibitory* synapse), of the membrane potential (recall that when the membrane potential reaches a certain upper threshold, the neuron is ready to fire a spike) of the post-synaptic neuron, after some delay, and these set of phenomena is usually referred to as *synaptic integration*. Thus, excitatory inputs from the neurons in the network increase the firing intensity, and inhibitory inputs decrease it.
One of the best strategies consists of representing the firing activity of each neuron with a counting process which records the successive times at which the neuron fires a spike; hence, a complete realization of these point processes makes up the spike trains "attached" to each neuron. Indeed, recall that the form and the duration of a spike is always the same: the main information is contained in the sequence of spiking times; that is why we need to focus on time series of spiking events, and this is usually done via counting processes.

More specifically, in the context of this thesis, we represent such a situation using the so-called *Hawkes processes*, widely used in (and born for) the insurance and banking context: due to the structure of their intensity processes,

Hawkes processes seem to be the best tool to model the kind of network we are here dealing with. Their formal definition, together with their properties, is deepened in the next chapter, when discussing all the needed *Mathematical Tools*.

Coming back to a more general situation, and having in mind the functioning of the neuronal activity as above mentioned, we can schematically say that a neuron has an input (its internal dynamics, taking place inside of the cell body) and an output, which generally consists of all functions of discrete or continuous time. For a real neuron, the input is through synapses onto the neuron's dendrites, or onto the neuron's cell body, and the output is through axons that connect it chemically to other neurons. Hence, a formal neuron can be described in terms of its "state" at each time $t$: it is a set of real numbers giving the output and values of internal dynamics at time $t$, and the output, i.e. the spike, can be coded as a 0 or a 1. When the output at time $t$ is 1 the neuron is said to have "fired" at time $t$, i.e., the neuron generated an action potential or spike. Then, one keeps track of all the successive spikes, building in this way the sequence of spiking times, which is the main object of study.
Once modelled the firing activity of the single neuron, it is natural to pass to consider large sets of identical and interacting particles.

We must always remember that, in this whole modelling process, *noise* and *delay* are important factors to be taken into account: for instance, for what concerns delay, it is essential to keep in mind, if considering the successive spiking times as a reference, that every neuron is thought to be subject to a refractory period (with a duration that we can indicate with $r$), after firing an action potential; hence, when looking at populations of interacting neurons we must acknowledge that the proportion of binary neurons that fire in the successive $r$ milliseconds depends both on the proportion of neurons that are not delayed and on the proportion that receives at least threshold excitation.

Considering all these elements, we can point out that are many different ways to model neurons and their functioning: one of the best known and of the first models about spiking neurons, which has also been awarded with the Nobel Prize in 1963, is for sure the *Hodgkin–Huxley model* (for a reference, see [36]): designed in 1952 to explain the mechanisms that undergo the firing of action potentials in the squid giant axon, it consists of the first mathematical description of how action potentials in neurons are initiated and propagated. It uses a continuous time dynamical system's approach, i.e. it approximates with a system of nonlinear differential equations the electrical characteristics of excitable cells such as neurons and muscle cells.

## 1.3 Rhythmic behaviours and oscillations

The analysis of real world and complex systems naturally involves the study of huge families of interacting components: this can be observed in a great variety of areas, such that biology, socioeconomics, medicine, physics, statistical mechanics, and so on and so forth. A component can be a cell, a person, a particle, or, in the case we are going to consider in this work, a neuron being part of a system of interacting populations of neuronal cells.

*How to approach the study of such a large number of interacting units?* One usually tries to model the system, first considering each single unit, and then building a more complex and larger model including $N$ of these -identical- "particles", maybe letting $N$ tend to infinity: complexity and eventual self-organization then arise on a macroscopic scale from the dynamics of these minimal components that evolve coupled by interaction terms.
Once that the model has been formalized, the main aim is to capture and highlight some key behaviours and phenomena, both including the single components and the system considered as a whole unique block-unit.

In our context, in which we consider families of interacting neurons, the key question to be answered, i.e., the key phenomenon to be highlighted, is the following: *can a system exhibit collective periodic behaviour even though the single units have no tendency to behave periodically?* The answer is, simply: *yes*! Indeed, many real systems consisting of large quantities of interacting parts, as, for example, neuronal networks, may exhibit collective periodic behaviour even if the single component itself has no tendency to behave periodically. More generally, we can say that living systems are characterized by the emergence of recurrent dynamical patterns at all scales of magnitude.

In particular, in this thesis we are going to provide evidence of the emergence of *collective periodic behaviour* in the spiking activity of large populations of interacting neurons. Notice that this issue could be widely applied and addressed to many other fields– not only neuronal oscillations: self-organized collective behaviours are as well observed observed both in large communities of microscopic components and on larger levels, such as predator-prey equilibria, applauding audiences etc.; indeed, periodic and oscillating behaviours are the most common observed ways of self-organization in biology, ecology, and socioeconomics, just to have in mind a few of them.

One could also ask: *what are the mechanisms that can enhance the emergence of rhythmic behaviours?* Between many factors, it is of interest -in our context of neuronal networks- to highlight the role of *noise* and *delay*. When talking about delay we refer to the standard refractory period that classically involves a neuron that has just fired a spike.
The concept of noise, instead, is quite more complicated: the activity of a population of neurons embodied in the brain inevitably occurs in the presence of *noise*, which essentially consists of stochastic fluctuations -arising

from internal and external sources- due to thermal energy, ion channel chattering, irregular synaptic inputs from other neurons, and other possible disturbances, such as unreliable neurotransmitter release, and identifying all noise sources may be challenging. The different sources of noise combine so that spiking activity is highly irregular: what we can say is just that the mean spike counts approximately equal to the variance.

Hence, when dealing with large populations of interacting neurons, the resulting ensemble dynamics is a mixture of non-linear neuronal dynamics, inter-neuronal coupling, delays and noise. And all these different factors combine in such a way that the resulting dynamics can turn out to be periodic.

*How to detect such a rhythmic behaviour in practice, when dealing, for instance, with the human brain?* The brain structural networks are typically assessed with diffusion tensor/spectrum imaging techniques, which can provide an image on the complex interconnections that are the paths along the neurophysiological dynamics takes place. One way to measure such neuronal activity is through *functional Magnetic Resonance Imaging (fMRI)*, that has widely highlighted the emergence of collective and coherent brain activities following some recurrent patterns.
It is clear that one of the main aims of the current studies on neurophysiology is to understand the relation between the structure of brain networks and their functions. Simple models of neuronal activity can provide an effective description of the features that enhance the whole-brain spontaneous activity when tuned at their critical point, but rhythms generating mechanism are poorly understood.

Hence, passing to the model object of this work, the objective is to specialize the current discussion about rhythmic behaviours to what really happens in the brain: it is evident that living beings are characterized by biological intrinsic and ubiquitous rhythms which are classically controlled by the brain (just think of the $24h$ for the day-night rhythm of most animals or of limb movements, which are composed of alternating flexions and extensions). The aim is then to investigate the emergence of such a rhythmic tendency, and to understand how these "oscillations" emerge, knowing that the single neuron emits its spikes without following a periodic dynamics: hence, the focus is on the collective behaviour of the whole model, looking for the emergence of self-organized rhythms.

To this purpose of study, the current thesis relies on the work carried out by Susanne Ditlevsen and Eva Löcherbach in their paper [17]. The main motivation for their work, besides the above highlighted interest in finding regular behaviours in complex and noisy systems, comes from the rhythmic scratch like network activity in the turtle, induced by a mechanical stimulus, and recorded and analyzed by Berg and co-workers (for a reference, see, for

example, [5], [6],[7] and [31]).

The objective is then to build a model consisting of large families of interacting neurons which can generate self-sustained periodic behaviour, i.e., oscillations. This will be here done, as above mentioned, by modelling the network via Hawkes processes, and the astonishing result, which turns out to be evident also through the simulations we are going to provide, is that, always having in mind the fact a single neuron doesn't follow a periodic dynamic when firing a spike, a collective periodic behaviour, i.e., oscillations, emerges, highlighting an intrinsic periodicity of the system and a natural tendency to self-organization, enhanced both by delay and noise.

# Chapter 2

# Mathematical Tools

In this chapter we provide a miscellany of preliminary results that can be consulted to better understand the mathematical tools needed in the current thesis. Clearly, the notation used here below is going to be as general as possible: in the subsequent chapters -regarding the specific object of this work- we will then specify notations and symbols, adapting them to the study of the neuronal network in which we are interested.

## 2.1 Point Processes, Counting Processes, Hawkes Processes

In order to understand the way we are going to model the multi-class system of interacting neurons which is the object of the current work, we need a few notation concerning the so-called *point processes* and *counting processes*: such a discussion will then lead us to talk about *Hawkes processes*, that are the kind of instruments we shall use to model properly such a complex network, due to their particular and specific nature.

Following the same notation as in [30], first of all we recall that a **point process** on the non-negative half-line (indexed on times) is a *strictly increasing* sequence of random times, let's call them $(T_i)_{i\geq 1}$, without any accumulation points; we can say that $T_i$ is the time at which a certain event verifies.
Analogously, the associated **counting process** $(N_t)_{t\geq 0}$, with $N_t$ a function defined for $t \geq 0$ and assuming integer and non-negative values, is defined as the number of events of the point process $(T_i)_{i\geq 1}$ that happen *before* time $t$. More formally, indeed, we can write

$$N_t := \sum_{i\geq 1} \mathbb{1}_{[t\geq T_i]}$$

where it is clear that $N_0 = 0$. Notice that $N_t$, counting the number of events happened until time $t$, provides a representation that is equivalent to the one of the correspondent point process.

The simplest -and also most known- type of point process is for sure the so-called **Poisson process** with intensity $\lambda$.

**Definition** (*Poisson Process*)**.** Let $(\tau_i)_{i\geq 1}$ be a sequence of random, independent variables having exponential distribution with parameter $\lambda$, and consider a sequence of times $T_n := \sum_{i=1}^{n} \tau_i$. Then, the process described by the sequence $T_n$ (or equivalently by the associated counting process $N_t$ for $t \geq 0$) is called *Poisson process* with intensity $\lambda$.

A first property of a Poisson process that we can point out, which is inherited by the fact the r.v. $(\tau_i)_{i\geq 1}$ are distributed exponentially, is the so-called *lack of memory*: indeed, the temporal location of future events is independent from the one of past events.

A second important characteristic is the fact that the above described Poisson process is a *homogeneous* point process: in fact, the events are observed with *constant* intensity $\lambda$. A different situation, which gives rise to the so-called *inhomogeneous* Poisson processes, happens when we consider a variable intensity $\lambda(t)$, depending on time and also on the past history of the process, considered up to time $t$.

**Remark.** When considering Poisson processes, and more generally point processes with deterministic intensity, the *waiting times* between two successive events are independent. This doesn't happen when the intensity depends on the history of the process itself, and is then random.

Hence, we can proceed via defining the variable *intensity* of a process as follows:

**Definition** (*Intensity*)**.** The intensity of a point process $(N_t)_{t\geq 0}$ -assuming the existence of the limit- is defined as follows:

$$\lambda(t) := \lambda(t \mid \mathcal{F}_t) = \lim_{h\to 0} \frac{\mathbb{P}(N_{t+h} - N_t = 1 \mid \mathcal{F}_t)}{h}$$

where $\mathcal{F}_t$ is the history of the process until time $t$ and therefore contains the sequence of times $T_i$.
Notice that this kind on intensity *completely characterizes* the distribution of the point process.

### Hawkes Processes

When we need to describe more complicated situations, it is handful to consider a new class of point processes in which the intensity function depends explicitly on past events, hence considering processes that are *self-exciting*: namely, observing an event causes the increment of the intensity function of the process itself.

A well-known example of this kind of processes in represented by the so-called **Hawkes processes**: named after Alan G. Hawkes, they are widely used for statistical modeling of events in mathematical finance, epidemiology, earthquake events, trade orders, bank defaults, gang violence, and other fields in which a random events may exhibit a self-exciting behaviour, like the neuronal network we are going to present in the current work.

We can give a first elementary definition, remaining in the uni-dimensional context:

**Definition.** Let $(N_t)_{t\geq 0}$ be a point process associated to its history $\mathcal{F}_t$, for $t \geq 0$. Recall it is completely determined by its intensity function $\lambda(t)$.
The process is said to be an *Hawkes process* if the intensity function $\lambda(t)$ is in the form:

$$\lambda(t) := \lambda(t \mid \mathcal{F}_t) = \lambda_0(t) + \sum_{i:t>T_i} \phi(t - T_i)$$

where $\lambda_0(t) : \mathbb{R} \to \mathbb{R}_+$ is a function determining the "baseline" intensity of the process, which is independent of other events, while $\phi : \mathbb{R} \to \mathbb{R}_+$ plays the role of a memory kernel.

Notice that Hawkes processes represent a particular class of *inhomogeneous* Poisson processes, in which the intensity function its explicitly depending on past events through the function $\phi(\cdot)$: indeed, all events happening at a time $T_i < t$, i.e. the ones already happened and observed before time $t$, contribute to the intensity function considered at time $t$; this means that each arrival increases the rate of future arrivals for some period of time.

Entering in more detail about the specific structure of the intensity of a Hawkes process we can say what follows:

- $\lambda_0(t) > 0$ describes the observation of those events which are triggered by external factors.
- $\phi(t - T_i)$ is the kernel that modulates the change that an event at time $t$ causes to the intensity function; hence, the *self-exciting* nature of the Hawkes process derives from the summation $\sum_{i:t>T_i} \phi(t - T_i)$.

### 2.1.1 Multivariate versions of the counting processes

We are now ready to specify all these processes to a multi-class context, like the one we are going to consider in this thesis when dealing with populations of interacting neurons. In order to be consistent with the notation we will use when dealing with the neuronal network in object, we generally indicate with $(Z_t^1, \dots, Z_t^N)_{t\geq 0}$ a multivariate counting process. This means that, as in the uni-dimensional case, each component $Z_t^i$ records the number of events related to the $i - th$ component that happens before time $t$, i.e., in

the interval $[0, t]$. As in the standard case, also the multivariate version of the counting process is characterized by its intensity process $(\lambda_t^1, \dots, \lambda_t^N)_{t\geq 0}$ via the relation:

$$\mathbb{P}(Z^i \text{ has a jump in } [t, t+dt] \mid \mathcal{F}_t) = \lambda_t^i \, dt, \quad i = 1, \dots, N.$$

with $\mathcal{F}_t$ denoting, as always, the sigma-field (i.e. the history) generated by the process $(Z^i)_{1\leq i\leq N}$ up to time $t$.

### Multivariate Hawkes Processes

*How to recover the multivariate version of a Hawkes process?* In order to model multivariate and self-exciting Hawkes processes (see [25] and [26]) the intensity process is given in the form:

$$\lambda_t^i = h_i\left(\sum_{j=1}^{N} \int_0^{t-} \phi_{ji}(t-s) dZ_s^j\right)$$

where $\phi_{ji} : [0, \infty) \to \mathbb{R}$ are functions modelling how $Z^j$ influences $Z^i$ via directly influencing its intensity process $\lambda^i$. Moreover, notice that when $\phi_{ji} = 0$, we immediately fall in the standard Poisson process case.

At this point we can pass to a formal definition of a general *multivariate Hawkes process*, which is the one we use from Chapter 3 on to present the neuronal model in object.

Before giving the formal definition, we need to open a brief parenthesis about the setting: we work on a general filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}), \mathbb{P})$, where we consider $(Z_t)_{t\geq 0}$ to be a counting process if it is:

- non-decreasing;
- *càdlàg*, i.e., defined on $\mathbb{R}$, everywhere right-continuous, having left limits everywhere;
- integer-valued and finite for all times;
- such that it has jumps of height 1.

Now, it is also useful to recall the notion of *compensator* of a process: having $(Z_t)_{t\geq 0}$ a $(\mathcal{F}_t)_{t\geq 0}$-adapted counting process, we know it exists a unique non-decreasing predictable process, let's denote it with $(\Lambda_t)_{t\geq 0}$ such that $(Z_t - \Lambda_t)_{t\geq 0}$ is a local martingale; such a process $(\Lambda_t)_{t\geq 0}$ is said the *compensator* of $(X_t)_{t\geq 0}$.

Hence, in this context, consider a directed and countable graph $\mathbb{G} = (\mathcal{S}, \mathcal{E})$ where $\mathcal{S}$ is the set of nodes, $\mathcal{E}$ the set of directed edges (we write $e = (j, i) \in \mathcal{E}$ for the oriented edge). Moreover, we take into account a kernel

$\boldsymbol{\phi} = (\phi_{ji},\ (j,i) \in \mathcal{E})$ with $\phi_{ji} : [0,\infty) \to \mathbb{R}$, and a non-linear intensity component $\boldsymbol{h} = (h_i, i \in \mathcal{S})$ with $h_i : \mathbb{R} \to [0,\infty)$.
Therefore, we can give the following definition of *Hawkes process*, which corresponds to Definition 1 in [15]:

**Definition 2.1** (*Hawkes process*)**.** A Hawkes process with parameters $(\mathbb{G}, \boldsymbol{\phi}, \boldsymbol{h})$ is a family of $(\mathcal{F}_t)_{t\geq 0}$-adapted counting processes $(Z_t^i)_{i\in\mathcal{S}, t\geq 0}$ such that:

1. almost surely, $\forall i \neq j$, $(Z_t^i)_{t\geq 0}$ and $(Z_t^j)_{t\geq 0}$ never jump simultaneously;
2. for every $i \in \mathcal{S}$, the compensator $(\Lambda_t^i)_{t\geq 0}$ of $(Z_t^i)_{t\geq 0}$ has the form $\Lambda_t^i = \int_0^t \lambda_s^i ds$, where the intensity process $(\lambda_t^i)_{t\geq 0}$ is given by

$$\lambda_t^i = h_i\left(\sum_{j\to i} \int_0^{t-} \phi_{ji}(t-s) dZ_s^j\right)$$

where $\sum_{j\to i}$ stands for a sum over the edges of type $\{j : (j,i) \in \mathcal{E}\}$.

Moreover, notice that we generally say that a Hawkes process is *linear* when $h_i(x) = \mu_i + x\ \forall\ x \in \mathbb{R}$, $i \in \mathcal{S}$, $\mu_i \geq 0$, $\phi_{ji} \geq 0$.

*How does the self-exciting process work?* Considering a Hawkes process like the one in the above definition, let's consider a vertex $i \in \mathcal{S}$: then, looking at $Z^i$, its rate of jump at time $t$, i.e., its intensity, is given by

$$\lambda_i(t) = h_i(\sum_{j\to i} \sum_{k\geq 1} \phi_{ji}(t - T_k^j) \mathbb{1}_{\{T_k^j < t\}}),$$

$(T_k^j)_{k\geq 1}$ being the time occurrences of the jumps of the process $Z^j$.

Looking at this situation in a self-excitation context, the fact is that each time one of the $Z^j$'s jumps, it excites its neighbours increasing their rate of jump (when considering $h$ increasing and $\phi$ positive).

We can give an equivalent definition of Hawkes process (to check the actual equivalence the reader is invited to refer to Proposition 3 of [15]); such a new definition is going to be useful, in particular in the proof of Theorem 3.6, where we need to study Hawkes processes in terms of a system of Poisson-driven stochastic differential equations, in order to speak of existence and uniqueness and to prove a propagation of chaos result via a coupling argument.

Then, the situation is the following: working on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and considering a family $(\pi^i(ds,dz), i \in \mathcal{S})$ of i.i.d. $(\mathcal{F}_t)_{t\geq 0}$-Poisson measures with intensity measure $dsdz$ on $[0,\infty)\times[0,\infty)$ (the reader can refer to Section 2.2 to deepen the concept of Poisson random measure), then:

**Definition 2.2.** A family $(Z_t^i)_{i\in\mathcal{S},t\geq 0}$ of càdlàg $(\mathcal{F}_t)_{t\geq 0}$-adapted processes is called a Hawkes process with parameters $(\mathbb{G}, \boldsymbol{\phi}, \boldsymbol{h})$ if a.s., for all $i \in \mathcal{S}$, all $t \geq 0$

$$Z^i = \int_0^t \int_0^\infty \mathbb{1}_{\{z\leq h_i(\sum_{j\to i}\int_0^{s-}\phi_{ji}(s-u)dZ_u^j)\}}\pi^i(dsdz)$$

## 2.2 Poisson Random Measures

Generally speaking, as *continuous processes* can be represented by means of a driving underlying Brownian Motion, *jump processes* can be treated using an underlying discrete noise, i.e., by a *Poisson Random Measure* (PRM).

**Definition 2.3** (*Poisson Random Measure*)**.** Let $(Z_n)_{n\geq 1}$ be random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values on $\mathbb{R}_+ \times \mathbb{R}_+$. It is possible to associate to this sequence of r.v. a random counting measure defining

$$N := \sum_{n=1}^{\infty} \delta_{Z_n}$$

If considering $\mu$ a $\sigma$-finite measure on $\mathbb{R}_+ \times \mathbb{R}_+$, then $N$ is said to be a *Poisson Random Measure* (PRM) having intensity measure $\mu$ if the followings hold:

- $\forall\ A$ in $\mathcal{B}(\mathbb{R}_+ \times \mathbb{R}_+)$,

  $$N(A) = \sum_{n=1}^{\infty} \mathbb{1}_A(Z_n) \sim Poiss(\mu(A))$$

  considering $N(A) = +\infty$ a.s. if it holds $\mu(A) = \infty$ and $N(A) = 0$ if $\mu(A) = 0$.

- $\forall\ n$, $\forall\ A_1, \ldots, A_n$ *in* $\mathcal{B}(\mathbb{R}_+ \times \mathbb{R}_+)$, mutually disjoint, it holds

  $$N(A_1), \ldots, N(A_n) \text{ are } \perp\!\!\!\perp$$

  where the symbol $\perp\!\!\!\perp$ indicates mutual independence between the variables.

In the current discussion it shall be convenient to set the measure $\mu$ as equal to the Lebesgue measure.

**Remark** (*About PRMs*)**.**

- Setting the filtration $\mathcal{F}_t^N := \sigma(N(A) : A \subset [0,t] \times \mathbb{R}_+)$ for $t \geq 0$, and letting $(\lambda(t))_{t\geq 0}$ be a $(\mathcal{F}_t^N)$-predictable process taking values in $\mathbb{R}_+$, put

  $$Z_t := \int_{[0,t]} \int_{\mathbb{R}_+} \mathbb{1}_{\{z\leq\lambda(s)\}} N(ds, dz)$$

  Then $\lambda(t)$ is an $\mathcal{F}_t^N$-intensity of $Z$, and we say the process $(Z(t))_{t\geq 0}$ is obtained from *thinning* of $N$.

- Let $\psi(s,z,w)$ be $(\mathcal{F}_t)$-predictable and such that

$$\mathbb{E}\left[\int_0^t \int_{\mathbb{R}_+} |\psi(s,z)| ds dz\right] < \infty \ \forall \ t \geq 0$$

Then,

$$\int_0^t \int_{\mathbb{R}_+} \psi(s,z) N(ds,dz) - \int_0^t \int_{\mathbb{R}_+} \psi(s,z) ds dz$$

is a $(\mathcal{F}_t^N)$-martingale.

The reader is invited to notice that such a construction via underlying PRMs helps for *coupling*, which is one of the most common available ways to prove a result of propagation of chaos; indeed, will use this technique in the proof of Theorem 3.6, to prove a propagation of chaos result related to our model of interacting neurons; we will come back on this topic in Section 3.3. For the time being, we invite the reader to consult [10], (Chapter 4, Section 4.1 -in particular Definition 4.1- which presents the best known coupling methods in the context of proving propagation of chaos) or [16] and [39] for a general overview on coupling methods in probability theory.

## 2.3 Basics on Renewal Theory

Generally speaking, *Renewal Theory* is the branch of probability theory that generalizes the Poisson processes for arbitrary holding times; indeed, instead of exponentially distributed holding times, a renewal process may have any i.i.d. holding times, which must have finite mean. It is concerned with the study of the so-called *renewal equation*, representing one of the essential theories of probability, being strictly connected, for example, with the study of regenerative processes.

In the current work, we use its tools to prove an essential Central Limit theorem result in Chapter 3; hence, we need to proceed giving some general basics -following the ones presented in [8] (Chapter 4)- about renewal theory that shall be henceforth useful.

### 2.3.1 Renewal Point Processes

We consider an i.i.d. sequence $\{S_n\}_{n\geq 1}$ of non-negative random variables with common cumulative distribution

$$F(x) := \mathbb{P}(S_n \leq x)$$

The distribution $F$ is either called *defective* if $F(\infty) < 1$ or *proper* if $F(\infty) = 1$. Given this, we call the sequence of variables $\{S_n\}_{n\geq 1}$ the *inter-renewal-sequence*.

We naturally associate to $\{S_n\}_{n\geq 1}$ another sequence $\{T_n\}_{n\geq 0}$, which is the *associated renewal sequence*, defined by:

$$T_n := T_{n-1} + S_n \quad for\ n \geq 1$$

where the *initial delay* $T_0$ is a finite and non-negative random variable, which is independent of the inter-renewal sequence $\{S_n\}_{n\geq 1}$, and time $T_n$ is called *renewal time* (or an *event*). Notice that if $T_0 = 0$ we say the renewal sequence is an *undelayed* sequence.

To any renewal sequence as the one above described, we can associate the following stochastic process

$$N([0,t]) := \sum_{n\geq 0} \mathbb{1}_{\{T_n \leq t\}} \quad t \geq 0$$

which is the counting process recording the number of events in the closed interval $[0,t]$. We can clearly point out that the function $t \mapsto N([0,t])$ is a right-continuous function having limit on the left for each $t > 0$, i.e., $N[0,t)$.

Having in mind these objects, we now present a list of important results, notations and definitions which we shall use in the sequel. For shortness, we omit the proofs.

The first result we can highlight, concerning the counting process $N([0,t])$, is the following:

**Theorem.** $\forall\ t \geq 0$, *it holds that* $\mathbb{E}[N([0,t])] < \infty$. *In particular, a.s.* $N([0,t]) < \infty\ \forall\ t \geq 0$.

*Proof.* For a proof of this result, see Theorem 4.1.1 in [8]. □

At this stage, we are ready to give the proper definition of *renewal function*, which is essential for the further developing of the theory:

**Definition.** The function $R : \mathbb{R} \to \bar{\mathbb{R}}_+$ defined as:

$$R(t) := \mathbb{E}[N([0,t])]$$

with $N$ being the counting process associated to the undelayed renewal sequence, is said the *renewal function*.

Since $R$ is right-continuous and non-decreasing, we can associated a unique measure -called the *renewal measure*, and sometimes denoted, again, with the letter $R$- $\mu_R$ on $\mathbb{R}_+$ in such a way that $\mu_R([0,t]) = R(t)$; notice that $\mu_R(\{0\}) = R(0) = 1$.

**Example** (*The Poisson Process*)**.** We report a first example in order to make everything clear: considering the case of the exponential inter-event times having cumulative distribution $F(t) = 1 - e^{-\lambda t}$ for $t \geq 0$; we have that the undelayed renewal process is then an Homogeneous Poisson Process with intensity $\lambda$, to which a point at time time $t$ is added; hence, it is clear that the correspondent renewal function is $R(t) = 1 + \lambda t$.

It is often convenient to express the function $R$ in terms of the cumulative distribution $F$: to this aim, just take the undelayed case and observe that we can express time $T_n$ as

$$T_n := S_1 + \cdots + S_n$$

i.e., as the sum of $n$ independent random variables having $F$ as cumulative distribution; hence we can write

$$\mathbb{P}(T_n \leq t) = F^{*n}(t)$$

where $F^{*n}$ stands for the $n$-fold convolution of the function $F$, and its classically defined in a recursive way as:

$$F^{*0}(t) = \mathbb{1}_{[0,\infty)}(t) \qquad F^{*n}(t) = \int_{[0,t]} F^{*(n-1)}(t-s)dF(s) \quad n \geq 1$$

The role of 0 in the above integral is the following:

$$\int_{[0,t]} \phi(s)dF(s) = \phi(0)F(0) + \int_{(0,t]} \phi(s)dF(s)$$

Given this notation, the key point of this discussion now emerges: indeed we can write

$$E[N([0,t])] = E[1 + \sum_{n\geq 1} \mathbb{1}_{\{T_n \leq t\}}] = 1 + \sum_{n\geq 1} \mathbb{P}(T_n \leq t)$$

which leads to

$$R(t) = \sum_{n=0}^{\infty} F^{*n}(t)$$

**Theorem.** *It holds that:*

$$\mathbb{P}(S_1 < \infty) < 1 \iff \mathbb{P}(N([0,\infty)) < \infty) = 1 \iff E[N([0,\infty))] < \infty$$

*Proof.* For a proof, see Theorem 4.1.4 in [8]. □

We conclude this section with a bit of nomenclature: a renewal process is either said *recurrent* or *transient* depending on the fact the cumulative function $F$ is proper of defective, i.e., respectively, if $\mathbb{P}(S_1 < \infty) = 1$ and $\mathbb{P}(S_1 < \infty) < 1$.

### 2.3.2 The Renewal Equation

In the previous section we set the basics about renewal theory; at this point, we have to get to the gist of it, via defining the *renewal equation*.

Let $F : \mathbb{R}_+ \to \mathbb{R}_+$ be a generalized cumulative distribution function on $\mathbb{R}_+$, meaning $F = cG$ with $c > 0$ a constant and $G$ a proper cumulative distribution function of a non-negative real random variable. Then, by means of convolution, we define the *renewal equation* as

$$f = g + f * F \qquad (\textit{Renewal equation})$$

i.e.,

$$f(t) = g(t) + \int_{[0,t]} f(t-s)dF(s) \quad t \geq 0$$

where $g : \mathbb{R}_+ \to \mathbb{R}$ is a measurable function said *data*.
The nomenclature related to the renewal equation is the following: we say it is *proper* if $F(\infty) = 1$, *defective* if $F(\infty) < 1$, *excessive* if $F(\infty) > 1$.

At this point, we get to the key result:

**Theorem.** *The renewal function $R$ satisfies the so-called Fundamental Renewal equation*

$$R = 1 + R * F$$

*Proof.* For a proof of this result, see Theorem 4.1.6 in [8] □

One of the best known examples of models represented towards means of renewal theory and renewal equations is for sure the *Lotka-Volterra population model*, which features the evolution of a population of women: much literature has been produced on it, this being one of the most famous models about populations dynamics.

Despite what one could think, it is relatively easy to obtain an expression for the solution of the renewal equation in terms of the renewal function $R$. Indeed, when taking the data $g$ as locally bounded, it holds the following:

**Theorem.** *If $F(\infty) \leq 1$, the renewal equation $f = g + f * F$ admits a unique and locally bounded solution $f : \mathbb{R}_+ \to \mathbb{R}$ given by $f = g * R$, i.e.,*

$$f(t) = \int_{(0,t]} g(t-s)dR(s)$$

*Proof.* For a proof, we invite the reader to refer to Theorem 4.1.14 in [8] □

We could say much more in relation to this topic, but these are the basics that suffice for us to understand the demonstration strategies used in Chapter 3, where we adapt these tools to the multidimensional case using matrix renewal equations strategies.

## 2.4 Basics about Markov processes: generators and semigroups

In order to deal properly with the main tools used in Chapter 5, we need some additional basic concepts related to Markov processes: the aim is to introduce the notions of *infinitesimal generator* and *semigroup*. This will be done using the same notation as in [9].
It will also be of interest for a further developing to deepen the relation between these objects and some specific PDEs, like the *Fokker-Planck* equation.

