ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

École doctorale n°84 : Sciences et Technologies de l'Information et de la Communication

Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

Spécialité « Informatique temps réel, robotique, automatique »

présentée et soutenue publiquement par

Mathieu Galtier

le 13 décembre 2011

A mathematical approach to unsupervised learning in recurrent neural networks

Directeur de thèse : **Olivier FAUGERAS**

Jury

M. Nicolas BRUNEL, research director, CNRS, Université Paris Descartes	Rapporteur
M. Stephen COOMBES, professor, University of Nottingham	Rapporteur
M. Herbert JAEGER, professor, Jacobs University Bremen	Rapporteur
M. Paul BRESSLOFF, professor, University of Utah	Examinateur
M. Bruno CESSAC, research director, INRIA	Examinateur
M. Olivier FAUGERAS, research director, INRIA / ENS Paris	Examinateur
M. Yves FREGNAC, research director, CNRS	Examinateur
M. Pierre ROUCHON, professor, MinesParisTech	Président

THÈSE

INRIA / ENS Paris
NeuroMathComp project team
23 avenue d'Italie, 75013 PARIS

Remerciements

Je remercie mon directeur de thèse, Olivier Faugeras, pour les conditions idéales dans lesquelles il m'a permis de faire ma thèse: un environnement de travail stimulant, une grande ouverture sur le monde de la recherche et une liberté de recherche grisante. Son approche visionnaire et son implication de tout instant dans le but de marier neurosciences et mathématiques resteront un exemple pour moi.

Je remercie chaleureusement Paul Bressloff qui m'a accueilli à Oxford pendant 8 mois lors de ma thèse. Son savoir, son artillerie technique de physicien, son ouverture scientifique, son humour, son scepticisme exacerbé et son intérêt pour les jeunes en font, a mes yeux, non seulement en grand chercheur mais aussi un ami.

Je remercie Gilles Wainrib et Jonathan Touboul pour avoir rajouter du bruit dans mon travail. Ces deux jeunes chercheurs ont été, non seulement de bons amis, mais aussi de précieux collaborateurs qui m'ont beaucoup appris et guidé pendant ces 3 ans. Nos travaux communs sont loin d'être finis et j'en suis ravi!

Je remercie le jury de ma thèse et plus particulièrement les rapporteurs pour avoir accepté de relire et de commenter ma thèse. Plus précisément merci Nicolas Brunel, Stephen Coombes, Herbert Jaeger, Paul Bressloff, Bruno Cessac, Yves Frégnac, Pierre Rouchon et Olivier Faugeras.

Beaucoup de gens ont contribué plus ou moins indirectement à mes travaux. Sans être exhaustif et en vrac: Geoffroy mon co-bureau qui n'est presque jamais d'accord avec moi, les sophipolitains Greg, Romain, Emilien, Pierre, Marie-Cécile, Joan (etc) avec qui j'ai appris la dure loi du barbecue sur la plage et de la coinche digestive, et les oxfordois Paul, Mark, Jara, Maria, Kostas, Jean-Charles qui m'ont appris a boire de bière en anglais.

Merci aussi à tous les gens qui m'ont simplement aidé à vivre bien pendant et avant la thèse. En particulier ma famille pour avoir fait de moi le teigneux rigolard que je suis et aussi mon groupe de zik (Sax, Diego, Yo, Ratau et Ines) pour la pratique de la rhétorique et de la démocratie. Et puis surtout merci Camille...

Ma thèse à été réalisée dans l'équipe NeuroMathComp a l'INRIA et l'ENS Paris. J'ai été financé par la bourse ERC NerVi # 227747 et la région PACA.

Contents

Chapter 0
Introduction

Brain and neocortex

The brain coordinates perceptions and behaviors of animals through its interactions with the spinal chord and the peripheral nervous system. From the first animals 600 million years ago, the brain has significantly developed and specialized in different parts. Its structure is very complicated and varies greatly with the different species. We are still very far away from understanding the brain of even one particular animal. Therefore, we focus only on one of its parts: the neocortex.

The neocortex is a part of the mammals' brain which is responsible for high-level cognitive tasks from sensory perceptions and generation of conscious motor commands to language and conscious thoughts. It does not mean non-mammals can not have such high-level processes since they have other brain parts that have the same functional roles [Butler and Hodos 1996]. It has been known from the seminal work of Brodmann [Brodmann 1909] that the neocortex is spatially organized in different parts called areas. Each area is devoted to a different kind of information processing, e.g. vision, touch, language, social interactions and even religious faith [Azari et al. 2001]. Anatomically these areas are similar and their specialization is thought to come from their wiring with the rest of the brain. Thus, at first order, the neocortex is more complex and less complicated than the brain. More complex because it process a lot of different types of information with roughly the same architecture. Less complicated because it is not made of anatomically different modules and it is a part of the brain.

The neocortex appeared with the first mammals 200 million years ago. It developed more particularly along the human phylogenetic branch. It can be said that the main feature which makes us different from animals from a neuroanatomical perspective is our large neocortex. The size of the human brain has kept increasing for millions of years (mainly due to an expansion of the neocortex) with a sharp increase 200 thousand years ago when we were homo habilis, see [Lee and Wolpoff 2003]. Our exceptional intellectual and cognitive skills are mainly due to our neocortex: we are comparatively very good at understanding, predicting, communicating and (most probably) thinking. Therefore, our "intelligence" or "humanity" as opposed to animality appeared in the blink of an eye on the timescale of evolution.

A focus on learning / plasticity

The brain and the neocortex have a purpose: give the animal a behavior adapted to its environment. It is a striking fact that the brain, in particular the neocortex, can adapt to so many different situations. Having the ability to learn during their lifetime is a crucial quality for animals.

It is generally believed that learning corresponds to a modification of the connections between neurons [Dayan and Abbott 2001]. Indeed, the neurons in the cortex are densely interconnected with approximately ten thousand connections per neuron. These connections are weighted: if the connection from neuron A to neuron B is strong then the excitation of neuron A will easily propagate to B. Changing the pattern of interconnections of the neurons changes the way information flows in the network: this is learning.

Through learning neurons specialize in some task. Therefore, each neuron in the neocortex is devoted to processing a certain kind of information which is defined by the way it is connected to its neighbors. Hubel and Wiesel won the Nobel prize in 1981 for showing that many neurons in the visual part of the neocortex were selective to small oriented edges in the visual field [Hubel and Wiesel 1962]. These neurons were very excited when an edge of a particular orientation and at a certain position was in the visual field and they were quiet if not. Without going into the details of terminology, it can be said that the emergence of *meaningful concepts* is related to this specialization. In a way learning can be seen as devoting several neurons to process information about the thing or concept being learned. Thus the specialized neurons might be thought to embody the *meaning* of the learned concept. This vision leads to the intuitive (and rigorously wrong) vision of neurons *coding for* a particular concept.

An interesting fact is that the specialization of neurons can change through time. The boundaries between cortical maps are not fixed and a neuron formerly devoted to vision can become associated with audition. This often happens when a young person becomes blind and his visual cortex becomes useless. For instance, it may be reused for audition through learning. A famous experiment on ferrets in [Sur et al. 1999], showed that rewiring the visual information path to the auditory cortex lead to the development of visual features in the auditory cortex. This suggests the animals learned to see through their auditory cortex. Another striking example is the product developed by the company Brainport technology: by plugging the output of a camera to a tactile stimulator to be put on the tongue, blind people roughly learned to see!

This suggests there is a universal learning mechanism that leads to understanding the cortical stimuli whatever they are. There are already a few candidate for the learning rules operating in the brain, see [Caporale and Dan 2008]. These learning rules modify the connections between two neurons solely based on the activity of both neurons. It is not clear how such local rules can lead to a global understanding of the stimuli. This thesis is devoted to looking for this universal mechanism.

A mathematical approach to understand the brain

Historically, the study of the nervous system was a branch of biology. Biologists have been focusing on understanding the mechanisms at the basis of the functioning of the nervous system. For instance, the first (and probably most important) principle of functioning is that information is carried and processed by hundreds of billions of neurons that form the computational substrate of the brain. With such complexity, it was difficult for biologists to go beyond the qualitative description of mechanisms. Recently, there has been two major developments in neuroscience which dragged the field to a more interdisciplinary framework: (i) imaging methods have brought new ways to observe the brain activity, e.g. EEG, MEG, fMRI, optical imaging (ii) computational models have made it possible to have another experimental approach: after experimenting *in vivo* or *in vitro* its is now possible to experiment *in silico*. These two scientific revolutions are in fact extensions of the biological approach and seem appropriate to analyze phenomena at larger scales.

The application of mathematics to neuroscience is another recent axis of research to which this thesis intends to contribute. In a rigorous sense, mathematics can be seen as a language. It consists in building propositions from initial hypotheses following a rigorous grammar. In a practical sense, the mathematical approach may not be exclusively rigorous: a relevant simplification or computer simulation can also shed light onto the conclusions of a mathematical reasoning. The essence of mathematics is the process of drawing (ideally dramatic) conclusions from (ideally weak) hypotheses, yet there may be different tools to do it. The hope of using mathematics to deal with the brain is that it is its natural language, as in physics. In the last century, mathematics have proved to be so relevant in physics that the Nobel prize Eugene Wigner talked about the *the unreasonable effectiveness of mathematics in the natural sciences* (where he meant physics). Indeed, modern physics is based on a very deep and abstract mathematical theory that proved its value by suggesting new experiments and predicting their results. Almost nobody doubts the deep mathematical nature of physics and it had become a strong belief that mathematics will explain biology someday. Yet, it is still a belief.

So far, there is no real evidence of *an unreasonable effectiveness of mathematics in neuroscience*. Although the brain and more generally all living organisms are ultimately physical objects, we do not have a satisfying mathematical understanding of them so far. Is it just a question of time? Is it due to fundamental difference between living and non-living objects which can not be bridged by the mathematic language? These are open questions which drive most of current research in mathematical biology. It is generally considered that the main (and sometimes only) success of mathematics in neuroscience so far dates back to the Nobel prize of Hodgkin and Huxley in 1963. In [Hodgkin and Huxley 1952], they designed a 4-dimensional nonlinear differential system which reproduced accurately the functioning of a neuron. This is qualitatively similar to Newton's discovery about the laws of motion. Would we say this is mathematics? It seems more appropriate to consider the Hogkin-Huxley model as a definition of neuron in the mathematical language. Indeed, their work had a biological purpose: describing the mechanisms governing the behavior of neurons, in mathematical terms. In a way, they defined the hypotheses needed to begin the mathematical work.

What would be a successful mathematical work in neuroscience? First, a success may correspond to the conclusion of a mathematical reasoning saying something true (and unexpected) about the real brain (not the mathematical object). To check the truth of a mathematical claim, it would have to be testable. In this perspective, a good work in mathematical neuroscience would lead to testable predictions as in physics. This approach is descriptive, predictive and well-suited for addressing medical questions for instance.
Second, there may be an intrinsic value in the mathematical objects or propositions derived from biologically inspired hypotheses. Let alone the mathematical beauty of a theory, these objects might reveal a principle of functioning of the brain. This principle might be only said in the mathematical language (as opposed to our current anatomical language). In this approach, there may

be no way back to biology yet there would be a lot of meaning in the theory. Put it bluntly, if a mathematical theory leads to the definition of a new kind of (artificial) intelligence then this would be a success. This approach is generative, theoretical and well-suited for addressing questions in data-analysis and/or robotics for instance.

This thesis mainly takes the second approach: we model and mathematically analyze learning neural networks to show how they can understand their environment. Although we relate our work to some experiments, we face the experimental impossibility to measure the strength of the connections between neurons on a large scale (which is the focus of this thesis).

Problematic and organization of the thesis

It is a common idea that the neocortex builds itself a model of the environment through learning. Actually, it is not more than a reformulation of the notion of understanding: we understand a phenomenon when we are no longer surprised by the way it might evolve. To do this we must have sampled the different possibilities so that we have a global knowledge of what is likely to happen next. In a way, we have built ourself a model of the phenomenon. This is the focus of this thesis and the problematic could be formulated as follow:

Give a mathematical sense to the claim:
the neocortex builds itself a model of the world.

This question is too ambitious to be answered rigorously, yet we think our approach is relevant and we propose an interesting perspective to address this issue.

This thesis is divided in three chapters and organized as follow

In the first chapter, we address the modeling of large population of interconnected neurons with learning. This implies having a look at biology and choosing or designing appropriate mathematical models for the building blocks of a learning neural network, e.g. neurons and synapses. This leads to defining a huge differential system which is too difficult to analyze mainly because of the intrinsic non-linearity of the neurons. Therefore, we propose a new method of spatial averaging to go from this large network of spiking neurons to a rate-based network of populations of neurons. In other words, we assume the neurons are so numerous that we can consider their mean to define a mean-field equation between populations of neurons. The goal is to have a simpler, i.e. more linear, model of a learning neural network.

In the second chapter, we analyze the dynamics of the previously averaged network for different learning rules. We make good use of the slowness of learning compared to the activity of the network. Indeed, considering the learning neural network as a slow/fast system makes it possible to apply temporal averaging tools to reduce it. In some cases, this reduce system appears to derive from an energy so that it always converge to an equilibrium connectivity. This equilibrium connectivity corresponds to the entire knowledge of the network about the external world. We show how it can be explicitly related to the stimuli.

In the third chapter, we claim that the network post-learning is a model of its environment. While former research established that learning in feed-forward networks may lead to the extraction of the principal components of the stimuli, we focus on the recurrent connectivity post-learning and study the way it processes and link these objects together. It mainly consists in transforming the equilibrium connectivity into a rich description of the stimuli. In particular, we address the dynamics of spontaneous activity, see how it relates to the stimuli.

In all chapters, there is both background and original work. The background is always in the beginning of the chapter (sections 1.1, 1.2.1, 2.1, 2.2, 3.1 and 3.2), whereas the original work is at the end (sections 1.2.2, 1.2.3, 2.3, 2.4, 3.3, 3.4).

Chapter 1
From spiking neurons to rate-based networks

Overview

This chapter has two goals: first, introducing some background biology to model spiking learning neural networks and, second, developing an original method to average these spiking networks into a rate-based equation.

The necessary biological background mainly consists in describing the building blocks of neural networks: neurons and synapses. Then, we motivate the choice of certain models and gather all these elements in a single system which is the starting point of our analysis. This is a spiking system.

After having introduced the usual motivations to consider rate-based equations, we propose a new method to get a mean-field description of this spiking network. Considering that the neurons belong to populations, the number of neurons per populations tends to infinity and averaging over the neurons in each population, we finally get an equation describing the smooth evolution of the averaged firing-rate in each population. This is a rate-based system derived from the initial spiking network.

In this chapter sections 1.1 and 1.2.1 are background, whereas sections 1.2.2 and 1.2.3 are original.

Résumé

Ce chapitre a deux buts: permièrement, il introduit les connaissances de bases pour modéliser les réseaux de neurones à potentiel d'action et, deuxièment, il présente une méthode originale pour moyeniser ces réseaux à potentiel d'action afin d'obtenir des réseaux à taux de décharge.

Les connaissances de bases en biologie nécessaires portent sur les différentes briques de bases des réseaux neurones: les neurones et les synapses. Parmis la diversité de modèles décrivant ces éléments nous en choisissons certains pour les rassembler dans un unique système qui est le point de départ de la thèse. Ce système est un réseau de neurones à potentiel d'action.

Après avoir motivé l'utilisation de modèles à taux de décharge, nous proposons une nouvelle méthode pour obtenir une decription en champ moyen du réseau de neurone à potentiel d'action. Pour ce faire nous supposons que les neurones sont rassemblés en populations dont le nombre de neurones tend vers l'infini. En moyenant les variables dans chaque population, nous obtenons une equation décrivant lévolution du taux de décharge moyen dans chaque population. Ce système est réseau de neurones à taux de décharge dérivé du réseau initial à potentiel d'action.

Dans ce chapitre les sections 1.1 et 1.2.1 sont issues de la littérature, alors que les sections 1.2.2 et 1.2.3 sont originales.

Collaborations, publications and ackowledgements

This part originates from numerous discussions with Jonathan Touboul, Geoffroy Hermann and Olivier Faugeras who are focusing on a rigorous development of a mean field approach. In particular, the mean field analysis we develop in part 1.2.2 is based on the main theorem of the recent papers [Touboul 2011, Baladron et al. 2011]. It will be extended to more general neurons in collaboration with Jonathan Touboul and submitted after the thesis.

Contents

1.1 From biology to microscopic models

There are few biological facts that we need to mention before modeling a cortical network. The purpose of this section is to give a big picture of the biological substrate and to introduce the assumptions that are necessary to get tractable models.

After a few general fact about the neocortex (which we call cortex for simplicity), we introduce the neurons and their models. Then, we focus on the synapse as a signal transmitter but also as the location where some learning mechanisms occur. Finally, we gather all these microscopic models into a large network equation which will be the starting point of the subsequent analysis.

1.1.1 Cortex

The cortex is a thin sheet of neural tissue located all around the mammalian brain. It corresponds to the folded shape we see in most brain illustrations. It is said to be the locus of most of the high-level processes taking place in the brain linked for instance with awareness, language, thought, senses, actions. Most of the cortex is made up of six density layers which are labeled from I to VI. It is not well understood yet what kind of processing is happening between these layers. For simplicity, we will not take them into account. With this considerable assumptions, the cortex can be seen as a 2-dimensional sheet of neural tissue.

The cortex is mainly made of 2 different kind of cells: neurons and glial cells. There are a very large number of these cells: an order of magnitude of 10^{10} neurons and 4 times as much glial cells. The former are said to be the principal vectors of information. The latter are generally though to be responsible for providing energy to neurons. This is why glial cells are almost always neglected in models of neural tissue. Being interested in the way the cortex processes information, we will completely neglect them too.

Neurons have the interesting property to be able to send signals over long distances to a huge number of other neurons (approximately 10^5 per neuron). Information is sent in an electrical form from the neuron's cell-body called soma, through its axon, to a synapse where the electrical signal is turned into a chemical one. Then, the signal is transmitted to a dendrite which carries an electrical signal to the soma of another neuron. Therefore, there are two building blocks in the design of simple neural networks: neurons and synapses.

Neurons and synapses come in many different shape and size. Even there functional behavior may vary from one to the other. In this work, we completely neglect these disparities and focus on large networks of similar neurons and synapses. We consider that the neurons can be excitatory and/or inhibitory at the same time. In particular, we do not restrict our approach to networks of excitatory neurons coupled with inhibitory neurons.

In conclusion, the brain (and even the cortex) is an extraordinarily complicated system which is well beyond our current mathematical techniques. To use the mathematical language, we must step away from biology to define simplistic objects whose biological relevance is weak. This is clearly a drawback of the use of mathematics to get biological insight.

1.1.2 Neurons

1.1.2.1 What is a neuron?

Biological mechanisms The main biological mechanisms involved in the neurons' behavior are well-understood and comprehensively discussed in [Hille 1992]. A pedagogical introduction for theoretical people is [Izhikevich 2007]. Actually, a neuron is a particular type of cell, endowed with a nucleus, whose membrane is excitable and which transmits information electrically or through synaptic realease. The neuron lies in a solution where there are ions of different types (e.g. K^+, Na^+ and Ca^{2+}). The neuron's membrane is porous and the pores are called ion-channels; they are selective to the ions type. The ions concentrations are different on the inside and the outside of a neuron. As shown in figure 1.1, this leads to two opposed forces that drive the ions through the membrane channels: electric potential gradient and diffusion gradient. The electric potential difference between the inside of the neuron and the outside is called the membrane potential. Without external perturbation, the neuron reaches an equilibrium called the resting state (where resting membrane potential is $-70\ mV$ or so for vertebrate brains).

The proportions of opened ion-channels non-linearly depend on the membrane potential, such that a sufficiently strong perturbation of the membrane potential (e.g. due to another neuron) may dramatically but temporarily change the value of the system variables. This corresponds to the generation of spike. A simple view is that neurons have a spike initiation threshold which

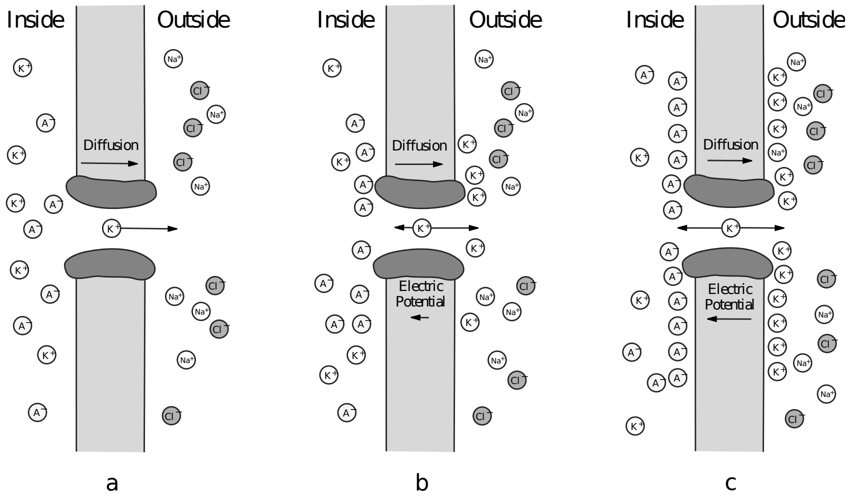


Figure 1.1: This represents the forces participating to the functioning of a potassium ion channel. Figure a shows that a neuron has a larger concentration of ions K^+ and A^- (a generic name for a negatively charged ion). Therefore, this induces a diffusion forces that tend to push potassium ions outside. Figure b shows the emergence of an electric potential because the outside becomes more positively charged than the inside. Eventually this leads to an equilibrium where the forces cancel out as shown in figure c. Taken from [Izhikevich 2007].

has to be exceeded for a spike to be generated.

Functional description A neuron exhibits a spiking behavior, i.e. a small perturbation can generate a transient non-linear amplification of the membrane potential. These spikes or action potentials have a particular shape, as shown in figure 1.2 which is robustly reproduced for each stimulation. Indeed, neurons process information in an all-or-nothing fashion (almost digital). Therefore, the strength of the stimulation can be seen in the intensity (would it be high frequency or precise timing) of the spike trains only. Yet, the neurons can not spike infinitely fast because a spike is always followed by a refractory period which is of fixed duration. The spiking frequency saturates at an order of magnitude of 10^3 Hz.

Mathematical description In this thesis, a neuron is defined by a dynamical system whose main variable is the membrane potential v. There might be

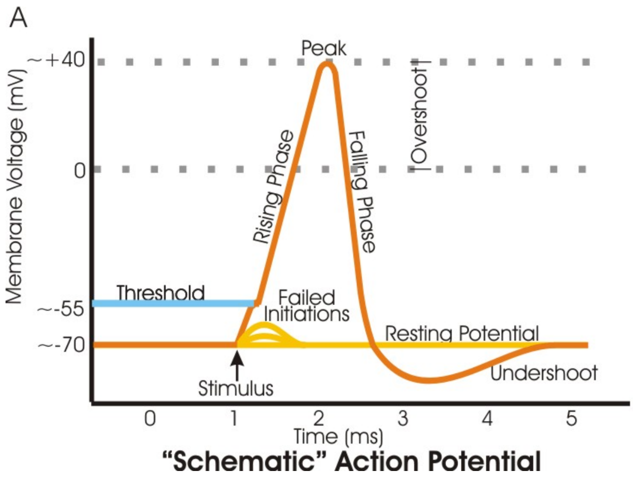


Figure 1.2: Typical shape of an action potential. Taken from Wikipedia.

other variables to describe its internal state (e.g. proportion of open ion-channels) and its dynamics is non-linear.

This is a different approach to the input-output vision of the neurons usually shared in the field of artificial Neural Networks.

1.1.2.2 Neuron models

There are many different models of neurons. Some of them only focus on reproducing the functional behavior while others care about the the biological representativity of the variables. Other models can be continuous or analogous (even linear in some cases) and have no spikes; we argue in the following that they are to be considered as neurons' population models.

Functional models are numerous and are not in the scope of this thesis, therefore we will just briefly review this part of the field. The simplest of these models is the integrate and fire neurons where the dynamics of the neuron is linear and stimulated by some inputs which increase the excitation of the neuron. When the neuron's membrane potential reaches a certain threshold, it is considered that the neuron initiates a spike and is then reseted to a resting value. This the basic mechanism of a hybrid neuron. Recently, a lot of work has been devoted to designing computationally efficient hybrid neurons that would mimic real neurons at best, e.g. [Brette and Gerstner 2005, Izhikevich 2003, Touboul 2009]. Although, not directly linked with biology, they prove to be relatively efficient in reproducing the behavior of a single neuron. However, the mathematical analysis of networks of such neurons is quite difficult [Tsodyks and Sejnowski 1995, Hansel et al. 1995, Gerstner 1995, Brunel 2000]. We believe the hybrid formalism is computationally easy but mathematically very hard to handle.

This is why, we will focus on other models which are built upon biological mechanisms. They are called conductance based models. They are intrinsically non-linear but their dynamics is continuous, bounded and does not require to be artificially reseted. Bifurcation theory seems to be a good language to address their dynamics which is well understood for isolated neurons. Yet, these models still pose significant mathematical difficulties when coupled together.

We start by sketching very briefly the mechanisms involved in the neuron's functioning and then introduce one the most complete and probably the most famous neuron model: the Hodgkin-Huxley model. We will gradually simplify the complexity of this mathematically intractable system trying to keep the main ingredients to stay close to the qualitative behavior of the neurons. We will end this partial review of neuron models by introducing the McKean model which is quite simple and will be the building block of large networks we consider in the rest of the chapter. The following description is based on [Ermentrout and Terman 2010] and [Izhikevich 2007].

Hodgkin-Huxley model [Hodgkin and Huxley 1952]

This model implements the preceding biological mechanisms into a differential system. The approach of Hodgkin and Huxley, which won them the Nobel prize for physiology and medicine, consisted in using new experimental methods on giant squid neurons to derive a mathematical system explaining the spikes generation and the dynamical features of the real neuron. This model corresponds to neurons with potassium and sodium channels and with a leak current. It describes the evolution of the membrane potential v together with 3 (in)activation variables n, m and h, such that n^4 (resp. m^3h) is the proportion of opened channels for the ions K^+ (resp. Na^+). This conductance based models can be written

$$\begin{cases} C\dot{v} &= I - \bar{g}_K n^4(v - E_K) - \bar{g}_{N_a} m^3 h(v - E_{N_a}) - \bar{g}_L(v - E_L) \\ \dot{n} &= \frac{n_\infty(v) - n}{\tau_n(v)} \\ \dot{m} &= \frac{m_\infty(v) - m}{\tau_m(v)} \\ \dot{h} &= \frac{h_\infty(v) - h}{\tau_h(v)} \end{cases} \tag{1.1}$$

where the membrane capacitance is $C = 1\ \mu F/cm^2$, the (shifted) Nernst potentials are $E_K = -12\ mV$, $E_{N_a} = 120\ mV$ and $E_L = 10.6\ mV$. The values of the maximal conductance are $\bar{g}_K = 36\ mS/cm^2$, $\bar{g}_{N_a} = 120\ mS/cm^2$ and $\bar{g}_L = 0.3\ mS/cm^2$. I is the external current in $\mu A/cm^2$. The profiles of the functions n_∞, m_∞, h_∞, τ_n, τ_m and τ_h are shown in figure 1.3.

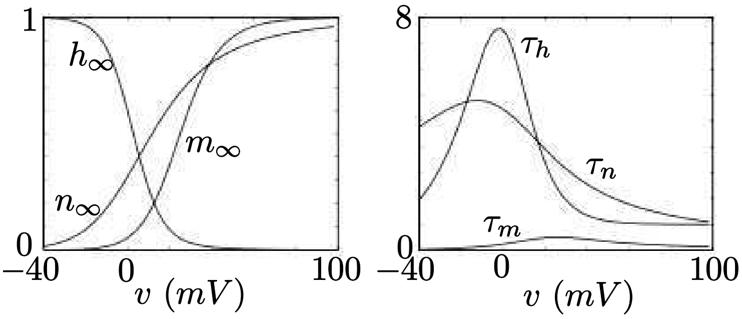


Figure 1.3: (left) Steady state (in)activation functions. (right) Voltage-dependent time constants for the Hodgkin-Huxley model. Illustration taken from [Izhikevich 2007].

This model's main drawback is its level of complication. Therefore, it is natural to attempt to reduce the Hodgkin-Huxley model to a 2-variable model. This would make possible to have a qualitative and geometric approach of the dynamics through the phase plane of the system. First, observe that τ_m is much smaller that both τ_n and τ_h. Therefore, it can be assumed that $m(t)$ immediately converges to its equilibrium value $m_\infty\big(v(t)\big)$. Replacing $m(t)$ by $m_\infty(v)$ in system (1.1) reduces the model by one dimension. Second, to rule out another variable we need to be able to compare them. Therefore, we change variable and for $x = n, h$ define v_x such that $x = x_\infty(v_x)$. v_x is called the equivalent potential of the (in)activation variable x. Numerically, we can observe that v_h and v_n are actually very close to each other. Therefore, although arbitrary, it sounds reasonable to set $n = n_\infty(v_h)$ which eliminates the variable v_n and lead to the following reduced model

$$\left\{\begin{array}{ll} C\dot{v} & = I - \bar{g}_K n_\infty^4(v_h)(V - E_K) - \bar{g}_{N_a} m_\infty^3(v) h_\infty(v_h)(v - E_{N_a}) - g_L(v - E_L) \\ \dot{v_h} & = \frac{h_\infty(v) - h_\infty(v_h)}{\tau_h(v) h'_\infty(v_h)} \end{array}\right. \tag{1.2}$$

We can geometrically analyze the dynamics by drawing the phase plane of this system as shown in figure 1.4.a. A convenient aspect of the equivalent potential method is that the v_h-nullcline is $v = v_h$. The cubic shape of the V nullcline is responsible for the spiking behavior of the neuron and is common to most (non-hybrid) spiking neuron models. In fact, it is the qualitative feature we are going to extract to build a simpler but mathematically tractable model.

Morris-Lecar model [Morris and Lecar 1981]

This is a simple model of spike production based on a neuron with potassium and calcium channel with a leak current. In this model, the calcium current depends instantaneously on the voltage. This system also belongs to the class of the conductance based models and can be written

$$\begin{cases} C\dot{v} &= I - \bar{g}_K n(v - E_K) - \bar{g}_{C_a} m_\infty(V)(v - E_{C_a}) - g_L(v - E_L) \\ \dot{n} &= \phi \frac{n_\infty(v) - n}{\tau_n(v)} \end{cases} \tag{1.3}$$

where the functions m_∞, n_∞ and τ_n have the same shape as in figure 1.3. The parameters, controlling for instance the offset of the steepness of the sigmoidal functions m_∞ and n_∞, have to be tuned with experiments depending on the particular neuron to be modeled.

The phase plane of this model is shown in figure 1.4.b. It is made of a cubic nullcline and a strictly increasing nullcline as in the reduced Hodgkin-Huxley model.

Fitzhugh-Nagumo model [Fitzhugh 1961, Nagumo et al. 1962]

This is an idealized model based on the previous observations that both the reduced Hodgkin-Huxley and the Morris-Lecar models have a cubic nullcline and a strictly increasing nullcline. It is made of two variable: v is to be seen as the membrane potential, w is called the adaptation variable. It has the form

$$\begin{cases} \dot{v} &= I + f(v) - w \\ \dot{w} &= \varepsilon_w(v - bw) \end{cases} \tag{1.4}$$

where f is a polynomial of the third degree, e.g. $f(v) = v - v^3$, and $\varepsilon_w, b \in \mathbb{R}_+$. It is often assumed that $\varepsilon_w \ll 1$. The phase plane of this system is shown in figure 1.4.c.

It is a famous model in the study of excitability and it has been applied to many different systems from neurons to heart. Actually, it is very close the van der Pol oscillator where the monotonic nullcline becomes a vertical line.

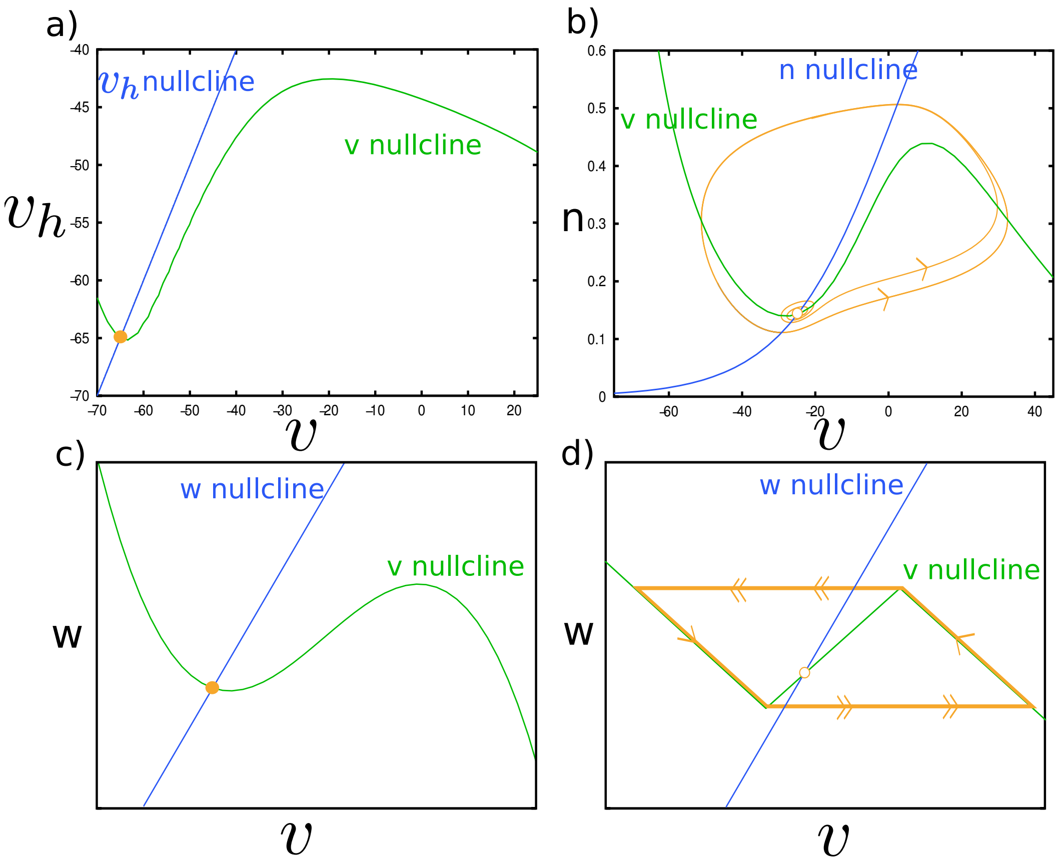


Figure 1.4: These figures represent the phase planes of four different neuron models. In all cases, the input acts as a vertical shifting of the cubic green curves. Therefore, it can be seen as a bifurcation parameter that can generate oscillations (spikes). Figures a and c correspond to a weak input that does not change the stability of the resting state. In figures b and d, the input is large enough to that the equilibrium point is unstable and there is a stable limit cycle. a) The reduced Hodgkin-Huxley model, see (1.2), for $I = 0$ and the parameters specified above. Adapted from [Ermentrout and Terman 2010]. b) The Morris-Lecar model, see (1.3) for $I = 100$. The other parameters are chosen so that the neuron is close to a Hopf bifurcation, see table 3.1 of [Ermentrout and Terman 2010] from which this figure is adapted. c) The Fitzhugh-Nagumo model, see (1.4). The input is below the neuron's threshold. d) The McKean model, see (1.5). The input is larger than the natural threshold. All these neuron's models share the same qualitative dynamics.

McKean model [McKean 1970, Tonnelier 2007]
This is a caricature of the Fitzhugh-Nagumo model. Actually, it keeps the notion of cubic shape for the nullcline of the membrane potential and approximates this shape by a piecewise linear function as shown in figure 1.4.d. Therefore, it is not better nor worse than the Fitzhugh-Nagumo model which is related to the conductance-based models in the same heuristic fashion. It is governed by system (1.4) with

$$f(v) = \begin{cases} -lv - (l+c)a & \text{if } v \leq -a \\ cv & \text{if } a < v < a \\ -lv + (l+c)a & \text{if } a \leq v \end{cases} \tag{1.5}$$

with $l, c, a \in \mathbb{R}_+$. To get oscillatory behaviors, we must set $bc < 1$.

This neuron model is mathematically much more simple to handle than the previous ones and it keeps the main feature of its dynamics. This why we choose it and later analyze its interaction with others.

1.1.2.3 A relaxation oscillator in a noisy environment

As shown in figure 1.4, these four neuron models share the same qualitative dynamics: they are relaxation oscillators or excitable systems. Indeed, the strength of the external input I, which corresponds to vertically translating the cubic curve, determines the dynamical regime of the neurons. If it small enough for the v-nullcline (in blue) to cut the cubic curve on its decreasing negative part then there is a stable fixed point represented by a plain orange disk in figures 1.4.a and 1.4.c. If the two nullclines intersect when the v-nullcline is increasing then the fixed point is unstable and there is a limit cycle along the branches of the cubic curve as shown in figures 1.4.b and 1.4.d. If it is so large that the nullclines intersect in the decreasing positive part of the cubic then there is also a stable fixed point. However, this last situation is not relevant to the normal functioning of a biological neural network.

It is usually assumed that the adaptation variable is slow compared to the membrane potential. This implies that the horizontal part of the vector field in the phase planes figure 1.4 is larger than the vertical part. In other words, in the oscillatory regime the neuron virtually "jumps" from one slow branch to the other giving the membrane potential evolution really sharp transitions from an up state to a down (or resting) state and reciprocally.

The time spent on each slow branch depends on its distance to the v-nullcline (in blue): the closer to the blue curve the slower. It means that, in the setting of equation (1.4), having a positive external input I leads to longer up-states than down-states as opposed to the behavior of a usual spike as shown in figure 1.2. Therefore, this regime is nor biologically plausible. And it makes sense to consider the absence of stimulation to the neuron corresponds to a negative input $-I_0$. A stimulation would correspond to a (still negative input) of $I - I_0$.

It is not clear what the value of I_0 and the range of I should be. In the deterministic case and for McKean's neuron, it seems reasonable to choose $I_0 \leq I^* = -(1-c)a$ so that a neuron without stimulation would be in a resting state. The maximal stimulation should be larger than $I^* - I_0$ so that the neuron is not always quiet. In a noisy environment the picture is a big more complicated.

Indeed, we choose to add noise only on the fast variable which corresponds to assuming that synaptic noise, which acts on the fast variable, is much stronger than the channel noise, which acts on the adaptation variable. This is also mathematically useful. Together with the previous remarks this leads to a new system with a stochastic part

$$\begin{cases} dv &= \big(I - I_0 + f(v) - w\big)dt + \sigma dB(t) \\ \dot{w} &= \varepsilon_w(v - bw) \end{cases} \tag{1.6}$$

where $\sigma \in \mathbb{R}_+$ and $B(t)$ is Brownian motion.

This system is a slow/fast excitable system with additive noise on the fast variable. This corresponds exactly to the framework of [Muratov et al. 2005]. In this paper, Muratov and colleagues suggest that when the time scale $\tau_w = \frac{1}{\varepsilon_w}$ is very large the effect of the noise can induce periodic behavior of an excitable system even if the noise is asymptotically small and the deterministic system is a priori on a stable equilibrium point. This is called "self induced stochastic resonance". Indeed, the barrier of potential to escape the slow manifold for a constant input such that $I - I_0 < -(1-c)a$, i.e. in the stable resting state, is $\Delta(I) = \frac{1+\frac{l}{c}}{1+l}\big(-I-(1-c)a\big)$. When the barrier is equal to $\sigma ln(\tau_w)$ (in other words $e^{-\frac{\Delta(\mathbf{x}_\alpha^*)}{\sigma}} \sim \varepsilon_w$), the noise has enough time to sample most of its distribution so that it almost always overcome the potential difference. Therefore, we define Muratov's critical input I^* corresponding to a probability one of spiking (even though $-I_0 < -(1-c)a$). It is

$$I^* = -(1-c)a - \frac{1+l}{1+\frac{l}{c}} \sigma ln(\tau_w) \tag{1.7}$$

When the effective input $I - I_0$ is below this critical value the neuron rarely spikes. It is also shown in [Muratov et al. 2005] that the spike times follow a Poisson distribution which is a common observation for resting cortical tissues. However, when the effective input is equal or larger than this critical value, the neuron fires almost periodically even if the nullclines of the deterministic system intersect on a stable equilibrium point. This behaviors are illustrated in figure 1.5.

With noise, excitable systems show new regimes that were not present in the deterministic case. As a consequence the choice of I_0 and the range of I may be very different than in the deterministic case. Indeed, $I - I_0$ can be always smaller than the deterministic critical value $-(1-c)a$ and the system would still exhibit a periodic behavior. In these systems noise has a important functional impact on the dynamics.

1.1.3 Synapses

Synapses link neurons together. They are parts of neurons that have specialized into transmitting chemical signals from the axon of pre-synaptic neuron to a dendrite of the post-synaptic neuron. They are of crucial importance in the network functioning since it is widely believed that the learning mechanisms occurring in the brain mostly affect the synapses' strength. However, the learning mechanisms (being central in our approach) will be described in the next section. This section is devoted to analyzing the synapse as a signal transmitter.

First, we introduce the synapse from a biological, functional but also mathematical point of view. Then we focus on modeling the transformation of the signal through this chemical channel. This part is also significantly based on [Ermentrout and Terman 2010].

1.1.3.1 What is a synapse?

Biological mechanisms A synapse is the connection from one pre-synaptic neuron to a post-synaptic neuron. In this thesis we focus on the most common

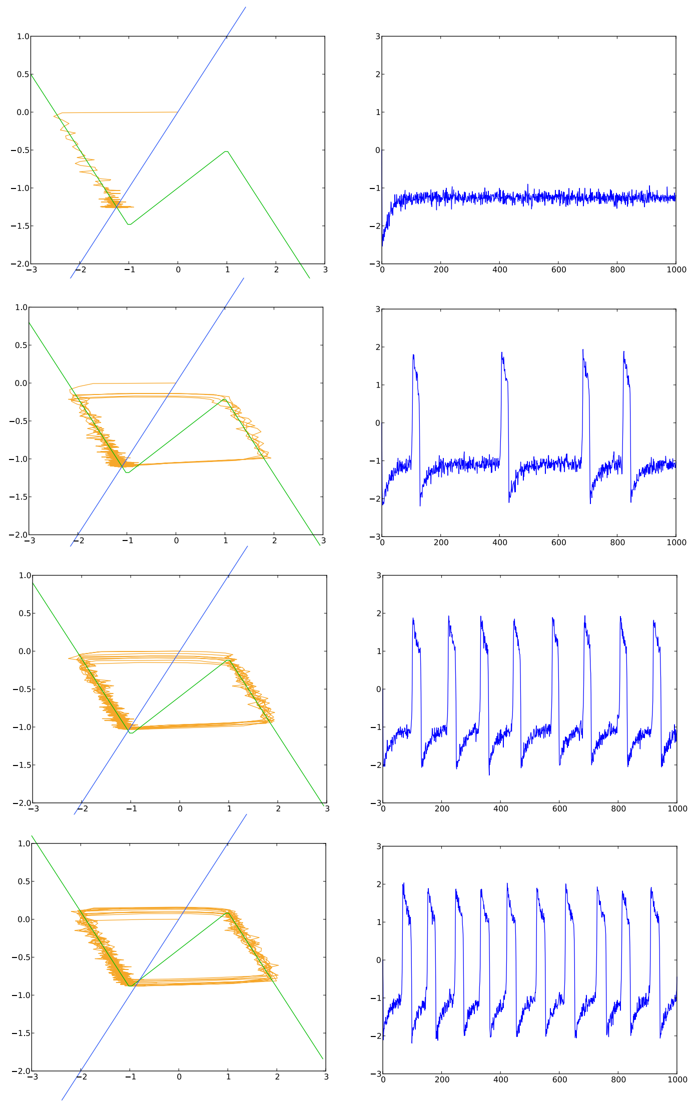


Figure 1.5: These figures correspond to the trajectories of a single noise-driven McKean neuron with inputs of increasing strength. The left column corresponds to the trajectories on the phase plane. The right column correspond to the time course of the membrane potential. The parameters used for these simulations are $a = b = l = 1$, $c = 0.5$, $\sigma = 0.1$, and $\varepsilon_w = 0.01$. a) $I - I_0 = -1 < I^*$. Below Muratov's critical input their membrane potential is almost always close to its resting state. b) $I - I_0 = -0.7$. The neuron fires in an irregular fashion. It is suggested in [Muratov et al. 2005] that this corresponds to a Poisson spike train. c) $I - I_0 = -0.6$. The neuron seems to be periodically spiking although the deterministic part of the system would converge to a stable equilibrium. The noise induces periodic oscillations far from a bifurcation. This is "self induced stochastic resonance". d) $I - I_0 = -0.4 > -(1 - c)a$. The neuron is regularly spiking. Even the deterministic part of the system would lead to oscillations. This regime may been thought of as a bursting regime.

type of synapse: chemical synapses. Chemical synapses are made of a pre-synaptic axon terminal and the closest part of the post-synaptic dendrite. These two parts are separated by a synaptic cleft. Figure 1.6 illustrates the activation of a synapse.

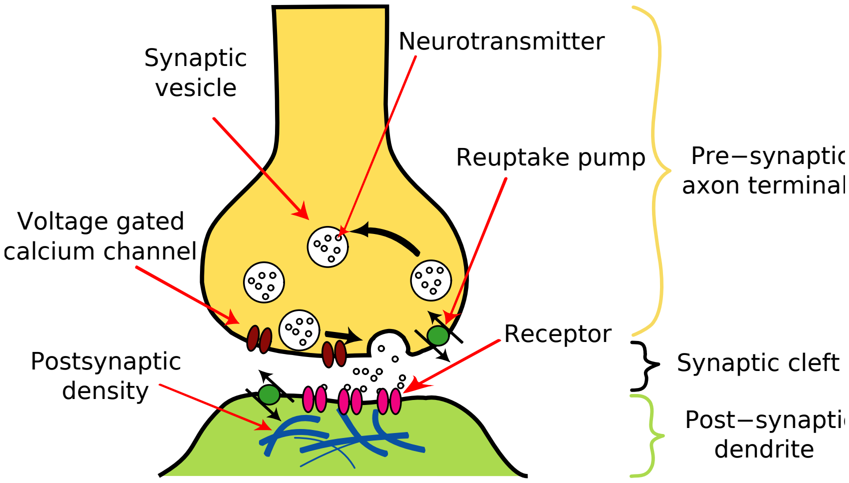


Figure 1.6: Drawing of a synapse with its difference elements. Taken from Wikipedia.

There are several steps which lead to the activation of a synapse. First, an action potential reaches the axon terminal of the pre-synaptic neuron. There are voltage-gated calcium channels which therefore release calcium in the axon. Then, the calcium activates some proteins which are located at the surface of vesicles full of neurotransmitters (the chemical messengers). These proteins change shape allowing the vesicle to fuse with the membrane of the axon, so that the neurotransmitters are dumped in the synaptic cleft. Some of them are then captured by the dendrite receptors, which activates them. This has the effect of opening ion-channels in the post-synaptic membrane which cause the injection of a positive or negative current depending on the nature of the neurotransmitter. Naturally, the more the receptors the stronger the connection between the two neurons. The neurotransmitters are then reabsorbed by the pre-synaptic cell or they may drift away by thermal shaking to be eventually metabolically broken down.

Actually, there are a lot of different transmitters, even on a single synapse, which have very different roles. The main transmitters associated with cortical neurons are glutamate and c-aminobutyric acid (GABA), which can be thought of as excitatory and inhibitory messengers, respectively. There are different kind of receptors (even for the same neurotransmitter). Besides, the generation of a post-synaptic potential, performed by the post-synaptic density, is complicated and varies significantly. For instance the glutamate may be captured by AMPA/Kainate receptors which are very fast or by NMDA receptors which have long lasting effects. Similarly, GABA receptors can be fast (e.g. GABA_A) or slow and long-lasting (e.g. GABA_B).

There are also non-linear effects in the release of neurotransmitters. Indeed, if the synapse is constantly excited, the amount of vesicles ready to be dumped in the synaptic cleft may decrease significantly before recovering. This mechanism, weakening the signal transmission, is called synaptic depression. On the contrary, pre-synaptic stimulation can lead to more vesicles being docked to the membrane. Therefore, the next pre-synaptic spike might release an unusually large amount of neurotransmitters. This is called facilitation. These two mechanisms are referred to as short-term plasticity because they impact the strength of the synapse on a short term.

Functional description Synapses propagate the excitation of a neuron to the others. In chemical synapses, the electrical signal is temporarily converted to a chemical signal before it is converted back to an electrical signal in the post-synaptic neuron. This chemical transformation increases the inertia of the synapse which is slower than the spike propagation. Besides, the release of neurotransmitters is probabilistic and occurs in discrete amounts. This gives the synapse a noisy behavior.

Actually, there are three different time scales in the mechanisms described above. The first one is the time scale of the axon propagation which is faster than the others. Second, there is the time scale of signal propagation through the synapse. Third, there is the time scale of depletion or facilitation which is slower than the two previous ones.

Mathematical description Synapses can be finely modeled by writing a differential system describing the evolution of the different quantities involved in the synapse activation. The different time scales can be directly translated in the dynamical system formalism: the synapse synapse can be seen as a singularly perturbed system. For simplicity, the synapse is often considered to be a linear time-delayed filter which converts an incoming spike into a (smoother) post-synaptic potential.

As discussed in the section 1.1.4, this is often modeled together with slower differential equations governing the strength of the connections (due to long term plasticity).

1.1.3.2 Synaptic dynamics

Although the synapse is a crucial element of the signal processing, there are less papers about synapse models than about neuron models [Tsodyks et al. 1998, Destexhe et al. 1998, Ermentrout and Terman 2010]. We present two models: the first one is partly based on biology and leads to a complicated system, the second, is a simple functional model which is widely used.

A non-linear dynamical system [Chapeau-Blondeau and Chambet 1995] We now present a model partly based on the biological mechanism of the synapse. Indeed, we can model the synaptic currents to the post-synaptic neuron as the product of a conductance with a voltage difference:

$$I_{\text{syn}} = g(t)(v_{\text{post}} - v_{\text{rest}})$$

where $g(t) = \bar{g}s(t)$ where $s(t)$ is the fraction of open-channels and $\bar{g} \in \mathbb{R}$ (its sign depends on the nature of the neurotransmitter). If $[T]$ is the concentration of neurotransmitter in the synaptic cleft, one can write an idealized equation for the evolution of s:

$$\dot{s} = a_r[T](1-s) - a_d s$$

where $a_r, a_d \in \mathbb{R}_+$. This equation is to be coupled with a model of the evolution of $[T]$ depending on the pre-synaptic excitations. There are different possibilities but as suggested in [Destexhe et al. 1998] (where they fit the parameters to data) this can be

$$[T](v_{\text{pre}}) = \frac{T_{\text{max}}}{1 + e^{-\frac{v_{\text{pre}} - v_T}{K_p}}}$$

This model is very simple compared to the biological mechanisms but already mathematically difficult. Therefore, we focus on a simpler one.

Linear functional model Synapse are often modeled as simple linear transforms of the pre-synaptic spike trains. The idea is to observe the temporal evolution of a post-synaptic potential after a spike hit the pre-synaptic axon terminal. The spike being close to a Dirac function it corresponds to considering the synapse as a input-output filter and measuring its impulse response. Actually, this function is close to

$$h_\tau : t \mapsto \frac{1}{\tau} e^{-\frac{t}{\tau}} H(t) \tag{1.8}$$

where $\tau \in \mathbb{R}_+$ and H is the Heaviside function. This means the input current could be written

$$I_{\text{syn}}(t) = \bar{g}\big((v_{\text{pre}} - v_{\text{rest}}) * h_\tau\big)(t) \tag{1.9}$$

where $\bar{g}$ is the strength of the synapse, v_{rest} is the value of the resting membrane potential and $*$ is the convolution. This formula holds even in the case of v_{pre} being a spike train.

If $\tau \ll 1$ then the synapse is almost instantaneous, i.e. $(h_\tau * v)(t) \simeq v(t)$. On the contrary, if τ is large then the synapse has a long memory and it has the effect of smoothing the incoming spike train into a more regular post-synaptic potential.

1.1.3.3 Evolution of the synapse strength

Most synapse models involve a term evaluating the strength of the synapse ($\bar{g}$ in the models above). Actually, the strength of the synapses is the (multi-dimensional) variable which stores memories. Thus, it is a primordial variable and we will intensely focus on it later. For now, we just mention the fact that it is activity dependent and evolves on three different time scales.

First, the short term plasticity of the synapse (facilitation of depletion) alters the synapse at the scale of the spike trains duration. We already have an interesting insight about its functional role [Varela et al. 1997, Pfister et al. 2010] but there dynamical role is less understood. Actually, these mechanisms involve non-linearities which makes them difficult to mathematically understand in large networks. Therefore, we will not take them into account in the rest of the thesis.

Second, there are learning mechanisms which change the synaptic strength on long time scales (several minutes or more). Indeed, it is widely believed that there are local learning rules which increase (potentiation) or decrease (depression) based on the activity of the network, which makes it possible to encode memories. These learning rules will be reviewed in parts 1.1.4.

Finally, there is even another mechanism called synaptic tagging which would alter the connectivity on even longer time scale, ([Frey and Morris 1997, Redondo and Morris 2010]). Indeed, long term plasticity (which is the focus of this thesis) does not last for many hours. Therefore, there is a need for another mechanism which would freeze the memories when they are judged valuable. The hypothesis is that there are external and global markers that are emitted only when the network has been exposed to a meaningful event. These markers, when spread in the network's medium, would consolidate the changes brought by functional learning and make them last longer. Therefore, the synaptic strength changes which are not followed by the arrival of these markers would eventually vanish. This kind of plasticity is tightly linked with the notion of reinforcement learning. The study of these mechanisms is beyond the scope of this thesis.

1.1.4 Learning for spiking networks

Learning and memory are generally though to be mediated by activity-dependent synapse modifications. The literature about learning rules is abundant and is better adapted to rate-based neural networks which are studied in the next chapter. Therefore, we only focus in this section on the learning mechanisms for spiking neurons. We will review these rate-based learning rules once the rate-based models are properly introduced.

1.1.4.1 What is spike timing dependent plasticity (STDP)?

There are two ways to identify the learning mechanisms occurring at the synapse level. Either one considers the neuron as a black box and injects various inputs to see how the connection from the pre-synaptic neuron to the post-synaptic neuron evolves. This methods leads to identifying functional learning rules. Or one could analyze the cellular mechanisms underlying such functional plasticity. However, it is so complicated that it is still a challenge in neuroscience. We briefly review these approaches in the following based on the review articles [Caporale and Dan 2008, Sjöström and Gerstner 2010].

Functional description It has been known for a long time that high-frequency stimulation of the pre-synaptic neuron leads to long-term potentiation, i.e. an increase in the synapse strength. Another way to induce long-term potentiation is by pairing low-frequency stimulation of the pre-synaptic neuron with large post-synaptic depolarization. On the contrary, long-term depression, i.e. a decrease in the synapse strength, can be induced by low-frequency stimulation of the pre-synaptic neuron, either alone or paired with a small post-synaptic depolarization. Together, these mechanisms allow a balanced modification of the network connectivity. Beyond this static view, numerous recent papers have established the importance of the relative timing of the pre-synaptic and post-synaptic spikes (among them [Bi and Poo 1998]). This leads to the definition of the canonical spike timing dependent plasticity (STDP): if a pre-synaptic spike reaches the the synapse shortly before the post-synaptic neuron fire a spike, then the synapse is potentiated. On the contrary, if a pre-synaptic spike reaches the the synapse shortly after the post-synaptic neuron fire a spike, then the synapse is depressed. The modification profiles are shown in figure 1.7. Actually, the canonical STDP captures the importance of causality in determining the direction of synaptic plasticity. This behavior has been robustly observed for many different synapses in different parts of the brain from insects to humans. However, some other temporal profiles where also observed as shown in figure 1.8.

In this thesis, we only focus on the canonical STDP which is always implied when we use the acronym STDP.

Biological mechanisms It is not completely clear yet what are the cellular mechanisms responsible for the functional observations above. It seems both potentiation and depression depend on the post-synaptic intracellular calcium transients. Brief and large calcium events lead to potentiation of the synapse, whereas longer and smaller calcium events lead to depression of the synapse. This known as the calcium hypothesis.

For glutamatergic (excitatory) synapses, the NMDA receptor is though to serve as a coincidence detector. When the pre-synaptic axon terminal releases glutamate and the post-synaptic neuron is also depolarized, this causes removal of the Mg^{2+} blocks which where blocking the NMDA receptor. Then, some calcium ions Ca^{2+} can go into the post-synaptic neuron through the open NMDA receptors.