First of all, recall that a stochastic process $X = \{X_t\}_{t\geq 0}$ taking values in $\mathbb{R}^d$ is said to be a *Markov process* if $\forall\ 0 \leq s \leq t < \infty$ and $\forall\ B \in \mathcal{B}(\mathbb{R}^d)$ it holds that:

$$\mathbb{P}(X_t \in B \mid \mathcal{F}_s) = \mathbb{P}(X_t \in B \mid X_s)$$

meaning that the conditional law of $X_t$ knowing the whole history of the process until time $s$ (represented by the filtration $\mathcal{F}_s$) is only a function of $X_s$: the future depends on the past only through the present state.

Given this, and for $0 \leq s < t < \infty$, we define the *transition nucleus* $Q_{s,t}(x, dy)$ as any regular version of $X_t$ with respect to $X_s$: hence, $Q$ is a probability nucleus such that

$$\mathbb{P}(X_t \in dy \mid \mathcal{F}_t) = Q_{s,t}(X_s, dy)$$

Moreover, we say $X$ is a *homogeneous* Markov process if $Q_{s,t}(x, dy) = Q_{t-s}(x, dy)$, meaning that the conditional law of $X_{t+h}$ with respect to $X_t$ doesn't depend on the specific time $t$, but only on $h$. For $t, h \geq 0$ we hence have:

$$\int_{y\in\mathbb{R}^d} Q_s(x, dy)Q_t(y, dz) = Q_{t+s}(x, dz)$$

also known as the *Chapman-Kolmogorov Equation*.

At this point, and from now on, assume to work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and let $X$ be a diffusion (see Section 2.5.1), i.e. such that $dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t$. Notice that in the following we shall indicate with $\mathbb{P} = \mathbb{P}_x$ the dependence on $x$ in the probability.

The key point -that is the reason why we widely use the notions we are here going to introduce in relation to diffusion processes in Chapter 5- is that *diffusion processes are Markov and homogeneous processes*, meaning that:

$$\mathbb{P}_x(X_{t+h} \in B \mid \mathcal{F}_t) = \mathbb{P}_z(X_h \in B)_{|_{z=X_t}} = Q_h(X_t, B)$$

where $Q_h(x, dy) = \mathbb{P}_x(X_h \in dy)$ is the transition semigroup.

This shall be more clear when getting to Section 2.5.1 and deepening the proper concept of diffusion process.

### Infinitesimal generators

A Markov process $X$ is characterized by its infinitesimal behaviour through the so-called *infinitesimal generator*, which we indicate with $A$: consider $C_0(\mathbb{R}^d, \mathbb{R})$ the space of continuous functions tending to zero as $|x| \to \infty$. The infinitesimal generator of $X$ is defined as follows:

$$Af(x) := \lim_{t \to 0} \frac{E_x[f(X_t)] - f(x)}{t}$$

for all those functions $f$ in $C_0(\mathbb{R}^d, \mathbb{R})$ such that the limit exists (and such a set of functions is the *domain* of the generator).
It is possible to prove that the generator determines completely the law of the Markov process!

In the case of a diffusion, the infinitesimal generator $A$ is the following differential operator:

$$Af(x) = \sum_{i=1}^{d} b_i(x) \frac{\partial}{\partial x_i} f(x) + \frac{1}{2} \sum_{i,j=1}^{d} (\sigma\sigma^t)_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x)$$

### Markov semigroups

In addition to the concept of infinitesimal generator, we naturally associate to a Markov process $X$ a family of operators $\{P_t\}_{t \geq 0}$ which are called *semigroups of the process*, acting on $f \in C_0(\mathbb{R}^d, \mathbb{R})$ as follows:

$$P_t f(x) := E_x[f(X_t)] = \int_{\mathbb{R}^n} f(y) Q_t(x, dy)$$

Together with the notion of infinitesimal generator, the semigroups will be essential to handle all the results of Chapter 5.

For what concerns the basic properties of these objects, first of all we can point out that $P_t f \in C_0(\mathbb{R}^d, \mathbb{R})$ $\forall f \in C_0(\mathbb{R}^n, \mathbb{R})$. Moreover, we can notice that the name "semigroup" comes from the fact that it holds that $P_t \circ P_s = P_{t+s}$, i.e., $P_t(P_s f) = P_{t+s}$ $\forall\ s, t \geq 0$, for all $f \in C_0(\mathbb{R}^d, \mathbb{R})$.

Notice that the generator $A$ completely determines the semigroup $P_t$!

### Kolmogorov equations

At this point, we can open a brief parenthesis related to the connection between the above mentioned objects and a particular class of PDEs, the *Kolmogorov backward and forward equations*, which are strictly connected with the *Fokker-Planck equation*.

Let $X$ be a diffusion with an absolutely continuous (for all $t > 0$) transition nucleus $Q_t(x, dy) = P_x(X_t \in dy)$; this simply means that it exists a measurable function $q_t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ such that

$$Q_t(x, dy) = q_t(x, y)dy \quad \forall\, t > 0, \forall\, x, y \in \mathbb{R}^d$$

Under suitable assumptions one can show that $q_t(x, y)$ satisfies the *Kolmogorov backward equation*, for all $y \in \mathbb{R}^d$ fixed:

$$\frac{\partial}{\partial t} q_t(x, y) = A_x q_t(x, y) \quad \forall\, t > 0, \forall\, x \in \mathbb{R}^d$$

where $A_x$ means the generator acts on the $x$-variable of $q_t(x, y)$.

Moreover -that's the key point- it also holds the so-called *Kolmogorov forward equation*, alternatively called exactly *Fokker-Planck equation*, for all $x \in \mathbb{R}^d$ fixed:

$$\frac{\partial}{\partial t} q_t(x, y) = A_y^* q_t(x, y) \quad \forall\, t > 0, \forall\, y \in \mathbb{R}^d$$

where $A^*$ is the adjoint operator of $A$, given by:

$$A^* f(y) = \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial y_i \partial y_j} [(\sigma\sigma^t)_{ij}(y) f(y)] - \sum_{i=1}^{d} \frac{\partial}{\partial y_i} [b_i(y) f(y)]$$

Hence, it is clear that the *Fokker-Planck* equation, whose importance in the context of mathematical modelling can be deepened by the reader by consulting [37], plays a crucial role when dealing with diffusion processes: all these connections between ensembles of interacting neurons and diffusion processes will be clearer in Chapter 5, where we build a diffusive approximation of the neuronal network model at stake.

## 2.5 Diffusions and Piecewise Deterministic Markov Processes

In the previous section the concept of diffusion has emerged: it is therefore necessary, at this point, to introduce a brief background on stochastic processes in order to better understand what we are dealing with.

Almost all continuous-time stochastic process models of applied probability consist of some combination of the following three main elements: *diffusion*, *deterministic motion*, *random jumps*. Usually, the main approach to model diffusion is related to Itô calculus and stochastic differential equations, while an heterogeneous collection of models is used to deal with deterministic motion and random jumps, i.e. to handle all non-diffusion applications; the most important historical change of pace in this context happened in 1984,

when M.H.A. Davis introduced the class of *PDMPs* (i.e. Piecewise Deterministic Markov Processes), providing a general family of stochastic models covering virtually all *non-diffusion applications*, that can be analyzed via methods that are analogous to the ones used in diffusion theory: the aim is putting all non-diffusion models on the same footing as diffusion theory. We explore this kind of processes in detail in the upcoming sections.

### 2.5.1 Basics about Diffusion Processes

Generally speaking, *diffusion processes* are a class of continuous time Markov processes with almost surely continuous sample paths, for which the Kolmogorov forward equation is the *Fokker-Planck* equation, and are used to model many real-life stochastic systems.

From a mathematical point of view, a $d$-vector valued process $(X_t)$ is generally represented as the solution of the Itô stochastic differential equation:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t \tag{2.1}$$

i.e. via a deterministic (drift) part $b$ and some noise which is driven by (or approximated by) a Brownian motion. Indeed, in (2.1), $B_t = (B_t^1, \cdots, B_t^r)'$ represents a vector of independent Brownian motions, while $b(t,x)$ and $\sigma(t,x)$ are respectively a $d$-vector and a $d \times r$-vector satisfying some smoothness conditions in the variables $(t,x)$ to ensure the uniqueness of the solution.

To enter in more detail, the increments $X_{t+dt} - X_t$ of the process $(X_t)$ have *mean* equal to $b(t, X_t)$ and covariance given by $a(t, X_t)dt = \sigma(t, X_t)\sigma'(t, X_t)dt$.

More precisely, and as already mentioned, $(X_t)$ is a Markov process with continuous sample paths and having differential generator $A$ given by:

$$Af(x) = \sum_{i,j=1} a_{ij}(t,x)\frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1} b_i(t,x)\frac{\partial f}{\partial x_i}$$

for $f \in C^2(\mathbb{R})$.

Under some smoothness assumptions, the main interpretation of what above is that the density function $p(t,x)$ of $(X_t)$ satisfies the *Fokker-Planck Equation*:

$$\begin{cases} \frac{\partial p}{\partial t}(t,x) = A^* p(t,x) & t > 0 \\ p(0,x) = p_0(x) \end{cases}$$

where $p_0(x)$ is the given density of $x_0$ and $A^*$ is the formal adjoint of the operator $A$.

### The Approximating Diffusion Technique

Since in Chapter 5 we are going to build an *approximating diffusion process* for the process at stake, we can now pass to a brief analysis of the widely used technique of the *diffusion approximation*. To deepen this kind of topic, and to know more about what we are here going to present from a general point of view, the reader is invited to consult [21].

The general idea underlying this technique is the following: we want to replace a complicated and analytically intractable stochastic process by an appropriate diffusion process, which simply is -as above mentioned- a Markov process having continuous sample paths; the approach under the application of a diffusion approximation can be compared, as we are going to see in a while, to the standard normal approximation for sums of random variables using to the well-known Central Limit Theorem.

Entering in more detail, assume to be dealing with $X$ an analytically intractable process: the aim is then finding a diffusion process $Y$ such that the distribution of $X$ can be approximated by the one of $Y$ is a weak sense (meaning convergence in distribution), in the sense that

$$X \overset{d}{\approx} Y$$

The standard approach in this case is to phrase an approximation in terms of a limit theorem: hence we want to find a sequence $Y_n = \{X_n(t) : t \geq 0\}$ such that $X$ may be identified with $Y_n$ for $n$ large, and then $X \overset{d}{\approx} Y$ equals to $Y_n \xrightarrow{w} Y$ in distribution.

Hence, first of all we need to make explicit the definition of *weak convergence*:

**Definition** (*Weak convergence*)**.** Consider $\{Y_n : n \geq 0\}$ a sequence of random variables. We say it converges weakly to $Y$ if and only if there exists $(\Omega, \mathcal{F}, \mathbb{P})$ probability space supporting a family of random variables $\{Y', Y'_n : n \geq 0\}$ such that:

- $Y_n \overset{d}{=} Y'_n$;
- $Y \overset{d}{=} Y'$;
- $Y'_n \xrightarrow[n\to\infty]{} Y'$ on $(\Omega, \mathcal{F}, \mathbb{P})$;

where the last item of the list refers to the notion of convergence in a function space; therefore, it is necessary to define a metric on such a space in order to have that $Y'_n \to Y' \iff d(Y'_n, Y) \to 0$. In order to do that, the most used function space is the space of càdlàg functions, which we denote with $D_E[0,\infty) = \{w : [0,\infty) \to E \mid w(\cdot)\ is\ right\ continuous\ for\ every\ t \geq$

$0$ *and has left limits at every* $t \geq 0\}$, with $E$ an Euclidean space. And the most suitable metric for this function space is for sure the *Skorokhod metric* $d$ on $D_E[0,\infty)$ (see [32] to get further information about this topology).

The most important result which comes into play at this stage is the famous *Donsker's Theorem*, which is a well-known limit theorem describing the behaviour of the $\mathbb{R}^d$-valued random walk over long intervals of time, indicating it is well approximated by Brownian motion.

Hence, consider $\{Z_n : n \geq 0\}$ an i.i.d. sequence of $\mathbb{R}^d$-valued random variables, and let $S_n = Z_1 + \cdots + Z_n$ (with $S_0 = 0$) be the $\mathbb{R}^d$-valued random walk.
Having in mind the classical results given by the Strong Law of Large Numbers (*SLLN*) and the Central Limit Theorem (*CLT*), the idea underlying the building of approximating diffusion process is to recover a process-valued version of them.

Indeed, considering the process $\bar{Y}n(t) = S_{[nt]}/n$, one can immediately show that a functional *SLLN* (i.e. a process-valued one) holds, in the sense that $d(\bar{Y}_n, \bar{Y}) \xrightarrow[n\to\infty]{} 0$ a.s. where $\bar{Y}(t) = \mu t$.
In the same context, a version of the *CLT* can be pointed out: just consider the process $Y_n(t) = \sqrt{n}(S_{[nt]}/n - \mu t) = \sqrt{n}(\bar{Y}_n(t) - \bar{Y}(t))$. It is possible to show that $Y_n(t) \xrightarrow{w} \Sigma^{1/2}N(0, t\mathbb{1})$ for all $t \geq 0$ (with $\Sigma$ the covariance matrix). And for finite-dimensional distributions of $Y_n$ we have that:

$$(Y_n(t_1), \cdots, Y_n(t_n))) \xrightarrow{w} (Y(t_1), \cdots, Y(t_n))$$

which means a convergence in distribution to a multivariate normal vector such that the mean is $E[Y(t_i)] = 0$ and variance is $E[X^t(t_i)X(t_j)] = \Sigma_{\{t_i,t_j\}}$.

At this stage the key point is the following: since the only process supported on $D_E[0,\infty)$ with finite dimensional distributions described by the above properties of the mean and the covariance is the Brownian-motion process, then, if $Y \in D_E[0,\infty)$, it must be $Y \overset{d}{=} \Sigma^{1/2}B$, where $B(\cdot)$ is a standard Brownian motion process on $\mathbb{R}^d$.

Having in mind what above, we are ready to state the famous *Donsker's Theorem*, also called *Functional CLT*, which gives a precise statement of the limit behaviour of the process $\{Y_n : n \geq 0\}$.

**Theorem.** *If* $E\|Z_n\|^2 < \infty$ *then* $Y_n \xrightarrow[n\to\infty]{w} \Sigma^{1/2}B$ *in* $D_E[0,\infty)$

Since $\Sigma^{1/2}B$ is the most fundamental diffusion process, this theorem gives the prototypical diffusion approximation, suggesting the common features a diffusion approximation should have, in the form:

$$S_{[nt]} \overset{d}{\approx} \mu nt + \sqrt{n}\Sigma^{1/2}B(t)$$

This means that the limit process used to approximate $S_{[nt]}$ depends on the random-walk only through the mean vector $\mu = E[Z_n]$ and the covariance matrix $\Sigma$ of $Z_n$, and this is essentially what we do in Section 5.2, adapting this kind of approximation to our neuronal context.

### 2.5.2 About Piecewise Deterministic Markov Processes

Once deeply investigated the class of diffusion processes, we can pass to the analysis of *non-diffusion models*, via introducing the already mentioned *Piecewise Deterministic Markov Processes* -from now on *PDMPs*. This kind of processes will be essential in Chapter 5 -Section 5.1- where we build an associated *PDMP* in relation to the neuronal model at stake.

Before passing to the formal definition, we start by referring the reader to [14] in order to deepen the simpler class of *piecewise-linear processes*, to become aware of several examples of models, e.g. in queuing theory, which are built using this kind of processes, and to get to know more in detail the huge class of *PMDPs*.

As already mentioned in the introduction of this section, a *PDMP* is a process in which the behaviour is generated by *random jumps and deterministic motion*, meaning that the motion consists of random jumps at points in time whose evolution is deterministically governed by an ordinary differential equation. These processes are characterized by three quantities: the *flow*, the *jump rate*, and a *transition measure*, which we are now going to analyze. The reader is also invited to notice that, for the time being, we are simply considering *piecewise deterministic processes*: Markovianity will naturally come later as a natural consequence/property.

Hence, let us consider the following state space

$$E = \cup_{\nu \in K} M_\nu = \{(\nu, \xi) : \nu \in K, \xi \in M_\nu\}$$

where $K$ is a countable set, $d : K \to \mathbb{N}$ a given function, and for all $\nu \in K$, $M_\nu$ is an open subset of $\mathbb{R}^{d(\nu)}$.
Then, take into account $\mathcal{E}$ the class of measurable sets in $E$, i.e.,

$$\mathcal{E} = \{\cup_{\nu \in K} A_\nu \in \mathcal{M}_\nu\}$$

with $\mathcal{M}_\nu$ being the Borel sets of $M_\nu$.
Hence, $(E, \mathcal{E})$ is a Borel space, and the state space of the process is denoted by $X_t = (\nu_t, \xi_t)$.

In this context, a probability law of $(X_t)$ is determined by the following objects:

- the vector fields $(\mathcal{H}_\nu, \nu \in K)$;

- A measurable function $\lambda : E \to \mathbb{R}_+$;
- A transition measure $Q : \mathcal{E} \times (E \cup \Gamma^*) \to [0, 1]$

where the vector fields must satisfy the following equation:

$$\begin{cases} \frac{d}{dt} f(\xi_t) = \mathcal{H} f(\xi_t) \quad \forall\ f \text{ smooth} \\ \xi_0 = z \end{cases} \tag{2.2}$$

with $\xi(t, x)$ being the integral curve of the vector field $\mathcal{H}$. More precisely, the vector fields $\mathcal{H}_\nu$ are supposed to be such that there is a unique integral curve $\phi_\nu(t, z)$ satisfying (2.2) with $\mathcal{H} = \mathcal{H}_\nu$ and $\xi_t = \phi_\nu(t, z)$.

Let now us consider:

- $\partial M_\nu$ as the boundary of $M_\nu$;
- $\partial^* M_\nu = \{z \in \partial M_\nu : \phi_\nu(t, \xi) = z, \exists\ (t, \xi) \in \mathbb{R}_+ \times M_\nu\}$, i.e. the boundary points at which the integral curves of $M_\nu$ exit from $M_\nu$.

Moreover, we set:

$$\Gamma^* = \cup_{\nu \in K} \partial^* M_\nu$$

For $x = (\nu, \xi) \in E$ we define:

$$t^*(x) = \inf\{t > 0 : \phi_\nu(t, \xi) \in \partial^* M_\nu\}$$

And to conclude we write $\mathcal{H}h(x)$ for the action of the vector fields $\mathcal{H}_\nu$ on functions $h : E \to \mathbb{R}$ at $x = (\nu, \xi) \in E$.

For what concerns the intensity $\lambda$, we suppose that for each $(\nu, \xi) \in E$ there exists $\epsilon > 0$ such that the function $s \mapsto \lambda(\nu, \phi_\nu(s, \xi))$ is integrable for $s \in [0, \epsilon[$.

It remains to specify the transition measure $\mathcal{Q}(A; x)$: it is a measurable function of $x$ for each fixed $A \in \mathcal{E}$, defined for $x \in E \cup \Gamma^*$, being essentially a probability measure on the state space $(E, \mathcal{E})$ for each $x \in E$.

At this point, everything has been defined, so that we can build the motion of the process $(X_t)$ starting from the point $x = (n, z) \in E$.

*How to build such a process?* We define the function $F$ as follows:

$$F(t) = \begin{cases} \exp\left( - \int_0^t \lambda(n, \phi_n(s, z)) ds \right) & \text{if } t < t^*(x) \\ 0 & \text{if } t \geq t^*(x) \end{cases} \tag{2.3}$$

Then, we select a random variable $T_1$ such that $\mathbb{P}[T_1 > t] = F(t)$ (i.e. the jump happens after time $t$ with a probability given by the function $F(t)$). After that, we choose, independently, an $E$-valued random variable $(N, Z)$

having distribution $\mathcal{Q}(\cdot;\phi_n(T_1,z))$. The trajectory of $(X_t)$ for $t \leq T_1$ hence is:

$$X_t = (\nu_t, \xi_t) \begin{cases} (n, \phi_n(t,z)) & \text{if } t < T_1 \\ (N, Z) & \text{if } t = T_1 \end{cases} \tag{2.4}$$

Starting from $X_{T_1}$, we then proceed by selecting the next inter-jump time $T_2 - T_1$ and post-jump location $X_{T_2}$ using the same procedure, and so on. This gives a piecewise deterministic trajectory $(X_t)$ with jump times $T_1, T_2, \ldots$.

**Remark** (About Markovianity)**.** It is possible to show that $(X_t)$ is a Markov process, in fact strong Markov, i.e. having Markov property holding at all stopping times and not just at fixed times. This means that the distribution of the jump time $T_{k+1}$ only depends on the current state $(\nu_t, \xi_t)$ of the process, and the process begins "anew" at $T_{k+1}$. And, for this reason, we actually talk about Piecewise Deterministic *Markov* processes.

For the time being, the current chapter is enough to handle the essential mathematical tools needed in the upcoming sections: to get further information about the objects at stake, and to deepen all those concepts we do not have time to present here in detail, the reader is invited to refer to the *Bibliography* and to use all the references that can be found throughout this thesis when introducing new notions and materials.

# Chapter 3

# Modelling systems of interacting neurons via Hawkes processes

The aim of this thesis is to study a microscopic model representing several large families of interacting neurons; as seen in the biological introduction, this situation can be modelled using *multi-class systems of nonlinear Hawkes processes*, which -by their nature- are particularly effective in modelling these neuronal patterns. In particular, we are interested in the mean field limits of such processes: in the upcoming sections we arrive, hence, to establish propagation of chaos of the finite system and an associated Central Limit Theorem.
Therefore, first of all we must proceed by giving the setting needed for this kind of model. After that, we deepen the study by considering the evolution of the system as the total number of neurons goes to infinity and in the long-time behaviour context.

## 3.1 General framework for the neuronal model

We consider a large network of interacting neurons, taking $n$ populations of neurons, where each of them contains exactly $N_k$ neurons, for $k = 1, \dots, n$. This means the total number of neurons in the model is $N = N_1 + \cdots + N_n$.

The activity of each neuron in the model is described by a counting process, in particular a Hawkes process, that represents the successive times at which each neuron emits a spike; hence, the whole model is represented by a *multivariate* Hawkes process, as we shall see in detail. More specifically, each neuron is characterized by its own spike train, since we consider as attached to each neuron its counting process

$$Z^N_{k,i}(t) \quad as \;\; k = 1, \dots, n, \;\; i = 1, \dots, N_k, \;\; t \geq 0$$

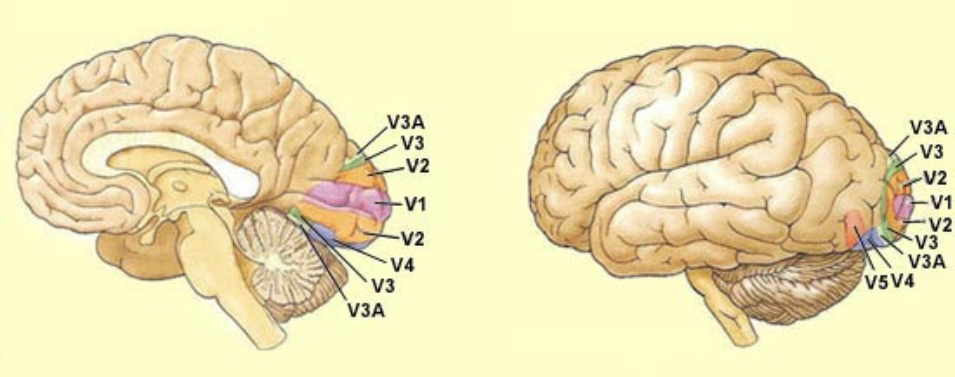



Figure 3.1: Recall that, when talking about "interacting populations" we mean that each population might represent a different functional group of neurons, or that populations can be pools of excitatory and inhibitory neurons in a network. A classical example is the different hierarchical layers in the visual cortex: indeed, researchers have discovered nearly 30 different cortical areas that contribute to visual perception. The primary area (V1) and the secondary area (V2) are surrounded by many other tertiary and associative visual areas: V3, V4, V5 (or MT), PO, etc., and all these distinct areas can be considered as interacting populations of neurons. Image credits: https://thebrain.mcgill.ca/intermediaire.php

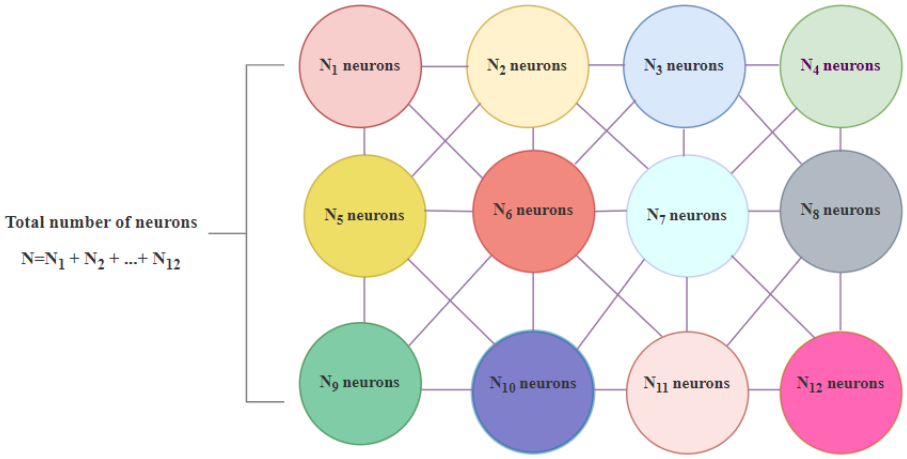



Figure 3.2: *Example* : a system made of $n = 12$ different interacting classes of neurons; each of them contains $N_i$ neurons, so that the total number of neurons of the model is $N = N_1 + \cdots + N_{12}$.

where the index $k$ indicates the population the neuron in object belongs to, $i$ is a label given to the $i-th$ neuron in population number $k$, and considering only positive times means assuming no spiking activity before time 0; therefore, $Z^N_{k,i}(t)$ records the number of spikes in the given interval of time, i.e., before time $t$, emitted by the $i-th$ neuron belonging to population $k$. This means that, for all k and $\forall\ i$ we consider the sequence

$$T_0^{k,i} = 0 < T_1^{k,i} < \cdots < T_n^{k,i} < \dots$$

as the sequence of successive spiking times of the $i$-th neuron belonging to population $k$. Hence, we could equivalently write

$$\begin{aligned} Z^N_{k,i}(t) &= \sharp\ of\ spikes\ of\ neuron\ i\ belonging\ to\ population\ k,\ before\ time\ t \\ &= \sum_{n=1}^{\infty} \mathbb{1}_{\{T_n^{k,i} \le t\}} \end{aligned}$$

As it is known, each counting process $Z^N_{k,i}(t)$ is identified by its intensity process $(\lambda^N_{k,i}(t))$, that depicts the probability of jumping -i.e. emitting a spike- of the process, in a given interval of time; namely:

**Definition 3.1** (*Stochastic intensity*)**.** Let $\mathcal{F}_t$ denote the filtration

$$\mathcal{F}_t = \sigma(Z^N_{k,i}(s), s \le t, 1 \le k \le n, 1 \le i \le N_k)$$

i.e., the internal history of the process $(Z^N_{k,i}(t))_{1\le k\le n, 1\le i\le N_k}$. Any $(\mathcal{F}_t)$-predictable positive process $(\lambda^N_{k,i}(t))_{t\ge 0}$ such that

$$\mathbb{E}[Z^N_{k,i}(t) - Z^N_{k,i}(s) \mid \mathcal{F}_t] = \mathbb{E}[\int_s^t \lambda^N_{k,i}(u)du \mid \mathcal{F}_s] \quad \forall\ 0 \le s \le t$$

is called an $\mathcal{F}_t$-intensity of $Z^N_{k,i}$.
Notice that, generally speaking, if $\lambda^N_{k,i} \equiv \mu$, with $\mu$ constant, then the processes in object are simply Poisson processes with intensity $\mu$.

In our case of study, we can write

$$\mathbb{P}(Z^N_{k,i} \text{ jumps in } ]t, t+dt] \mid \mathcal{F}_t) = \lambda^N_{k,i}(t)dt \quad for\ k = 1, \dots, n,\ i = 1, \dots, N_k$$

This means that $\lambda^N_{k,i}(t)$ is the instantaneous jump rate of neuron $i$ in population $k$, at time $t$.

When explicating the intensities in our neuronal context, they are the classical form of the ones of a multivariate nonlinear Hawkes process:

**Definition 3.2** (*Hawkes intensities*)**.**

$$\lambda_{k,i}^N(t) = f_k\left(\sum_{l=1}^{n} \frac{1}{N_l} \sum_{j=1}^{N_l} \int_{]0,t]} h_{kl}(t-s) dZ_{l,j}^N(s)\right) \tag{3.1}$$

The two functions $f_k$ and $h_{kl}$ appearing the previous formula are essential for the description of the model, indeed:

- $f_k : \mathbb{R} \to \mathbb{R}_+$ is the spiking rate function associated to a given population;
- $\{h_{kl} : \mathbb{R}_+ \to \mathbb{R}\}$, is a family of *synaptic weight functions*, also called *memory kernels*, modelling how population $l$ influences population $k$. More precisely, the leak terms $h.(t-s)$ describe how an event lying back $t-s$ time units in the past influences the present time t, and we assume this influence vanishes at $+\infty$.

It is commonly admitted that spike trains should be processes with infinite memory, and this in fact achieved if the kernels $h_{kl}$ are not compactly supported: in that case the framework is the one of truly infinite memory processes. Moreover, from the form of the intensities in (3.1) it is evident that the setting is the one of self-exciting processes, the firing activity of each neuron being influenced by the past history of the neuron itself and also by the activity of the other neurons in the network.

It is now needed to set two essential properties we have to assume about the quantities at stake.

**Assumption 1.** The functions $f_k$ and $h_{kl}$ need to satisfy the following:

- Assumption for $h_{kl}$: $h_{kl} \in L^2_{loc}(\mathbb{R}_+;\mathbb{R})$, $1 \leq k, l \leq n$.
- Assumptions for $f_k$:
  - $f_k \in C^1(\mathbb{R}, \mathbb{R}_+)\ \forall\ k = 1, \ldots, N$;
  - *Lipschitzianity condition*: there exists a finite constant $L$ such that for all $x, x' \in \mathbb{R}$, for all $1 \leq k \leq n$,

$$|f_k(x) - f_k(x')| \leq L|x - x'| \tag{3.2}$$

Since we work within a *mean field* framework of interacting units, it is important to notice that population $l$ influences population $k$ *only* through its *empirical measure*; roughly speaking, we recall the empirical measure of a set of random variables is a random measure arising from a particular realization of a (usually finite) sequence of those variables, usually computed collecting a certain number of realizations of the variables and calculating the relative frequencies. Formally, the definition is the following:

**Definition 3.3** (*Empirical measure*)**.** Consider $Z_1, \dots, Z_n$ a set of i.i.d. random variables, taking values on a state space which we call $S$ and with probability distribution $P$. Then, the empirical measure $P_n$ is defined for measurable subsets $M$ of the state space $S$ as:

$$P_n(M) := \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i}(M)$$

Hence, specifying this random measure to the generic $k$-th population of neurons, we get that the empirical measures we need are

$$\frac{1}{N_k} \sum_{i=1}^{N_k} \delta_{(Z_{k,i}^N(t))_{t\geq 0}} \quad k = 1, \dots, n$$

The mean field assumption implies, in addition, that all neurons belonging to the same population behave in the same way and are, therefore, exchangeable. More precisely, in the multi-class context we deal with, we consider the following notion of *multi-exchangeability*:

**Definition 3.4** (*Multi-exchangeability*)**.** The sequence of random variables $(Z_{k,i})_{1\leq k\leq n, 1\leq i\leq N_k}$ is said to be *multi-exchangeable* if its law is invariant under any permutation of the indexes within the classes; namely, for any $1 \leq k \leq n$ and any permutation $\sigma_k$ of $\{1, \dots, N_k\}$ the following equality in distribution holds:

$$(Z_{k,\sigma_k(i)}, 1 \leq k \leq n, 1 \leq i \leq N_k) \overset{\mathcal{L}}{=} (Z_{k,i}, 1 \leq k \leq n, 1 \leq i \leq N_k)$$

In this context, we are interested in studying the large population limit, in which the total number of neurons $N \to \infty$; we find out that the evolution within a class is described by a nonlinear *limit* differential equation driven by a Poisson random measure, arriving to state a Central Limit Theorem result. Plus, as we shall see in the subsequent chapters, this study leads us to look for the emergence of self-sustained oscillations in the limit system even in absence of periodic behaviour of the single neuron.
To these aims, we assume that for all $k = 1, \dots, N$, the limit

$$\lim_{N\to\infty} \frac{N_k}{N} = p_k \text{ exists and } \in ]0, 1[$$

## 3.2 Probabilistic setting and Hawkes processes

Given the general setting as described above, we now need to give further information about the proper probabilistic framework we work in.
The model is studied in a filtered probabilistic space

$$(\Omega, \mathcal{A}, \mathbb{F}) = (\mathbb{M}, \mathcal{M}, (\mathcal{M}_t)_{t\geq 0})^I$$

where:

- The probability space is $\mathbb{M} := \{m = (t_n)_{n\in\mathbb{N}}$ such that $t_1 > 0, t_n \leq t_{n+1}, t_n < t_{n+1}$ if $t_n < \infty$, with $\lim_{n\to\infty} t_n = +\infty\}$; this is the canonical path space of a simple point process, representing the "time path" of the process in object as a sequence of successive time steps tending to infinity as the number of neurons in the network goes to infinity.

- Given $m \in \mathbb{M}$, $n \in \mathbb{N}$, set $T_n(m) := t_n$; each element $m \in \mathbb{M}$ is so identified by its associated point measure $\mu := \sum_n \delta_{T_n(m)}$.

- The filtration is $\mathcal{M}_t := \sigma\{\mu(A) : A \in \mathcal{B}(\mathbb{R}), A \subset [0,t]\}$.

- The considered $\sigma$-algebra is simply $\mathcal{M} = \mathcal{M}_\infty$.

- $I = \cup_{k=1}^n \{(k,i), i \geq 1\}$, i.e. the union on the $k$ populations of neurons of the couples (given population, neurons of that population).

Then, we look at the multivariate point measure $(Z^N_{k,i}(t))_{1\leq k\leq n, 1\leq i\leq N_k}$ on the *finite dimensional subspace*

$$(\mathbb{M}, \mathcal{M}, (\mathcal{M}_t)_{t\geq 0})^{I^N} \text{ where } I^N = \cup_{k=1}^n \{(k,i), 1 \leq i \leq N_k\}$$

considering neurons just up to the bound $N_k$, which is the exact number of neurons within each population.

Therefore we can now say that in the neuronal context of our interest an *Hawkes process* with parameters $(f_k, h_{kl})_{k\geq 1, l\leq n}$ is a probability measure $P$ on the filtered probability space $(\mathbb{M}, \mathcal{M}, (\mathcal{M}_t)_{t\geq 0})^I$ such that:

- $\mathbb{P}$-almost surely, for each couple $(k,i)$, $(l,j) \in I$ such that $(k,i) \neq (l,j)$, the processes $Z^N_{k,i}(t)$, $Z^N_{l,j}(t)$ never jump simultaneously; this means that when considering different neurons belonging to the same population, or different neurons in distinct populations, we don't expect to have simultaneous firings.

- Recalling the notion of compensator for a counting process, for all the couples $(k,i)$ in $I^N$, the compensator of $Z^N_{k,i}(t)$ is $\int_0^t \lambda^N_{k,i}(s)ds$, where $\lambda^N_{k,i}(t)$ are the intensity processes in (3.1).

Notice that this definition is consistent with the first definition of Hawkes process the reader can find in the Chapter dedicated to *Mathematical tools*, which is reported from [15].

Given this background, we are sure about the uniqueness of such a process: more precisely, provided Assumption 1, there exists a path-wise unique Hawkes process $(Z^N_{k,i}(t))_{(k,i)\in I^N}$ for all $t \geq 0$. (For a detailed proof of this result see [15], (Theorem 6)).

## 3.3 Towards Propagation of Chaos

Given the finite model, the aim is to study the limit behaviour of the multivariate process $(Z^N_{k,i}(t))_{1\leq k\leq n, 1\leq i\leq N_k}$ modelling the spiking activity of neurons as the total number of neurons itself tends to infinity.

To this purpose, we need to set the framework that leads to build an approximation of the finite system with a system of *inhomogeneous and independent* Poisson processes.

We need to consider the classical notions of chaoticity and propagation of chaos for single-class systems and to focus on their extension to multi-class systems of particles in order to establish a propagation of chaos result for the finite system. Generally speaking, with the term *chaoticity*, it is intended the convergence in law of a starting set of exchangeable random variables to independent and identically distributed random variables; directly connected with this concept, we find the equally important notion of *propagation of chaos*: it is said to hold when the initial chaos, i.e., the initial chaotic distribution, is propagated at later times. In other words, propagation of chaos means that *the stochastic independence of two random particles in a many-particle system persists in time, as the number of particles tends to infinity.* Namely, handling these concepts, we will show that, generally speaking, the initial chaos of the system -intended as the fact that, in the large population limit, within each class the neurons become statistically independent- is propagated at later times. (To deepen this kind of topics related to propagation of chaos and chaoticity, we refer the reader to [10], [22], and [38]).

To this aim we consider the space of càdlàg functions $D(\mathbb{R}_+, \mathbb{R}_+)$ (recall this means they are everywhere right-continuous and have left limits everywhere) endowed with the Skorokhod topology (the interested reader can deepen the choice of this type of topology in place of the more common uniform topology by consulting [32]). Hence, we work on the set of probability measures on this space, i.e. with $(\mathcal{P}(D(\mathbb{R}_+, \mathbb{R}_+))$, equipped with the weak convergence topology associated to the Skorokhod topology on $D(\mathbb{R}_+, \mathbb{R}_+)$.

The notions of *multi-exchangeability* and *multi-chaoticity* play a fundamental role in this study. Therefore, in order to make explicit the definition of *multi-chaoticity* for the finite-system $(Z^N_{k,i}(t))_{(k,i)\in I^N}$, we consider $P_1, \ldots, P_n \in \mathcal{P}(D(\mathbb{R}_+, \mathbb{R}_+))$ and get the following :

**Definition 3.5** (*Multi-chaoticity*)**.** The system $(Z^N_{k,i}(t))_{(k,i)\in I^N}$ is said to be $P_1 \otimes \cdots \otimes P_n$-multi-chaotic if for any $m \geq 1$,

$$\lim_{N\to\infty} \mathcal{L}((Z^N_{k,i}(t))_{(k=1,\ldots,n,\ i=1,\ldots,m)}) = P_1^{\otimes m} \otimes \cdots \otimes P_n^{\otimes m}$$

In this framework of mean field study, the limit process $\bar{Z}_1(t), \ldots, \bar{Z}_n(t)$ associated to the finite system $(Z^N_{k,i}(t))_{(k,i)\in I^N}$ has distribution $P_1, \ldots, P_n$

and models the number of spikes fired by a neuron in the interval $[0,t]$ belonging to population $k$ as $N \to \infty$.

To get the form of the limit system, notice that, re-writing (3.1) in the following way:

$$\lambda_{k,i}^N(t) = f_k\left(\sum_{l=1}^n \int_{]0,t]} h_{kl}(t-s) \underbrace{\left[\frac{1}{N_l}\sum_{j=1}^{N_l} dZ_{l,j}^N(s)\right]}_{\substack{\downarrow\ LLN \\ d\mathbb{E}(\bar{Z}_l(s))}}\right)$$

(where $LLN$ stands for Law of Large numbers), it follows that $\bar{Z}_k(t)$ is a process with intensity:

$$f_k\left(\sum_{l=1}^n \int_0^t h_{kl}(t-s) d\mathbb{E}(\bar{Z}_l(s))\right) dt \tag{3.3}$$

So that the limit system consists of a family of counting processes $\bar{Z}_k(t)$ for $k = 1, \ldots, n$, that are solutions of the inhomogeneous equations given by:

$$\bar{Z}_k(t) = \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\{z \le f_k(\sum_{l=1}^n \int_0^s h_{kl}(s-u) d\mathbb{E}(\bar{Z}_l(u)))\}} N^k(ds, dz),\ 1 \le k \le n \tag{3.4}$$

where the function $f_k$ is the spiking rate of population $k$ in the original system, the kernels $h_{kl}$ model the influence between populations, and the $N^k$ are i.i.d. PRM (Poisson Random measures, see Chapter 2 (Section 2.2)) on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity measure $dsdz$.

It is now natural to associate to the limit process its mean-value as follows:

$$\mathbb{E}(\bar{Z}_1(t), \ldots, \bar{Z}_n(t)) = m_t = (m_t^1, \ldots, m_t^n) \tag{3.5}$$

Having in mind Definition 3.4 and taking expectations, we easily see that the mean values in (3.5) are solutions of the form:

$$m_t^k = \int_0^t f_k\left(\sum_{l=1}^n \int_0^s h_{kl}(s-u) dm_u^l\right) ds \quad k = 1, \ldots, n. \tag{3.6}$$

Having set up the whole context, the question to be investigated concerns how the large time behaviour of the limit system represented by the integrated intensities $(m_t^1, m_t^2 \ldots, m_t^n)$ predicts the large time behaviour of the finite system we started with; in particular, the upcoming chapter will focus on the limit system, studying the emergence of oscillations.

For the time being, equipped with this piece of information, we are going to state the convergence of the process $(Z^N_{k,i}(t))_{(k,i)\in I^N}$ in the large population limit in terms of the empirical measures

$$\frac{1}{N_k}\sum_{i=1}^{N_k}\delta_{(Z^N_{k,i}(t))_{t\geq 0}}, \quad k=1,\dots,n$$

A direct consequence of the main theorem of this section is that a weak law of large numbers holds for the empirical measures, which converge in distribution to the law $P_k$ of the limit system (see Corollary 3.7). As it is known, this convergence result is *equivalent* to the (multi)-chaoticity of the system (for a proof of this result, see Proposition 2.2 in [38], in which the topic is discussed in one-dimension).

Therefore, we are ready to state and prove the main result of this part: it assures the multi-chaoticity of the finite system and, by showing the convergence in law of the finite system to the limit system, it establishes the *propagation of chaos* as foreseen.
This means that, in the large population limit, *neurons belonging to the same class converge to i.i.d. copies of the same limit law*, or, better, that neurons of class $k$ behave according to the law $P_k$ of the correspondent limit system $\bar{Z}_k$.

Moreover, as a consequence of multi-chaoticity and multi-exchangeability, the restriction to each class of neurons is chaotic itself, and any finite subsystem of neurons is asymptotically independent with neurons of class $k$ having the law $P_k$; indeed, there is an asymptotic independence between the different classes, the interactions between classes only surviving in law.

**Theorem 3.6.** *Provided Assumption 1, then:*

1. *There exists a path-wise unique solution to (3.4) such that the function* $t \to \mathbb{E}(\sum_{k=1}^n \bar{Z}_k(t))$ *is locally bounded.*
2. *Multi-chaoticity The system of processes* $(Z^N_{k,i}(t))_{(k,i)\in I^N}$ *is* $P_1\otimes\cdots\otimes P_n$*-multi-chaotic, with* $P_k = \mathcal{L}(\bar{Z}_k)$ *for* $k=1,\dots,n$.
3. *Propagation of chaos: For any* $i \geq 1$,

$$((Z^N_{1,i}(t),\dots,(Z^N_{n,i}(t)))_{t\geq 0}) \xrightarrow[N\to\infty]{\mathcal{L}} ((\bar{Z}_1(t),\dots,\bar{Z}_n(t))_{t\geq 0})$$

*meaning convergence in* $D(\mathbb{R}_+,\mathbb{R}^n_+)$*, endowed with the Skorokhod topology.*

*Proof.* We proceed by splitting the proof into two main steps and retracing what has been done in *Theorem 8* of [15], extending it to a multi-class context.

i. We must prove the path-wise uniqueness of the solution to the equation for the limit system (3.4) which makes the function $t \to \mathbb{E}(\sum_{k=1}^{n} \bar{Z}_k(t))$ locally bounded.

   Therefore, let $(\bar{Z}_1(t), \dots, \bar{Z}_n(t))$ be any solution of (3.4) and set $m_t = (m_t^1, \dots, m_t^n) = \mathbb{E}(\bar{Z}_1(t), \dots, \bar{Z}_n(t))$ as the associated vector of the integrated intensities. By construction, this vector is solution of (3.6).

   At this point, we need to apply the multi-dimensional version of the following auxiliary lemma, which corresponds to Lemma 24 in [15]:

   **Lemma.** *Let $h : \mathbb{R} \to [0, \infty)$ be a Lipschitz-continuous function and let $\phi : [0, \infty) \to \mathbb{R}$ be locally integrable. Then the equation*

$$m_t = \int_0^t h\left(\int_0^s \phi(s-u) dm_u\right) ds$$

   *has a unique non-decreasing and bounded solution; moreover $m$ is of class $C^1[0, \infty)$.*

   Looking at the general notation used in the above lemma, using the spiking rates $f_k$ in place of $h$ and the weights $h_{kl}$ as $\phi$ (noticing we are allowed to do that thanks to Assumption 1), we immediately get that equation (3.6) has a unique, locally bounded, non-decreasing solution (in each coordinate), that moreover is of class $C^1$.

   It is now left to prove uniqueness and well-posedness of the solution: considering $(\bar{Z}_k(t))_{(1 \le k \le n, t \ge 0)}$ a solution of (3.4), its mean value $m_t = (m_t^1, \dots, m_t^n) = \mathbb{E}(\bar{Z}_1(t), \dots, \bar{Z}_n(t))$ is a solution to (3.6) by construction; therefore it is defined uniquely, and this implies that the right hand side in (3.4) is also unique, leading to uniqueness.

   To prove existence, it is sufficient to consider $m$ as the unique solution to (3.6) and to set

$$\bar{Z}_k(t) = \int_0^t \int_{\mathbb{R}_+} \mathbb{1}_{\{z \le f_k(\sum_{l=1}^n \int_0^s h_{kl}(s-u) d\mathbb{E}(\bar{Z}_l(u))\}} N^k(ds, dz), \ 1 \le k \le n$$

   We just need to prove that $m_t = (m_t^1, \dots, m_t^n) = \mathbb{E}(\bar{Z}_1(t), \dots, \bar{Z}_n(t))$. But it is clear that

$$\mathbb{E}(\bar{Z}_k(t)) = \int_0^t f_k\left(\sum_{l=1}^n \int_0^s h_{kl}(s-u) dm_u^l\right) ds \quad k = 1, \dots, n$$

   which is exactly $m_t^k$ since $m$ solves (3.6).

ii. *Propagation of chaos:* first of all we need to consider the i.i.d.PRMs $(N^N_{k,i}(ds,dz))_{(k,i)\in I^N}$ on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity $dsdz$, i.e., the associated measure $\mu$ being the Lebesgue measure, representing $\perp\!\!\!\perp$ noises.
We look at the process $(Z^N_{k,i}(t))_{(k,i)\in I^N, t\geq 0, N\geq 1}$:

$$\begin{aligned} Z^N_{k,i}(t) &= \int_0^t \int_0^\infty \mathbb{1}_{\{z\leq \lambda^N_{k,i}(s)\}} N^N_{k,i}(ds,dz) \\ &= \int_0^t \int_0^\infty \mathbb{1}_{\{z\leq f_k(\sum_{l=1}^n \frac{1}{N_l}\sum_{j=1}^{N_l}\int_0^{s-} h_{kl}(s-u)dZ^N_{l,j}(u))\}} N^N_{k,i}(ds,dz) \end{aligned} \tag{3.7}$$

Thanks to the equivalence of the two definitions of Hawkes processes, as already pointed out when discussing about the *Mathematical tools*, the process defined above is clearly a Hawkes process also according to our notion of Hawkes process in this neuronal context.

At this point, we need to apply a *coupling method*, hence we shall open a brief parenthesis:

**Remark.** Roughly speaking, *coupling* means the joint construction of two or more random elements (in our case, processes), in order to deduce properties of the individual item; this terminology is used in two different ways in standard probability theory: in the general sense of constructing joint versions of random elements (as mentioned) and intending the construction of "intermediate" versions of two stochastic processes in such a way that their paths coincide.
In the context of using underlying PRMs, we can briefly focus on the so-called *synchronous coupling*, which consists of building two stochastic processes using the same underlying PRM, in order to favour common jumps: more specifically, suppose to have a first process $Z$ with intensity $\lambda$ (w.r.to its own filtration), and a second process $\tilde{Z}$ with intensity $\tilde{\lambda}$ (w.r.to its own filtration); then, we build these two processes as thinnings of the same underlying PRM, and see that the uncommon jumps are caused by atoms $(s,z)$ such that either $\lambda(s) < z < \tilde{\lambda}(s)$ or $\tilde{\lambda}(s) < z < \lambda(s)$. This kind of procedure turns out to be useful to prove a convergence result on the paths of these processes (since, indeed, common jumps are favoured), leading to prove propagation of chaos.

Hence, the idea is -having in mind the finite system $Z^N_{k,i}(t))_{(k,i)\in I^N}$- constructing $N$ nonlinear processes by taking the same Poisson random measures as those defining the particle system, and to this aim, we proceed coupling the finite system $Z^N_{k,i}$ to the limit process through the following:

$$\bar{Z}^N_{k,i}(t) = \int_0^t \int_0^\infty \mathbb{1}_{\{z\leq f_k(\sum_{l=1}^n \int_0^s h_{kl}(s-u)dm^l_u)\}} N^N_{k,i}(ds,dz) \tag{3.8}$$

where $m_t$ is the unique solution to (3.6) and $(N^N_{k,i})_{\{1\leq k\leq n,\ 1\leq i\leq N_k\}}$ is the same Poisson random measure used in the dynamics of the process (3.7); notice also that the processes in (3.8), with $k$ running from 1 to $n$ (populations of neurons) and $i$ running between 1 and $N_k$ (neurons within each population) are all independent processes.

We can also point out that, recalling the form of the limit process $\bar{Z}_k$ in (3.4), the convergence in law $\bar{Z}^N_{k,i} \xrightarrow{\mathcal{L}} \bar{Z}_k$ is realized.

The key point in the proof is proving that the processes $(\bar{Z}^N_{k,i}(t))$ and $(Z^N_{k,i}(t))$ are asymptotically close, i.e. that for all $T \geq 0$, $1 \leq k \leq n$, $1 \leq i \leq N_k$

$$\lim_{N\to\infty} \mathbb{E}\left[\sup_{u\in[0,t]} |\bar{Z}^N_{k,i}(u) - Z^N_{k,i}(u)|\right] = 0$$

Therefore, it is convenient to set up the following notation:

- $\Delta^N_{k,i}(t) = \int_0^t |d[\bar{Z}^N_{k,i}(u) - Z^N_{k,i}(u)]|$, i.e., the total variation distance;
- $\delta^N_{k,i}(t) = \mathbb{E}(\Delta^N_{k,i}(t))$, i.e. the expectation of the previously defined quantity. Notice that, by construction, the neurons within each class are exchangeable, so that this quantity doesn't depend on $i$; therefore, we are allowed to write $\delta^N_k(t)$ in place of $\delta^N_{k,i}(t)$.

We can now immediately notice that the following easy inequalities hold:

$$\begin{aligned} \Delta^N_{k,i}(t) &\geq \sup_{u\in[0,t]} |\bar{Z}^N_{k,i}(u) - Z^N_{k,i}(u)| \\ \delta^N_k(t) &\geq \mathbb{E}\left[\sup_{u\in[0,t]} |\bar{Z}^N_{k,i}(u) - Z^N_{k,i}(u)|\right] \end{aligned} \tag{3.9}$$

Once understood what has been done so far, we can start controlling the quantity $\Delta^N_{k,i}(t)$; first of all we write it explicitly using the definitions of the processes $\bar{Z}^N_{k,i}$ and $Z^N_{k,i}$:

$$\begin{aligned} \Delta^N_{k,i}(t) = \int_0^t \int_{\mathbb{R}_+} \Big| &\mathbb{1}_{\{z\leq f_k(\sum_{l=1}^n \int_0^s h_{kl}(s-u)dm^l_u)\}} \\ &- \mathbb{1}_{\{z\leq f_k(\sum_{l=1}^n \frac{1}{N_l}\sum_{j=1}^{N_l} \int_0^{s-} h_{kl}(s-u)dZ^N_{l,j}(u))\}} \Big| N^N_{k,i}(ds,dz) \end{aligned}$$

Then, we compute the expectation $\mathbb{E}$ on both sides, exploiting the lipschitzianity with constant $L$ of the spiking rate functions $f_k$ and the

triangular inequality to get:

$$\frac{\mathbb{E}(\Delta^N_{k,i}(t))}{L} \leq \int_0^t \sum_{l=1}^n \mathbb{E}\left|\frac{1}{N_l}\sum_{j=1}^{N_l}\int_0^{s^-} h_{kl}(s-u)(dm^l_u - d\bar{Z}^N_{l,j}(u))\right|ds$$
$$+ \int_0^t \sum_{l=1}^n \mathbb{E}\left|\frac{1}{N_l}\sum_{j=1}^{N_l}\int_0^{s^-} h_{kl}(s-u)d[\bar{Z}^N_{l,j}(u) - Z^N_{l,j}(u)]\right|ds \tag{3.10}$$

Looking at the right hand side of the previous inequality, we denote with $A$ the integral on the first line, with $B$ the one on the second line; with this notation we can re-write:

$$\frac{1}{L}\mathbb{E}(\Delta^N_{k,i}(t)) \leq A + B$$

where $A$ is a variance term; from now on our aim is controlling $A$ and $B$ separately, going towards the final estimate that allows to conclude with the propagation of chaos result.

Estimate for $B$: it is simple to bound this term -recalling the notation for $\Delta$ and $\delta$- just by using some evident inequalities and the following auxiliary lemma, which corresponds to Lemma 22 in [15], whose proof relies on Fubini's theorem :

**Lemma.** *Given the functions $\phi : [0,\infty) \to \mathbb{R}$ locally integrable and $\alpha : [0,\infty) \to \mathbb{R}$ such that $\alpha(0) = 0$, both with finite variations on compact intervals, the following equalities hold $\forall\ t \geq 0$:*

$$\int_0^t \int_0^{s^-} \phi(s-u)d\alpha(u)ds = \int_0^t \int_0^s \phi(s-u)d\alpha(u)ds = \int_0^t \phi(t-s)\alpha(s)ds$$

Therefore, coming back to the needed estimate for the $B$, let's apply the previous lemma recalling the definition of term $B$, in order to get to the following chain of inequalities:

$$B \leq \int_0^t \mathbb{E}\int_0^{s^-}\left[\sum_{l=1}^n |h_{kl}(s-u)|d\Delta^N_{l,1}(u)\right]ds$$
$$\leq \int_0^t \left[\sum_{l=1}^n |h_{kl}(t-u)|\delta^N_l(u)\right]du$$

Estimate for $A$: to bound this term, we introduce the following random variables:

$$X^N_{k,l,j}(t) = \int_0^{t^-} h_{kl}(t-u)d\bar{Z}^N_{l,j}(u) \quad 1 \leq j \leq N_l \tag{3.11}$$

We can easily notice that these are i.i.d. random variables having mean value $\int_0^t h_{kl}(t-u)dm_u^l$ by definition of $m$.

Let's now substitute in the previous equation (3.11) the definition of $\bar{Z}_{k,i}^N(t)$ as in (3.8), therefore getting:

$$X_{k,l,1}^N(s) = \int_0^{s^-} \int_0^\infty \mathbb{1}_{\{z \le f_l(\sum_{m=1}^n \int_0^u h_{lm}(u-r)dm_r^m)\}} h_{kl}(s-u) N_{l,1}^N(du,dz)$$

where we are allowed to consider $j=1$ thanks to the exchangeability property of neurons within the same population.
At this stage we are allowed, thanks to the fact the integrand is deterministic, to compute the difference $Y_{k,l,1}^N(s) = X_{k,l,1}^N(s) - \mathbb{E}(X_{k,l,1}^N(s))$, which is useful to compute the variance:

$$Y_{k,l,1}^N(s) = \int_0^{s^-} \int_0^\infty \mathbb{1}_{\{z \le f_l(\sum_{m=1}^n \int_0^u h_{lm}(u-r)dm_r^m)\}} h_{kl}(s-u) \tilde{N}_{l,1}^N(du,dz)$$

with $\tilde{N}_{l,1}^N(du,dz)$ simply being the difference $N_{l,1}^N(du,dz) - dsdz$, i.e. the compensated PRM.

Recalling the definition of variance for a random variable, the following equality comes up:

$$\begin{aligned} Var(X_{k,l,1}^N(s)) &= \int_0^s f_l\left(\sum_{m=1}^n \int_0^u h_{lm}(u-r)dm_r^m\right) h_{kl}^2(s-u)du \\ &\overset{(3.6)}{=} \int_0^s h_{kl}^2(s-u)dm_u^l \end{aligned}$$

This turns out to be essential to bound $A$, since by definition of $A$ itself and of $X_{k,l,j}^N(s)$ the following holds:

$$A \le \sum_{l=1}^n \frac{1}{\sqrt{N_l}} \int_0^t \sqrt{Var(X_{k,l,1}^N(s))}ds$$

The path is now clear to give an estimate of the whole sum $A+B$; indeed, remember that our interest is to give a bound to

$$\frac{1}{L}\mathbb{E}(\Delta_{k,i}^N(t)) = \frac{1}{L}\delta_k^N(t)$$

To work more comfortably over the different population of neurons ($k = 1, \dots, n$), we introduce the 1-norm of the noticeable involved quantities, i.e. we set

$$\|\delta^N(t)\|_1 = \sum_{k=1}^n \delta_k^N(t), \quad \|m_t\|_1 = \sum_{k=1}^n m_k^t, \quad \|h(t)\|_1 = \sum_{k,l=1}^n |h_{kl}(t)|$$

Hence, recalling that $\frac{1}{L}\mathbb{E}(\Delta_{k,i}^N(t)) \leq A+B$ and combining the estimates found for $A$ and $B$ respectively, we get:

$$\begin{aligned}\frac{\|\delta^N(t)\|_1}{L} \leq \left(\sum_{k=1}^{n} \frac{1}{\sqrt{N_k}}\right) \int_0^t \sqrt{\left(\int_0^s \|h(s-u)\|_1^2 d\|m_u\|_1\right)} ds \\ + \int_0^t \|h(t-u)\|_1 \|\delta^N(u)\|_1 du\end{aligned} \tag{3.12}$$

Then,

$$\sup_{t\leq T} \|\delta^N(t)\|_1 \leq C_T \left(\sum_{k=1}^{n} \frac{1}{\sqrt{N_k}}\right) \tag{3.13}$$

Which comes from the direct application of the following auxiliary lemma, corresponding to point (i) in Lemma 23 of [15], in its generalization to the multi-class context:

**Lemma.** *Let $\phi : [0,\infty) \to [0,\infty)$ be a locally integrable function and $g : [0,\infty) \to [0,\infty)$ be a locally bounded function. Moreover, consider a locally bounded and non-negative function $u$ such that for all $t \geq 0$:*

$$u_t \leq g_t + \int_0^t \phi(t-s) u_s ds \quad \forall\ t \geq 0$$

*Then,*

$$\sup_{[0,T]} u_t \leq C_T \sup_{[0,T]} g_t$$

*for some constant $C_T$ only depending on $T$ and $\phi$.*

Indeed, starting from (3.12), the function

$$\int_0^t \left(\int_0^s \|h(s-u)\|_1^2 d\|m_u\|_1\right)^{1/2} ds$$

is locally bounded thanks to the fact the weights $h$ are locally square integrable (Assumption 1), and at point *i.* it was proved that the solution $m$ is of class $C^1$.

Therefore from (3.13), passing to the limit for the number of neurons $N \to \infty$, considering any fixed couple $(k,i) \in I^N$ and thanks to the bounds in (3.9), we arrive to the following convergence -more precisely, uniform convergence on compact time intervals- result:

$$\mathbb{E}\left[\sup_{u\in[0,T]} |\bar{Z}_{k,i}(u) - Z_{k,i}^N(u)|\right] \leq C_T \left(\sum_{k=1}^{n} \frac{1}{\sqrt{N_k}}\right) \xrightarrow[N\to\infty]{} 0 \tag{3.14}$$

Given this, to conclude it is sufficient to notice that the topology of uniform convergence on compact time intervals is finer than the Skorokhod topology we are here working with.

Hence, considering the fixed length sequences $l_1, \dots l_k$ for $k = 1, \dots, n$, the following convergence in law (in $\mathcal{D}(\mathbb{R}_+, \mathbb{R}_+^{l_1+\dots+l_n})$) takes place, directly thanks to the bound in (3.14):

$$\begin{gathered}((Z^N_{1,1}(t))_{t\geq 0}, \dots, (Z^N_{1,l_1}(t))_{t\geq 0}, \dots, (Z^N_{n,1}(t))_{t\geq 0}, \dots, (Z^N_{n,l_n}(t))_{t\geq 0}) \\ \downarrow \mathcal{L} \\ (\underbrace{\bar{Z}_1, \dots, \bar{Z}_1}_{l_1 \text{ i.i.d. copies}}, \dots, \dots, \dots, \underbrace{\bar{Z}_n, \dots, \bar{Z}_n}_{l_n \text{ i.i.d. copies}})\end{gathered}$$

Therefore, also propagation of chaos was established.

□

As a direct consequence, and specializing to our neuronal context what has been done in [23], (Corollary (5.2)), we get immediately the following mean field approximation, which is equivalent, as already mentioned, to the chaoticity of the system:

**Corollary 3.7.** *Assume multi-exchangeability and multi-chaoticity of the system and take the hypotheses of Theorem 3.6 as given. Recall the form of the limit system $\bar{Z}_k(t)$ with associated distribution $P_k$. Then, the convergence in distribution of the empirical distributions*

$$\frac{1}{N_k}\sum_{i=1}^{N_k} \delta_{(Z^N_{(k,i)})} \xrightarrow[N\to\infty]{\mathcal{L}} P_k \quad \forall\ 1 \leq k \leq n \tag{3.15}$$

*holds for the weak topology on $\mathcal{P}(D(\mathbb{R}_+, \mathbb{R}_+))$ with $\mathcal{D}(\mathbb{R}_+, \mathbb{R}_+)$ endowed with the Skorokhod topology.*

## 3.4 Associating a Central Limit Theorem

We now go straight towards to study the large time behaviour. Since the solution to the limit system equation (3.4) is in the form of an inhomogeneous Poisson process, provided we have enough information about the integrated intensities in (3.3), its large-time behaviour can be easily described. In this context, we ask whether it is possible to use the large time estimates of the *mean field limit system* to describe the large-time behaviour of the *original Hawkes process* with a large number of particles, but of finite size.

Hence, in this section we get to state and prove a Central Limit Theorem result not only looking at the large population limit as $N \to \infty$, but also considering the convergence of time $t$ to infinity; that is, we study the behaviour of the limit system by looking at the integrated intensities $(m^1_t, \dots, m^n_t)$ and controlling their longtime behaviour.