Actually, the synapse is rarely close to the soma of the post-synaptic neuron. Therefore, there are back-propagating action potentials which goes up

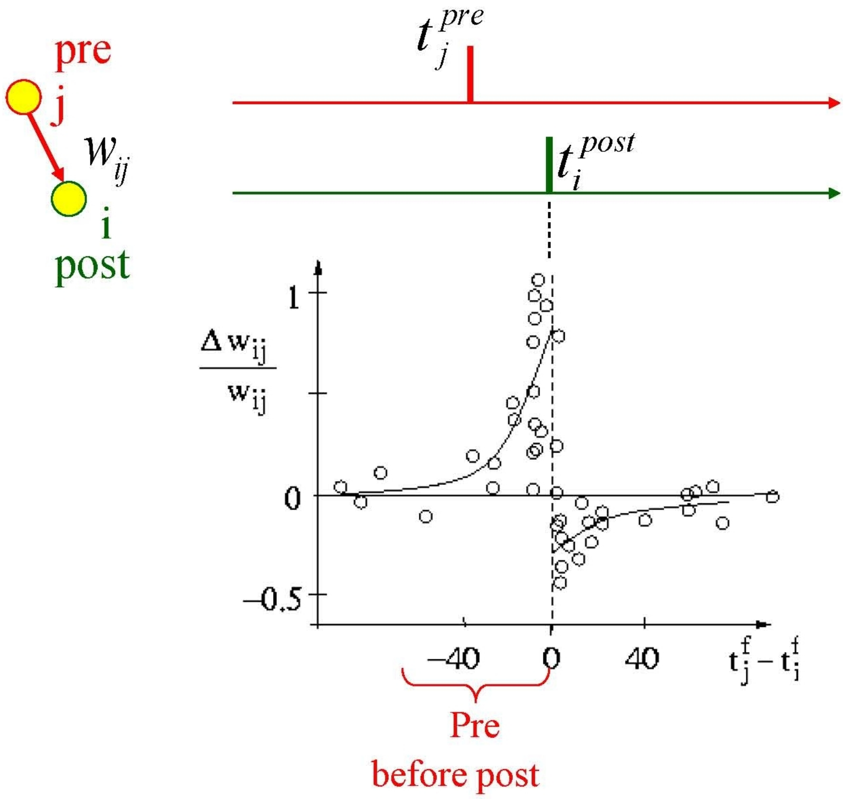


Figure 1.7: Spike-Timing Dependent Plasticity: The STDP function shows the change of synaptic connections as a function of the relative timing of pre- and post-synaptic spikes after 60 spike pairings. Taken from [Sjöström and Gerstner 2010], originally adapted from [Bi and Poo 1998].

the dendrite when the neuron fires so that the coincidence of pre and post-synaptic spikes can be detected.

Of course, this is a simplistic picture of the real mechanisms which can be very diverse. Yet, cellular mechanisms are not the topic of this thesis. Thus, we will not go beyond this description.

1.1.4.2 Models of STDP

Recently, many papers have been devoted to model and analyze the role of STDP [Van Rossum et al. 2000, Song et al. 2000, Gerstner and Kistler 2002a, Izhikevich and Desai 2003, Masquelier et al. 2009]. Because, it is a learning rule based on the time of the spikes and also because some popular neuron models are hybrid, many papers start with a formulation of STDP which ex-

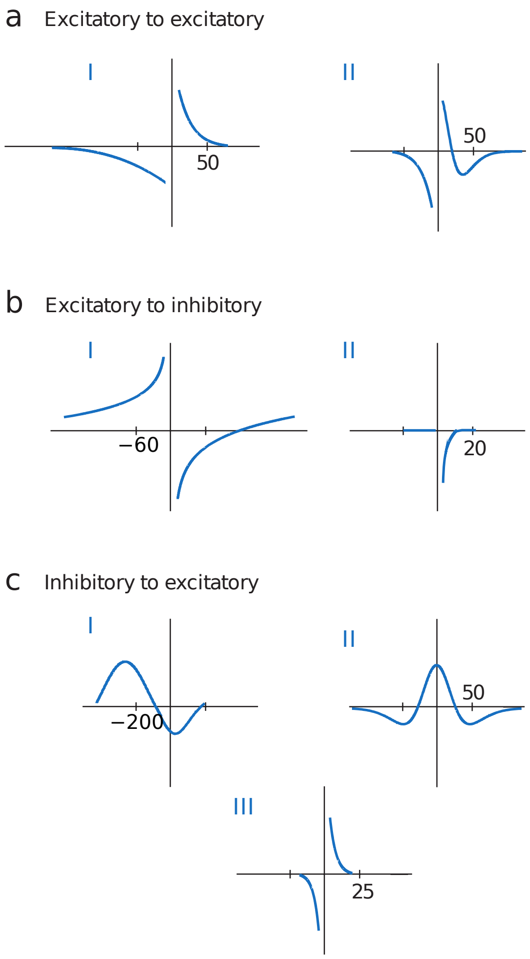


Figure 1.8: Diversity of temporal windows for STDP. Temporal axes in milliseconds. Taken from [Caporale and Dan 2008].

plicitly uses the times of the spikes. Although, we briefly recall what a model of STDP is in this framework, we will take a very different approach in this thesis. We will not define the times of the spikes and derive a continuous learning rule which implements STDP, whatever the neurons' dynamics.

Spike times models Consider the connection from neuron j to neuron i, which we write $\mathbf{J}_{ij}$. Assume that neuron i fires at times t_i^m for $m = 1, 2, ..$ (and similarly for j). Therefore, the modification of the connectivity $\Delta\mathbf{J}_{ij}$ is

$$\Delta\mathbf{J}_{ij} = \sum_m \sum_n W(t_i^m - t_j^n) \tag{1.10}$$

where W defines the learning window as illustrated in figure 1.7. We can model it by

$$\begin{aligned} W(x) &= a_+ e^{-\frac{x}{\tau_+}} && \text{if } x > 0 \\ W(x) &= -a_- e^{\frac{x}{\tau_-}} && \text{if } x < 0 \end{aligned} \tag{1.11}$$

with $a_\pm, \tau_\pm \in \mathbb{R}_+$ and $a_\pm$ may depend on the value of $\mathbf{J}_{ij}$.

Continuous models Assume that the membrane potential of neuron i (resp. j) is $\mathbf{v}_i$ (resp. $\mathbf{v}_j$). First, we assume that the evolution of the membrane potential is governed by a model which generates pure train spikes, i.e. sums of Dirac functions. In this framework the following system equation is exactly similar to (1.10)

$$\Delta \mathbf{J}_{ij}(t) = a_+ \bar{\mathbf{v}}_i(t)(\bar{\mathbf{v}}_j * h_{\tau_+})(t) - a_-(\bar{\mathbf{v}}_i * h_{\tau_-})(t)\bar{\mathbf{v}}_j(t) \tag{1.12}$$

where $\bar{\mathbf{v}}_i = \mathbf{v}_i - v_{\text{rest}}$ and $h_{\tau_\pm}(t) = e^{-\frac{t}{\tau_\pm}} H(t)$ with H the Heaviside function. Indeed, if the neuron has a pure spiking behavior, then the term $a_+\bar{\mathbf{v}}_i(t)(\bar{\mathbf{v}}_j * h_{\tau_+})(t)$ is non-null when the post-synaptic neuron i is spiking at time t, and then, via the factor $\bar{\mathbf{v}}_j * h_{\tau_+}$, it counts the number of previous spikes from the pre-synaptic neuron j that might have caused the post-synaptic spike. This calculus is weighted by an exponentially decaying function h_+. This accounts for the left part of figure 1.7. The last term $-a_-(\bar{\mathbf{v}}_i * h_{\tau_-})(t)\bar{\mathbf{v}}_j(t)$ takes the opposite perspective. It is non-null when the pre-synaptic neuron j is spiking and counts he number of previous spikes from the post-synaptic neuron i that are not likely to have been caused by the pre-synaptic neuron. The computation is also weighted by the opposite of an exponentially decaying function $-h_-$. This accounts for the right part of figure 1.7.

Actually, we can relax the assumption about the pure spike trains and this formula still makes sense and implements the STDP rule, provided the neuron has a spiking behavior. This rule is interesting, first, because it can be easily coupled with any neuron model (we do not need to known the times of the spikes). Second, because it has been observed that the STDP is in fact more dependent on the voltage rather than the exact times of the spike. In a way, it is more natural than the previous definition.

In this thesis, we assume that $h_\tau = h_{\tau_+} = h_{\tau_-}$, and that the terms $a_\pm$ do not depend on the current value $\mathbf{J}_{ij}$ and we add a simple linear decay term which says that without excitation the strength of the synapse goes to zero.

This reads

$$\dot{\mathbf{J}}_{ij} = a_+ \bar{\mathbf{v}}_i (\bar{\mathbf{v}}_j * h_\tau) - a_- (\bar{\mathbf{v}}_i * h_\tau) \bar{\mathbf{v}}_j - \kappa \mathbf{J}_{ij} \tag{1.13}$$

with $\kappa \in \mathbb{R}_+$.

1.1.5 Recurrent networks

Now, we can build networks from the different models described above. We could combine any model together as long as they have a variable for the membrane potential (which is the case for all those we introduced). Yet, we are going to focus on a simple network model which we are able to analyze afterwards. We choose the Mc Kean neuron model together with linear functional synapses and continuous STDP modeling.

Therefore, we consider a network of $n \in \mathbb{N}^*$ identical McKean neurons. Each neuron is described by the membrane potential $\mathbf{v}_i$ and its adaptation variable $\mathbf{w}_i$. Thus, the network is characterized by its field of membrane potential $\mathbf{v} \in \mathbb{R}^n$ and its field of adaptation $\mathbf{w} \in \mathbb{R}^n$. The adaptation variable is generally considered to be slow compared to the activity variable. Therefore, we assume $\varepsilon_w \ll 1$ in (1.6).

The neurons are connected with synapses which differ only by their strength, i.e. their impulse responses have the same shape. This makes it possible to consider that each pair of neurons is connected via a single effective synapse whose strength is the mean of all real synapses linking the pair. Therefore, we can define the connectivity matrix $\mathbf{J} \in \mathbb{R}^{n \times n}$ of the network such that the strength of the synapse between neuron i and j is $\mathbf{J}_{ij}$. A common assumption, it to consider that the contribution of the different synapse coming to a neuron are summed at the neuron's soma. Together with the assumption of linear synapses (with time constant $\tau_s \in \mathbb{R}_+$), this leads to considering that the communication term for neuron i is $\sum_{j=1}^{n} \mathbf{J}_{ij} (\bar{\mathbf{v}}_j * h_{\tau_s}) = \{\mathbf{J} \cdot (\bar{\mathbf{v}} * h_{\tau_s})\}_i$ where $\bar{\mathbf{v}}_i = \mathbf{v}_i - v_{\text{rest}}$. Later we will be analyzing the behavior of such a network when the number of neurons n tends to infinity. Therefore, we need to scale the communication term by $\frac{1}{n}$ such that adding neurons does not increase the global excitation.

The learning mechanisms are embodied by another differential equation on the connectivity variable $\mathbf{J}$. It corresponds to equation (1.13). We choose the decay time of the learning window to be $\tau_l \in \mathbb{R}_+$. It turns out the time scale of learning is much smaller than that of the activity variables. Therefore, we introduce a small parameter $\varepsilon_J \ll 1$ which is pre-multiplying the right hand side of (1.13).

We assume that the external inputs interact with the neurons in the form of an additive current $\mathbf{u}(t) \in \mathbb{R}^n$.

Some uncorrelated additive noise is added to the membrane potential to take into account neglected microscopic fluctuations for instance. For biological relevance (if any), the adaptation variable $\mathbf{w}$ is to be considered slow compared to the membrane potential $\mathbf{v}$. This why we do not add noise on this variable: we assume it has been averaged over time.

Here, we neglect the propagation of the action potential along the axon. It is assumed that the neurons are punctual and the signals propagate at an infinite speed. This is a strong assumption since the action potential travel at different speed in the cortex which may correspond to a functional role of delays in information propagation. In particular, the dynamics of a neural network is modified by adding delays, see [Veltz and Faugeras 2011] for instance. Yet, these propagation delays significantly increase the complexity of the subsequent mathematical analysis and we could not take them into account. However, we believe the delays in the propagation will not change significantly the slow dynamics of learning which is the focus of this thesis.

After rescaling the time twice to express the system in the learning time-scale, this leads to the following system

$$\begin{cases} \varepsilon_J \varepsilon_w \, d\mathbf{v}_i &= \Big(f(\mathbf{v}_i) - \mathbf{w}_i + \frac{1}{n} \sum_{i=1}^n \mathbf{J}_{ij} (\bar{\mathbf{v}}_j * h_{\tau_s}) + \mathbf{u}_i(t) - I_0 \Big) dt + \sigma dB_i(t) \\ \varepsilon_J \, \dot{\mathbf{w}}_i &= \mathbf{v}_i - b\mathbf{w}_i \\ \dot{\mathbf{J}}_{ij} &= a_+ \bar{\mathbf{v}}_i (\bar{\mathbf{v}}_j * h_{\tau_l}) - a_- (\bar{\mathbf{v}}_i * h_{\tau_l}) \bar{\mathbf{v}}_j - \kappa \mathbf{J}_{ij} \end{cases} \tag{1.14}$$

where $B(t)$ is a n-dimensional Brownian noise, $\bar{\mathbf{v}}_i = \mathbf{v}_i - v_{\text{rest}}$, $h_\tau(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}} H(t)$ (where $H(t)$ is the Heaviside function) and f is defined in (1.5). We recall ε_J, ε_w, l, c, a, τ_l, τ_s, I_0, σ, b, $\kappa \in \mathbb{R}_+$ with $\varepsilon_J, \varepsilon_w \ll 1$ and $a_\pm \in \mathbb{R}$.

We can rewrite this equation in a vector form by observing that $\bar{\mathbf{v}}_i(\bar{\mathbf{v}}_j * h_\tau) = \{\bar{\mathbf{v}} \otimes (\bar{\mathbf{v}} * h_\tau)\}_{ij}$ where $\otimes$ is the tensor (or Kronecker) product. Indeed,

$$\begin{cases} \varepsilon_J \varepsilon_w \, d\mathbf{v} &= \Big(f(\mathbf{v}) - \mathbf{w} + \frac{1}{n} \mathbf{J} \cdot (\bar{\mathbf{v}} * h_{\tau_s}) + \mathbf{u}(t) - I_0 \Big) dt + \sigma dB(t) \\ \varepsilon_J \, \dot{\mathbf{w}} &= \mathbf{v} - b\mathbf{w} \\ \dot{\mathbf{J}} &= a_+ \bar{\mathbf{v}} \otimes (\bar{\mathbf{v}} * h_{\tau_l}) - a_- (\bar{\mathbf{v}} * h_{\tau_l}) \otimes \bar{\mathbf{v}} - \kappa \mathbf{J} \end{cases} \tag{1.15}$$

This system is the starting point of our analysis. It is a high-dimensional, non-autonomous, non-linear, delayed, stochastic, integro-differential system.

In the following, we briefly come back to some background information before working on this system to get its mean field equation.

1.2 Networks of populations: rate-based approach

There is a fundamental gap between networks of spiking neurons and networks of rate-based neurons. The former were discussed in the previous section: they are relatively close to biology but very difficult to analyze mathematically. The latter are models of networks where the building blocks do not have a spiking behavior. On the contrary, they tend to have a smooth "analogous" profile. They are designed to represent the average behavior of populations of neurons. This is why they are also called population models. Note that their physical meaning may not be the actual firing rate of a neuron, but something close to it.

There are many reason to use firing rate models, first for computation complexity reasons, second for their direct relation with macroscopic measurements of the brain (EEG, MEG and Optical imaging) which average over numerous neurons and third because the biological structure of the network seems to define such coherent groups called cortical columns.

A central issue in mathematical neuroscience is to establish links between these two kinds of models. Of course the complex behavior of spiking networks can not be totally represented by a rate-based network but it is generally thought that rate-based models could be a good approximation of the average behavior of populations of spiking neurons. In this section, we propose a method and a rate-based equation to approximate the evolution of populations of neurons described by system (1.15).

First, we will briefly explain how the firing rate of a neuron can be computed in a simple case. We will also show how this helps deriving rate-based equations such as the Hopfield or Wilson-Cowan equations. Second, we will develop our own method to derive another rate-based equation from first principles (corresponding to system (1.15)).

In this entire part, we neglect the learning equation: we assume it is occurring on such a slow timescale that the connectivity can be considered constant. Finally, we will summarize the results and argue to extend this rate-based approach to a learning neural network.

1.2.1 Usual derivation

1.2.1.1 Computing the firing rate of leaky integrate and fire neuron

Here, we first define a very simple neuron characterized only by its membrane potential v and suggest how to find an analytical expression for its firing rate ν. This popular approach was initiated in [Ricciardi and Smith 1977] and is widely used in recent papers. For precise references and a pedagogical introduction see chapter 15 of [Feng 2004] written by Renart, Brunel and Wang. Here, we closely follow this reference.

Consider a leaky integrate an fire neuron driven by a constant input I and some white noise:

$$\tau_m dv = \big(-v + v_{ss} \big) dt + \sigma \sqrt{\tau_m} dB(t)$$

with $\sigma, \tau_m \in \mathbb{R}_+$. The constant v_{ss} is the sum of a a leak and an external input. Besides define a threshold $v_{th} \in \mathbb{R}$. When the former Ornstein-Uhlenbeck process crosses this threshold, it is considered that the neuron emits a spike. The membrane potential is therefore reset to $v_r \in \mathbb{R}$. To compute the firing of such a neuron, it is necessary to compute the Fokker-Planck equation associated with this process which governs the time evolution of the density probability of the random variable $V(t)$. To mimic the reset mechanism define an absorbing barrier corresponding to the threshold v_{th} and inject an extra probability current at v_r. As shown in chapter 15 of [Feng 2004], it is possible to compute the stationary probability density function of the neuron model $\rho(v)$. Taking into account the fraction $\nu\tau_{ref}$ of neurons in a refractory state, the steady state firing rate can be found by the normalization condition $\int_{-\infty}^{v_{th}} \rho(x)dx + \nu\tau_{ref} = 1$. This leads to the analytical expression

$$\nu = \left(\tau_{ref} + \tau_m \sqrt{\pi} \int_{\frac{v_r - v_{ss}}{\sigma}}^{\frac{v_{th} - v_{ss}}{\sigma}} e^{x^2} \big(1 + \mathrm{erf}(x)\big) dx \right)^{-1} \tag{1.16}$$

where $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$. Because, $v_{ss} = I + v_l$, it is possible to plot the dependence of the firing rate on the constant input I as shown in figure 1.9.

The complexity of formula (1.16) makes it difficult to use in networks of neurons. However, Brunel and colleagues [Amit and Brunel 1997, Brunel 2000, Fourcaud-Trocmé et al. 2003] have managed to get interesting insights in the modeling of networks based on this formula. Yet, our approach will be different and we only needed to highlight the existence of a simple functional dependence of a neuron's firing rate on its inputs , shown in figure 1.9.

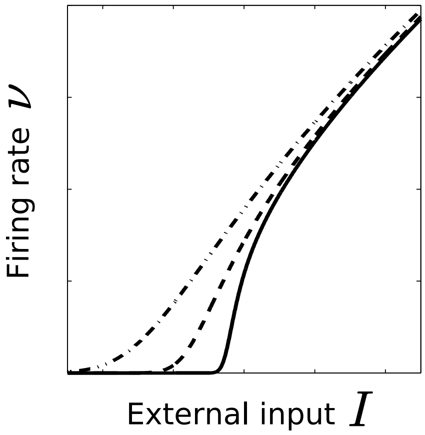


Figure 1.9: This shows the function $\nu(I)$ for the leaky integrate and fire neuron. The three curves correspond to different level of noise with $0 < \sigma_{\text{solid}} < \sigma_{\text{dashed}} < \sigma_{\text{dot-dashed}}$. Adapted from chapter 15 of [Feng 2004] where the numerical values can be found.

1.2.1.2 Deriving Hopfield and Wilson-Cowan equations

The present part consist in a heuristic derivation of a system describing the evolution of a network of rate-based based on the assumption that there exists a function S such that $\nu = S(I)$.

This derivation was advocated in early works of Jack Cowan and later formalized in [Ermentrout and Cowan 1980]. A nice introduction can be found in chapter 11 of Ermentrout's book [Ermentrout and Terman 2010]. It is an interesting derivation since the Hopfield or Wilson-Cowan equations are essentially the underlying "biology" in the popular theory of neural networks, even if the link with biology is rather tenuous.

Let us highlight the intrinsic assumptions needed to use this method.

Assumptions

- The synapses correspond to the linear filters described above. Only the strength of the synapse $\mathbf{J}_{ij}$ is supposed to vary among synapses: the transfer functions of the synapses are proportional.

- The synapses are slow compared to the duration of a spike.

- As suggested in the previous part, the firing rate $\boldsymbol{\nu}_i$ of each neuron i can be written $\boldsymbol{\nu}_i(t) = S\big(\mathbf{v}_i(t)\big)$ where S is a positive sigmoid.

There are several assumptions that we add to make the analysis easier. They are not fundamentally necessary and dropping them leads to similar results. First, we assume the spikes are Dirac functions produced by a neuron model which is not specified. This replaces the second assumption. Second, we assume the spikes propagate infinitely fast. Third, we assume the finite impulse response of the synapse is proportional to $h_{\tau_l}(t) = \frac{1}{\tau_l} e^{-\frac{t}{\tau_l}} H(t)$.

First, we observe that under these assumptions the total potential that neuron i receives from neuron j is $\Phi_{ij}(t) = \sum_m h_{\tau_l}(t - t_j^m)$, where t_j^m are the times of the spikes emitted by neuron j. Then, on can say that the firing rate $\boldsymbol{\nu}_i(t)$ determines the instantaneous number of spikes to a emitted by neuron i. Therefore, $\boldsymbol{\nu}_i(t)dt$ can be seen as the probability of a spike being emitted in the time interval $]t, t + dt[$. Thus the total potential brought to i by j is

$$\Phi_{ij}(t) = \int_{t_0}^{t} h_{\tau_l}(t - s)\boldsymbol{\nu}_j(s)ds = \int_{t_0}^{t} h_{\tau_l}(t - s)S\big(\mathbf{v}_j(s)\big)ds$$

If the contributions of the different pre-synaptic neurons and the external input $\mathbf{u}$ sum at the post-synaptic soma, the membrane potential of neuron i becomes

$$\mathbf{v}_i(t) = \int_{t_0}^{t} h_{\tau_l}(t - s)\Big(\sum_{j=1}^{n} \mathbf{J}_{ij} S\big(\mathbf{v}_j(s)\big) + \mathbf{u}_i(s)\Big)ds$$

This is the solution of the differential equation

$$\tau_l \dot{\mathbf{v}}_i = -\mathbf{v}_i + \sum_{j=1}^{n} \mathbf{J}_{ij} S\big(\mathbf{v}_j\big) + \mathbf{u}_i(t) \tag{1.17}$$

which can be rewritten in a vector form

$$\tau_l \dot{\mathbf{v}} = -\mathbf{v} + \mathbf{J} \cdot S\big(\mathbf{v}\big) + \mathbf{u}(t) \tag{1.18}$$

This equation is a well-know rate-based equation, sometimes called the continuous Hopfield equation ([Hopfield 2007]). The neurons rates are not exactly $\mathbf{v}$ but $S(\mathbf{v})$.

Remark 1.

- *It is interesting to observe that the non-linearity intrinsically characterizing the dynamics of the neurons only is now in the communication term. Actually, this model has mixed the dynamics of the different building blocks.*

- *In fact, there is no need to define populations of neurons in this derivation.*

- *It is also possible to derive (with the same method) an equation for the synaptic drive of the neurons defined by $\mathbf{z}_i = \int_{t_0}^{t} h_1(t-s)\boldsymbol{\nu}_i(s)ds$. This equation is*

  $$\tau_l \dot{\mathbf{z}} = -\mathbf{z} + S\big(\mathbf{J}\cdot\mathbf{z} + \mathbf{u}(t)\big) \tag{1.19}$$

  Although not directly linked to the firing rate, it is often preferred to the voltage-based equation (1.18). Separating excitatory and inhibitory neurons leads to the famous Wilson-Cowan equations, ([Wilson and Cowan 1972, Wilson and Cowan 1973]).

- *A space continuous version of these equations with a fixed convolutional connectivity is called a neural field ([Coombes 2005]).*

These equations are highly based on the arbitrary assumption that $\boldsymbol{\nu}_i(t) = S\big(\mathbf{v}_i(t)\big)$. The shape of the sigmoid S is only suggested by figure 1.9 and conflicts the modeling choice of having neurons emitting Dirac spike trains. Besides, these rate-based equations are not explicitely related to a underlying microscopic spiking network and therefore, it is not possible to compare these equations to a spiking network to assess the accuracy of the averaging. In the following, we derive another rate-based equation which avoids these three drawbacks.

1.2.2 Derivation of a rate model averaging a network of McKean neurons

This part is entirely original. It corresponds to recent research and therefore is not completely formalized.

The purpose of this section is to derive a rate-based equation which approximates the average behavior of populations of McKean neurons. Therefore, we consider a network of p populations (labeled with Greek letters) with a total of pn neurons. Thus, we intend to derive a system of p equations which approximates the average behavior of a spiking system of np equations. For each neuron i, we define $p(i) = 1, .., p$ the population it belongs to. Besides, for each population β and $i = 1, .., n$, $\beta(i)$ design the i_{th} neuron of the population.

Assume that the connectivity $\mathbf{J}$ is constant, i.e. $\varepsilon_J = 0$. Besides, assume $b = 1$ for simplicity. In a first time, assume that $I_0 = 0$; the case $I_0 \neq 0$ will be studied in part 1.2.2.6. Together with the notations above, we rewrite system (1.15)

$$\begin{cases} d\mathbf{v}_i & = \Big(f(\mathbf{v}_i) - \mathbf{w}_i + \frac{1}{p} \displaystyle\sum_{\beta=1}^{p} \frac{1}{n} \sum_{j=1}^{n} \mathbf{J}_{i\beta(j)} (\bar{\mathbf{v}}_{\beta(j)} * h_{\tau_s}) + \mathbf{u}_{p(i)}(t) \Big) dt + \sigma dB_i(t) \\ \dot{\mathbf{w}}_i & = \varepsilon_w (\mathbf{v}_i - \mathbf{w}_i) \end{cases} \tag{1.20}$$

with f cubic piecewise linear function given by (1.5) and $\bar{\mathbf{v}}_i = \mathbf{v}_i - v_{\text{rest}}$

1.2.2.1 Spatial averaging

Now, we apply spatial averaging methods to get a system describing the evolution of the neural network in the limit $n \to +\infty$. We have identified two kinds of averaging methods which have not been merged yet. Both of them ask for several hypotheses. For instance, we will assume that the inputs coming to a population α is the same for all the neurons in the population, i.e. $\mathbf{u}_i(t) = \mathbf{u}_{p(i)}(t)$. There are additional hypotheses on the connectivity matrix $\mathbf{J}$ which depend on the method chosen. Breaking these assumptions will be numerically studied in part 1.2.2.5.

1. **The McKean-Vlasov equation** reviewed and studied in [Dawson and Gartner 1987, Sznitman 1991] and precisely formalized by [Touboul 2011, Baladron et al. 2011] in the case of neural networks. The main assumption is that all the connections from population α to population β are the same, with $\alpha, \beta = 1, .., p$. This reads $\mathbf{J}_{ij} = \mathbf{J}_{p(i)p(j)}$. It can also be considered that there is a dynamic noise, e.g. a centered Ornstein-Uhlenbeck process, added to the connection between each pair of neurons. Under these assumptions it can be rigorously proved (see the references above) that, asymptotically, the neurons of population α are stochastic processes following the same law written $\mathbf{v}_\alpha$ and governed by system (1.21) below. Besides, if the neurons' initial conditions of all the neurons are independent then any finite number of neurons will remain independent to any other finite group of neuron during the system's evolution. This is called propagation of chaos.

$$\begin{cases} d\mathbf{v}_\alpha &= \Big(f(\mathbf{v}_\alpha) - \mathbf{w}_\alpha + \frac{1}{p}\sum_{\beta=1}^{p} \mathbf{J}_{\alpha\beta}\big(\mathbb{E}(\bar{\mathbf{v}}_\beta) * h_{\tau_s}\big) + \mathbf{u}_\alpha(t)\Big)dt + \sigma dB_i(t) \\ \dot{\mathbf{w}}_\alpha &= \varepsilon_w(\mathbf{v}_\alpha - \mathbf{w}_\alpha) \end{cases} \tag{1.21}$$

where $\mathbf{J}_{\alpha\beta} \in \mathbb{R}$ is a component of matrix (abusively written) $\mathbf{J} \in \mathbb{R}^{p\times p}$ linking the populations together. The main difficulty in this system is the presence of the term $\mathbb{E}(\bar{\mathbf{v}}_\beta)$ which the expectation of the random variable $\bar{\mathbf{v}}_\beta(t)$: this system in integro-differential.

2. **The Ben Arous equation** inspired by [Arous and Guionnet 1995] and applied to a rate based neural network in [Faugeras et al. 2009a]. The main assumption here is that the distribution of the connection neuron j in population β to neuron i in population α is

$$\mathbf{J}_{ij} \sim \mathcal{N}\Big(\frac{\mathbf{J}_{\alpha\beta}}{n}, \frac{\mathbf{\Lambda}^2_{\alpha\beta}}{n}\Big)$$

where $\mathbf{\Lambda} \in \mathbb{R}_+^{p\times p}$. This corresponds to frozen noise. The case $\mathbf{\Lambda} = 0$ corresponds to the McKean-Vlasov method (without dynamic noise). The results proven in the two reference above tend to show that a neuron of population α would have its law governed by the following equation

$$\begin{cases} d\mathbf{v}_\alpha &= \Big(f(\mathbf{v}_\alpha) - \mathbf{w}_\alpha + \frac{1}{p}\sum_{\beta=1}^{p} \mathbf{U}_{\alpha\beta}^{\bar{\mathbf{v}}*h_{\tau_s}(t)} + \mathbf{u}_\alpha(t)\Big)dt + \sigma dB_i(t) \\ \dot{\mathbf{w}}_\alpha &= \varepsilon_w(\mathbf{v}_\alpha - \mathbf{w}_\alpha) \end{cases} \tag{1.22}$$

where $\mathbf{U}^V(t) \in \mathbb{R}^{p\times p}$ is the effective interaction process, a Gaussian process of parameters

$$\begin{cases} \mathbb{E}\left[\mathbf{U}^X_{\alpha\beta}(t)\right] = \mathbf{J}_{\alpha\beta}\mathbb{E}[X_\beta(t)]; \\ Cov(\mathbf{U}^X_{\alpha\beta}(t), \mathbf{U}^X_{\alpha\beta}(s)) = \mathbf{\Lambda}^2_{\alpha\beta}\mathbb{E}\Big[X_\beta(t)X_\beta(s)\Big]; \\ Cov(\mathbf{U}^X_{\alpha\beta}(t), \mathbf{U}^X_{\gamma\delta}(s)) = 0 \text{ if } \alpha \neq \gamma \text{ or } \beta \neq \delta. \end{cases} \tag{1.23}$$

The proof of this is not rigorous and would need some work to be made reliable. Yet, it seams reasonable at first sight and many authors ([Amari et al. 1977, Sompolinsky et al. 1988, Samuelides and Cessac 2007])have used this heuristic in a similar way.

Observe that the two averaged systems (1.21) and (1.22) are very similar: the expectation of $\mathbf{v}_\alpha$ is governed by the same system.

The second variable of both system is simply linked to the first: it can be written $\mathbf{w}_i = \mathbf{v}_i * h_{\tau_w}$ where $\tau_w = \frac{1}{\varepsilon_w}$ and $h_\tau(t)e^{-\frac{t}{\tau}}H(t)$. This makes it possible to remove the the second equation by adding a delayed term in the first.

In the following, we need to assume we are in the McKean-Vlasov case if we want to be rigorous. However, we believe the results extend to mixture of frozen and dynamical noises. Actually, the simulations in part 1.2.2.5 tend to confirm this idea.

Actually, systems (1.21) and (1.22) are very complicated and there is no satisfying formalism to analyze their solutions. Indeed, they look like classical stochastic differential systems but the right hand side includes a term depending on the law of the process, e.g. the expectation of $\mathbf{v}_\alpha$ for system (1.21). They are fundamentally very different and more complicated than usual differential systems. Besides, usual Monte-Carlo methods do not work to numerically compute the time-evolving law of the solutions. Indeed, if one wants to compute one trajectory of the system, one needs to already know the law of the solution. Therefore, it is way beyond the scope of this thesis to analyze rigorously the solutions of such a system.

However, we show in the following it is possible to derive a simple system describing the firing rate of each population starting from systems (1.21) or (1.22).

1.2.2.2 Firing rate

We need to define what we mean by firing rate of a population. An instantaneous firing rate measurement could correspond to counting the spikes emitted by neurons in a population at time t. However, for spiking trains, this variable would be almost binary which does not correspond to the smooth variable we wish to build. It is therefore natural to count the number of spikes emitted during a fixed time window centered at time t. Therefore, the firing rate implies two integrations: the first is spatial, the second is temporal. Counting the number of spikes may not be an easy task for the continuous neurons we have chosen. However, we can use the fact that the spikes approximately have the same shape: integrating the value of the membrane potential during a time window and dividing by the area under a spike gives the number of emitted spikes.

Therefore, we define the firing rate $\boldsymbol{\nu}_\alpha$ of population α as the spatial and temporal mean of the $\mathbf{v}_\alpha$ over the neurons in the populations and over a time window of width θ larger than the duration of the spikes. Note that this is not exactly the firing rate of a population but, because the spikes have always the same shape, we claim it is highly correlated to firing rate. Therefore, we will abusively refer to it as the firing rate although it might be better to call it spatio-temporal running average of the membrane potential. Observe that we have to make sure $\tau_w \ll \theta$. We choose to use a Gaussian time window, i.e. $g(s) = \frac{1}{K} e^{-\frac{t^2}{\sigma^2}}$. We choose σ so that $Kg(\theta/2) = 0.01$ and K so that $\int_\mathbb{R} g(s)ds = 1$. This means the variable $\boldsymbol{\nu}_\alpha$ is only proportional to the real firing rate.

The mean over all the neurons in population α can be replaced by the expectation of the stochastic process $\mathbf{v}_\alpha$ when the number of neurons is large enough.

This leads to the definitions

$$\begin{aligned}
\boldsymbol{\nu}_\alpha(t) &\overset{def}{=} \big(\mathbb{E}(\mathbf{v}_\alpha) * g\big)(t) \\
\bar{\boldsymbol{\nu}}_\alpha(t) &\overset{def}{=} \big(\mathbb{E}(\mathbf{v}_\alpha - v_{\text{rest}}) * g\big)(t) = \boldsymbol{\nu}_\alpha(t) - v_{\text{rest}}
\end{aligned} \tag{1.24}$$

Besides, because the convolution is commutative we can derive the following from equation (1.21)

$$\dot{\boldsymbol{\nu}}_\alpha = \mathbb{E}\big(f(\mathbf{v}_\alpha)\big) * g - \boldsymbol{\nu}_\alpha * h_{\tau_w} + \frac{1}{p}\sum_{\beta=1}^{p} \mathbf{J}_{\alpha\beta}\big(\bar{\boldsymbol{\nu}}_\beta * h_{\tau_s}\big) + (\mathbf{u}_\alpha * g)(t) \tag{1.25}$$

In fact, the only unknown term in the formula above is $\mathbb{E}\big(f(\mathbf{v}_\alpha)\big) * g$. If we manage to express it in terms of $\boldsymbol{\nu}_\alpha$, then the system would be closed and we would have built a rigorous rate-based model from a spiking network.

1.2.2.3 Finding the sigmoids

So far the derivation was rigorous but it is now necessary to find a good approximation of the term $\mathbb{E}\big(f(\mathbf{v}_\alpha)\big) * g$ in equation (1.25). We show it leads to the computation of sigmoidal functions.

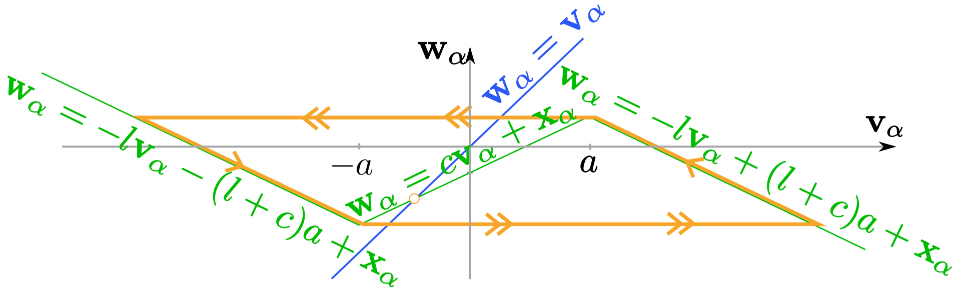
First, we get back to system (1.21) and define the effective input

$$\mathbf{x}_\alpha = \frac{1}{p}\sum_{\beta=1}^{p} \mathbf{J}_{\alpha\beta}\big(\mathbb{E}(\bar{\mathbf{v}}_\beta) * h_{\tau_s}\big) + \mathbf{u}_\alpha(t)$$

This leads to the system

$$\begin{cases} d\mathbf{v}_\alpha &= \Big(f(\mathbf{v}_\alpha) - \mathbf{w}_\alpha + \mathbf{x}_\alpha(t) \Big) dt + \sigma dB_i(t) \\ \dot{\mathbf{w}_\alpha} &= \varepsilon_w (\mathbf{v}_\alpha - \mathbf{w}_\alpha) \end{cases}$$

If $\mathbf{x}_\alpha$ is constant then we have a 2-dimensional isolated neuron model that we discussed in part 1.1.2.3. Its dynamics can be summarized by the phase space shown in figure 1.10. Actually, this is not a usual phase plane since $\mathbf{x}_\alpha$ depends on time. Therefore, the green curve in figure 1.10 is moving vertically when the $\mathbf{x}_\alpha$ changes.



$\mathbf{w}_\alpha$
$\mathbf{v}_\alpha$
$-a$
a
$\mathbf{w}_\alpha = \mathbf{v}_\alpha$
$\mathbf{w}_\alpha = -l\mathbf{v}_\alpha - (l+c)a + \mathbf{x}_\alpha$
$\mathbf{w}_\alpha = c\mathbf{v}_\alpha + \mathbf{x}_\alpha$
$\mathbf{w}_\alpha = -l\mathbf{v}_\alpha + (l+c)a + \mathbf{x}_\alpha$

Figure 1.10: Phase place of the deterministic part of system (1.21) where x_a is supposed to be a negative constant. If $\varepsilon_w \ll 1$ (which is assumed for this figure) then the horizontal part of the system trajectory (in orange with doubled arrows) goes much faster that the two others slow parts.

The first assumption is that $\varepsilon_w \ll 1$, or equivalently $1 \ll \tau_w$, so that the neurons in population α, whose law is described by (1.21), keep jumping from the two slow branches represented in the phase plane 1.10. This reads $f(\mathbf{v}_\alpha) = -l\mathbf{v}_\alpha \pm (l+c)a + \mathbf{x}_\alpha$ such that $\mathbb{E}\big(f(\mathbf{v}_\alpha)\big) = -l\mathbb{E}(\mathbf{v}_\alpha) + \mathbb{E}(\pm)(l+c)a + \mathbf{x}_\alpha$ with $\mathbb{E}(\pm) = \int_{\mathbf{v}_\alpha > 0} d\mathbf{v}_\alpha(t) - \int_{\mathbf{v}_\alpha < 0} d\mathbf{v}_\alpha(t) = \mathbb{P}(\mathbf{v}_\alpha > 0) - \mathbb{P}(\mathbf{v}_\alpha < 0)$. Therefore,

$$\mathbb{E}\big(f(\mathbf{v}_\alpha)\big) * g = -l\boldsymbol{\nu}_\alpha + (l+c)a\big(\mathbb{P}(\mathbf{v}_\alpha > 0) - \mathbb{P}(\mathbf{v}_\alpha < 0)\big) * g$$

The second assumption consists is assuming that the inputs and the timescale of the synapses are slow compared to the duration of the temporal integration window θ: in particular, $\theta \ll \tau_s$. Therefore, over a time window of size θ, $\mathbf{x}_\alpha$ can be considered constant so that system (1.21) reduces to a single isolated neuron as in equation (1.6) for which the computation of the term above is much easier. Indeed, a McKean neuron with a constant input I has a behavior such that the function $t \mapsto \Big(\big(\mathbb{P}(\mathbf{v}_\alpha > 0) - \mathbb{P}(\mathbf{v}_\alpha < 0) \big) * g \Big)(t)$ converges to a fixed value which we write $S_\sigma\big(I\big) \in \mathbb{R}$ (provided θ is sufficiently large). Therefore, we can write

$$\Big(\big(\mathbb{P}(\mathbf{v}_\alpha > 0) - \mathbb{P}(\mathbf{v}_\alpha < 0)\big) * g\Big)(t) \simeq S_\sigma\big(\mathbf{x}_\alpha(t)\big) \tag{1.26}$$

To compute the sigmoid we need to compute $\mathbb{P}(v \gtrless 0) * g$, for v the membrane potential of a noisy McKean neuron with a constant input I. This amounts to computing the proportion of time system a McKean neuron spends on (or close to) the slow manifolds $w = -lv \pm (l+c)a + I$.

Without noise $\sigma = 0$

In this case, it is possible to compute explicitly $S_0(I)$.

If $I \leq -(1-c)a$ (resp. $I \geq (1-c)a$), then the system has a single stable fixed point on the negative (resp. positive) slow manifold. In figure 1.10, this corresponds to the blue curve crossing the green piecewise cubic where the latter is decreasing. In this case, $S_0(I) = -1$ (resp. $S(I) = 1$).

If $-(1-c)a < I < (1-c)a$, then the system is oscillating on a deterministic limit cycle. In this case, $S_0(I) = \frac{T_0^+(I) - T_0^-(I)}{T_0^+(I) + T_0^-(I)}$ where $T_0^-(I)$ (resp. $T_0^+(I)$) is the duration it takes for the system to go along the negative (resp. positive) part of the slow manifold. As shown in [Coombes 2001], it is simple to compute these values . Indeed, assume the fast membrane potential immediately goes to one of the slow nullclines. This gives the equation: $-l\mathbf{v}_a \pm (l+c)a - \mathbf{w}_\alpha + I = 0$. Injecting this in the slow equation and integrating along the orange path in figure 1.10 leads to

$$T_0^+(I) = l\int_{-ca+I}^{ca+I} \frac{dw}{-(1+l)w + (l+c)a + I} = \frac{l}{1+l} ln\Big(\frac{(1+2c/l+c)a - I}{(1-c)a - I}\Big)$$

Similarly,

$$T_0^-(I) = \frac{l}{1+l} ln\Big(\frac{(1+2c/l+c)a + I}{(1-c)a + I}\Big)$$

Therefore, for $I \in]-(1-c)a, (1-c)a[$

$$S_0(I) = \frac{ln\left(\frac{\big((1+2c/l+c)a-I\big)\big((1-c)a+I\big)}{\big((1-c)a-I\big)\big((1+2c/l+c)a+I\big)}\right)}{ln\left(\frac{\big((1+2c/l+c)a-I\big)\big((1+2c/l+c)a+I\big)}{\big((1-c)a-I\big)\big((1-c)a+I\big)}\right)} \tag{1.27}$$

This function is shown in figure 1.11. It is a non-smooth sigmoidal function with vertical tangents at $-(1-c)a$ and $(1-c)a$. This corresponds to the transition from a fixed point to the oscillatory pattern.

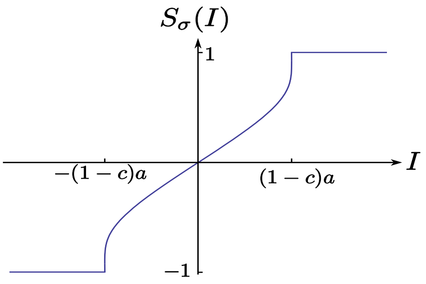


Figure 1.11: This is the representation of the function S_0 whose explicit expression is (1.27).

With noise $\sigma > 0$

In this case, it does not seem possible to compute explicitly $S_\sigma(I)$. However, we can numerically compute $S_\sigma(I)$ by simulating a single neuron, whose law is given by (1.21), over a long time period with a constant effective input I and a fixed σ. From the history of its membrane potential it is possible to compute $S_\sigma(I) = \frac{T_\sigma^+(I) - T_\sigma^-(I)}{T_\sigma^+(I) + T_\sigma^-(I)}$. From a mathematical perspective, the objects $T_\sigma^\pm(I)$ are the expectation of a hitting time to a non-linear boundary. Concretely, we compute them as the mean of the time spent of each decreasing branch of teh cubic. We have reported the results in figures 1.12 and 1.13.

We see that the global slope of the sigmoids decreases when the noise increases. The negative part of the sigmoids is in accordance with the classical results reviewed in part 1.2.1.1 and shown in figure 1.9. These results link this approach to the fully formed sigmoids which are commonly chosen for the usual rate equations described in part 1.2.1.2.

In accordance with part 1.1.2.3, the presence of noise makes it possible for the system to generate spikes even though the deterministic system driven by the same input would only have a stable equilibrium point. Following the lines of [Muratov et al. 2005], it is possible to define a critical input value (see equation (1.7)) which is the largest value for which the sigmoid is equal to -1. The value of these critical values is the red line in figure 1.13.

1.2.2.4 A model for the populations firing rate

We are now in position to define an averaged model describing the evolution of the populations firing rate. Deriving this system is the main achievement of this chapter. It takes the form of a self consistent non autonomous, delayed

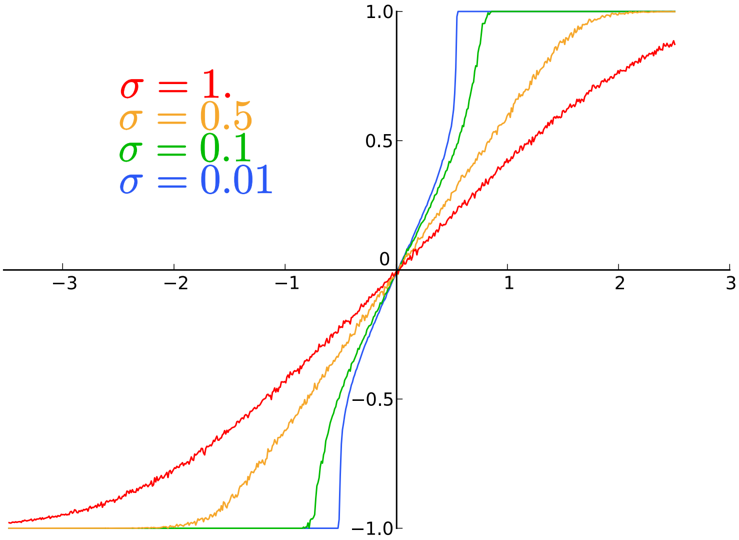


Figure 1.12: These are numerical computations of the functions $S_\sigma(I)$ with the parameters for these simulations are $a = l = 1$, $c = \frac{1}{2}$, $\tau_w = 100$. The profile of the sigmoids are shown for $\sigma \in \{0.01, 0.1, 0.5, 1.\}$. The noisy aspect is due to the computing of the the ratio $\frac{T_\sigma^+(I) - T_\sigma^-(I)}{T_\sigma^+(I) + T_\sigma^-(I)}$ over fixed a fixed time interval which is not always a multiple of the spike duration. We see that the first value of I for which the sigmoid is not null decreases with σ.

differential system.

Indeed, putting back all the pieces leads to defining an averaged system.

$$\dot{\bar{\boldsymbol{\nu}}}_\alpha = -\bar{\boldsymbol{\nu}}_\alpha * (l\delta + h_{\tau_w}) + \tilde{S}_\sigma\Big(\frac{1}{p}\sum_{\beta=1}^{p} \mathbf{J}_{\alpha\beta}\big(\bar{\boldsymbol{\nu}}_\beta * h_{\tau_s}\big) + (\mathbf{u}_\alpha * g)(t)\Big) \tag{1.28}$$

In a vector form this is

$$\dot{\bar{\boldsymbol{\nu}}} = -\bar{\boldsymbol{\nu}} * (l\delta + h_{\tau_w}) + \tilde{S}_\sigma\Big(\frac{1}{p}\mathbf{J} \cdot (\bar{\boldsymbol{\nu}} * h_{\tau_s}) + \mathbf{u} * g\Big) \tag{1.29}$$

where δ is the Dirac function and $\tilde{S}_\sigma$ is a function from $\mathbb{R}^p$ to $\mathbb{R}^p$ (with $\tilde{S}(\bar{\boldsymbol{\nu}})_\alpha = \tilde{S}(\bar{\boldsymbol{\nu}}_\alpha)$) and

$$\tilde{S}_\sigma = I_d + (l + c)aS_\sigma - (1 + l)v_{\text{rest}} \tag{1.30}$$

1.2.2.5 Numerical simulations

In this part, we show the result of several simulations to evaluate the accuracy of the averaging process and to test the robustness of the approximation to

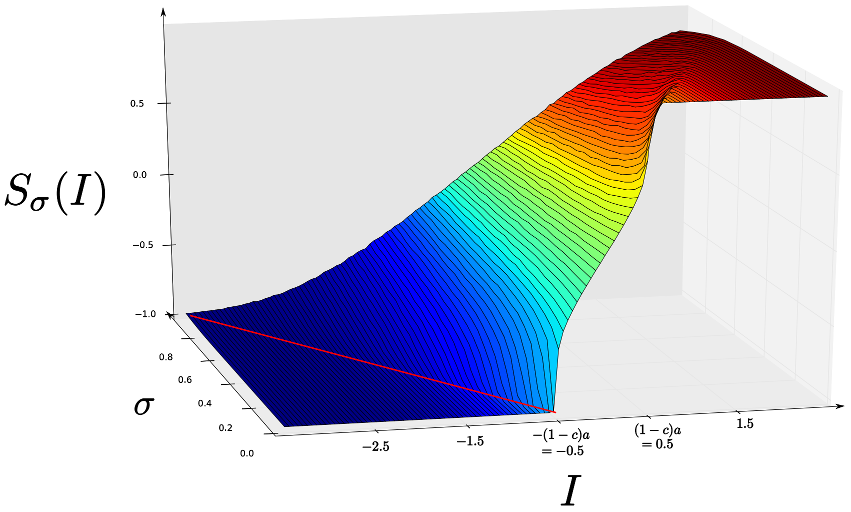


Figure 1.13: Surface representing the sigmoids for a noise amplitude ranging from 0 to 1. The surface was smoothed by a convolution with Gaussian kernel to avoid noise induced artifacts as in figure 1.12. The red line corresponds to the location of Muratov criticality points defined by (1.7). We see it accurately corresponds to the take off locus of the sigmoids.

the variation of different parameters. In this goal, we simulate both the exact system 1.20 and the averaged system 1.29. A posteriori , we compute the firing rate evolution of each population for the solutions of the exact system thanks to the definition 1.24. We then plot it together with the solutions of the averaged system to see that the curves match. We define the error between the solutions $\boldsymbol{\nu}_{\text{exact}} : t \in [0, T[\mapsto \boldsymbol{\nu}_{\text{exact}}(t) \in \mathbb{R}^p$ and $\boldsymbol{\nu}_{\text{averaged}} : t \in [0, T[\mapsto \boldsymbol{\nu}_{\text{averaged}}(t) \in \mathbb{R}^p$ of the two systems by

$$\text{error}(\boldsymbol{\nu}_{\text{exact}}, \boldsymbol{\nu}_{\text{averaged}}) = \frac{1}{Tp} \int_0^T \sum_{i=1}^{p} \left| \boldsymbol{\nu}_{\text{exact}\,i}(t) - \boldsymbol{\nu}_{\text{averaged}_i}(t) \right| dt$$

- The default **parameters** we used in the following simulations are: $a = b = l = 1$, $c = \frac{1}{2}$, $\varepsilon_w = 0.01$, $\varepsilon_s = 0.001$, $\theta = 1001$, $\sigma = 0.1$, $n = 200$, $p = 5$. The simulations are run in the fast time scale with a stochastic Euler method with a time-step of $dt = 1.$ and with a number of iteration of $T = 20000$.

- The default **inputs** are defined by $p = 5$ functions from $[0, T[$ to $\mathbb{R}$.

$$\begin{aligned}
\mathbf{u}_1(t) &= 0.1\sin(\tfrac{2\pi mt}{T}) - I_0 \\
\mathbf{u}_2(t) &= 0.1\cos(\tfrac{4\pi mt}{T}) - I_0 \\
\mathbf{u}_3(t) &= 0.2\big(\tfrac{T}{2} - |t - \tfrac{T}{2}|\big) - 0.1 - I_0 \\
\mathbf{u}_4(t) &= 0.1 - I_0 \\
\mathbf{u}_5(t) &= -0.1 - I_0
\end{aligned}$$

  where m is the number of oscillations of $\mathbf{u}_1$ and is set to $m = 2$ by default. The default value of I_0 is set to Muratov's critical value given in 1.7.

- The default **connections** between the populations are randomly drawn according to a normal distribution $\mathcal{N}(0, 0.04)$. According to the hypotheses for the McKean-Vlasov method all the neurons between two population are connected with a link of the strength $J_{\alpha\beta}$.

Figure 1.14 is the first numerical simulation which corresponds to two different values for the noise parameter σ. It shows the necessity of temporal integration (i.e. the convolution with g) to go from the fast oscillations (at the spike time scale) to the smoother plain colored curves which are well matched by the averaged solutions in dashed, black. A better proof of the accuracy of the averaged equations is the top picture of figure 1.15 which corresponds to a lot more neurons in each population.

Number of neurons per population The definition of the firing rate of a population according to 1.24 involves the expectation of the process $\mathbf{v}_\alpha$. This approximation of the real firing rate is even better if the number of neurons is larger. We test the impact of the number of neurons in figure 1.15. As expected the larger the number of neurons per population the better is the approximation.

In figure 1.15, we also check the accuracy of the method when the number of populations varies. It seems the number of populations does not change significantly the accuracy of the averaged system.

Frozen noise The underlying hypothesis in McKean-Vlasov's method is that all the neurons between two populations are connected with the same connectivity strength. Here we relax this hypothesis and add some frozen noise to

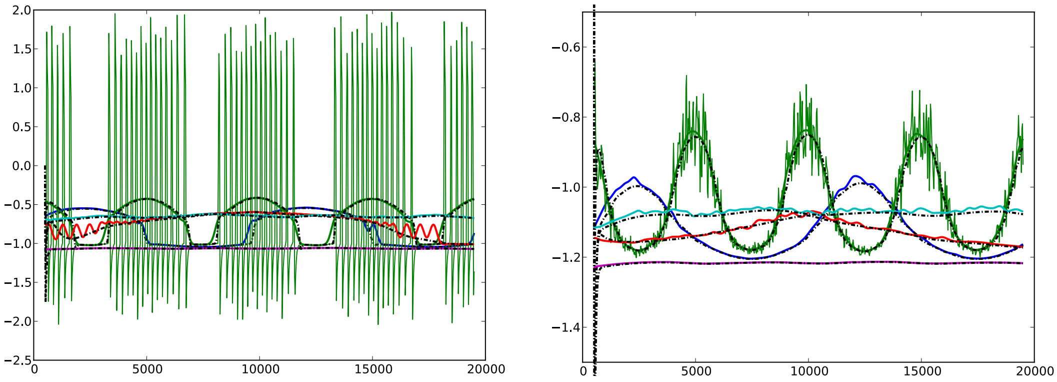


Figure 1.14: Comparison of the exact and averaged systems for different values of σ: (left) $\sigma = 0$, (right) $\sigma = 0.1$. The other parameters have their default value. In both figure, the fast oscillating green curves correspond to the value of the instantaneous firing rate for the first population only of the exact system 1.20, i.e. equation 1.24 without the convolution with g. The smooth, plain colored curves correspond to the temporally averaged firing rate of the exact system, i.e. equation 1.24. The red oscillations are border effects due to the fact that the time window of size θ may not match the period of oscillations. The evolution of the populations firing rate according to the averaged model 1.29 are plotted in black, dashed curves. The deterministic system shows large oscillations in the instantaneous firing rate because the neurons are perfectly synchronized. Adding some noise obviously reduces this synchrony and lead to a much smaller oscillations as can be seen in the right figure.

the connections in the spirit of Ben Arous' method. More precisely, we assume that the connection from neuron j in population β to neuron i in population α is

$$\mathbf{J}_{ij} \sim \mathbf{J}_{\alpha\beta}\Big(1 + \mathcal{N}(0, \sigma_J^2)\Big)$$

Similarly, we add some frozen noise to the inputs such that the input to a neuron i in population α is

$$\mathbf{u}_i \sim \mathbf{u}_\alpha\Big(1 + \mathcal{N}(0, \sigma_u^2)\Big)$$

In figure 1.16 we evaluate the accuracy of the averaged equation when these two parameters vary. It appears the frozen noise on the connections has little to no impact on the accuracy whereas the frozen noise on the inputs significantly reduces the accuracy of the averaged system. This can be partly

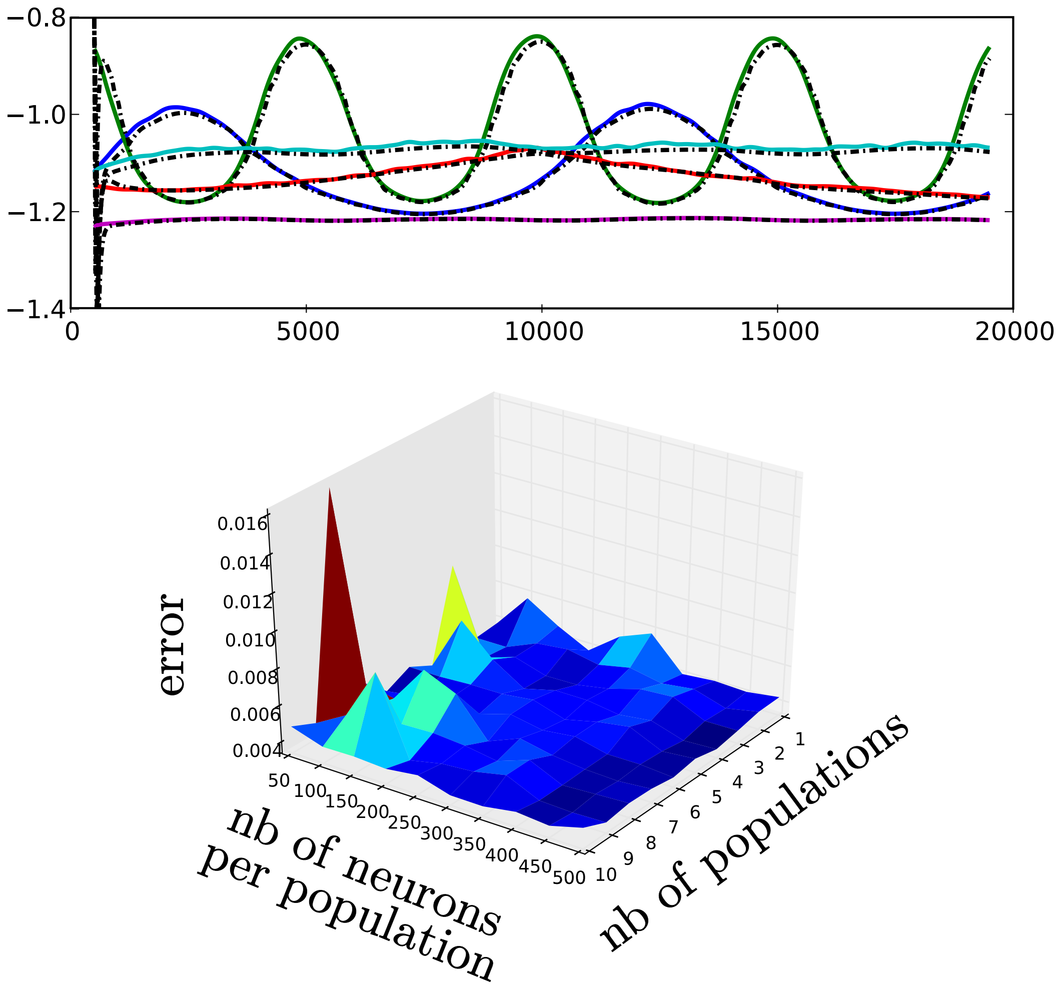


Figure 1.15: Comparison of the exact and averaged systems when the number of neurons in the network vary. (top) Time evolution of the populations firing rates of the exact (plain colored curves) and averaged (dashed black) system with the default parameters except from the number of neurons per population which is set to $n = 1000$. (bottom) Error between the 2 models when the number of neuron per population n and the number of populations p vary. The inputs where chosen to be constant with and equal to 0.8 I^*.

explained by the fact that the noise on the connectivity is averaged in the communication term which is a sum of the contribution of all the other neurons.

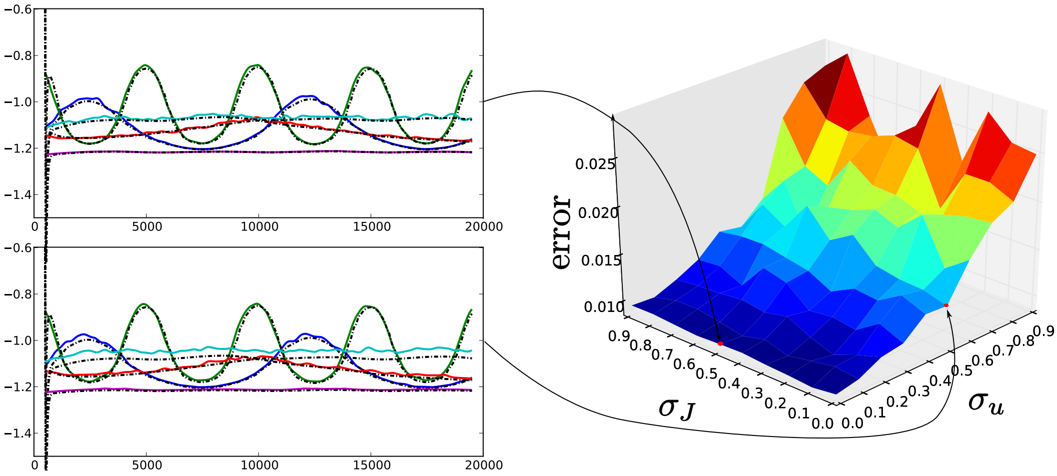


Figure 1.16: Comparison of the exact and averaged systems when the frozen noise on the connectivity and the inputs vary. (left) Time evolution of the populations firing rates of the exact (plain colored curves) and averaged (dashed black) system with the default parameters. Some frozen noise was added to the connectivity and inputs: (top left) $\sigma_J = 0.5$, $\sigma_u = 0$ (bottom left) $\sigma_J = 0$, $\sigma_u = 0.5$. (right) Error between the two models when the frozen noise on the connectivity and the inputs vary. Adding noise on the inputs makes the approximation worse than adding noise on the connectivity. In particular, it seems the population number 4 that receives the ”positive” constant input is very affected by the noise on the inputs.

Speed of the effective inputs The semi-analytic derivation of the averaged equation was based on the fact that the effective input $\mathbf{x}_\alpha$ was slow. Here, we try to see if the results extend to faster inputs and/or faster synapses by varying the number of oscillations of the inputs m and the time scale of the synapses. The results are shown in figure 1.17. Actually, the speed of the synapse does not impact on the accuracy in a monotonic way suggesting our averaged system might be valid even where the synapse are fast.

The top picture in figure 1.17 shows that the averaged system has difficulties following the fast evolution of the exact system. Indeed, in this regime the approximation (1.26) breaks down and the averaged system become less accurate.

When the number of oscillation in the inputs m goes beyond 10 then the second input period is larger than the size of the time window θ. Therefore, these oscillations are averaged out in both systems and the accuracy improves as shown in the bottom picture of figure 1.17. There seems to be a similar effect for the speed of the synapses: beyond a critical value of the synapse time constant $\varepsilon_w^* = 10^{-2}$ the fast variations are averaged in both system thanks to the convolution window of size $\theta = 1001$.

1.2.2.6 Biological regime of McKean's neuron

As said in part 1.1.2.3, an important drawback of McKean's neuron is that the spikes look like square functions when the effective input $\mathbf{x}_\alpha$ is larger than 0. This problem can be avoided if the inputs and the connectivity are assumed small enough. In particular, the choice of I_0 and the range of the inputs are important parameters that control the proximity of the model to biology. Yet, we do not know exactly which values to choose for the parameters and a further study based on experimental data would be needed to assess the biological relevance of this result.

In this part, we show that in a certain regime the average system is close to be linear in the sense that the saturating parts of the sigmoid S_σ may not be involved in the neural computation. Indeed, in certain cases it is possible to compute the histogram of the effective inputs value for the exact system. It turns out this histogram may not span the entire sigmoid as shown in the right column of figure 1.18. Actually, the sigmoid may even be approximated by a linear function on the effective value of $\mathbf{x}_\alpha$.

Figure 1.18 shows two different choices for I_0 and the range of the inputs that can make this linear approximation valid (right column) or false (left column). We do not pretend that the biological reality corresponds more to the linear case than the sigmoidal case. Yet, we believe it is a reasonable motivation to study the linear system defined by system (1.29) with S_σ being a linear function.

In this section we have assumed so far that the connectivity was constant, we break this assumption and try to generalize the approach to learning neural networks in the following.

1.2.3 A tentative to average a learning network

Unlike the previous section, we assume here that the connectivity is being learned according to the STDP learning rule (1.13). We intend to derive an

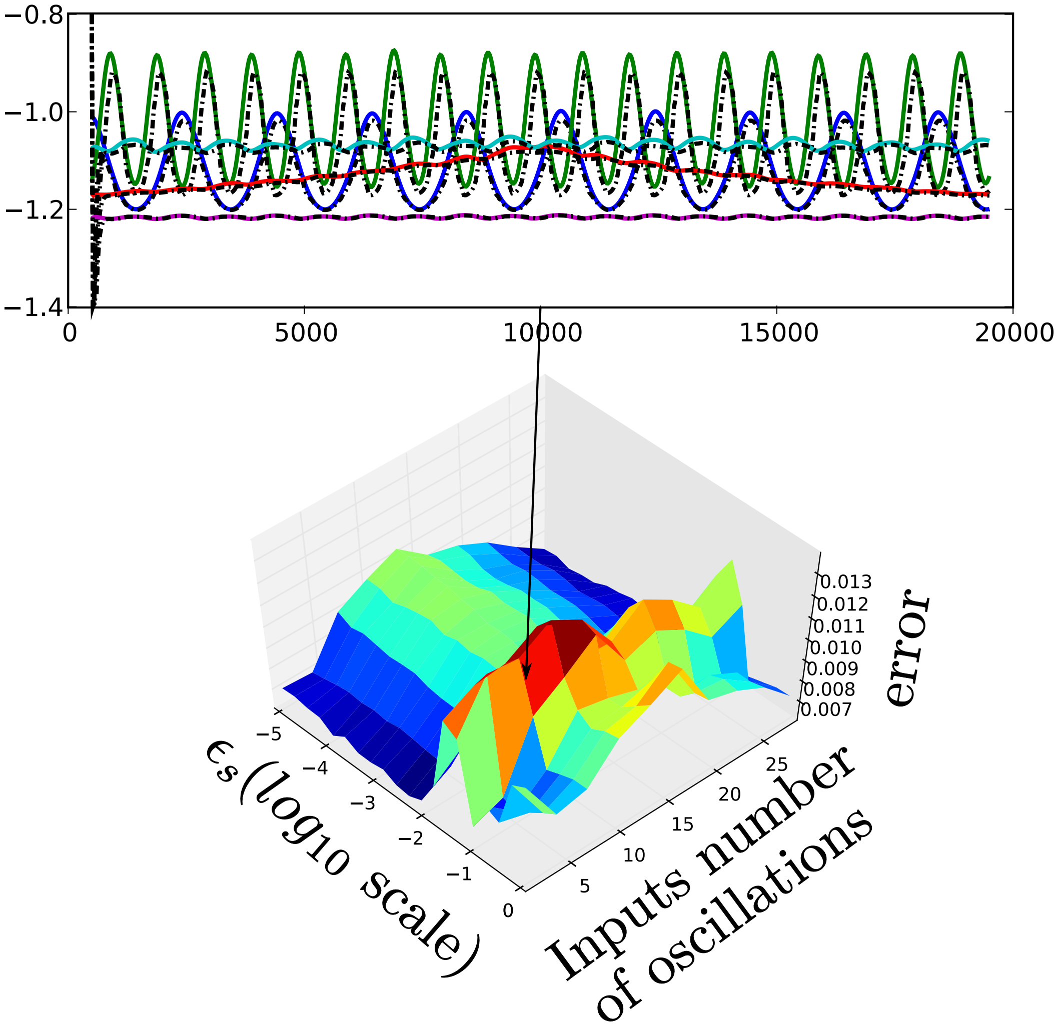


Figure 1.17: Comparison of the exact and averaged systems when the speed of the synapse and the speed of the connectivity vary. (top) Time evolution of the populations firing rates of the exact (plain colored curves) and averaged (dashed black) system with the default parameters except that the number of oscillations of the inputs is $m = 10$ and $\varepsilon_w = 0.05$. (bottom) Error between the two models when the speed of the effective inputs varies. There are two way it can vary: the inputs speed varies and/or the time scale of the synapses varies.

averaged equation in the same spirit as in the previous part for system (1.15) where the membrane potential and the connectivity are coupled. Unfortunately, we had not enough time to formalize this approach precisely so we will

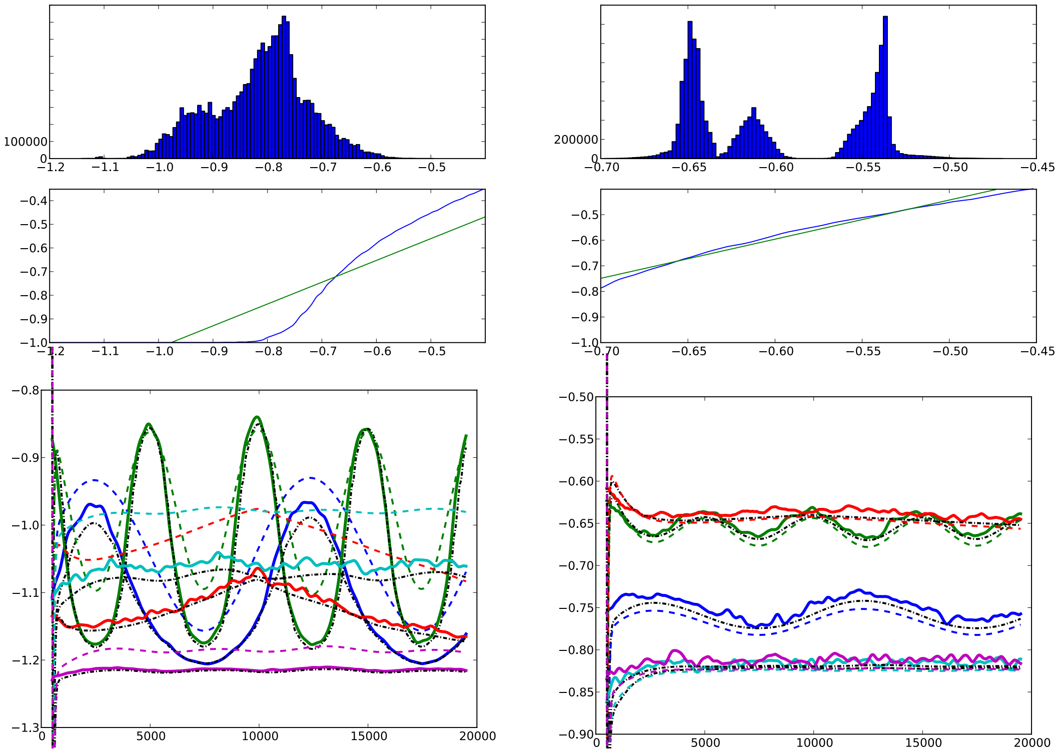


Figure 1.18: Comparison of the exact, averaged and linear system for two kinds of inputs: (left column) Default input with $I_0 = I^*$ (see equation (1.7)). (right column) The input has a smaller range and is centered above Muratov's critical input: we choose the default input divided by 10 with $I_0 = 6$. (top row) Histograms of the effective input $\mathbf{x}_\alpha$ in the exact system. (middle row) The effective sigmoid S_σ is shown in blue for the values corresponding to the histogram above. In green we have plotted a linear approximation of the effective sigmoid on the histogram interval. (bottom row) Comparison of the solutions of the exact (plain colored curves), averaged (dot-dashed black curves) and linear (dashed color curves) systems. The parameters have their default value.

only sketch the main arguments that may lead to a solution.

There are two links we need to investigate to incorporate learning to the previous approach: the role of the evolving connectivity on the averaging of the activity and, reciprocally, the modification of the connectivity under the influence of the activity. Under the assumption that learning occurs on a much slower scale than the activity, i.e. $\varepsilon_J \ll 1$, the first link is trivial. Indeed, at the scale of the activity the connectivity can be though of as a constant so that the derivation of the previous part holds. Therefore, we only need to understand the way the connectivity is averaged under the influence of the activity to close the loop.

Therefore, the question is: Is it possible to express the evolution of $\mathbf{J}_{\alpha\beta}$, the averaged connection between the population α and β, as a function of the firing rate of population α and β only? Indeed, if it is true this would close the average learning system and our goal would be reached. We believe this question is very important since it corresponds to knowing whether the learning mechanisms intrinsically depend on the timing of the spikes or if a firing rate approach might capture the essence of learning. Note that the latter does not discard the fact that the information processing in the brain may be based on the times of the spikes. It only suggests that the modification of the synaptic strength might be understood at the rate level. Of course, it contradicts the terminology itself since the learning rule we introduced in part 1.1.4 is called "spike timing depend plasticity". But it is not absurd since the firing rates might gather enough information about the spike timing to be considered as the exclusive information vector for learning.

We start by recalling the microscopic learning equation between two neurons i and j in populations α and β respectively:

$$\dot{\mathbf{J}}_{ij} = a_+\bar{\mathbf{v}}_i(\bar{\mathbf{v}}_j * h_{\tau_l}) - a_-(\bar{\mathbf{v}}_i * h_{\tau_l})\bar{\mathbf{v}}_j - \kappa\mathbf{J}_{ij}$$

where $\bar{\mathbf{v}}_i = \mathbf{v}_i - v_{\text{rest}}$ $h_{\tau_l}(t) = \frac{1}{\tau_l}e^{-\frac{t}{\tau_l}}H(t)$ with H the Heaviside function. Because $\varepsilon_J \ll 1$ it is clear that $\mathbf{J}_{\alpha\beta}$ can be approximated by

$$\mathbf{J}_{\alpha\beta}(t) \simeq \left(\frac{1}{n^2}\sum_{i\in\alpha,\ j\in\beta}\mathbf{J}_{ij}\right) * g \tag{1.31}$$

When $n \to +\infty$ this leads to

$$\dot{\mathbf{J}}_{\alpha\beta} = a_+\Big(\mathbb{E}(\bar{\mathbf{v}}_\alpha)\big(\mathbb{E}(\bar{\mathbf{v}}_\beta) * h_{\tau_l}\big)\Big) * g - a_-\Big(\big(\mathbb{E}(\bar{\mathbf{v}}_\alpha) * h_{\tau_l}\big)\mathbb{E}(\bar{\mathbf{v}}_\beta)\Big) * g - \kappa\mathbf{J}_{\alpha\beta}$$

We might as well replace h by $\tilde{h} = h_{\tau_l} * g$ in the previous equation. Indeed, this does not significantly change the profile of the learning because g is a symmetric function (typically a Gaussian or a step function).

To close the equation as a function of $\bar{\boldsymbol{\nu}}$, the main problem consists in showing that

$$\Big(\mathbb{E}(\bar{\mathbf{v}}_\alpha)\big(\mathbb{E}(\bar{\mathbf{v}}_\beta) * \tilde{h}_{\tau_l}\big)\Big) * g \simeq \big(\mathbb{E}(\bar{\mathbf{v}}_\alpha) * g\big)\big(\mathbb{E}(\bar{\mathbf{v}}_\beta) * \tilde{h}_{\tau_l}\big) \tag{1.32}$$

- **First tentative:**
Actually, the previously relation is rigorously true if $\mathbb{E}(\bar{\mathbf{v}}_\beta) * \tilde{h}_{\tau_l}$ is a constant function. Therefore, we can observe that $\tilde{h}_{\tau_l} = h_{\tau_l} * g$ is slower than g, so that it can be considered that $\mathbb{E}(\bar{\mathbf{v}}_\beta) * \tilde{h}_{\tau_l}$ is almost constant on $[t - \frac{\theta}{2}, t + \frac{\theta}{2}]$ for all $t \in \mathbb{R}_+$. Therefore, it seems reasonable to consider that (1.32) holds.

- **Second tentative:**
Here, we assume that the neurons have a pure spiking behavior, i.e.

$$\bar{\mathbf{v}}_i(t) = \sum_k \delta(t - t_k^{(i)})$$

where the $t_k^{(i)}$ are the times of the spikes of neuron i. Of course, this assumption only captures the spiking behavior of the neurons in a caricatural (and rigorously wrong) way. Yet, we think it is a interesting first step to extend the spatial averaging methods to learning neural networks.

Under this assumption, observe that

$$\begin{aligned}\bar{\mathbf{v}}_i(t)\big(\bar{\mathbf{v}}_j * \tilde{h}\big)(t) &= \sum_k \delta(t - t_k^{(i)}) \sum_q \tilde{h}(t - t_q^{(j)}) = \sum_{k,q} \delta(t - t_k^{(i)}) \tilde{h}(t_k^{(i)} - t_q^{(j)}) \\ &= \sum_k \Big(\sum_q \tilde{h}(t_k^{(i)} - t_q^{(j)}) \Big) \delta(t - t_k^{(i)})\end{aligned}$$

Because, we are interested by the value of $\bar{\mathbf{v}}_i\big(\bar{\mathbf{v}}_j * \tilde{h}\big) * g$ at time t and g is supported on $]-\frac{\theta}{2}, \frac{\theta}{2}[$, we naturally focus on the spikes emitted during the time interval $[t - \frac{\theta}{2}, t + \frac{\theta}{2}]$, so that we can write

$$\bar{\mathbf{v}}_i(t)\big(\bar{\mathbf{v}}_j * \tilde{h}\big)(t) = \sum_{t_k^{(i)} \in [t - \frac{\theta}{2}, t + \frac{\theta}{2}]} \Big(\sum_q \tilde{h}(t_k^{(i)} - t_q^{(j)}) \Big) \delta(t - t_k^{(i)})$$

Such that

$$\Big(t \mapsto \bar{\mathbf{v}}_i(t)\big(\bar{\mathbf{v}}_j * \tilde{h}\big)(t)\Big) * g = \sum_{t_k^{(i)} \in [t - \frac{\theta}{2}, t + \frac{\theta}{2}]} \Big(\sum_q \tilde{h}(t_k^{(i)} - t_q^{(j)}) \Big) \Big[\Big(t \mapsto \delta(t - t_k^{(i)})\Big) * g \Big]$$

We now claim that the sum $\sum_q \tilde{h}(t_k^{(i)} - t_q^{(j)})$ does not really depend on $t_k^{(i)} \in [t - \frac{\theta}{2}, t + \frac{\theta}{2}]$ so that, for $t_k^{(i)} \in [t - \frac{\theta}{2}, t + \frac{\theta}{2}]$

$$\sum_q \tilde{h}(t_k^{(i)} - t_q^{(j)}) \simeq \sum_q \tilde{h}(t - t_q^{(j)})$$

> This assumption is supported by several arguments: (i) during the time interval neither the input nor the connectivity change significantly suggesting that the spike times may be considered stationary (ii) any pair of neurons i and j is independent (propagation of chaos), excluding causation effects between i and j (iii) the function $\tilde{h}$ is almost constant on the interval. If the previous approximation is considered to be valid then summing over all the neurons in population $\alpha = p(i)$ and $\beta = p(j)$ scaling by $\frac{1}{n^2}$ leads to approximation (1.32).

Such that it seems reasonable to have

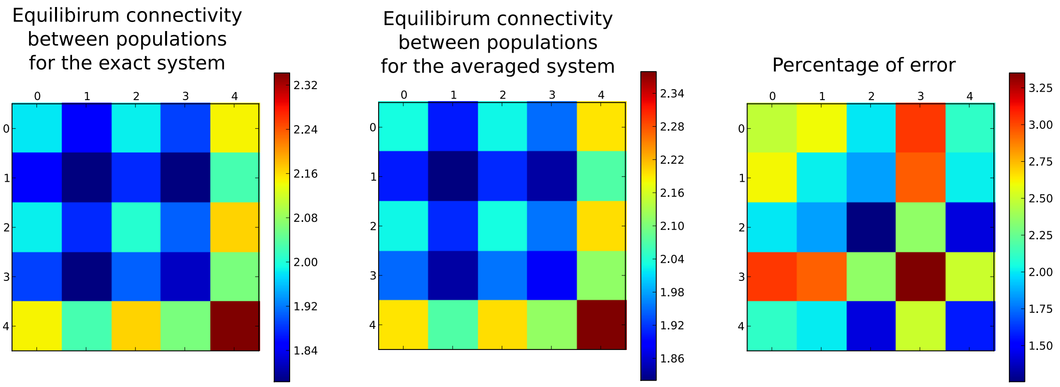
$$\dot{\mathbf{J}}_{\alpha\beta} = a_+ \bar{\boldsymbol{\nu}}_\alpha \left(\bar{\boldsymbol{\nu}}_\beta * h_{\tau_l}\right) - a_- \left(\bar{\boldsymbol{\nu}}_\alpha * h_{\tau_l}\right) \bar{\boldsymbol{\nu}}_\beta - \kappa \mathbf{J}_{\alpha\beta} \tag{1.33}$$

A numerical simulation to assess the accuracy of approximating the learning rule. Figure 1.19 shows equilibrium connectivities for a learning neural network in the default parameters, inputs and connections (except $T = 10000$). We alos chose the following values for the learning related parameters: $\tau_l = 1$, $\kappa = 1$, $a_+ = 2$ and $a_- = 1$. It compares the connectivity between the populations of neurons in two cases: (left) the connections of the exact system of $5 \times 200 = 100$ neurons are averaged a posteriori and (middle) the connections between the populations of the averaged system of 5 populations. Actually, we computed these matrices using the entire history of the activity which was simulated with a constant connectivity. This is an approximation which is based on the fact that learning is very slow.

We observe that the difference on each link is less than 4% which is an encouraging result which would need to be developped as a perspective. There are two sources of errors that prevent the comparison to be perfect: (i) the exact an averaged system are slightly different as shown in the previous simulation in section 1.2.2.5. (ii) the averaging of the learning rule itself is not perfect and induces some small errors.

1.2.4 Summary, conclusions and immediate extensions

In this section we have derived an averaged system (1.29) approximating the evolution of the firing rate of coupled populations of McKean neurons described by (1.15). The derivation was semi-analytic and was mainly based on an assumption of slow synapses. We have numerically explored the limits of


Equilibirum connectivity between populations for the exact system

Equilibirum connectivity between populations for the averaged system

Percentage of error

Figure 1.19: (left) equilibrium connectivity (averaged) between each population of the exact system according to the learning rule (1.2.3) . (middle) equilibrium connectivity between each unit of the averaged system according to the learning rule (1.33). (right) This is the error in percent between the two equilibirum connectivities shown in the left and right pictures.

the approximation and we believe we have shown it is accurate and robust to the variation of parameters.

We have also suggested that it is possible to extend this approach to a slowly learning neural network. This means we have defined an approximate equation for the evolution of the averaged connectivity between populations starting from the canonical STDP learning. It turns out the structure of the macroscopic learning rule is the same as the microscopic case.

This final averaged system is

$$\boxed{\begin{aligned} \dot{\boldsymbol{\nu}} &= -\bar{\boldsymbol{\nu}} * (l\delta + h_{\tau_w}) + \tilde{S}_\sigma\left(\tfrac{1}{p}\mathbf{J}\cdot(\bar{\boldsymbol{\nu}} * h_{\tau_s}) + \mathbf{u} * g\right) \\ \dot{\mathbf{J}} &= a_+\bar{\boldsymbol{\nu}} \otimes \left(\bar{\boldsymbol{\nu}} * h_{\tau_l}\right) - a_-\left(\bar{\boldsymbol{\nu}} * h_{\tau_l}\right) \otimes \bar{\boldsymbol{\nu}} - \kappa\mathbf{J} \end{aligned}} \tag{1.34}$$

1.2.4.1 Links with previous results

Many authors have introduced derivations from spiking to rate-based equations, see [Amit and Tsodyks 1991, Rinzel and Frankel 1992, Shriki et al. 2003, Camera et al. 2004, Ostojic et al. 2009] based on methods quite similar to the material in part 1.2.1.1. Our approach is closer to Ermentrout's work in [Ermentrout 1994], in the sense that we relate two dynamical systems: the exact and the averaged, which makes it possible to compare them dynamically.

We believe it gives a generic way to address the averaging of spiking neurons whatever their model (although we only applied it on McKean neurons) and we leave this task as a persepctive.

The derivation involved the computation of a sigmoidal function parametrized by the level of noise in the network of neurons. In the same spirit as traditional results briefly reviewed in part 1.2.1.1, we have observed that the noise has the effect of smoothing the non-linearity. For appropriate inputs, we show that the averaged equation could be considered linear. However, we leave as an open question the biological relevance of this linear approximation.

We believe our approach is more based on biology and needs less hypotheses than the heuristic derivation motivating the Hopfield and Wilson-Cowan equations. In particular, we do not assume a priori that there is a sigmoidal relationship between the frequency and the total input to a neuron: this relation emerges from the computations. Eventually, we define an average system (1.29), which is quite similar to the usual synaptic-drive equation (1.19). Unlike Wilson-Cowan equations systems (1.29) parameters relate to the initial spiking neurons' parameters.

1.2.4.2 Perspectives

This approach raises many questions and would need a significant amount of additional work to be made reliable enough to be taken as a starting point by other scientists. Here, we highlight some directions to develop this result in the future.

Are synapses slow enough? The derivation of the averaged system (1.29) is based on different assumptions which need to be biologically checked. For instance, is the hypothesis of slow synapses broken by fast AMPA receptors? Is it possible to improve the derivation to get a more complicated but more relevant system?

Non homogeneity of the network In this derivation, we have assumed that all the neurons and synapses were the same. Is it possible to break this assumption and consider populations made of several types of neurons? Can we model the diversity of neurons and synapses observed experimentally with this approach?

Dynamical properties of such equations The study of the dynamics of Wilson-Cowan equations has been a hot topic of research (see [Coombes 2005] for a review). Since the averaged system (1.29) is slightly different from these equations, it seems necessary to develop a rigorous and systematic analysis of its dynamical properties. The difference between our model and the synaptic-drive equation (1.19) is essentially the presence of terms with (exponential) delays. Does it change significantly the dynamical behavior of the equations? We believe it does not, yet it would be necessary to check this rigorously.

Including learning to the derivation In section 1.2.3, we have tried to extend the derivation of an averaged system to a learning neural network. However, we are not satisfied yet with the derivation and we think it would be necessary to find a more rigorous path. We think this point has to be studied deeply since it addresses the question of the usefulness of spikes for learning. Are they a fundamental mechanism which cannot be neglected in the study of learning. Or can we grab the essence of learning in a formalism without spiking neurons? Actually, we believe the latter hypothesis is the best and we have tried to convince the reader in section 1.2.3. Besides, there is a huge field of research devoted to learning rules in rate-models, which we try to contribute to in the rest of the thesis. Therefore, it is a crucial question which needs to be rigorously addressed to build a theory of neural systems from first principles.

Chapter 2

The slow dynamics of learning in fast rate-based models

Overview

This chapter is devoted to studying the dynamics of rate-based learning neural networks under the assumption that learning is very slow compared to the activity. After having introduced the appropriate framework and the different learning rules, we use various temporal averaging methods to get a reduced system asymptotically governing the behavior of the connectivity. Finally, we prove it always converges to an equilibrium point under some assumptions. In a case of linear activity, we can even compute explicitly an expansion for the equilibrium connectivity.

In this chapter sections 2.1 and 2.2 are background, whereas sections 2.3 and 2.4 are original.

Résumé

Ce chapitre porte sur l'étude de la dynamique de réseaux de neurones à taux de décharge couplés à un mécanisme d'apprentissage. L'hypothèse de base dans ce chapitre est que l'apprentissage est beaucoup plus lent que l'activité des neurones. Après avoir revu les différentes règles d'apprentissage et avoir introduit le cadre mathématique approprié, nous utilisons plusieurs méthodes de moyennisation temporelle pour obtenir un système réduit gouvernant l'évolution de la connectivité. Finalement nous prouvons que ce système converge toujours vers un unique point d'équilibre sous certaines conditions techniques. Dans le cas où l'activité du réseau est linéaire nous fournissons une formulation explicite de la connectivité d'equilibre.

Dans ce chapitre les sections 2.1 et 2.2 sont issues de la littérature, alors que les sections 2.3 et 2.4 sont originales.

Collaborations, publications and ackowledgements

This part is mainly based on collaborations with Paul Bressloff for section 2.3 (which lead to a paper in Neural Computation) and Gilles Wainrib for section 2.4 (lead to a paper in the journal SIAM mutliscale modeling). In particular, my contribution was to use a theorem for generic non-autonomous systems derived in [Wainrib 2011] and presented in appendix B in the case of learning neural networks. To do so, I have developed a mathematical result which was general enough to be published as an autonomous paper in the Comptes rendus de l'académie des sciences, it is attached in appendix E.

Contents

Rate-based models of populations of neurons are widespread because of their dynamical simplicity. In these models, the evolution of the network's activity is smooth as opposed to spiking neural networks. In the previous chapter, we have shown how rate-based and spiking networks can be related: rate based models are spatial averages of populations of spiking neurons. However, it is habitual refer to the building blocks of rate-based models not as populations of neurons but simply as neurons.

There exists a formal way to define a neuron as a function. This is qualitatively different from the dynamical models considered in this thesis since it gives a input-output aspect of neurons and therefore mainly applies to feed-forward networks. Yet, the intrinsic mechanism of propagation of information through a weighted network can still be taken into account in this formalism. Indeed, the activity of a readout neuron written $y \in \mathbb{R}$, over a layer of multiple neurons, whose activity is $\mathbf{x} \in \mathbb{R}^n$, may be described by $y = S(\sum_{i=1}^n \mathbf{j}_i \mathbf{x}_i)$ where S is a nonlinear function, typically a Heaviside with offset or a sigmoid, and $\mathbf{j} \in \mathbb{R}^n$ is the vector of connections of the bottom layer to the readout neuron. This approach is at the heart of the traditional research in *neural networks* which belongs to the field of computer science and was originated by [McCulloch and Pitts 1943]. It is widely used for supervised learning in hierarchical networks which are out of the scope of this thesis. However, this also an assumption in early works about unsupervised learning in hierarchical structure [Oja 1982, Bienenstock et al. 1982, Földiák 1991, Miller and MacKay 1994, Wallis and Baddeley 1997] which are extensively reviewed in [Hertz et al. 1991].

There are two canonical dynamical rate-based models: the Hopfield or voltage-based model and the Wilson-Cowan or synaptic-drive-based. The first was introduced in [Amari 1977, Hopfield 1982, Hopfield 1984] and is of the form $\dot{\mathbf{v}}_i = -\mathbf{v}_i + \sum_{i=1}^n \mathbf{J}_{ij} S(\mathbf{v}_j) + \mathbf{u}_i$. The second was introduced in [Wilson and Cowan 1972] and is of the form $\dot{\mathbf{z}}_i = -\mathbf{z}_i + S(\sum_{i=1}^n \mathbf{J}_{ij} \mathbf{z}_j) + \mathbf{u}_i$. Their heuristic derivation was detailed in 1.2.1.2. We show in the following that they are identical after change of variable. It is sometimes considered that the inputs are not a dynamical forcing but only a specification of the initial conditions. This corresponds to the denomination *attractor neural networks*. We do not consider these kind of networks since the inputs are introduced as a non homogeneous term, yet the mathematical structure of our problem is quite similar. The study of learning rules in this framework has also a been a topic of interest mainly in the wake of Hopfield's works ([Dong and Hopfield 1992, Hopfield 2007]).

The dynamical study of these spatially extended networks, without learning, has been very rich starting from the contributions of Wilson and Cowan ([Wilson and Cowan 1973]). It is often assumed that the neurons are continuously located on a given geometrical shape, e.g. the plane $\mathbb{R}^2$, so that the connectivity can be defined to be convolutional. These continuous models are called *neural fields*. In this new framework the tools from Fourier theory, in particular, have led to substantial results of to characterize the solutions of such equations and their stability, [Pinto and Ermentrout 2001, Laing et al. 2002, Folias and Bressloff 2004, Faugeras et al. 2009b, Veltz and Faugeras 2009] and see [Coombes 2005] for a review. A few papers have dealt with learning the feed-forward connectivity to a recurrent neural field without learning [Takeuchi and Amari 1979, Bressloff 2005].

There have been different approaches to unsupervised learning in neural networks. Some of them were based on the reduction of redundancy [Barlow 1989, Barlow 2001], or on the maximization of mutual information [Linkser 1988, Linkser 1992, Atick and Redlich 1990] or the maximization of sparseness [Olshausen and Field 1996]. Most of these approaches are based on a functional principle but may be partially implemented with explicit learning rules tuning the connectivity. Another important approach to unsupervised learning, which is quite close to the approach we develop in the following, is the Boltzmann-machine approach [Ackley et al. 1985, Amari et al. 2002]. They show that in an idealized network close to the Hopfield equations a certain learning rule corresponds to the opposite of the gradient of the Kullback-Leibler divergence between the law of the inputs and that generated by the network spontaneous activity. This is a mathematically dramatic result although it implies huge computations to compute the learning rule. We were really inspired by the philosophy of this result which shows that the network copies the inputs. Chapters 2 and 3 present a similar approach where we start from biological learning rules instead.

The topic of this chapter is the study of dynamical unsupervised learning according to correlation or causation based learning rules inspired from biology in recurrent rate-based networks and reviewed in [Dayan and Abbott 2001, Gerstner and Kistler 2002b, Gerstner and Kistler 2002a, Chen et al. 2007].

This chapter consists in a mathematically rigorous analysis of the dynamics of rate-based learning networks. A priori, the main functional question that we may ask ourself is: does learning lead to a fixed understanding of the world? We believe a good learning rule should converge to a single equilibrium connectivity and should not oscillate or diverge. We will see that in our non-autonomous framework (the inputs is constantly moving) the notion of equilibrium point is not trivial. Basically, we rely on the slow-fast dynamics of learning neural networks to go around this problem.

The first section consists in defining and reviewing the possible models for rate-based learning neural networks significantly influenced by the references [Dayan and Abbott 2001, Gerstner and Kistler 2002b, Gerstner and Kistler 2002a]. The second part is a first step toward the slow-fast treatment of the problem where the inputs are assumed to be slow. The third section generalizes this approach to fast inputs and noisy neurons and for different learning rules.

2.1 Rate-based learning neural networks

We now introduce a large class of rate-based stochastic neuronal networks with learning models. They are defined as coupled systems describing the simultaneous evolution of the activity of $n \in \mathbb{N}$ neurons and the connectivity between them. We define $\mathbf{v}(t) \in \mathbb{R}^n$ the *activity field* of the network and $\mathbf{J}(t) \in \mathbb{R}^{n\times n}$ the *connectivity matrix* at time $t \in \mathbb{R}_+$.

2.1.1 Activity field

The activity field is assumed to evolve according to

$$d\mathbf{v}_i = \left(\mathcal{F}(\mathbf{v}_i) + \mathcal{S}\Big(\sum_{j=1}^{n} \mathbf{J}_{ij}\mathcal{H}(\mathbf{v}_j) + \mathbf{u}_i(t)\Big)\right) dt + \sum_{j=1}^{n} \Sigma_{ij}(\mathbf{v}, \mathbf{J}) dB_j(t)$$

where the function $\mathcal{F}$ characterizes the intrinsic dynamical behavior of the neurons. In this rate-based approach it is often considered to be linear with a negative leak constant (and possibly with time delays). Functions $\mathcal{S}$ and $\mathcal{H}$ characterize the non-linear communication term. In the most general case, these three functions are more than functions from $\mathbb{R} \to \mathbb{R}$. In fact, their argument might be the entire history of the variable (not only its value at time t). This makes it possible to take into account functional or time-delayed problems. Mathematically speaking, they are functionals from $C^1(\mathbb{R}_-, \mathbb{R}^n)$ to $\mathbb{R}^n$. It turns out this formalism does not pose substantial problems and allows a useful generalization of traditional models. Actually, these three functions may depend on the neurons and therefore should be indexed by $i \in \mathbb{N}$, however, we choose not to write them for simplicity. The noise-related matrix Σ is linked to the spatial correlations of the white noise. It this general framework, it may depend on $\mathbf{v}$ and $\mathbf{J}$.

Abusively redefining $\mathcal{F}$, $\mathcal{S}$ and $\mathcal{H}$ as vector valued operators corresponding to the element-wise application of their real counterpart leads to the following vector valued system.

$$d\mathbf{v} = \big(\mathcal{F}(\mathbf{v}) + \mathcal{S}\big(\mathbf{J}.\mathcal{H}(\mathbf{v}) + \mathbf{u}(t)\big)\big)\, dt + \Sigma(\mathbf{v}, \mathbf{J}).dB(t) \tag{2.1}$$

where the "." operator is a matrix-vector multiplication. This system is a non-autonomous, stochastic, possibly time-delayed, non-linear, slow-fast system.

Voltage-based or synaptic-drive-based equations In the previous chapter, we introduced in part 1.2.1 two types of rate-based models called the voltage-based (or Hopfield) model in (1.18) and the synaptic-drive-based (or Wilson-Cowan) model in (1.19). After a change of time, they are

$$\begin{array}{lrcl} \textbf{voltage-based} & \dot{\mathbf{v}} & = & -\mathbf{v} + \mathbf{J}\cdot S\big(\mathbf{v}\big) + \mathbf{u}(t) \\ \textbf{synaptic-drive-based} & \dot{\mathbf{z}} & = & -\mathbf{z} + S\big(\mathbf{J}\cdot\mathbf{z} + \mathbf{u}(t)\big) \end{array}$$

where S is traditionally chosen to be a positive sigmoidal function.

In the formalism of (2.1), they correspond to the choices $\mathcal{F} = -I_d$, $\mathcal{S} = I_d$, $\mathcal{H} = S$ and $\mathcal{F} = -I_d$, $\mathcal{S} = S$, $\mathcal{H} = I_d$ respectively.

When the (invertible) connectivity and the inputs are constant, the change of variable $\mathbf{v} = \mathbf{J}.\mathbf{z} + \mathbf{u}$ makes it possible to go from one to the other and reciprocally. It is just a choice of modelisation to select one or the other.

Frequency based system from averaging In the previous chapter, we also derived an averaged equation of the behavior of populations of mcKean neurons, see 1.2.2. It leads to system (1.34), which can be written as follows without learning where we simplified slightly the notations ($\bar{\boldsymbol{\nu}}$ is replaced by $\boldsymbol{\nu}$, $\frac{1}{p}\mathbf{J}$ is replaced by $\mathbf{J}$, $\mathbf{u} * g$ is replaced by $\mathbf{u}$, and $\tilde{S}_\sigma$ is replaced by $\tilde{S}$).

$$\textbf{frequency-based} \quad \dot{\boldsymbol{\nu}} = -\boldsymbol{\nu} * (l\delta + h_{\tau_w}) + \tilde{S}\Big(\mathbf{J}\cdot(\boldsymbol{\nu} * h_{\tau_s}) + \mathbf{u}\Big)$$

This corresponds to $\mathcal{F}(\mathbf{v}) = -\mathbf{v} * (l\delta + h_{\tau_w})$, $\mathcal{H}(\mathbf{v}) = \mathbf{v} * h_{\tau_s}$, $\mathcal{S} = \tilde{S}$ defined in equation (1.30) and $\boldsymbol{\Sigma} = 0$.

This system looks like the synaptic-drive-based equation and we would like to find the associated (pseudo) voltage-based equation. Therefore, we assume the connectivity is fixed and invertible and define $\mathbf{a} = \mathbf{J}\cdot(\boldsymbol{\nu} * h_{\tau_s}) + \mathbf{u}$. To invert this relation observe that $\boldsymbol{\nu} * h_{\tau_s} = \mathbf{J}^{-1}.(\mathbf{a} - \mathbf{u}(t))$. We recall $h_{\tau_s}(t) = \frac{1}{\tau_s} e^{-\frac{t}{\tau_s}} H(t)$ whose Fourier transform is $\hat{h}_{\tau_s}(\xi) = \frac{1}{1+2i\pi\tau_s\xi}$, so that the convolution of $\boldsymbol{\nu} * h_{\tau_s}$ with the inverse Fourier transform of $\xi \mapsto 1 + 2i\pi\tau_s\xi$ (which we write $h_{\tau_s}^{(-1)}$) is $\boldsymbol{\nu}$. This leads to $\boldsymbol{\nu} = \mathbf{J}^{-1}.\big(\mathbf{a} - \mathbf{u}\big) * h_{\tau_s}^{(-1)}$.

This leads to $\dot{\mathbf{a}} = \mathbf{J}.\big(-\boldsymbol{\nu} * (l\delta + h_{\tau_w}) + \tilde{S}(\mathbf{a})\big) * h_{\tau_s} + \dot{\mathbf{u}}$, i.e.

$$\textbf{pseudo voltage-based} \quad \dot{\mathbf{a}} = -\mathbf{a} * (l\delta + h_{\tau_w}) + \mathbf{J}.\tilde{S}(\mathbf{a}) * h_{\tau_s} + (\mathbf{u} + \dot{\mathbf{u}})$$

Note that we have not supposed that the inputs were fixed in this derivation. The link between voltage-based and synaptic-drive-based with evolving inputs would give the same result i.e. replace $\mathbf{u}$ by $\mathbf{u} + \dot{\mathbf{u}}$ when going from the synaptic-drive-based to the voltage-based.

We have also assumed a fixed connection which might look irrelevant to our learning framework, however, the learning will be assumed so small that this equivalence will hold.

In the previous chapter, we motivated the definition of such rate-based models to account for the evolution of population of neurons. In a framework when the number of neurons tended to infinity, the averaged equations where deterministic. In the formalism we have just introduced there is some additional noise. Actually this noise could come from finite size effects. Indeed, if the number of neurons in the populations in not infinite but large then the evolution of the network may be described by a noisy version of the averaged equation for an infinity of neurons. This area of research is in great development [Bressloff 2009, Buice et al. 2010]. Although the "shape" of the noise could be tuned to represent better these finite size effects, we won't go in this level of detail and mainly deal with simple additive noise with a constant matrix $\mathbf{\Sigma}$.

2.1.2 Connectivity matrix

The connectivity matrix is assumed to evolve according to a slow unsupervised learning rule. This reads

$$d\mathbf{J} = \varepsilon\ \boldsymbol{\chi} \odot G(\mathbf{v}, \mathbf{J})\ dt \tag{2.2}$$

where ε is very small. The matrix $\boldsymbol{\chi} \in \mathbb{R}^{n \times n}$ is made of binary coefficients, i.e. $\boldsymbol{\chi}_{ij} \in \{0, 1\}$. It represents the physical connectivity of the network. Indeed, there might by no physical axons going from one neuron to another. Therefore, the learning rule cannot change this non-existing link. The component $\boldsymbol{\chi}_{ij}$ is 1 if there is a link from j to i and it is 0 if there is no link. Besides, the initial connectivity of non physically connected neurons is assumed to be null.

The operator $\odot$ is the Hadamard or element-wise product:

$$\begin{array}{lcl} \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} & \rightarrow & \mathbb{R}^{n\times n} \\ \mathbf{X}, \mathbf{Y} & \mapsto & \mathbf{X} \odot \mathbf{Y} \text{ such that } \{\mathbf{X} \odot \mathbf{Y}\}_{ij} = \mathbf{X}_{ij}\mathbf{Y}_{ij} \end{array}$$

In the part 2.2, we review the traditional possibilities for the choice of the learning rule G.

2.1.3 Existence and Uniqueness of the solution

The first question that arises when considering the coupled system of equations (2.1) and (2.2) is about the well posedness of the system. Is there a unique solution to the coupled system? What is the maximal interval of definition of the solutions?

Actually the system must be set as an initial value problem or Cauchy-problem to address these questions. In this thesis, we assume that the initial time is 0 and the value of the variables on $\mathbb{R}_-$ is also null. Therefore, we rigorously define the system

$$\left\{ \begin{array}{lcl} d\mathbf{X} & = & f(t, \mathbf{X})dt + \tilde{\boldsymbol{\Sigma}}(\mathbf{X})dB(t) \\ \mathbf{X}(\mathbb{R}_-) & = & 0 \end{array} \right.$$

where $\mathbf{X} = (\mathbf{v}, \mathbf{J})$, $\tilde{\boldsymbol{\Sigma}} = (\boldsymbol{\Sigma}, 0)$ and $f : \big(t, (\mathbf{v}, \mathbf{J})\big) \in \mathbb{R} \times (\mathbb{R}^n \times \mathbb{R}^{n\times n}) \mapsto \Big(\mathcal{F}(\mathbf{v}) + \mathcal{S}\big(\mathbf{J}.\mathcal{H}(\mathbf{v}) + \mathbf{u}(t)\big), \varepsilon\, \boldsymbol{\chi} \odot G(\mathbf{v}, \mathbf{J})\Big) \in \mathbb{R}^n \times \mathbb{R}^{n\times n}$.

Although, the formalism is rather general (non-linear, stochastic and time-delayed) the main requirement for the system to be well posed is the Lipschitzianity of the functions (see chap 2 of [Hale and Lunel 1993], [Mao 1997] and [Da Prato and Zabczyk 1992]).

Uniqueness If f is locally Lipschitz (i.e. Lipschitz on any compact subset of $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n\times n}$), there exists a unique solution to the Cauchy system above. This means $\mathcal{F}$, $\mathcal{S}$, $\mathcal{H}$, $\boldsymbol{\Sigma}$ and G have to be locally Lipschitz which will be verified in all the situations we are treating in this thesis.

Global existence The only remaining pathological behavior is the explosion in finite time. Actually, we cannot prevent this from happening in the most general case because Hebbian learning rules (detailed below) are quadratic rules. In particular, if $\mathcal{F} = -Id$, $\mathcal{S} = Id$, $\mathcal{H} = Id$, and $G(\mathbf{v}) = \mathbf{v} \otimes \mathbf{v}$ then the system explodes in finite time. Adding a strong linear decay to the learning rule may prevent this problematic behavior from happening, as shown in part 2.4. This implies a systematic dynamic analysis of the different learning rules which will shed light on their biological plausibility: a diverging rule is not biologically plausible and should be excluded from our analysis.

2.1.4 A word about the dynamics

This chapter addresses the dynamical behavior of such learning neural networks. But, what kind of dynamics do we expect to find?

From a functional point of view, we are interested in the statistical information the network has managed to extract from the inputs. Therefore, we need to assume that the inputs are evolving in time: a single frozen input in a learning system is irrelevant.

Besides, the network is expected to extract the regularities in these inputs. Therefore, they must be structured in some way and it makes no sense studying the system with random inputs. The easiest way to impose the existence of recurrent patterns in the inputs in a generic way is to assume they are periodic. A more general framework would include inputs evolving stochastically in a bounded energetic landscape.

In any case, the system is forced by time-evolving inputs. Therefore, talking about equilibrium points makes no sense. However, we expect the system to converge to a fixed "understanding" of the inputs, a final statistical knowledge about inputs. Therefore, we expect the connectivity (which is the variable for learning) to converge through the learning process.

The main ingredient to have a converging connectivity coupled with a input-driven activity is the separation of time-scales between the two variables. The activity is considered to be much faster than the connectivity. Therefore in sections 2.3 and 2.4, we will apply reduction methods for slow-fast systems before analyzing the dynamics.

2.2 A guided tour of learning rules

Over the last 60 years, the fields of experimental and theoretical neuroscience have been significantly influenced by Hebb's postulate:

> When an axon of cell A is near enough to excite cell B or repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased. [Hebb 1949]

All the rules below implement with certain subtleties Hebb's postulate. This is why this class of correlation/causation based rules are sometimes called Hebbian learning rules.

However, there is also a precise rule which is also called Hebbian learning. It corresponds to the simplest learning rule of the class, i.e. particular form of G, which only takes into account the correlation effects implied by Hebb's words above. Actually, Hebb quotation also introduced a notion of causality between the neurons. This causality link is at the heart of the STDP learning rule which we detail later.

In this thesis, we choose to call *Hebbian learning rule* the simplest learning rule presented above. We call *correlation-based learning rules* the class of rules built upon the Hebbian learning rule which only track the correlations between neurons. And we call *causation-based learning rules* those which take into account the causality links between neurons.

Appart from Hebbian learning, most of these rules were originally introduced in a feed-forward formalism. One of the main interest of this part is to extend these definitions to recurrent neural networks which is the framework of this entire thesis.

Before starting we need to introduce properties these learning rules may have which are qualitatively very important to classify them: (i) the locality of the rule (ii) the stability of the rule.

1. A learning rule is said to be **local** if the evolution of the connection from neuron j to neuron i only depends on the value of the connection and the activities of neurons i and j. Mathematically, this reads $G_{ij}(\mathbf{v}, \mathbf{J}) = G_{ij}(\mathbf{v}_i, \mathbf{v}_j, \mathbf{J}_{ij})$. This property is necessary for biological plausibility because it seems unlikely the synapse from neuron j to neuron i is influenced by the activity of other neurons (though the proximity

of synapse might lead to such behavior which is irrelevant to the underlying population approach we take here).

2. A learning rule is said to be **diverging** if it leads to the explosion of the connectivity variable. This description gathers the cases of explosion in finite and infinite time. Obviously, these rules are not biologically plausible.

The most common learning rules are often diverging whereas plausible learning rules should maintain homeostasis in the brain. Therefore, something more is needed to prevent the network from diverging: synaptic scaling [Abbott et al. 2000, Turrigiano and Nelson 2004]. As explained in this papers, different mechanisms correspon to the notion of synaptic scaling from artifically bounding the synaptic weights and adding an energetical constraint to the wieghts strength growth to heterosynaptic constraints or temporal normalizations. Although, we review some of these mechanisms in the following we will chose as simple linear decay embodying energetical constraints on the connections between population of neurons in the rest of the thesis.

In the following, we present three types of learning mechanisms. A learning rule can be a combination of these mechanisms.

2.2.1 Correlation-based and Hebbian rules

- The **Hebbian learning rule** was introduced in the context of linear networks, i.e. when the nonlinearity $\mathcal{F}$, $\mathcal{S}$ and $\mathcal{H}$ in equation (2.1) are the identity. It corresponds to choosing $G_{ij}(\mathbf{v}, \mathbf{J}) = \mathbf{v}_i \mathbf{v}_j$ or

$$G(\mathbf{v}) = \mathbf{v} \otimes \mathbf{v}$$

  in a vector form (where $\otimes$ the tensor product). This rule is obviously local. However, it is diverging because the common activation of two neurons lead to the strengthening of their connection which increase their activity and so on: the connectivity explodes. Besides, this is only a mechanism for the increase of the synapse strength.

- **Anti-Hebbian** learning simply corresponds to inserting a minus in front of the Hebbian rule. It has an opposed functional effect to the Hebbian case.

- The first idea to upgrade the Hebbian learning rule is to consider a rule made of the tensor product of two functions of the activity, i.e.,

  $$G(\mathbf{v}) = \phi_1(\mathbf{v}) \otimes \phi_2(\mathbf{v})$$

  where the functions $\phi_k : \mathbb{R}^n \to \mathbb{R}^n$ correspond to the element-wise application of a function (which we abusively write ϕ_k), i.e. $\phi_1(\mathbf{v})_i = \phi_1(\mathbf{v}_i)$. In Hebb's philosophy and to support the experiments on long term potentiation (see 1.1.4), $\phi_1(\mathbf{v}_i)$ should be a function the firing rate of neuron i. We detail below three examples for the choice of the ϕ_i:

  1. When the activity is voltage-based, we would expect $\phi_1 = \phi_2 = S$ since $S(\mathbf{v}_i)$ is assumed to be the firing rate of i see 1.2.1.2.
  2. **Rules with gating** make it possible for the synaptic strength to decrease. Post-synaptic gating $G(\mathbf{v}, \mathbf{J}) = (\mathbf{v} - \theta) \otimes \mathbf{v}$, and pre-synaptic gating $G(\mathbf{v}, \mathbf{J}) = \mathbf{v} \otimes (\mathbf{v} - \theta)$, involve the choice of a fixed threshold θ, which is often chosen somewhat arbitrarily. Sliding threshold rules (see part 2.2.3) go one step further and consider that this threshold is a function of the state and its past. This rule is local but does not prevent the system from exploding in certain cases.
  3. Bienenstock et al. improved the rules with gating to include saturation as well. In [Bienenstock et al. 1982], they proposed to choose

     $$G(\mathbf{v}, \mathbf{J}) = \phi(\mathbf{v}) \otimes \mathbf{v} \tag{2.3}$$

     where $\phi_i(\mathbf{v}) = \arctan\big(\mathbf{v}_i(\mathbf{v}_i - \theta)\big)$, as shown in figure 2.1. The **BCM learning rule** has more interesting features and part 2.2.3 deals with one of them. This rule is local. It is not clear yet what kind of dynamics it leads to.

- To prevent the Hebbian learning rule from diverging, a simple and effective idea is to add an elastic negative feedback to the connectivity strength increment. This rule is referred as the **Hebbian learning rule with linear decay, or M$_0$**. It is written as follows

  $$G(\mathbf{v}, \mathbf{J}) = -\kappa \mathbf{J} + \mathbf{v} \otimes \mathbf{v} \tag{2.4}$$

  where $\kappa \in \mathbb{R}_+$. This rule is local and not diverging if κ is sufficiently large (see part 2.4).

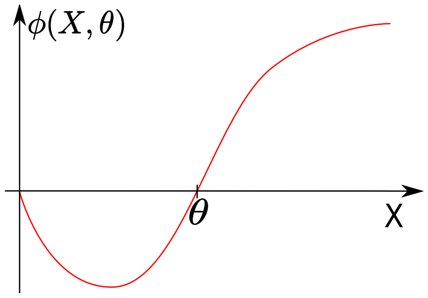


Figure 2.1: The function ϕ introduced by Bienenstock et al. [Bienenstock et al. 1982].

2.2.2 Heterosynaptic constraints

We introduce now other rules that are **non-local** and implement a notion of competition between synapses, [Malsburg and Cowan 1982, Oja 1982, Miller 1996], which we call heterosynapty, see section 1.5 of [Rolls and Deco 2002] for details. They implement mechanisms of spatial averaging of the synapses' strength. It corresponds to the fact that two neurons sending their axon to a third one may have an influence on each other. For instance, if the first one is very active and excites the target neuron, while the second is quiet, it is likely that the strength of the connection of the second neuron to the target neuron decreases.