Since Hawkes processes can be seen as branching processes, as we are going to deepen in a while, we do this by looking at the "transition matrix" $\Lambda$ -which we now introduce- and studying its spectral properties.

Indeed, recalling for a moment the form of a classical uni-dimensional Hawkes process, we know it can generically be represented as a point process whose intensity $\lambda(t)$ is in the form:

$$\lambda(t) = \lambda_0(t) + \sum_{i:t>T_i} \phi(t - T_i)$$

where $\lambda_0(t)$ is a function modelling the initial intensity of the process, and $\phi$ plays the role of the memory kernel, in which it is evident that all events happening before time $t$ contribute to the total intensity of the process (accordingly to the fact Hawkes processes are truly self-exciting processes).

Given this easy notation, and interpreting the classical "arrivals" involved in point processes as the spike trains of each neuron, it is clear that due to the structure of such a process, the neuronal firings in the process of intensity $\phi(t - t_k)$ are "sons" of the spike that is emitted at time $t_k$. Therefore, just by integrating the memory kernels on the positive real line, i.e, computing $\int_0^\infty \phi(t)dt$, we get the "average number of sons" of each spike, and such a number is classically called *branching ratio*. From an intuitive point of view, it represents the number of events appearing in the whole process, or, to use a more modern term stolen from the medical world, the "virality" of the process.

Thus, viewing the successive emissions of a spike as descendants of earlier neuronal firings allows to compare such a Hawkes process with a Galton-Watson type branching process (the reader is invited to refer to [1] to deepen the field of branching processes). We recall that these processes, originally born to study the propagation of family names, are now generalized and simple models to investigate populations of individuals evolving in time, in which each $n-th$ generation gives birth to a random number of individuals, called offspring, in a i.i.d. way.

Having in mind this comparison, it useful to perform an analysis on the number of the descendants: if the branching ratio is $< 1$ -the so called *subcritical case*-, then the number of "descendant" spikes is finite with probability 1; on the other hand, when the ratio is $> 1$ -*supercritical case*- then each firing has a positive probability of having infinitely many descendants.

**Remark.** Even though the rate functions $f_k$ are not linear, we are allowed to use such a procedure of "comparison" with a branching process thanks to the assumption lipschitzianity (with constant $L$) of the rates. Moreover, notice that the fact the rates $f_k$ are not supposed to be linear also implies that the similarity between the classical results in [15] has to be adapted to the case in

object using matrix renewal equations and matrix-convolution equations. A general overview about the so-called *renewal theory* is reported in Chapter 2 (Section 2.3), in order to give the reader the basics to understand the upcoming proofs.

Hence, passing to our multi-class and nonlinear system, we must specify the matrix of the integrated memory kernels $\Lambda := (\Lambda_{ij})$ with $i, j$ running in the set of populations $1, \dots, n$, to the model in object, hence obtaining:

$$\Lambda_{ij} = L \int_0^\infty |h_{ij}(t)|dt \quad 1 \leq i, j \leq n \tag{3.16}$$

In our notations we re-write $\Lambda$ using an auxiliary matrix as follows:

$$\Lambda = \int_0^\infty H(t)dt \quad with \quad H(t) = \Big(L|h_{ik}(t)|\Big)_{1\leq,i,k\leq n} \text{ for any } t \geq 0 \tag{3.17}$$

The study we are about to do concerns the spectral properties of such a matrix $\Lambda$; considering its eigenvalues, and performing a criticality analysis, we look at the subcritical and supercritical case, indicating with $\mu_1$ the largest eigenvalue of $\Lambda$.

We can immediately notice a difference between the context we are working with (spiking rates just Lipschitz) and the more classical results obtained, for example in [15], when assuming the rates to be linear.
Indeed, if the rates were linear, as $N, t \to \infty$, we would not have to assume any additional conditions on $N, t$ in the subcritical case, while in the supercritical we would just need the assumption $e^{\gamma t/N} \to 0$, for some constant $\gamma$. In our model, instead, we are just able to obtain a Central Limit Theorem in the subcritical case in the regime $N/t \to 0$, while the supercritical situation is more difficult, needing to assume $e^{\gamma t} t^{-1} N^{-1/2} \to 0$.

Moreover, as we shall see, the rate of convergence emerging in the Central Limit Theorem is $(m_t^k)^{1/2}$ for $k = 1, \dots, n$, but in our framework we don't dispose of general asymptotical equivalents of $t \to m_t^k$; thus, in this context, to get the final result, it is essential to bound the mean values $m_t^k$ in order to have them of at least linear growth within populations (actually, as we are going to see, we work in a situation in which the growth is linear if considering the subcritical framework, exponential in the supercritical). In other words, this means we assume that within each population there is always some *minimal strictly positive* spiking intensity.

We are almost ready to proceed with the analysis of the two cases as above, needing to find two (one for each criticality region) preliminary bounds on the growth of the integrated intensities $m_t^k$; as we shall see, this is more tricky in the supercritical framework.

### 3.4.1 Preliminary work: Subcritical case

**Subcritical case ($u_1 < 1$)**: consider the synaptic weights $h_{kl} \in L^1(\mathbb{R}_+;\mathbb{R}) \cap L^2(\mathbb{R}_+;\mathbb{R})$ for $1 \leq k, l \leq n$, and take into account Assumption 1. Then,

**Proposition 3.8.** *Given the above hypotheses, there exists a constant $\gamma_0$ such that*

$$m_t^k \leq \gamma_0 t \quad \forall\ k = 1, \dots, n$$

*Moreover, given the $\bar{Z}_{k,i}^N$ as in (3.8), there exists a constant $C$ such that the following inequality holds:*

$$\mathbb{E}\left(\sup_{s\leq t} |Z_{k,i}^N(s) - \bar{Z}_{k,i}^N(s)|\right) \leq \frac{Ct}{\sqrt{N}} \quad \textit{for } 1 \leq k \leq n. \tag{3.18}$$

*Proof.* First of all the aim is finding a linear bound for the growth of the integrated intensities $m_t^k$; looking at their form in (3.6), we get that the intensities $\lambda_t^k$ are:

$$\lambda_t^k := \frac{dm_t^k}{dt} = f_k\left(\int_0^t \sum_{l=1}^n h_{kl}(t-s)dm_s^l\right) \tag{3.19}$$

Just by applying the fact the spiking rate functions are Lipschitz with constant $L$, we get to the following preliminary bound:

$$\lambda_t^k = f_k(0) + \sum_{l=1}^n L\left(\int_0^t |h_{kl}(t-s)|\lambda_s^l\right) \tag{3.20}$$

Now, recall that the matrices $H$ and $\Lambda$ were defined as

$$H(t) = (L|h_{ik}(t)|)_{1\leq,i,k\leq n} \qquad \Lambda = \int_0^\infty H(t)dt$$

So that the bound (3.20) becomes:

$$\lambda_t \leq f(0) + H * \lambda(t)$$

where $f(0)$ has to be intended as an n-dimensional vector, i.e. $(f_1(0), \dots, f_n(t))^T$, and the symbol $*$ denotes the convolution operation.
The main quantity we now take into account is the following:

$$\Gamma_l := \sum_{k=1}^l H^{*k}$$

where $H^{*k}$ is the $k$-fold convolution.
In order to give an explicit formula for its integral, recall the following result,

up to well-posedness of the integrals- for matrix-valued functions convolutions: $\int_0^\infty A * B(t)dt = (\int_0^\infty A(t)dt)(\int_0^\infty B(t)dt)$
Hence, it follows that

$$\int_0^\infty \Gamma_l(t)dt = \sum_{k=1}^{l} \Lambda^k \tag{3.21}$$

Since we are dealing with the subcritical case, i.e. the maximal eigenvalue $\mu_1$ of $\Lambda$ is strictly smaller than 1, it is clear that the unique maximal eigenvalue of (3.21) must be $\sum_{k=1}^{l} \mu_1^k \leq \frac{u_1}{1-\mu_1}$ as it is a geometric series with reason smaller than 1.

Therefore, the function $\Gamma(t) = \sum_{l=1}^{n} H^{*l}(t)$ is locally bounded and well-defined; moreover, we can conclude that the integral function $\int_0^\infty \Gamma(t)dt$ is exactly $\frac{\mu_1}{1-\mu_1}$.

At this point recall that, given a general renewal equation $f = g + f * F$, with $f, g, F$ functions, $g : \mathbb{R} \to \mathbb{R}_+$, the so called "data", measurable and locally bounded and $F : \mathbb{R}+ \in \mathbb{R}_+$ the generalized cumulative distribution function, we know it admits a unique locally bounded solution $f : \mathbb{R}_+ \to \mathbb{R}$ given by $f = g * R$, where $R$ is the renewal function, i.e. $R(t) = \sum_{n\geq 0} F^{*n}(t)$.

Therefore, specifying this fact to our case of study, we notice that any solution $y(t)$ of the equation $y = f(0)+H*y$ is in the form $y(t) = f(0)+\Gamma*f(0) = f(0) + \int_0^t \Gamma(s)f(0)ds$, since, in our case, the renewal function is given by $\Gamma$, and the "data" playing the role of $g$ is $f(0)$.
Hence, the following bound on $\lambda$, to be intended as holding component-wise, comes out:

$$\lambda(t) \leq f(0) + \left(\int_0^t \Gamma(s)ds\right) f(0) \leq f(0) + \left(\int_0^\infty \Gamma(s)ds\right) f(0)$$

This allows to conclude with the first point of the proposition: indeed, since we found out $\lambda(t)$ is a bounded function of $t$, recalling relation (3.19), we get that also the wanted linear bound on the growth of the intensities $m_t^k$ is achieved.

At this stage, we want to prove (3.18) holds; recalling the notation used in the proof of Theorem 3.6 for $\delta_k^N(t)$ and $\Lambda_{k,i}^N(t)$, and recalling the inequalities found in (3.10) and (3.12) of that proof, we get the following chain of

relations, with $\delta^N(t) = (\delta_1^N(t), \dots, \delta_n^N(t))^T$:

$$\begin{aligned}
\delta_k^N(t) &\leq \frac{C}{\sqrt{N}} \int_0^t \sum_{l=1}^n \left[ \int_0^s h_{kl}^2(s-u)\lambda_u^l du \right]^{1/2} + (H * \delta^N)_k(t) \\
&\leq \frac{Ct}{\sqrt{N}} + (H * \delta^N)_k(t) \\
&\overset{(\star)}{\leq} \frac{Ct}{\sqrt{N}} + \frac{C}{\sqrt{N}} \int_0^t \sum_{l=1}^n \Gamma_{kl}(t-s) s ds \\
&\leq \frac{Ct}{\sqrt{N}}
\end{aligned}$$

where, in general, the constant $C$ might change value from line to line and even within the same equation; inequality $(\star)$ comes from the fact we proved the intensities $\lambda_u^l$ are bounded and thanks to Assumption 1, in which we supposed the synaptic weights $h_{kl} \in L^2(\mathbb{R}_+; \mathbb{R})$.
Hence, as we did in the proof of Theorem 3.6, from the bound found in the last line of the previous chain of inequalities we get:

$$\mathbb{E}\left( \sup_{s \leq t} |Z_{k,i}^N(s) - \bar{Z}_{k,i}^N(s)| \right) \leq \frac{Ct}{\sqrt{N}}$$

for $k = 1, \dots, n$. Therefore, the proof of the proposition is done. □

### 3.4.2 Preliminary work: Supercritical case

**Supercritical case ($u_1 > 1$)**: take into account Assumption 1 and consider the functions $h_{kl}$ belonging to $L^1(\mathbb{R}_+; \mathbb{R})$ for $1 \leq k, l \leq n$; moreover assume that there exist $p \geq 1$ and $C$ constants such that $|h_{kl}(t)| \leq C(1 + t^p)$ for all $t \geq 0$. Then,

**Proposition 3.9.** *Given the above hypotheses, there exists a unique constant $\gamma_0$ such that $\int_0^\infty e^{-\gamma_0 t} H(t) dt$ has greatest eigenvalue $\mu = +1$, such that:*

$$m_t^k \leq c e^{\gamma_0 t} \quad \forall\ k = 1, \dots, n$$

*Moreover, there exists a constant $C$ such that the following inequality holds:*

$$\mathbb{E}\left( \sup_{s \leq t} |Z_{k,i}^N(s) - \bar{Z}_{k,i}^N(s)| \right) \leq \frac{C e^{\gamma_0 t}}{\sqrt{N}} \quad \text{for } 1 \leq k \leq n. \tag{3.22}$$

To prove this proposition, first of all we need the following auxiliary result involving classical notions about systems of renewal equations:

**Lemma 3.10.** *Consider the matrix $H(t) = (L|h_{ik}(t)|)_{1 \leq i,k \leq n}$ for $t \geq 0$ as in (3.17) and take as valid the assumptions made for the supercritical case*

*above.*
*Moreover, set the following notation:* $\Gamma(t) := \sum_{k\geq 1} H^{*k}(t)$*, where the notation* $(.)^{*k}$ *stands for the k-fold convolution.*
*Then, the followings hold:*

1. *There exists a unique constant* $\gamma_0$ *such that* $\int_0^\infty e^{-\gamma_0 t} H(t)dt$ *has as largest eigenvalue the value* $+1$*;*

2. *The function* $\Gamma$ *is locally bounded and there exists a constant* $C$ *such that the entries of* $\Gamma$ *are bounded in such a way that:*

$$\Gamma_{ij}(t) \leq Ce^{\gamma_0 t} \tag{3.23}$$

3. *For all pairs of locally bounded functions* $u, h : \mathbb{R}_+ \to \mathbb{R}^n$ *such that* $u = h + H * u$*, it holds the following equality:* $u = h + \Gamma * h$

*Proof.* The proof is essentially based on results on systems of renewal equations that can equivalently be found in Corollary 3.1 and Theorem 3.1 in [13], to which we refer the reader. □

We can now easily prove Proposition 3.9:

*Proof.* It is sufficient to recall that, as in the proof of Proposition 3.8, we can get the bound (3.20) on the intensities $\lambda_t^k$ for $k = 1, \ldots, n$. Therefore, we use the above Lemma 3.10, point *2.* to get the chain of inequalities:

$$\lambda_t^k \leq f_k(0) + \sum_{l=1}^{n} \left( \int_0^t \Gamma_{kl}(s)ds \right) f_l(0) \leq f_k(0) + Ce^{\gamma_0 t} \leq ce^{\gamma_0 t}$$

where it is sufficient to take $c = \max_{1\leq k\leq n} f_k(0) + C$. In this way, the exponential bound to the growth of the intensities, and hence of the integrated intensities $m_t^k$, is obtained.

Given this bound, repeating the very same procedure as in the second part of the proof of Proposition 3.8, we get to inequality (3.22) as wanted. □

### 3.4.3 Main Result: *CLT* for systems of interacting neurons

Thanks to this preliminary work, we get to the main result of this section, which is, as foreseen, a *Central Limit Theorem* that takes into account the convergence to infinity of both the number of neurons $N$ and the time $t$. It shows clearly that the finite size system is close to the system of integrated intensities $(m_t^1, \ldots, m_t^n)$, and that the rate of convergence is, indeed, $(m_t^k)^{1/2}$.

**Theorem 3.11.** *Assume to be either in the situation described in the subcritical case or in the supercritical case. Consider the limits for* $N \to \infty$ *and*

*for $t \to \infty$, with the additional assumptions that $\frac{t}{N} \to 0$ if working in the subcritical situation, $\frac{e^{\gamma_0 t}}{t\sqrt{N}} \to 0$ if in the supercritical.*
*Suppose also that:*

$$\liminf_{t\to\infty} \frac{m_t^k}{t} \geq \gamma_k \quad (\exists\ \gamma_k > 0) \text{ for } k = 1, \dots, n \tag{3.24}$$

*Then,*

1. *For any fixed index $i$, for some constant $C$ and for $1 \leq k \leq n$ it holds the following:*

$$\lim \sup_{N,t\to\infty} \sqrt{m_t^k}\ \mathbb{E}\left[\left|\frac{Z_{k,i}^N(t)}{m_t^k} - 1\right|\right] \leq C$$

   *meaning that the ratio $\frac{Z_{k,i}^N(t)}{m_t^k}$ converges to 1 in probability.*

2. *Central limit result Assume in addition that $\lim_{N\to\infty} \frac{N_k}{N} > 0$ for all population $k = 1, \dots, n$. Then, for all fixed $(l_1, \dots, l_n)$, the following convergence in law holds:*

$$\left(\left(\frac{Z_{1,i}^N(t) - m_t^1}{\sqrt{m_t^1}}\right)_{i=1,\dots,l_1}, \dots, \left(\frac{Z_{n,i}^N(t) - m_t^n}{\sqrt{m_t^n}}\right)_{i=1,\dots,l_n}\right) \xrightarrow[(t,N)\to\infty]{\mathcal{L}} \mathcal{N}$$

   *where $\mathcal{N}$ is a multivariate normal variable with mean vector 0 and covariance matrix $I_{l_1+\dots+l_n}$, i.e $\mathcal{N} \sim \mathcal{N}(0, I_{l_1+\dots+l_n})$*

*Proof.* The proof exploits the two propositions we have already proved, so that we must distinguish between subcritical and supercritical situation, adapting the proof of Theorem 10 in [15] to our nonlinear study:

- **Subcritical case**: first of all we define the following process

$$U_{k,i}^N(t) := Z_{k,i}^N(t) - m_t^k \text{ for } (k,i) \in I^N$$

  Moreover, recalling the form of the process $\bar{Z}_{k,i}^N(t)$ as in (3.8), we introduce the martingales:

$$M_{k,i}^N(t) = \int_0^t \int_0^\infty \mathbb{1}_{\{z \leq f_k(\int_0^s h_{kl}(s-u)dm_l(u))\}} \tilde{N}_{k,i}^N(ds,dz) = \bar{Z}_{k,i}^N(t) - m_t^k$$

  where $\tilde{N}_{k,i}^N = N_{k,i}^N(ds,dz) - dsdz$ is the compensated PRM. Hence, the following relations hold:

$$U_{k,i}^N(t) = M_{k,i}^N(t) + R_{k,i}^N(t), \qquad R_{k,i}^N(t) = Z_{k,i}^N(t) - \bar{Z}_{k,i}^N(t)$$

Now, remember the bound (3.18) found in Proposition 3.8, that using the previous notation results in:

$$\mathbb{E}\left(\sup_{s\leq t}|R^N_{k,i}(s)|\right) \leq \frac{Ct}{\sqrt{N}} \tag{3.25}$$

At this point we need to recall the notion of *quadratic covariation* for two Càdlàg martingales $M, N$, defined by $[M,N]_t = M_tN_t - \int_0^t M_{s-}dN_s - \int_0^t N_{s-}dM_s$. From this, and considering the fact we are assuming that the processes $Z^N_{k,i}$, $Z^N_{l,j}$ never jump simultaneously for $(k,i)\neq(l,j)$, we immediately get that $[M^N_{k,i}, M^N_{l,j}]_t = 0$ if $(k,i)\neq(l,j)$.
Plus, by definition of quadratic covariation, the variation $[M^N_{k,i}, M^N_{k,i}]_t$ equals to $\bar{Z}^N_{k,i}(t)$. Hence, the expectation $\mathbb{E}[(M^N_{k,i}(t))^2] = \mathbb{E}(\bar{Z}^N_{k,i}(t)) = m^k_t$.
Putting everything together:

$$\frac{\mathbb{E}|U^N_{k,1}(t)|}{m^k_t} \leq \frac{\mathbb{E}|M^N_{k,1}|}{m^k_t} + \frac{CtN^{-1/2}}{m^k_t} \leq (m^k_t)^{-1/2} + \frac{Ct}{m^k_t\sqrt{N}} \tag{3.26}$$

At this point, we must bring into play the hypotheses $t/N\to\infty$ and $\liminf_{t\to\infty} m^k_t = \gamma_k > 0$; using the bound achieved above and recalling the definition of $U^N_{k,i}(t)$, we get to:

$$\limsup_{N,t\to\infty}\sqrt{m^k_t}\,\mathbb{E}\left[\left|\frac{Z^N_{k,1}(t)}{m^k_t}-1\right|\right] = \limsup_{N,t\to\infty}\sqrt{m^k_t}\left[\frac{\mathbb{E}|U^N_{k,1}(t)|}{m^k_t}\right] \leq 1$$

And since the first part of the theorem consists, in fact, of proving that the quantity on the left hand side of the previous line is bounded by a constant, the first item in the theorem is done.

It is now time to prove point *2.*, meaning we want to get the convergence in law of the vector

$$\left(\left(\frac{Z^N_{1,i}(t)-m^1_t}{\sqrt{m^1_t}}\right)_{i=1,\dots,l_1},\dots,\left(\frac{Z^N_{n,i}(t)-m^n_t}{\sqrt{m^n_t}}\right)_{i=1,\dots,l_n}\right)$$

to a multivariate normal variable $\mathcal{N}(0, I_{l_1+\dots l_n})$ for any fixed $l_1,\dots l_n$.
To this aim, for all $k$ and considering $i = 1,\dots l_k$ with $l_k \leq N_k$, we can write:

$$\sqrt{m^k_t}\left(\frac{Z^N_{k,i}(t)}{m^k_t}-1\right) = \frac{U^N_{k,i}(t)}{\sqrt{m^k_t}} = \frac{M^N_{k,i}(t)}{\sqrt{m^k_t}} + \frac{R^N_{k,i}(t)}{\sqrt{m^k_t}}$$

At this stage, it is immediate to notice, as in (3.25) that the second term in the last summation is such that

$$\mathbb{E}((m^k_t)^{-1/2}|R^N_k(t)|) \leq (m^k_t)^{-1/2}CtN^{-1/2}$$

which tends to zero as $N, t \to \infty$.
Hence, to give an estimate of $\sqrt{m_t^k}\left(\frac{Z_{k,i}^N(t)}{m_t^k} - 1\right)$, it is sufficient to show that the vector

$$\left(\left(\frac{M_{1,i}^N(t) - m_t^1}{\sqrt{m_t^1}}\right)_{i=1,\dots,l_1}, \dots, \left(\frac{M_{n,i}^N(t) - m_t^n}{\sqrt{m_t^n}}\right)_{i=1,\dots,l_n}\right)$$

tends in law to $\mathcal{N}(0, I_{l_1} + \dots I_{l_n})$ as $t, N \to \infty$. To this end, we can directly use the following Central Limit Theorem for martingales, generalizing it to our multi-class situation (the reader can refer to Lemma 12 of [15] to see the proof of this result):

**Lemma.** *Let $l \geq 1$, $N \geq 1$ be fixed numbers and consider a family $(M_1^N(t), \dots, M_l^N(t))_{t\geq 0}$ of l-dimensional local martingales satisfying the condition $M_i^N(0) = 0$.*
*Assume they have bounded jumps and that $[M_i^N, M_j^N]_t = 0$, $i \neq j$ for all $N \geq 1$ and for $t \geq 0$.*
*Plus, assume it exists a function $(v_t)_{t\geq 0} : [0, \infty) \to [0, \infty)$,continuous and increasing, such that $\forall\ i = 1, \dots, l$ it holds:*

$$\lim_{t,N\to\infty} v_t^{-2}[M_i^N, M_i^N]_t \to 1 \text{ in probability}$$

*Then, the vector $v_t^{-1}(M_1^N(t), \dots, M_n^N(t))$ converges in law, as $t, N \to \infty$, to the Gaussian distribution $\mathcal{N}(0, I_l)$, with $I_l$ being the $l \times l$ identity matrix.*

Since all the hypotheses are verified taking as the equivalent of $v_t$ the functions $(m_t^k)^{-1/2}$, and we have already seen that the covariation $[M_{k,i}^N(t), M_{l,j}^N(t)]_t = 0$ for $(k,i) \neq (l,j)$ we just have to prove that the convergence $\frac{[M_{k,i}^N(t), M_{k,i}^N(t)]_t}{m_t^k} \to 1$ in probability as $t, N \to \infty$ takes place.
But we have know that $[M_{k,i}^N(t), M_{k,i}^N(t)]_t = \bar{Z}_{k,i}^N(t)$: hence, such a convergence is an immediate consequence of the previous point. Thus, directly applying the lemma, the proof for the subcritical case is done.

- **Supercritical case**: given the proof in the subcritical case, it is quite easy to end this step in the supercritical situation; indeed, we use the very same quantities and notations as in the previous case, and recall that in Proposition 3.9 we got the following bound:

$$\mathbb{E}\left(\sup_{s\leq t} |R_{k,i}^N(s)|\right) \leq \frac{Ce^{\gamma_0 t}}{\sqrt{N}} \tag{3.27}$$

Hence, as we did in (3.26) for the subcritical situation, we easily reach the following control:

$$\frac{\mathbb{E}|U_{k,1}^N(t)|}{m_t^k} \leq \frac{1}{m_t^k} + \frac{Ce^{\gamma_0 t}}{m_t^k\sqrt{N}} \tag{3.28}$$

Now, remember we are assuming the constraint $\frac{e^{\gamma_0 t}}{t\sqrt{N}} \to 0$ as $N, t \to \infty$. Then, it is immediate to prove the first part of the theorem, just by applying a lim sup:

$$\limsup_{N,t\to\infty} \sqrt{m_t^k}\ \mathbb{E}\left[\left|\frac{Z_{k,1}^N(t)}{m_t^k} - 1\right|\right] = \limsup_{N,t\to\infty} \sqrt{m_t^k}\ \left[\frac{\mathbb{E}|U_{k,1}^N(t)|}{m_t^k}\right] \leq C$$

, for some constant $C$ and for $k = 1, \ldots, n$. And, with this, point *1.* of the theorem is done.

To prove *2.*, it is sufficient to repeat the same procedure used in the proof of the subcritical case.

In this way, the theorem is completely proved.

□

# Chapter 4

# Emerging oscillations in the Monotone Cyclic Feedback system

In the previous chapter, the Central Limit Theorem revealed that the finite size system is close to the system of integrated intensities within each population, i.e $(m_t^1, \dots, m_t^n)$, and showed that the rate of convergence is of the order of $(m_t^k)^{1/2}$.

At this point, we are ready to talk about the emergence oscillations, which is the main landing of this work. The setting is the same as in the previous section, but henceforth we need further assumptions, in particular on the form of the memory kernels.

As mentioned in the introduction, we know that many real systems composed of interacting units, such as the neuronal network model in object, may exhibit collective periodic behaviour even if the single components have no natural tendency to behave periodically; it is known that biological and internal rhythms are ubiquitous in living beings, and one of the main questions concerns how do these rhythms arise and are controlled by the brain. The whole situation turns out in the emergence of macroscopic oscillations, which are one of the most common self-organizing behaviours observed in living systems.

In this chapter, knowing that such oscillations are observed in the *way* information is processed in the brain, we present conditions under which the *limit process* presents solutions periodic in law. This is done characterizing the interaction between classes accordingly to a *monotone cyclic feedback* system: using an approach which is typical of the study of dynamical systems, the main theorem of the section shows situations in which the system of the integrated intensities previously introduced ($m_t^k$ for $k$ running between populations, $k = 1, \dots, n$), possesses *non-constant, attractive, periodic orbits*, i.e., it exhibits an *oscillatory* behaviour.

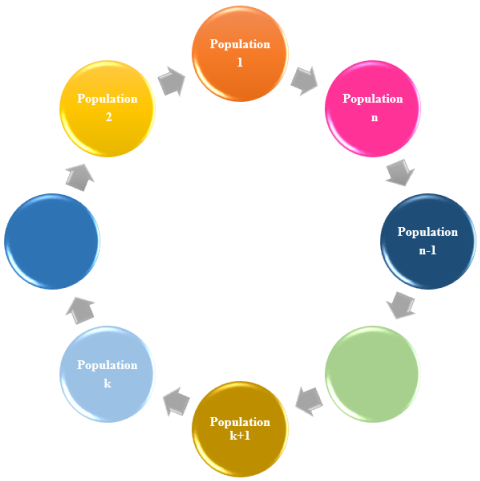



Figure 4.1: *How the cyclic system works*: population $k+1$ is in some sense "connected" just with population $k$, and it influences population $k$ according to the form of the memory kernels $h_{kk+1}$.

## 4.1 General setting for the Monotone Cyclic Feedback System

To investigate situations in which oscillations related to the limit system $(\bar{Z}_k(t))_{(1\leq k\leq n)}$ and $(m_t^k)_{(1\leq k\leq n)}$ (in the forms (3.4),(3.6)) occur, we describe the way information is transported through the system according to a *Monotone Cyclic Feedback* (from now on, *MFC*) system (similarly to what has been done in [3] and [34]), dealing with a special type of memory kernels $h_{kl}$, given in the form of Erlang kernels.
First of all we need to explain what does *MFC* mean:

- **Monotone**: we suppose that the rate functions $f_k$ for $k = 1, \dots, n$ are non-decreasing.
- **Cyclic**: meaning that population $k$ is only influenced by population $k+1 \; \forall \; k$; this leads directly to the fact $h_{kl} = 0 \; \forall \; k, l$ such that $l \neq k+1$.
- **Feedback**: in the sense that population $n$ is only influenced by population 1 (i.e. we identify $n+1$ with 1).

In this context, it is evident that, due to the cyclicity assumption, the intensity of the process associated to the $i-th$ neuron belonging to population $k$ is given by the following expression, which comes from an adaptation of (3.1) to this new context:

$$\lambda_{k,i}^N(t) = f_k\left(\frac{1}{N_{k+1}}\sum_{j=1}^{N_{k+1}}\int_0^t h_{kk+1}(t-s)dZ_{k+1,j}^N(s)\right) \tag{4.1}$$

Having in mind this background, we can pass to the examination of the role of the memory kernels $h_{kk+1}$, which, we recall, describe how population $k+1$ influences population $k$. We consider a special form of these synaptic weights: the functions $h_{kk+1}$ will, from now on, be *Erlang kernels*, i.e. given in the form of an Erlang distribution, as follows:

$$h_{kk+1}(s) = c_k e^{-\nu_k} \frac{s^{\eta_k}}{\eta_k!} \qquad h_{n1}(s) = c_n e^{-\nu_n} \frac{s^{\eta_n}}{\eta_n!} \tag{4.2}$$

where the parameters $c_k \in \{+1, -1\}$ describe the influence between populations, taking the value $+1$ if the influence of population $k+1$ on population $k$ is excitatory, $-1$ if inhibitory, and $\eta_k \in \mathbb{N}_0, \nu_k > 0$ are fixed constants, described in the picture below: $\eta_k + 1$ is the *order of delay*, while the parameter $\nu_k$ is also called the *rate*.
In addition, we set the quantity $\delta = \prod_{k=1}^{n} c_k$, and distinguish between two cases: if $\delta > 0$, we say the system is of total *positive feedback*, if $\delta < 0$ of *negative feedback*; this is exactly the case we take into consideration from now on along this discussion.

### 4.1.1 Focus on the Erlang Kernels

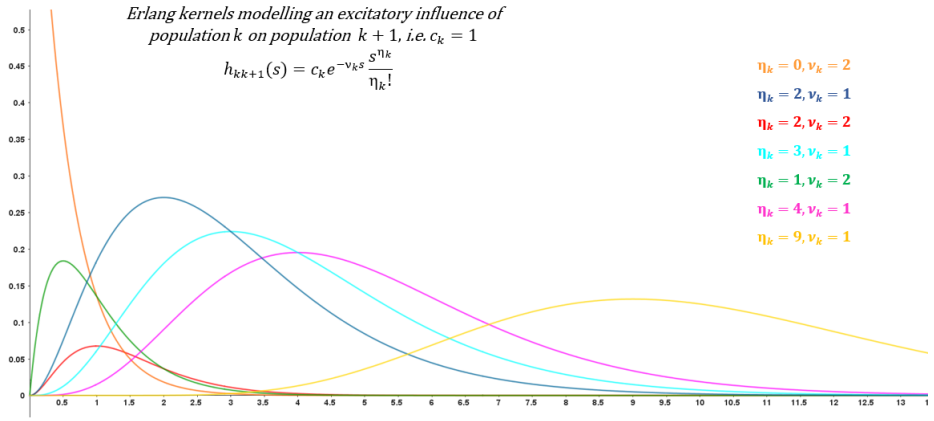



Figure 4.2: **Examples of Erlang kernels** as the parameters $\eta_k, \nu_k$ change value, modelling the *excitatory* ($c_k = 1$) influence of population $k$ on population $k+1$. The graphs describing, instead, an inhibitory influence between populations would be symmetrical with respect to the horizontal axis.