These heterosynaptic learning rules are modifications of the Hebbian paradigm by subtracting a global term from function G which has the effect of projecting the connectivity onto a subspace of $\mathbb{R}^{n\times n}$, see figure 2.2. The impact of these constraints was successfully analyzed in [Miller and MacKay 1994] in the case of a linear perceptron (simplest feed-forward network). We follow the nomenclature introduced in [Miller and MacKay 1994]

1. **Subtractive normalization** (S_1) It consists in subtracting from the Hebbian learning rule the spatial mean of the incoming connection:

$$G_{ij}(\mathbf{v},\mathbf{J},\chi) = \mathbf{v}_i\mathbf{v}_j - \frac{\mathbf{v}_i\sum_{k=1}^{n}\chi_{ik}\mathbf{v}_k}{\sum_{k=1}^{n}\chi_{ik}} \tag{2.5}$$

With this rule the vector $\mathbf{a}^{\{1\}} \in \mathbb{R}^n$ such that $\mathbf{a}_i^{\{1\}} = \sum_{k=1}^{n}\mathbf{J}_{ik}$ is kept

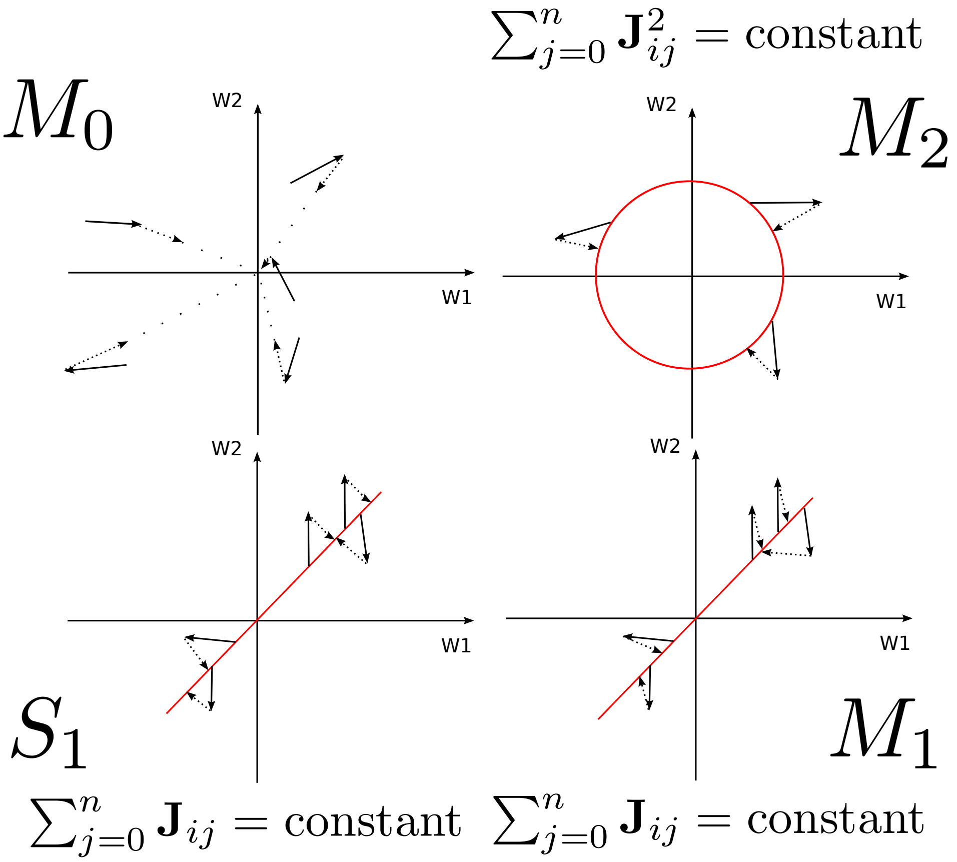

$\sum_{j=0}^{n} \mathbf{J}_{ij}^{2} = \text{constant}$
M_0
W2
W1
W2
M_2
W1
W2
W1
W2
W1
S_1
M_1
$\sum_{j=0}^{n} \mathbf{J}_{ij} = \text{constant}$
$\sum_{j=0}^{n} \mathbf{J}_{ij} = \text{constant}$

Figure 2.2: Shows the role of different constraints in a geometric way on two dimensions. Actually, these constraints correspond to n projections of subspaces of $\mathbb{R}^n$, one for each row of the connectivity matrix. The pictures above correspond to one of these projections. The learning rule M_0 is a linear decay term. The learning rules S_1, M_1, M_2 correspond to projection on invariant subspaces in red.

constant during learning. Indeed,

$$\begin{aligned}\frac{d}{dt}\mathbf{a}_i^{\{1\}} &= \sum_{k=1}^{n}\left(\boldsymbol{\chi}_{ik}\mathbf{v}_i\mathbf{v}_k - \frac{\boldsymbol{\chi}_{ik}\mathbf{v}_i\sum_{p=1}^{n}\boldsymbol{\chi}_{ip}\mathbf{v}_p}{\sum_{p=1}^{n}\boldsymbol{\chi}_{ip}}\right)\\ &= \mathbf{v}_i\left(\sum_{k=1}^{n}\boldsymbol{\chi}_{ik}\mathbf{v}_k - \frac{\sum_{k=1}^{n}\boldsymbol{\chi}_{ik}}{\sum_{p=1}^{n}\boldsymbol{\chi}_{ip}}\sum_{p=1}^{n}\boldsymbol{\chi}_{ip}\mathbf{v}_p\right) = 0\end{aligned}$$

In other words for the rule S_1, the sum of the incoming synaptic weights to a neural mass remains constant. Therefore, if the Hebbian learning mechanism makes only one synapse increase then the other will necessarily decrease. In fact, as shown in the bottom left picture of figure 2.2 it corresponds to the orthogonal projection of Hebbian learning on an affine subspace. However, this does not prevent the system from diverging because, the affine subspace is not bounded. Therefore, the rule is often used with artificial bounds on the connectivity and it is observed that all the synapses tend to saturate eventually, see [Miller and MacKay 1994].

2. **Multiplicative normalization of the first kind** (M_1) To prevent the system from diverging it is useful to weight the subtractive term by $\mathbf{J}_{ij}$. Therefore, the learning rule is

$$G_{ij}(\mathbf{v},\mathbf{J},\boldsymbol{\chi}) = \mathbf{v}_i\mathbf{v}_j - \frac{\mathbf{v}_i \sum_{k=1}^n \boldsymbol{\chi}_{ik}\mathbf{v}_k}{\sum_{k=1}^n \mathbf{J}_{ik}}\mathbf{J}_{ij} \tag{2.6}$$

This rule also conserves $\mathbf{a}^{\{1\}} \in \mathbb{R}^n$. Indeed,

$$\begin{aligned}\frac{d}{dt}\mathbf{a}_i^{\{1\}} &= \sum_{k=1}^n \Big(\boldsymbol{\chi}_{ik}\mathbf{v}_i\mathbf{v}_k - \frac{\boldsymbol{\chi}_{ik}\mathbf{J}_{ik}\mathbf{v}_i\sum_{p=1}^n \boldsymbol{\chi}_{ip}\mathbf{v}_p}{\sum_{p=1}^n \mathbf{J}_{ip}}\Big)\\ &= \mathbf{v}_i\Big(\sum_{k=1}^n \boldsymbol{\chi}_{ik}\mathbf{v}_k - \frac{\sum_{k=1}^n \boldsymbol{\chi}_{ik}\mathbf{J}_{ik}}{\sum_{p=1}^n \mathbf{J}_{ip}}\sum_{p=1}^n \boldsymbol{\chi}_{ip}\mathbf{v}_p\Big) = 0\end{aligned}$$

because $\mathbf{J}_{ik} = \boldsymbol{\chi}_{ik}\mathbf{J}_{ik}$. The main difference with the previous rule is that the rule is no more diverging. In fact, the multiplication by $\mathbf{J}_{ij}$ forces the projection to be non-orthogonal and inward pointing as shown in figure 2.2. Actually, [Gütig et al. 2003] have studied the intermediary case when $\mathbf{J}_{ij}$ is raised to the fractional power $\mu \in]0,1[$.

3. **Multiplicative normalization of the second kind** (M_2) Instead of projecting the connectivity on an affine subspace, this constraint project it on the sphere. It turns out it corresponds to a simple modification of the previous rule:

$$G_{ij}(\mathbf{v},\mathbf{J},\boldsymbol{\chi}) = \mathbf{v}_i\mathbf{v}_j - \frac{\mathbf{v}_i \sum_{k=1}^n \boldsymbol{\chi}_{ik}\mathbf{J}_{ik}\mathbf{v}_k}{\sum_{k=1}^n \mathbf{J}_{ik}^2}\mathbf{J}_{ij} \tag{2.7}$$

This rule correspond to keeping $\mathbf{a}^{\{2\}} \in \mathbb{R}^n$ constant along the trajectories, where $\mathbf{a}_i^{\{2\}} = \sum_{j=1}^k \mathbf{J}_{ik}^2$. Indeed,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\mathbf{a}_i^{\{2\}} &= \sum_{k=1}^n \Big(\boldsymbol{\chi}_{ik}\mathbf{J}_{ik}\mathbf{v}_i\mathbf{v}_k - \frac{\boldsymbol{\chi}_{ik}\mathbf{J}_{ik}^2\mathbf{v}_i\sum_{p=1}^n \boldsymbol{\chi}_{ip}\mathbf{J}_{ip}\mathbf{v}_p}{\sum_{p=1}^n \mathbf{J}_{ip}^2}\Big)\\ &= \mathbf{v}_i\Big(\sum_{k=1}^n \boldsymbol{\chi}_{ik}\mathbf{J}_{ik}\mathbf{v}_k - \frac{\sum_{k=1}^n \boldsymbol{\chi}_{ik}\mathbf{J}_{ik}^2}{\sum_{p=1}^n \mathbf{J}_{ip}^2}\sum_{p=1}^n \boldsymbol{\chi}_{ip}\mathbf{J}_{ip}\mathbf{v}_p\Big) = 0\end{aligned}$$

This rule corresponds to the orthogonal projection of each row of the connectivity on the centered sphere of radius $\mathbf{a}_i^{\{2\}}$ as shown in the top right picture of figure 2.2. It is often called the Oja learning rule in reference to the work of [Oja 1982] which will be detailed in part 3.1.2.2.

In the asymptotic limit of slow learning, it can be shown that this rule is the first order of

$$G_{ij}(\mathbf{v}, \mathbf{J}, \boldsymbol{\chi}) = \frac{\mathbf{J}_{ij} + \varepsilon \mathbf{v}_i \mathbf{v}_j}{\sum_{j=1}^{n} \mathbf{J}_{ij} + \varepsilon \mathbf{v}_i \mathbf{v}_j}$$

Remark: In the rules S_1, M_1, M_2 the averaging is done with respect to the pre-synaptic neurons. It is straightforward to build equivalent rules by averaging with respect to the post-synaptic neurons or even over all the neurons (post and pres-synaptic).

2.2.3 History dependent rules

So far, we have been considering time instantaneous rules, where G was only depending on the value of $\mathbf{v}$ at time t. However, there are evidences that time averaging may be included in some way in the learning process, see part 1.1.4. These learning rules are all local.

1. **Trace learning** introduced by [Földiák 1991] is a functional learning rule directly derived from the Hebbian learning rule. Roughly speaking, it averages (though it is more a geometric weighting) the mean firing rates of neural mass over running temporal windows, instead of considering the instantaneous firing rate of the post-synaptic neuron. With $g_\tau : t \mapsto \frac{1}{\tau} e^{-\frac{t}{\tau}} H(t)$ where H is the Heaviside function, this reads

$$G(\mathbf{v}, \mathbf{J}) = (\mathbf{v} * g_\tau) \otimes \mathbf{v}$$

   This rule is local but there is no reason it might converge since it has not corrected the inherent problem of Hebbian learning, i.e. the fact that there is no mechanism for decreasing the synapse's strength.

2. There are functional implementations of the learning rules with gating. For instance, the threshold θ might be a running mean of the activity variable. This gives

$$G(\mathbf{v}, \mathbf{J}) = \big(\mathbf{v} * (\delta - g_\tau)\big) \otimes \mathbf{v}$$

where δ is a Dirac function. It is similar to the covariance learning introduced in [Sejnowski et al. 1977]. One of the main features of this rule is that it subtracts a running average to the post-synaptic term. Therefore, the rule implements **differential learning** and not "absolute" learning.

3. We go back to the **BCM learning rule** [Bienenstock et al. 1982, Intrator and Cooper 1992 Castellani et al. 1999, Blais and Cooper 2008] because it includes a mechanism of temporal averaging. In their original article Bienenstock, Cooper and Munro considered the learning rule (2.3) where the choice of θ is crucial. Actually, they were in a formalism of n inputs being periodically shown to the network, so that they could choose θ a sort of average of the $\mathbf{v}$ over the m inputs. Actually, we are not ready yet to deal with this ensemble averaging formalism. So we will define a temporal version of BCM learning rule which is

$$\begin{aligned} G_{ij}(\mathbf{v},\mathbf{J}) &= \phi(\mathbf{v}_i,\theta_i)\mathbf{v}_j \\ \theta \in \mathbb{R}^n \text{ such that } \theta_i &= \left(\tfrac{\mathbf{v}_i * g_\tau}{c}\right)^2 \mathbf{v}_i * g_\tau \end{aligned} \tag{2.8}$$

where $c \in \mathbb{R}_+^*$. This is another example of differential learning.

Actually, this elarning rule was introduced in an ensemble averaging framework which does not clearly corresponds to the functionning of a biological network. Actually, this is through temporal averaging of time delayed learning rule that the ensemble averaging can be performed as shown in part 2.4. This is why we prefer to consider this learning rule as a time delayed one as a starting point.

Some recent works have shown that this learning rules shares similarities with the STPD rule for spiking neurons [Izhikevich and Desai 2003, Pfister and Gerstner 2006]. They have not proven a rigorous equilavence but qualitive similarities.

4. In part 1.1.4, we have shortly reviewed some recent results about the causation-based learning rule called **spike timing dependent plasticity (STDP)**. In particular, we have derived a learning rule which does not depend on the spiking behavior of the neurons. In 1.2.3, we have pleaded that this rule may be extended to rate-based networks. Indeed, let us define the STDP learning rule as

$$G(\mathbf{v},\mathbf{J}) = a_+\mathbf{v} \otimes (\mathbf{v} * g_\tau) - a_-(\mathbf{v} * g_\tau) \otimes \mathbf{v}$$

Note that changing g_τ in both terms (to possible different functions) makes it possible for this last equation to take into account the simple Hebbian learning rule and the first two example above: trace learning and differential learning. Actually, this rule is quite generic and most functional rules would fit in this formalism.

2.3 Dynamics of Hebbian learning with slow inputs

This section is devoted to studying the dynamics of the connectivity when the network is circularly exposed to $m \in \mathbb{N}$ inputs (with $m > 1$). Indeed, we assume here that the input signal $\mathbf{u} : \mathbb{R} \to \mathbb{R}^n$ is a piecewise constant function switching between m constant points of $\mathbb{R}^n$ which can be though of as pictures. This will provide a link with most methods of machine learning which aim at a better statistical description of a cloud of m points.

A significant assumption we make here is that these pictures are exposed slowly so that we can neglect the transient effects due to the switching. For each picture, the network will be assumed to converge to the corresponding equilibrium state (this will be proved to be true later).

In a first time, we assume the network activity is voltage-based and deterministic. We will discuss the extension to other models at the end of the section.

Therefore we assume the activity of the networks is governed by

$$\dot{\mathbf{v}} = -\mathbf{v} + \mathbf{J} \cdot S\big(\mathbf{v}\big) + \mathbf{u}(t)$$

where $S : \mathbb{R}^n \to \mathbb{R}^n$ is a positive sigmoidal function which saturates at $s_m \in \mathbb{R}_+$ and whose derivative maximum is $s'_m \in \mathbb{R}_+$. Typically, it can be considered to be the element-wise application of $S(\mathbf{v})_i = S(\mathbf{v}_i) = \frac{s_m}{1+e^{-4s'_m(\mathbf{v}_i-\vartheta)}}$ where $\vartheta \in \mathbb{R}_+^*$ is the threshold of the sigmoid.

The synaptic weights are assumed to evolve according to the Hebbian learning rule with decay of the form

$$\dot{\mathbf{J}} = \varepsilon(S(\mathbf{v}) \otimes S(\mathbf{v}) - \kappa\mathbf{J})$$

where ε is the learning rate. In this section we assume that there is a physical connection between all the neurons, i.e. $\boldsymbol{\chi}_{ij} = 1$. We can then rewrite the combined voltage and weight dynamics as the following non–autonomous (due to time–dependent inputs) dynamical system:

$$\Sigma : \quad \begin{cases} \dot{\mathbf{v}} & = & -\mathbf{v} + \mathbf{J} \cdot S(\mathbf{v}) + \mathbf{u}(t) \\ \dot{\mathbf{J}} & = & \varepsilon\Big(S(\mathbf{v}) \otimes S(\mathbf{v}) - \kappa\mathbf{J}\Big). \end{cases} \tag{2.9}$$

Existence and uniqueness of the solutions As said before, we need to check the solutions are not exploding in finite time. In this case, there is no explosion because the solutions are bounded. Indeed, boundedness of S and $\mathbf{u}$ implies boundedness of the system Σ. To prove this, note that the right hand side of the equation for $\mathbf{J}$ is the sum of a bounded term and a linear decay term in $\mathbf{J}$. Therefore, $\mathbf{J}$ is bounded and hence the term $\mathbf{J} \cdot S(\mathbf{v})$ is also bounded. The same reasoning applies to $\mathbf{v}$. S being Lipschitz continuous implies that the right hand side of the system is globally Lipschitz. This is sufficient to prove existence and uniqueness of the solution on $\mathbb{R}_+$ by applying the Cauchy-Lipschitz theorem. In the following, we will derive an averaged autonomous dynamical system Σ', which will be well-defined for the same reasons.

2.3.1 Averaging the system

We will show that system Σ can be approximated by an autonomous Cauchy problem which will be much more convenient to handle. This averaging method makes the most of multiple time–scales in the system. First, it is natural to consider that learning occurs on a much slower time-scale than the evolution of the membrane potentials, i.e.

$$\varepsilon \ll 1. \tag{2.10}$$

Second, an additional time-scale arises from the rate at which the inputs are sampled by the network. That is, the network cycles periodically through m fixed inputs, with the period of cycling given by τ. It follows that $\mathbf{u}$ is $\tau-$periodic, piecewise constant. We assume that the sampling rate is also much slower than the evolution of the membrane potentials,

$$\frac{m}{\tau} \ll 1. \tag{2.11}$$

Finally, we assume that the period τ is small compared to the time-scale of the learning dynamics,

$$\varepsilon \ll \frac{1}{\tau}. \tag{2.12}$$

We can now simplify the system Σ by applying Tikhonov's theorem for slow/fast systems, and then classical averaging methods for periodic systems.

2.3.1.1 Tikhonov's theorem

Tikhonov's theorem ([Tikhonov 1952] and [Verhulst 2007] for a clear introduction) deals with slow/fast systems. It says the following:

Theorem 2.3.1. *Consider the initial value problem*

$$\begin{aligned}&\dot{\mathbf{x}} = f(\mathbf{x},\mathbf{y},t),\ \mathbf{x}(0)=\mathbf{x}_0,\ \mathbf{x}\in\mathbb{R}^p, t\ \in\mathbb{R}_+\\&\varepsilon\dot{\mathbf{y}} = g(\mathbf{x},\mathbf{y},t),\ \mathbf{y}(0)=\mathbf{y}_0,\ \mathbf{y}\in\mathbb{R}^q\end{aligned}$$

Assume that:

1. *A unique solution of the initial value problem exists and we suppose, this holds also for the reduced problem*

$$\begin{aligned}&\dot{\mathbf{x}} = f(\mathbf{x},\mathbf{y},t),\ \mathbf{x}(0)=\mathbf{x}_0\\&0 = g(\mathbf{x},\mathbf{y},t)\end{aligned}$$

   with solutions $\bar{\mathbf{x}}(t)$, $\bar{\mathbf{y}}(t)$.

2. *The equation $0 = g(\mathbf{x},\mathbf{y},t)$ is solved by $\bar{\mathbf{y}}(t) = \phi(\mathbf{x},t)$, where $\phi(\mathbf{x},t)$ is a continuous function and an isolated root. Also suppose that $\bar{\mathbf{y}}(t) = \phi(\mathbf{x},t)$ is an asymptotically stable solution of the equation $\frac{d\mathbf{y}}{d\tau} = g(\mathbf{x},\mathbf{y},\tau)$ that is uniform in the parameters $\mathbf{x}\in\mathbb{R}^p$ and $t\in\mathbb{R}_+$.*

3. *$\mathbf{y}(0)$ is contained in an interior subset of the domain of attraction of $\bar{\mathbf{y}}$.*

Then we have

$$\begin{aligned}&\lim_{\varepsilon\to 0}\mathbf{x}_\varepsilon(t) = \bar{\mathbf{x}}(t),\ 0\le t\le r\\&\lim_{\varepsilon\to 0}\mathbf{y}_\varepsilon(t) = \bar{\mathbf{y}}(t),\ 0\le d\le t\le r\end{aligned}$$

with d and r constants independent of ε.

In order to apply Tikhonov's theorem directly to the system Σ, we first need to rescale time according to $t\to\varepsilon t$. This gives

$$\begin{aligned}\varepsilon\dot{\mathbf{v}} &= -\mathbf{v}+\mathbf{J}\cdot S(\mathbf{v})+\mathbf{u}\\ \dot{\mathbf{J}} &= S(\mathbf{v})\otimes S(\mathbf{v})-\kappa\mathbf{J}.\end{aligned}$$

Tikhonov's theorem then implies that solutions of Σ are close to solutions of the reduced system (in the unscaled time variable)

$$\begin{cases}\mathbf{v}(t) &= \mathbf{J}\cdot S\big(\mathbf{v}(t)\big)+\mathbf{u}(t)\\ \dot{\mathbf{J}} &= \varepsilon\Big(S(\mathbf{v})\otimes S(\mathbf{v})-\kappa\mathbf{J}\Big),\end{cases}\tag{2.13}$$

provided that the dynamical systems Σ in equation (6), and equation (2.13) are well defined. It is easy to show that both systems are Lipschitz because of the properties of S. Following [Faugeras et al. 2008], we know that if

$$s'_m \|\mathbf{J}\| < 1, \tag{2.14}$$

then there exists an isolated root $\bar{\mathbf{v}} : \mathbb{R}_+ \to \mathbb{R}^n$ of the equation $\mathbf{v} = \mathbf{J} \cdot S(\mathbf{v}) + \mathbf{u}$ and $\bar{\mathbf{v}}$ is asymptotically stable. Equation (2.14) corresponds to the weakly connected case. Moreover, the initial condition belongs to the basin of attraction of this single fixed point.

Note that we require $\frac{m}{\tau} \ll 1$ so that the membrane potentials have sufficient time to approach the equilibrium associated with a given input before the next input is presented to the network. In fact, this assumption makes it reasonable to neglect the transient activity dynamics due to the switching between inputs.

2.3.1.2 Periodic averaging

The system given by equation (2.13) corresponds to a differential equation for $\mathbf{J}$ with τ-periodic forcing due to the presence of $\mathbf{v}$ on the right-hand side. Since $\tau << \varepsilon^{-1}$, we can use classical averaging methods (see [Sanders et al. 2007]) to show that solutions of (2.13) are close to solutions of the following autonomous system on the time-interval $[0, \frac{1}{\varepsilon}]$ (which we suppose large because $\varepsilon << 1$)

$$\Sigma_0 : \quad \begin{cases} \mathbf{v}(t) &= \mathbf{J} \cdot S(\mathbf{v}(t)) + \mathbf{u}(t) \\ \dot{\mathbf{J}} &= \varepsilon\Big(\dfrac{1}{\tau}\displaystyle\int_0^\tau S(\mathbf{v}(s)) \otimes S(\mathbf{v}(s)) ds - \kappa \mathbf{J}(t)\Big). \end{cases}$$

It follows that solutions of Σ are also close to solutions of Σ_0. Finding the explicit solution $\mathbf{v}(t)$ for each input $\mathbf{u}(t)$ is difficult and requires fixed points methods, e.g. a Picard algorithm. Therefore, we will consider yet another system Σ' whose solutions are also close to Σ_0 and hence Σ. In order to construct Σ' we need to introduce some additional notation.

Let us label the m inputs by $\mathbf{u}^{(a)}, a = 1, \ldots, m$ and denote by $\mathbf{v}^{(a)}$ the fixed point solution of the equation $\mathbf{v}^{(a)} = \mathbf{J} \cdot S(\mathbf{v}^{(a)}) + \mathbf{u}^{(a)}$. Given the periodic sampling of the inputs, we can rewrite the system Σ_0 above as

$$\begin{aligned} \mathbf{v}^{(a)} &= \mathbf{J} \cdot S(\mathbf{v}^{(a)}) + \mathbf{u}^{(a)} \\ \dot{\mathbf{J}} &= \varepsilon\Big(\frac{1}{m}\sum_{a=1}^{m} S(\mathbf{v}^{(a)}) \otimes S(\mathbf{v}^{(a)}) - \kappa \mathbf{J}(t)\Big). \end{aligned} \tag{2.15}$$

If we now introduce the $n \times m$ matrices $\mathbf{V}$ and $\mathbf{U}$ with components $\mathbf{V}_{ia} = \mathbf{v}_i^{(a)}$ and $\mathbf{U}_{ia} = \mathbf{u}_i^{(a)}$, then we can eliminate the tensor product and simply write (2.15) in the matrix form

$$\begin{aligned} \mathbf{V} &= \mathbf{J} \cdot S(\mathbf{V}) + \mathbf{U} \\ \dot{\mathbf{J}} &= \varepsilon\Big(\frac{1}{m} S(\mathbf{V}) \cdot S(\mathbf{V})' - \kappa \mathbf{J}(t)\Big), \end{aligned} \tag{2.16}$$

where $S(\mathbf{V}) \in \mathbb{R}^{n \times m}$ such that $[S(\mathbf{V})]_{ia} = s(\mathbf{v}_i^{(a)})$. A second application of Tikhonov's theorem (in the reverse direction) then establishes that solutions of the system Σ_0 (written in the matrix form (2.16)) are close to solutions of the matrix system

$$\Sigma' : \quad \begin{cases} \dot{\mathbf{V}} = -\mathbf{V} + \mathbf{J} \cdot S\big(\mathbf{V}\big) + \mathbf{U} \\ \dot{\mathbf{J}} = \varepsilon\Big(\frac{1}{m} S(\mathbf{V}) \cdot S(\mathbf{V})' - \kappa \mathbf{J}(t)\Big) \end{cases} \tag{2.17}$$

In the remainder of the section, we will focus on the system Σ' whose solutions are close to those of the original system Σ provided condition (2.14) is satisfied, i.e. the network is weakly connected. Clearly, the fixed points $(\mathbf{v}^*, \mathbf{J}^*)$ of system Σ satisfy $\|\mathbf{J}^*\| \leq \frac{s_m^2}{\kappa}$. Therefore, equation (2.14) says that if $\frac{s_m^2 s_m'}{\kappa} < 1$ then Tikhonov's theorem can be applied and systems Σ and Σ' can be reasonably considered as good approximations of each other. The advantage of the averaged system Σ' is that it is given by autonomous ordinary differential equations. Moreover, since it is Lipschitz continuous, it leads to a well-posed Cauchy problem.

2.3.1.3 Simulations

To illustrate the above approximation, we simulate a simple network with both exact, i.e. Σ, and averaged ,i.e. Σ', evolution equations. For these simulations, the network consists of $n = 10$ fully-connected neurons and is presented with $m = 10$ different random inputs taken uniformly in the intervals $[0,1]^n$. For this simulation we use $s(x) = \frac{1}{1+e^{-4(x-1)}}$, and $\kappa = 10$. Figure 2.3. shows the percentage of error between final connectivities for different values of ε and τ/m. Figure 2.4 shows the temporal evolution of the norm of the connectivity for both the exact and averaged system for $\tau = 10^3$ and $\varepsilon = 10^{-3}$.

2.3.2 Stability

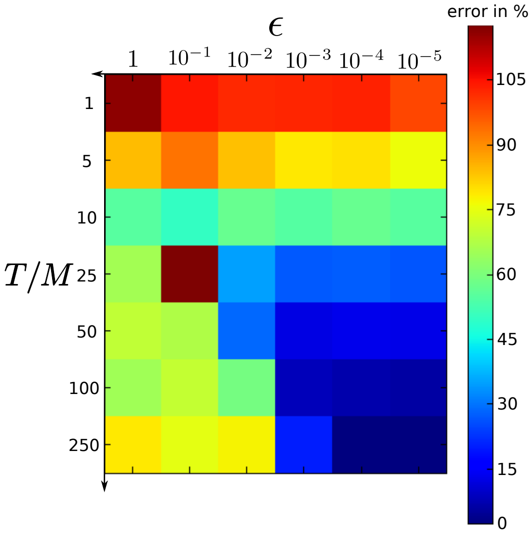


Figure 2.3: Percentage of error between final connectivities for the exact and averaged system.

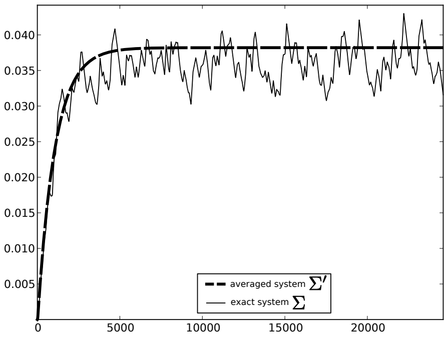


Figure 2.4: Temporal evolution of the norm of the connectivities of the exact system Σ and averaged system Σ'.

2.3.2.1 Lyapunov function

In the case of a single fixed input ($m = 1$), the systems Σ and Σ' are equivalent and reduce to the neural network with adapting synapses previously analyzed by [Dong and Hopfield 1992]. Under the additional constraint that the weights are symmetric ($\mathbf{J}_{ij} = \mathbf{J}_{ji}$), these authors showed that the simultaneous evolution of the neuronal activity variables and the synaptic weights can be re-expressed as a gradient dynamical system that minimizes a Lyapunov or energy function of state. We can generalize their analysis to the case of multiple inputs ($m > 1$) and non-symmetric weights using the averaged system Σ'. That is, following along similar lines to [Dong and Hopfield 1992], we introduce the energy function

$$E(\mathbf{X}, \mathbf{J}) = -\frac{1}{2}\langle \mathbf{X}, \mathbf{J} \cdot \mathbf{X} \rangle - \langle \mathbf{U}, \mathbf{X} \rangle + \langle 1, \overline{S^{-1}}(\mathbf{X}) \rangle + \frac{m\kappa}{2}\|\mathbf{J}\|^2 \tag{2.18}$$

where $\mathbf{X} = S(\mathbf{V})$, $\|\mathbf{J}\|^2 = \langle \mathbf{J}, \mathbf{J} \rangle = \sum_{i,j} \mathbf{J}_{ij}^2$,

$$\langle \mathbf{X}, \mathbf{J} \cdot \mathbf{X} \rangle = \sum_{a=1}^{m} \sum_{i=1}^{n} \mathbf{U}_i^{(a)} \mathbf{J}_{ij} \mathbf{U}_j^{(a)}, \quad \langle \mathbf{U}, \mathbf{X} \rangle = \sum_{a=1}^{m} \sum_{i=1}^{n} \mathbf{u}_i^{(a)} \mathbf{U}_i^{(a)} \tag{2.19}$$

and

$$\langle 1, \overline{S^{-1}}(\mathbf{X}) \rangle = \sum_{a=1}^{m} \sum_{i=1}^{n} \int_0^{\mathbf{U}_i^{(a)}} S^{-1}(\xi) d\xi. \tag{2.20}$$

In contrast to [Dong and Hopfield 1992], we do not require *a priori* that the weight matrix is symmetric. However, it can be shown that the system always converges to a symmetric connectivity pattern.

More precisely,

Proposition 2.3.2. *The connectivity becomes symmetric through Hebbian learning with linear decay.*

$$\mathcal{A} = \left\{ (\mathbf{V}, \mathbf{J}) \in \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times n} : \ \mathbf{J} = \mathbf{J}' \right\} \textit{ is an attractor of system } \Sigma'$$

Proof. We need to prove the 2 points: (i) $\mathcal{A}$ is an invariant set, and (ii) for all $\big(\mathbf{V}(0), \mathbf{J}(0)\big) \in \mathbb{R}^{n \times m} \times \mathbb{R}^{n \times n}$, $\big(\mathbf{V}(t), \mathbf{J}(t)\big)$ converges to $\mathcal{A}$ as $t \to +\infty$. Since $\mathbb{R}^{n \times n}$ is the direct sum of the set of symmetric connectivities and the set of anti-symmetric connectivities, we write $\mathbf{J}(t) = \mathbf{J}_S(t) + \mathbf{J}_A(t)$, $\forall t \in \mathbb{R}_+$, where $\mathbf{J}_S$ is symmetric and $\mathbf{J}_A$ is anti-symmetric.

(i) In (2.17), the right hand side of the equation for $\dot{\mathbf{J}}$ is symmetric. Therefore, if $\exists t_1 \in R_+$ such that $\mathbf{J}_A(t_1) = 0$, then $\mathbf{J}$ remains in $\mathcal{A}$ for $t \geq t_1$.

(ii) Projecting the expression for $\dot{\mathbf{J}}$ in equation (2.17) on to the anti-symmetric component leads to

$$\frac{d\mathbf{J}_A}{dt} = -\varepsilon\kappa \mathbf{J}_A(t) \tag{2.21}$$

whose solution is $\mathbf{J}_A(t) = \mathbf{J}_A(0) \exp(-\varepsilon\kappa t), \forall t \in \mathbb{R}_+$. Therefore, $\lim_{t \to +\infty} \mathbf{J}_A(t) = 0$. The system converges exponentially to $\mathcal{A}$. □

It can then be shown that on $\mathcal{A}$ (symmetric weights), E is a Lyapunov function of the dynamical system Σ', that is,

$$\frac{dE}{dt} \leq 0, \quad \text{and} \quad \frac{dE}{dt} = 0 \implies \frac{d\mathcal{Y}}{dt} = 0, \quad \mathcal{Y} = (\mathbf{V}, \mathbf{J})'.$$

The boundedness of E and the Krasovskii-LaSalle invariance principle then implies that the system converges to an equilibrium [Khalil and Grizzle 1996]. We thus have

Theorem 2.3.3. *The initial value problem for the system Σ' with $\big(\mathbf{V}(0), \mathbf{J}(0)\big) \in \mathcal{H}$, converges to an equilibrium state.*

Proof. See appendix C.1.1 □

It follows that neither oscillatory nor chaotic attractor dynamics can occur.

2.3.2.2 Linear stability

Although we have shown that there are stable fixed points, not all of the fixed points are stable. However, we can apply a linear stability analysis on the system Σ' to derive a simple sufficient condition for a fixed point to be stable. The method we use in the proof could be extended to more complex rules. The proof reveals the significant role played by the Kronecker product in Hebbian learning.

Theorem 2.3.4. *The equilibria of system Σ' satisfy:*

$$\begin{cases} \mathbf{V}^* = \frac{1}{\kappa m} S(\mathbf{V}^*) \cdot S(\mathbf{V}^*)' \cdot S(\mathbf{V}^*) + \mathbf{U} \\ \mathbf{J}^* = \frac{1}{\kappa m} S(\mathbf{V}^*) \cdot S(\mathbf{V}^*)' \end{cases} \tag{2.22}$$

and a sufficient condition for stability is

$$3 s'_m \|\mathbf{J}^*\| < 1 \tag{2.23}$$

provided $1 > \varepsilon\kappa$ which is most probably the case since $\varepsilon << 1$.

Proof. See appendix C.1.2 □

This condition is strikingly similar to that derived in [Faugeras et al. 2008]. In fact, condition (2.23) is stronger than the contracting condition (2.14). It says the network may converge to a weakly connected situation. It justifies the averaging method by saying that we remain in the domain of validity of the averaging method. It also says that the dynamics of $\mathbf{V}$ is likely (because the condition is only sufficient) to be contracting and therefore subject to no bifurcations: a fully recurrent learning neural network is likely to have a "simple" dynamics.

2.3.3 Summary, conclusions and immediate extensions

Analyzing the dynamics of the connectivity $\mathbf{J}$ in a network slowly exposed to a finite number of pictures was achieved in two steps: (i) the derivation of a matrix-valued system Σ' from a vector-valued system Σ through a slow-fast reduction and periodic averaging and (ii) the proof that the averaged system Σ' derived from an energy.

Generalizing this approach is possible for the first point yet derivation from an energy corresponds to deep dynamical relationship between the communication term (basically a dot product) and Hebbian learning rule (basically a Kronecker product).

- The first point is easily generalizable to more general systems of the form (2.1) and (2.2). The main underlying assumption is that the activity should converge to a single equilibrium state for each picture. If this is verified then it is possible to apply Tikhonov theorem to show the proximity of the respective connectivities of system Σ and Σ'. If it is not verified then Σ and Σ' have different behaviors. We have just shown this assumption hold for a voltage-based activity with a small connectivity. Therefore, it extends to synaptic-drive-based because of the change of variable linking the two. The case of frequency-based is very similar and seems to hold (at least numerically), transferring the property to the pseudo voltage-based. Then any kind of learning rule can be applied as long as it keeps the connectivity small enough for inequation (2.14) to hold.

  Short time delays in the learning rule can be averaged out in the process. Indeed, a system with delays has identical equilibrium points as a constant input as the same system without delays (yet it might change their stability, see [Veltz and Faugeras 2011]). But there might be a problem at the transition between several inputs. In fact, system Σ' does not take delays with an order of magnitude smaller than $\frac{\tau}{m}$ into account. Given that we have assumed the input is so slow that we can neglect the transients, the learning window will almost never overlap the transition between two pictures. The reasoning above can therefore be extend to multiple inputs. So this means the frequency-based and synaptic-drive-based have the same associated system Σ' if the delays are short enough.

If the delays are long enough, then there might be a sort of ensemble averaging. All the inputs may be taken into account in the delay window and this can sum to an average over the m pictures. In particular, in the case of differential learning, this might turn to a learning rule made of the difference of the instantaneous activity for one input and the average activity for all the inputs: when this is the a^{th} input then $G_{ij}(\mathbf{v}) = \big(\mathbf{v}_i^{(a)} - \frac{1}{m}\sum_{b=1}^{m}\mathbf{v}_i^{(b)}\big)\mathbf{v}_j$. While spatial constraint (see section 2.2.2) subtract a spatial average of the neurons activity, differential learning may subtract a temporal average over all the inputs.

- However, the second point is very specific to Hebbian learning with linear decay. The stability analysis should be carried with ad hoc methods for each different case. For the particular setting we chose to expose, the fact that system Σ' in (2.17) derives from an energy (see theorem 2.3.3) means there is a deep mathematical link between the dot product due the communication in the activity variable and the tensor product inherent to Hebbian learning. Indeed, these two terms are the only coupling terms in system Σ' and it is exceptional that they derive from the same energy: $\langle \mathbf{V}, \mathbf{J}\cdot\mathbf{V}\rangle$. There is unusual mathematical singularity and simplicity in the relationship between Hebbian learning and network communication.

2.4 Generalization to fast inputs and intrinsic noise

In this section, we address the more general case of noisy learning neural networks exposed to arbitrarily fast inputs. We still want to study the dynamics of the connectivity through learning in this more general framework.

To study fast inputs, we have developed mathematical tools to go beyond Tikhonov theorem 2.3.1. In this new framework, we take into account the transient trajectories of the noisy activity variable. These new methods for temporal averaging are pedagogically exposed in appendix B. They lead to a reduced differential equation for the connectivity which gives an approximate solution of the initial slow-fast system.

This formula is complicated and we have not found a way to use it in a generic non-linear case. However, based on a new representation of the solutions of non-autonomous, linear, functional and stochastic differential equations, developed in appendix E, we are able to use it in the linear case. As in the previous section, we are able to prove the convergence to a single equilibrium point in a *weakly connected regime*. Besides, this approach makes it possible to have an explicit expansion for the equilibrium points of the connectivity.

Let us consider a generic learning neural network as described by equations (2.1) and (2.2), i.e.

$$\left\{\begin{array}{lcl} d\mathbf{v} & = & \frac{1}{\varepsilon_1}\Big(\mathcal{F}(\mathbf{v}) + \mathcal{S}\big(\mathbf{J}^\varepsilon.\mathcal{H}(\mathbf{v}) + \mathbf{u}(\frac{t}{\varepsilon_2})\big)\Big)\,dt + \frac{1}{\sqrt{\varepsilon_1}}\Sigma(\mathbf{v},\mathbf{J}^\varepsilon).dB(t) \\ d\mathbf{J}^\varepsilon & = & G(\mathbf{v},\mathbf{J}^\varepsilon)\;dt \end{array}\right. \tag{2.24}$$

for simplicity we have assumed that the matrix of physical connection between the neurons $\boldsymbol{\chi}$ is full, i.e. $\boldsymbol{\chi}_{ij} = 1$ for all i, j, as in the previous section. The notation $\mathbf{J}^\varepsilon$ corresponds to finite but non-null ε_1 and ε_2.

Note that we have added a small parameter ε_2 in the input function. The value of this parameter controls the speed of the inputs. The fact that we assume that ε_2 is small says that the input evolves on a much faster scale than the connectivity. In fact, by controlling the ratio $\frac{\varepsilon_1}{\varepsilon_2}$ we can also compare the speed of the inputs to the speed of the activity variable. We develop tools to work in the asymptotic regime where both ε_1 and ε_2 tend to zero (which we write $\varepsilon \to 0$) keeping the ratio constant. This motivates the definition of

$$\mu := \lim_{\varepsilon\to 0}\frac{\varepsilon_1}{\varepsilon_2}$$

For the same reasons as in section 2.3, we assume the inputs are τ-periodic. We will discuss later the extension of the results to stochastic inputs with enough ergodicity to sample sufficiently their probability distribution in a time τ.

2.4.1 Averaging principles : theory

2.4.1.1 General result

Following, the lines of appendix B, we have the following result

Theorem 2.4.1.*Definition and distance to an averaged system.*
Assuming the following hypotheses are verified:

- *The functions $\mathcal{F}, \mathcal{S}, \mathcal{H}, \mathbf{\Sigma}$ and G are locally Lipschitz continuous.*
- *System (2.24) has a globally attracting subspace centered on 0. In other words, there exists a constant $R > 0$ such that the right hand side of system (2.24) applied at $(\mathbf{v}, \mathbf{J})$, such that $\|(\mathbf{v}, \mathbf{J})\| > R$, is pointing inward.*
- *The diffusion matrix $\mathbf{\Sigma}$ is bounded, i.e. $\exists M_{\mathbf{\Sigma}} > 0$ such that $\forall (\mathbf{v}, \mathbf{J})$, $\|\mathbf{\Sigma}(\mathbf{v}, \mathbf{J})\| < M_{\mathbf{\Sigma}}$, and uniformly non-degenerate, i.e. $\exists \eta_0 > 0$ such that $\forall \mathbf{x}, \mathbf{v} \in \mathbb{R}^n$ and $\mathbf{J} \in \mathbb{R}^{n \times n}$, $< \mathbf{\Sigma}(\mathbf{v}, \mathbf{J}).\mathbf{\Sigma}(\mathbf{v}, \mathbf{J})'.\mathbf{x}, \mathbf{x} > \geq \eta_0 ||\mathbf{x}||^2$.*
- *For a constant input, the activity has a unique equilibrium point which turns out to be stable. In a time-dependent input framework, it can be written $\exists r_0 < 0$ such that for all $t \geq 0$ and for all $\mathbf{v}, \mathbf{x} \in \mathbb{R}^n$ and $\mathbf{J}, \mathbf{W} \in \mathbb{R}^{n \times n}$:*

$$\Big\langle \nabla_{\mathbf{x},\mathbf{W}} F(\mathbf{v}, \mathbf{J}, t), (\mathbf{v}, \mathbf{J}) \Big\rangle \leq r_0 ||\mathbf{x}||^2$$

  where F is the right hand side of the activity equation (2.1).

Let $\mu \in [0, \infty]$. If $\mathbf{J}^\varepsilon$ is the solution of system (2.24) and $\bar{\mathbf{J}}$ is solution of

$$\frac{d\bar{\mathbf{J}}}{dt} = \bar{G}_\mu(\bar{\mathbf{J}}) \text{ with } \bar{\mathbf{J}}(0) = \mathbf{J}^\varepsilon(0)$$

where $\bar{G}_\mu : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ is defined as

$$\bar{G}_\mu(\bar{\mathbf{J}}) := \left(\frac{\tau}{\mu}\right)^{-1} \int_0^{\frac{\tau}{\mu}} \int_{\mathbf{v} \in \mathbf{R}^n} G(\mathbf{v}, \bar{\mathbf{J}}) \nu_\mu^{\bar{\mathbf{J}}}(t, d\mathbf{v}) dt \tag{2.25}$$

where $\nu_\mu^{\mathbf{J}_0}(t, d\mathbf{v})$ is the $\frac{\tau}{\mu}$-periodic evolution system of measures corresponding to the rescaled time-inhomogeneous frozen process

$$d\mathbf{v} = F(\mathbf{v}, \mathbf{J}_0, \mu t)dt + \mathbf{\Sigma}(\mathbf{v}, \mathbf{J}_0)dB(t) \tag{2.26}$$

Then the following convergence result holds, for all $T > 0$ and $\delta > 0$:

$$\overset{\mu}{\lim_{\varepsilon \to 0}} \mathbb{P}\left[\sup_{t\in[0,T]} |\mathbf{J}^\varepsilon(t) - \bar{\mathbf{J}}(t)|^2 > \delta\right] = 0$$

Proof. See [Wainrib 2011] for a rigorous proof or appendix B for a sketch of the proof. □

Remark 2.The last hypothesis is the strongest. As in Tikhonov theorem 2.3.1, it corresponds to asking the fast variable to be roughly converging to a single equilibrium point. For instance, spiking neurons do not verify this hypothesis. However, traditional rate-based models fulfill this assumption when the slope of the sigmoid is not too sharp.

2.4.1.2 Periodic measure for linear activity

In concrete situations, it is difficult to compute the measure $\nu_\mu^{\mathbf{J}}(t, d\mathbf{v})$. It corresponds to finding the time-dependent measure associated to the fast activity equation (2.26) with a frozen connectivity $\mathbf{J}$. Computing explicitly this measure in a non-linear framework is out of the scope of this thesis. Here, as suggested by section 1.2.2.6, we will narrow our study to linear activities with additive noise:

$$d\mathbf{v} = \big((\mathbf{J} - \mathbf{L}).\mathbf{v}(t) + \mathbf{u}(\mu t)\big)dt + \mathbf{\Sigma}.dB(t)$$

where $\mathbf{L} \in \mathbb{R}^{n\times n}$ is proportional to the identity, i.e. $\mathbf{L} = lI_d$ with $l > 0$. $\mathbf{L}$ accounts for the intrinsic dynamics of the neurons. In this case, it is possible to compute $\nu_\mu^{\mathbf{J}}(t, d\mathbf{v})$ explicitly. It is a Gaussian measure centered on the $\frac{\tau}{\mu}$ periodic solution of

$$\dot{\mathbf{x}} = (\mathbf{J} - \mathbf{L}).\mathbf{x}(t) + \mathbf{u}(\mu t)$$

with covariance $\mathbf{Q}$ where

$$(\mathbf{J} - \mathbf{L}).\mathbf{Q} + \mathbf{Q}.(\mathbf{J} - \mathbf{L})' + \mathbf{\Sigma}.\mathbf{\Sigma}' = 0 \tag{2.27}$$

This reads,

$$\nu_\mu^{\mathbf{J}}(t, d\mathbf{v}) = \mathcal{N}_{\mathbf{x}(t),\mathbf{Q}}(d\mathbf{v})$$

2.4.1.3 Asymptotic well-posedness for linear activities

When the activity is assumed to be linear there is a risk the solution explodes in finite time. Indeed, if the real part of the eigenvalues of the evolving connectivity grow larger than l then the fast activity diverges exponentially. As a consequence, the learning rule would explode increasing again the activity and a the retroactive action from one to the other would lead to explosion of the solutions.

Therefore, we need to consider learning rules for which the connectivity never grows larger than $\mathbf{L}$. For simplicity, we call the *subspace of well-posedness*, the set of matrices of $\mathbb{R}^{n\times n}$ whose eigenvalues have real part smaller than l. For instance, we always remain in the subspace of well-posedness for the Hebbian learning rule with linear decay, i.e. $G(\mathbf{v},\mathbf{J}) = -\kappa\mathbf{J} + \mathbf{v}\otimes\mathbf{v}$, when κ is large enough. In that case, we show in the following that the connectivity cannot cross the critical value $\mathbf{L}$ and the system is globally well-posed (in a deterministic framework).

However, if we add some noise to the system, it may push the solution out of subspace of well-posedness and therefore, the system would diverge. Actually, this will always be the case on a long time-scale. Note that this is only due to the linear framework (a saturation would prevent this from happening). Yet, we can ask the learning rule to keep the connectivity sufficiently far away from subspace of well-posedness border so that it takes a very long time for the system to be stochastically pushed to diverge. To analyze this rigorously let us introduce the definition of an asymptotically well-posed system (driven by a perturbation of amplitude ε)

Definition 2.4.2. *A stochastic differential equation with a given initial condition is asymptotically well-posed in probability if for the given initial condition:*

1. *a unique solution exists until a time* τ_ε

2. *for all* $T > 0$,
$$\lim_{\varepsilon\to 0}\mathbb{P}\left[\tau_\varepsilon \geq T\right] = 1$$

We now need to introduce a proposition which will prove the learning neural networks with linear activity we are considering are asymptotically well-posed. It is

Proposition 2.4.3. *If there exists a subset* $\mathcal{E}$ *of* $\mathbb{R}^{n\times n}$ *such that:*

1. *The functions $\mathcal{F}, \mathcal{S}, \mathcal{H}, G, \mathbf{\Sigma}$ satisfy the assumptions of theorem 2.4.1 restricted on $\mathbb{R}^n \times \mathcal{E}$.*

2. *For all $\mathbf{J}_0 \in \mathcal{E}$, the solutions of (2.26) are bounded.*

3. *$\mathcal{E}$ is invariant under the flow of $\bar{G}_\mu$, as defined in (2.25)*

Then for any initial condition $\mathbf{J}_0 \in \mathcal{E}$ system (2.24) is asymptotically well-posed in probability and $\mathbf{J}^\varepsilon$ satisfies the conclusion of theorem 2.4.1.

Proof. See appendix B.6. □

Remark 3.In the case of linear activity, the second hypothesis means $\mathcal{E} \subset \{\mathbf{J} \in \mathbb{R}^{n\times n} : \mathbf{J} < \mathbf{L}\}$.

Therefore, to prove well posedness of a learning neural network with linear activity, it is sufficient to prove that the subspace of well-posedness is invariant under the averaged learning rule $\bar{G}_\mu$.

2.4.2 Symmetric Hebbian learning

One of the simplest, yet non-trivial, stochastic learning model is obtained when considering:

- a linear model for neuronal activity, namely $\mathcal{F}(\mathbf{v}_i) = (-\mathbf{L}.\mathbf{v})_i = -l\ \mathbf{v}_i$ with l a positive constant.

- a linear model for the synaptic transmission, namely $\mathcal{S}(\mathbf{v}_i) = \mathbf{v}_i$ and $\mathcal{H}(\mathbf{v}_i) = \mathbf{v}_i$.

- a constant diffusion matrix $\mathbf{\Sigma}$ (additive noise) proportional to the identity $\mathbf{\Sigma} = \sigma Id$ (spatially uncorrelated noise).

- a Hebbian learning rule with linear decay, namely $G_{ij}(\mathbf{J}, \mathbf{v}) = -\kappa \mathbf{J}_{ij} + \mathbf{v}_i \mathbf{v}_j$. Actually, it corresponds to the tensor product: $\{\mathbf{v} \otimes \mathbf{v}\}_{ij} = \mathbf{v}_i \mathbf{v}_j$.

This model can be written as follows:

$$\begin{cases} d\mathbf{v} &= \frac{1}{\varepsilon_1}\Big(-\mathbf{L}.\mathbf{v} + \mathbf{J}^\varepsilon.\mathbf{v} + \mathbf{u}(\frac{t}{\varepsilon_2})\Big)dt + \frac{\sigma}{\sqrt{\varepsilon_1}} dB(t) \\ \frac{d\mathbf{J}^\varepsilon}{dt} &= -\kappa \mathbf{J}^\varepsilon + \mathbf{v} \otimes \mathbf{v} \end{cases} \tag{2.28}$$

where neurons are assumed to have the same decay constant: $\mathbf{L} = lI_d$; $\mathbf{u}$ is a periodic continuous input; $\sigma, \varepsilon_1, \varepsilon_2, \kappa \in \mathbb{R}_+$ with $\varepsilon_1, \varepsilon_2 \ll 1$ and $B(t)$ is a n-dimensional Brownian noise.

The first question that arises is about the well-posedness of the system: what is the definition interval of the solutions of system (2.28)? Do they explode in finite time? At first sight, it seems there may be a runaway of the solution if the largest real part among the eigenvalues of $\mathbf{J}$ grows bigger than l. In fact, it turns out this scenario can be avoided if the following assumption linking the parameters of the system is satisfied.

Assumptions 2.4.4.

$$\textit{There exists } p \in]0,1[\text{ , } \textit{such that } \left(\frac{\sigma^2 l}{2p(1-p)} + \frac{u_m^2}{p(1-p)^2}\right) < \kappa l^3$$

where $u_m = \sup_{t\in\mathbb{R}_+} \|\mathbf{u}(t)\|_2$.

It corresponds to making sure the external (i.e. u_m) or internal (i.e. σ) excitations are not too large compared to the decay mechanism (represented by κ and l). Note that if $p \in]0,1[$, u_m and d are fixed it is sufficient to increase κ or l for this assumption to be satisfied.

A space of well-posed defined before can be defined in the following. It is written E_p and indexed by $p \in]0,1[$ verifying assumption 2.4.4

$$E_p = \Big\{\mathbf{J} \in \mathbb{R}^{n\times n} : \ \mathbf{J} \text{ is symmetric, } \mathbf{J} \geq 0 \text{ and } \mathbf{J} < p\mathbf{L}\Big\}$$

is invariant by the flow of the averaged system $\bar{G}$. Therefore, the averaged system is defined and bounded on $\mathbb{R}_+$. The slow/fast system being asymptotically close to the averaged system, it is therefore asymptotically well-defined in probability. This summarized is in the following

Theorem 2.4.5. *If assumption 2.4.4 is verified for* $p \in]0,1[$, *then system* (2.28) *is asymptotically well-posed in probability and the connectivity matrix* $\mathbf{J}^\varepsilon$ *solution of system* (2.28) *converges to* $\bar{\mathbf{J}}$, *in the sense that for all* $\delta, T > 0$,

$$\overset{\mu}{\lim_{\varepsilon\to 0}} \mathbb{P}\left[\sup_{t\in[0,T]} |\mathbf{J}_t^\varepsilon - \bar{\mathbf{J}}_t|^2 > \delta\right] = 0$$

where $\bar{\mathbf{J}}$ *is the deterministic solution of:*

$$\frac{d\bar{\mathbf{J}}_{ij}}{dt} = \bar{G}(\bar{\mathbf{J}})_{ij} = \underbrace{-\kappa\bar{\mathbf{J}}_{ij}}_{decay} + \underbrace{\frac{\mu}{\tau}\int_0^{\frac{\tau}{\mu}} \mathbf{v}_i(s)\mathbf{v}_j(s)\; ds}_{correlation} + \underbrace{\frac{\sigma^2}{2}(\mathbf{L}-\bar{\mathbf{J}})_{ij}^{-1}}_{noise} \qquad (2.29)$$

where $\mathbf{v}(t)$ *is the* $\frac{\tau}{\mu}$*-periodic attractor of* $\frac{d\mathbf{v}}{dt} = (\bar{\mathbf{J}} - \mathbf{L}).\mathbf{v} + \mathbf{u}(\mu t)$, *where* $\mathbf{J} \in \mathbb{R}^{n\times n}$ *is supposed to be fixed.*

Proof. It consists in applying formally the temporal averaging tools presented in appendix B. It is to be be checked that E_p verifies the hypotheses of proposition 2.4.3 to asymptotic well-posedness of the solutions.

See theorem C.2.1 in appendix C.2.2 for details. □

In the following, we focus on the averaged system described by (2.29). Its right hand side is made of three terms: a linear and homogeneous decay, a correlation term and a noise term. It is exceptional that correlations and noise decouple in the averaged system. This is due to the structure of the system: a linear activity and a quadratic learning rule (without non-linearities ϕ_i). These 2 terms are made explicit in the following

2.4.2.1 Noise term

As seen in section 2.4.1 in the linear case, the noise term $\mathbf{Q}$ is the unique solution of the Lyapunov equation (2.27) with $\Sigma = \sigma Id$. Because the noise is completely uncorrelated and identical for each neuron and also because the connectivity is symmetric observe that $\mathbf{Q} = \frac{\sigma^2}{2}(\mathbf{L} - \bar{\mathbf{J}})^{-1}$ is the unique solution of the system.

In more complicated cases, the computation of this term appears to be much more difficult as we will see in part 2.4.4.

2.4.2.2 Correlation term

This term corresponds to the autocorrelation of the neuronal activity. It is only implicitly defined, thus, this part is devoted to find an explicit form depending only on the parameters l, μ, τ, the connectivity $\mathbf{J}$ and the inputs $\mathbf{u}$.

The autocorrelation term of a $\frac{\tau}{\mu}$-periodic function can be rewritten as

$$\{\mathbf{v}.\mathbf{v}'\}_{ij} = \int_0^{\frac{\tau}{\mu}} \mathbf{v}_i(s)\mathbf{v}_j(s)\ ds$$

With this notation, it is simple to think of $\mathbf{v}$ as a "semi-continuous matrix" of $\mathbb{R}^{n\times[0,\frac{\tau}{\mu}[}$. Hence, the operator "." can be though of as a matrix multiplication. Similarly the transpose operator turns a matrix $\mathbf{v} \in \mathbb{R}^{n\times[0,\frac{\tau}{\mu}[}$ into a matrix $\mathbf{v}' \in \mathbb{R}^{[0,\frac{\tau}{\mu}[\times n}$. See part A for details about the notations.

It is common knowledge, see [Gerstner and Kistler 2002a] for instance, that this term gathers information about the correlation of the inputs. Indeed, if we assume that the input is sufficiently slow, then $\mathbf{v}$ has enough time to converge to $\mathbf{u}(t)$ for all $t \in [0, +\infty[$. Therefore, in the first order $\mathbf{v}(t) \simeq (\mathbf{J}-\mathbf{L})^{-1}.\mathbf{u}(t)$. This leads to $\mathbf{v}.\mathbf{v}' \simeq (\mathbf{J}-\mathbf{L})^{-1}.\mathbf{u}.\mathbf{u}'.(\mathbf{J}'-\mathbf{L})^{-1}$. In the weakly connected regime one can assume that $\mathbf{J}-\mathbf{L} \simeq -\mathbf{L}$ leading to $\mathbf{v}.\mathbf{v}' \simeq \frac{1}{l^2}\mathbf{u}.\mathbf{u}'$ which is the autocorrelation of the inputs. Actually, without the assumption of a slow input, the lagged correlations of the input appear in the averaged system.

Before giving the expression of these temporal correlations, we need to introduce some notations. First, define the convolution filter $g_{\mu/l} : t \mapsto \frac{l}{\mu}e^{-\frac{l}{\mu}t}H(t)$. This family of functions is displayed for different values of $\frac{l}{\mu}$ in figure 2.8.a. It appears that $g_{\mu/l} \longrightarrow \delta_0$ when $\frac{\mu}{l} \longrightarrow 0$, where δ_0 is the Dirac distribution centered at the origin. In this asymptotic regime, the convolution filter and its iterates $g_{\mu/l} * .. * g_{\mu/l}$ are equal to the identity.

We also define the filtered correlation of the inputs $\mathbf{C}^{k,p} \in \mathbb{R}^{n\times n}$ by

$$\mathbf{C}^{k,q} \overset{def}{=} \frac{1}{u_m^2\tau}\big(\mathbf{u} * g_{\mu/l}^{(k+1)}\big).\big(\mathbf{u} * g_{\mu/l}^{(q+1)}\big)'$$

where $g_{\mu/l}^{(k+1)} = g_{\mu/l} * ... * g_{\mu/l}$ is the kth convolution of $g_{\mu/l}$ with itself and $u_m = \sup_{t\in\mathbb{R}_+} \|\mathbf{u}(t)\|_2$. This is the correlation matrix of the inputs filtered by two different functions. It is easy to show that this is similar to computing the cross-correlation of the inputs with the inputs filtered by a another function:

$$\mathbf{C}^{k,q} = \frac{1}{u_m^2\tau}\Big(\mathbf{u} * \big(g_{\mu/l}^{(k+1)} * g_{\mu/l}^{(q+1)'}\big)\Big).\mathbf{u}' = \frac{1}{u_m^2\tau}\mathbf{u}.\Big(\mathbf{u} * \big(g_{\mu/l}^{(q+1)} * g_{\mu/l}^{(k+1)'}\big)\Big)' \quad (2.30)$$

which motivates the definition of the (k,p)-*temporal profile* $g_{\mu/l}^{(k+1)} * g_{\mu/l}'^{(q+1)}$, where $(g_{\mu/l}')^{(k)}(t) = (g_{\mu/l}^{(k)})'(t) = g_{\mu/l}^{(k)}(-t)$. This notation is deliberately similar to that of the transpose operator we use in the proofs. These functions are integrable on $\mathbb{R}$ and infinitely differentiable, therefore, they tend to zero for $t \rightarrow \pm\infty$. Actually, they are bell-shaped functions as shown in figure 2.5. We have not found a way to make them explicit yet, therefore, the next remarks are simply based on numerical illustrations. When $k = q$ the temporal profiles are centered. The larger the difference $k - q$, the larger the center of the bell. The larger the sum $k + q$, the larger the standard deviation. This motivates the idea that $\mathbf{C}^{k,p}$ can be thought of as the $k - q$ lagged correlation of the inputs. One can also say that $\mathbf{C}^{10,10}$ is more blurred than $\mathbf{C}^{0,0}$ in the sense that the inputs are temporally integrated over a ”wider“ window in the first case.

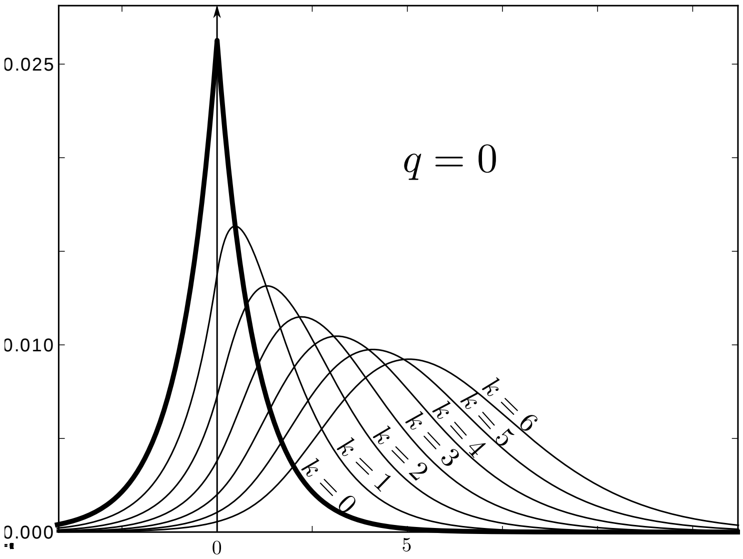


Figure 2.5: This shows the $(k,q)-$temporal profiles with $\frac{\mu}{l} = 1$, i.e. the functions $g_1^{(k+1)} * g_1'^{(q+1)}$, for $q = 0$ and k ranging from 0 to 6. For $k = q = 0$, the temporal profile is even and this also occurs to be true for any $k = q$. When $k > q$ the function reaches its maximum for strictly positive values that grow with the difference $k - q$. Besides, the temporal profiles are flattened when $k + q$ increases.

Observe that $g_{\mu/l}^{(k+1)}(t) = \frac{l^{k+1}}{\mu^{k+1}k!}t^k e^{-\frac{l}{\mu}t}H(t)$. Therefore, $\|g_{\mu/l}^{(k+1)}\|_1 = \frac{\Gamma(k+1)}{k!} = 1$. Thanks to Young's inequality for convolutions, which says that $\|\mathbf{u}*g_{\mu/l}^{(k)}\|_2 \leq \|\mathbf{u}\|_2\|g_{\mu/l}^{(k)}\|_1$, it can be proved that $\|\mathbf{C}^{k,q}\|_2 \leq 1$.

We intend to express the correlation term $\mathbf{v}.\mathbf{v}'$ as an infinite converging sum involving these filtered correlations. In this perspective, we use a new technical result we have proved for this purpose (and published in a separate paper) to write the solution of a general class of non-autonomous linear systems (e.g. $\frac{d\mathbf{v}}{dt} = (\mathbf{J} - \mathbf{L}).\mathbf{v} + \mathbf{u}(t)$) as an infinite sum, in the case $\mu = 1$:

Lemma 2.4.6.*If* $\mathbf{v}$ *is the solution of* $\frac{d\mathbf{v}}{dt} = (\mathbf{J} - \mathbf{L}).\mathbf{v} + \mathbf{u}(t)$ *it can be written by the sum below which converges if* $\mathbf{J}$ *is in* E_p *for* $p \in]0,1[$.

$$\mathbf{v} = \sum_{k=0}^{+\infty} \frac{\mathbf{J}^k}{l^{k+1}} \cdot \mathbf{u} * g_{1/l}^{(k+1)}$$

where $g_{1/l} : t \mapsto le^{-lt}H(t)$.

Proof. See appendix E □

This is a decomposition of the solution of linear differential system on a basis of operators where the spatial and temporal part are decoupled. It is the main result of a technical and more general result which can be found in appendix E. This important step in the detailed study of the averaged equation cannot be achieved easily in models with non-linear activity. Everything is now set up to introduce the explicit expansion of the correlation we are using in the following. Indeed, we use the previous result to rewrite the correlation term as follow

Proposition 2.4.7. *The correlation term can be written*

$$\frac{\mu}{\tau}\mathbf{v}.\mathbf{v}' = \frac{u_m^2}{l^2}\sum_{k,q=0}^{+\infty}\frac{\mathbf{J}^k}{l^k}\cdot\mathbf{C}^{k,q}\cdot\frac{\mathbf{J}'^q}{l^q}$$

Proof. See proposition C.2.3 in appendix C.2.2 for details. □

The speed of the inputs characterized by μ only appears in the temporal profiles $g_{\mu/l}^{(k)} * {g'_{\mu/l}}^{(q)}$. In particular, if the inputs are much slower than neuronal activity time-scale , i.e. $\mu = 0$, then $g_{+\infty} = \delta_0$ and $\mathbf{u} * g_{+\infty} = \mathbf{u}$. Therefore, $\mathbf{C}^{k,q} = \mathbf{C}^{0,0}$ and the sums in the formula of proposition 2.4.7 are separable, leading to $\mathbf{v}.\mathbf{v}' = (\mathbf{L}-\mathbf{J})^{-1}.\mathbf{u}.\mathbf{u}'.(\mathbf{L}-\mathbf{J}')^{-1}$ which corresponds to the heuristic result previously explained.

Therefore, the averaged equation can be explicitly rewritten

$$\frac{d\mathbf{J}}{dt} = \bar{G}(\mathbf{J}) = -\kappa\mathbf{J} + \frac{u_m^2}{l^2}\sum_{k,q=0}^{+\infty}\frac{\mathbf{J}^k}{l^k}\cdot\mathbf{C}^{k,q}\cdot\frac{\mathbf{J}'^q}{l^q} + \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{J})^{-1} \tag{2.31}$$

In figure 2.6, we illustrate this result by comparing, for different ε, the stochastic system and its averaged version. The above decomposition has been used as a basis to numerically compute trajectories of the averaged system.

2.4.2.3 Global stability of the equilibrium point

Now that we have found an explicit formulation for the averaged system, it is natural to study its dynamics. Actually, we prove in the following that if the connectivity is kept smaller than $\frac{l}{3}$, i.e. assumption 2.4.4 is verified with $p \leq 3$, then the dynamics is trivial: the systems converges to a single equilibrium point. Indeed, under the previous assumption, the system can be written $\bar{G}(\mathbf{J}) = -\kappa\mathbf{J} + F(\mathbf{J})$, where F is a contraction operator on $E_{\frac{1}{3}}$. Therefore, one can prove the uniqueness of the fixed point with the Banach fixed point argument and exhibit an energy function for the system.

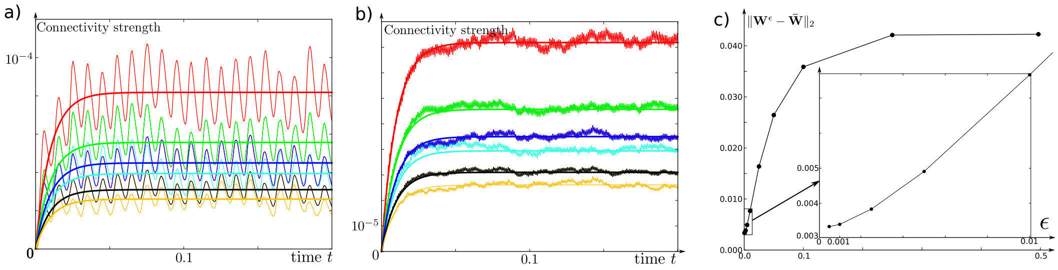


Figure 2.6: The first two figures, a) and b), represent the evolution of the connectivity for the original stochastic system (2.28), superimposed with the averaged system (2.29), for two different values of ε: respectively $\varepsilon = 0.01$ and $\varepsilon = 0.001$. Each color correspond to the weight of an edge in a network made of $n = 3$ neurons. As expected, it seems that the smaller ε the better the approximation. This can be seen in picture c) where we have plotted the precision on the y-axis and ε on the x-axis. The parameters used here are $l = 12.$, $\mu = 1.$, $\kappa = 100$, $\sigma = 0.05$. The inputs have a random (but frozen) spatial structure and evolve according to a sinusoidal function.

Theorem 2.4.8.*If assumption 2.4.4 is verified for $p \leq \frac{1}{3}$ then there is a unique equilibrium point in the invariant subset E_p which is globally, asymptotically stable.*

Proof. See theorem C.2.4 in appendix C.2.2 for details. □

2.4.2.4 Explicit expansion of the equilibrium point

When the network is weakly connected, the high order terms in expansion (2.31) may be neglected. In this part, we follow this idea and find an explicit expansion for the equilibrium connectivity where the strength of the connectivity is the small parameter enabling the expansion. The weaker the connectivity the larger the number of negligible terms in the expansion.

In fact, it is not natural to speak about a weakly connected learning network since the connectivity is a variable. However, we are able to identify a *weak connectivity index* which controls the strength of the connectivity. We say the connectivity is weak when it is negligible compared to the intrinsic leak term, i.e. $\frac{|||\mathbf{J}|||}{l}$ is small. We show in the appendix that this weak connectivity index depends only on the parameters of the network and can be written

$$\tilde{p} = \frac{u_m^2}{\kappa l^3} + \frac{\sigma^2}{2\kappa l^2}$$

In the asymptotic regime $\tilde{p} \to 0$ we have $\frac{\mathbf{J}}{\tilde{p}l} = \mathcal{O}(1)$. This index is the "small" parameter needed to perform the expansion.

We also define $\lambda = \frac{\sigma^2 l}{2u_m^2}$, which has information about the way $\tilde{p}$ is converging to zero. In fact, it is the ratio of the two terms of $\tilde{p}$.

With these, we can prove that the equilibrium connectivity $\mathbf{J}^*$ has the following asymptotic expansion in $\tilde{p}$:

Theorem 2.4.9.

$$\begin{aligned}\mathbf{J}^* = &\frac{\tilde{p}l}{1+\lambda}(\lambda + \mathbf{C}^{0,0}) \\ &+ \frac{\tilde{p}^2 l}{(1+\lambda)^2}\Big(\lambda^2 + \lambda(\mathbf{C}^{0,0} + \mathbf{C}^{1,0} + \mathbf{C}^{0,1}) + \mathbf{C}^{0,0}.\mathbf{C}^{1,0} + \mathbf{C}^{0,1}.\mathbf{C}^{0,0}\Big) \\ &+ \mathcal{O}(\tilde{p}^3)\end{aligned}$$

Proof. See theorem C.2.5 in appendix C.2.2 for details. □

At first order, the final connectivity is $\mathbf{C}^{0,0}$, the filtered correlation of the inputs convolved with a bell-shaped centered temporal profile. In the case of figure 2.7, this is a good approximation of the final connectivity.

Not only the spatial correlation are encoded in the weights but there is also some information about the temporal correlation, i.e. two successive but orthogonal events occurring in the inputs will be wired in the connectivity although they do not appear in the spatial correlations, see figure 2.7 for an example.

2.4.3 Trace learning: Band-pass filter effect

In this section, we study an improvement of the learning model by adding temporal delays in the system and explain the way it changes the results of the previous section. In biological terms, this corresponds to several specific features:

- Trace learning:
  It is likely that a biological learning rule will integrate the activity over

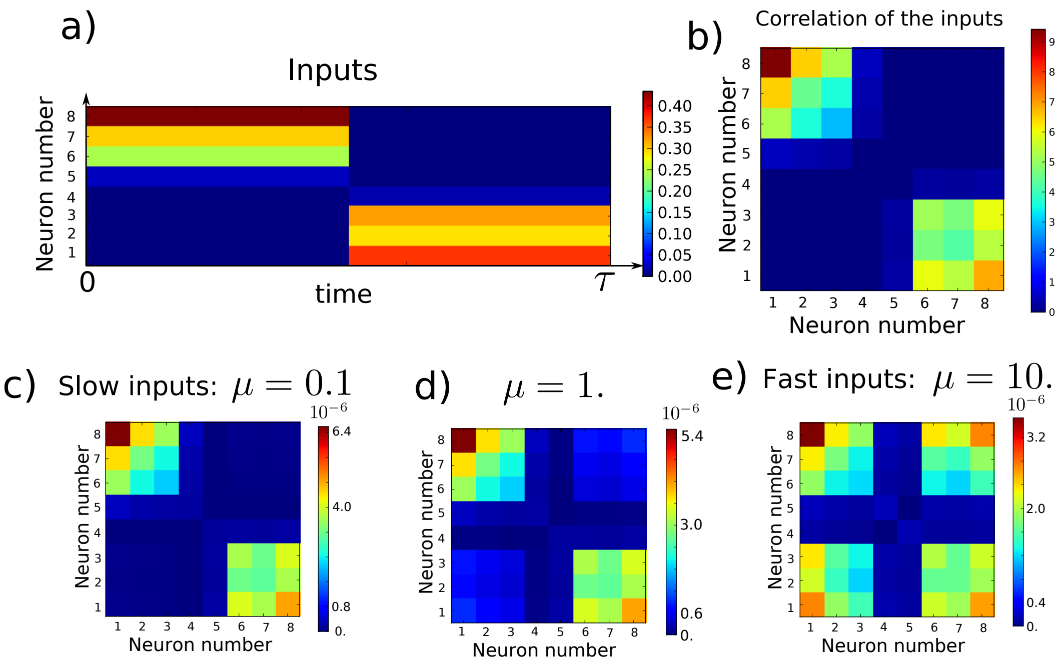


Figure 2.7: Figure a) shows the temporal evolution of the input to a $n = 8$ neurons network. It is made of two spatially random patterns that are shown alternatively. Figure b) shows the correlation matrix of the inputs. The off-diagonal terms are null because the two patterns are spatially orthogonal. The figures c), d) and e) represent the first order of theorem 2.4.9 expansion for different μ. Actually, this approximation is quite good since the percentage of error between the averaged system and the first order, computed by error $= \frac{\|\bar{\mathbf{J}} - \text{order } 1\|_1}{\|\bar{\mathbf{J}}\|_1}$, are $1.92 \times 10^{-4}\%$ for the three figures. These figures make it possible to observe the role of μ. If μ is small, i.e. the inputs are slow, then the transient can be neglected and the learned connectivity is roughly the correlation of the inputs, see figure a). If μ increases, i.e. the inputs are faster, then the connectivity starts to encode a link between the two patterns that were flashed circularly and elicited responses that did not fade away when the other pattern appeared. The temporal structure of the inputs is also learned when μ is large. The parameters used in this figure are $\varepsilon = 0.001$, $l = 12$, $\kappa = 100$, $\sigma = 0.02$.

a short time. As Földiàk suggested in [Földiák 1991], it makes sense to consider the learning equation as being

$$\frac{d\mathbf{J}}{dt} = -\kappa\mathbf{J} + (\mathbf{v} * g_1) \otimes (\mathbf{v} * g_1)$$

where $*$ is the convolution and $g_1 : t \in \mathbb{R} \mapsto \beta_1 e^{-\beta_1 t} H(t)$. Rolls numerically shows in chap 8 of [Rolls and Deco 2002] that the temporal convolution, leading to a spatio-temporal learning, makes it possible to perform invariant object recognition. Besides, trace learning appears to be the symmetric part of the biological STDP rule that we detail in section 2.4.4.

- Damped oscillatory neurons:
  Many neurons have an oscillatory behavior. Although we can not take this into account in a linear model, we can model a neuron by a damped oscillator, which also introduces a new important time-scale in the system. Adding adaptation to neuronal dynamics is an elementary way to implement this idea. The neurons, or populations of neurons, of the final averaged system (1.34) of chapter 1 have such a damped oscillatory behavior which motivates this choice.

  This corresponds to modeling a single neuron without inputs by the equivalent formulations

$$\left\{ \begin{array}{rl} \frac{dv}{dt} &= -lz \\ \frac{dz}{dt} &= \beta_2(v - z) \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rl} \frac{dv}{dt} &= -lv * g_2 \\ \text{where } g_2(t) &= \beta_2 e^{-\beta_2 t} H(t) \end{array} \right.$$

- Dynamic synapses:
  The electro-chemical process of synaptic communication is very complicated and non-linear. Yet, one of the features of synaptic communication we can take into account in a linear model is the shape of the post-synaptic potentials. In this section, we consider that each synapse is a linear filter whose finite impulse response (i.e. the post-synaptic potential) has the shape $g_3(t) = \beta_3 e^{-\beta_3 t} H(t)$. This is a common assumption which, for instance, is at the basis of traditional rate based models, see section 1.2.1.

For mathematical tractability, we assume in the following that $\beta = \beta_1 = \beta_2 = \beta_3 \in \mathbb{R}_+$ such that $g_{1/\beta} = g_1 = g_2 = g_3$, i.e. the time scales of the neurons, those of the synapses and those of the learning windows are the same. Actually, there is a large variety of the temporal scales of the neurons, synapses, and learning windows, which makes this assumption not absurd. Yet, investigating the impact of breaking this assumption would be necessary to model better biological networks. This leads to the following system

$$\begin{cases} d\mathbf{v} &= \frac{1}{\varepsilon_1}\Big((\mathbf{J}^\varepsilon - \mathbf{L}).\mathbf{v} * g_{1/\beta} + \mathbf{u}(\frac{t}{\varepsilon_2})\Big)dt + \frac{\sigma}{\sqrt{\varepsilon_1}}dB(t) \\ \frac{d\mathbf{J}^\varepsilon}{dt} &= -\kappa\mathbf{J}^\varepsilon + (\mathbf{v} * g_{1/\beta}) \otimes (\mathbf{v} * g_{1/\beta}) \end{cases} \tag{2.32}$$

where the notations are the same as in section 2.4.2. In fact, the behavior of a single neuron will be oscillatory damped if $\Delta = \sqrt{1 - 4\frac{l}{\beta}}$ is a pure imaginary number, i.e. $4l > \beta$. This is the regime on which we focus. Actually, the Hebbian linear case of section 2.4.2 corresponds to $\beta = +\infty$ in this delayed system.

It turns out most of the results of section 2.4.2 remain true for system (2.32) as detailed in the following. The existence of the solution on $\mathbb{R}_+$ is proved in theorem C.2.6. The computations show that, in the averaged system, the noise term remains identical, whereas the correlation term is to be replaced by $\frac{\mu}{\tau}(\mathbf{v} * g_{1/\beta}).(\mathbf{v} * g_{1/\beta})'$. Besides, lemma 2.4.6 can be extended to our delayed system by changing only the temporal filters, see appendix E. This helps proving the following (see C.2.8 and C.3.3 for details).

$$\frac{\mu}{\tau}(\mathbf{v} * g_{1/\beta}).(\mathbf{v} * g_{1/\beta})' = \frac{u_m^2\|v\|_1^2}{l^2}\sum_{k,q=0}^{+\infty}\frac{\mathbf{J}^k}{(l/\|v\|_1)^k}\cdot\tilde{\mathbf{C}}^{k,q}\cdot\frac{\mathbf{J}'^q}{(l/\|v\|_1)^q}$$

where

$$\tilde{\mathbf{C}}^{k,q} = \frac{1}{u_m^2\tau\|v\|_1^{k+q+2}}(\mathbf{u} * v^{(k+1)}) \cdot (\mathbf{u} * v^{(q+1)})'$$

where $v : t \to \frac{l}{\mu\Delta}\Big(e^{-\frac{\beta}{2\mu}(1-\Delta)t} - e^{-\frac{\beta}{2\mu}(1+\Delta)t}\Big)H(t)$. Observe that applying Young's inequality for convolutions leads to $\|\tilde{\mathbf{C}}^{k,q}\|_2 \leq 1$. Actually, lemma C.3.3 shows that $v^{(k)} = v_k : t \mapsto \frac{\sqrt{\pi\beta}}{k!}e^{-\frac{\beta}{2}t}\Big(\frac{t}{|\Delta|}\Big)^{k+\frac{1}{2}}J_{k+\frac{1}{2}}\big(\frac{\beta|\Delta|}{2}t\big)H(t)$ where $J_n(z)$ is the Bessel function of the first kind. The value of the L1 norm of v is computed in appendix C.3.3, it leads to $\|v\|_1 = coth\big(\frac{\pi}{2\Delta}\big)$ if Δ is a pure imaginary number and $\|v\|_1 = 1$ else.

Therefore, the averaged system can be rewritten

$$\frac{d\mathbf{J}}{dt} = \bar{G}(\mathbf{J}) = -\kappa\mathbf{J} + \frac{u_m^2\|v\|_1^2}{l^2}\sum_{k,q=0}^{+\infty}\frac{\mathbf{J}^k}{(l/\|v\|_1)^k}\cdot\tilde{\mathbf{C}}^{k,q}\cdot\frac{\mathbf{J}'^q}{(l/\|v\|_1)^q} + \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{J})^{-1}$$

As before the existence and uniqueness of a globally attractive equilibrium point is guaranteed if assumption 2.4.4 is verified for $p \leq \frac{1}{3}$, see theorem C.2.9.

Besides, the weakly connected expansion of the equilibrium point we did in part 2.4.2.4 can be derived in this case (see theorem C.2.10). At first order, this leads to the equilibrium connectivity

$$\mathbf{J}^* = \frac{\tilde{p}l}{1+\lambda}(\lambda + \|v\|_1^2 \tilde{\mathbf{C}}^{0,0}) + \mathcal{O}(\tilde{p}^2 \|v\|_1)$$

The second order is given in the appendix C.2.10. The only difference with the Hebbian linear case is the shape of the temporal filters. Actually the temporal filters have an oscillatory damped behavior if Δ is purely imaginary. These two cases are compared in figure 2.8.

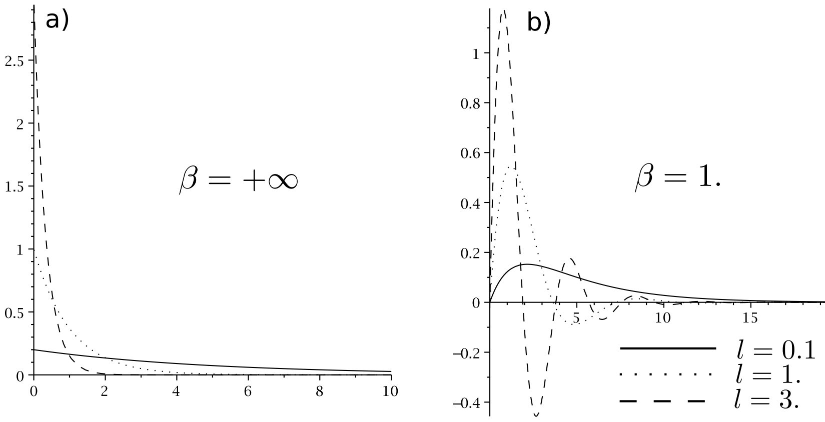


Figure 2.8: These represent the temporal filter $v : t \mapsto v(t)$ for different parameters. a) When $\beta = +\infty$, we are in the Hebbian linear case of section C.2.2. The temporal filters are just decaying exponentials which averaged the signal over a past window. b) When the dynamics of the neurons and synapse are oscillatory damped, some oscillations appear in the temporal filters. The number of oscillations depends on Δ. If Δ is real, then there are no oscillations as in the previous case. However, when Δ becomes a pure imaginary number it creates a few oscillations which are even more numerous if $|\Delta|$ increases.

These oscillatory damped filters have the effect of amplifying a particular frequency of the input signal. As shown in figure 2.9, if Δ is a pure imaginary number then $\tilde{\mathbf{C}}^{0,0}$ is the cross-correlation of the band-pass filtered inputs with themselves. This band-pass filter effect can also be observed in the higher order terms of the weakly connected expansion. This suggests that the bio-physical oscillatory behavior of neurons and synapses leads to selecting the

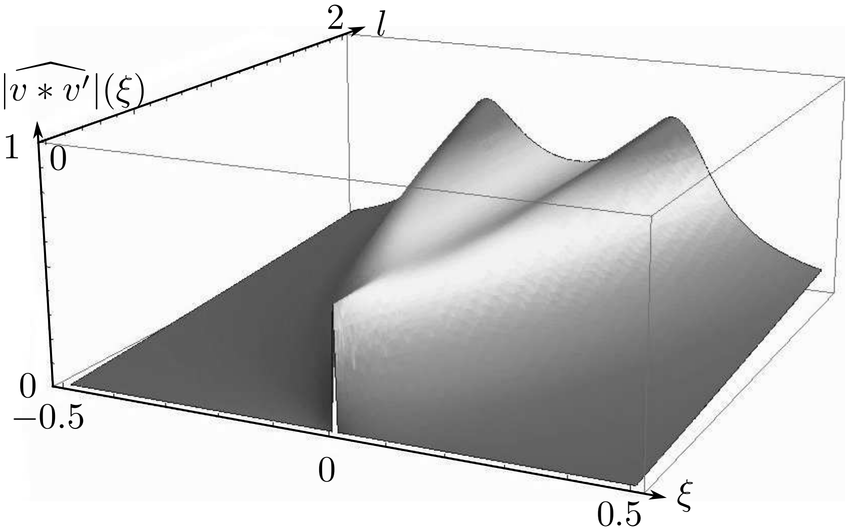


Figure 2.9: This is the spectral profile $|\widehat{v*v'}|(\xi)$ for $\beta = 1$ and $l \in]0, 2]$. When $4l < \beta$ the filter reach its maximum for the null frequency, but if l increases beyond $\frac{\beta}{4}$ the filter becomes of band-pass filter with long tails in $\frac{1}{\xi^2}$.

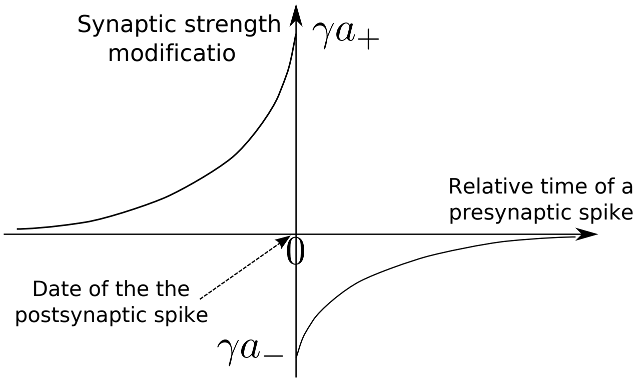
corresponding frequency of the inputs and performing the same computation as for the Hebbian linear case of the previous section.

2.4.4 Asymmetric STDP learning with correlated noise

Here, we extend the results to temporally asymmetric learning rules and spatially correlated noise. We consider a learning rule that is similar to the Spike Timing Dependent Plasticity (STDP) which is closer to biological experiments than the previous Hebbian rules, see part 1.1.4 for details. It has been observed that the strength of the connection between two neurons depends mainly on the difference between the time of the spikes emitted by each neuron as shown in figure 2.10, see [Caporale and Dan 2008]. Note that our approach is valid for any value of $a_+, a_- \in \mathbb{R}_+$, in particular we will not need to have a stronger depression than facilitation to prove stability of the system. As suggested by section 1.2.3, we consider here that this learning rule also apply to rate based models.

This leads to the coupled system

$$\left\{ \begin{array}{ll} d\mathbf{v} & = \frac{1}{\varepsilon_1}\Big(- \mathbf{L}.\mathbf{v} + \mathbf{J}^\varepsilon.\mathbf{v} + \mathbf{u}(\frac{t}{\varepsilon_2})\Big)dt + \frac{1}{\sqrt{\varepsilon_1}}\mathbf{\Sigma}.dB(t) \\ \frac{d\mathbf{J}^\varepsilon}{dt} & = -\kappa \mathbf{J}^\varepsilon + a_+ \mathbf{v} \otimes (\mathbf{v} * g_{1/\gamma}) - a_-(\mathbf{v} * g_{1/\gamma}) \otimes \mathbf{v} \end{array} \right. \tag{2.33}$$



Synaptic strength modificatio
γa_+
Relative time of a presynaptic spike
0
Date of the the postsynaptic spike
γa_-

Figure 2.10: This figure represents the synapse strength modification when the post-synaptic neuron emits a spike. The y-axis corresponds to an additive or multiplicative update of the connectivity. For instance, in [Bi and Poo 1998], this is $\frac{\Delta \mathbf{J}_{ij}}{\mathbf{J}_{ij}}$. However, we assume an additive update in this thesis. The x-axis is the time at which a pre-synaptic spike reaches the synapse, relatively to the time of post-synaptic time chosen to be 0.

In this framework, the method exposed in section C.2.2 holds with small changes. First, the well-posedness assumption becomes

Assumptions 2.4.10.

There exists $p \in]0,1[$, *such that* $\frac{|a_+|+|a_-|}{p(1-p)}\left(\frac{s^2\gamma}{2(1+\gamma/l-p)}+\frac{u_m^2}{(1-p)}\right) < \kappa l^3$

where s^2 *is the maximal eigenvalue of* $\mathbf{\Sigma}.\mathbf{\Sigma}'$.

Under this assumption the system is asymptotically well-posed in probability (theorem C.2.12). And we show the averaged system is

$$\frac{d\mathbf{J}}{dt} = \bar{G}(\mathbf{J}) = -\kappa \mathbf{J} + \frac{u_m^2\big(|a_+|+|a_-|\big)}{l^2}\sum_{k,q=0}^{+\infty}\frac{\mathbf{J}^k}{l^k}\cdot \mathbf{D}^{k,q}\cdot\frac{\mathbf{J}'^q}{l^q} + \mathbf{Q} \tag{2.34}$$

where we have used theorem C.2.13 to expand the correlation term. The noise term $\mathbf{Q}$ is equal to $\mathbf{Q}_{11}.(\mathbf{L}+\gamma-\mathbf{J}')^{-1}$ where $\mathbf{Q}_{11}$ is the unique solution of the Lyapunov equation $(\mathbf{J}-\mathbf{L}).\mathbf{Q}_{11}+\mathbf{Q}_{11}.(\mathbf{J}'-\mathbf{L})+\mathbf{\Sigma}.\mathbf{\Sigma}'=0$. Lemma D.3.2 gives a solution for this equation which leads to $\mathbf{Q}=\gamma\sum_{k=0}^{+\infty}\mathbf{J}^k.\mathbf{\Sigma}.\mathbf{\Sigma}'.(2\mathbf{L}-\mathbf{J}')^{-(k+1)}.(\mathbf{L}+\gamma-\mathbf{J}')^{-1}$. In equation (2.34), the correlation matrices $\mathbf{D}^{k,q}$ are given by

$$\mathbf{D}^{k,q} = \frac{1}{u_m^2\tau\big(|a_+|+|a_-|\big)}\Big(\mathbf{u}*g_{\mu/l}^{(k+1)}*\big(a_+g'_{1/\gamma}-a_-g_{1/\gamma}\big)\Big)\cdot(\mathbf{u}*g_{\mu/l}^{(q+1)})'$$

According to theorem C.2.14, The system is also globally asymptotically convergent to a single equilibrium which we study in the following.

We perform a weakly connected expansion of the equilibrium connectivity of system (2.34). As shown in theorem C.2.15, the first order of the expansion is

$$\mathbf{J}^* = \frac{\tilde{p}l}{1+\lambda}\big(\lambda(\alpha_+ - \alpha_-)\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d} + \mathbf{D}^{0,0}\big) + \mathcal{O}(\tilde{p}^2)$$

Writing $\mathbf{D}^{0,0} = \mathbf{S} + \mathbf{A}$, where $\mathbf{S}$ is symmetric and $\mathbf{A}$ is skew-symmetric, leads to

$$\begin{aligned}\mathbf{S} &= \tfrac{a_+ - a-}{2\tau}\mathbf{u} * g_{\mu/l} * \big(g'_{1/\gamma} + g_{1/\gamma}\big) \cdot (\mathbf{u} * g_{\mu/l})' \\ \mathbf{A} &= \tfrac{a_+ + a-}{2\tau}\mathbf{u} * g_{\mu/l} * \big(g'_{1/\gamma} - g_{1/\gamma}\big) \cdot (\mathbf{u} * g_{\mu/l})'\end{aligned}$$

According to lemma C.3.1, the symmetric part is very similar to trace learning in section 2.4.3. Applying lemma C.3.2 leads to

$$\boxed{\begin{array}{rcl}\mathbf{S} & = & \tfrac{a_+ - a-}{\tau}\big(\mathbf{u} * g_{\mu/l} * g_{1/\gamma}\big) \cdot \big(\mathbf{u} * g_{\mu/l} * g_{1/\gamma}\big)' \\ \mathbf{A} & = & \tfrac{a_+ + a-}{\gamma\tau}\big(\tfrac{d\mathbf{u}}{dt} * g_{\mu/l} * g_{1/\gamma}\big) \cdot \big(\mathbf{u} * g_{\mu/l} * g_{1/\gamma}\big)'\end{array}} \tag{2.35}$$

Therefore, the STDP learning rule simply adds an antisymmetric part to the final connectivity keeping the symmetric part as the Hebbian case. Besides, the antisymmetric part corresponds to computing the cross-correlation of the inputs with its derivative. For high order terms this remains true although the temporal profiles are different from the first order. This results are in line with previous works underlying the similarity between STDP learning and differential Hebbian learning, where $G(\mathbf{v}) \sim \dot{\mathbf{v}} \otimes \mathbf{v}$, see [Xie and Seung 2000].

2.4.5 Summary, conclusions and immediate extensions

We have applied temporal averaging methods on slow/fast systems modeling learning mechanisms occurring in linear stochastic neural networks. When we make sure the connectivity remains small, the dynamics of the averaged system appears to be simple: the connectivity always converges to a unique equilibrium point. Then we performed a weakly connected expansion of this final connectivity to get explicit approximations. The terms in the expansion are combinations of the noise covariance and the lagged correlations of the inputs: the first order term is simply the sum of the noise covariance and the correlation of the inputs.

- As opposed to the former Input/Output vision of the neurons, we have considered the membrane potential $\mathbf{v}$ to be the solution of a dynamical system. The consequence of this modeling choice is that not only the spatial correlations are learned but also the temporal correlations. Due to the fact we take the transients into account, the activity never converges but it lives between the representation of the inputs. Therefore, it links concepts together.

  The parameter μ is the ratio of the timescales between the inputs and the activity variable. If $\mu = 0$ the inputs are infinitely slow and the activity variable has enough time to converges towards its equilibrium point. When μ grows, the dynamics becomes more and more transient, it has no time to converge. Therefore, if the inputs are extremely slow the networks only learns the spatial correlation of the inputs. If the inputs are fast, it also learns the temporal correlations. This is illustrated in figure 2.7.

  This suggests that learning associations between concepts, for instance learning a words in two different languages, may be optimized by presenting the two words to be associated circularly with a certain frequency. Indeed, increasing the frequency (with a fixed duration of exposition to each word) amounts to increasing μ. Therefore, the networks learns better the temporal correlations of the inputs and thus strengthens the link between these 2 concepts.

- Section 2.4.3 suggests that neurons and synapses with a preferred frequency of oscillation will preferably extract the correlation of the inputs filtered by a band pass filter centered on the intrinsic frequency of the neurons.

  Actually, it has been observed that the auditory cortex is tonotopically organized, i.e. the neurons are arranged by frequency [Romani et al. 1982]. It is traditionally thought that this is achieved thanks to a particular connectivity between the neurons. We exhibit here another mechanism to select this frequency which is solely based on the parameters of the neurons: a network with a lot of different neurons whose intrinsic frequencies are uniformly spread is likely to perform a Fourier-like operation: decomposing the signal by frequency.

In particular, this emphasizes the fact that the network does not treat space and time similarly. Roughly speaking, associating several pictures or several sounds are therefore two different tasks which involves mechanisms of two different kinds.

- In this section, the original hierarchy of the network has been neglected: the network is made of neurons which receives external inputs. A natural way to include a hierarchical structure (with layers for instance), without changing the setup of the section, is therefore to remove the external input to some neurons. However, according to theorem 2.4.9 (and its extensions C.2.10 and C.2.15), one can see that these neurons will be disconnected from the others at first order (if the noise is uncorrelated). Linear activities imply that the high level neurons disconnect from the others, which is a problem. In fact, one can observe that the second order term in 2.4.9 is not null if the noise matrix $\mathbf{\Sigma}$ is not diagonal. In fact, this is the noise between neurons which will recruit the high level neurons to build connections from and to them.

  It is likely that a significant part of the noise in the brain is locally induced, e.g. local perturbations due to blood vessels or local hormonal signals. In a way, the neurons close to each other share their noise and it seems reasonable to choose the matrix $\mathbf{\Sigma}$ so that it reflects the biological proximity between neurons. In fact, $\mathbf{\Sigma}$ specifies the original structure of the network and makes it possible for close neurons to recruit each other.

- It is also interesting to observe that most of the noise contribution to the equilibrium connectivity for STDP learning (see C.2.15) vanishes if the learning is purely skew-symmetric, i.e. $a_+ = a_-$. The analysis proves that it is only the symmetric part of learning, i.e. the Hebbian mechanism, which writes the noise in the connectivity. Indeed, the role of a stationary noise in a purely antisymmetric network would be null as far as learning is concerned. Indeed, noise will be integrated equally on both negative and positive sides of the learning window. Because, these two parts of the windows are of opposite sign (in the antisymmetric case) then the integration of the noise cancels out.

- We have shown that there is a natural analogue of the STDP learning for spiking neurons in our case of linear neurons. This asymmetric rule converges to a final connectivity which can be decomposed into symmetric and skew-symmetric parts. The first one is similar to the symmetric Hebbian learning case, emphasizing that the STDP is in fact an asymmetric Hebbian-like learning rule. The skew-symmetric part of the final connectivity is the cross correlation between the inputs and their derivative. In the next part we explore the functional implications of this statement.

Chapter 3

Neural networks post-learning as models of their environment

Overview

This chapter is devoted to understanding the dynamics of the network post-learning and compare it to the inputs. We first review the notion of hierarchy, which we lack in our fully recurrent formalism. Second, we introduce cortical maps and various models for their emergence. Then we start our analysis of the equilibrium connectivity. We claim that interpreting its symmetric part as a distance matrix leads to giving a position to the neurons which corresponds to the underlying geometrical structure of the inputs. We then show how it may relate to cortical maps. Finally, we study the anti-symmetric part of the connectivity and show it can be seen as a vector field which corresponds to that of the dynamical system defining the inputs. In a way the network sees the inputs as the solution of a dynamical system and learn the structure of the dynamical system. It can then be a predictor of the inputs.

In this chapter sections 3.1 and 3.2 are mainly background, whereas sections 3.3 and 3.4 are original

Résumé

Ce chapitre porte sur la dynamique d'un réseau de neurones post-apprentissage et la compare aux entrées externes présentées lors de l'apprentissage. Dans un premier temps nous faisons une revue de la notion de hierarchie qui n'est pas bien prise en compte dans notre cadre de travail complètement récurrent. Ensuite nous introduisons les cartes corticales et les différents modèles décrivrant leur émergence. Après ces deux revues, nous sommes en position pour analyser la connectivité à l'equilibre résultat du chapitre précédent. Nous proposons que la partie symmétrique de cette matrice peut être vue comme une matrice de distance ce qui correspond à donner une position aux neurones. Nous suggérons que les positions des neurones correspondent à la structure géométrique sous-jacente des entrées. En particulier nous montrons comment ceci peut être relié aux cartes corticales. Enfin nous étudions la partie anti-symmétrique de la connectivité à l'équilibre et montrons comment elle peut être interprétée comme un champ de vecteur qui correspond au système dynamique décrivant les entrées. Ceci suggère que le réseau voit les entrées comme la solution d'un système dynamique et apprend la structure de ce système. En fait le réseau devient un prédicteur des entrées.

Dans ce chapitre les sections 3.1 et 3.2 sont principalement issues de la littérature, alors que les sections 3.3 et 3.4 sont originales.

Collaborations, publications and ackowledgements

Section 3.3 was written in collaboration with Paul Bressloff and Olivier Faugeras when I visited Paul in Oxford for a semester and lead to the publication of an article in neural computation. Section 3.4 is work in progress and will be submitted after the thesis.

.

Contents

What kind of information is learned from the stimuli? How is it stored in the network? How is it used to process new inputs? How does it discriminate between inputs belonging to the training set or original stimuli? We still lack a coherent and satisfying answer to these simple questions. The previous chapter may bring some elements to understand the role of learning in recurrent neural networks. Indeed, it was assumed that learning corresponds to the slow modification of the connectivity of a neural network. Therefore, the global knowledge about the environment is gathered in the equilibrium connectivity only. More precisely, it was shown that the equilibrium connectivity of a STDP learning rule is made of two parts: (i) a symmetric part corresponding to the spatio-temporal auto-correlation of the inputs (without time-lag), (ii) an antisymmetric part corresponding to the spatio-temporal cross-correlation between the inputs and their derivative. This may sound as a satisfying answer to the initial questions, yet, we believe it is possible and necessary to go further.

A key idea is to consider the inputs as a solution of a dynamical system. Indeed, because we want to give a mathematical sense to the claim: ”a learning neural network builds a model of its environment“, we are looking for a relevant way to compare the inputs to a neural network, in order to show that learning corresponds to a convergence of the neural network to the inputs. If we assume that the inputs are generated by a single dynamical system, then the entire information about the inputs is contained in the dynamical system which is defined by a vector-field upon a manifold. The right hand side of the dynamical system at a certain point on the manifold is the value of the vector field at this position. Therefore, the entire knowledge of the inputs is gathered in a vector field on a manifold which both define the dynamical system generating the inputs. Our goal is to show that these two mathematical objects, the manifold and the vector field associated to the inputs, are eventually embedded in the connectivity of the neural network. In other words, the thesis of this chapter is the following: through slow learning the spontaneous activity becomes statistically and dynamically identical to the inputs. We believe the spontaneous activity post-learning is the solution to a differential system with a very similar manifold and vector field. There are three main reasons to support this approach: (i) both the inputs and the spontaneous activity can be seen as the solution of a dynamical system, making them two objects of the same nature and therefore comparable (ii) its has been rigorously shown that any dynamical system (in particular the system that generates the inputs) can be approximated by a recurrent neural network [Funahashi and Nakamura 1993] (iii) the statistical features of spontaneous activity seem to reproduce that of the inputs, suggesting that the spontaneous activity replays the inputs [Kenet et al. 2003, Han et al. 2008, Berkes et al. 2011]. Actually, these papers tend to show that the spontaneous activity is similar to the attractors evoked by natural inputs and no the inputs themselves because the latter link is impossible to measure. Yet, with our semi-analytical approach we can compare the spontaneous activity to the inputs which will be the approach in the following.

Unfortunately, the rigorous proof of the problematic of this thesis is beyond our capabilities and we have had to simplify the problem significantly. In particular, the cortex is processing information at multiple scales which we are not yet able to take into account. Understanding the interactions of these different temporal and spatial scales is a current topic of research[1]. It is generally though that the hierarchical organization of the cortex is responsible for this multi-scale processing. Although most studies on the development of cortical structures were done in such a hierarchical context, this thesis focuses on completely recurrent network (without hierarchy).

We use a semi-analytic, semi-computational approach to show the network learns to extract the underlying geometry (i.e. manifold) and dynamics (i.e. vector field) of the inputs on some examples. This chapter is therefore nothing more than a proof of concept. We present some inputs with an explicit geometry and/or dynamics, then we show the network can retrieve them. It has been suggested in [Swindale 1996] that the geometry of the inputs may be embedded in the experimentally observed spatial cortical patterns called cortical maps. We show how our approach can reduce to the development of cortical maps and qualitatively match the experiments.

This chapter is divided in four parts, first we shortly talk about the role of hierarchy which seems to be responsible for the multi-scale processing reported above in section 3.1. The functional role of learning was originally addressed in the framework of feed-forward network which we shortly review before

[1] for instance the author belongs to a European Project called BrainScaleS (http://brainscales.kip.uni-heidelberg.de/) which focuses on the study of multi-scale phenomena in the brain.

switching completely to recurrent networks which are the building blocks of more relevant hierarchies. Then, we shortly introduce cortical maps and their developmental models in section 3.2. The third section 3.3 corresponds to the analysis of the symmetric part of the connectivity. We show that interpreting it as a distance matrix leads to extracting the underlying geometry of the inputs. Finally, we address the skew-symmetric part of the connectivity in 3.3 which will be seen as a vector field defined on the geometric shape introduced in the previous part.

3.1 A glimpse of hierarchy

Here, we shortly present the notion of hierarchical neural network. First, we motivate it as a necessary feature to account for multiple scales. Second, we introduce some important results about correlation-based learning rules in feed-forward networks (probably the simplest hierarchical networks). Finally, we present the different ingredients of a canonical hierarchical network whose study is beyond the scope of this thesis. In this framework, recurrent networks (which are the topic of this thesis) appear to be the building blocks of such hierarchical structures.

3.1.1 Functional role of hierarchy

The organization of the cortex is hierarchical: the sensory inputs are not uniformly plugged to the neurons. Some neurons are significantly influenced by the inputs; whereas others are far away from the inputs (meaning the information has to pass through a lot of intermediary neurons). Actually, biological experiments report a roughly vertical organization of the different areas in the cortex, see figure 3.1 for instance, allowing the definition of low-level and high-level areas.

The notion of receptive field gives a heuristic intuition about the functional role of hierarchy. In a feed-forward framework, the receptive field is defined as the region of space in which the presence of a stimuli will influence the behavior of a neuron. As an example, we focus on vision in the following and we refer to [Rolls and Deco 2002] for an easy introduction to the organization of the visual system. Roughly speaking, neurons in the retina or in the lateral geniculate nucleus (LGN) (the gray boxes on the left picture of figure 3.1) are excited when some light reaches the part of the retina they are in charge of representing. Their receptive field is a small part of the visual field. More precisely, they a have a center-surround receptive field: they are maximally excited when the stimulus corresponds to a dot of light in the middle and dark surrounds (or reciprocally). One step higher in the hierarchy is the primary visual cortex (V1). Hubel and Wiesel have received a Nobel prize for showing that some neurons in V1 where tuned to detecting edge. Indeed, they are optimally excited by small oriented edges with a given orientation in the visual field, see [Hubel and Wiesel 1962]. This discovery led to the analysis

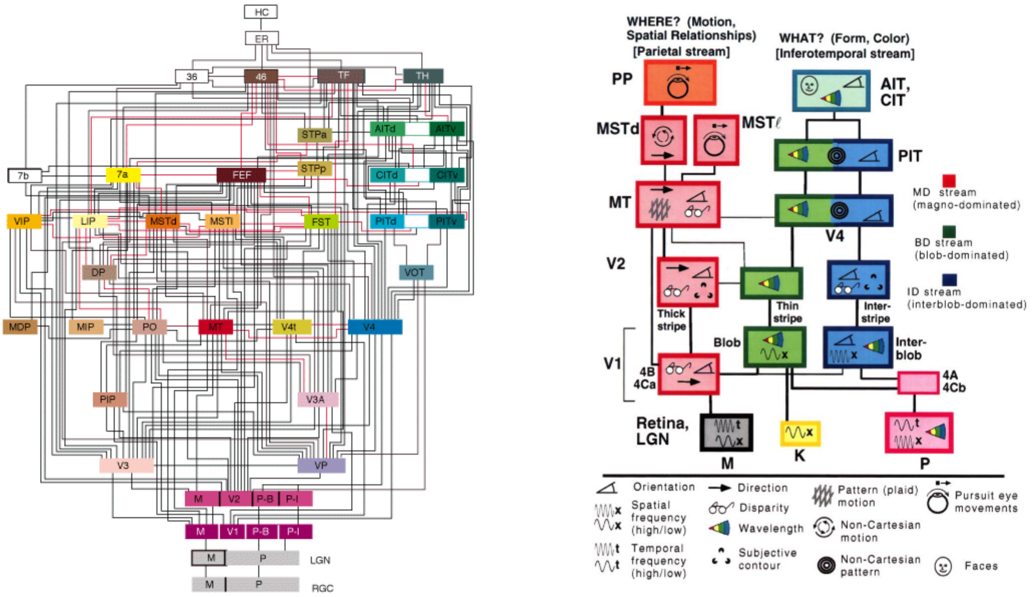


Figure 3.1: (left) Hierarchical organization of the macaque visual system in 32 areas connected via more than 300 pathways. The inputs arrive on the bottom box which correspond to the Retinal Ganglion Cells (in the retina). Taken from [Felleman and Van Essen 1991] (right) This is a simplification of Felleman diagram that highlights the existence of two different pathways: the ventral (what) and dorsal (where) streams. Taken from [Van Essen and Gallant 1994]

of the receptive field of neurons higher in the hierarchy. It was observed that their shape grew larger and more complex at each step in the hierarchy. It was even observed that high enough in the hierarchy some neurons where excited when human faces appeared in the visual field, as shown the right picture of figure 3.1. These results tend to show that neurons *code for* some concepts and that these concepts grow more abstract and complicated when we go up the hierarchy.

However, this view of neural processing is not relevant to describe recurrent network subject to spontaneous activity. Indeed, the brain is to be seen as a dynamical system perturbed by the inputs. The neurons are not only influenced by the inputs but also by the internal state of the brain. More concretely, a particular neuron can be excited because of internal variability or connections from high-level neurons. Besides, the notion of receptive field is irrelevant for high-level neurons which might be triggered by a particular event anywhere in the visual field. In a way, the notion of receptive field can only make sense if the neuron receives most of its excitation from the raw inputs (yet, it does not mean that these sensory neurons all have a well-defined receptive field). The notion of receptive field is only a nice intuition for the low-level part of the network.

The role of hierarchy is to build concepts on multiple scales. Based on a low-level description of the world hierarchical learning makes it possible to define the meaning a higher-order concept. Based on the lower-level understanding of the world the neurons in the area above would code for a frequent pattern observed in the lower layer(s). In a hierarchical network the total connectivity can be separated in two parts: the vertical and the lateral connectivity. The former links neurons from different layers, while the second links neurons from the same layer. A intuitive vision of their respective role is that learning of the vertical connections defines the meaning of a neuron (what it codes for in terms of the lower-level concepts), while learning the lateral connections defines the way these concepts are statistically linked together.

3.1.2 Unsupervised perceptron: a study case

3.1.2.1 What is a perceptron?

The study of learning rules has been introduced and extensively studied in the case of a perceptron network, see [Dayan and Abbott 2001] for a review. It is the simplest feed-forward network: $n \in \mathbb{N}$ pre-synaptic neuron sending their axon to a single post-synaptic neuron, see figure 3.2. This structure is at the heart of any hierarchical structure, therefore the results on perceptrons correspond to the learning of vertical connections in the cortex.

The storage capacity of the perceptron under various rules has been widely studied [Gardner and Derrida 1988, Engel and Broeck 2001, Brunel et al. 2004] but is out of the scope of this thesis.

The evolution of the variables still evolve according to equations (2.1) and (2.2) with a particular connectivity mask $\boldsymbol{\chi}$ detailed below, so that most of

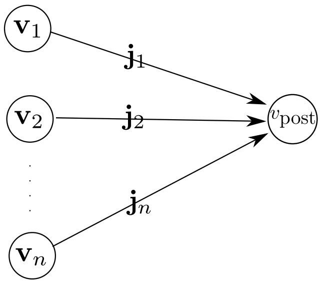


Figure 3.2: A perceptron is a network with a single post-synaptic neuron, the connectivity is a vector of $\mathbb{R}^n$ (thus noted with a lowercase letter).

the results of the previous part hold.

$$\boldsymbol{\chi} = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \tag{3.1}$$

A great deal of work about the (multilayer) perceptron has been done in the field of supervised learning. This is out of the scope of this thesis. We focus on unsupervised correlation-causation based learning rules on this structure. Actually, all the rules in part 2.2 can be studied in this feed-forward framework. However, we will only focus on three of them for the striking conclusions they lead to.

3.1.2.2 Oja learning rule and PCA extraction

This learning rule introduced in [Oja 1982] was described in equation (2.7) for a recurrent network. We also refer to [Miller and MacKay 1994] for a generalization to other rules. We follow the lines of this reference in this short section.

In the traditional approach used for most studies, it is assumed that the post-synaptic activity is a linear combination of the inputs (which is opposed the dynamical system we have in this thesis), i.e. $v_{\text{post}} = \sum_{k=1}^{n} \mathbf{j}_k \mathbf{v}_k = \mathbf{j}'.\mathbf{v}$

In the perceptron framework, it reads

$$\frac{d\mathbf{j}}{dt} = v_{\text{post}}\mathbf{v} - v_{\text{post}}^2\mathbf{j} = (\mathbf{j}'.\mathbf{v})\mathbf{v} - (\mathbf{j}'.\mathbf{v})^2\mathbf{j}$$

As shown in part 2.2.2, it is easy to show that the L_2 norm of the connectivity is kept constant (equal to 1) during learning.

If this equation is slow, it seems reasonable to average the equation over the inputs. Besides, using the assumption that $v_{\text{post}} = \mathbf{j}'.\mathbf{v}$ leads to defining the averaged system

$$\frac{d\mathbf{j}}{dt} = \mathbf{C}.\mathbf{j} - (\mathbf{j}'.\mathbf{C}.\mathbf{j})\mathbf{j}$$

where $\mathbf{C}$ is the temporal correlation matrix of the inputs $\mathbf{U}.\mathbf{U}' = \sum_i \mathbf{U}_i'.\mathbf{U}_i$ with the notations defined in section 2.3.1. Oja proved the last two systems were identical (in the asymptotic regime). The eigenvectors of $\mathbf{C}$ are obviously equilibrium points of the previous equation and it can be shown that the principal eigenvector is the only one to be stable.

The conclusion to this is dramatic: the Hebbian learning rule (with a multiplicative normalization, see 2.2) in a perceptron leads to the extraction of the first principal component of the inputs! And there is more: if one consider n' post-synaptic neurons in the high-level layer negatively connected, then the n' vector of connectivity from the low-layer to each neuron in the top layer will converge the n' principal component of the inputs, see [Rubner and Tavan 1989]. The learning of vertical connections has made it possible to represent the inputs in the best way possible in the upper layer. This suggests hierarchy corresponds to successive representations of the inputs at different scales.

3.1.2.3 BCM learning rule and input selections

The BCM learning rule [Bienenstock et al. 1982, Intrator and Cooper 1992, Blais and Cooper 2008] was originally developed in a perceptron framework and is generalized to a recurrent network in part 2.2. As opposed to our temporal approach, this rule was originally introduced as an ensemble learning rule:

$$\begin{aligned}\tfrac{d\mathbf{j}}{dt} &= \phi(v_{\text{post}}, \theta)\mathbf{v} - \kappa\mathbf{j}\\ \theta &= \mathbb{E}(v_{\text{post}}/c)^2\end{aligned}$$

where ϕ is shown in figure 2.1, $c \in \mathbb{R}_+$ and the expectation is taken over all the inputs. It is also assumed that the activity of the post synaptic neuron is a linear function of the inputs $v_{\text{post}} = \mathbf{j}'.\mathbf{v}$. It can be proved that if the inputs are linearly independent then the learning rule converges to a state so that the post-synaptic neuron is becoming selective to only one of the inputs. If there are multiple post-synaptic neurons negatively connected, the neurons of the top layer become selective to different inputs, see [Castellani et al. 1999].