**Remark 1** (*Erlang distribution*)**.** We recall some basics about the Erlang distribution, in order to better understand the choice of such a type of kernels: originally born to model the number of calls which might be made at

the same time to the operators of a switching station, it is a two-parameters family of continuous probability distributions with support $[0, \infty)$, whose parameters are $k, \lambda$, respectively called the "shape" and the "rate" of the distribution. It represents the sum of $k$ independent exponential variables, or, equivalently and more compellingly to our discussion, it is the distribution of the time until the $k-th$ event of a Poisson process with rate $\lambda$ takes place. The density function is:

$$f(x; k, \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$$

with mean value and mode that are respectively $k/\lambda$ and $(k-1)/\lambda$. Within this framework, the parameters that come into play in the description of the memory kernels, i.e $\eta_k, \nu_k$, correspond to, respectively, parameters $k-1$ and $\lambda$ in the classical Erlang distribution.

Hence, considering the generalities about the Erlang distribution, the delay of influence of population $k+1$ on population $k$ takes its maximum absolute value at $\eta_k/\nu_k$ time units back in time, while the mean, if we consider the normalization to a probability density and according to the notation associated to the Erlang distribution, is $(\eta_k + 1)/\nu_k$.

**Remark 2** (*About delay*)**.** One could ask: *what about the delay? Why do we need to introduce it in such a distributed form?* To answer to these questions, it is useful to open a brief parenthesis about the usual choices of delays while modelling systems like the one we are considering in our discussion.

We recall that, usually, e.g. in population dynamics, a lot of situations are often modelled by introducing a *discrete* time delay into the equations; this means obtaining a (DDE) -Delay Differential Equation-, which has the property that the time derivative at the current time depends on the solution and maybe on the derivatives at a previous time. Hence, there is a dependence on the past history, and for this reason this type of equations are quite common in many models, since they allow to build more realistic scenarios than the ones independent of past history.
When considering a discrete delay, it is implicitly assumed that the delays are distributed over the population by a $\delta$-Dirac distribution, and the resulting (DDEs) are often used to study the properties of the model, like stability of equilibria or existence of stability switches, as function of the discrete delay taken as parameter.

Hence, why should we choose a *continuous* or *distributed* delay, instead of a discrete one? That's because, sometimes, introducing a discrete delay is a vague approximation of a real situation of delay over a large size population; in these contexts it is more reasonable to choose a continuously distributed delay: the key point is that the mean delay, in this case, is equal to the

discrete delay, and the model is built in such a way that the variance is positive, to keep in mind the delay difference among individuals, as we are doing in this thesis when choosing the Erlang distribution. Moreover, notice that the higher the order of the delay, the more concentrated is the delay around its mean value: this means that, in our context, as $\eta_k \to \infty$ (keeping fixed the mean value $(\eta_k+1)/\nu_k$), the distributed delay converges to a discrete delay: thus, we recover the discrete delay model.

## 4.2 The Monotone Cyclic Feedback system

At this stage, we are ready to start properly the discussion that leads to the discovery of the emergence of internal oscillations in the system. First of all we need to introduce the following memory variables, for $k = 1, \dots, n$:

$$x_t^k = \int_0^t h_{kk+1}(t-s) dm_s^{k+1} \tag{4.3}$$

Hence, recalling the form (3.6) of the variables $(m_t^k)_{1\leq k\leq n}$, it is clear that within this new framework, they become:

$$m_t^k = \int_0^t f_k(x_s^k) ds \tag{4.4}$$

It is now essential to understand the reason below the choice of the kernels in the Erlang form: by choosing the synaptic weights in this way, we allow the developing of the system of memory variables introduced above into a high dimensional (precisely, of dimension $\tau$, which we describe hereafter) system (see, below, *MFC*) of differential equations without delay, which we call a *Markovian cascade* of successive memory terms. Moreover, this choice allows - as we shall see in the upcoming chapter- a Markovian description of the Hawkes processes in object, helping us to study much more easily how the behaviour of the limit system we are now dealing with is observed within the finite size system $Z_{k,i}^N$.

Starting by differentiating the kernels with respect to time, we get the following expression:

$$h'_{kk+1}(t) = -\nu_k h_{kk+1}(t) + c_k \frac{t^{\eta_k - 1}}{(\eta_k - 1)!} e^{-\nu_k t}$$

And the identification $x_t^k = x_t^{k,0}$ brings us to the natural definition of the auxiliary variables

$$x_t^{k,l} = \int_0^t c_k e^{-\nu_k (t-s)} \frac{(t-s)^{\eta_k - l}}{(\eta_k - l)!} dm_s^{k+1} \tag{4.5}$$

for $k = 1, \dots, n$ and for $l = 0, \dots, \eta_k$. Therefore, we obtain a dynamics in which a delay kernel acts on the interactions between populations: at any time $t$ the influence of population $k+1$ depends on the trajectory up to time $t$, and it is weighted through the delay kernel.
At this point, with the aim of obtaining a differential equation, we differentiate the variables in (4.5) with respect to time, getting:

$$\frac{dx_t^{k,l}}{dt} = -\nu_k x_t^{k,l} + x_t^{k,l+1}, \quad l < \eta_k \tag{4.6}$$

We are now ready to present the proper *MFC* system as foresaw; it is sufficient to iterate the previous differentiation argument to obtain the following system of coupled equations:

$$\begin{cases} \frac{dx_t^{k,l}}{dt} = -\nu_k x_t^{k,l} + x_t^{k,l+1}, \quad 0 \le l < \eta_k \\ \frac{dx_t^{k,\eta_k}}{dt} = -\nu_k x_t^{k,\eta_k} + c_k f_{k+1}(x_t^{k+1,0}), \quad l = \eta_k \\ x_0^{k,l} = 0 \quad \text{(initial conditions)} \end{cases} \tag{MFC}$$

where it is useful to notice that the second equation comes from the fact the $m_t^k$ are in the form (4.4), and to recall the identification $x^k = x^{k,0}$.

Hence, we have found a *monotone cyclic negative feedback* system that drives the way information is transmitted into the network; its dimension, or better the total order of delay, is clearly given by

$$\tau := n + \sum_{k=1}^{n} \eta_k$$

since $x = (x^{1,0}, \dots, x^{1,\eta_1}, x^{2,0} \dots, x^{n,\eta_n})^T$; hence, it is the sum of the total number of populations in the finite size regime and of the memory length, i.e., related to the delay- $\sum_{k=1}^{n} \eta_k$.

From now on, the framework will be the one of the negative feedback, $\delta < 0$; since all the needed notation is set, we can start investigating the stability properties of such a system, looking for equilibria and periodicity.

## 4.3 Existence of a unique equilibrium

We are ready to analyze the whole situation pointing towards the discovery of oscillations: we need to use the typical tools proper of dynamical systems theory, intended as systems whose dynamics is driven by a differential equation involving time derivatives, usually in the typical $\dot{x} = F(x)$; for us, the corresponding differential equations controlling the neuronal dynamics are the ones found in *MFC*. This means we must talk about spectral stability and equilibrium points, flows and invariant surfaces, Lyapunov functions and

so on and so forth, calling into play some relevant and well-known results in dynamical systems theory. With the aim of understanding better this kind of results, the reader can refer to [2], which gives an overview about general theory on dynamical systems. For the time being, we hereafter report a "glossary" containing the essentials we will need throughout this chapter:

**Glossary** (*Basics on Dynamical Systems Theory*)**.** As mentioned, a *Dynamical system* is generally defined as a particle or ensemble of particles whose state varies over time, obeying some differential equations involving time derivatives (in the typical form $\dot{x} = F(x)$, with $F$ a vector field having a certain regularity) describing how the state variables change as a function of the current states and the system parameters. In other words, it consists of a system whose state evolves with time over a state space according to a fixed rule.

The basic notation and nomenclature in dynamical systems theory which is going to emerge in the subsequent sections is the following:

- *Phase space*: the phase space is the geometric space spanned by all of the system's equations. If the system has $N$ state variables, then its phase space has dimension $N$.

- *Flow*: the flow of a dynamical system corresponds to the vectors for all points in phase space. More specifically, given a $C^1$ system $\dot{x} = F(x)$, the associated flow is the function $\phi_t(x_0)$, which is the solution to the initial value problem
$$\begin{cases} \dot{x} = F(x) \\ x(0) = x_0 \end{cases}$$
clearly depending on the initial position $x_0$.

- *Orbits:* an orbit is a curve in the phase space that follows the system's flow, being a geometric representation of solution of the system under a particular set of initial conditions, as the system evolves. As a phase space trajectory is uniquely determined for any given set of coordinates, different orbits can't intersect in the phase space, hence constituting a partition of the phase space.
A particular type of orbits corresponds to the so-called *periodic* orbits, i.e., to special types of solutions which repeat themselves in time. If a nearby orbit indefinitely stays close to a given orbit, then it is said to be *stable*; moreover, if the nearby orbit converges to the given orbit, then it is called *asymptotically stable* and the given orbit is said to be *attracting*.

- *Fixed points*: an equilibrium (or fixed point) $x^*$ satisfies the condition $F(x^*) = 0$, i.e., it is a stationary (constant in time) solution of the

system's equations. When talking about fixed points, it is of interest characterizing their stability properties: intuitively, if considering a system in a stable equilibrium state, then a small perturbation results in a localized motion, e.g. in small oscillations; in a system with damping, a stable equilibrium state is moreover asymptotically stable. Instead, if looking at unstable equilibria, a small disturbance results in a motion with a large amplitude that may not converge to the original state, i.e., the perturbation grows in time. Stability analysis can be performed, for example, via Lyapunov arguments or spectral analysis of the Jacobian matrix of $F$.

- *Spectral stability of equilibria*: the stability of equilibria (which we usually indicate with $x^*$) of smooth ODEs can be determined by the sign of the real parts of the eigenvalues of the Jacobian matrix computed at the equilibrium point $x^*$, i.e. of the matrix $DF(x^*)$. The equilibrium is said to be *asymptotically stable* if all eigenvalues have negative real parts, while it is *unstable* if at least one eigenvalue has positive real part.

- *Lyapunov function*: when speaking about Lyapunov theory, we consider Lyapunov functions as particular class of scalar functions that can be used to prove the stability of an equilibrium; indeed they provide a fundamental tool to establish Lyapunov stability or asymptotic stability. For certain types of ODEs, the existence of Lyapunov functions is a necessary and sufficient condition for stability, and in many specific cases the construction of Lyapunov functions is known.
  Considering the dynamical system $\dot{x} = F(x)$, with $F : \mathbb{R}^n \to \mathbb{R}^n$, and supposing it has an equilibrium point at $x^* = 0$, we say $V : \mathbb{R}^n \to \mathbb{R}$ is a Lyapunov function for the dynamical system in object if it is continuous, of class $C^1$, such that $V(x) > 0 \ \forall \ x \neq 0$, $V(x^*) = 0$ and $\nabla V \cdot F \leq 0$. Via studying the function $V$ it is possible to characterize different types of equilibria and to get to their stability features (*Lyapunov stability theory*).

- *Limit sets*: we denote with $\omega(x)$ the so-called $\omega$-limit set of $\phi_t(x)$, i.e., the set of points $y \in \mathbb{R}^n$ satisfying the property that it exists a strictly increasing sequence of times $(t_n)_{n\geq 0}$ such that $\lim_{n\to\infty} t_n = \infty$ and $\lim_{n\to\infty} \phi_{t_n} = y$; namely, it is the set containing all those points to which the flow starting from $x$ converges. Similarly, we define $\alpha(x)$, the $\alpha$-limit set of $\phi_t(x)$, as the set of points $y \in \mathbb{R}^n$ such that there exists a strictly decreasing sequence of times $(t_n)_{n\geq 0}$ with the properties $\lim_{n\to\infty} t_n = -\infty$ and $\lim_{n\to\infty} \phi_{t_n}(x) = y$.

- *Invariant sets for the flow*: a set $A$ is said to be an invariant set if it holds that any trajectory entering in $A$, or starting in $A$, keeps staying

in the set $A \; \forall \; t$. In terms of the flow, $A$ is an invariant set if $\forall \; x \in A$ $\phi_t(x) \in A \; \forall \; t \in \mathbb{R}$. Typical examples of invariant sets are the $\omega$-limit set and the $\alpha$-limit set.

- *Attractors and basins of attraction*: an attractor is simply a set of states toward which a system tends to evolve for a wide range of starting conditions, meaning that system values that get close enough to the attractor remain close to it even if subjected to some disturbance. We say an attractor is structurally stable when a small change in the system parameters only leads to a small change in its morphology; otherwise it is said to undergo a bifurcation.
  Each attractor is surrounded by its own basin of attraction, i.e., all of the points in the phase space that flow onto that specific attractor.

- *Repellors*: they are the contrary of the concept of attractor; indeed, if a set of points is periodic or chaotic, but the flow in the neighborhood is away from the set itself, the set is not an attractor, but it is instead called a repellor.

Given the above basics about dynamical systems, we can properly start our discussion with the following intermediate result: it assures the existence -in our hypotheses- of a *unique* equilibrium $x^*$ for the *MFC* system, which is going to turn out to be unstable.

**Proposition 4.1.** *The MFC system admits a unique equilibrium $x^*$ when working in the negative feedback situation and assuming the hypothesis of monotonicity of the rates is true.*

*Proof.* The aim of the proof is to characterize *uniquely* the existence of an equilibrium $x^*$; knowing that the dimension of the system is $\tau = n + \sum_{k=1}^{n} \eta_k$, we must find the $\tau$ values for the $\tau$ coordinates of $x^*$.

Along the proof, we show how to find the coordinate $(x^*)^{n,\eta_n}$, exploiting a recursive relation and the definition of equilibrium, i.e. a constant -in time- solution to the differential equations system in *MFC*; given the procedure to find such a coordinate, all the other coordinates can be found in a similar way.

Therefore, having in mind the *MFC* system, we must recall that an equilibrium $((x^*)^{k,l})_{(1 \leq k \leq n, \; 0 \leq l \leq \eta_k)}$, must satisfy, for all $k = 1, \dots, n, l = 1, \dots, \eta_k$

$$\frac{d(x^*)^{k,l}}{dt} = 0$$

Hence, considering $k = n$, $l = \eta_n$, if $(x^*)^{n,\eta_n}$ has to be an equilibrium, then it must hold that:

$$\begin{aligned} \frac{d(x^*)^{n,\eta_n}}{dt} = 0 \overset{MFC}{\iff} \nu_n (x^*)^{n,\eta_n} &= c_n f_1 (x^*)^{1,0} \\ \iff (x^*)^{n,\eta_n} &= \frac{c_n}{\nu_n} f_1((x^*)^{1,0}) \end{aligned} \tag{4.7}$$

At this point, we must then consider the first equation in *MFC* to compute explicitly $(x^*)^{1,0}$:

$$\frac{d(x^*)^{1,0}}{dt} = 0 \overset{MFC}{\iff} \nu_1(x^*)^{1,0} = (x^*)^{1,1} \iff (x^*)^{1,0} = \frac{(x^*)^{1,1}}{\nu_1}$$

Inserting this expression for the coordinate $(x^*)^{1,0}$ into the one in (4.7) for $(x^*)^{n,\eta_n}$, we get:

$$(x^*)^{n,\eta_n} = \frac{c_n}{\nu_n} f_1\big((x^*)^{1,0}\big) = \frac{c_n}{\nu_n} f_1\left(\frac{(x^*)^{1,1}}{\nu_1}\right)$$

Now, the mechanism is quite clear: from the first equation in *MFC* we get the expression for $(x^*)^{1,1} = (x^*)^{1,2}/\nu_1$, so that:

$$(x^*)^{n,\eta_n} = \frac{c_n}{\nu_n} f_1\big((x^*)^{1,0}\big) = \frac{c_n}{\nu_n} f_1\left(\frac{(x^*)^{1,1}}{\nu_1}\right) = \frac{c_n}{\nu_n} f_1\left(\frac{(x^*)^{1,2}}{\nu_1^2}\right)$$

We iterate this procedure until we reach the values $k = 1$, $l = \eta_1$, i.e.

$$(x^*)^{n,\eta_n} = \frac{c_n}{\nu_n} f_1\left(\frac{(x^*)^{1,\eta_1}}{\nu_1^{\eta_1}}\right)$$

And then, in order to find the explicit expression for $(x^*)^{1,\eta_1}$, we consider again the second equation in the *MFC* system; taking into account stationarity, we obtain:

$$\frac{d(x^*)^{1,\eta_1}}{dt} = 0 \overset{MFC}{\iff} \nu_1(x^*)^{1,\eta_1} = c_1 f_2\big((x^*)^{2,0}\big) \iff (x^*)^{1,\eta_1} = \frac{c_1}{\nu_1} f_2\big((x^*)^{2,0}\big)$$

At this point it seems evident how to continue: we have to iterate this procedure, combining the two equations in *MFC* system and the stationarity condition for equilibria, until we get to:

$$(x^*)^{n,\eta_n} = \frac{c_n}{\nu_n} f_1 \circ \frac{c_1}{\nu_1^{\eta_1+1}} f_2 \circ \frac{c_2}{\nu_2^{\eta_2+1}} f_3 \circ \cdots \circ \frac{c_{n-1}}{\nu_{n-1}^{\eta_{n-1}+1}} f_n\big((x^*)^{n,0}\big) \tag{4.8}$$

To conclude, it is sufficient to notice that $(x^*)^{n,0}$ is a solution of the following:

$$\nu_n(x^*)^{n,0} = (x^*)^{n,1} \iff (x^*)^{n,0} = \frac{(x^*)^{n,1}}{\nu_n} = \frac{(x^*)^{n,2}}{\nu_n^2} = \cdots = \frac{(x^*)^{n,\eta_n}}{\nu_n^n}$$

And so, the entry $(x^*)^{n,\eta_n}$ must satisfy the relation:

$$(x^*)^{n,\eta_n} = \frac{c_n}{\nu_n} f_1 \circ \frac{c_1}{\nu_1^{\eta_1+1}} f_2 \circ \frac{c_2}{\nu_2^{\eta_2+1}} f_3 \circ \cdots \circ \frac{c_{n-1}}{\nu_{n-1}^{\eta_{n-1}+1}} f_n\left(\frac{(x^*)^{n,\eta_n}}{\nu_n^{\eta_n}}\right)$$

Notice that we are sure about the fact it exists exactly only one solution $(x^*)^{n,\eta_n}$ in $\mathbb{R}$ because $\frac{c_n}{\nu_n} f_1 \circ \frac{c_1}{\nu_1^{\eta_1+1}} f_2 \circ \cdots \circ \frac{c_{n-1}}{\nu_{n-1}^{\eta_{n-1}+1}} f_n(\frac{1}{\nu_n^{\eta_n}})$ is a decreasing

function: this follows from the fact we are considering the negative feedback case ($\delta = c_1 \cdot \cdots \cdot c_n < 0$) and because of the assumption of monotonicity of the rates.

Moreover, as mentioned, the values for the other components of the equilibrium $x^*$ can be determined -in a unique way- using a similar procedure; having this remark done, the proof is concluded. □

## 4.4 Emergence of periodic behaviour: main theorem

We are on the right way to state and prove the main result of this section; in order to proceed, we need the following essential assumption, containing the hypotheses of the previous proposition; hence, working in such a context, the existence of the unique equilibrium $x^*$ is assured.

**Assumption 2.** Assume the followings are true:

- The spiking rates $f_k$, for $k = 1, \ldots, n$, are non-decreasing, bounded and analytic functions.
- The quantity $\rho := \prod_{k=1}^{n} c_k f_k'((x^*)^{k,0})$ is $< 0$.

At this point, recalling the roles of the parameters $\eta_k$ and $\nu_k$ in the Erlang kernels, we arrive to the main theorem of this section:

**Theorem 4.2.** *Accepting as valid Assumption 2, considering the solutions $\mu$ to the equation*

$$\prod_{i=1}^{n} (\nu_i + \mu)^{\eta_i + 1} = \rho \tag{4.9}$$

*and supposing that there exist at least two solutions $\mu$ to the previous one such that*

$$Re(\mu) > 0 \tag{4.10}$$

*then:*

1. *The equilibrium $x^*$ is linearly unstable; moreover the MFC system possesses at least one -and not more than a finite number- of periodic orbits. Within them, at least one is asymptotically stable.*
2. *If $\tau$ is equal to 3, then it exists a globally attracting invariant surface $\Sigma$ such that the unstable equilibrium $x^*$ is a repellor for the flow on the surface $\Sigma$ itself.*
   *Plus, every solution of the MFC system will be attracted to a non constant periodic orbit.*

*Proof.* First of all, we immediately notice that, thanks to Assumption 2, which assures the rates $f_k$ to be bounded, we are sure about the existence of an invariant and compact subset $K$ for the system $MFC$.
Then, we can rewrite $MFC$ in the typical form of a dynamical system, i.e. as $\dot{x} = F(x)$, omitting the subscript $t$ to indicate the variables $(x_t^{k,l})$. Recall that we work in a vectorial setting of dimension $\tau = n + \sum_{k=1}^{n} \eta_k$, where $x = (x^{1,0}, \dots, x^{1,\eta_1}, x^{2,0} \dots, x^{n,\eta_n})^T$.

In order to study the stability of the system, we need the spectral analysis of the Jacobian matrix of the linearization around the equilibrium $x^*$, whose coordinates have already been explicitly computed in the proof of Proposition 4.1. Therefore, we need the Jacobian matrix $DF(x^*)$ and, then, its characteristic polynomial $P(\mu)$.
First of all, we look at $DF(x^*)$: recalling the equations in $MFC$, we have that:

$$\dot{x} = (-\nu_1 x^{1,0} + x^{1,1}, \dots, -\nu_1 x^{1,\eta_1} + c_1 f_2(x^{2,0}), \dots, -\nu_n x^{n,\eta_n} + c_n f_1(x^{1,0}))^T$$

Therefore, the Jacobian matrix $DF(x)$ is a $\tau \times \tau$ matrix in the form:

$$\begin{pmatrix}
-\nu_1 & 1 & 0 & 0 & 0 & \dots & \dots & \dots \\
0 & -\nu_1 & 1 & 0 & 0 & \dots & \dots & \dots \\
 & & \ddots & & & \dots & & \\
\dots & \dots & 0 & -\nu_1 & c_1 f_2'(x^{2,0}) & 0 & \dots & 0 \\
 & & & & \ddots & & & \\
\dots & \dots & \dots & \dots & 0 & -\nu_n & 1 & \dots \\
 & & & & & & \ddots & \\
c_n f_1'(x^{1,0}) & 0 & \dots & \dots & \dots & \dots & 0 & -\nu_n
\end{pmatrix}$$

in which the cyclic fashion is evident.

At this point, we are interested in the characteristic polynomial of this matrix computed at the equilibrium point, i.e.

$$P(\mu) = \det(DF(x^*) - \mathbb{1}\mu)$$

Looking at the form of the matrix computed in $x^*$, and recalling that $\rho = \prod_{k=1}^{n} c_k f_k'((x^*)^{k,0})$, we get:

$$P(\mu) = \prod_{k=1}^{n} (\nu_k - \mu)^{\eta_k + 1} - (-1)^\tau \rho = (-1)^\tau [\prod_{k=1}^{n} (\nu_k + \mu)^{\eta_k + 1} - \rho]$$

Now, it is sufficient to recall hypothesis (4.10) of the theorem, which assures the existence of at least two eigenvalues such that $Re(\mu)<0$: from basic theory in stability and dynamical systems, it is immediate to conclude that the point $x^*$ in unstable.

At this point, we must use Assumption 2 and recall that the quantity $\rho$ is supposed to be negative; considering the matrix $DF(x^*)$ with reversed signs, it follows that it holds:

$$\det(-DF(x^*)) > 0 \tag{4.11}$$

Hence, it is immediate to conclude that part *1.* of the theorem is true thanks to the following auxiliary result in dynamical system, which corresponds to Theorem 4.3 in [34]:

**Theorem.** *Consider an analytic monotone cyclic feedback system in $\mathbb{R}^n$ (like the one discussed so far) with $\delta = -1$ in $\mathbb{R}^n_+$, which possesses a compact attractor $K \subset \mathbb{R}^n_+$. Moreover, suppose that $K$ contains a single equilibrium $x^*$, and that $DF(x^*)$ satisfies the following condition*

$$\delta \det(-DF(x^*)) < 0$$

*and has at least two eigenvalues with positive real part.*
*Then, the monotone cyclic feedback system has at least one, but no more than a finite number of, nontrivial periodic orbits. Plus, at least one of these is orbitally asymptotically stable.*

It remains to prove point *2.*: therefore, from now on we work with $\tau = 3$. In this situation, there are only three different possibilities, since $3 = \tau = n + \sum_{k=1}^{n} \eta_k$:

1. $n = 1$ population, and so $\eta_1 = 2$ necessarily; notice that this case is not of our interest since it would be trivial to take into account a model of *interacting* populations having just one of them; however, we include it in our discussion for completeness. In this case:

$$DF(x) = \begin{pmatrix} -\nu_1 & 1 & 0 \\ 0 & -\nu_1 & 1 \\ c_1 f_1'(x^{1,0}) & 0 & -\nu_1 \end{pmatrix}$$

2. $n = 2$ populations, and without loss of generality we can assume $\eta_1 = 0$, $\eta_2 = 1$; therefore we have:

$$DF(x) = \begin{pmatrix} -\nu_1 & c_1 f_2'(x^{2,0}) & 0 \\ 0 & -\nu_2 & 1 \\ c_2 f_1'(x^{1,0}) & 0 & -\nu_2 \end{pmatrix}$$

3. $n = 3$ populations, with $\eta_1 = \eta_2 = \eta_3 = 0$; hence,

$$DF(x) = \begin{pmatrix} -\nu_1 & c_1 f_2'(x^{2,0}) & 0 \\ 0 & -\nu_2 & c_2 f_3'(x^{3,0}) \\ c_3 f_1'(x^{1,0}) & 0 & -\nu_3 \end{pmatrix}$$

At this point it is important to recall that $\rho$ must be strictly negative by assumption; then: in the first case, $c_1 f_1'(x^{1,0}) < 0$, in the second either one between $c_1 f_2'(x^{2,0})$ and $c_2 f_1'(x^{1,0})$ is negative, in the third either or the three factors are all negative or there is just one of them with this property. However, in all the three situations, we can reduce ourselves to the case in which all the non-zero entries which are outside the diagonal (which are, by definition, already negative since $\nu_k > 0$), are strictly negative (it is indeed sufficient to act a change of variables to fall in this situation).

Therefore, when $\tau = 3$, we have that $-DF(x^*)$ is a positive irreducible matrix; we are so allowed to use the popular Perron-Frobenius Theorem, which assures that the Jacobian matrix $DF(x^*)$ has a unique largest eigenvalue $\mu < 0$, whose associated eigenvector has all the components that are, again, strictly negative. The theorem also implies that the remaining eigenvectors, corresponding to the remaining pair of complex conjugate eigenvalues - that have positive real part thanks to condition (4.10)- don't have all components of the same sign.

We are almost done: it is sufficient to adapt the following theorem -originally born in the context of competitive differential equations- to conclude with the second statement (for further information about this auxiliary result, the reader is invited to see Theorem 1.7 in [27]):

**Theorem.** *Consider n species represented by a system of differential equations:*

$$\dot{x}_i = x_i N_i(x_1, \ldots, x_n) = F_i(x_1, \ldots, x_n) \quad i = 1, \ldots, n$$

*with* $(x_1, \ldots, x_n) \in \mathbb{R}^n_+$. *We assume the following three conditions hold for the system:*

- *Dissipation: there is a compact invariant set K -the fundamental attractor- which uniformly attracts each compact set of initial values.*
- *Irreducibility: the community matrix* $DN = [\partial N_i / \partial x_j]$ *is irreducible at every point in the interior of* $\mathbb{R}^n_+$.
- *Competition: the rates* $N_i$ *are taken to be* $C^1$ *functions satisfying the competition condition* $\partial N_i / \partial x_j \leq 0$ *for* $i \neq j$.

*Then, assuming additionally that the origin is a source for the flow in* $\mathbb{R}^n_+$, *and that at every equilibrium in* $\mathbb{R}^n_+ \backslash \{0\}$ *it holds* $\partial N_i / \partial x_j < 0$ *for all* $i, j$, *we have that every trajectory in* $\mathbb{R}^n_+ \backslash \{0\}$ *is asymptotic to one in* $\Sigma$, *which is homeomorphic to the* $n-1$*- dimensional simplex* $\Delta_{n-1}$ *by radial projection.*

Adapting the above mentioned result to out neuronal context, we notice that all the hypotheses hold thanks to the previous considerations about the entries of Jacobian matrix $DF(x^*)$ (that would correspond to the community matrix) and thanks to the fact we have already proved that an invariant set

$K$ exists. Hence, we are sure about the existence of a globally attracting invariant surface $\Sigma$ such that every trajectory in the compact invariant set $K$ is, sooner or later, attracted to $\Sigma$. Notice that in the present context this shows that the two-dimensional unstable manifold of the equilibrium $x^*$ is a neighborhood of $x^*$ in $\Sigma$, and, therefore, $x^*$ is a repellor for the flow on the compact invariant set $K$, since it lies in the global attractor $\Sigma$ (the reader can refer to [3] to deepen these results, since what we are here going is actually a generalization of what has been done in Theorem 6.3 of the mentioned paper).

The essential tool to conclude, at this point, is the well-known Poincarè-Bendixson theorem, which characterizes the long-term behaviour of orbits of continuous dynamical systems. We report it in its simpler version, for a generic dynamical system on the plane. (For a proof of this classical result in dynamical systems theory, we refer the reader to [20]).

**Theorem** (**Poincarè-Bendixson**). *Let $\dot{x} = F(x)$ be a $C^1$ autonomous system in $\mathbb{R}^2$. Suppose that we have $x \in \mathbb{R}^2$ such that $\omega(x)$ is compact and doesn't contain any equilibria. Then, $\omega(x)$ is a periodic orbit. Similarly, if $\alpha(x)$ is compact and doesn't contain any equilibria, then $\alpha(x)$ is a periodic orbit.*

Thanks to it the thesis immediately follows; indeed, it implies that every connected attractor-free set is either a periodic orbit (or eventually a set of equilibria); hence, each of the aforementioned trajectories will finally converge to a *non constant periodic* orbit. □

Well, the bottom line is: Theorem 4.2, thanks to Poincarè-Bendixson, implies that every solution to *MFC* is sooner or later be attracted by a non constant and periodic orbit; here they are, oscillations for the limit system!

The question, at this point, is: since, by now, we have been dealing with the limit system, to which extent are these oscillations felt by the finite size system? A first answer is directly given by the Central Limit theorem we proved in the previous chapter; indeed, it follows as an immediate corollary of Theorem 4.2 that hypothesis (3.24) needed to state Theorem 3.11 (i.e. *CLT*) is satisfied when considering two populations, ($n = 2$).