As in the case of Oja learning, learning of feed-forward connections leads to good representation of the inputs in the upper layer.

3.1.2.4 Hebbian learning rule with linear decay

This is a (small) contribution of this thesis to the study of hierarchical neural network. Here, we consider that the network follows a differential equation of the type system Σ in equation (2.9) where the learning equation is multiplied by $\boldsymbol{\chi}$ (this is a Hadamard product). For simplicity, we also assume that $S = I_d$. The idea is to analyze the "shape" of the fixed points of this system using the slow inputs approximation Σ' derived in section 2.3.

Because the matrix $\boldsymbol{\chi}$ for a perceptron in equation (3.1) is not trivial, we can not be sure there is a stable fixed point. However, we propose to mathematically analyze the fixed points of the system. Then, we numerically show that the system may converge to such particular equilibrium points.

Recall the network is exposed to m inputs in $\mathbb{R}^n$, i.e. the inputs matrix $\mathbf{U}$ and the activity matrix $\mathbf{V}$ have $n+1$ rows and m columns (each column corresponds to a different input). Note that the last row of $\mathbf{U}$ is null. The fixed point of system Σ' are

$$\begin{cases} \mathbf{V}^* &= \left(\frac{1}{\kappa}\boldsymbol{\chi} \odot \mathbf{V}^*.\mathbf{V}^{*\prime}\right).\mathbf{V}^* + \mathbf{U} \\ \mathbf{J}^* &= \frac{1}{\kappa}\boldsymbol{\chi} \odot \mathbf{V}^*.\mathbf{V}^{*\prime} \end{cases}$$

Therefore, the fixed points of the activity of the pre-synaptic neurons verifies $\mathbf{V}^*_{ia} = \mathbf{U}_{ia}$ for i a neuron in $\{1..n\}$ and a a input in $\{1..m\}$. The equilibrium value of the post-synaptic activity is

$$\mathbf{V}_{n+1,a} = \frac{1}{\kappa}\sum_{b=1}^{m}\sum_{i=1}^{n+1}\boldsymbol{\chi}_{n+1,i}\mathbf{V}^*_{n+1,b}\mathbf{V}^*_{i,b}\mathbf{V}^*_{i,a}$$
$$= \frac{1}{\kappa}\sum_{b=1}^{m}\sum_{i=1}^{n}\mathbf{V}^*_{n+1,b}\mathbf{U}^*_{i,b}\mathbf{U}^*_{i,a} = \frac{1}{\kappa}\{\mathbf{V}^*.\mathbf{U}^{*\prime}.\mathbf{U}\}_{n+1,a}$$

For simplicity, we write $\mathbf{v}_{\text{post}} \in \mathbb{R}^m$ the vector of the equilibrium values of the post-synaptic neuron for all the inputs such that $\mathbf{v}_{\text{post}_a} = \mathbf{V}_{n+1,a}$. This is the last row of matrix $\mathbf{V}$. According to the previous equation its equilibrium value verifies

$$\mathbf{v}^*_{\text{post}} = \frac{1}{\kappa l_{\text{post}}}\mathbf{U}'.\mathbf{U}.\mathbf{v}^*_{\text{post}} \tag{3.2}$$

Here, we have added $l_{\text{post}} \in \mathbb{R}$ the time constant of the decay of the activity of the post-synaptic neuron which is assumed to be different from the time-constants of the pre-synaptic neurons, i.e. the decay term of the post-synaptic term in the activity equation of system (2.9) is multiplied by $l_{\text{post}} \in \mathbb{R}$.

Due to the very sparse shape of $\boldsymbol{\chi}$, the information in $\mathbf{J}^*$ can be summarized in $\mathbf{j} \in \mathbb{R}^n$ which corresponds to the connectivity of the n pre-synaptic neurons to the post-synaptic neuron (it is the last row of $\mathbf{J}$). With this notation, the equilibrium point of the connectivity is

$$\mathbf{j}^* = \frac{1}{\kappa}\mathbf{U}.\mathbf{v}^*_{\text{post}}$$

Multiply this equation by $\mathbf{U}.\mathbf{U}'$ on the leads to

$$\mathbf{U}.\mathbf{U}'.\mathbf{j}^* = \frac{1}{\kappa}\mathbf{U}.\underbrace{\mathbf{U}'.\mathbf{U}.\mathbf{v}^*_{\text{post}}}_{=\kappa l_{\text{post}}\mathbf{v}^*_{\text{post}}} = l_{\text{post}}\underbrace{\mathbf{U}.\mathbf{v}^*_{\text{post}}}_{=\kappa \mathbf{j}^*} = \kappa l_{\text{post}}\mathbf{j}^*$$

The conclusion is that the system converges to both the eigenvectors of the spatial and temporal correlation of the inputs: $\mathbf{v}^*_{\text{post}}$ is an eigenvector of $\mathbf{U}'.\mathbf{U}$ and $\mathbf{j}^*$ is an eigenvector of $\mathbf{U}.\mathbf{U}'$ (both of them associated with the eigenvalue κl_{post}). This approach gives an interesting interpretation of the connectivity versus the membrane potential: $\mathbf{j}$ extracts the temporal correlation whereas $\mathbf{v}_{\text{post}}$ extracts the spatial correlation of the inputs. It also says that a sequence of output neurons with different decay time constants, could extract all the eigen-modes of the inputs. Note that if κl_{post} is not an eigenvalue of $\mathbf{U}.\mathbf{U}'$, then the network must either converge to the null solution or diverge.

Simulations of the linear perceptron are illustrated in figure 3.3.

These results show that it is not necessary to have hetero-synaptic constraints as described in [Oja 1982, Miller and MacKay 1994] and in section 2.2.2 to converge to the eigenvectors of the time-correlation of the inputs. The eigenvector extraction is possible even for a local learning rule. The main difference with hetero-synaptic constraints lies in the fact that they systematically converge to the maximum eigenvector, whereas this system with a local learning rule needs κl_{post} to be well tuned to converge to this eigenvector.

3.1.3 A canonical hierarchical network

As suggested by figure 3.1, the hierarchical structure of the cortex is very complicated. We believe a computational approach would be able to address the

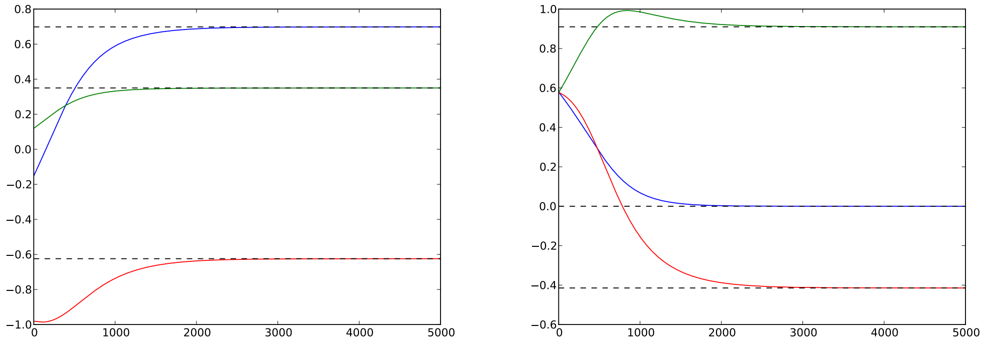


Figure 3.3: Evolution of the membrane potential of the post-synaptic neuron (left) and the connectivity of pre-synaptic neurons to the post-synaptic neuron (right) for a linear perceptron. Plain curves correspond to the (normalized) membrane potential (left) and strength of the connectivity from 3 input neurons to a single post-synaptic neurons (right) elicited by different inputs. Dotted curves correspond to each components of the eigenvector of $\mathbf{U}'.\mathbf{U}$ (left) and $\mathbf{U}.\mathbf{U}'$ (right), with $\mathbf{U}$ being a 3×3 random matrix. For this simulation, $\kappa = 1$, $\varepsilon = 0.01$, $T = 10$ and κl_{post} is set to be the maximal eigenvalue of $\mathbf{U}'.\mathbf{U}$.

real hierarchies, but a mathematical approach needs a simplified architecture. Therefore, we define a canonical hierarchical network to be analyzed mathematically (though it is beyond the scope of this thesis which focuses on fully recurrent networks). We believe the structure of figure 3.4 is a good candidate for the canonical hierarchical because it gathers the main features of the biological hierarchical structures described before while being quite simple. It is made of a cascade of layers which correspond to the different areas in the cortex. There are three types of connections: lateral or recurrent connections that link neurons of the same layer, feed-forward connections which link neurons from low-level layers to high-level layers and feedback connections which go in the opposite side propagating top-down information.

This structure is very simplified because an underlying abstraction axis was assumed (from low-level to high-level). In other words, it has a ”one-dimensional“ structure which corresponds to a single pyramid of abstraction. For instance, it does not take into account double streams structure like in

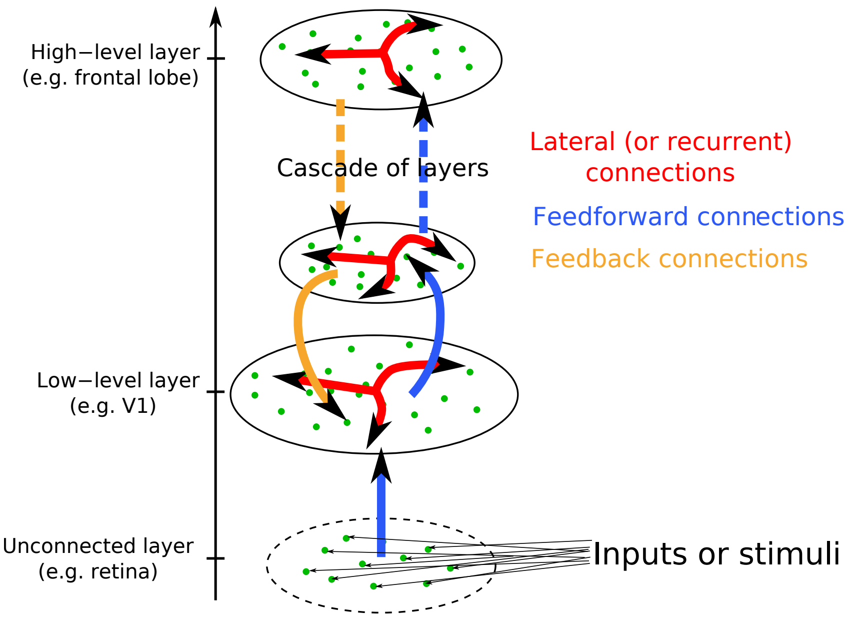


Figure 3.4: Idealization of a hierarchical network.

the right picture of figure 3.1.

Considering that he lowest layer (e.g. the retina) is unconnected and does not receive feedbacks is not a restrictive assumption. Indeed, choosing appropriately the feed-forward connection from the lowest layer to the first above makes it possible to get rid of the retina which may become a transparent filter. The two reasons why we defined this canonical structure with a retina are (i) there are biological evidence that it is true in the brain (ii) the results for the perceptron only work for this stage of the hierarchy.

A priori, the simplest way to keep this hierarchical structure invariant through learning is to consider a roughly ”one-dimensional“ physical structure, i.e. the binary matrix $\boldsymbol{\chi}$ which pre-multiply the connectivity evolution. However, we have shown in section 2.4.5, that the addition to spatially correlated noise can induce and define a hierarchical structure. The noise matrix which specifies the spatial correlation of the noise plays the role of the physical connection matrix $\boldsymbol{\chi}$.

To our knowledge, there is no satisfying mathematical approach of learning in hierarchical network in the sense defined above. Learning the first feed-forward connections (e.g. from retina to V1) has been extensively studied since it corresponds to a simple generalization of the perceptron reviewed before. There has been many computational studies on network without feedback and where the lateral connectivities were fixed [Linsker 1988, Földiák 1991, Stringer and Rolls 2002, Serre 2005]. Several studies have addressed the simultaneous learning of feed-forward and lateral connectivity [Rubner and Tavan 1989, Bartsch and Van Hemmen 2001, Miikkulainen et al. 2005]. Yet, there does not seem to be a generalizable and compelling mathematical principle emerging for larger networks, e.g. with feedbacks. It seems we are not yet able to understand deeply the consequence of having learning recurrent or feedback connections. In this thesis, we address the case of fully recurrent networks, which can be seen as the building blocks of the canonical hierarchical network. We hope this will eventually help studying these hierarchical structure from a mathematical perspective.

3.2 Emergence of cortical maps

Over the last fifty years, the field of neuroscience has developed significantly thanks to new imaging methods. They have made it possible to observe and understand patterns in the brain. One of the most important results has been the discovery of cortical maps.

3.2.1 What is a cortical map?

They are spatially structured patterns observed in the cortex. They are closely linked with the functional connectivity of the cortex which specifies which neuron codes for what. There is probably an underlying structure to any cortical area, however we are not able to understand all of them because we do not know precisely what is encoded in every part of the cortex. Therefore, the best examples of cortical maps can be found among the low-level sensory area. In the following, we detail three of them. Before proceeding, note that the existence of such structured patterns depends not only on the task the area is devoted to but also on the species. For instance, if it is possible to have beautiful results for the orientation columns for tree shrew, they do not arise in rats for instance.

- **Rat barrel cortex** an analogous to **retinotopy**:
  In rat's sensory cortex, every whisker is associated with a sharply bounded cortical site in layer IV (see figure 3.5). However such a barrel structure is less evident in other layers so that it is harder to distinguish columns and label them with a given whisker.

  This kind of cortical map is qualitatively very similar to the retinotopy observed in the visual cortex of many animals. Indeed, it has been observed that neurons in V1 (the lowest-level area of the visual cortex) tend to be closer in the cortex when their receptive fields are close in the visual field. Because, the visual field is roughly two dimensional then the cortex also has a two dimensional structure. This is similar to the rat's barrel cortex since the whiskers are aligned on a two dimensional grid and we observe that the associated area in the cortex also has this geometry. These are good examples of the cortex learning the underlying geometrical structure of the inputs (which is the same in both cases: a plane).

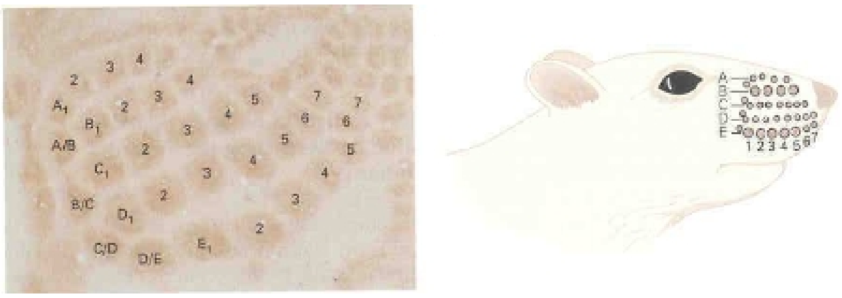


Figure 3.5: Rat's barrel cortex (actually this is layer IV). Every whisker of the rat corresponds to a well defined area of the cortex mostly responsible for processing information from it. These processing units have the same distribution as whiskers on the muzzle. Taken from [Kandel et al. 1991].

- **Ocular dominance columns**:

  In some species, it has also been observed that neurons of low-level visual layers tend to have a specific eye preference: they are almost not excited when their preferred eye is closed. It is then possible to project this information on the cortex by coloring both eyes in different colors. To be more precise, this applies only in layer IV of V1 by injecting radioactive tracers in one eye and finding in the laminar plane of layer IV the grouping in radioactive bands of the same occularity, given that enough time has been given to the tracer to diffuse with axonal transport. Figure 3.6 represent this organization.

  In this case, the information to be stored has a binary value and therefore is spread along a one-dimensional axis. The reason why figure 3.6 does not have only two regions for each eye is that this binary information is considered on top of others, e.g. position in the visual field. Therefore, there must be a trade-off between retinotopy and ocular dominance which ends up in the layered structure above.

- **Orientation columns**:

  The discovery of Hubel and Wiesel that some neurons in the primary visual cortex were tuned to select an edge of a particular orientation at a particular position in the visual field also has a nice interpretation in terms of cortical map. Using an imaging method called *optical imaging*, it is possible to observe the orientation preference of the neurons at the scale of the area. For some species, neurons coding for the same orien-

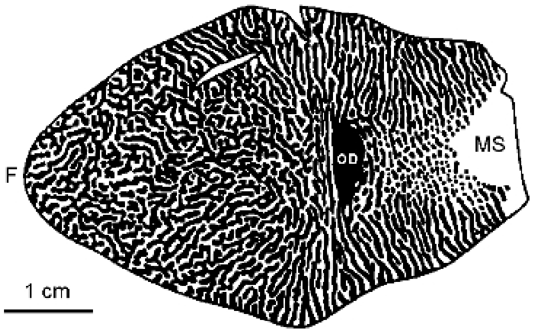


Figure 3.6: Ocular dominance columns in macaque monkey. It shows the pattern over nearly the complete visual hemi-field in a macaque monkey. The outer boundaries of the pattern correspond to the vertical mid-line of the visual field; F indicates the fovea; OD the optic disc, and MS the monocular segment. The pattern is a drawing made from a montage of sections stained for cytochrome oxidase in a monkey which had lost one eye over a year prior to sacrifice. Taken from [Florence and Kaas 1992].

tation tend to gather together to form orientation columns. As shown in figure 3.7, it possible to draw a map of these orientation columns. Superposing the distribution of axons from one column to its neighbors leads to an interesting observation. Neurons tend to have co-aligned connections: (i) neurons coding for the same orientation tend to be preferentially connected, (ii) the direction of the axons roughly match the preferred direction of the neurons. As shown later, we believe this is a consequence of a Hebbian learning rule with inputs made of a lot of straight lines (covering more than a single receptive field).

These orientation columns are superposed to the retinotopic organization and the ocular dominance columns, see [Swindale 2000].

- Face columns in the Infero-Temporal cortex:
  It has even been observed [Tanaka 1996, Wang et al. 1998] that some cells were specifically excited by faces in the infero-temporal cortex. Besides, neighboring columns code for close-by views of the same face, therefore there is a notion of cortical space as in the previous cases.

Although very irregular, the two-dimensional cortical maps observed at a given stage of development, can be unfolded in higher dimensions to get smoother geometrical structures. Indeed, [Bressloff et al. 2001] suggested that

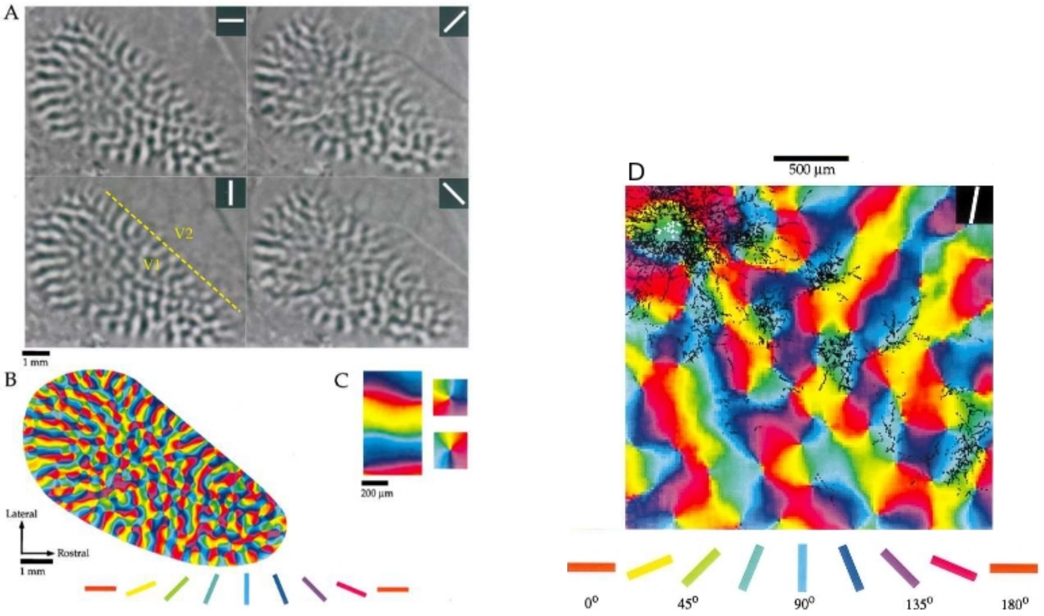


Figure 3.7: Optical imaging in tree shrew visual cortex. A, Difference images obtained for four stimulus angles. Dark colors indicates areas that were active during presentation of the stimulus. B, Orientation preference map. Orientation preference of each location is color-coded according to the indications below the map. C, Portions of the orientation preference map shown in B have been enlarged to demonstrate that the orientation preference maps contained both linear zones (left), and pinwheel arrangements (right) that are functional discontinuities. D, reconstruction of axon terminal distribution (black dots) from a marker (biocytin) injected in neurons (white crosses) in (layer II/III of) the tree shrew visual cortex. Although the distribution is locally isotropic, an orientation preference emerges at longer scales. The neurons connect preferentially to others of the same orientation preference. Besides, the direction of the axon seems to correspond to the orientation preference. This is called co-linearity. Taken from [Bosking et al. 1997].

the network of orientation pinwheels in V1 is a direct product between a circle for orientation preference and a plane for position, based on a modification of the ice-cube model of Hubel and Wiesel [Hubel and Wiesel 1977]. From a more abstract geometrical perspective, [Petitot 2003] has associated such a structure to a 1-jet space and used this to develop some applications to computer vision. More recently, more complex geometrical structures such as spheres and hyperbolic surfaces that incorporate additional stimulus features such as spatial frequency and textures, were considered respectively in [Bressloff and Cowan 2003] and [Chossat and Faugeras 2009]. To our knowledge, there is no research yet about the emergence of such high dimensional structures through biologically inspired learning. We intend to address this problem in the following.

3.2.2 Through learning of the feed-forward connectivity

Unsupervised learning of feed-forward connections forms the basis of most weight-based models of cortical development, assuming fixed lateral connectivity (e.g. Mexican hat) and modifiable vertical connections (see the reviews of [Erwin et al. 1995, Swindale 1996, Dayan and Abbott 2001]). In these developmental models, the statistical structure of input correlations provides a mechanism for spontaneously breaking some underlying symmetry of the neuronal receptive fields leading to the emergence of feature selectivity [Miller and MacKay 1994]. When such correlations are combined with fixed intra-cortical interactions, there is a simultaneous breaking of translation symmetry across cortex leading to the formation of a spatially periodic cortical feature map. A related mathematical formulation of cortical map formation has been developed in [Takeuchi and Amari 1979, Bressloff 2005] using the theory of self–organizing neural fields.

3.2.3 Through learning of the recurrent connectivity

There is less literature about the emergence of cortical maps through the learning of lateral connections. Swindale has shown how to use Kohonen maps (see [Kohonen 1990]) to compute cortical maps in [Swindale 1996, Swindale 2000]. He assumes the inputs he plugs to the Kohonen map correspond to the output of the feed-forward processing and let the neurons' position evolve. They converge to biologically plausible maps (when appropriately initialized). As shown later, the method we present in the following is quite close to his, the main difference being that we use correlation based learning rules when he uses Kohonen maps. Another approach similar in the philosophy is the dimension reduction approach called elastic net approach [Goodhill and Willshaw 1990, Durbin and Mitchison 1990].

There have only been a few computational studies that consider the joint development of lateral and vertical connections: [Bartsch and Van Hemmen 2001, Miikkulainen et al. 2005] and we are not aware of any mathematical treatment of this question.

This thesis focuses on learning of the recurrent (or horizontal or lateral) connectivity and, in the following, we propose a generic way to extract the underlying geometry of a connectivity matrix.

3.3 Symmetric part of the recurrent connectivity: a distance matrix

We have seen in the previous chapter that the symmetric part of the connectivity of a recurrent network roughly converges to the temporal correlation matrix of the inputs: $\mathrm{sym}(\mathbf{J}^*) \sim \mathbf{u}.\mathbf{u}'$. In the case of a voltage-based network with slow inputs, it is precisely given by theorem 2.3.4. For a linear, weakly connected network with Hebbian learning is is given by theorem 2.4.9. In all the cases described in chapter 2, the symmetric part of $\mathbf{J}^*$ can be thought of as a spatial correlation matrix of a positive time-series related in some way to the inputs.

We believe this matrix is related to the underlying geometry of the inputs or the manifold on which is defined the dynamical system generating the inputs. This part is devoted to showing how to extract it from the connectivity. In this section, we focus on the symmetric part of the connectivity therefore on the correlation-based mechanism of the learning rule.

3.3.1 From a symmetric connectivity matrix to a convolutional network

So far neurons have been identified by a label $i \in \{1..n\}$; there is no notion of geometry or space in the preceding results. However, as we show below, the inputs may contain a spatial structure that can be encoded by the final connectivity. In this section, we show that the network behaves as a convolutional network on this geometrical structure. The idea is to interpret the final connectivity as a matrix describing the distance between the neurons living in a k-dimensional space. This is quite natural since $\mathbf{J}^*$ is symmetric and has positive coefficients, properties shared with a Euclidean distance matrix. More specifically, we want to find an integer $k \in \mathbb{N}$ and n points in $\mathbb{R}^k$, denoted by $\mathbf{x}_i, i \in \{1, \dots, n\}$, so that the connectivity can roughly be written as $\mathbf{J}^*_{ij} \simeq g(\|\mathbf{x}_i - \mathbf{x}_j\|^2)$, where g is a positive decreasing real function. If we manage to do so then the interaction term in system Σ becomes

$$\{\mathbf{J} \cdot S(\mathbf{v})\}_i \simeq \sum_{j=1}^{n} g(\|\mathbf{x}_i - \mathbf{x}_j\|^2) S\big(\mathbf{v}(\mathbf{x}_j)\big) \tag{3.3}$$

where we redefine the variable $\mathbf{v}$ as a field such that $\mathbf{v}(\mathbf{x}_j) = \mathbf{v}_j$. This equation says that the network is convolutional with respect to the variables $\mathbf{x}_i$, $i = 1, .., n$ and the associated convolution kernel is $g(\|\mathbf{x}\|^2)$.

In practice, it is not always possible to find a geometry for which the connectivity is a distance matrix. Therefore, we project the appropriate matrix on the set of Euclidean distance matrices. This is the set of matrices $\mathbf{W}$ such that $\mathbf{W}_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2$ with $\mathbf{x}_i \in \mathbb{R}^k$. More precisely, we define $\mathbf{D} = g^{-1}(\mathbf{J}^*)$, where g^{-1} is applied to the coefficients of $\mathbf{J}^*$. We then search for the distance matrix $\mathbf{D}_\perp$ such that $\|\mathbf{D}_\parallel\|^2 = \|\mathbf{D} - \mathbf{D}_\perp\|^2$ is minimal. The minimization turns out to be a least square minimization whose parameters are the $\mathbf{x}_i \in \mathbb{R}^k$. This can be implemented by a set of methods known as multidimensional scaling, which are reviewed in [Borg and Groenen 2005]. In particular, we use the stress majorization or SMACOF algorithm for the stress1 cost function throughout the thesis. This leads to writing $\mathbf{D} = \mathbf{D}_\perp + \mathbf{D}_\parallel$ and therefore $\mathbf{J}^*_{ij} = g(\mathbf{D}_{\perp ij} + \mathbf{D}_{\parallel ij})$ where $\mathbf{D}_\perp$ is a distance matrix, i.e. there exists n vectors $\mathbf{x}_i \in \mathbb{R}^k$ such that $\mathbf{D}_{\perp ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2$ with $\mathbf{x}_i \in \mathbb{R}^k$

We now consider two particular choices of the function g:

1. If $g(x) = a\big(1 - \frac{x}{\lambda^2}\big)$ with $a > 1$ and $\lambda \in \mathbb{R}_+$, then one can always write

$$\mathbf{J}^*_{ij} = \mathbf{J}^*(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{M}(\mathbf{x}_i, \mathbf{x}_j) + \ g(\|\mathbf{x}_i - \mathbf{x}_j\|^2) \tag{3.4}$$

   such that $\mathbf{M}(\mathbf{x}_i, \mathbf{x}_j) = -\frac{a}{\lambda^2}\mathbf{D}_{\parallel ij}$.

2. If $g(x) = ae^{-\frac{x}{\lambda^2}}$ with $a, \lambda \in \mathbb{R}_+$, then one can always write

$$\mathbf{J}^*_{ij} = \mathbf{J}^*(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{M}(\mathbf{x}_i, \mathbf{x}_j)\ g(\|\mathbf{x}_i - \mathbf{x}_j\|^2) \tag{3.5}$$

   such that $\mathbf{M}(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\mathbf{D}_{\parallel ij}}{\lambda^2}}$,

where $\mathbf{J}^*$ is also redefined as a function over the $\mathbf{x}_i$, i.e. $\mathbf{J}^*(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{J}^*_{ij}$. For obvious reasons, $\mathbf{M}$ is called the non-convolutional connectivity. It is the role of multidimensional scaling methods to minimize the role of the undetermined function $\mathbf{M}$ in the previous equations, i.e. ideally having $\mathbf{M} \equiv 0$ (resp. $\mathbf{M} \equiv 1$) for the first (resp. second) assumption above. The ideal case of a fully convolutional connectivity can alway be obtained if k is large enough. Indeed, proposition 3.3.1 shows that $\mathbf{D} = g^{-1}(\mathbf{J}^*)$ satisfies the triangular inequality for matrices (i.e. $\mathbf{D}_{ij} \leq (\sqrt{\mathbf{D}_{ik}} + \sqrt{\mathbf{D}_{kj}})^2$) for both g under some plausible assumptions. Therefore, it has all the properties of a distance matrix (symmetric, positive coefficients and triangular inequality) and one can find points in $\mathbb{R}^k$ such that it is the distance matrix of these points if $k \leq n-1$ is large enough. In this case, the connectivity on the space defined by these points is fully convolutional, i.e. equation (3.3) is exactly verified.

For the following proposition, we assume that we are the the situation of an equilibrium connectivity $\mathbf{J}^*$ of a voltage-based network with Hebbian learning, i.e. $\mathbf{J}^*$ is the solution of the fixed point equations in theorem 2.3.4. The important detail about this setup is that the sigmoid S is always positive (as opposed to the linear case). This hypothesis makes it possible to prove

Proposition 3.3.1.*If the neurons are equally excited on average (i.e.* $\|S(\mathbf{V}_{i.})\| = c \in \mathbb{R}_+$*), and*

1. *if* $g(x) = a\big(1 - \frac{x}{\lambda^2}\big)$ *with* $a, \lambda \in \mathbb{R}_+$*, then* $\mathbf{D} = g^{-1}(\mathbf{J}^*)$ *satisfies the triangular inequality.*

2. *if* $g(x) = ae^{-\frac{x}{\lambda^2}}$ *with* $a, \lambda \in \mathbb{R}_+$*, then* $\mathbf{D} = g^{-1}(\mathbf{J}^*)$ *satisfies the triangular inequality if the following assumption is satisfied*

$$arcsin\big(S(0)\big) - arcsin\Big(\sqrt{a^3 - \sqrt{a^6 - a^3}}\Big) \geq \frac{\pi}{8} \tag{3.6}$$

Proof. We recall the notation $\mathbf{V} \in \mathbb{R}^{n\times m}$ such that $\mathbf{V}_{ia} = \mathbf{v}_i^{(a)}$ the value of the equilibrium activity of the ith neuron for the ath input. We also note $\mathbf{V}_{i.} \in \mathbb{R}^m$ the vector of the activity of neuron i for all the inputs. The uniform excitement of neurons on average reads, for all $i = 1, .., n$, $\|S(\mathbf{V}_{i.})\| = c \in \mathbb{R}_+$. For simplicity and without loss of generality we can assume $c = 1$ and $a \geq 1$ (we can play with a and λ to generalize this to any $c, a \in \mathbb{R}_+$ as long as $a > c$). The triangular inequality we want to prove can therefore be written

$$\sqrt{g^{-1}\big(S(\mathbf{V}_{i.}).S(\mathbf{V}_{j.})'\big)} \leq \sqrt{g^{-1}\big(S(\mathbf{V}_{i.}).S(\mathbf{V}_{k.})'\big)} + \sqrt{g^{-1}\big(S(\mathbf{V}_{j.}).S(\mathbf{V}_{k.})'\big)} \tag{3.7}$$

for readability we rewrite $\mathbf{x} = S(\mathbf{V}_{i.})$, $\mathbf{y} = S(\mathbf{V}_{j.})$ and $\mathbf{z} = S(\mathbf{V}_{k.})$. These three vectors are on the unity sphere and have only positive coefficients. Actually, there is a distance between these vectors which consists in computing the geodesic angle between them. In other words, consider the intersection of $vect(\mathbf{x}, \mathbf{y})$ and the m-dimensional sphere. This is a circle where $\mathbf{x}$ and $\mathbf{y}$ are located. The angle between the two points is written $\theta_{\mathbf{x},\mathbf{y}}$. It is a distance on the sphere, thus, it satisfies the triangular inequality:

$$\theta_{\mathbf{x},\mathbf{y}} \leq \theta_{\mathbf{x},\mathbf{z}} + \theta_{\mathbf{z},\mathbf{y}} \tag{3.8}$$

Actually, all these angles belong $[0, \frac{\pi}{2}[$ because $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ only have positive coefficients.

Observe that $g^{-1}(\mathbf{x}.\mathbf{y}') = g^{-1}\big(cos(\theta_{\mathbf{x},\mathbf{y}})\big)$ and separate now the two case for the choice of function g:

1. If $g(x) = a\big(1 - \frac{x}{\lambda^2}\big)$ with $a \geq 1$, then $g^{-1}(x) = \lambda^2\big(1 - \frac{x}{a}\big)$. Therefore, define $h_1 : x \mapsto \lambda\sqrt{1 - \frac{cos(x)}{a}}$. We now want to apply h_1 to (3.8) but h_1 is monotonic only if $x \leq \frac{\pi}{2}$. Therefore, divide (3.8) by 2 and apply h_1 on both sides to get
$$h_1(\frac{\theta_{\mathbf{x},\mathbf{y}}}{2}) \leq h_1(\frac{1}{2}\theta_{\mathbf{x},\mathbf{z}} + \frac{1}{2}\theta_{\mathbf{z},\mathbf{y}})$$
Now consider the function $\eta_a(x, y) = h_1(x + y) - h_1(x) - h_1(y)$ for $x, y \in [0, \frac{\pi}{4}[$. Because, h_1 is increasing it is clear that $\frac{\partial \eta_a}{\partial x}(x, y) \leq 0$ (and similarly for y), such that η_a reaches its maximum for $x = y = \frac{\pi}{4}$. Besides, $\eta_a(\frac{\pi}{4}, \frac{\pi}{4}) \leq \eta_1(\frac{\pi}{4}, \frac{\pi}{4}) < 0$. This proves that $2h_1(\frac{1}{2}\theta_{\mathbf{x},\mathbf{z}} + \frac{1}{2}\theta_{\mathbf{z},\mathbf{y}}) \leq h_1(\theta_{\mathbf{x},\mathbf{z}}) + h_1(\theta_{\mathbf{z},\mathbf{y}})$. Moreover, it is easy to observe that $2h_1(\frac{x}{2}) \geq h_1(x)$ for all $a > 1$. This concludes the proof for $g(x) = a\big(1 - \frac{x}{\lambda^2}\big)$.

2. If $g(x) = ae^{-x}$ then $g^{-1}(x) = \lambda^2 ln(\frac{a}{x})$. As before, define $h_2 : x \mapsto \lambda\sqrt{ln(\frac{a}{cos(x)})}$. We still want to apply h_2 to (3.8), but h_2 is not defined for $x > \frac{\pi}{2}$, which is likely for the right hand side of (3.8). Therefore, we apply h_2 to (3.8) divided by two and use the fact that h_2 is increasing on $[0, \frac{\pi}{2}[$. This leads to $h_2(\frac{\theta_{\mathbf{x},\mathbf{y}}}{2}) \leq h_2(\frac{1}{2}\theta_{\mathbf{x},\mathbf{z}} + \frac{1}{2}\theta_{\mathbf{z},\mathbf{y}})$. First, we use the convexity of h_2 to get $2h_2(\frac{\theta_{\mathbf{x},\mathbf{y}}}{2}) \leq h_2(\theta_{\mathbf{x},\mathbf{z}}) + h_2(\theta_{\mathbf{z},\mathbf{y}})$ and then we use the fact that $2h_2(\frac{x}{2}) \geq h_2(x)$ for $x \in [0, \delta[$. With $\delta < \frac{\pi}{2}$. This would conclude the proof but we have to make sure the angles remain in $[0, \delta[$. Actually, we can compute δ which verifies $2h_2(\frac{\delta}{2}) = h_2(\delta)$. This leads to, $\delta = 2arccos(\sqrt{a^3 - \sqrt{a^6 - a^3}})$.
In fact, the coefficients of $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ are strictly positive and larger than $S(0)$. Therefore, the angles between them are strictly smaller than $\frac{\pi}{2}$. More precisely, $\theta_{\mathbf{x},\mathbf{y}} \in [0, \frac{\pi}{2} - 2arcsin\big(S(0)\big)[$. Therefore, a necessary condition for the result to be true is $2arccos(\sqrt{a^3 - \sqrt{a^6 - a^3}}) \geq \frac{\pi}{2} - 2arcsin\big(S(0)\big)$. Using the fact that $arccoS(x) = \frac{\pi}{2} - arcsin(x)$ leads to $arcsin\big(S(0)\big) - arcsin(\sqrt{a^3 - \sqrt{a^6 - a^3}}) \geq \frac{\pi}{8}$.

□

Remark 4.It is necessary that $S > 0$ for the previous to hold. In a way, the sigmoid has the role of making sure the activity remains positive so that the equilibrium connectivity can be a distance matrix. Therefore, we can not be sure that the equilibrium connectivity of a linear system as considered in part 2.4. Yet, we believe that the result may still hold in more general cases.

3.3.2 Unveiling the geometrical structure of the inputs

We hypothesize that the shape defined by the set of the $\mathbf{x}_i$ reflects the underlying geometrical structure of the inputs. We have not found a way to prove this so we only provide numerical examples that illustrate this claim. Therefore, the following is only a (numerical) proof of concept. For each example, we feed the network with inputs having a defined geometrical structure and then show how this structure can be extracted from the lateral connectivity by the method outlines in section 3.1.

In particular, we assume that the inputs are uniformly distributed Gaussians over a submanifold $\Omega \subset \mathbb{R}^n$ with fixed geometry which we expect the network to retrieve after learning. This corresponds to associating to the ath input a location $\mathbf{z}_a \in \Omega$ and assuming that it has a bell shape around this center.

We introduce now a strong assumption, which corresponds to saying that the feed-forward connectivity (which we do not consider here) has already properly filtered the information coming from the sensory organs. Indeed, as said in part 3.1, the feedforward connections tend to isolate the meaningful features of the input by designing relevant receptive fields for the cortical neurons. A inspiring example is that of neurons in V1 which are selective for positions in the visual field: the feedforward connections from the retina to V1 give a position $\mathbf{y}_i$ to the ith neuron in the visual field. There has been numerous papers describing the evolution of such feedfoward connections leading to retinotopic receptive field in the low-level areas of the cortex [Miller and MacKay 1994, Dayan and Abbott 2001, Gerstner and Kistler 2002b, Miikkulainen et al. 2005]. In our more general framework, we can generalize it by saying that each neuron is assigned a position on the manifold Ω. All the neurons are therefore assumed to be localized on the manifold Ω but this piece of information is only in the inputs to the network.

With these notations we define the set of inputs by the matrix $\mathbf{U} \in \mathbb{R}^{n\times m}$ such that $\mathbf{U}_{ia} = \mathbf{u}_i^{(a)} = f(\|\mathbf{y}_i - \mathbf{z}_a\|_\Omega)$ where f is a decreasing function on $\mathbb{R}_+$. The norm $\|.\|_\Omega$ is the natural norm defined over the manifold Ω. For simplicity, assume $f(x) = f_\sigma(x) = Ae^{-\frac{x^2}{\sigma^2}}$ so that the inputs are localized bell-shaped bumps on the shape Ω.

These stimuli are different from natural pictures of the daily life even if they may capture some essential regularity in them. However, they appear to be good approximations of retinal (resp. cortical) waves which are pre-natal spontaneously emerging waves of activity propagating in the retina (resp. the cortex) [Wong 1999]. In a way, the waves in the retina train the cortex to understand their 2 dimensional geometry even before the eyes open. It was suggested that these local waves are a good motivation for considering localized inputs, [Miikkulainen et al. 2005]. Although we do not capture the dynamics of these waves in the part (see 3.3) the spatial structures of these waves are close to be localized Gaussians in the visual field.

3.3.2.1 Planar retinotopy

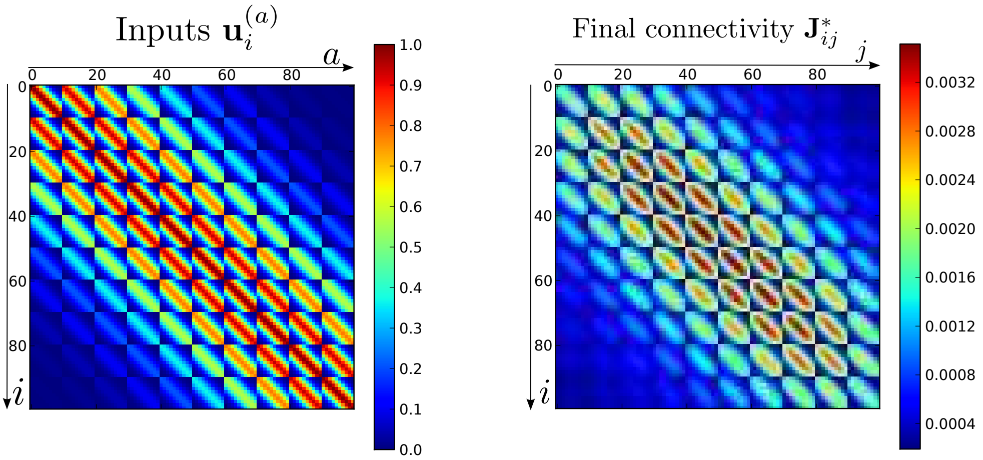


Figure 3.8: Plot of planar retinotopic inputs on $\Omega = [0,1] \times [0,1]$ (left) and final connectivity matrix of the system Σ' (right). The parameters used for this simulation are $S(x) = \frac{1}{1+e^{-4(x-1)}}$, $l = 1$, $\kappa = 10$, $n = m = 100$, $\sigma = 4$. As can be seen on the left picture, these inputs correspond to $u_m = 1$.

First, we consider a set of spatial Gaussian inputs uniformly distributed over a two-dimensional plane, e.g. $\Omega = [0,1] \times [0,1]$. For simplicity, we take $n = m = K^2$ and set $\mathbf{z}_a = \mathbf{y}_i$ for $i = a$, $a \in \{1, \dots, m\}$. (The numerical results show an identical structure for the final connectivity when the $\mathbf{y}_j$ correspond to random points, but the analysis is harder). In the simpler case of one-dimensional Gaussians with $n = m = K$, the input matrix takes the form $\mathbf{U} = T_{f_\sigma}$, where T_f is a symmetric Toeplitz matrix:

$$T_f = \begin{pmatrix} f(0) & f(1) & f(2) & \cdots & \cdots & f(K) \\ f(1) & f(0) & f(1) & f(2) & \cdots & f(K-1) \\ f(2) & f(1) & f(0) & f(1) & \cdots & f(K-2) \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ f(K) & f(K-1) & f(K-2) & \cdots & \cdots & f(0) \end{pmatrix} \tag{3.9}$$

In the two-dimensional case, we set $\mathbf{z} = (x, y) \in \Omega$ and introduce the labeling $\mathbf{y}_{k+(l-1)K} = (x_k, y_l)$ for $k, l = 1, \dots, K$. It follows that $\mathbf{U}_{ia} = \mathbf{u}_i^{(a)} \sim \exp(-\frac{(x_k - x_{k'})^2}{\sigma^2}) \exp(-\frac{(y_l - y_{l'})^2}{\sigma^2})$ for $i = k + (l-1)K$ and $a = k' + (l'-1)K$. Hence, we can write $\mathbf{U} = T_{f_\sigma} \otimes T_{f_\sigma}$, where $\otimes$ is the Kronecker product; the Kronecker product is responsible for the $K \times K$ sub-structure we can observe in figure 3.8 with $K = 10$. Note that if we were interested in a n-dimensional retinotopy, then the input matrix could be written as a Kronecker product between n Toeplitz matrices. As previously mentioned, the final connectivity matrix roughly corresponds to the correlation matrix of the input matrix. It turns out that the correlation matrix of $\mathbf{U}$ is also a Kronecker product of two Toeplitz matrix generated by a single Gaussian (with a different standard deviation). Thus, the connectivity matrix has the same basic form as the input matrix when $\mathbf{z}_a = \mathbf{y}_i$ for $i = a$. The inputs and stable equilibrium points of the simulated system are shown in figure 3.8. The positions $\mathbf{x}_i$ of the neurons after multidimensional scaling are shown in figure 3.9 for different parameters. Note that we find no significant change in the position $\mathbf{x}_i$ of the neurons when the convolutional kernel g varies (as will be also shown in section 3.3.3.1). Thus, we only show results for one of these kernels, namely, $g(x) = e^{-x}$.

If the standard deviation of the inputs is properly chosen as in figure 3.9.b, we observe that the neurons are distributed on a regular grid which is retinotopically organized. In other words, the network has learned the geometric shape of the inputs. This can also be observed in figure 3.9.d which corresponds to the same connectivity matrix as in figure 3.9.b but represented

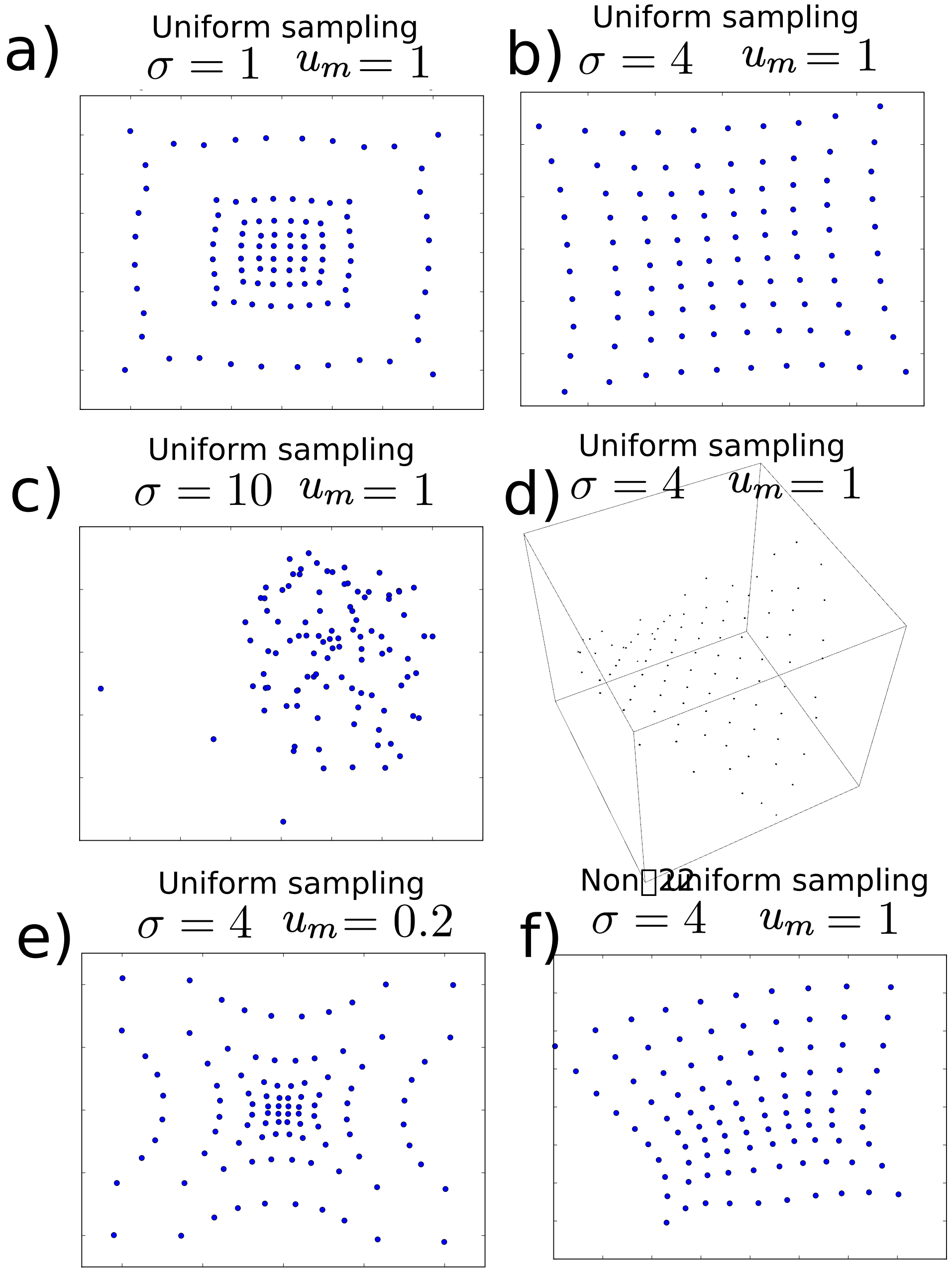


Figure 3.9: Positions $\mathbf{x}_i$ of the neurons after having applied multidimensional scaling to the equilibrium connectivity matrix of a learning network of $n = 100$ neurons driven by planar retinotopic inputs as described in figure 3.8. In all figures, the convolution kernel $g(x) = e^{-x}$; this choice has virtually no impact on the shape of the figures. (a) Uniformly sampled inputs with $m = 100$, $\sigma = 1$, $u_m = 1$ and $k = 2$. (b) Uniformly sampled inputs with $m = 100$, $\sigma = 4$, $u_m = 1$ and $k = 2$. (c) Uniformly sampled inputs with $m = 100$, $\sigma = 10$, $u_m = 1$ and $k = 2$. (e) Uniformly sampled inputs with $m = 100$, $\sigma = 4$, $u_m = 0.2$ and $k = 2$. (d) Same as b) but in three dimensions, i.e. $k = 3$. (f) Non-uniformly sampled inputs with $m = 150$, $\sigma = 4$, $u_m = 1$ and $k = 2$. The first 100 inputs are as in (b) but 50 more inputs of the same type are presented to one half of the visual field. This corresponds to the denser part of the picture.

in three dimensions. The neurons self-organize on a 2-dimensional saddle shape that accounts for the border distortions that can be observed in two dimensions (which we discuss in the next paragraph). If σ is too large, as can be observed in figure 3.9.c, the final results is poor. Indeed, the inputs are not local anymore and cover most of the visual field. Therefore, the neurons saturate, i.e. $S(\mathbf{V}_{ia}) \simeq S_m$, for all the inputs and no structure can be read in the activity variable. On the other hand, if σ is small then the neurons still seem to self organize (as long as the inputs are not completely localized on single neurons) but with significant border effects.

There are several reasons why we observe border distortions in figure 3.9. We believe the most important is due to an unequal average excitation of the neurons. Indeed, the neurons corresponding to the border of the "visual field" are less excited than the others. For example, consider a neuron on the left border of our artificial visual field. It has no neighbors on its left and therefore is less likely to be excited by its neighbors and therefore less excited on average. The consequence is that it is less connected to the rest of the network (see for instance the top row of the right picture of figure 3.8), because their connection depends on their level of excitement through the correlation of the activity. Therefore, it is further away from the other neurons, which is what we observe. When the inputs are really localized the border neurons are even less excited on average and thus are further away as shown in figure 3.9.a. Note that that neurons near a map border have neighbors not only in the considered visual area but also in the symetrical map extending in the next (or previous) area. For instance, the map of the visual field flip symmetrically arround the representation of the vertical meridian of the hemifield (e.g. border V1-V2) and the periphery (e.g. border V2-V3). However, we did not model the hierachy in our articificial cortex and the space in which we represent the neurons is abstract and does not necessarily correspond to the physical space. Therefore, these biological observations do not concern us in the following.

Another way to get distortions in the positions $\mathbf{x}_i$ is to reduce or increase excessively the amplitude $u_m = \max_{i,a} |\mathbf{u}_i^{(a)}|$ of the inputs. Indeed, if it is small, the equilibrium activity described by equation (2.22) is also small and likely to be the flat part of the sigmoid. In this case, neurons tend to be more homogeneously excited and less sensitive to the particular shape of the inputs. Therefore, the network loses some information about the underlying structures of the inputs. Actually, the neurons become relatively more sensitive to the neighborhood structure of the network and the border neurons have a different behavior as the rest of the network as shown in figure 3.9.e. The parameter κ has much less impact on the final shape since it only corresponds to a homogeneous scaling of the final connectivity.

So far, we have assumed that the inputs were uniformly spread on the manifold Ω. If this assumption is broken the final position of the neurons will be affected. As shown in figure 3.9.f, where 50 inputs were added to the case of figure 3.9.b in only half of the visual field, the neurons that code for this area tend to be closer. Indeed, they tend not to be equally excited on average (as supposed in proposition 3.3.1) and a distortion effect occurs. This means that a proper understanding of the role of the vertical connectivity would be needed to complete this geometrical picture of the functioning of the network. This is, however, beyond the scope of this thesis.

3.3.2.2 Toroïdal retinotopy

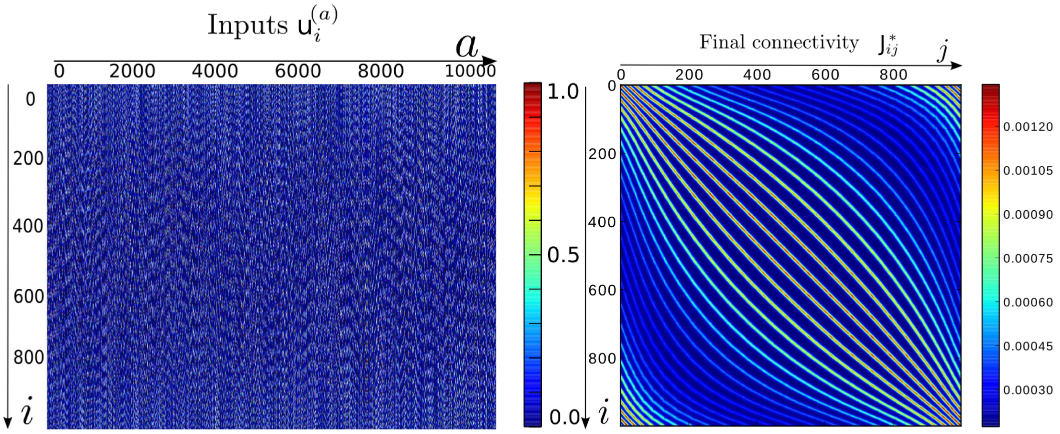


Figure 3.10: Plot of retinotopic inputs on $\Omega = \mathbb{T}^2$ (left) and the final connectivity matrix (right) for the system Σ'. The parameters used for this simulation are $S(x) = \frac{1}{1+e^{-4(x-1)}}$, $l = 1$, $\kappa = 10$, $n = 1000, m = 10,000$, $\sigma = 2$.

We now assume that the inputs are uniformly distributed over a two-dimensional torus, i.e. $\Omega = \mathbb{T}^2$. That is, the input labels $\mathbf{z}_a$ are randomly distributed on the torus. The neuron labels $\mathbf{y}_i$ are regularly and uniformly distributed on the torus. The inputs and final stable weight matrix of the

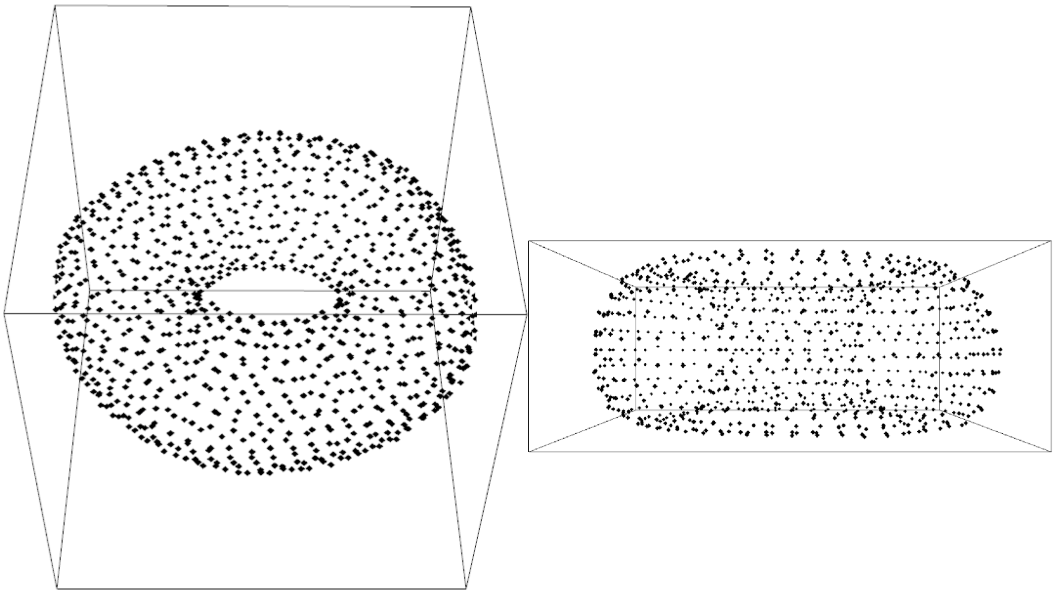


Figure 3.11: Positions $\mathbf{x}_i$ of the neurons for $k = 3$ after having applied multidimensional scaling methods presented in section 3.3.1 to the final connectivity matrix shown in figure 3.10.

simulated system are shown in figure 3.10. The positions $\mathbf{x}_i$ of the neurons after multidimensional scaling for $k = 3$ are shown in figure 3.11, and appear to form a cloud of points distributed on a torus. In contrast to the previous example, there are no distortions now because there are no borders on the torus. In fact, the neurons are equally excited on average in this case which makes property 3.3.1 valid.

3.3.3 Links with neuroanatomy

The brain is subject to energy constraints which are completely neglected in the above formulation. These constraints most likely have a significant impact on the positions of real neurons in the brain. Indeed, it seems reasonable to assume that the positions and connections of neurons reflect a trade-off between the energy costs of biological tissue and their need to process information effectively. For instance, it has been suggested that a principle of wire length minimization may occur in the brain (see [Swindale 1996, Chklovskii et al. 2002]). In our neural mass framework, one may consider that the stronger two neural masses are connected, the larger the number of real axons linking the neurons together. Therefore, minimizing axonal length can be read as: the stronger the connection the closer, which is consistent with the convolutional part of the weight matrix. However, the underlying geometry of natural inputs is likely to be very high-dimensional, whereas the brain lies in a three-dimensional world. In fact, the cortex is so flat that it is effectively two-dimensional. Hence, the positions of real neurons are different from the positions $\mathbf{x}_i \in \mathbb{R}^k$ in a high dimensional vector space; since the cortex is roughly two-dimensional, the positions could only be realized physically if $k = 2$. Therefore, the three-dimensional toric geometry or any higher dimensional structure could not be perfectly implemented in the cortex without the help of non-convolutional long-range connections. Indeed, we suggest that the cortical connectivity is made of two parts: i) a local convolutional connectivity corresponding to the convolutional term g in (3.4) and (3.5), which is consistent with the requirements of energy efficiency, and ii) a non-convolutional connectivity corresponding to the factor $\mathbf{M}$ in equations (3.4) and (3.5), which is required in order to represent various stimulus features. If the cortex were higher-dimensional ($k \gg 2$) then there would no non-convolutionnal connectivity $\mathbf{M}$, i.e. $\mathbf{M} \equiv 0$ for the linear convolutional kernel or $\mathbf{M} \equiv 1$ for the exponential one.

We illustrate the above claim by considering two examples based on the functional anatomy of the primary visual cortex: the emergence of ocular dominance columns and orientation columns, respectively. We proceed by returning to the case of planar retinotopy (section 5.3.1) but now with additional input structure. In the first case, the inputs are taken to be binocular and isotropic, whereas in the second case they are taken to be monocular and anisotropic. The details are presented below. Given a set of prescribed inputs, the network evolves according to equation (2.17) and the lateral connections converge to a stable equilibrium. The resulting weight matrix is then projected onto the set of distance matrices for $k = 2$ (as described in section 3.3.2.1) using the stress majorization or SMACOF algorithm for the stress1 cost function as described in [Borg and Groenen 2005]. We thus assign a position $\mathbf{x}_i \in \mathbb{R}^2$ to the ith neuron, $i = 1, \ldots, n$. (Note that the position $\mathbf{x}_i$ extracted from the weights using multidimensional scaling is distinct from the "physical" position $\mathbf{y}_i$ of the neuron in the retino-cortical plane; the latter determines the center of its receptive field). The convolutional connectivity

(g in equations (3.4) and (3.5)) is therefore completely defined: on the planar map of points $\mathbf{x}_i$, neurons are isotropically connected to their neighbors; the closer the neurons are the stronger is their convolutional connection. Moreover, since the stimulus feature preferences (orientation, ocular dominance) of each neuron i, $i = 1, \dots, n$, are prescribed, we can superimpose these feature preferences on to the planar map of points $\mathbf{x}_i$. In both examples, we find that neurons with the same ocular or orientation selectivity tend to cluster together: interpolating these clusters then generates corresponding feature columns. It is important to emphasize that the retino-cortical positions $\mathbf{y}_i$ do not have any columnar structure, that is, they do not form clusters with similar feature preferences. Thus, in contrast to standard developmental models of vertical connections, the columnar structure emerges from the recurrent weights post-learning which are interpreted as a Euclidean distances. It follows that neurons coding for the same feature tend to be strongly connected; indeed, the multidimensional scaling algorithm has the property that it positions strongly connected neurons close together . Equations (3.4) and (3.5) also suggest that the connectivity has a non-convolutional part, $\mathbf{M}$, which is a consequence of the low-dimensionality ($k = 2$). In order to illustrate the structure of the non-convolutional connectivity, we select a neuron i in the plane and draw a link from it at position $\mathbf{x}_i$ to the neurons at position $\mathbf{x}_j$ for which $\mathbf{M}(\mathbf{x}_i, \mathbf{x}_j)$ is maximal in figures 3.13, 3.14 and 3.15. We find that $\mathbf{M}$ tends to be patchy, i.e. it connects neurons having the same feature preferences. In the case of orientation, $\mathbf{M}$ also tends to be co-aligned, i.e. connecting neurons with similar orientation preference along a vector in the plane of the same orientation.

3.3.3.1 Ocular dominance columns and patchy connectivity

In order to construct binocular inputs, we partition the n neurons into two sets $i \in \{1, \dots, n/2\}$ and $i \in \{n/2+1, \dots, n\}$ that code for the left and right eyes, respectively. The ith neuron is then given a retino-cortical position $\mathbf{y}_i$, with the $\mathbf{y}_i$ uniformly distributed across the line for figure 3.12 and across the plane for figures 3.13 and 3.13. We do not assume *a priori* that there exist any ocular dominance columns, that is, neurons with similar retino-cortical positions $\mathbf{y}_i$ do not form clusters of cells coding for the same eye. We then take the ath input to the network to be of the form

$$\begin{aligned} &\text{Left eye} \quad \mathbf{u}_i^{(a)} &&= (1+\gamma(a))e^{-\frac{(\mathbf{y}_i-\mathbf{z}_a)^2}{\sigma'^2}}, \quad && i = 1, \ldots, n/2 \\ &\text{Right eye} \quad \mathbf{u}_i^{(a)} &&= (1-\gamma(a))e^{-\frac{(\mathbf{y}_i-\mathbf{z}_a)^2}{\sigma'^2}}, \quad && i = n/2+1, \ldots, n, \end{aligned} \tag{3.10}$$

where the $\mathbf{z}_a$ are randomly generated from $[0, 1]$ in the 1-dimensional case and $[0, 1]^2$ in the 2-dimensional case. For each input a, $\gamma(a)$ is a randomly and uniformly taken in $[-\gamma, \gamma]$ with $\gamma \in [0, 1]$ (see [Bressloff 2005]). Thus, if $\gamma(a) > 0$ ($\gamma(a) < 0$) then the corresponding input is predominantly from the left (right) eye.

First, we illustrate the results of ocular dominance simulations in one dimension in figure 3.12. Although not biologically realistic, taking the visual field to be one dimensional makes it possible to visualize the emergence of ocular dominance columns more easily. Indeed, in figure 3.12 we analyze the role of the binocular disparity of the network, i.e. we change the value of γ. If $\gamma = 0$ (blue curves in figures 3.12.a and 3.12.b and top pictures in figures 3.12.c and 3.12.d), there are virtually no differences between left and right eyes and we observe much less segregation than in the case $\gamma = 1$ (green curves in figures 3.12.a and 3.12.b and bottom pictures in figures 3.12.c and 3.12.d). Increasing the binocular disparity between the two eyes results in the emergence of ocular dominance columns. Yet, there does not seem any spatial scale associated with these columns: they form on various scales as shown in figure 3.12.d.

In figures 3.13 and 3.14, we plot the results of ocular dominance simulations in two dimensions. In particular, we illustrate the role of changing the binocular disparity γ, changing the standard deviation of the inputs σ' and using different convolutional kernels g. We plot the points $\mathbf{x}_i$ obtained by performing multidimensional scaling on the final connectivity matrix for $k = 2$, and superimposing upon this the ocular dominance map obtained by interpolating between clusters of neurons with the same eye preference. The convolutional connectivity (g in equations (3.4) and (3.5)) is implicitly described by the position of the neurons: the closer the neurons, the stronger their connections. We also illustrate the non-convolutional connectivity ($\mathbf{M}$ in equations (3.4) and (3.5)) by linking one selected neuron to the neurons it is most strongly connected to. The color of the link refers to the color of the target neuron. The multidimensional scaling algorithm was applied for each set of parameters with different initial conditions and the best final solution

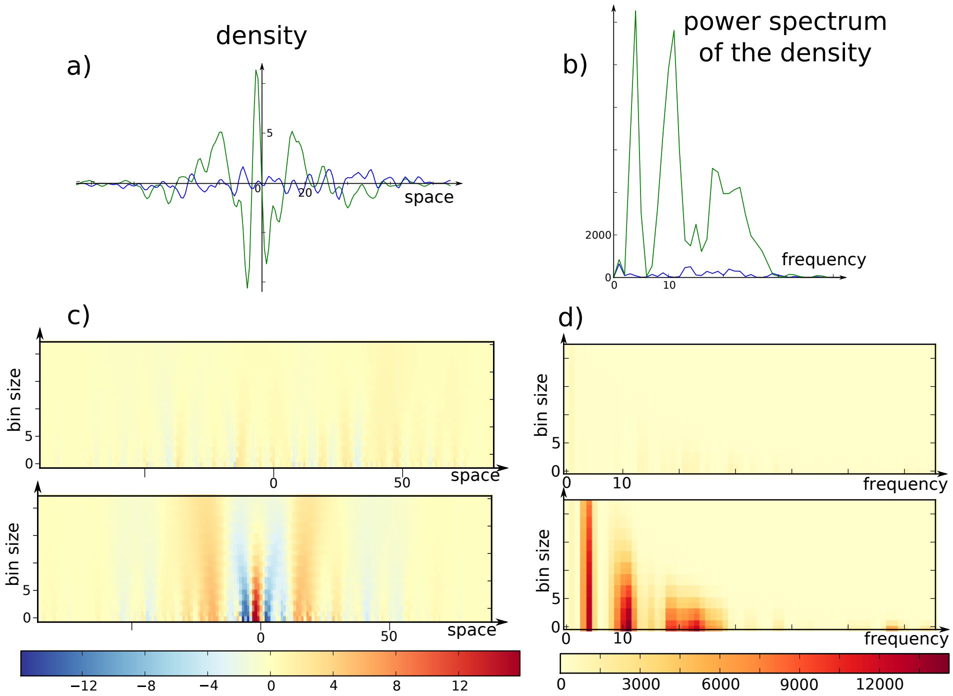


Figure 3.12: Emergence of ocular dominance columns. Analysis of the equilibrium connectivity of a network of $n = 1000$ neurons exposed to $m = 3000$ inputs as described in equation (3.10) with $\sigma' = 10$. The parameters used for this simulation are $u_m = 1$, $\kappa = 10$ and $S(x) = \frac{1}{1+e^{-4(x-1)}}$. (a) Relative density of the network assuming that the "weights" of the left neurons are +1 and the "weights" of the right eye neuron are −1. Thus, a positive (resp. negative) lobe corresponds to a higher number of left neurons (resp. right neurons) and the presence of oscillations implies the existence of ocular dominance columns. The size of the bin to compute the density is 5. The blue (resp. green) curve corresponds to $\gamma = 0$ (resp. $\gamma = 1$). It can be seen that the case $\gamma = 1$ exhibits significant oscillations consistent with the formation of ocular dominance columns. (b) Power spectra of curves plotted in (a). The dependence of the density and power spectrum on bin size is shown in (c) and (d), respectively. The top pictures correspond to the blue curves, i.e. no binocular disparity and the bottom pictures correspond to the green curves $\gamma = 1$, i.e. a higher binocular disparity.

was kept and plotted. The initial conditions were random distributions of neurons or artificially created ocular dominance stripes with different numbers of neurons per stripe. It turns out the algorithm performed better on the latter. (The number of tunable parameters was too high for the system to converge to a global equilibrium for a random initial condition). Our results show that non-convolutional or long–range connections tend to link cells with the same ocular dominance provided the inputs are sufficiently strong and different for each eye.

3.3.3.2 Orientation columns and collinear connectivity

In order to construct oriented inputs, we partition the n neurons into four groups Σ_θ corresponding to different orientation preferences $\theta = \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. Thus, if neuron $i \in \Sigma_\theta$ then its orientation preference is $\theta_i = \theta$. For each group, the neurons are randomly assigned a retino-cortical position $\mathbf{y}_i \in [0,1] \times [0,1]$. Again, we do not assume *a priori* that there exist any orientation columns, that is, neurons with similar retino-cortical positions $\mathbf{y}_i$ do not form clusters of cells coding for the same orientation preference. Each cortical input $\mathbf{u}_i^{(a)}$ is generated by convolving a thalamic input consisting of an oriented Gaussian with a Gabor–like receptive field (as in [Miikkulainen et al. 2005]). Let $\mathcal{R}_\theta$ denote a 2-dimensional rigid body rotation in the plane with $\theta \in [0, 2\pi)$. Then

$$\mathbf{u}_i^{(a)} = \int G_i(\boldsymbol{\xi} - \mathbf{y}_i) I_a(\boldsymbol{\xi} - \mathbf{z}_a) d\boldsymbol{\xi}, \tag{3.11}$$

where

$$G_i(\boldsymbol{\xi}) = G_0(\mathcal{R}_{\theta_i}\boldsymbol{\xi}) \tag{3.12}$$

and $G_0(\boldsymbol{\xi})$ is the Gabor–like function

$$G_0(\boldsymbol{\xi}) = A_+ e^{-\boldsymbol{\xi}'.\boldsymbol{\Lambda}^{-1}.\boldsymbol{\xi}} - A_- e^{-(\boldsymbol{\xi}-\mathbf{e}_0)'.\boldsymbol{\Lambda}^{-1}.(\boldsymbol{\xi}-\mathbf{e}_0)} - A_- e^{-(\boldsymbol{\xi}+\mathbf{e}_0)'.\boldsymbol{\Lambda}^{-1}.(\boldsymbol{\xi}+\mathbf{e}_0)}$$

with $\mathbf{e}_0 = (0,1)$ and

$$\boldsymbol{\Lambda} = \begin{pmatrix} \sigma_{large} & 0 \\ 0 & \sigma_{small} \end{pmatrix}.$$

The amplitudes $A_+, A_- \in \mathbb{R}$ are chosen so that $\int G_0(\boldsymbol{\xi}) d\boldsymbol{\xi} = 0$. Similarly, the thalamic input $I_a(\boldsymbol{\xi}) = I(\mathcal{R}_{\theta'_a}\boldsymbol{\xi})$ with $I(\boldsymbol{\xi})$ the anisotropic Gaussian

$$I(\boldsymbol{\xi}) = e^{-\boldsymbol{\xi}'.\boldsymbol{\Lambda}'^{-1}.\boldsymbol{\xi}}, \qquad \Lambda' = \begin{pmatrix} \sigma'_{large} & 0 \\ 0 & \sigma'_{small} \end{pmatrix}.$$

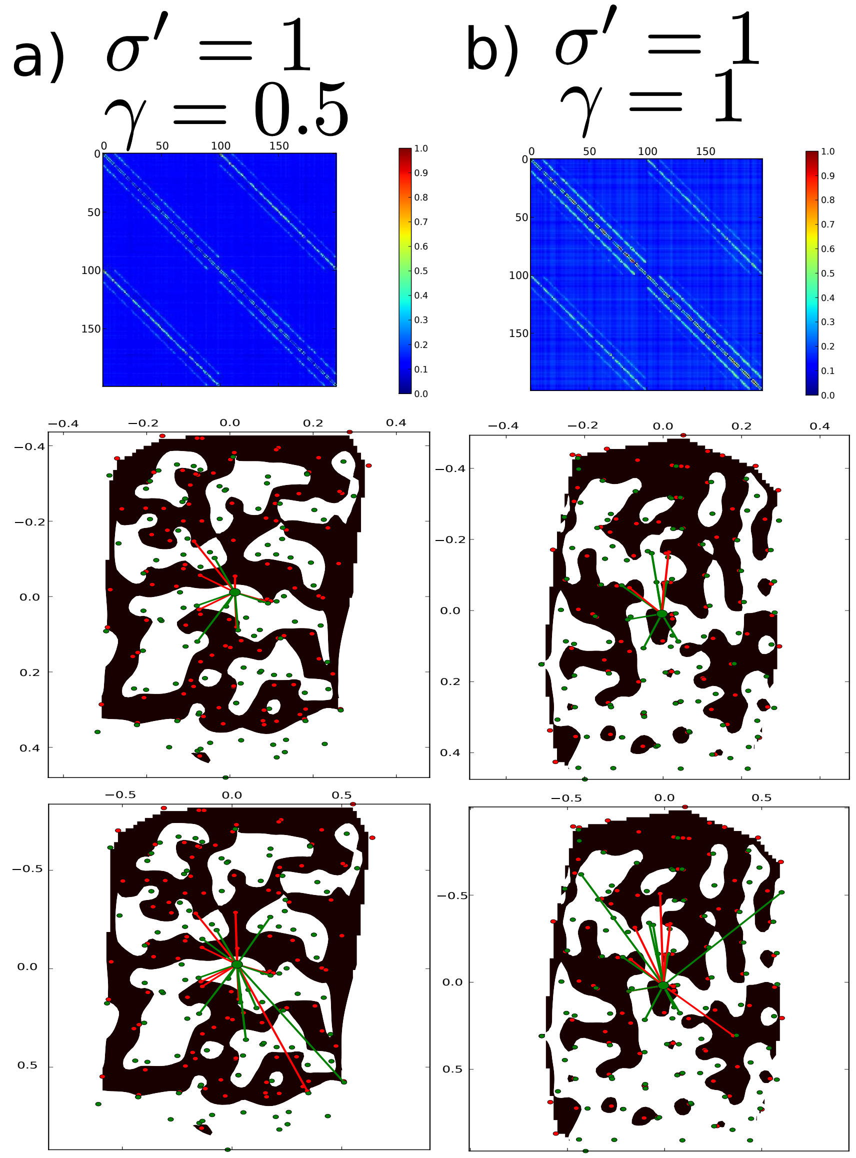


Figure 3.13: Analysis of the equilibrium connectivity of a modifiable recurrent network driven by 2-dimensional binocular inputs. This figure and figure 3.14 correspond to particular values of the disparity γ and standard deviation σ'. Each cell shows the profile of the inputs (top), the position of the neurons for a linear convolutional kernel (middle) and an exponential kernel (bottom). The parameters of the kernel (a and λ) were automatically chosen to minimize the non-convolutional part of the connectivity. It can be seen that the choice of the convolutional kernel has little impact on the position of the neurons. (a) and (c) correspond to $\gamma = 0.5$, which mean there is little binocular disparity. Therefore, the non-convolutional connectivity connects neurons of opposite eye preference more than for $\gamma = 1$, as shown in (b) and (d). The inputs for (a) and (b) have a smaller standard deviation than for (c) and (d). It can be seen that the neurons coding for the same eye tend to be closer when σ' is larger. The other parameters used for these simulations are $S(x) = 1/(1 + e^{-4(x-1)})$, $l = 1$, $\kappa = 10$, $n = m = 200$.

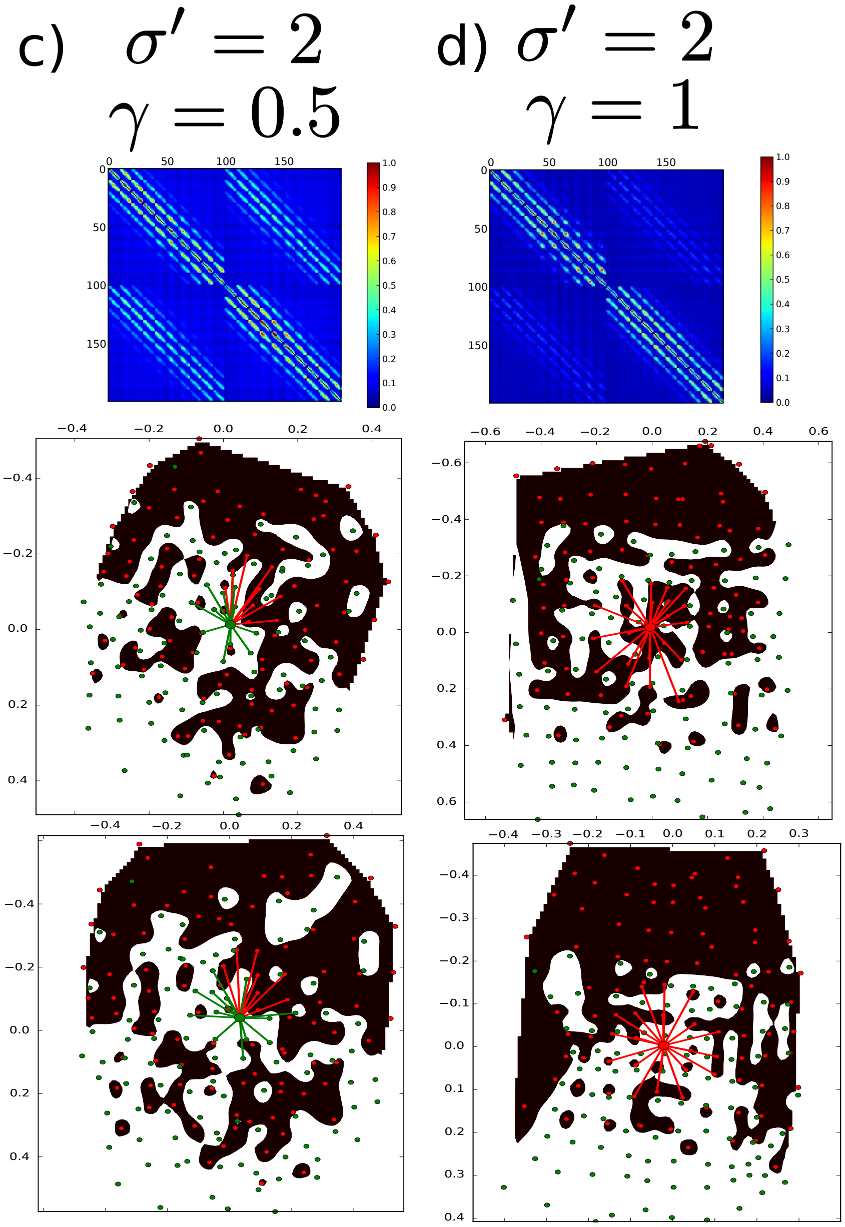


Figure 3.14: See figure 3.13 for the commentaries.

The input parameters θ'_a and $\mathbf{z}_a$ are randomly generated from $[0, \pi)$ and $[0, 1]^2$ respectively. In our simulations we take $\sigma_{large} = 0.133...$, $\sigma'_{large} = 0.266...$ and $\sigma_{small} = \sigma'_{small} = 0.0333....$ The results of our simulations are shown in the left picture of figure 3.15. In particular, we plot the points $\mathbf{x}_i$ obtained by performing multidimensional scaling on the final connectivity matrix for $k = 2$, and superimposing upon this the orientation preference map obtained by interpolating between clusters of neurons with the same orientation preference. To avoid border problems we have zoomed on the center on the map. We also illustrate the non-convolutional connectivity by linking a group of neurons gathered in an orientation column to all other neurons for which $\mathbf{M}$ is maximal. The patchy, anisotropic nature of the long–range connections is clearly seen. The anisotropic nature of the connections is further quantified in the histogram of figure 3.15.

3.3.4 Summary, conclusions and immediate extensions

In this section, we have shown how a neural network can learn the underlying geometry of a set of inputs.

We have considered the solution of a fully recurrent neural network whose connections were slowly learned through Hebbian learning with decay. The equilibrium connectivity is the starting point of this chapter. The approach could be extend to any symmetric learning rule and we would get the same qualitative results. If the learning rule was asymmetric this method is to be applied on the symmetric part of the connectivity.

We have then demonstrated how the connectivity matrix can be expressed as a distance matrix in $\mathbb{R}^k$ for sufficiently large k when the sigmoid is positive, which can be related to the underlying geometrical structure of the inputs. Indeed, this methods gives a position in $\mathbb{R}^k$ to all the neurons. On the geometrical shape suggested by the neurons' distribution in space, the network appears to be convolutional: the connectivity is only a function on the distance between the neurons.

If the connectivity matrix is embedded in a lower two-dimensional space ($k = 2$), then the emerging patterns are similar to experimentally observed cortical feature maps. That is, neurons with the same feature preferences tend to cluster together forming cortical columns within the embedding space. Moreover, the recurrent weights decompose into a local isotropic convolu-

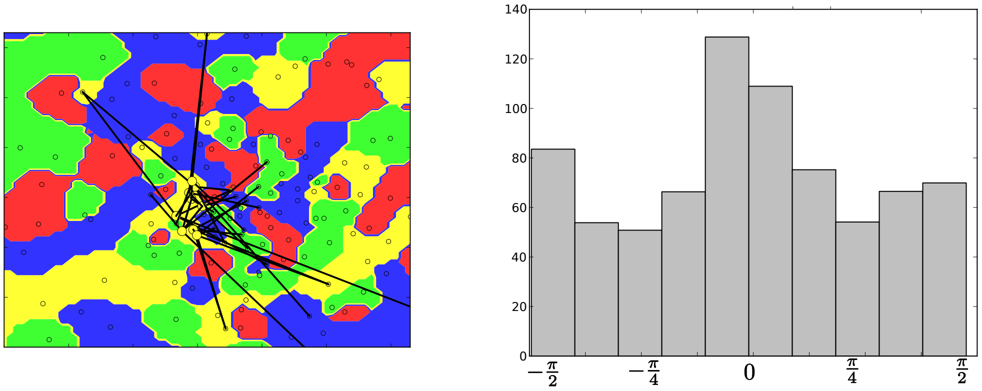


Figure 3.15: Emergence of orientation columns. (Left) Plot of the positions $\mathbf{x}_i$ of neurons for $k = 2$ obtained by multidimensional scaling of the weight matrix. Neurons are clustered in orientation columns represented by the colored areas, which are computed by interpolation. The strongest components of the non-convolutional connectivity ($\mathbf{M}$ in equation (3.5)) from a particular neuron in a yellow area are illustrated by drawing black links from this neuron to the target neurons. Since the yellow color corresponds to an orientation of $\frac{3\pi}{4}$, the non-convolutional connectivity shows the existence of a co-linear connectivity as exposed in [Bosking et al. 1997]. The parameters used for this simulation are $S(x) = \frac{1}{1+e^{-4(x-1)}}$, $l = 1$, $\kappa = 10$, $n = 900$, $m = 9000$. (Right) Histogram of the 5 largest components of the non-convolutional connectivity for 80 neurons randomly chosen among those shown in the left picture. The abscissa corresponds to the difference in radian between the direction preference of the neuron and the direction of the links between the neuron and the target neurons. This histogram is weighted by the strength of the non-convolutional connectivity. It shows a preference for co-aligned neurons but also a slight preference for perpendicularly-aligned neurons (e.g. neurons of the same orientation but parallel to each other).

tional part, which is consistent with the requirements of energy efficiency, and a longer–range non-convolutional part that is patchy. This suggest a new interpretation of the cortical maps: they correspond to two-dimensional embeddings of the underlying geometry of the inputs.

Geometric diffusion methods (see [Coifman et al. 2005]) are also an efficient way to reveal the underlying geometry of sets of inputs. There are several reasons why multidimensional scaling of the lateral connectivity is preferred. First, the focus of this thesis is not the direct analysis of the inputs but the study of the final lateral connectivity of a learning network. The advantage is that the connectivity is a $n \times n$ matrix whereas the size of the inputs is $n \times m$ which is potentially much higher. Besides, the present approach allocates a position to the neurons, as opposed to the inputs (which is the case for geometric diffusion methods). This makes these two techniques different in nature. Second, we are interested in decomposing the connectivity between a convolutionnal and non-convolutionnal part and this is why we focus not only on the spatial structure but also on the shape of the activity equation on this structure. This two results come together when decomposing the connectivity. Actually, this focus on the connectivity was necessary to use the energy minimization argument of part 2.3.2.1 and compute the cortical maps in part 3.3.3. This would have made no sense in the diffusion geometric framework. In conclusion, these two approach share the same philosophy but diffusion geometry is focused on inputs whereas ours is focused on the connectivity of the network.

One of the limitations of applying simple Hebbian learning to recurrent cortical connections is that it only takes into account excitatory connections, whereas 20% of cortical neurons are inhibitory. Indeed, in most developmental models of feed-forward connections, it is assumed that the local and convolutional connections in cortex have a Mexican hat shape with negative (inhibitory) lobes for neurons that are sufficiently far from each other. From a computational perspective, it is possible to obtain such a weight distribution by replacing Hebbian learning with some form of covariance learning (see [Sejnowski and Tesauro 1989]). However, it is difficult to prove convergence to a fixed point in the case of the covariance learning rule, and multidimensional scaling method cannot be applied directly unless the Mexican hat function is truncated so that it is invertible. Another limitation of rate-based Hebbian learning is that it does not take into account causality, in contrast to more biologically detailed mechanisms such as spike timing dependent plasticity.

The approach taken here is very different from standard treatments of cortical development (as in [Miller et al. 1989, Swindale 1996]), in which the recurrent connections are assumed to be fixed and of convolutional Mexican hat form whilst the feed-forward vertical connections undergo some form of correlation-based Hebbian learning. In the latter case, cortical feature maps form in the physical space of retino-cortical coordinates $\mathbf{y}_i$, rather than in the more abstract planar space of points $\mathbf{x}_i$ obtained by applying multidimensional scaling to recurrent weights undergoing Hebbian learning in the presence of fixed vertical connections. A particular feature of cortical maps formed by modifiable feed-forward connections is that the mean size of a column is determined by a Turing-like pattern forming instability, and depends on the length scales of the Mexican hat weight function and the two-point input correlations (see [Miller et al. 1989, Swindale 1996]). No such Turing mechanism exists in our approach so that the resulting cortical maps tend to be more fractal-like (many length scales) compared to real cortical maps. Nevertheless, we have established that the geometrical structure of cortical feature maps can also be encoded by modifiable recurrent connections. This should have interesting consequences for models that consider the joint development of feed-forward and recurrent cortical connections. One possibility is that the embedding space of points $\mathbf{x}_i$ arising from multidimensional scaling of the weights becomes identified with the physical space of retino-cortical positions $\mathbf{y}_i$. The emergence of local convolutional structures together with sparser long-range connections would then be consistent with energy efficiency constraints in physical space.

This section also draws a direct link between the recurrent connectivity of the network and the positions of neurons in some vector space such as $\mathbb{R}^2$. In other words, learning corresponds to moving neurons or nodes so that their final position will match the inputs' geometrical structure. Similarly, the Kohonen algorithm detailed in [Kohonen 1990] describes a way to move nodes according to the inputs presented to the network. It also converges toward the underlying geometry of the set of inputs. Although not formally equivalent, it seems that both of these approaches have the same qualitative behavior. However, our method is more general in the sense that no neighborhood structure is assumed *a priori*; such a structure emerges via the embedding into $\mathbb{R}^k$.

Finally, note that we have used a discrete formalism based on a finite number of neurons. However, the resulting convolutional structure obtained by expressing the weight matrix as a distance matrix in $\mathbb{R}^k$, see equations (3.4) and (3.5), allows us to take an appropriate continuum limit. This then generates a continuous neural field model in the form of an integro-differential equation whose integral kernel is given by the underlying weight distribution. Neural fields have been used increasingly to study large–scale cortical dynamics (see [Coombes 2005] for a review). Our geometrical learning theory provides a developmental mechanism for the formation of these neural fields. One of the useful features of neural fields from a mathematical perspective is that many of the methods of partial differential equations can be carried over. Indeed, for a general class of connectivity functions defined over continuous neural fields, a reaction-diffusion equation can be derived whose solution approximates the firing rate of the associated neural field [Degond and Mas-Gallic 1989, Cottet 1995, Edwards 1996]. It appears that the necessary connectivity functions are precisely those that can be written in the form (3.5). This suggests that a network that has been trained on a set inputs with an appropriate geometrical structure behaves as a diffusion equation in a high-dimensional space together with a reaction term corresponding to the inputs.

3.4 Anti-symmetric part of the recurrent connectivity: a vector field

We have shown in section 2.4.4 than a learning rule for a rate-based, linear neural networks inspired from STDP lead to an equilibrium connectivity which has both a symmetric and an anti-symmetric part. The analysis of the symmetric part was done in the previous part and leads to the definition of a distribution of points which, we believe, corresponds to the underlying geometry of the inputs. This part is devoted to the analysis of the anti-symmetric part of the connectivity.

What we want to prove here is that the antisymmetric part of the connectivity codes for a vector field on top of the geometrical structure defined in the previous section. We believe this vector field corresponds to that of the inputs.

This part is not finalized yet. We did not manage to use rigorously the mathematical language to get the expected result. Besides, we have not merged yet the results of this section and the previous one. In particular, we work with a predefined geometry instead of inferring it thanks to the previous section. This part has to be considered as a detailed perspective section.

3.4.1 The input as the solution of a dynamical system

First, we need to assume that the inputs $\mathbf{u} : t \in \mathbb{R} \mapsto \mathbf{u}(t) \in \mathbb{R}^n$ correspond to the solution of an autonomous dynamical system, i.e. there exists $\xi : \mathbb{R}^n \to \mathbb{R}^n$ a vector field such that

$$\dot{\mathbf{u}}(t) = \xi\big(\mathbf{u}(t)\big) \tag{3.13}$$

Actually, there a several ways to make this more general:

- An explicit time dependence of ξ can be considered so that $\mathbf{u}$ is the solution of a non-autonomous system.

- The network's inputs $\mathbf{u}$ can be seen as an observable of a hidden state of the environment written $\mathbf{x} \in \mathbb{R}^{\tilde{n}}$ in higher dimension, i.e. $\tilde{n} \geq n$. Redefining ξ on $\mathbb{R}^{\tilde{n}}$ and with $\zeta : \mathbb{R}^{\tilde{n}} \to \mathbb{R}^{\tilde{n}}$, the system would be

$$\begin{array}{rcll} \dot{\mathbf{x}} & = & \xi\big(\mathbf{x}(t)\big) & \text{evolution of the environment} \\ \mathbf{u}(t) & = & \zeta\big(\mathbf{x}(t)\big) & \text{observable of the environment} \end{array}$$

In this framework, the hidden cortical layers may be responsible of defining and estimating the hidden environment variables.

- Some noise could be added to the inputs equation to take into account their intrinsic variability and/or the noisy behaviors of the sensors. This may give
$$d\mathbf{u}(t) = \xi(\mathbf{u}(t))dt + \eta(\mathbf{u}) \cdot dB(t)$$
where $\eta : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ and B is a n-dimensional Brownian noise.

- The inputs need not to be defined on a Euclidean space (such as $\mathbb{R}^n$) but can be defined on a n-dimensional manifold $\mathcal{M}$. The dynamical (3.13) can therefore be redefined on a manifold by considering that $\xi : \mathcal{M} \to \mathbb{R}^n$.

All these additional level of complexity could also be combined. Yet, in the following, we focus on the initial formulation (3.13) for simplicity.

3.4.2 An analytic tentative to extract the vector field

In this part, we remove the inputs' term and, therefore, consider the spontaneous activity of the network. It is governed by

$$d\mathbf{v} = \big(- l\mathbf{v} + \mathbf{J}^*.\mathbf{v}\big)dt + \mathbf{\Sigma}.dB(t)$$

We focus on the communication term $\mathbf{J}^*.\mathbf{v}$ to show how the vector field of the inputs $\mathbf{u}$ emerges in the activity equation.

Assume that the connectivity matrix is the result of purely anti-symmetric STDP learning, i.e. $a_\pm = a_+ = a_-$, for inputs which where shown very slowly, i.e. $\mu = 0$. According to section 2.4.4, and more precisely equation (2.35) the final connection is

$$\mathbf{J}^* = \frac{2a_\pm}{\gamma\tau}\Big(\frac{d\mathbf{u}}{dt} * g_{1/\gamma}\Big).\Big(\mathbf{u} * g_{1/\gamma}\Big)'$$

When γ tends to infinity $\mathbf{J}^*$ tends to zero. From a biological stand-point, it is absurd to take this limit. However, we want to exploit the fact that $\lim_{\gamma\to+\infty} g_{1/\gamma} = \delta$ so that it is reasonable to consider that, in a certain regime,

$$\mathbf{J}^* \propto \mathbf{J}_\infty \overset{def}{=} \frac{d\mathbf{u}}{dt}.\mathbf{u}'$$

Therefore, using equation (3.13) the computation of the communication term $\mathbf{J}^*.\mathbf{v}$ is uniformly proportional to

$$\mathbf{J}_\infty.\mathbf{v} = \xi(\mathbf{u}).\mathbf{u}'.\mathbf{v}(t) = \int_0^\tau \xi(\mathbf{u})(s) \otimes \mathbf{u}(s)ds.\mathbf{v}(t) = \int_0^\tau \xi(\mathbf{u})(s)\big\langle \mathbf{u}(s), \mathbf{v}(t)\big\rangle ds$$

The idea is to see the last integral as an expectation for a given measure. To do so we must introduce some assumptions and definitions:

- Here, we need to assume that the inputs and the activity are always positive. We think it is a useful property granted in more realistic networks by the positive sigmoid we have neglected here for simplicity. Indeed, considering a voltage-based equation (see section 2.1) instead of a linear equation would immediately give this property. Under this assumption and for a fixed $\mathbf{v}(t) \in \mathbb{R}^n$, the scalar product $\big\langle \mathbf{u}(s), \mathbf{v}(t)\big\rangle$ is always positive and can be seen as a probability density function.

- We assume that τ is large enough so that the (stochastic) inputs have had enough time to sample their distribution. Yet, the inputs may not sample $\mathbb{R}^n$ entirely. Therefore, we introduce the function

  $$\begin{array}{rccc} \chi_\mathbf{u}: & \mathbb{R}^n & \rightarrow & \mathbb{R}_+ \\ & \mathbf{x} & \mapsto & \int_0^\tau \delta\big(\mathbf{x} - \mathbf{u}(s)\big)ds \end{array}$$

  where δ is the Dirac function. In a way, $\chi_\mathbf{u}$ counts the number of times the inputs have passed through a point in $\mathbb{R}^n$. Note that this definition extends very easily to $\mathbf{u}$ being a stochastic process by changing the Dirac function by the probability density function of $\mathbf{u}$ at time s.

This motivates the definitions

$$\begin{array}{rcl} \rho_\mathbf{v}\big(\mathbf{x}\big) & = & \frac{\chi_\mathbf{u}(\mathbf{x})\langle \mathbf{x},\mathbf{v}\rangle}{Z_\mathbf{v}} \\ \text{where} \quad Z_\mathbf{v} & = & \int_{\mathbb{R}^n} \chi_\mathbf{u}(\mathbf{x})\langle \mathbf{x}, \mathbf{v}\rangle d\mathbf{x} = \int_0^\tau \langle \mathbf{u}(s), \mathbf{v}\rangle ds \end{array}$$

such that

$$\mathbf{J}^*.\mathbf{v} \propto \mathbf{J}_\infty.\mathbf{v} = Z_\mathbf{v} \int_{\mathbb{R}^n} \xi(\mathbf{x})\rho_\mathbf{v}\big(\mathbf{x}\big)d\mathbf{x} = Z_\mathbf{v}\ \mathbb{E}_{\rho_\mathbf{v}(\mathbf{x})}\big(\xi(\mathbf{x})\big)$$

The factor $\mathbb{E}_{\rho_\mathbf{v}(\mathbf{x})}\big(\xi(\mathbf{x})\big)$ is the expectation of the inputs' vector field for a measure $\rho_\mathbf{v}$ proportional to $\mathbf{x} \mapsto \chi_\mathbf{u}(\mathbf{x})\langle \mathbf{x}, \mathbf{v}\rangle$. It is interesting to observe that $\rho_\mathbf{v}(\mathbf{x})$ is maximal when $\mathbf{x}$ belong to the trajectory of the inputs and is proportional to $\mathbf{v}$. If $\mathbf{v}$ belongs to the trajectory of the inputs (and the inputs are assumed to be normalized by the feed-forward connectivity, i.e. $\|\mathbf{u}(t)\|$ is a constant) then the probability density function is peaked on $\mathbf{v}$ such that $\mathbb{E}_{\rho_{\mathbf{v}}(\mathbf{x})}\big(\xi(\mathbf{x})\big) \simeq \xi(\mathbf{v})$. On the contrary, if $\mathbf{v}$ does not belong to the trajectory of the inputs then the communication term will infer the value of the vector field from the elements of the trajectory of the inputs that are close to $\mathbf{v}$.

The factor $Z_{\mathbf{v}}$ corresponds to the fact that the right hand side can not be an uniform expectation since the result is going to be stronger when $\mathbf{v}(t)$ is close to a very frequent input. For instance, if the inputs converges to an equilibrium point, the value of the integral above is going to be larger when $\mathbf{v}(t)$ is close to this equilibrium point. In other words, the network would need the inputs to sample uniformly $\mathbb{R}^n$ for this integral to be the empirical expectation.

Therefore, the communication term can be though of as a smoothed and weighted version of the inputs vector field.

The network as a predictor of the future of the inputs The activity of such a network post-learning can therefore be seen as

$$\frac{d\mathbf{v}}{dt} = \Big(- l\mathbf{v} + \mathbf{\Sigma}.\frac{dB(t)}{dt}\Big) + \xi(\mathbf{v}) + \mathbf{u}(t)$$

where $\frac{dB(t)}{dt}$ is a white noise. Of course this formulation is not rigorously true but we believe it captures the idea of this section. The first two terms (in the large parenthesis) correspond to a random excitation of the neurons field which rapidly and uniformly fades away. In a way, this is a mechanisms to propose different patterns according to matrix $\mathbf{\Sigma}$. The third term $\xi(\mathbf{v})$ propagates the propositions of the random term as the inputs would have done. It is the predicting term. It models the dynamics of the inputs. The last term $\mathbf{u}$ is an external input which is null when considering the spontaneous activity of the network. It can also trigger some neural representations if switched on briefly. For instance, if $\mathbf{u}$ is the beginning of an input (e.g. a movie) the network is used to seeing and then stops suddenly, the network will still temporarily see the rest of the input because of the term $\xi(\mathbf{v})$.

3.4.3 A computational approach to extract the vector field

Chapter 23 of [Borg and Groenen 2005] reviews the traditional methods to visually represent an asymmetric matrix. Among several methods, they show that any asymmetric matrix can be seen a drift field defined over a distribution of points. The symmetric part of the matrix leads to defining the position of the points whereas the anti-symmetric part of the matrix defines a vector attached to each point to indicate the statistical direction embedded in the matrix. This methods gives a way to convert a matrix to a discretized spatial structure equipped with a vector field. We illustrate this mechanism in the two following examples.

We will also simulate the network post-learning with a punctual stimulation both in time and space to illustrate the predicting capabilities of such networks.

Example 1: A flow along a cylinder In this example we consider inputs described by figure 3.16. The inputs $\mathbf{u}(t) \in \mathbb{R}^{n^2}$ corresponds to a small Gaussian drifting on a cylinder. Its center is written $\mathbf{z}(t) \in [0, 10]^2$ and is assumed to verify

$$d\mathbf{z} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} dt + 0.01 \begin{pmatrix} dB_1(t) \\ dB_2(t) \end{pmatrix}$$

where we assume that $\mathbf{z}(t + dt) = (c, 0)'$ when $\mathbf{z}(t) = (c, 10)'$ to emulate a cylinder.

In the same spirit as section 3.3, we assume the $n^2 \in \mathbb{N}$ neurons are "labeled" by a position $\mathbf{y}_i \in [0, 10]^2$, which are uniformly spread. For simplicity here, we assume that $\mathbf{y}_i = \frac{2}{n}(i \mathbin{//} n, i \,\%\, n)' + (\frac{1}{n}, \frac{1}{n})$ (where we consider $//$ is the integer division and $\%$ the modulo). Therefore, the inputs can be written

$$\mathbf{u}_i(t) = e^{-\frac{\|\mathbf{z}(t) - \mathbf{y}_i\|}{\sigma^2}}$$

where $\sigma = \frac{2}{n}$. Actually, we consider 100 different initial conditions for the inputs (uniformly spread along the left border) and concatenate the results in a single function $\mathbf{u}$.

We then compute the equilibrium connectivity matrix according to section 2.4.4 with $n = 51$, $a_+ = 4$, $a_- = 2$, $\gamma = 0.05$, $\mathbf{\Sigma} = 0.01 I_d$, $\mu = 0$, and we choose τ to correspond to the entire history of the inputs.

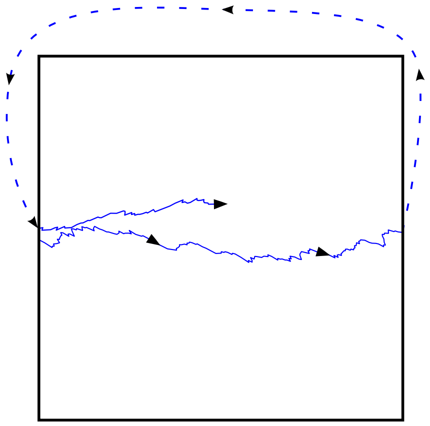


Figure 3.16: The inputs consists in a drifting point with noise in one direction of a square. When the points reaches the right border it is immediately teleported to the left border. This has the effect of emulating a drift on a cylinder.

- In a first time, we apply the methods of [Borg and Groenen 2005] on the equilibrium connectivity $\mathbf{J}^*$. The neurons are assumed to lie on a regular grid on the square $[10, 10]^2$. Actually, we would have had the same retinotopic distribution by applying the methods of section 3.3.2.1. This gives the vector field shown in figure 3.17.

  We see that the left and right extremity of the square have small vectors. This is due to the fact that we emulated a flow on a torus. At the right border the neurons want to send their piece of information to the first neuron on the left. Yet they also strongly inhibits the neurons immediately on their left. Therefore, the two effects cancel out because both of them are in the same direction. A similar reasoning applies to neurons on the right border.

  In this case, the vector field inferred by equation (3.4.3) is well retrieved by the network.

- We now compute the activity field for any input. We use a (stochastic) Euler method for 6000 time-steps with $dt = 5$. The initial condition is null. The parameter l is chosen so that the largest eigenvalues real part of $-lI_d + \mathbf{J}^*$ is null. We choose an input which corresponds to a temporary excitation of a single neuron of $\mathbf{u}_i = 0.2$ between $t =$ and $t = 1200$ and $t = 2400$. The results are shown in figure 3.18. We

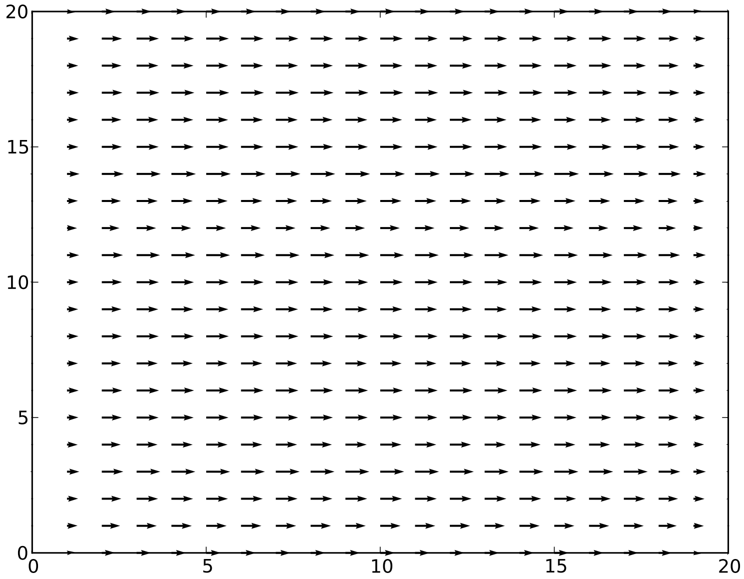


Figure 3.17: This is the vector field deduced from the anti-symmetric part of the equilibrium connectivity of the system in example 1 and according to chapter 23 of [Borg and Groenen 2005]. For readability we have chosen $n = 21^2$ for this simulation.

have located the neurons according to their label position $\mathbf{y}_i$ although the structure of the symmetric part of the equilibrium connectivity is exactly the same as that of section 3.3.2.1 so that the previous method would find the planar geometry of the inputs.

We see that the temporary excitation generates a wave of excitation in the same direction as the learned inputs. In a way, the network propagates the information it expects to see. Actually, it is possible to see the propagation of the noisy patterns due to the white noise in the same direction as the flow on videos of such a system. When the external excitation stops there is still a remaining propagating bump which travels at the same speed as the inputs.

We think the retro-propagating oscillations in figure 3.18 are due to the linearity of the model. The positive spot inhibits the neurons immediately in the opposite direction as the flow which become negatively excited (due to the linearity). Then these negative neurons have the effect of positively exciting the next neurons in the direction opposite to the flow. This mechanism repeats to give these oscillations.

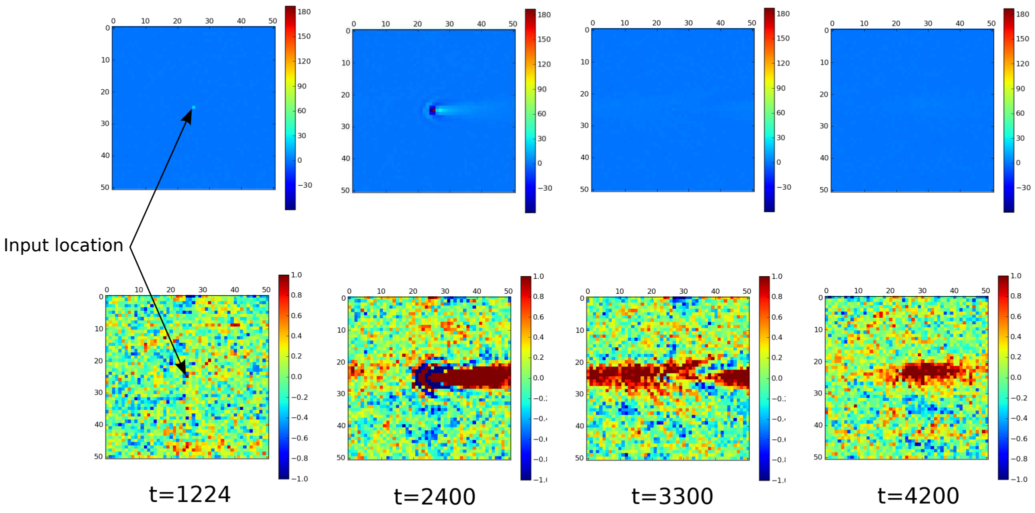


Figure 3.18: The two rows correspond to the same simulation with two different scales. (top) The scale is $[-50, 185]$, it corresponds to the extrema of the field. (bottom) The scale is $[-1, 1]$ and the values larger or smaller than the extrema are thresholded.

Example 2: a spiral in 2 dimensions We consider an input made of localized Gaussians on a square converging to the middle of the square with damped two dimensional oscillations . In other words, the inputs $\mathbf{u}(t) \in \mathbb{R}^{n^2}$ corresponds to a small Gaussian drifting on plane along a spiral. Its center is written $\mathbf{z}(t) \in [0, 10]^2$ and is assumed to verify

$$d\mathbf{z} = \begin{pmatrix} -0.5 & 1 \\ -1 & -0.5 \end{pmatrix} .\mathbf{z}dt + 0.01 \begin{pmatrix} dB_1(t) \\ dB_2(t) \end{pmatrix}$$

200 inputs are drawn randomly, uniformly on the square $[10, 10]^2$. The parameters to compute the equilibrium connectivity are $n = 21$, $a_+ = 1$, $a_- = 0.9999$, $\gamma = 0.5$, $\mathbf{\Sigma} = 0.01 I_d$, $\mu = 0$, and we choose τ to correspond to the entire history of the inputs.

- The vector field extracted from the equilibrium connectivity is shown in figure 3.19.

  We observe that the strength of the vector field is much stronger close to the middle of the square. This is due to the fact that a lot of inputs trajectories have been in this area and therefore the term Z_v is very

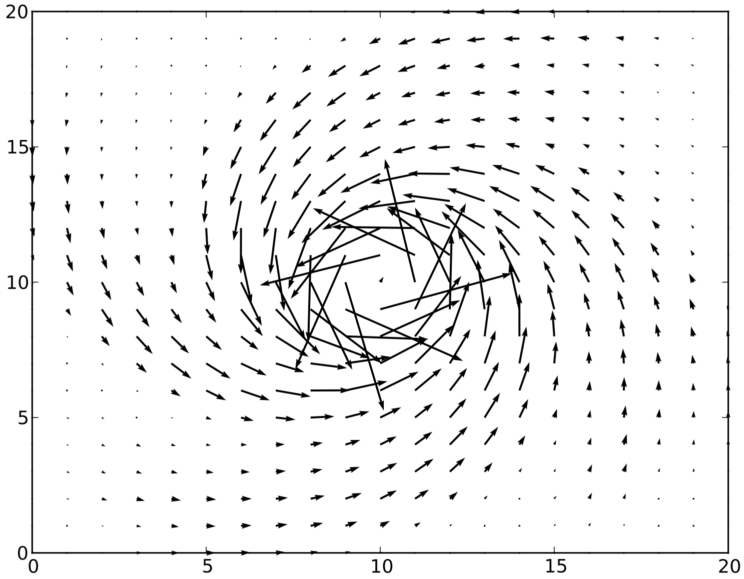


Figure 3.19: This is the vector field deduced from the anti-symmetric part of the equilibrium connectivity of the system in example 2 and according to chapter 23 of [Borg and Groenen 2005].

strong in this area. Put differently, the inputs do not cover uniformly the square and therefore the network knows some area (the center) better than others (the borders).

In this case, also learning has made it possible for the network to learn the vector field of the inputs.

- We now compute the activity field for any input. We use a (stochastic) Euler method for 6000 time-steps with $dt = 0.5$. The initial condition is null. The parameter l is chosen so that the largest eigenvalues real part of $-lI_d + \mathbf{J}^*$ is null. We choose an input which corresponds to a temporary excitation of a single neuron of $\mathbf{u}_i = 10.$ between $t =$ and $t = 1200$ and $t = 5600$. The results are shown in figure 3.4.3. We have located the neurons according to their label position $\mathbf{y}_i$ although the structure of the symmetric part of the equilibrium connectivity is exactly the same as that of section 3.3.2.1 so that the previous method would find the planar geometry of the inputs.

  The pictures in the extreme left and right columns show 2 different states that are often visited by the dynamics of the stochastic network. They

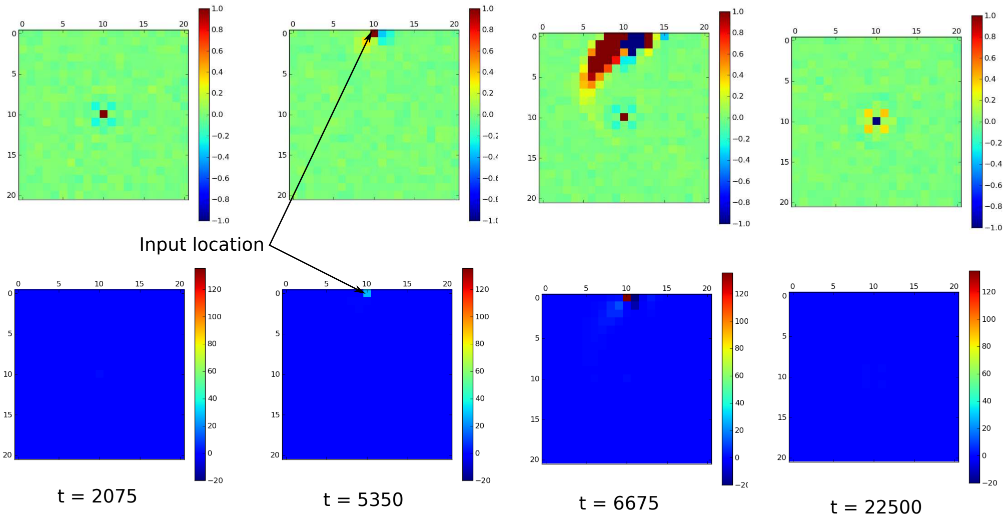


Figure 3.20: The two rows correspond to the same simulation with two different scales. (top) The scale is $[-20, 130]$, it corresponds to the extrema of the field. (bottom) The scale is $[-1, 1]$ and the values larger or smaller than the extrema are thresholded.

can alternate because the network is linear. They have an oscillatory spatial pattern as in the example 1.

We still see a propagation of the inputs activity in the same direction as what the inputs would have done. However, with these parameters there is no residual excitation when the stimulation is stopped.

As in the previous case, we see that the temporary excitation generates a wave which goes in the same direction as the inputs: it describes a spiral.

3.4.4 Summary, conclusions and immediate extensions

In this section, we have shown that the anti-symmetric part of the equilibrium connectivity after STDP learning codes for the vector field generating the inputs. Considering the inputs as the solution of a dynamical system amounts to defining them completely by a manifold and a vector field upon this manifold. While the previous section was devoted to showing the manifold could be extracted from the symmetric connectivity, this section showed how the vector field was extracted from the anti-symmetric connectivity. The technical core of this approach is given by lemma C.3.2 which says that it is the cross-correlation of the inputs with their derivative (the vector field comes in the computation through this mechanism). With this formulation, we have been able to show that the communication term acts as the action of the vector field on the activity.

This approach reveals how the spontaneous activity can "replay" the inputs. We believe equation (3.4.2) is an interesting expression of the spontaneous activity (with $\mathbf{u} = 0$) which is comparable to the inputs given by system (3.13).

The simulations suggest that the network post-learning acts as a predictor of the inputs. In a way, it fills in the gaps left by a potentially noisy input and its observations smoother. In a framework where the inputs $\mathbf{u}$ are just an observable of a real underlying system, we believe the activity field is a more reliable observable.