**Corollary 4.3.** *Consider, for the model in object, a situation with $n = 2$ populations, and suppose that assumptions and conditions needed for Theorem 4.2 hold true. Then, there exist $\gamma_1, \gamma_2 > 0$ such that:*

$$\liminf_{t\to\infty} \frac{m_t^k}{t} \geq \gamma_k, \qquad k = 1, 2$$

*Proof.* We are dealing with a *negative* feedback system, so that -since $k = 1, 2$- we have $c_1 c_2 < 0$. Therefore, thanks to Theorem 4.2, every solution is

-sooner or later- attracted to a non constant periodic orbit.
Now recall the form of the integrated intensities in (4.4), i.e.

$$m_t^k = \int_0^t f_k(x_s^k)ds = \int_0^t f_k(x_s^{k,0})ds$$

Moreover, take into account that the spiking rates $f_k$, for $k = 1, 2$, are supposed to be functions $f_k : \mathbb{R} \to \mathbb{R}_+$, that means they are strictly positive. Hence, thanks to the well-known Mean Value Theorem for integrals, we have:

$$\frac{1}{t}m_t^k = \frac{1}{t}\int_0^t f_k(x_s^{k,0})ds = f_k(x_{\bar{s}}^{k,0}) \quad \exists\ \bar{s} \in [0, t]$$

And it directly follows that:

$$\liminf_{t\to\infty} \frac{m_t^k}{t} > 0$$

□

Notice that this turns out to be relevant for our discussion, since in the followings (see section 5.3 and on) the focus will be on the two-class situation. Before coming back to the main question about how oscillations in the limit system are felt by the finite size system and studying the two-populations case in detail, in the following section we analyze for a while the interesting phenomena of stability switches.

## 4.5 About phase transitions and stability switches

We have already pointed out that the assumption of distributed delay, inherent in the use of the Erlang kernels, is a matter of fact that is necessary to take into account when modelling in biological sciences. As we are dealing with an unstable equilibrium, in our notation $x^*$, we can pass to the examination of the so called *stability switches*, or *phase transitions*. Indeed, in the current section, we analyze how the stability properties of the model change when delay is increased, since it is known that the stability of an equilibrium may be lost when delay is increased, and that a further increase in the dimension of the delay may result in restabilization. (For an in depth analysis of these phenomena, the reader is invited to refer to [4] and [12]).

Therefore, in our context, and recalling that the distributed delay converges to a discrete delay as $\eta_k \to \infty$ (while keeping $\eta_k/\nu_k$ constant), we investigate what happens to the equilibrium when increasing the memory (i.e. $\eta_k$ for some $k = 1, \ldots, n$), or, equivalently, the whole dimension $\tau$ of the system.

In order to do that, we need a brief parenthesis on the instability condition pointed out in (4.10).

**Remark.** We consider the particular situation in which the parameters $\nu_k$ for $k = 1, \dots, n$ are such that $\nu_1 = \nu_2 = \cdots = \nu_n = \nu$, and let the whole dimension of the system be $\tau \geq 3$.
Then, the characteristic polynomial that turned out in the proof of Theorem 4.2 becomes:

$$P(\lambda) = (-\nu - \mu)^\tau - (-1)^\tau \rho = (-1)^\tau [(\nu + \mu)^\tau - \rho]$$

where we recall that, by assumption, $\rho < 0$. Therefore, to compute explicitly the $\tau$ eigenvalues, we need to distinguish two cases:

- *Case $\tau$ odd*: the condition to obtain the eigenvalues, i.e. being a zero of the characteristic polynomial, is the following:

  $$(\nu + \mu)^\tau = \rho \iff (\nu + \mu)^\tau = (-1)|\rho| \iff \mu = -\nu - \sqrt[\tau]{|\rho|}$$

  Hence, recalling how to take square roots of complex numbers, we obtain that the eigenvalues in this case are:

  $$\mu_j = -\nu - |\rho|^{\frac{1}{k}} e^{i\frac{2j\pi}{\tau}} \quad j = 0, \dots, \tau - 1$$

  At this point, since $\tau$ is odd, it is important to notice that between the $\tau$ eigenvalues, there is one real root (the one corresponding to $j = 0$, i.e. $\mu = -\nu - |\rho|^{\frac{1}{\tau}}$, which is $< 0$), while the remaining are all complex conjugates pairs with real part $-\nu - |\rho|^{\frac{1}{\tau}} \cos(\frac{2j\pi}{\tau})$. Recall that we are interested in the real parts of such eigenvalues in order to give conditions for instability as the one in (4.10). The maximal value for these real parts is assumed in correspondence of the indexes $j \in \{(k-1)/2, (k+1)/2\}$, since, due to the presence of the sign minus, we need to minimize the function $cos(\alpha)$, which means keeping the angle $\alpha$ near to $\pi$.
  Hence, we arrive to the main point of this remark: we want to state a condition concerning explicitly the real part of the eigenvalues, which would imply condition (4.10). To this aim, we consider the maximal value for the real parts of the eigenvalues $\mu$ which has been found above: of course, if such a maximal value is $> 0$, recalling the solutions are complex conjugate pairs, then there are for sure at least two $\mu_j$ such that $Re(\mu_j) > 0$ as the hypotheses of Theorem 4.2 asked. Therefore, we have:

  $$Re(\mu) > 0 \iff -\nu + |\rho|^{\frac{1}{\tau}} \cos(\frac{\pi}{\tau}) > 0 \iff |\rho| > \frac{\nu^\tau}{(cos(\frac{\pi}{\tau}))^\tau} \quad (4.12)$$

- *Case $\tau$ even*: in this case, the previous conditions become

  $$(\nu + \mu)^\tau = (-1)|\rho| \iff \mu = -\nu + \sqrt[\tau]{-|\rho|} = -\nu + \sqrt[\tau]{-1}\sqrt[\tau]{|\rho|}$$

And so, the eigenvalues are:

$$\mu_j = -\nu + |\rho|^{\frac{1}{k}} e^{-i\frac{\pi}{\tau}} e^{i\frac{2j\pi}{\tau}} = -\nu + |\rho|^{\frac{1}{k}} e^{i\frac{(2j-1)\pi}{\tau}} \quad j = 1, \dots, \tau$$

Since $\tau$ is even, the eigenvalues in this situation are all complex conjugate pairs, whose real part is given by $-\nu + |\rho|^{\frac{1}{\tau}} \cos(\frac{(2j-1)\pi}{\tau})$. The maximal value for such eigenvalues is assumed when $j \in \{1, \tau\}$: this comes directly from the fact there is a plus in the formula $-\nu + |\rho|^{\frac{1}{\tau}} \cos(\frac{(2j-1)\pi}{\tau}$; thus, since the function $\cos(\alpha)$ assumes the maximal values in correspondence of angles $\alpha \in \{0, 2\pi\}$, then the correspondent indexes to be considered to achieve the maximum are exactly $j = 1, \tau$. In correspondence of these indexes -as in the odd case- the maximal value for the real parts is $-\nu + |\rho|^{\frac{1}{\tau}} \cos(\frac{\pi}{\tau})$; hence, again, condition (4.12) found out in the case $\tau$ odd implies (4.10).

Recalling the previous remark, the study of phase transitions due to increasing memory can now start, working in a situation as the one described above, i.e. considering $\nu_1 = \cdots = \nu_n = \nu$ while modelling $n$ populations of interacting neurons.

In the first example -recalling $\tau$ is the whole dimension of the system *MFC*- we assume $\tau = 2$: hence, we are in the even situation, and the eigenvalues in this case are exactly $\mu_{1,2} = -\nu \pm i\sqrt{|\rho|}$; the real part is, in both cases, $-\nu$, which is strictly negative: therefore, the equilibrium point in stable, and so, at the latest, there will be only damped oscillations. Therefore, this situation is not of particular interest.

Instead, focusing on the case $\tau \geq 3$, $\nu_1 = \cdots = \nu_n = 1$, a relevant threshold comes out: first of all, notice that, in this situation, increasing the memory (i.e. the values of some $\eta_k$ or the dimension $\tau$ of the system), the coordinates $(x^*)^{k,0}$ of the equilibrium $x^*$ don't change, due to the fact $\nu = 1$ stands at the denominator counting with a power of order $k$ (the reader can refer to the proof of Proposition 4.1 to get convinced about that); hence, recalling that $\rho = c_1 f_1'((x^*)^{1,0}) \cdot \cdots \cdot c_n f_n'((x^*)^{n,0})$, also its value doesn't change if increasing the memory. Therefore, recalling (4.12):

$$|\rho| > \frac{\nu^\tau}{(cos(\frac{\pi}{\tau}))^\tau}$$

it is immediate to notice that, on the right, as $\tau \to \infty$, the value of the fraction tends to one, and considering $|\rho| \in (1, 8)$ (which comes out from the fact, given $\tau = 3$,the previous inequality on $|\rho|$ turns out in : $|\rho| > \frac{1}{(\cos(\frac{\pi}{3}))^3} = 8$), we immediately get that for $\rho \in (-8, -1)$ there exists $\bar{\tau} > 3$, minimal, such that:

$$|\rho| > \frac{\nu^\tau}{(cos(\frac{\pi}{\tau}))^\tau} \quad \forall \quad \tau \geq \bar{\tau}$$

whence,

$$|\rho| \leq \frac{\nu^\tau}{(cos(\frac{\pi}{\tau}))^\tau} \quad for \quad \tau < \bar{\tau}$$

It is evident that, increasing the memory up to a value that overcomes the threshold $\bar{\tau}$, the equilibrium $x^*$, from being stable and attracting, becomes unstable: a *phase transition* occurs.

In a more general situation, only assuming $\nu_1 = \cdots = \nu_n$, we can study what happens as taking limits due to increasing memory in relation (4.12); the l.h.s $|\rho|$ keeps being always bounded thanks to the fact the spiking rates are assumed to be Lipschitz continuous, whence the r.h.s tends to $\infty$ as $\nu$ increases, $\forall\ \tau$. Thus, the situation is the following:

- If the parameter $\nu$ is large enough, then $|\rho| \leq \frac{\nu^\tau}{(cos(\frac{\pi}{\tau}))^\tau}$, so that the system is stable and with no oscillations for any value of the memory.
- Instead, if $\nu > 1$ is fixed, the r.h.s of the previous inequality tends to $\infty$ if increasing the dimension $\tau$ of the memory: hence, it is possible that for small values of $\tau$ condition (4.12) holds -meaning the system is unstable- whereas the increase of $\tau$ leads to reversing the sign of the inequality and having a stabilization.

The main implication of this general discussion is the following: recalling that a discrete delay is attained as a limit of the parameters in the game, i.e. sending $\eta_k \to \infty$ when keeping constant the ratio $\eta_k/\nu_k$, and noticing that $\tau = n + \sum_{k=1}^n \eta_k$, then $\eta_k \to \infty \Rightarrow \tau \to \infty$: this results in the fact that a discrete delay system never exhibits oscillations, being always stable. This is another reason below the choice of the Erlang kernels, modelling the memory/delay in a distributed and not in a discrete way.

Thereupon, one of the take-home messages of this discussion about the emergence of oscillations and possibly of phase transitions is that, in the context of modelling such a system of interacting neurons, the choice of the Erlang kernels plays a crucial role; indeed, as we have already anticipated and we are now going to see in the next chapter, the very specific structure of this type of memory kernels -which play the role, in this context, of synaptic weights- allows us to adopt a Markovian description of the Hawkes processes in object; this turns out to be essential to simplify the discussion that will follow from here on, allowing us to give a more in-depth answer to the main question that had previously arisen: *how does the behaviour of the limit system affect the finite size system?* As we have seen, a first result to this issue is given by *CLT*; the rest is investigated in depth in the following section through the construction of an approximating diffusion process.

# Chapter 5

# A Diffusion Approximation for the Hawkes Process

Once investigated the oscillatory behaviour of the *limit system*, it is of interest analyzing the influence of such a behaviour on the *finite size system*. The credit for the following results relies on the choice of the very specific structure of the Erlang kernels; indeed they allow a Markovian description, through their intensities, of the Hawkes process playing this game. This is a crucial point: we have already pointed out that, in general, Hawkes processes are *infinite memory* processes, while the Markovian world relies on memorylessness, dealing just with the fact that, given the present state, the future does not depend on the past events. Moreover, we know that working in a Markovian world simplifies and allows a lot of studies, and different considerations, due to the strong theory that stands below.

Recalling that the setting is the one of the previous section, dealing with Erlang kernels, in a monotone, cyclic and negative feedback context, in the subsequent section we build a Piecewise Deterministic Markov Process (from now on, *PDMP*), that allows a new -equivalent- representation of the multiclass system in object.

Then, in a later section, we deepen the analysis of the model building an approximating diffusion process, which we show to be close "in a weak sense" to the previously defined *PDMP*.

We conclude our study looking at just two interacting populations: the mindblowing result we show at the end of the chapter is that the fixed population size system exhibits the same type of oscillations as the limit system.

## 5.1 Building an associated Piecewise Deterministic Markov Process

Recalling the form (4.2) of the Erlang memory weights, we shall see in a moment that, just by deriving them successively with respect to time, it is possible to write the *intensity processes* (remember the form of the intensities as introduced in the previous chapter (4.1)) to get an equivalent *high dimensional* -of dimension $\tau$- system of *PDMPs*.
The procedure is the following: first of all we need to consider the stochastic processes

$$X_k^N(t) := \frac{1}{N_{k+1}} \sum_{j=1}^{N_{k+1}} \int_{]0,t[} h_{kk+1}(t-s) dZ_{k+1,j}^N(s) = \int_{]0,t[} h_{kk+1}(t-s) d\bar{Z}_{k+1}^N(s) \tag{5.1}$$

with $\bar{Z}_{k+1}^N(s) = \frac{1}{N_{k+1}} \sum_{j=1}^{N_{k+1}} Z_{k+1,j}^N(s)$ being the empirical spike counting measure, which has not to be confused with the limit process $\bar{Z}_k(t)$.
These processes $(X_k^N(t-))_{t\geq 0}$ are Càdlag adapted processes (of which we have to consider the left continuous version due to the fact the intensities are assumed to be predictable processes) that determine *completely* the dynamics of the system: to get convinced about that it is sufficient to recall the form of the intensities. Notice that, if we hadn't chosen the synaptic weights in the form of an Erlang distribution, these variables wouldn't have been Markovian; this is a key point in the further developing of the system of stochastic processes, since it goes beyond the fact Hawkes processes are truly infinite memory processes for a general choice of the kernels.

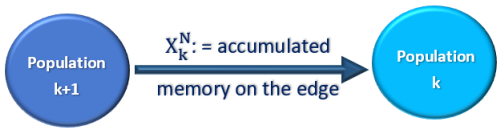



Figure 5.1: The intensities $X_k^N$ describe the accumulated memory belonging to the directed edge pointing from population $k+1$ to population $k$.

At this point, having in mind the very same procedure we followed to derive system *MFC* and the successive derivatives with respect to time of the Erlang kernels

$$h'_{kk+1}(t) = -\nu_k h_{kk+1}(t) + c_k \frac{t^{\eta_k - 1}}{(\eta_k - 1)!} e^{-\nu_k t}$$

we introduce the "auxiliary" processes

$$X^N_{k,l}(t) := c_k \int_{]0,t[} e^{-\nu_k(t-s)} \frac{(t-s)^{\eta_k - l}}{(\eta_k - l)!} d\bar{Z}^N_{k+1}(s) \tag{5.2}$$

Then, following the same procedure as in the derivation of *MFC*, we directly obtain the following system of stochastic differential equations, that hence corresponds to the stochastic version of *MFC*, which we call, due to the structure of the interaction between populations, a *cascade* of memory terms. As *MFC*, the resulting total dimension is $\tau = n + \sum_{k=1}^{n} \eta_k$.

$$\begin{cases} dX^N_{k,l}(t) = [-\nu_k X^N_{k,l}(t) + X^N_{k,l+1}(t)]dt, & 0 \leq l < \eta_k \\ dX^N_{k,\eta_k} = -\nu_k X^N_{k,\eta_k}(t)dt + c_k d\bar{Z}^N_{k+1}(t), & l = \eta_k \end{cases} \tag{PDMP}$$

with the usual identification $X^N_k = X^N_{k,0}$.

Let's immediately point out that, in the subsequent discussion, we shall need to consider $A^X$ and $P^X_t$ as, respectively, the infinitesimal generator and the Markovian semigroup of such a process as defined in *PDMP*.

Notice we are using the same notation introduced above, i.e. $\bar{Z}^N_{k+1}(s) = (\sum_{j=1}^{N_{k+1}} Z^N_{k+1,j})/N_{k+1}$, where $Z^N_{k,j}$ "jumps" with intensity $f_{k+1}(X^N_{k+1,0}(t-))$: to get convinced about that, it is sufficient to recall the form of the intensities (4.1) and what we had found out in *MFC*, comparing the "last" equations of each cycle in *MFC* and in *PDMP*:

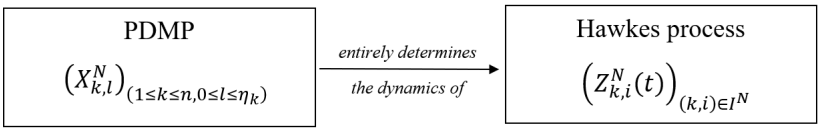
$$\begin{aligned} \frac{dx^{k,\eta_k}_t}{dt} &= -\nu_k x^{k,\eta_k}_t + c_k f_{k+1}(x^{k+1,0}_t), \quad l = \eta_k \quad (MFC) \\ dX^N_{k,\eta_k} &= -\nu_k X^N_{k,\eta_k}(t)dt + c_k d\bar{Z}^N_{k+1}(t), \quad l = \eta_k \quad (PDMP) \end{aligned}$$

Hence, the "mean" process $\bar{Z}^N_{k+1}(t)$ has "jumps" (recalling that in this context they can be considered as fired spikes) of size $1/N_{k+1}$, whence the "jump" rate equals to $N_{k+1} f_{k+1}(X^N_{k+1,0}(t-))$ (just looking at it as a superposition of processes).
It directly follows -and this is essential to build a diffusion approximation- that the variance of such a process is

$$\begin{aligned} Var(\bar{Z}^N_{k+1}(t)) = Var\left(\frac{1}{N_{k+1}} \sum_{j=1}^{N_{k+1}} Z^N_{k+1,j}\right) &= \frac{1}{N^2_{k+1}} \sum_{j=1}^{N_{k+1}} Var(Z^N_{k+1,j}(s)) \\ &= \frac{N_{k+1} f_{k+1}(X^N_{k+1,0}(t-))}{N^2_{k+1}} = \frac{f_{k+1}(X^N_{k+1,0}(t-))}{N_k + 1} \end{aligned} \tag{5.3}$$

where the fact $Var(\bar{Z}^N_{k+1,j}(t)) = f_{k+1}(X^N_{k+1,0}(t-))$ comes directly from the fact we are considering counting processes with exactly these intensities.

At this point, the essential fact is the following: due to the equivalence between the description of the system offered by the process in (5.1) and the original Hawkes process in the finite size regime $(Z^N_{k,i}(t))_{(k,i)\in I^N}$ (see Section (3.2)), we have that:




## 5.2 The Approximating Diffusion Process

At this stage we are ready to build an approximating diffusion process *ADP* that is henceforth going to substitute the role of *PDMP* in the study of the model.
But *why should we pass to an approximating diffusion process?* The complete answer to this huge issue can be found either in the "biological" introduction in Chapter 1, or more in detail from a mathematical point of view in the *Mathematical Tools* chapter, but briefly we can say that diffusion is used to deal with noise more easily and to replace complicated and analytically intractable stochastic processes with an appropriate process, much easier to handle.

Hence, having in mind the variance computed in the previous section and the definition of the process $\bar{Z}^N_{k+1}(t)$, we can consider the generalities about the well-known technique of the *approximating diffusion* in the neuronal context, related to the *Fokker-Planck* equation, which we henceforth call $Y^N$:

$$\begin{cases} dY^N_{k,l}(t) = [-\nu_k Y^N_{k,l}(t) + Y^N_{k,l+1}(t)]dt, \quad 0 \le l < \eta_k \\ dY^N_{k,\eta_k} = -\nu_k Y^N_{k,\eta_k}(t)dt + c_k f_{k+1}(Y^N_{k+1,0}(t))dt + c_k \frac{\sqrt{f_{k+1}(Y^N_{k+1,0})(t)}}{\sqrt{N_{k+1}}} dB_{k+1}(t) \end{cases} \tag{ADP}$$

where the "noise" of each population -intended as the stochastic fluctuations the activity of a population of neurons inevitably occurs in due to many different factors, as deepened in Chapter 1- is approximated by the Brownian motions $(B_k(t))_{1\le k\le n}$ (which we assume to be standard and independent).

As we can see, we are approximating the process $\bar{Z}^N_{k+1}(t)$ via capturing the drift (of the mean, i.e the rate $f_{k+1}(Y^N_{k+1,0}(t))$) and diffusion (change in the variance), modelled by the term $\sqrt{Var}\ dB_{k+1}(t)$, as prescribed by the approximating diffusion operation, and accordingly to the more general fact

that a stochastic random process often consists in a deterministic (drift) part, and of a random component (or noise).

Notice that as *MFC* and *PDMP*, the dimension of the state space of the diffusion approximation process $Y^N$ is $\tau$; moreover, in the following we shall use $A^Y$ and $P_t^Y$, accordingly to the notation used for $X^N$, to indicate respectively the infinitesimal generator and the Markov semigroup of process $Y^N$.

**Notation.** Denoting an element $x \in \mathbb{R}^\tau$ as $x = (x^1, \dots, x^\tau)$, we consider the following norm for a generic function $f$ defined on $\mathbb{R}^\tau$:

$$\|f\|_{l,\infty} := \sum_{j=0}^{l} \sum_{|i|=j} \|\partial^i f\|_\infty$$

where we recall that $\|f\|_\infty = \sup_x |f(x)|$, $x \in \mathbb{R}^\tau$.

The setting is now ready to give the statement and the proof of the main theorem of this section: it is a first step towards a convergence in law result, showing that that approximating diffusion process $Y^N$ is a good approximation of $X^N$.

**Theorem 5.1.** *Suppose the spiking rates $f_k \in C_b^5 := \{f | f$ is a bounded function having derivatives up to order 5 which are bounded themselves$\}$. Then, there exists $C$ constant (depending on the rates $f_1, \dots, f_n$ and the bounds holding for their derivatives) such that, for any function $\phi \in C_b^4(\mathbb{R}^\tau; \mathbb{R})$, the following is true:*

$$\|P_t^X \phi - P_t^Y \phi\|_\infty \leq Ct \frac{|\phi\|_{4,\infty}}{N^2} \tag{5.4}$$

*where $N$ is the total number of neurons within the model.*

Hence, being able to approximate the original process with the approximating diffusion process $Y^N$ allows to find out that *ADP* is a good small noise approximation of *PDMP* when $N \to \infty$, being close in "weak sense" to the original *PDMP* defining the Hawkes process. (However, we must notice that for $N \to \infty$ the diffusive component of *ADP* tends to zero as the total number of neurons tends to infinity, so that we can't say anymore *ADP* is properly a diffusion).

In any case, the fact the *PDMP* $X^N$ -which, we recall, entirely determines the dynamics of the original Hawkes process- is well approximated by *ADP* is essential: indeed, in the subsequent section, we study the approximating diffusion $Y^N$ in the case of two populations of neurons, in order to see how their collective behaviour is similar to the oscillatory one predicted -for the limit system- in the previous chapter.

*Proof.* To prove the theorem in just a few steps, we need the following three auxiliary results about generators (which we denote with $A^{\cdot}$) and Markov semigroups (denoted by $P_t^{\cdot}$); all the three of them need, in the hypotheses, that the assumptions for Theorem (5.1) hold, so that we give them for grant and omit to repeat them in each statement.
For a proof of these subsidiary lemmas, we refer the reader to [17].

**Lemma 1.** There exists a constant $C$ such that for all $\phi \in C_b^3(\mathbb{R}^\tau;\mathbb{R})$ it holds:

$$\|A^X\phi - A^Y\phi\|_\infty \leq C\frac{\|\phi\|_{3,\infty}}{N^2} \tag{5.5}$$

**Lemma 2.** There exists a constant $C$ such that for all $\phi \in C_b^4(\mathbb{R}^\tau;\mathbb{R})$, for any $\delta > 0$ the following two estimates hold:

$$\|P_\delta^Y\phi - \delta A^Y\phi\|_\infty \leq C\delta^2\|\phi\|_{2,\infty} \tag{5.6}$$

And,

$$\|P_\delta^X\phi - \phi - \delta A^X\phi\|_\infty \leq C\delta^2\|\phi\|_{2,\infty} \tag{5.7}$$

**Lemma 3.** There exists a constant $C$ such that for all $\phi \in C_b^4(\mathbb{R}^\tau;\mathbb{R})$, for any $t > 0$ the following holds:

$$\|P_t^Y\phi\|_{4,\infty} \leq C\|\phi\|_{4,\infty} \tag{5.8}$$

Given the above results, it is easy to conclude with the actual proof of the Theorem in object.

First of all, set $\delta > 0$ and consider the times $t_k$ defined as $t_k = k\delta \wedge t$ for $k \geq 0$.
Then, define the quantity:

$$\Delta_\delta\phi(x) = P_\delta^X\phi(x) - P_\delta^Y\phi(x)$$

Given the above notation, the following estimate -which relies on a telescopic sum- comes out:

$$\|P_t^X\phi - P_t^Y\phi\|_\infty \leq \sum_{k=0}^{t/\delta}\|P_{t-t_{k+1}}^X\Delta_\delta P_{t_k}^Y\phi\|_\infty \leq \sum_{k=0}^{t/\delta}\|\Delta_\delta P_{t_k}^Y\phi\|_\infty \tag{5.9}$$

At this point we can notice that, combining the two estimates in (5.6) and (5.7) and using a version of the triangular inequality for the norm we are dealing with, we get that:

$$\|\Delta_\delta\phi(x)\|_\infty = \|P_\delta^X\phi(x) - P_\delta^Y\phi(x)\|_\infty \leq \delta\|A^X\phi - A^Y\phi\|_\infty + \delta^2\|\phi\|_{4,\infty}$$

Hence, applying to the previous chain of inequalities the bound (5.5) we got in Lemma 1, we obtain:

$$\|\Delta_\delta \phi(x)\|_\infty = \|P_\delta^X \phi(x) - P_\delta^Y \phi(x)\|_\infty \leq \left[\delta C \frac{1}{N^2} + \delta^2\right] \|\phi\|_{4,\infty}$$

We are now ready to conclude: it is sufficient to use (5.8) and (5.9) to get:

$$\|P_t^X \phi(x) - P_t^Y \phi(x)\|_\infty \leq C\left(\frac{1}{N^2} + \delta\right) \|\phi\|_{4,\infty} \left(\sum_{k=0}^{t/\delta} \delta\right)$$

Noticing that $|\{k : k\delta \leq t\}| \leq \frac{t}{\delta}$ and choosing $\delta = \frac{1}{N^2}$, the last inequality can be estimated as:

$$\|P_t^X \phi(x) - P_t^Y \phi(x)\|_\infty \leq Ct \frac{\|\phi\|_{4,\infty}}{N^2}$$

that is exactly the bound we wanted to prove. □

## 5.3 Study for the fixed population size: two interacting populations

Let's recap for a moment what we have done so far: while studying the limit system, we built *MFC* and we discovered the emergence of collective periodic behaviour in the limit system itself, using proper tools of dynamical systems theory. After that, we asked ourselves to which extent the discovered oscillations are felt by the finite size system, and we saw a first answer was given by *CLT*.
Afterwords, we moved our attention to the study of the finite size system: first of all, thanks to the specific structure of the chosen memory weights, we built system *PDMP*, which entirely determines the dynamics of the original Hawkes process modelling the interacting neurons. Then, through building an approximating diffusion process $Y^N$ and setting system *ADP*, we found that the process $Y^N$ is indeed a good small noise approximation of the *PDMP*.
Hence, we are now ready to study in more detail the relation between the periodic behaviour we investigated for the limit system and the finite size system; or better, we would like to understand -and discover if- the highlighted oscillations of the limit system in *MFC* affect the behaviour of the finite system, which we shall, from now on, identify with the diffusion approximation process $Y^N$ described by *ADP* (we are allowed to to that thanks to convergence result found in Theorem (5.1)).
In order to perform such a study, henceforth we consider just $n = 2$ interacting populations of neurons: the synaptic weights are the in the usual

Erlang form, hence populations 1 and 2 are respectively identified by the parameters $\eta_1, \nu_1$ and $\eta_2, \nu_2$, by the respective spiking rates $f_1$ and $f_2$, and by the constants $c_{1,2} \in \{+1, -1\}$ describing the excitatory/inhibitory influence between the two populations.
First of all we can point out the *drift vector* $b(x)$, corresponding to the "deterministic" part of the process and the *diffusion matrix* $\sigma(x)$ (which is, more precisely, a $\tau \times 2$ matrix, corresponding to noise) of $ADP$, in the specific case of $n = 2$ interacting populations, with the generic notation $x = (x^1, \dots, x^\tau) \in \mathbb{R}^\tau$ ($\tau = n + \eta_1 + \eta_2 = 2 + \eta_1 + \eta_2$) in place of the extended version $x = (x^{1,0}, \dots, x^{1,\eta_1}, x^{2,0}, \dots, x^{2,\eta_2})$:

$$\text{Drift vector} = \begin{pmatrix} -\nu_1 x^1 + x^2 \\ -\nu_1 x^2 + x^3 \\ \vdots \\ -\nu_1 x^{\eta_1+1} + c_1 f_2(x^{\eta_1+2}) \\ -\nu_2 x^{\eta_1+2} + x^{\eta_1+3} \\ \vdots \\ -\nu_2 x^\tau + c_2 f_1(x^1) \end{pmatrix} =: b(x) \tag{5.10}$$

$$\text{Diffusion matrix} = \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & \frac{c_1}{\sqrt{N_2}}\sqrt{N f_2(x^{\eta_1+2})} \\ 0 & 0 \\ \vdots & \vdots \\ \frac{c_2}{\sqrt{N_1}}\sqrt{N f_1(x^1)} & 0 \end{pmatrix} =: \sigma(x) \tag{5.11}$$

Given this setting it is obvious to write the following SDE (Stochastic Differential Equation), corresponding to system $ADP$, specified to the case $n = 2$, in vectorial form:

$$dY^N(t) = b(Y^N(t))dt + \frac{1}{\sqrt{N}}\sigma(Y^N(t))dB(t) \qquad (ADP(2))$$

where the label *ADP(2)* is used to recall to the reader the fact we are dealing with the approximating diffusion process for $n = 2$ populations. Moreover, notice that, of course, when writing $B(t)$ in this context, we are considering the bi-dimensional Brownian motion $(B^1(t), B^2(t))^T$.

### 5.3.1 Oscillations in the diffusion approximation

The framework to work with $ADP$ in the case of 2 interacting populations is, at this point, set.
Throughout this section, with the purpose of investigating behaviours, we build the scaffolding that allows to get to the final result: *the diffusion*

*approximation ADP(2) has the same type of oscillations as the limit system* $(m_t^1, m_t^2)$ *in (MFC)*!

**Part 1 - The approximating diffusion visits the oscillatory region**

First of all we require the following:

**Assumption 3.** We need to assume as true all the assumptions made to state Theorem 4.2, in particular working in a negative feedback situation in which $\delta = c_1 c_2 < 0$, and dealing with $f_1 > 0$, $f_2 > 0$, smooth functions and non decreasing. Moreover, accordingly to Theorem 4.2, we need condition (4.10) to hold for the limit system, in order to be sure about the existence of a non constant, periodic, asymptotically orbitally stable orbit, let it be $\Gamma$, for the limit system itself.
This means: we work in a situation in which oscillations for the limit system occur.

**Notation.** Given a generic vector $x$, we write $s(x)$ to consider a *smoothed version* of $|x|$, meaning that:

- $s(x) = |x| \ \forall \ |x| \geq 1$;
- $|s'(x)| \leq C \ \exists \ C, \forall \ x$;
- $|s''(x)| \leq C \ \exists \ C, \forall \ x$;

At this point, we must have in mind that (as more explicitly pointed out in [11] and [40] which, respectively, present the theory below generalized Lyapunov functions and invariant set theorems for nonlinear dynamical systems and generalized Lyapunov criteria on stability of stochastic nonlinear systems) the existence of a generalized Lyapunov function for a stochastic process, given using a characterization in terms of infinitesimal generator of the process, directly leads to the existence of a compact and invariant set, which we shall call $K$ from now on; an invariant set like this has the important property of being such that the process visits it almost surely, infinitely many times. Hence, in the following proposition we set the definition of such a Lyapunov function for the system $Y^N = (Y_1^N, Y_2^N)$.

**Proposition 5.2.** *Consider Assumption 3 to hold true. Set:*

$$G(x) := \sum_{k=1}^{2} \sum_{l=0}^{\eta_k} \frac{l+1}{\nu_k^l} s(x^{k,l})$$

*Then, G is a Lyapunov-function for (ADP(2)) in the sense:*

$$A^Y G(x) \leq -cG(x) + d \tag{5.12}$$

*for some constants* $c, d > 0$ *constants depending on the quantity* $\max_{k=1,2} \|f_k\|_\infty$

*Proof.* Recalling that, for a generic uni-dimensional diffusion process of the type:

$$dY_t = bdt + \sigma dB_t$$

the infinitesimal generator, given $f$ a function with the needed regularity, is

$$Af(x) = bf'(x) + \frac{\sigma^2}{2} f''(x)$$

we have that the drift part of $A^Y G$ is given by the following dot product between $\tau$-dimensional vectors:

$$\sum_{k=1}^{2} \sum_{j=0}^{\eta_k - 1} \left[ -\nu_k x^{k,j} + x^{k,j+1} \right] \frac{\partial G}{\partial x^{k,j}} + \sum_{k=1}^{2} \left[ -\nu_k x^{k,\eta_k} + c_k f_{k+1}(x^{k+1,0}) \right] \frac{\partial G}{\partial x^{k,\eta_k}}$$

Notice that, at this point, we are interested in giving a bound just on the drift part (since the diffusion part of $A^Y G$ is itself bounded due to the structure of the matrix), and, in particular, we just look at those entries $x^{k,j}$ such that $|x^{k,j}| \leq 1$, since the other ones are for sure already bounded.

Hence, considering $|x^{k,j}| \leq 1$, we have:

$$\begin{aligned} \left[ -\nu_k x^{k,j} + x^{k,j+1} \right] \frac{\partial G}{\partial x^{k,j}} &= \left[ -\nu_k x^{k,j} + x^{k,j+1} \right] \frac{j+1}{\nu_k^j} sgn(x^{k,j}) \\ &\leq -\frac{j+1}{\nu_k^{j-1}} |x^{k,j}| + \frac{j+1}{\nu_k^j} |x^{k,j+1}| \end{aligned}$$

While, considering the components $x^{k,\eta_k}$, for $|x^{k,\eta_k}| < 1$ it holds:

$$\begin{aligned} &\left[ -\nu_k x^{k,\eta_k} + c_k f_{k+1}(x^{k+1,0}) \right] \frac{\partial G}{\partial x^{k,\eta_k}} = \\ &\qquad \left[ -\nu_k x^{k,\eta_k} + c_k f_{k+1}(x^{k+1,0}) \right] \frac{\eta_k + 1}{\nu_k^{\eta_k}} sgn(x^{k,\eta_k}) \end{aligned}$$

By splitting the two cases, $j = 0$, $j = 1, \ldots, \eta_k$, we see that, in conclusion, the drift part of $A^Y G$ can be bounded, $\forall\ k, j$, by:

$$-\sum_{k=1}^{2} \nu_k |x^{k,0}| - \sum_{k=1}^{2} \sum_{j=1}^{\eta_k} \frac{1}{\nu_k^{j-1}} |x^{k,j}| + 2 \max\left( \frac{\eta_1 + 1}{\nu_1^{\eta_1}} \|f_1\|_\infty, \frac{\eta_2 + 1}{\nu_2^{\eta_2}} \|f_2\|_\infty \right) = *$$

Now, recalling that $G(x) := \sum_{k=1}^{2} \sum_{j=0}^{\eta_k} \frac{j+1}{\nu_k^j} s(x^{k,j})$, we immediately notice that the previous line is such that:

$$* \leq -cG(x) + d$$

with $c, d$ constants evidently just depending on $\max_k \|f_k\|_\infty$; hence this ends the proof. □

As mentioned, as a consequence of the existence of the global Lyapunov function $G$ we are sure about the existence of a *compact invariant set* $K$ such that *ADP(2)* visits it infinitely often and almost surely; we can make the particular choice of considering

$$K = \{G \leq 2d/c\}$$

with $c, d$ corresponding to the ones in the above proposition, to get immediately that inequality (5.12) becomes:

$$A^Y G(x) \leq -\frac{c}{2} G(x) + d\mathbb{1}_K(x) \tag{5.13}$$

Notice that, in general, when studying the recurrence properties of continuous-time Markov processes, it is well-known (in particular, see [18]) and [19] that a "drift" condition on the generator like the one above obtained in (5.13) is a reliable criterion to provide an explicit bound on the exponential behaviour of the hitting times of a specific test-set. Hence, defining the hitting time of our interest, i.e. the one related to the compact invariant set $K$:

$$\theta_K = \inf\{t \geq 0 : Y^N(t) \in K\}$$

it holds (as in [19]) that

$$\mathbb{E}_x(e^{\frac{c}{2}\theta_K}) \leq G(x)$$

meaning that the excursions out of the invariant set $K$ have exponential moments. Therefore, the situation can be summarized as follows:

| **Limit system** | **Approximating diffusion** |
|---|---|
| Theorem 4.2 and assumption (4.10) assure the existence of $\Gamma$ non-constant orbit which is asymptotically orbitally stable, with period $T$. Without loss of generality we can assume $\Gamma$ *lives in the compact invariant set* $K$. | The existence of the global Lyapunov function $G$ allows to deduce the approximating diffusion possesses a compact invariant set $K$ to which it comes back infinitely often, with external excursions having exponential moments. |

What we shall show in a while is that each time $Y^N$ comes back to $K$, it visits the neighborhoods of the periodic orbit $\Gamma$.

**Notation.** In the sequel, we consider the followings:

- $S := S(\epsilon, \Gamma) := \{x : d(x, \Gamma) < \epsilon\}$, the "tube" around the orbit $\Gamma$ (the periodic orbit for the limit system);
- $Q_x^Y :=$ the law of the solution $(Y^N(t))_{t\geq 0}$ of *ADP(2)*, with initial condition $Y^N(0) = x$;

- $U = \{\phi \in C(\mathbb{R}_+; \mathbb{R}^\tau) : \phi(t) \in S(\epsilon, \Gamma) \ \forall \ 1 \leq t \leq t^*\}$ for fixed $t^* > 1$, i.e. the set of continuous functions that keep being in the tube $S(\epsilon, \Gamma)$ in the whole interval $[1, t^*]$; notice that we could easily shift the time and work in any interval of the type $[t_0, t^*]$ for any fixed $t_0 > 0$.

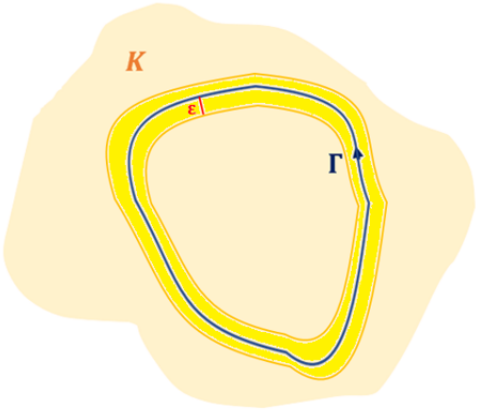



Figure 5.2: Representation of the compact subset $K$, on which we assume the periodic orbit $\Gamma$ lies on. The yellow region, of radius $\epsilon$, corresponds to the tube $S(\epsilon, \Gamma) := \{x : d(x, \Gamma) < \epsilon\}$

At this point, we have the following preliminary result:

**Proposition 5.3.** *Consider Assumption 3 to hold true. Then:*

$$\forall \ x \in \mathbb{R}^\tau, \ Q_x^Y(U) > 0$$

The above proposition is powerful: it claims that the approximating diffusion $Y^N$ visits the tube $S$, starting from an arbitrary initial point $x$ in $\mathbb{R}^\tau$ and in the chosen interval $[1, t^*]$ with positive probability! This is a first step towards discovering the same oscillatory behaviour it has already been detected for the limit system also for the two-populations model.

*Proof.* In order to prove this Proposition we need some additional notation, which we need to introduce a control argument essential for the proof.

First of all we have to deal with the *Cameron-Martin functional space*, which we denote with letter $\mathcal{H}$, whose definition is the following: setting an arbitrary but fixed time $T^* < \infty$, the functions $h \in \mathcal{H}$ are such that $h : [0, T^*] \to \mathbb{R}^2$ verifies the following properties:

- $h$ is measurable;
- $h$ has absolutely continuous components, i.e., $h^i(t) = \int_0^t \dot{h}^i(s)ds$, for $i = 1, 2$;
- It holds that $\int_0^{T^*} (\dot{h}^i)^2 ds < \infty$, for $i = 1, 2$;

In this framework, we consider the following deterministic system on $[0, T^*]$:

$$\begin{cases} d\phi(t) = b(\phi(t))dt + \frac{1}{\sqrt{N}}\sigma(\phi(t))\dot{h}(t)dt \\ \phi(0) = x \end{cases} \tag{5.14}$$

where $\phi : [0, T^*] \to \mathbb{R}^\tau$ is a function depending on $N$, i.e. "the number of neurons", $x$, the initial position and $h$, i.e., the choice of the function in the Cameron-Martin space. (We immediately notice the analogy with system *ADP(2)*).

Proceeding as in [28], (Theorem 3.1, part (a)), we consider the following control argument, originally taken from [35], to which we refer the reader to deepen the field of support theorems for diffusion processes. Notice that the proof of the following auxiliary result is quite long and technical: hence, we invite the reader to see the Appendix of [28], not to further burden the discussion.

**Proposition.** *Assume that the hypotheses beyond Theorem 4.2 are fulfilled and let $Q_x^{t^*}$ be the law of the solution $(Y^N(t))_{0\le t\le t^*}$ of ADP with initial position $Y^N(0) = x$.*
*Consider $\phi = \phi_{(N,h,x)}$ a solution of system (5.14) and fix $x \in K$ -where $K$ is the compact invariant subset already discussed- and $h \in \mathcal{H}$ such that $\phi = \phi_{(N,h,x)}$ exists on a time interval $[0, T^*]$ for $T^* > t^*$.*
*Then, it holds the following:*

$$(\phi_{(N,h,x)})_{|[0,t^*]} \in \overline{supp(Q_x^{t^*})}$$

The above result is the fundamental tool we now use to prove Proposition 5.3 in object: considering the same setting of the above proposition and recalling we are working with the two-populations model, from now on, as we did in the definitions of the drift and the diffusion matrices in (5.10) and (5.11), for $x \in \mathbb{R}^\tau$ we write $x = (x^1, \dots, x^\tau)$ with $\tau = n + \eta_1 + \eta_2 = 2 + \eta_1 + \eta_2$ in place of the extended version $x = (x^{1,0}, \dots, x^{1,\eta_1}, x^{2,0}, \dots, x^{2,\eta_2})$.

With this notation, it is important to bear in mind that the "first" component in each population (i.e. $x^{1,0}$ for population 1 and $x^{2,0}$ for population 2) corresponds, respectively, to $x^1$ and $x^{\eta_1+2}$, while the only two components interested by Brownian noise (i.e. $x^{1,\eta_1}$ for the first population and $x^{2,\eta_2}$ for the second) are respectively denoted with $x^{\eta_1+1}$ and $x^\tau$.

Hence, we fix an initial configuration $x \in \mathbb{R}^\tau$ and a time $t^* > 1$, and we recall that, in our notation, we use $\Gamma$ to denote the periodic (of period $T$) orbit of the limit system; thus, it seems natural to consider, hereafter, $\Gamma(t)$, for $t \in [0, T]$ a parametrization of the orbit.

At this point, the control argument takes hold of the proof: the aim is finding a *smooth* trajectory controlling the dynamics of the two components $\phi^1$ and $\phi^{\eta_1+2}$ (in terms of the solution $Y^N$, respectively corresponding to $Y^N_{1,0}$ and $Y^N_{2,0}$) from the initial position $x$ to a position on the periodic orbit $\Gamma$, considering a time period of unitary length.

To this aim, we show that we can choose a control $h$ (with the conventions introduced in the auxiliary Proposition above), which allows to write the equalities $\phi^1 = \gamma^1$, $\phi^{\eta_1+2} = \gamma^2$, with $\gamma : \mathbb{R}_+ \to \mathbb{R}^2$ being a $C^\infty$-solution to:

$$\begin{cases} \gamma(t) \equiv (\Gamma^1(t), \Gamma^{\eta_1+2}(t)) \ for \ t \in [1, \infty) \\ \gamma^1(0) = x^1 \\ \gamma^2(0) = x^{\eta_1+2} \end{cases} \tag{5.15}$$

Having in mind that the only coordinates driven by Brownian noise are $x^{\eta_1+1}$ and $x^\tau$ (since the diffusion coefficient $\sigma$ is null on the other coordinates), we must notice that any choice of such a control $h$ allows to control directly these coordinates, while, in system (5.15) we are prescribing the trajectories for $x^1$, $x^{\eta_1+2}$.

Hence, given the above remark, we must prove we are allowed to choose a control $h$ such that $\phi^1 = \gamma^1$, $\phi^{\eta_1+2} = \gamma^2$: to this aim, assume we have already found this control function $h$. Then, recalling the form (5.10) of the drift vector $b(x)$, and having prescribed, i.e. fixed, $\phi^1$ and $\phi^{\eta_1+2}$ towards the action of $\gamma$, for sure we have that all the other coordinates are entirely determined as and measurable functions of $\gamma$ as follows:

- *Population 1*:

$$\begin{aligned} \phi^2(t) &= \frac{d\phi^1(t)}{dt} + \nu_1 \phi^1(t) \\ &\cdots\cdots\cdots\cdots \\ \phi^{\eta_1+1}(t) &= \frac{d\phi^{\eta_1}(t)}{dt} + \nu_1 \phi^{\eta_1}(t) \end{aligned} \tag{5.16}$$

- *Population 2*:

$$\begin{aligned} \phi^{\eta_1+3}(t) &= \frac{d\phi^{\eta_1+2}(t)}{dt} + \nu_2 \phi^{\eta_1+2}(t) \\ &\cdots\cdots\cdots\cdots \\ \phi^{\tau}(t) &= \frac{d\phi^{\tau-1}(t)}{dt} + \nu_2 \phi^{\tau-1}(t) \end{aligned} \tag{5.17}$$

**Remark.** The previous expressions for the coordinates of $\phi$ directly come from (5.14), which we recall being:

$$d\phi(t) = b(\phi(t))dt + \frac{1}{\sqrt{N}}\sigma(\phi(t))\dot{h}(t)dt$$

and from the structure of $b$ and $\sigma$ in (5.10) and (5.11), just by moving the derivatives on the right hand sides and making explicit the correspondent entry of $\phi$ on the left.

At this point we must recall the cyclic structure of the limit system in *MFC*, in order to look at its periodic orbit $\Gamma$:

$$\begin{cases} \frac{dx_t^{k,l}}{dt} = -\nu_k x_t^{k,l} + x_t^{k,l+1}, \quad 0 \leq l < \eta_k \\ \frac{dx_t^{k,\eta_k}}{dt} = -\nu_k x_t^{k,\eta_k} + c_k f_{k+1}(x_t^{k+1,0}), \quad l = \eta_k \end{cases} \tag{MFC}$$

Recalling the previous relations for $\phi$ and *MFC*, we immediately see the relations are the same, meaning that $\phi(t) = \Gamma(t)$ for all $t \geq 1$, i.e, after time 1, *the trajectory evolves on the periodic orbit* $\Gamma$ (which is exactly what we would like to prove).

Hence, we must show we can find a control $h$ which allows the above considerations for $\phi$: as mentioned, the control $h$ can directly act on the coordinates $\phi^{\eta_1+1}$ and $\phi^{\tau}$ towards drift part and diffusion coefficient $\sigma$ as follows:

$$\begin{aligned} \frac{d\phi^{\eta_1+1}(t)}{dt} &= -\nu_1\phi^{\eta_1+1}(t) + c_1 f_2(\phi^{\eta_1+2}(t)) + \frac{c_1}{\sqrt{N_2}}\sqrt{N f_2(\phi^{\eta_1+2}(t))}\dot{h}^1(t) \\ \frac{d\phi^{\tau}(t)}{dt} &= -\nu_2\phi^{\tau}(t) + c_2 f_1(\phi^1(t)) + \frac{c_2}{\sqrt{N_1}}\sqrt{N f_1(\phi^1(t))}\dot{h}^2(t) \end{aligned} \tag{5.18}$$

with $\dot{h}^1(t), \dot{h}^2(t)$ which have, of course, to be chosen, depending on the choice of the control function $h$, and where all the $\phi$'s are measurable functions of the trajectory $\gamma$.

The game is nearly closed; indeed, let's look at the following possible choice for the control $h : [0, \infty[ \to \mathbb{R}^2$:

$$\begin{cases} \dot{h}^1(t) = \frac{\frac{d\phi^{\eta_1+1}(t)}{dt} + \nu_1\phi^{\eta_1+1}(t) - c_1 f_2(\phi^{\eta_1+2}(t))}{c_1\sqrt{N}/\sqrt{N_2}\sqrt{f_2(\phi^{\eta_1+2}(t))}} \\ \dot{h}^2(t) = \frac{\frac{d\phi^{\tau}(t)}{dt} + \nu_2\phi^{\tau}(t) - c_2 f_1(\phi^1(t))}{c_2\sqrt{N}/\sqrt{N_1}\sqrt{f_1(\phi^1(t))}} \end{cases} \tag{5.19}$$

Why does this choice hold? First of all, notice that inserting these derivatives in (5.18), the equality is trivially verified; moreover, recall that the spiking rates for the two populations, $f_1$ and $f_2$, are assumed to be (strictly) positive

and lower bounded on the invariant subset $K$, hence the derivatives are well-defined $\in \mathcal{H}$.

Since we have $\phi^1(t) = \Gamma^1(t)$ and $\phi^{\eta_1+2}(t) = \Gamma^{\eta_1+2}(t)\ \forall\ t \geq 1$, i.e. the two components move on the periodic orbit of the limit system, looking at the dynamics of all the other components of $\phi$ as prescribed by (5.16) for population 1 and (5.17) for population 2, it necessarily must be

$$\phi(t) = \Gamma(t) \quad \forall\ t \geq 1$$

Moreover, looking at (5.19), we notice that

$$\begin{aligned} \dot{h}^1(t) = 0 \iff \frac{d\phi^{\eta_1+1}(t)}{dt} = \nu_1\phi^{\eta_1+1}(t) + c_1 f_2(\phi^{\eta_1+2}(t)) \\ \dot{h}^2(t) = 0 \iff \frac{d\phi^{\tau}(t)}{dt} = \nu_2\phi^{\tau}(t) + c_2 f_1(\phi^1(t)) \end{aligned} \tag{5.20}$$

and since $\phi(t) = \Gamma(t)$ for all $t \geq 1$, hence $\phi$ verifies *MFC*, it is immediate to notice that the above equalities trivially hold, so that $\dot{h}^1(t) = \dot{h}^2(t) = 0$ for all $t \geq 1$.

The moral is: *we built a control function $h \in \mathcal{H}$ that forces the trajectory of the system to evolve on the periodic orbit $\Gamma$ of the limit system after a fixed time.*

To conclude, recall that $U = \{\phi \in C(\mathbb{R}_+;\mathbb{R}^\tau) : \phi(t) \in S(\epsilon,\Gamma)\ \forall\ 1 \leq t \leq t^*\}$ for fixed $t^* > 1$; hence, if we consider the set $R = R(\epsilon,\phi) := \{\psi \in C(\mathbb{R}_+;\mathbb{R}^\tau) : d(\phi(t),\psi(t)) < \epsilon\ \forall\ 1 \leq t \leq t^*\}$, having $\phi(t) = \Gamma(t)\ \forall\ t \leq 1$, it is evident that $R \subset U$.

At this point the auxiliary proposition comes into play; indeed recall:

$$\phi_{(N,x,h)}|[0,t^*] \in \overline{supp(Q_x^{t^*})}$$

Wherefore we get $Q_x^{t^*}(R(\epsilon,\phi)) > 0$, meaning that:

$$R \subset U, \quad Q_x^{t^*}(R(\epsilon,\phi)) > 0 \implies Q_x^{t^*}(U) > 0$$

Hence, $\forall\ x \in \mathbb{R}^\tau,\ Q_x^Y(U) > 0$, $\forall\ t^* > 1$, as wanted.

□

**Part 2 - The visits happen within a finite time horizon**

At this point, the situation is the following: there is a positive probability $Y^N$ visits the tube around the periodic orbit $\Gamma$, which, let us again recall, is the periodic orbit of the limit system, but the question is: *what about the time period in which this visit takes place?* Or better, *can we already say that this visit happens almost surely within a finite time horizon?* The answer is: not yet. We need to proceed with our discussion and to introduce a few more definitions and notation in order to be (almost) sure about the visit to the tube to take place in a finite time horizon.

**Definition 5.4** (*Lie Algebra* $\Delta_{\mathcal{L}_M^*}$)**.** Consider $\sigma^1, \sigma^2 : \mathbb{R}^\tau \to \mathbb{R}^\tau$ the two columns in the diffusion matrix $\sigma$ in (5.11) of *ADP(2)*, and let $b$ be the drift vector in (5.10). Define a set $\mathcal{L}$ of vector fields by the initial condition $\sigma^1, \sigma^2$ and an arbitrary number of iterates as

$$L \in \mathcal{L} \Rightarrow [b, L], [\sigma^1, L], [\sigma^2, L] \in \mathcal{L} \tag{5.21}$$

Consider a natural number $M \in \mathbb{N}$ and define $\mathcal{L}_M$ as the subset of $L$ determined by the same initial conditions and at most $M$ iterations (5.21).
Use the notation $\mathcal{L}_M^*$ to indicate the closure of the subset $\mathcal{L}_M$ under the Lie bracketing operation.
Then, define:

$$\Delta_{\mathcal{L}_M^*} := LA(\mathcal{L}_M)$$

as the linear span of $\mathcal{L}_M^*$; we call $\Delta_{\mathcal{L}_M^*}$ the *Lie algebra* spanned by the subset $\mathcal{L}_M$.

**Remark** (*Lie bracket*)**.** Recall that the Lie bracket between two smooth vector fields $g, h : \mathbb{R}^\tau \to \mathbb{R}^\tau$ is:

$$[g, h]^j = \sum_{i=1}^{\tau} \left( h^i \frac{\partial g^j}{\partial x^i} - g^i \frac{\partial h^j}{\partial x^i} \right) \quad for\ j = 1, \dots, \tau$$

**Definition 5.5** (*Full weak Hörmander dimension*)**.** A point $x^*$ in $\mathbb{R}^\tau$ is said to be of *full weak Hörmander dimension* if it exists $M \in \mathbb{N}$ such that:

$$(\dim \Delta_{\mathcal{L}_M^*})(x^*) = \tau$$

meaning that the dimension of the Lie algebra is the maximal possible, i.e. the whole dimension $\tau$ of the state space.

Given the above notation, the crucial result is the following:

**Proposition 5.6.** *Considering the spiking rates $f_1(\cdot) > 0, f_2(\cdot) > 0$ accordingly to Assumption 3, it holds that, for all $x \in \mathbb{R}^\tau$, given $M = \max(\eta_1, \eta_2)$:*

$$(\dim \Delta_{\mathcal{L}_M^*})(x) = \tau$$

Hence, the *weak Hörmander dimension* condition holds for all point $x \in \mathbb{R}^\tau$! Notice that there is an implicit dependence on the orders of delay of the two populations, i.e. $\eta_1, \eta_2$, since the maximum number of iterations $M$ to "generate" the Lie algebra is $M = \max(\eta_1, \eta_2)$.

*Proof.* The proof simply follows just by computing the Lie-brackets $[\sigma^1, b]$, $[\sigma^2, b]$ and then bracketing the results successively with $b$. Everything works straightforward thanks to the structure of the drift vector $b$ in (5.10) and thanks to the assumption $f_1, f_2$ strictly positive. Since it consists of elementary but quite long operations about computing Lie brackets, let us omit this calculation in order not to further burden the discussion. □

**Remark.** Since the weak Hörmander condition holds, we know from [29] (Lemma 5.1, part (i)) that the following is true: considering $P_t$ the semigroup of a diffusion process, then it is a Feller semigroup, i.e., it maps the space of bounded continuous functions into itself; furthermore, *if the weak Hörmander condition is satisfied*, then $P_t$ is a *strongly Feller semigroup*, mapping the space of bounded measurable functions to the one of bounded continuous functions. The reader can deepen this kind of topics, strictly related to the Feller property, in [24].

Hence, this implies that the process in object is *strong Feller*, and this immediately allows to get to the following corollary:

**Corollary 5.7.** *Consider a set* $B \in \mathcal{B}(C(\mathbb{R}_+;\mathbb{R}^\tau))$ *such that* $B = \{\phi : \phi(t) \in A \ \forall \ t^* \leq t \leq t_1, t^* < t_1\}$, $\exists \ A \in \mathcal{B}(\mathbb{R}^\tau)$. *Then,*

$$x \mapsto Q_x^Y(B)$$

*is a continuous function.*

*Proof.* Our aim is to prove the continuity of the function $x \mapsto Q_x^Y(B)$ for sets $B$ of the type $B = \{\phi : \phi(t) \in B \ \forall \ t^* \leq t \leq t_1\}$, but since we are dealing with a Markovian and homogeneous context, what matters is the increment $t_1 - t^*$; hence we are allowed to write:

$$Q_x^Y(B) = \int Q_y^Y(\{\phi : \phi(t) \in B \ \forall \ 0 \leq t \leq t_1 - t^*\}) P_{t^*}^Y(x, dy) = P_{t^*}^Y \varphi(x)$$

where the second inequality comes directly from the definition of Markov semigroup since $\varphi(x) = Q_x^Y(\{\phi : \phi(t) \in B \ \forall \ 0 \leq t \leq t_1 - t^*\}$, $P_{t^*}^Y(x, dy)$ is the transition probability of process $Y^N$ and $P_{t^*}^Y$ corresponds to the Markov semigroup.

We are ready to conclude: it is sufficient to notice that $\varphi(x)$ is a measurable and bounded function, and the fact $Y^N$ has be proven to be strong Feller directly implies, by definition, the continuity of $P_{t^*}^Y \varphi(x)$. Hence, having $Q_x^Y(B) = P_{t^*}^Y \varphi(x)$, we have that $x \mapsto Q_x^Y(B)$ is a continuous function, as wanted. □

The consequence we are interested in is a direct application of the previous corollary; indeed, in order to prove that the visits to the periodic region arrive a.s. within a finite time horizon, the aim is to show that $x \mapsto Q_x^Y(U)$ is *lower-bounded* on compact sets, which means giving a control on the starting configuration $x$. And of course, in order to get the lower-boundedness on compacts, it is sufficient to get a continuity result on Borel sets $((\mathcal{C}(\mathbb{R}_+;\mathbb{R}^\tau)))$, as the above corollary exactly does.
Indeed, taking an initial condition $x \in K$ -recalling that $K$ is the invariant compact subset already cited- and considering the above introduced type of

sets $U = \{\phi \in C(\mathbb{R}_+;\mathbb{R}^\tau) : \phi(t) \in S(\epsilon,\Gamma) \in \mathcal{B}(C(\mathbb{R}_+;\mathbb{R}^\tau)) \ \forall\ 1 \le t \le t^*\}$, the corollary allows to conclude that

$$K \ni x \mapsto Q_x^Y(U)$$

is strictly lower bounded for all $t^* > 1$, which assures the visits to the oscillatory region to happen within a finite time horizon.

**Conclusion - Oscillations for the diffusion approximation**

At this point, the game is almost closed: the following theorem allows to show that, indeed, the approximating diffusion for $n = 2$ populations *ADP(2)* has the *same type of oscillations* as the limit system $(m_t^1, m_t^2)$, which have been already investigated in the previous chapter. In particular, we show that the process $Y^N$ *visits the tube* centered on the periodic orbit $\Gamma$ - i.e. the oscillatory region- *infinitely often* during time intervals of fixed length $t^*$ and that the waiting times between the visits possess exponential moments.

**Theorem 5.8.** *Let $\Gamma$ be a non constant, periodic, asymptotically orbitally stable orbit of the limit system of period $T$; moreover, consider Assumption 3 to hold true.*
*Define the following*

$$\theta_\Gamma(t^*) := \inf\{t \ge 0 : Y^N(s) \in S(\epsilon,\Gamma) \ \forall\ t \le s \le t + t^*\}$$

*which is the smaller time $t$ such that in the interval $[t, t + t^*]$ the solution $Y^N$ keeps being inside the tube centered at the periodic orbit.*
*Then $\forall\ \epsilon > 0$ and $\forall\ t^* > 1$, there exists $C, \lambda$ strictly positive constants such that:*

$$\mathbb{E}_x(e^{\lambda\theta_\Gamma(t^*)}) \le CG(x)$$

*where $G$ is the Lyapunov function as defined in Proposition 5.2.*
*In addition, we have that:*

$$\limsup_{t\to\infty} \mathbb{1}_{\{Y^N(s)\in S(\epsilon,\Gamma)\ \forall\ t\le s\le t+t^*\}} = 1 \quad \mathbb{P}_x - a.s.\ \forall\ x \in \mathbb{R}^\tau$$

*Proof.* Thanks to Proposition 5.3, we already know that $Q_x^Y(U) > 0$ for all initial positions $x$ and for each fixed time $t^*$. Moreover, taking into account $K$ is a compact set and having the continuity of the function $x \mapsto Q_x^Y(U)$ from Corollary 5.7, we can immediately deduce that $\inf_{x\in K} Q_x^Y(U) > 0$.

At this point, recall that, as a consequence of Proposition 5.2, we got that $\mathbb{E}_x(e^{\frac{c}{2}\theta_K}) \le G(x)$, so that we know the process $Y^N$ visits the compact set $K$ infinitely often and almost surely, with excursions outside of $K$ possessing exponential moments. Henceforth, equipped with this piece of information, and needing to prove a result concerning a lim sup, it is natural to apply

the conditional version of the well-known Borel-Cantelli lemma (we refer the reader to [33] to deepen this classical probability argument) to conclude that:

$$\limsup_{t\to\infty} \mathbb{1}_{\{Y^N(s)\in S(\epsilon,\Gamma)\ \forall\ t\leq s\leq t+t^*\}} = 1 \quad \mathbb{P}_x - a.s.\ \forall\ x \in \mathbb{R}^\tau$$

i.e., $Y^N(s)$ visits the oscillatory tube $\mathbb{P}_x$-almost surely during time intervals of length $t^*$, infinitely often. □

**Remark** (*On the choice of* $t^*$)**.** As we have already pointed out, we can choose the fixed time $t^*$ freely: we know that the periodic orbit $\Gamma$ has period $T$, and that the process $Y^N$ visits the oscillatory region infinitely often during a time interval of length $t^*$; hence, if we choose $t^*$ in such a way that $t^* \geq kT$, $\exists\, k \leq 1, k \in \mathbb{N}$, it is clear to notice that the process $Y^N$ will *oscillate infinitely often a.s.*

It is now natural to regroup ideas to summarize what has been done so far throughout this section by comparing the dynamics of the limit system $(m_t^1, m_t^2)$ and the one of the approximating diffusion process $(Y_1^N, Y_2^N)$.

Looking at what has been done in the previous chapter, Theorem 4.2 assured the existence of an orbit $\Gamma$, non constant and periodic, asymptotically orbitally stable. Indeed, it was shown that the limit system exhibits oscillations. Hence, the general dynamical systems theory assures the existence of a Lyapunov function in a neighborhood of such an orbit, that decreases along the trajectories of the limit system itself, therefore describing the "attraction" the periodic orbit $\Gamma$ exerts on the system.

Moving on to the study that has been carried out in the current chapter, in which the approximating diffusion process has been derived, the procedure has been the following: towards a Lyapunov argument it was shown that the diffusion process comes back to an invariant compact set $K$ infinitely often and almost surely, and, without loss of generality it was assumed $\Gamma \subset K$. After that, the cascade structure of drift vector in *ADP(2)* allowed to conclude that the weak Hörmander condition holds on the whole space, hence getting that the diffusion $Y^N$ is strong Feller.

These whole considerations enabled to show that that once the diffusion enters $K$, it also enters the basin of attraction of the periodic orbit $\Gamma$ (i.e. the tube $S$ around the orbit), being attracted by $\Gamma$ itself; at the end, towards a control argument, Proposition 5.3 and Theorem 5.8 allowed to prove that the entrances in the basin of attraction (hence, the visits to the periodic region) happen - infinitely often and almost surely- and within a finite time horizon, with waiting times possessing exponential moments.

Therefore, summarizing, it is clear that Theorem 5.8 gives a staggering result: *the Approximating Diffusion process has the very same type oscillations as the Limit System!*

# Chapter 6

# Numerical simulations: detecting oscillations

Throughout this chapter we focus on checking what has been predicted in the previous sections, via performing two different sets of simulations:

- In the first set, we fix the total number $N$ of neurons and look at the behaviour, as time $t$ increases, of both the limit system described by *MFC* and the approximating diffusion process *ADP*, in order to compare them.

- In the second set, we further investigate the behaviour of *ADP*: the objective is to convince the reader that it follows the oscillatory trend of the limit system *MFC*. To this aim, we perform new simulations of *ADP*, progressively increasing the total number of neurons $N$: we look at this situation at longer times with respect to the first set of studies.

## 6.1 General setting of the parameters

We must set all the needed parameters and quantities to perform our simulations. Since the aim is, beyond checking how the total number $N$ of neurons in the model affects the approximating diffusion, comparing the behaviours of *MFC* and *ADP*, we proceed reminding the reader the form of the two systems; indeed, recall that the limit system is described by the following monotone cyclic (negative) feedback system:

$$\begin{cases} \frac{dx_t^{k,l}}{dt} = -\nu_k x_t^{k,l} + x_t^{k,l+1}, \quad 0 \leq l < \eta_k \\ \frac{dx_t^{k,\eta_k}}{dt} = -\nu_k x_t^{k,\eta_k} + c_k f_{k+1}(x_t^{k+1,0}), \quad l = \eta_k \\ x_0^{k,l} = 0 \end{cases} \tag{MFC}$$

While the approximating diffusion process $Y^N$ follows

$$\begin{cases} dY^N_{k,l}(t) = [-\nu_k Y^N_{k,l}(t) + Y^N_{k,l+1}(t)]dt, \quad 0 \leq l < \eta_k \\ dY^N_{k,\eta_k} = -\nu_k Y^N_{k,\eta_k}(t)dt + c_k f_{k+1}(Y^N_{k+1,0}(t))dt + c_k \frac{\sqrt{f_{k+1}(Y^N_{k+1,0})(t)}}{\sqrt{N_{k+1}}} dB_{k+1}(t) \end{cases} \tag{ADP}$$

In order to be consistent with the fixed population size study performed in Section 5.3 and for an easier handling of the whole simulations, we study the case in which the total number of interacting populations is $n = 2$, in a context of negative feedback, $\rho < 0$. This means that, without loss of generality, having $n = 2$ and needing $c_1 c_2 < 0$, we can take $c_1 = -1$ (meaning the first population in object is inhibitory), $c_2 = 1$ (that is, the second one is excitatory).

Moreover, to fall into the hypotheses we have assumed throughout the whole discussion, we must set the spiking rates $f_k$, for $k = 1, 2$, to be bounded, Lipschitz and strictly increasing functions: to this aim, we take the followings:

$$f_1(x) = \begin{cases} 10e^x & \text{if } x < log(20) \\ \frac{400}{1+400e^{-2x}} & \text{if } x \geq log(20) \end{cases} \tag{$f_1$}$$

and,

$$f_2(x) = \begin{cases} e^x & \text{if } x < log(20) \\ \frac{40}{1+400e^{-2x}} & \text{if } x \geq log(20) \end{cases} \tag{$f_2$}$$

To conclude with the settings, we recall that the two systems we are dealing with live in a state space of dimension $\tau = n + \sum_{k=1}^{n} \eta_k$, that is, in our case, $\tau = 2 + \eta_1 + \eta_2$. Hence, in the three different sets of upcoming simulations, it is important to choose and specify the values for the parameters $\eta_1, \eta_2 \in \mathbb{N}_0$ (related to delay/memory) and $\nu_1, \nu_2 > 0$. The complete choice of all parameter values is then allowing us, at each step, to specify the systems *MFC* and *ADP* to the case in object, for $n = 2$.

It is also important to point out that, in order not to unbalance the model between the two interacting populations, it is assumed that $N_1 = N_2$, hence $N = 2 \cdot N_1$, throughout all the sets of simulations.

## 6.2 Sets of numerical simulations

We are now ready to deepen our study and look in detail at the following different sets of simulations.

### 6.2.1 First set: *Approximating Diffusion Process* vs *Monotone Cyclic Feedback system*, fixing the total number of neurons

In this set of simulations we fix the dimension $\tau$ of the system at $\tau = 7$, via setting $\eta_1 = 3$ and $\eta_2 = 2$, and we put $\nu = \nu_1 = \nu_2 = 1$. The number of neurons is set equal to 20 in each population, so that $N_1 = N_2 = 20$.

**What do we expect**: we want to make evident the oscillatory behaviour of both *MFC* and *ADP*.

Having in mind the above choice of the parameters, we would like to make explicit the value for the coordinates of the equilibrium $x^*$. Solving the *MFC* system, with these choices we obtain that $(x^*)^{1,l}$ for $l = 0, \dots \eta_1 = 3$ equals to $-2.424$; on the other side, we get that $(x^*)^{2,l}$ for $l = 0, \dots \eta_2 = 2$ is 0.885.

Hence, we can proceed with the computation of the quantity $\rho$:

$$\begin{aligned}\rho :&= \prod_{k=1}^{2} c_k f_k'((x^*)^{k,0}) \\ &= -f_1'((x^*)^{1,0})) \cdot f_2'((x^*)^{2,0}) \\ &= -f_1'(-2.424) \cdot f_2'(0.885) \\ &= -2.15\end{aligned}$$

At this point we would like to check if condition (4.12) is verified for such a value of $\rho$. Therefore, recalling the condition in object was

$$|\rho| > \frac{\nu^\tau}{(\cos(\frac{\pi}{\tau}))^\tau}$$

and that we are considering the special case $\tau = 7$, $\nu = \nu_1 = \nu_2 = 1$ we get that, indeed

$$2.15 > \frac{1}{(\cos(\frac{\pi}{7}))^7)} = 2.08$$

Hence, recalling that this condition corresponds to condition (4.10) of Theorem 4.2, we can conclude that the equilibrium $x^*$ is unstable, having at least one but a finite number of periodic orbits; moreover, amongst them there is one orbit which is asymptotically stable.

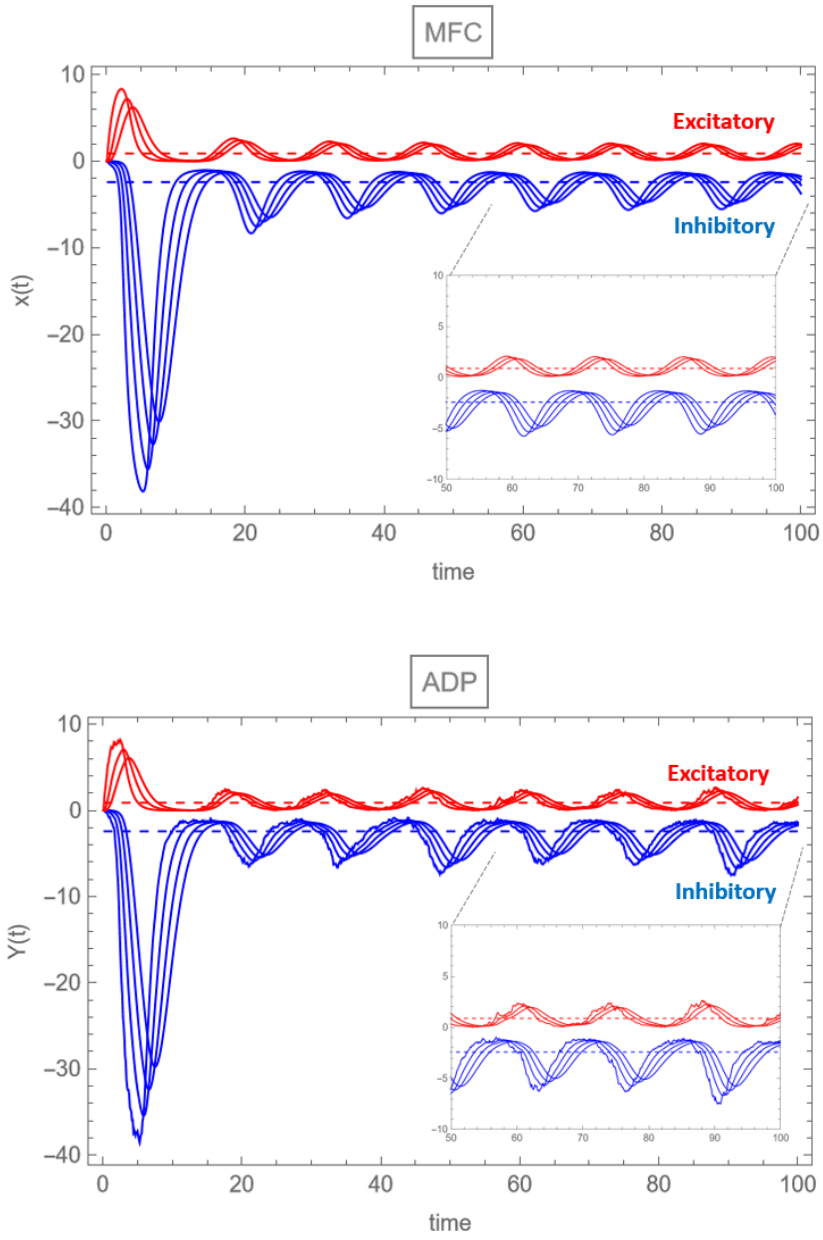



Figure 6.1: *Comparison MFC vs ADP*: it is evident that the two systems follow the very same type of oscillations. The dotted lines are in correspondence of the location of the critical point $x^*$; the inset is a blow-up of the last part of the simulation.

### 6.2.2 Second set: *Approximating Diffusion Process*, increasing the total number of neurons

In this second set of simulations we fix the dimension $\tau$ of the system at $\tau = 7$, via setting $\eta_1 = 3$ and $\eta_2 = 2$, and we put $\nu = \nu_1 = \nu_2 = 1$, exactly as we did in the previous set. We study $ADP$ as the total number $(N)$ of neurons increases, examining four different cases passing from a starting value $N_1 = N_2 = 10, 50, 100, 500$. Each of the following panels of simulations contains 10 realizations of the process $ADP$.

**What do we expect**: if $N$ is small, and with respect to the oscillations concerning the limit system $MFC$, we see $ADP$ oscillates but changing phase randomly; instead, increasing $N$, it should be evident that $ADP$ has the very same type of oscillations as $MFC$ (it is sufficient to look at the first set of simulations to get convinced about that). As $N$ increases, the graphs of the oscillations become even more evident, and are practically superimposable to the ones of $MFC$.

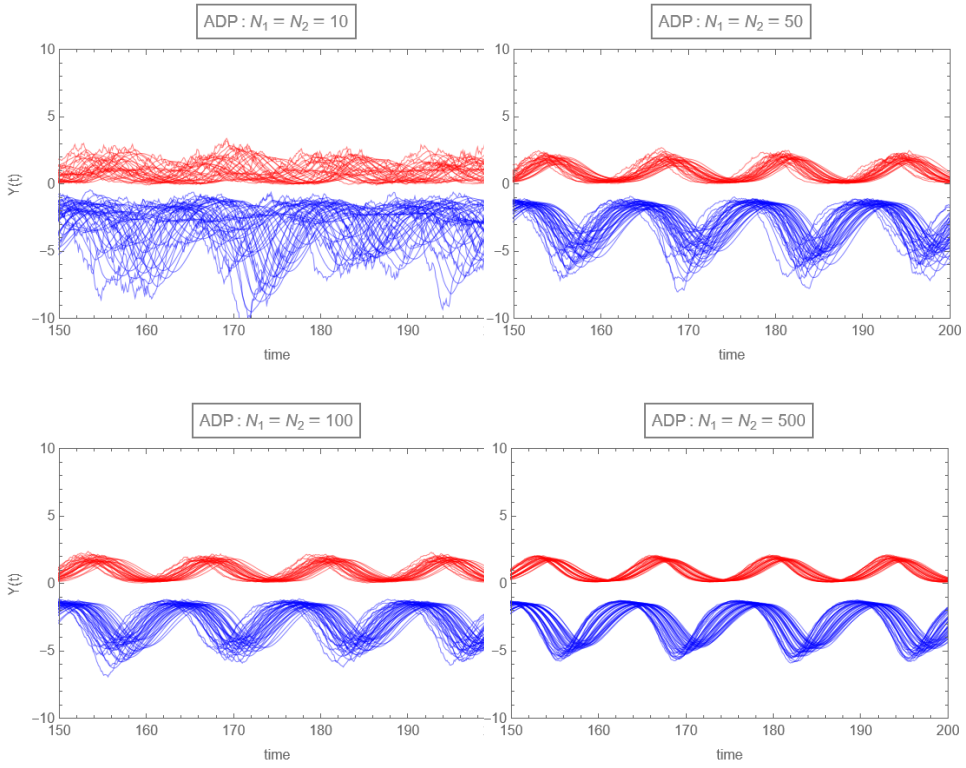



Figure 6.2: The Diffusion Approximation $ADP$, for increasing $N$. Each panel contains 10 realizations of the process.

# Conclusions

In this thesis we studied the emergence of a collective periodic behaviour in a multi-population system of interacting neurons: knowing that periodic patterns and rhythms are pervasive in complex and living systems, we set our study to the purpose of investigating the appearance of oscillations in the collective spiking activity of the neurons at stake in the model.

With this aim, modelling the network via a multivariate system of non-linear Hawkes processes recording the number of spikes of the neurons in a given interval of time, we firstly established Propagation of Chaos and a Central Limit Theorem in the large population regime.

After that -focusing on the limit system- we studied the interaction between classes according to a Monotone Cyclic Feedback System *(MFC)*, considering the memory kernels given in the form of an Erlang distribution: this enabled us, via a dynamical systems approach, to evidence an oscillatory behaviour for the limit system, which possesses attracting, non-constant, periodic orbits.

Then we considered, in association to the original system finite size system, a Piecewise Deterministic Markov Process *(PDMP)* whose dynamics entirely determines the dynamics of the original Hawkes process. Having at disposal this Markovian tool, we hence built an Approximating Diffusion Process *(ADP)* which is shown to be a good approximation of the *(PDMP)*. Moreover, we discovered the tendency -in the large population limit- of both the processes *(PDMP)* and *(ADP)* to the limit system.

Lastly, we investigated the behaviour of the Approximating Diffusion Process in the fixed population size case, for $n = 2$, where $n$ is the total number of interacting populations. The result, after some work, is that the finite size system *(ADP)* is attracted to the same periodic orbit of the limiting system: not only, but also we discovered that, once *(ADP)* enters the basin of attraction of such an orbit, it continues visiting the vicinities of the orbit infinitely often and almost surely. Hence, the diffusion approximation possesses the same kind of oscillations of the limit system; to end up, we have shown these results in the last chapter, the one dedicated to numerical simulations.

To cite the famous poetess Maya Angelou (1928-2014), *"everything in the universe has a rhythm"*: this thesis is just an attempt to make evidence of the fact that apparently complex and chaotic systems can be sometimes reduced to recurrent, periodic and also simple patterns; Mathematics is, as always, the best language to tell such an astonishing story.

# Bibliography


[1] Søren Asmussen, Heinrich Hering, et al. *Branching processes*, volume 3. Springer, 1983.

[2] Luis Barreira and Claudia Valls. *Dynamical systems: An introduction*. Springer Science & Business Media, 2012.

[3] Michel Benaım and Morris W Hirsch. Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior*, 29(1-2):36–72, 1999.

[4] Edoardo Beretta and Dimitri Breda. Discrete or distributed delay? effects on stability of population growth. *Math. Biosci. Eng*, 13(1):19–41, 2016.

[5] Rune W Berg, Aidas Alaburda, and Jørn Hounsgaard. Balanced inhibition and excitation drive spike activity in spinal half-centers. *Science*, 315(5810):390–393, 2007.

[6] Rune W Berg and Susanne Ditlevsen. Synaptic inhibition and excitation estimated via the time constant of membrane potential fluctuations. *Journal of Neurophysiology*, 110(4):1021–1034, 2013.

[7] Rune W Berg, Susanne Ditlevsen, and Jørn Hounsgaard. Intense synaptic activity enhances temporal resolution in spinal motoneurons. *PLoS one*, 3(9):e3218, 2008.

[8] Pierre Brémaud. *Point process calculus in time and space*. Springer, 2020.

[9] Francesco Caravenna. Moto browniano e analisi stocastica, 2011.

[10] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. i. models and methods. *arXiv preprint arXiv:2203.00446*, 2022.

[11] VijaySekhar Chellaboina, Alexander Leonessa, and Wassim M Haddad. Generalized lyapunov and invariant set theorems for nonlinear dynamical systems. In *Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251)*, volume 5, pages 3028–3032. IEEE, 1999.

[12] Kenneth L Cooke and Zvi Grossman. Discrete delay, distributed delay and stability switches. *Journal of Mathematical Analysis and Applications*, 86(2):592–627, 1982.

[13] Kenny S Crump. On systems of renewal equations. *Journal of Mathematical Analysis and Applications*, 30(2):425–434, 1970.

[14] Mark HA Davis. Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):353–376, 1984.

[15] Sylvain Delattre, Nicolas Fournier, and Marc Hoffmann. Hawkes processes on large networks. *Annals of Applied Probability*, 2016.

[16] Frank Den Hollander. Probability theory: The coupling method. *Lecture notes available online (http://websites. math. leidenuniv. nl/probability/lecturenotes/CouplingLectures. pdf)*, 2012.

[17] Susanne Ditlevsen and Eva Löcherbach. Multi-class oscillating systems of interacting neurons. *Stochastic Processes and their Applications*, 127(6):1840–1869, 2017.

[18] Randal Douc, Gersende Fort, and Arnaud Guillin. Subgeometric rates of convergence of f-ergodic strong markov processes. *Stochastic processes and their applications*, 119(3):897–923, 2009.

[19] Douglas Down, Sean P Meyn, and Richard L Tweedie. Exponential and uniform ergodicity of markov processes. *The Annals of Probability*, 23(4):1671–1691, 1995.

[20] Noah Geller. Dynamics in the plane and the poincaré-bendixson theorem. *Available at math.uchicago.edu*, 2021.

[21] Peter W Glynn. Diffusion approximations. *Handbooks in Operations research and management Science*, 2:145–198, 1990.

[22] Carl Graham. Chaoticity for multiclass systems and exchangeability within classes. *Journal of Applied Probability*, 45(4):1196–1203, 2008.

[23] Carl Graham and Philippe Robert. Interacting multi-class transmissions in large stochastic networks. *Annals of Applied Probability*, 2009.

[24] Martin Hairer and Jonathan Mattingly. The strong feller property for singular stochastic pdes. *projecteuclid.org*, 2018.

[25] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[26] Alan G Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of applied probability*, 11(3):493–503, 1974.

[27] Morris W Hirsch. Systems of differential equations which are competitive or cooperative: Iii. competing species. *Nonlinearity*, 1(1):51, 1988.

[28] Reinhard Höpfner, Eva Löcherbach, and Michèle Thieullen. Ergodicity and limit theorems for degenerate diffusions with time periodic drift. application to a stochastic hodgkin- huxley model. *ESAIM: Probability and Statistics*, 20:527–554, 2016.

[29] Kanji Ichihara and Hiroshi Kunita. A classification of the second order degenerate elliptic operators and its probabilistic characterization. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 30(3):235–254, 1974.

[30] Darko Ivanovski. I processi di hawkes e loro applicazioni in finanza. *Thesis and Dissertation Padua Archive*, 2019.

[31] Patrick Jahn, Rune W Berg, Jørn Hounsgaard, and Susanne Ditlevsen. Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *Journal of computational neuroscience*, 31(3):563–579, 2011.

[32] Julian Kern. The skorokhod topologies: What they are and why we should care. *arXiv preprint arXiv:2210.16026*, 2022.

[33] Jicheng Liu and BLS Prakasa Rao. On conditional borel–cantelli lemmas for sequences of random variables. *Journal of Mathematical Analysis and Applications*, 399(1):156–165, 2013.

[34] John Mallet-Paret and Hal Smith. The poincaré-bendixson theorem for monotone cyclic feedback systems. *Journal of Dynamics and Differential Equations*, 2(4):367–421, 1990.

[35] Annie Millet and Marta Sanz-Solé. A simple proof of the support theorem for diffusion processes. In *Séminaire de probabilités XXVIII*, pages 36–48. Springer, 2006.

[36] Mark Nelson and John Rinzel. The hodgkin-huxley model. *Bower, JM and Beeman, editors, The book of Genesis*, pages 27–51, 1995.

[37] Hannes Risken and Hannes Risken. *Fokker-planck equation*. Springer, 1996.

[38] Alain-Sol Sznitman. Topics in propagation of chaos. *Ecole d'été de probabilités de Saint-Flour XIX—1989*, 1464:165–251, 1991.

[39] Hermann Thorisson. Coupling methods in probability theory. *Scandinavian journal of statistics*, pages 159–182, 1995.

[40] Xin Yu, Juliang Yin, and Suiyang Khoo. Generalized lyapunov criteria on finite-time stability of stochastic nonlinear systems. *Automatica*, 107:183–189, 2019.

# Acknowledgements

Desidero innanzitutto ringraziare sinceramente il mio relatore, il Professor Marco Formentin, che mi ha accompagnato in questo percorso con grande professionalità e disponibilità: La ringrazio dal profondo del cuore per tutti i consigli e le indicazioni, per avermi seguito fin dal giorno zero di stesura di questa tesi, e per avermi consigliato in previsione delle scelte di vita future. Lavorare insieme è stato un onore e un piacere, e ha rappresentato per me un grande arricchimento sia dal punto di vista personale, sia dal punto di vista accademico.

Non è facile riordinare le idee arrivati a questo punto: non è retorica se dico che questa avventura è stata -forse- la più difficile e sfidante della mia vita, ma anche la più stimolante. Sono arrivata in facoltà cinque anni fa, carica di passione e di aspettative, e mi sono trovata di fronte, inizialmente, a un mondo nuovo e affascinante, del tutto inaspettato, che però sembrava una montagna quasi impossibile da scalare. I momenti di sconforto, di certo, non sono mancati: a partire da quel pomeriggio del gennaio di ormai quattro anni fa, in cui, seduta sulla poltrona del salotto, ho pensato di mollare tutto e cambiare strada. Ed è qui che il mio ringraziamento più grande va a mamma Laura e papà Fabio, che -spesso anche senza rendersene conto- con le loro parole mi hanno sostenuta, supportata, sopportata, aiutata, rassicurata (giorno e -poveri loro- notte, feriali e festivi), motivata ad andare avanti, consapevoli che con la mia testardaggine, se avessi voluto, ce l'avrei fatta: e ringrazio anche me stessa per aver ascoltato quella testardaggine (e anche quella passione) e non aver mollato, dimostrando a me stessa che certe sfide sono fatte per essere superate. Sarà scontato da dire, ma tutto quello che sono e che ho fatto lo devo a loro: con i loro insegnamenti, la loro voglia di farmi viaggiare e scoprire il mondo, senza dimenticare le mie radici, mi hanno sicuramente trasmesso, tra le altre cose, la passione per la scienza che mi ha portato fin qui. Grazie infinite per la pazienza, perché solo voi sapete trattare e capire le mie debolezze e fragilità, e avete il magico potere di farle fiorire come nessuno saprà mai fare: sono grata per le persone che siete.

Un grazie speciale va anche a mia sorella Irene, che in questi cinque anni è cresciuta tanto, e che con la sua spensieratezza e il suo carattere, così diverso dal mio, mi ha sempre accompagnato, mostrandomi la vita da un altro punto

di vista; siete la famiglia più bella che potessi avere accanto.

Ringrazio i miei nonni, Sandro e Marì, il mio più grande esempio di amore, perché, silenziosamente, sono sempre stati al mio fianco, partecipi e orgogliosi dei miei risultati: cosa non darei per averli qui oggi insieme per festeggiare questo traguardo, che è anche loro.

Giunge ora il momento di aprire il capitolo dei compagni di Università: sono tante le persone che mi hanno accompagnato in questo viaggio, chi fin dall'inizio, chi poi ha deciso di cambiare strada, chi è arrivato in seconda battuta ma ha lasciato in ogni caso il segno.

Non posso che cominciare ringraziando Niccolò e Fulvio, che hanno iniziato con me -che pazzi siamo stati!- questa grande avventura... per passare poi a Irene M. che, presente fin dal primo giorno -con la sua camicetta verde a stelle- ha alleggerito ogni momento con la sua simpatia e serenità, ed è sempre stata in grado di smorzare qualsiasi situazione con il suo sorriso. Ringrazio Sofia, mia compagna di studio, con la quale nei momenti di sconforto ci piaceva ricordare, amareggiate, ridendo: *"Pensa che c'è ancora un altro anno..."* e gli anni sembravano non finire mai... ed invece eccoci qua, arrivate alla fine, con una certa malinconia. E poi Lucia, anima buona, il nostro genio, la nostra guida, senza le cui spiegazioni non avremmo passato più di qualche esame; Caterina ed Erika, senza le cui storie i nostri pranzi non avrebbero avuto senso, Anna D., che mi è sempre stata accanto, con cui ho condiviso tanti dolori ma in fin dei conti anche tante meritate gioie, e infine Marta, la mia romana preferita, che conosco da soli due anni, che però sembrano una vita.

Grazie per le amiche che siete e per quello che mi avete lasciato, spero di non perdervi mai: questi anni non sarebbero stati gli stessi senza di voi; siete state le mie più grandi muse e ancore di salvezza, e ve ne sarò sempre grata.

E come non citare il mitico gruppo allargato dei "Lorenzi": sempre presenti dai tempi del banco di fronte in Luzzatti, disponibili ad aiutare tutti noi, sono stati la compagnia migliore e più simpatica che si potesse sperare per questi cinque anni. Un grazie speciale va a loro, e in particolare ad Alberto, mio compagno di studio dell'ultimo anno, senza il quale sicuramente non sarei qui oggi.

Ultimo ma non meno importante, impossibile non menzionare il nostro Adrian: fin dal primo giorno leader del Dipartimento, con il suo spirito è stato in grado di renderlo casa.

Se penso a quante ne abbiamo passate, anni del Covid a parte, faccio anche fatica a ricordarle tutte: dai pranzi di Natale agli aperitivi, alle pause in Torre, agli esami provati e riprovati, agli appunti scambiati e restituiti, agli scherzi e alle risate e ai momenti in cui tutto sembrava più grande di noi, pare che il tempo sia volato. Mi auguro che in un modo o nell'altro riusciremo

a mantenere i contatti, e che ognuno, dopo questo percorso, trovi la strada migliore per sé.

Un grazie speciale va a Sara, mia spalla, mio mezzo cuore e colonna portante, perché è sempre al mio fianco, c'è stata e spero ci sarà sempre: il tuo sostegno e la tua amicizia sono le cose più preziose che mi potessi regalare fin da quel giorno di agosto di qualche anno fa.

Ringrazio Alessia, entrata da poco nella mia vita, ma che è come ci fosse da sempre.

Ringrazio le mie amiche Giorgia e Dayana, pronte ad ascoltarmi e a sostenermi in ogni momento, e le mie ex-atlete del cuore Carlotta C., Marta, Claudia: mi siete sempre state accanto in questi anni, condividendo con me tutte le emozioni universitarie e non: sono grata di avervi tutte ancora al mio fianco.

Grazie a Martina C., che in questo ultimo anno ha condiviso con me le gioie e i dolori connessi alla fine dell'università, e che, pur conoscendomi da poco, mi ha dato tanto.

Ringrazio Roberta, che, accompagnandomi ormai da anni, mi ha dato la forza di alzarmi dalla scrivania per intraprendere una carriera da studentesse-atlete, e che con la sua allegria contagiosa e spirito combattivo è una persona preziosa da avere accanto.

Ringrazio la mia Eli -detta Sbetz- che, seppur a distanza, è un pilastro importante della mia vita, con me fin dai tempi spensierati delle sculture di sabbia, e con cui ora ci ritroviamo a discutere di futuro e contratti di lavoro: esserci state vicine nei nostri percorsi di vita finora è stato un regalo speciale, e spero che continueremo così per sempre.

Grazie a tutto il gruppo Jesolo, *"23 agosto ferragosto"*: grazie a voi le estati -universitarie e pre-universitarie- mi hanno donato alcuni dei momenti più belli della mia vita. E come non citare i magici Iscampaa, con leader Agostino: siete delle persone stupende e genuine, vi prego di non cambiare mai, e vi ringrazio di essere qui con me a festeggiare.

Un grazie speciale alle mie amiche del liceo, Giada e Alina, e in particolare alla mia Giulietta, con me fin da quando ho deciso di intraprendere questo percorso universitario: siamo cresciute insieme e ci siamo viste diventare donne, ed è bellissimo essere qui insieme oggi.

Un grazie a Fede e Simo, la mia coppia preferita, parte della famiglia, qui oggi per condividere con me la gioia di questo traguardo.

Un grazie inoltre a tutti quelli che non ho citato; grazie a chi -mi scuso immensamente- ho dimenticato, e a chi ha iniziato con me questo viaggio ma ha poi deciso di prendere una strada diversa: avete tutti lasciato un segno e un ricordo indelebile.

E infine grazie anche a chi è entrato nella mia vita da poco, ed in così poco tempo l'ha già sconvolta.

Concludo citando una delle mie canzoni preferite: *"E comunque vada, per quanta strada c'è da fare, amerai il finale . . . "* E di certo amo questo finale, arrivato così in fretta e guadagnato anche con fatica. . . ma ancora di più, nonostante tutto, ho amato la strada che ho percorso, insieme a tutti voi: vi voglio bene.

*Ilaria*