Chapter 4
Conclusion and perspectives

4.1 Conclusions

The goal of this thesis is to introduce a mathematical theory to show how a neural network with unsupervised learning can create a model of its environment. It consists in defining a learning neural network in mathematical terms from the observations of the biological mechanisms taking place in the brain. These mathematical neural networks are most probably closer to the neocortex than any other part of the brain. Yet, there is still a gap between the simplified mathematical models and the reality of the biological tissues, which calls into question the biological relevance of our approach. Thus, we believe that the interest of this thesis mainly lies in the theory itself. We propose a functional mechanism describing the functioning of a recurrent learning neural network: it copies the statistics and dynamics of the stimuli it has been exposed to. Whatever the nature of the inputs, it learns the way they are statistically distributed and also the direction in which they will most probably evolve. Therefore, this works only if the statistics of the stimuli is not changing through time. In other words, if the network is exposed to visual stimuli first and then for some reason these stimuli change in nature (e.g. the network becomes blind) then the network will forget its knowledge about vision to learn that about a new kind of stimuli, e.g. audition. We believe this thesis is a crucial increment to the literature devoted to unravel the generic mechanism which leads us to learn the ontological structure of any kind of stimuli.

Chapter 1 was devoted to the mathematical modelisation of a learning neural network, i.e. going from biological facts to mathematical models. First we introduced the building blocks of a network: the neurons and the synapse. The molecular mechanisms at the basis of their functioning and their functional behavior were shortly sketched before introducing their mathematical models. The synapse appears to have two different roles: it is a chemical filter at the heart of neural communication and its strength is the variable subject to learning. With these building bocks, we design a fully recurrent spiking neural network made of **noisy McKean neurons with linear synapses and STDP learning**, in system (1.15). This system is very non-linear because of the **spiking** behavior of the neurons and therefore mathematically difficult to handle. To simplify it, we developed our own **mean field theory**. The idea is to consider the neurons belong to different populations and let the number of neurons in each population tend to infinity. We based our analysis on the theoretical result of [Touboul 2011, Baladron et al. 2011], and after a few approximations, we managed to derive a closed differential system describing the evolution of the **firing rate** of each population. This averaged system is close to the traditional rate-based equations. To our knowledge, it is for the first time derived from first principles. It is much more simple to analyze than the former spiking network because the non-linearity is much smoother: it is a sigmoidal function whose precise shape emerged from the computations. In a certain regime (high noise, small input range) it can even be assumed to be linear; yet, this may not correspond to the biological regime. Although we got rid of the spikes in the averaged system, this derivation does not discard the possibility that the spikes carry all the information and that the firing rates do not. Indeed, it only says that the behavior of the network at a larger scale can be described by a firing-rate model. We may have lost some small-scale information.

We also tried to take learning into account during this spatial averaging step. We have suggested a heuristic method to show that the learning equation can also be written as a function of the firing rate only. Yet, this is not satisfying enough and we can not conclude yet because this would have a very deep meaning in terms of learning. Is the spiking behavior of neurons necessary to understand learning in neural tissue? Or is it just a epiphenomenon so that only the firing rate matters, as far as learning is concerned? We can not answer yet, nevertheless, we suggest that the STDP learning rule has an interesting counterpart for rate-based networks.

In chapter 2, we studied the dynamics of the spatially averaged system with learning under the assumption that learning is very slow compared to the evolution of the activity. It makes it possible to study the coupled system connectivity/activity as a **slow/fast system** and apply tools of the **temporal averaging theory**. In the initial system, there is a non-autonomous input to the fast variable corresponding to the stimuli, therefore, the notion of equilibrium point of the system is not well-defined. Yet, the temporally averaged system is a reduced system about the evolution of the connectivity only. It has no external forcing because the fast non-autonomous input is averaged. Therefore, tools from temporal averaging theory make it possible to define an equilibrium point for the slow variable, i.e. the connectivity.

In a first time, we used the Tikhonov theorem and periodic averaging to derive an averaged system in the case of a voltage-based network with Hebbian learning with decay and slow inputs. It appears this system **derives from an energy**, i.e. the right hand side is the opposite of the gradient of an energy function. Dynamically speaking, this is a striking fact, since it implies that the system always converges to a equilibrium point. It is due to a deep mathematical relationship between the communication term (a dot product) and a Hebbian learning (a tensor product): both of them derive from the same term in the energy.

In a second time, we generalized this approach to noisy activities with arbitrarily fast inputs. Based of the recent results developed in [Wainrib 2011] and introduced in appendix B, we were able to derive an averaged, reduced system for the evolution of the connectivity (see theorem 2.4.1). However, this system is only formally defined and can only be made explicit if the activity equation is linear. Therefore, we considered linear activities as motivated by the previous chapter. In this framework, we were able to compute explicitly the right hand side of the reduced system based on the result of a technical paper we wrote in this purpose and exposed in appendix E. It can be shown that the reduced system always converges if the parameters of the system verify a certain assumption. In a **weakly connected regime**, it is even possible to compute explicitly an expansion of the equilibrium connectivity. Because the network is linear, the neurons which do not receive a direct input tend to disconnect from the rest of the network. This problem is irrelevant to the non-linear case since there is always a small intrinsic resting activity, which draws the connections away from zero by the Hebbian mechanism. Another way to go around this problem is to consider that the **noise is spatially correlated**. Then, the neurons which share their noise connect preferentially. The structure that is learned from the inputs is superposed to this artificial structure. It is an *ad hoc* way to impose a hierarchical structure on the network.

We also compared the equilibrium connectivity for different neuron models and different learning rules.

- In the simplest case of **Hebbian learning**, the first order of the equilibrium connectivity corresponds to the correlation matrix of the inputs temporally smoothed by a decreasing exponential function. The temporal decay of this exponential is linked to the speed of the inputs so that:

(i) if the inputs are very slow, the temporal decay goes to zero and the equilibrium connectivity is only the correlation matrix of the inputs, which is a well known result(ii) if the inputs are fast the equilibrium connectivity corresponds to the spatial but also temporal correlations of the inputs. It corresponds to the fact that the excitation elicited by a past stimulus does not have enough time to fade away and might be still strong enough to be learned by association with responses to newer stimuli through the Hebbian mechanism.

- When the neurons are assumed to be **damped oscillators** with **trace learning**, it appears that the equilibrium connectivity corresponds to the correlation of the filtered input. The filter applied to the inputs is a band pass filter centered on the intrinsic oscillatory frequency of the neurons. This suggests a new mechanism for the tonotopic organization of neural tissues (e.g. in the primary auditory cortex): a network of neurons with a lot of different intrinsic frequencies would decompose the signal in a Fourier-like fashion.

- We proved that **STDP learning** is nothing more than an asymmetric rule generalizing Hebbian learning. Indeed, the symmetric part of the equilibrium connectivity is the same as in the Hebbian case, whereas there is an additional anti-symmetric part of the connectivity that was null in the Hebbian case. We observed that only the symmetric part of the STDP rule was responsible for writing the noise in the connectivity, while the anti-symmetric part is noise free. One of the main results of this thesis is that the first order of the anti-symmetric part of the equilibrium connectivity (which is the core of STDP learning) is the cross-correlation of the inputs with their time derivative.

In chapter 3, we focused on the network activity post-learning and showed how it is a model of its environment.
First, we put the recurrent networks that we consider into the context of hierarchical neural networks. Most theoretical research on unsupervised learning in neural networks has been devoted to **feed-forward** connections and more precisely on the **perceptron**. In particular, it was previously shown that learning the feed-forward or retino-cortical connections with the Oja learning rule lead to the extracting the principal components of the inputs, i.e. the

principal eigenvectors of the temporal correlation matrix of the inputs. Using the results of the previous chapter applied to a perceptron, we showed that a simple Hebbian learning rule with decay would extract the eigenvector of the inputs temporal correlation matrix corresponding to a combination of the decay parameters of the system. We also showed that the activity of the post-synaptic neuron extracts the corresponding eigenvector of the spatial correlation matrix of the inputs. This suggests that the previous results of unsupervised learning in a perceptron were over-constrained and partial: a **local** learning rule can extract the **eigenvectors of both the temporal and spatial correlation matrix of the inputs**.

The study of the **recurrent or lateral** connectivity is the next step in the understanding of hierarchical learning networks. Obviously the inputs to this kind of networks have been filtered by the feed-forward connectivity but it is not clear how it restricts the class of cortical inputs. In the spirit of the retinal waves being the inputs to V1 during development, we assumed that the inputs are localized bell-shaped bumps of excitation in a visual field of a given geometry. These inputs might have a certain dynamics, i.e. the bumps may move on the geometry according to a vector field. Actually, this description is generalizable to any kind of inputs: we can assume that these inputs are the solutions of an autonomous dynamical system which is entirely characterized by a vector field upon a manifold (as any dynamical system).

We propose that the network is a model of its environment in the sense that **learning leads to copying the manifold and the vector field of the inputs**. We have not been able to prove rigorously this claim, yet, we used a semi-analytic approach coupled to numerical simulations on simple examples to support it.

- We proposed that the **symmetric part** of the connectivity is responsible for encoding the underlying **geometry or manifold** of the inputs. Indeed, we showed that the symmetric part of the equilibrium connectivity matrix could be interpreted as a **distance matrix** provided the sigmoid was positive. This is equivalent to giving to the neurons a position such that the distances between all the neurons correspond to the coefficients of the connectivity matrix. Actually, we argued that the distance had to be equal to a decreasing function applied to the coefficients so that the stronger the connections between neurons, the closer they have to be. Based on **multidimensional scaling** methods, we were able to give a position in $\mathbb{R}^k$ to the neurons. If k is large enough, i.e. larger than the dimension of the underlying geometry of the inputs, we showed that the geometrical structure of the inputs was retrieved by the networks whose **neurons sample the manifold**. If k was too small, in particular $k = 2$ as in the cortex, the geometry could not be retrieved, yet we saw the emergence of columnar group of neurons which were very similar to **cortical maps**. In fact, we argued that applying the principle "the stronger the connection, the closer the neurons" was similar to the energetic principle of **wire length minimization** in the brain. Indeed, in our framework of populations of neurons, if two populations are strongly connected, it is likely that a lot of axons will go from one to the other. Therefore, we suggested the cortical maps are two dimensional embeddings of high-dimensional geometries corresponding to the underlying structure of the inputs.

- We proposed that the **anti-symmetric part** of the connectivity is responsible for encoding the underlying **vector field** governing the evolution of the inputs. First, we had a semi-analytic approach to show how the communication term due to the anti-symmetric part of the connectivity can be approximated by the vector field of the inputs applied to the activity vector. Second, we showed how to define a vector field from any anti-symmetric matrix and then we illustrated on simple examples how this can be applied to the anti-symmetric part of the equilibrium connectivity to retrieve efficiently the vector field governing the inputs. This vector field can be superposed to the geometry extracted from the symmetric part of the connectivity to provide a discretized version of the dynamical system generating the inputs.

Therefore, the network post-learning corresponds to a dynamical copy of the inputs shown during learning. Thus, the **spontaneous activity** of the network, i.e. when no input is presented, replays in a way the inputs. More precisely, the noise term in the activity equation generates patterns randomly. Then the communication terms measure a sort of distance with the inputs and propagates the excitation according to the statistical behavior that the corresponding inputs have had.

Similarly, if an input (belonging to those that have been learned) is temporarily shown to the network and then suddenly removed, the excitation in the network will propagate so as to predict the evolution of the input: the network is therefore a dynamical and statistical **predictor** of the inputs.

4.2 Perspectives

Heterogeneous networks In this thesis, we have always considered that the networks were made of neurons with the same properties. However, we have mentioned several times the potential usefulness of having a lot of different neurons with different parameters. First, in section 2.4.3 we have shown that, in the case of neurons being damped oscillators, learning lead to computing the correlation matrix of the filtered inputs. The filter was a band-pass filter centered on the intrinsic frequency oscillation of the neurons. Thus having a heterogeneous network would make it possible to process information at different frequencies, which seems a priori very useful. Second, in part 3.1.2.4 we have shown that a perceptron with Hebbian learning with linear decay would lead to the extraction of the eigenvectors of the (temporal and spatial) correlation matrices with an eigenvalue corresponding to a combination of the decay parameters of the neurons and connections. Again, it seems that a heterogeneous network would process a much broader information than a homogeneous one. Finally, it appears compelling from a biological perspective that a homogeneous network does not exist.

However, the study of heterogeneous networks increases significantly the difficulty of proving some mathematical results. For instance, one may assume that the matrix gathering the time constants of all the neurons is not proportional to the identity, i.e. $\mathbf{L} = \text{diag}(l_i)$ with $l_i \neq l_j$ for some i and j. This implies that this matrix is not co-diagonalizable with the connectivity $\mathbf{J}$ and therefore most of the results of appendix C would fall.

Learning the symmetries of the inputs It is common to assume the visual inputs to the retina have a translational and rotational symmetry. In other words, one may assume that the group of rigid motions $E(2)$ leaves the set of inputs invariant. It seems an interesting perspective to study what this claim would mean in term of learning. Would the connectivity learn the symmetry? Actually, we believe that the equilibrium connectivity would be invariant under the action of the symmetries of $E(2)$ in some way. Put differently, the propagation of the action of symmetry groups from the set of inputs to the connectivity is an interesting conjecture to check.

Geometrical role of constraints We have given a geometrical interpretation of the connectivity post-learning. In particular, we gave a position to each neuron in the network and observed that it samples the geometrical support of the inputs. Yet, the density of neurons on this geometrical shape is very dependent on the variety of inputs. In a way, very frequent inputs will be "approximated" by a lot of neurons.

As reviewed in section 2.2.2, there are different constraints that can be added to keep the connectivity in certain subspace. Do these constraints have an interesting interpretation in the geometrical structure we have defined? Does it lead to a more homogeneous spread of the neurons?

Internal hierarchy of cortical layers The neocortex is a biological tissue made of 6 different density layers. These layers are said to have different functional roles. In this thesis, we have completely neglected this additional structure for modeling the cortex. However, we believe that this thesis sets the pace for a new way to analyze this internal structure. Indeed, the learning rules between the different layers seem to be different as shown in figure 1.8. The machinery for studying and explicitly computing the equilibrium connectivities we have built in section 2.4 could be extended to manipulate several learning rules at the same time, and we believe that it would be possible to compute and analyze the different connections from one layer to the others. This may reveal what is the computational role of such a structure.

Note that extending the theory in this direction would also be relevant to studying balanced networks, where neurons are grouped by pair of excitatory and inhibitory neurons.

Hierarchy: networks, with feed-forward, feedback and lateral connections In section 3.1.3, we defined a canonical hierarchical network and argued that modeling recurrent or lateral networks was the step we took toward a better understanding of the full hierarchy. Indeed, these hierarchical networks are very interesting since they seem to build abstraction pyramids with high-level concepts at the top of the hierarchy and low-level concepts at the bottom. This thesis was devoted to learning the lateral connectivity of raw inputs and therefore low-level information. It is a fascinating problem to study the interaction between feed-forward, feedback and recurrent connectivity.

Synaptic tagging: Coupling with reinforcement learning Actually, the modifications of the connectivity due to activity-based unsupervised learning, as described in this thesis, fade away in a few hours. In fact, there is an additional mechanism called synaptic tagging which makes it possible to store definitively the modifications of the connectivity. This roughly corresponds to a global signal modulating the learning rule: if, for some reason, the signal is active then the network will learn "for ever" and else if the signal is inactive the network will forget. The criterion to release the global signal would be very interesting to study. We may suspect that the global signal is active if the network is exposed to some crucial information.

This modulatory signal seems to be a way to perform reinforcement learning. It would be especially interesting to be studied in a network were not only perceptive cortical areas are modeled, but also motor areas. In this context the activity of some (motor) neurons would have an impact on the environment and one could think that unsupervised learning with a modulatory effect based on some criterion may lead to the emergence of behavior.

Does it work? This thesis is theoretical. It intends to build a mixed mathematical and computational theory of unsupervised learning in neural networks. The final claims are independent of the mathematical formalism and define functional principles of functioning: learning leads to copying the inputs, the spontaneous activity replays the inputs and the networks behaves as a predictor of the inputs. The last chapter is nothing more than a proof of concept.

Yet, we did not study the effectiveness of such networks on problems of real life. Can this approach be applied to concrete problems? Can it be useful for traditional data-mining topic, e.g. classification, estimation, prediction? What kind of problems does it best apply to? In other words, does it work?

Appendix A

Notations and definitions

Throughout the thesis we use the standard conventions:

- Bold symbols in upper case like $\mathbf{J}, \mathbf{W}$ are matrices (which may depend on time).
- Bold symbols in lower case like $\mathbf{v}, \mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{j}$ are unidimensional vectors (which may depend on time).
- Normal letters are either real constants or functions.

In particular we use the following symbols

- $l,\ \kappa,\ \tau,\ \varepsilon_1,\ \varepsilon_2,\ \mu,\ \sigma^2,\ \beta,\ \gamma,\ a_\pm \in \mathbb{R}_+$ are parameters of the network. We also define $\Delta = \sqrt{1 - 4\frac{l}{\beta}}$ for section 2.4.3 and $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$, a fixed noise matrix, for section 2.4.4.
- $n \in \mathbb{N}$ is the number of neurons in the network.
- $m \in \mathbb{N}$ is the number of inputs or stimuli to the network in section 2.3.
- $\mathbf{v} \in C^1(\mathbb{R}_+, \mathbb{R}^n)$ is the field of membrane potential in the network.
- $\mathbf{u} \in C^1(\mathbb{R}_+, \mathbb{R}^n)$ is the field of inputs to the network. We write $u_m = \sup_{t \in \mathbb{R}_+} \|\mathbf{u}(t)\|_2$.
- $\mathbf{v} \otimes \mathbf{u} \in C^1(\mathbb{R}_+, \mathbb{R}^{n \times n})$ is the tensor product between $\mathbf{u}$ and $\mathbf{v}$, which simply means $\{\mathbf{u} \otimes \mathbf{v}\}_{ij}(t) = \mathbf{u}_i(t)\mathbf{v}_j(t)$.
- $\mathbf{J} \in C^1(\mathbb{R}_+, \mathbb{R}^{n \times n})$ is the connectivity of the network. Throughout the thesis we assume $\mathbf{J}(\mathbb{R}_-) = 0$.
- $\|\mathbf{u}(t)\|_p$ for $p = 1, 2$ is the L^p norm of $\mathbf{u}(t) \in \mathbb{R}^n$, i.e. $\|\mathbf{u}(t)\|_p = \left(\sum_{i=1}^n |\mathbf{u}_i(t)|^p\right)^{\frac{1}{p}}$. And similarly for the connectivity matrices of $\mathbb{R}^{n \times n}$ with a double sum.
- $|||\mathbf{J}||| = \sup_{\mathbf{x} \in \mathbb{C}^n,\ \|\mathbf{x}\|=1} |\langle \mathbf{x}, \mathbf{J}.\mathbf{x} \rangle| = \max_{i \in \{1,..,n\}} \big\{ |\lambda_i| : \lambda_i \text{ is an eigenvalue of } \mathbf{J} \big\}$.
- $\mathbf{x}.\mathbf{y}' \in \mathbb{R}^{n \times n}$ is the cross-correlation matrix of two compactly supported and differentiable functions from $\mathbb{R}$ to $\mathbb{R}^n$, i.e.

$$\{\mathbf{x}.\mathbf{y}'\}_{ij} = \int_{-\infty}^{+\infty} \mathbf{x}_i(t)\mathbf{y}_j(t)dt$$

- H is the Heaviside function, i.e. $H(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ 1 & \text{if } t > 0 \end{cases}$.

- The real functions

$$\begin{aligned}
&g_{1/\gamma} : t \mapsto \gamma e^{-\gamma t} H(t) \\
&v : t \mapsto \tfrac{l}{\mu\Delta}\Big(e^{-\frac{\beta}{2\mu}(1-\Delta)t} - e^{-\frac{\beta}{2\mu}(1+\Delta)t}\Big) H(t) \\
&w : t \mapsto \tfrac{l}{2\mu\Delta}\Big((1+\Delta)e^{-\frac{\beta}{2\mu}(1-\Delta)t} - (1-\Delta)e^{-\frac{\beta}{2\mu}(1+\Delta)t}\Big) H(t)
\end{aligned}$$

  are integrable on $\mathbb{R}$.

Appendix B

Introduction to temporal averaging theory

Contents

In this section, we present multiscale theoretical results concerning stochastic averaging of periodically forced SDEs (section B.3). This part is a pedagogical introduction to the main theorem of [Wainrib 2011] which we use in section 2.4.1. These results combine ideas from singular perturbations, classical periodic averaging and stochastic averaging principles. Therefore, we recall briefly in section B.1 and in B.2 several basic features of these principles, providing several examples that are closely related to the application developed in section 2.4.

B.1 An elementary example of averaging for periodically forced slow-fast ODE

We present here an example of a slow-fast ODE perturbed by a fast external periodic input. We have chosen this example since it readily illustrates many ideas that will be developed in the following sections. In particular, this example shows how the ratio between the time-scale separation of the system and the time-scale of the input appears as a new crucial parameter.

Example 1. Consider the following linear time-inhomogeneous dynamical system, with $\varepsilon_1, \varepsilon_2 > 0$ two parameters:

$$\begin{aligned} \tfrac{dv}{dt} &= \tfrac{1}{\varepsilon_1}\Big(-v + \sin(\tfrac{t}{\varepsilon_2})\Big) \\ \tfrac{dw}{dt} &= -w + v^2 \end{aligned}$$

This system is particularly handy since one can solve analytically the first ODE, that is:

$$v(t) = \frac{1}{1+\mu^2}\left(\sin(\frac{t}{\varepsilon_2}) - \mu\cos(\frac{t}{\varepsilon_2})\right) + v_0 e^{-\frac{t}{\varepsilon_1}}$$

where we have introduced the *time-scales ratio*

$$\mu := \frac{\varepsilon_1}{\varepsilon_2}$$

In this system, one can distinguish various asymptotic regimes when ε_1 and ε_2 are small, according to the asymptotic value of μ:

- **Regime 1 : Slow input** $\mu = 0$:

First, if $\varepsilon_1 \to 0$ and ε_2 is fixed, then $v(t)$ is close to $\sin(\frac{t}{\varepsilon_2})$, and from *geometric singular perturbation theory* [Fenichel 1979, O'Malley 1991] one can approximate the slow variable w by the solution of

$$\frac{dw}{dt} = -w + (\sin(\frac{t}{\varepsilon_2}))^2$$

Now taking the limit $\varepsilon_2 \to 0$, and applying the classical *averaging principle* [Arnold and Levi 1988] for periodically driven differential equations, one can approximate y by the solution of

$$\frac{dw}{dt} = -w + \frac{1}{2}$$

since $\frac{1}{2\pi}\int_0^{2\pi} \sin(s)^2 ds = \frac{1}{2}$.

- **Regime 2 : Fast input** $\mu = \infty$:

  If $\varepsilon_2 \to 0$ and ε_1 is fixed, then the classical averaging principle implies that v is close to the solution of

  $$\frac{dv}{dt} = -\frac{v}{\varepsilon_1}$$

  so that w can be approximated by

  $$\frac{dw}{dt} = -w + \left(x_0 e^{-t/\varepsilon_1}\right)^2$$

  and when $\varepsilon_1 \to 0$, one does not recover the same asymptotic behavior as in Regime 1.

- **Regime 3 : Time-scale matching** $0 < \mu < \infty$:

  Now consider the intermediate case where ε_1 is asymptotically proportional to ε_2. In this case, v can be approximated on the fast time-scale t/ε_1 by the periodic solution $\tilde{v}_\mu(t) = \frac{1}{1+\mu^2}(\sin(t) - \mu\cos(t))$ of $\frac{dv}{dt} = -v + \sin(\mu t)$. As a consequence, w will be close to the solution of

  $$\frac{dw}{dt} = -w + \frac{1}{2(1+\mu^2)}$$

  since $\frac{1}{2\pi}\int_0^{2\pi} \tilde{v}_\mu(t)^2 dt = \frac{1}{2(1+\mu^2)}$.

Thus, we have seen in this example that

1. the two limits $\varepsilon_1 \to 0$ and $\varepsilon_2 \to 0$ do not commute

2. the ratio μ between the internal time-scale separation ε_1 and the input time-scale ε_2 is a key parameter in the study of slow-fast systems subject to a time-dependent perturbation.

B.2 Stochastic averaging principle

Time-scales separation is a key property to investigate the dynamical behavior of non-linear multi-scale systems, with techniques ranging from averaging principles to geometric singular perturbation theory. This property appears to be also crucial to understand the impact of noise. Instead of carrying a small noise analysis, a multi-scale approach based on the *stochastic averaging principle* [Khas' minskii 1968] can be a powerful tool to unravel subtle interplays between noise properties and non-linearities. More precisely, consider a system of stochastic differential equations (SDEs) in $\mathbb{R}^{p+q}$:

$$\begin{aligned} d\mathbf{v}_t^\varepsilon &= \frac{1}{\varepsilon} F(\mathbf{v}_t^\varepsilon, \mathbf{w}_t^\varepsilon)dt + \frac{1}{\sqrt{\varepsilon}} \mathbf{\Sigma}(\mathbf{v}_t^\varepsilon, \mathbf{w}_t^\varepsilon).dB(t) \\ d\mathbf{w}_t^\varepsilon &= G(\mathbf{v}_t^\varepsilon, \mathbf{w}_t^\varepsilon)dt \end{aligned}$$

with initial conditions $\mathbf{v}^\varepsilon(0) = \mathbf{v}_0,\ \mathbf{w}^\varepsilon(0) = \mathbf{w}_0$, and where $\mathbf{w}^\varepsilon \in \mathbb{R}^q$ is called the slow variable, $\mathbf{v}^\varepsilon \in \mathbb{R}^p$ is the fast variable, with $F, G, \mathbf{\Sigma}$ smooth functions ensuring existence and uniqueness for the solution $(\mathbf{v}^\varepsilon, \mathbf{w}^\varepsilon)$, and $B(t)$ a p-dimensional standard Brownian motion. Time-scale separation in encoded in the small parameter $\varepsilon = (\varepsilon_1, \varepsilon_2)$.

In order to approximate the behavior of $(\mathbf{v}^\varepsilon, \mathbf{w}^\varepsilon)$ for small ε, the idea is to average out the equation for the slow variable with respect to the stationary distribution of the fast one. More precisely, one first assumes that, for each $\mathbf{w} \in \mathbb{R}^q$ fixed, the *frozen* fast SDE:

$$d\mathbf{v}_t = F(\mathbf{v}_t, \mathbf{w})dt + \sigma(\mathbf{v}_t, \mathbf{w}).dB(t)$$

admits a unique invariant measure, denoted $\rho^y(dx)$. Then, one defines the averaged drift vector field $\bar{G}$:

$$\bar{G}(\mathbf{w}) := \int_{\mathbb{R}^m} G(\mathbf{v}, \mathbf{w}) \rho^{\mathbf{w}}(d\mathbf{v}) \tag{B.1}$$

and $\bar{\mathbf{w}}$ the solution of $\frac{d\bar{\mathbf{w}}}{dt} = \bar{G}(\bar{\mathbf{w}})$ with initial condition $\bar{\mathbf{w}}(0) = y_0$. Under some dissipativity assumptions, the stochastic averaging principle [Khas' minskii 1968] states:

Theorem B.2.1. *For any $\delta > 0$ and $T > 0$,*

$$\lim_{\varepsilon \to 0} \mathbf{P} \left[\sup_{t \in [0,T]} ||\mathbf{w}_t^\varepsilon - \bar{\mathbf{w}}_t||^2 > \delta \right] = 0 \tag{B.2}$$

As a consequence, analyzing the behavior of the deterministic solution $\bar{\mathbf{w}}$ can help to understand useful features of the stochastic process $(\mathbf{v}^{\varepsilon}, \mathbf{w}^{\varepsilon})$.

Example 2. In this example we consider a similar system as in Example 1, but with a noise term instead of the periodic perturbation. Namely, we consider $(\mathbf{v}^{\varepsilon}, \mathbf{w}^{\varepsilon})$ the solution of the system of SDEs:

$$\begin{aligned} dv^{\varepsilon} &= -\frac{1}{\varepsilon} v^{\varepsilon} dt + \frac{\sigma}{\sqrt{\varepsilon}} .dB(t) \\ dw^{\varepsilon} &= \left(-w^{\varepsilon} + (v^{\varepsilon})^2\right) dt \end{aligned}$$

with $\varepsilon > 0$ a small parameter and $\sigma > 0$ a positive constant. From Theorem B.2.1, the stochastic slow variable w^{ε} can be approximated in the sense of (B.2) by the deterministic solution $\bar{w}$ of:

$$\frac{dw}{dt} = \int_{v \in \mathbb{R}} (-w + v^2) \rho(dv)$$

where $\rho(dv)$ is the stationary measure of the linear diffusion process:

$$dv = -v dt + \sigma dB(t)$$

that is :

$$\rho(dv) = \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{v^2}{\sigma^2}}$$

Consequently, w^{ε} can be approximated in the limit $\varepsilon \to 0$ by the solution of:

$$\frac{dw}{dt} = -w + \frac{\sigma^2}{2}$$

Applying (B.2) leads to the following result: for any $T > 0$ and $\delta > 0$,

$$\lim_{\varepsilon \to 0} \mathbf{P} \left[\sup_{t \in [0,T]} |w_t^{\varepsilon} - (y_0 - \frac{\sigma^2}{2}) e^{-t} + \frac{\sigma^2}{2}|^2 > \delta \right] = 0$$

Interestingly, the asymptotic behavior of w^{ε} for small ε is characterized by a deterministic trajectory that depends of the strength σ of the noise applied to the system. Thus, the stochastic averaging principle appears particularly interesting to unravel the impact of noise strength on slow-fast systems.

Many other results have been developed since, extending the set-up to the case where the slow variable has a diffusion component or to infinite-dimensional settings for instance, and also refining the convergence study, providing *homogenization* results concerning the limit of $\varepsilon^{-1/2}(\mathbf{w}^\varepsilon - \bar{\mathbf{w}})$ or establishing large deviation principles (see [Kifer 2009] for a recent monograph). However, fewer results are available in the case of non-homogeneous SDEs, that is when the system is perturbed by an external time-dependent signal. This setting is of particular interest in the framework of stochastic learning models and we present the main relevant mathematical results in the following section.

B.3 Stochastic averaging in the non-homogeneous case

Combining ideas of periodic and stochastic averaging introduced previously, we present here theoretical results concerning multiscale SDEs driven by an external time-periodic input. Consider $(\mathbf{v}^\varepsilon, \mathbf{w}^\varepsilon)$ solution of:

$$\begin{array}{rcl} d\mathbf{v}^\varepsilon & = & \frac{1}{\varepsilon_1}\left[F(\mathbf{v}^\varepsilon, \mathbf{w}^\varepsilon, \frac{t}{\varepsilon_2})\right] dt + \frac{1}{\sqrt{\varepsilon_1}}\mathbf{\Sigma}(\mathbf{v}^\varepsilon, \mathbf{w}^\varepsilon).dB(t) \\ d\mathbf{w}^\varepsilon & = & G(\mathbf{v}^\varepsilon, \mathbf{w}^\varepsilon)dt \end{array} \tag{B.3}$$

with $t \to F(\mathbf{v}, \mathbf{w}, t) \in \mathbb{R}^p$ a τ-periodic function and $\varepsilon = (\varepsilon_1, \varepsilon_2) \in \mathbb{R}_+^2$. Parameter ε_1 represents the internal time-scale separation and ε_2 the input time-scale. We consider the case where both ε_1 and ε_2 are small, that is a strong time-scale separation between the fast variable $\mathbf{v}^\varepsilon \in \mathbb{R}^p$ and the slow one $\mathbf{w}^\varepsilon \in \mathbb{R}^q$, and a fast periodic modulation of the fast drift $F(\mathbf{v}, \mathbf{w}, .)$.
We further denote $\mathbf{z} = (\mathbf{v}, \mathbf{w})$.

Definition B.3.1. *We define the asymptotic time-scale ratio*

$$\mu := \lim_{\varepsilon \to 0} \frac{\varepsilon_1}{\varepsilon_2} \tag{B.4}$$

Accordingly, we denote $\overset{\mu}{\lim_{\varepsilon \to 0}}$ the distinguished limit when $\varepsilon_1 \to 0$, $\varepsilon_2 \to 0$ with $\varepsilon_1/\varepsilon_2 \to \mu$.
The following assumption is made to ensure existence and uniqueness of a strong solution to system (B.3).

Assumptions B.3.2. *Existence and uniqueness of a strong solution*
(i) The functions F, G and $\mathbf{\Sigma}$ are locally Lipschitz continuous in the space variable $\mathbf{z}$

(ii) There exists a constant $R > 0$ such that:

$$\sup_{||\mathbf{z}||>R,\ t>0} \frac{<(F(\mathbf{z},t),G(\mathbf{z})),\mathbf{z}>}{||\mathbf{z}||^2} < 0$$

To control the asymptotic behavior of the fast variable, one further assumes:

Assumptions B.3.3. *Asymptotic behavior of the fast process:*

(i) The diffusion matrix $\mathbf{\Sigma}$ is bounded:

$$\exists M_{\mathbf{\Sigma}} > 0 \ s.t \ \forall \mathbf{z}, \ ||\mathbf{\Sigma}(\mathbf{z})|| < M_{\mathbf{\Sigma}}$$

and uniformly non-degenerate:

$$\exists \eta_0 > 0 \ s.t \ \forall \mathbf{v}, \mathbf{z} \ \ < \mathbf{\Sigma}(\mathbf{z})\mathbf{\Sigma}(\mathbf{z})'\mathbf{v}, \mathbf{v} > \geq \eta_0 ||\mathbf{v}||^2$$

(ii) There exists $r_0 < 0$ such that for all $t \geq 0$ and for all $\mathbf{z}, \mathbf{x} \in \mathbb{R}^{p+q}$:

$$< \nabla_{\mathbf{z}} F.\mathbf{x}, \mathbf{x} > \leq r_0 ||\mathbf{x}||^2$$

According to the value of $\mu \in \{0, \mathbb{R}_+^*, \infty\}$, the stochastic averaging principle is based on a description of the asymptotic behavior of various rescaled fast frozen processes. More precisely, under Assumptions B.3.2 and B.3.3, one can deduce that:

- For any fixed $\mathbf{w}_0 \in \mathbb{R}^q$ and $t_0 > 0$ fixed, the law of the rescaled time-homogeneous frozen process:

  $$d\mathbf{v} = F(\mathbf{v}, \mathbf{w}_0, t_0)dt + \mathbf{\Sigma}(\mathbf{v}, \mathbf{w}_0).dB(t)$$

  converges exponentially fast to a unique invariant probability measure denoted by $\rho^{\mathbf{w}_0,t_0}(d\mathbf{v})$.

- For any fixed $\mathbf{w}_0 \in \mathbb{R}^q$, there exists a $\frac{T}{\mu}$-periodic evolution system of measures $\nu_\mu^{\mathbf{w}_0}(t, d\mathbf{v})$, different from $\rho^{\mathbf{w}_0,t}(d\mathbf{v})$ above, such that the law of the rescaled time-inhomogeneous frozen process

  $$d\mathbf{v} = F(\mathbf{v}, \mathbf{w}_0, \mu t)dt + \mathbf{\Sigma}(\mathbf{v}, \mathbf{w}_0).dB(t) \tag{B.5}$$

  converges exponentially fast towards $\nu_\mu^y(t,x)$, uniformly with respect to $\mathbf{w}_0$ (cf. Appendix Theorem B.3.5)

- For any fixed $\mathbf{w}_0 \in \mathbb{R}^q$, the law of the rescaled time-homogeneous frozen process:
$$d\mathbf{v} = \bar{F}(\mathbf{v}, \mathbf{w}_0)dt + \mathbf{\Sigma}(\mathbf{v}, \mathbf{w}_0).dB(t)$$
where $\bar{F}(x,y) := \tau^{-1}\int_0^\tau F(x,y,t)dt$, converges exponentially fast towards a unique invariant probability measure denoted by $\bar{\rho}^{\mathbf{w}_0}(d\mathbf{v})$.

According to the value of $\mu \in \{0, \mathbb{R}_+^*, \infty\}$, we introduce a vector field $\bar{G}_\mu$ which will play a role similar to $\bar{G}$ introduced in eq. (B.1).

Definition B.3.4. *We define $\bar{G}_\mu : \mathbb{R}^q \to \mathbb{R}^q$ as follows. In the **time-scale matching** case, that is when $0 < \mu < \infty$, then*

$$\bar{G}_\mu(\mathbf{w}) := \left(\frac{\tau}{\mu}\right)^{-1} \int_0^{\frac{\tau}{\mu}} \int_{\mathbf{v}\in\mathbf{R}^p} G(\mathbf{v}, \mathbf{w})\nu_\mu^{\mathbf{w}}(t, d\mathbf{v})dt \tag{B.6}$$

In the extremal cases $\mu \in \{0, \infty\}$:

- ***Slow input** $\mu = 0$:*
$$\bar{G}_0(\mathbf{w}) := \tau^{-1}\int_0^\tau \int_{\mathbf{v}\in\mathbf{R}^p} G(\mathbf{v}, \mathbf{w})\rho^{\mathbf{w},t}(d\mathbf{v})dt$$

- ***Fast input** $\mu = \infty$:*
$$\bar{G}_\infty(\mathbf{w}) := \int_{\mathbf{v}\in\mathbf{R}^p} G(\mathbf{v}, \mathbf{w})\bar{\rho}^{\mathbf{w}}(d\mathbf{v})$$

Notation: We may denote the periodic system of measures $\nu_\mu^{\mathbf{w}}(t, d\mathbf{v})$ associated with (B.5) by $\nu_\mu^{\mathbf{w}}[F, \mathbf{\Sigma}](t, d\mathbf{v})$ to emphasize its relationship with F and $\mathbf{\Sigma}$. Accordingly, we may denote $\bar{G}_\mu(\mathbf{w})$ by $\bar{G}_\mu^{[F,\mathbf{\Sigma}]}(\mathbf{w})$.

Preliminaries before the main result Under regularity and dissipativity conditions, [Lorenzi et al. 2010] prove the following result about the asymptotic behavior of the solution of:

$$\begin{aligned} dX_t^{s,x} &= b(X_t^{s,x}, t)dt + \sigma(X_t^{s,x}, t)dW_t, \quad t > s \\ X_s &= x \end{aligned}$$

Theorem B.3.5. *[Lorenzi et al. 2010]*

1. *There exist a unique τ-periodic family of probability measures $\{\mu(s,.), s \in \mathbf{R}\}$, such that:*

$$\int_{x\in\mathbf{R}^p} \mathbf{E}\left[f(X_t^{s,x})\right]\mu(s,dx) = \int_{x\in\mathbf{R}^p} f(x)\mu(t,dx)$$

Such a family is called evolution systems of measures.

2. *Furthermore, under stronger dissipativity condition, the convergence of the law of X to μ is exponentially fast. More precisely, for any $r \in (1,+\infty)$ there exist $M > 0$ and $\omega < 0$, such that for all $\phi \in L^r(\mathbf{R}^p, \mu(s,.))$:*

$$\int_{x\in\mathbf{R}^p} ||\mathbf{E}\left[\phi(X_t^{s,x})\right] - \int_{x'\in\mathbf{R}^p} \phi(x')\mu(t,dx')||^r \mu(s,dx) \leq Me^{\omega(t-s)} \int_{x\in\mathbf{R}^p} ||\phi(x)||^r \mu(t,dx)$$

We are now able to present our main mathematical result. Extending Theorem B.2.1, the following theorem describes the asymptotic behavior of the slow variable $\mathbf{w}^\varepsilon$ when $\varepsilon \to 0$ with $\varepsilon_1/\varepsilon_2 \to \mu$.

Theorem B.3.6. *Let $\mu \in [0,\infty]$. If $\bar{\mathbf{w}}$ is solution of*

$$\frac{d\bar{\mathbf{w}}}{dt} = \bar{G}_\mu(\bar{\mathbf{w}}) \text{ with } \bar{\mathbf{w}}(0) = \mathbf{w}^\varepsilon(0) \tag{B.7}$$

then the following convergence result holds, for all $T > 0$ and $\delta > 0$:

$$\lim_{\varepsilon\to 0}^{\mu} \mathbb{P}\left[\sup_{t\in[0,T]} |\mathbf{w}_t^\varepsilon - \bar{\mathbf{w}}_t|^2 > \delta\right] = 0$$

Proof. This is just an idea of the proof, whose full version can be found in [Wainrib 2011].

- We start by splitting $[0,t]$ as the union of $L_k = [kt/n, (k+1)t/n]$ for $k = 0, ..., n-1$. Within each L_k we define $\hat{x}^\varepsilon$ solution of:

$$d\hat{x}_s^\varepsilon = \frac{1}{\varepsilon} g(\hat{x}_s^\varepsilon, y_{kt/n}, \frac{s}{\varepsilon})ds + \frac{1}{\sqrt{\varepsilon}}\sigma(x_s^\varepsilon, y_{kt/n}^\varepsilon)dW_s$$

  for $kt/n \leq s \leq (k+1)t/n$ and $\hat{x}_{kt/n}^\varepsilon = x_{kt/n}^\varepsilon$.

- We write the difference $y_t^\varepsilon - \bar{y}_t$ as a sum:

$$\begin{aligned} y_t^\varepsilon - \bar{y}_t &= \sum_{k=0}^{n-1} \int_{kt/n}^{(k+1)t/n} \left(f(x_s^\varepsilon, y_s^\varepsilon) - \bar{f}(\bar{y}_s)\right) ds \\ &= \sum_{k=0}^{n-1} (I_{1,k} + I_{2,k}) + \int_0^t \left(f(\hat{x}_s^\varepsilon, y_s^\varepsilon) - f(\hat{x}_s^\varepsilon, \bar{y}_s)\right) ds \end{aligned}$$

with

$$I_{1,k} := \int_{kt/n}^{(k+1)t/n} (f(x_s^\varepsilon, y_s^\varepsilon) - f(\hat{x}_s^\varepsilon, y_s^\varepsilon))\, ds$$
$$I_{2,k} := \int_{kt/n}^{(k+1)t/n} \left(f(\hat{x}_s^\varepsilon, \bar{y}_s) - \bar{f}(\bar{y}_s)\right) ds$$

- Inspired from [Khas' minskii 1968], the idea is to select the value of $n(\varepsilon)$ so that the subintervals size $\Delta(\varepsilon)$ would be:

  1. sufficiently small to be able to approximate x^ε by $\hat{x}^\varepsilon$ during a time $\Delta(\varepsilon)$.
  2. sufficiently large for the mixing to occur.

- Estimate 1.

$$\sup_{s\in[0,t]} \mathbf{E}\left[||x_s^\varepsilon - \hat{x}_s^\varepsilon||^2\right] \le C\left(\frac{1}{\varepsilon^2 n^3} + \frac{1}{\varepsilon n^2}\right) \exp\left[C\left(\frac{1}{\varepsilon^2 n^2} + \frac{1}{\varepsilon n}\right)\right]$$

- Estimate 2. (Consequence of [Lorenzi et al. 2010])

$$\int_{x\in\mathbf{R}^p} \mathbf{E}\left[||\frac{1}{\xi}\int_{t_0}^{t_0+\xi} (f(X_s^{\varepsilon,x,y}, y) - \bar{f}(y))ds||^2\right] \mu_{t_0}(dx) \le M\varepsilon/\xi$$

- Now we are able to select

$$n(\varepsilon) = \frac{1}{\varepsilon \ln(1/\varepsilon)^{1/4}}$$

and conclude the proof.

□

Remark 5.

1. The case $\mu = 0$ (slow input) can be deduced from a combination of Theorem B.2.1 and of the classical averaging principle. More precisely, in this case the limit $\varepsilon_1 \to 0$ is taken *first*, so that from Theorem B.2.1 with fast variable $\mathbf{v}^\varepsilon$ and slow variables $\mathbf{w}$ and t (with the trivial equation $\dot{t} = 1$), $\mathbf{w}^\varepsilon$ is close in probability on finite time-intervals to the solution of the following inhomogeneous ODE:

$$\frac{d\tilde{\mathbf{w}}}{dt} = \int_{\mathbf{v}\in\mathbb{R}^p} G(\mathbf{v}, \tilde{\mathbf{w}})\rho^{\tilde{\mathbf{w}},t/\varepsilon_2}(d\mathbf{v}) := \tilde{G}(\tilde{\mathbf{w}}, t/\varepsilon_2)$$

Then taking the limit $\varepsilon_2 \to 0$, one can apply the deterministic averaging principle to the fast periodic vector field $\tilde{G}(\mathbf{w}, t/\varepsilon_2)$, so that $\tilde{w}$ converges when $\varepsilon_2 \to 0$ to the solution of:

$$\frac{d\bar{\mathbf{w}}}{dt} = \tau^{-1} \int_0^\tau \tilde{G}(\mathbf{v}, \bar{\mathbf{w}})dt = \bar{G}_0(\bar{\mathbf{w}})$$

which is precisely the statement of Theorem B.3.6.

2. The case $\mu = \infty$ (fast input) can again be deduced from the same tools, but in the reverse order. As the limit $\varepsilon_2 \to 0$ is taken first, one has to perform first a classical averaging of the periodic drift $F(\mathbf{v}, \mathbf{w}, t/\varepsilon_2)$, leading to the homogeneous system of SDEs (B.3) but whith $\bar{F}(\mathbf{v}, \mathbf{w})$ instead of $F(\mathbf{v}, \mathbf{w}, t/\varepsilon_2)$. Then, an application of Theorem B.2.1 on this system produces exactly the statement of Theorem B.3.6.

3. The case $0 < \mu < \infty$ is more complicated in the sense that it combines simultaneously both the periodic and stochastic averaging principles. A particular role is played by the frozen periodically forced SDE (B.5). The equivalent of the quasistationary measure $\rho^{\mathbf{w}}$ of Theorem B.2.1 is given by the asymptotically periodic behavior of Eq. (B.5), represented by the periodic family of measures $\nu_\mu^{\mathbf{w}}(t, d\mathbf{v})$.

4. By a rescaling of the frozen process (B.5), one deduces the following *scaling* relationships :

$$\nu_\mu^{\mathbf{w}}[F, \mathbf{\Sigma}](t, d\mathbf{v}) = \nu_1^{\mathbf{w}}[\frac{F}{\mu}, \frac{\mathbf{\Sigma}}{\sqrt{\mu}}](\mu t, d\mathbf{v})$$

and

$$\bar{G}_\mu^{[F,\mathbf{\Sigma}]}(\mathbf{w}) = \bar{G}_1^{[\frac{F}{\mu}, \frac{\mathbf{\Sigma}}{\sqrt{\mu}}]}(\mathbf{w})$$

Therefore, if one knows in the case $\mu = 1$ the averaged vector field associated with the fast process generated by a drift F and a diffusion coefficient σ, denoted $\bar{G}_1[F, \mathbf{\Sigma}]$, it is possible to deduce $\bar{G}_\mu$ in the general case $\mu \in (0, \infty)$ with a change $F \to \mu F$ and $\mathbf{\Sigma} \to \sqrt{\mu}\mathbf{\Sigma}$.

B.4 Case of a fast linear SDE with periodic input

We present here an elementary case where one can compute explicitly the quasi-stationnary time-periodic family of measures $\nu_\mu^y(t, x)$, when the equation for the fast variable is linear. Namely, we consider $\mathbf{v} \in \mathbb{R}^n$ solution of:

$$d\mathbf{v}(t) = (-\mathbf{A}.\mathbf{v}(t) + \mathbf{u}(\mu t))\, dt + \mathbf{\Sigma}.dB(t)$$

with initial condition $\mathbf{v}(0) = \mathbf{v}_0 \in \mathbb{R}^n$, and where $\mathbf{A}$ is a $n \times n$ positive definite matrix and $\mathbf{u}(.)$ a τ-periodic function.

We are interested in the large time behavior of the law of $\mathbf{v}(t)$, which is a time-inhomogeneous Ornstein-Uhlenbeck process. From [Lorenzi et al. 2010] we know that its law converges to a τ-periodic family of probability measures $\nu(t, d\mathbf{v})$. Due to the linearity in the previous equation, $\nu(t, d\mathbf{v})$ is Gaussian with time-dependent mean and constant covariance matrix:

$$\nu(t, d\mathbf{v}) = \mathcal{N}_{\mathbf{x}(t),\mathbf{Q}}(d\mathbf{v})$$

where $\mathbf{x}$ is the $\frac{\tau}{\mu}$-periodic attractor of $\frac{d\mathbf{x}}{dt} = -\mathbf{A}.\mathbf{x}(t) + \mathbf{u}(\mu t)$, i.e.

$$\mathbf{x}(t) = \int_{-\infty}^{t} e^{-A(t-s)}\mathbf{u}(\mu s)ds$$

and $\mathbf{Q}$ is the unique solution of:

$$\mathbf{A}.\mathbf{Q} + \mathbf{Q}.\mathbf{A}' + \mathbf{\Sigma}.\mathbf{\Sigma}' = 0 \tag{B.8}$$

Indeed, if one denotes $\mathbf{c}(t) = \mathbf{v}(t) - \mathbf{x}(t)$, then $\mathbf{c}(t)$ is solution of a classical homogeneous Ornstein-Uhlenbeck equation:

$$d\mathbf{c}(t) = -\mathbf{A}\mathbf{c}(t)dt + \mathbf{\Sigma}.dB(t)$$

whose stationary distribution is known [Risken 1996] to be a centered Gaussian measure with covariance matrix $\mathbf{Q}$ solution of (B.8). Notice that if $\mathbf{A}$ is self-adjoint with respect to $\mathbf{D}^{-1} = (\mathbf{\Sigma}.\mathbf{\Sigma}')^{-1}$ (i.e. $\mathbf{A}.\mathbf{D} = \mathbf{D}.\mathbf{A}'$) , then the solution is $\mathbf{Q} = \frac{\mathbf{A}^{-1}.\mathbf{D}}{2} = \frac{\mathbf{D}.\mathbf{A}'^{-1}}{2}$, which will be used in section C.2.2.

Hence, in the linear case, the averaged vector field of equation (B.6) becomes

$$\bar{G}_\mu(y) := \left(\frac{\tau}{\mu}\right)^{-1} \int_0^{\frac{\tau}{\mu}} \int_{\mathbf{v}\in\mathbb{R}^n} G(\mathbf{v}(t) + \mathbf{v}, y)\varphi_{0,\mathbf{Q}}(d\mathbf{v})dt \tag{B.9}$$

where $\varphi_{\mathbf{x},\mathbf{Q}}$ is the probability density function of the Gaussian law with mean $\mathbf{x} \in \mathbb{R}^q$ and covariance $\mathbf{Q} \in \mathbb{R}^{q\times q}$.

Therefore, due the linearity of the fast SDE, the periodic system of measure ν is just a constant Gaussian distribution shifted by a periodic function of time

$\mathbf{v}(t)$. In case G is quadratic in $\mathbf{v}$, this remark implies that one can perform independently the integral over time and over $\mathbb{R}^n$ in formula B.9 (noting that the crossed term has a zero average). In this case, contributions from the periodic input and from the noise appear in the averaged vector field in an additive way.

Example 3. In this last example, we consider a combination between Example 1 and Example 2, namely we consider the following system of periodically forced SDEs:

$$\begin{aligned} dv^\varepsilon &= \frac{1}{\varepsilon_1}\left[-v^\varepsilon + \sin\left(\frac{t}{\varepsilon_2}\right)\right] dt + \frac{\sigma}{\sqrt{\varepsilon_1}}.dB(t) \\ dw^\varepsilon &= \left(-w^\varepsilon + (v^\varepsilon)^2\right) dt \end{aligned}$$

As in Example 1 and as shown above, the behavior of this system when both ε_1 and ε_2 are small depends on the parameter μ defined in (B.4). More precisely, applying Theorem B.3.6, we have the following three regimes:

- **Regime 1 : slow input** $\mu = 0$:

$$\bar{G}_0(y) = -y + \frac{\sigma^2}{2} + \frac{1}{2}$$

- **Regime 2 : fast input** $\mu = \infty$:

$$\bar{G}_\infty(y) = -y + \frac{\sigma^2}{2}$$

- **Regime 3 : time-scale matching** $0 < \mu < \infty$:

$$\bar{G}_\mu(y) = -y + \frac{\sigma^2}{2} + \frac{1}{2(1+\mu^2)}$$

B.5 Asymptotic well-posedness

In some cases, assumptions B.3.2-B.3.3 may not be satisfied on the entire phase space $\mathbb{R}^p \times \mathbb{R}^q$, but only on a subset. Such situations when considering learning models. We introduce here a more refined set of assumptions ensuring that theorem B.3.6 still applies.

Let us start with an example, namely the following bidimensional system with white noise input:

$$\begin{cases} dv^\varepsilon &= \frac{1}{\varepsilon}\Big(-lv^\varepsilon + w^\varepsilon v^\varepsilon \Big) dt + \frac{\sigma}{\sqrt{\varepsilon}}.dB(t) \\ dw^\varepsilon &= \Big(-\mu w^\varepsilon + (v^\varepsilon)^2 \Big) dt \end{cases} \tag{B.10}$$

with $\varepsilon > 0,\ \sigma > 0,\ l > 0,\ \mu > 0$.

For the fast drift $-(l-w)v$ to be non-explosive, it is necessary to have $w < l-\alpha$ with $\alpha > 0$ for all time. The concern about this system comes from the fact that the slow variable w may reach l due to the fluctuations captured in the term v^2, for instance if μ is not large enough. Such a system may have exponentially growing trajectories. However, we claim that for small enough ε, w^ε will remain close to its averaged limit $\bar{w}$ for a very long time, and if this limit remains below $l-\alpha$, then w^ε can be considered as well-posed in the asymptotic limit $\varepsilon \to 0$. To make this argument more rigorous, we suggest the following definition:

Definition B.5.1. *A stochastic differential equation with a given initial condition is asymptotically well-posed in probability if for the given initial condition:*

1. *a unique solution exists until a time τ_ε*

2. *for all $T > 0$,*
$$\lim_{\varepsilon \to 0} \mathbb{P}\left[\tau_\varepsilon \geq T\right] = 1$$

We give in the following proposition sufficient conditions for system (B.3) to be asymptotically well-posed in probability and to satisfy conclusions of Theorem B.3.6.

Proposition B.5.2. *If there exists a subset $\mathcal{E}$ of $\mathbb{R}^q$ such that:*

1. *The functions $F, G, \mathbf{\Sigma}$ satisfy Assumptions B.3.2-B.3.3 restricted on $\mathbb{R}^p \times \mathcal{E}$.*

2. *$\mathcal{E}$ is invariant under the flow of $\bar{G}_\mu$, as defined in* (B.6)

Then for any initial condition $\mathbf{w}_0 \in \mathcal{E}$ system (B.3) *is asymptotically well-posed in probability and $\mathbf{w}^\varepsilon$ satisfies the conclusion of Theorem B.3.6.* The proof of prop. B.5.2 can be found in appendix B.6.

Here, we show that it applies to system (B.10). First, with $\mathcal{E}_\alpha = \{w \in \mathbb{R},\ w < l - \alpha\}$, for some $\alpha \in]0, l[$, it is possible to show that assumptions B.3.2-B.3.3 are satisfied on $\mathbb{R}^p \times \mathcal{E}_\alpha$. Then as a special case of (B.9), we obtain the following averaged system:

$$\frac{d\bar{w}}{dt} = -\mu\bar{w} + \frac{\sigma^2}{2(l-\bar{w})} := \bar{G}(\bar{w})$$

It remains to check that the solution of this system satisfies:

$$\exists \alpha > 0,\ s.t\ \bar{w}(0) < l - \alpha \Rightarrow \forall t > 0,\ \bar{w}(t) < l - \alpha$$

that is the subset $\mathcal{E}_\alpha$ is invariant under the flow of $\bar{G}$.
This property is satisfied as soon as

$$\eta := \frac{2\sigma^2}{\mu l^2} < 1$$

Indeed, one can show that $\bar{G}(w) = 0$ admits two solutions iff $\eta < 1$:

$$w_\pm = \frac{l}{2}(1 \pm \sqrt{1-\eta}) \in (0, l)$$

and that w_- is stable whereas w_+ is unstable. Thus, if $\bar{w}(0) < l - \alpha$ with $\alpha = l - w_+ > 0$, then $\bar{w}(t) < l - \alpha$ for all $t > 0$. In fact, the invariance property is true for all $\alpha \in]l - w_-, l - w_+[$.

B.6 Proof of asymptotical well-posedness, proposition B.5.2

Recall prop. B.5.2

Proposition B.6.1. *If there exists a subset $\mathcal{E}$ of $\mathbb{R}^q$ such that:*

1. *The functions $f, g, \mathbf{\Sigma}$ satisfy Assumptions B.3.2-B.3.3 restricted on $\mathbb{R}^p \times \mathcal{E}$.*

2. *$\mathcal{E}$ is invariant under the flow of $\bar{f}_\mu$, as defined in* (B.6)

Then, for any initial condition $\mathbf{w}_0 \in \mathcal{E}$, system (B.3) *is asymptotically well-posed in probability and $\mathbf{w}^\varepsilon$ satisfies the conclusion of Theorem B.3.6.*

Proof. First we introduce $(\tilde{\mathbf{v}}^{\varepsilon,\beta}, \tilde{\mathbf{w}}^{\varepsilon,\beta})$ solution of the auxiliary system:

$$\begin{aligned} d\mathbf{v} &= \frac{1}{\varepsilon_1}\left[g(\mathbf{v},\mathbf{w},\frac{t}{\varepsilon_2})\right]dt + \frac{1}{\sqrt{\varepsilon_1}}\mathbf{\Sigma}(\mathbf{v},\mathbf{w}).dB(t) \\ d\mathbf{w}^{\varepsilon} &= \tilde{f}_\beta(\mathbf{v},\mathbf{w})dt \end{aligned}$$

with the same initial condition as $(\mathbf{v}^{\varepsilon}, \mathbf{w}^{\varepsilon})$ and where:

$$\begin{aligned} \tilde{f}_\beta(\mathbf{v},\mathbf{w}) &= f(\mathbf{v},\mathbf{w})\psi_\beta(\mathbf{w}) \\ \psi_\beta(\mathbf{w}) &= e^{-\beta d(\mathbf{w},\partial\mathcal{E})} \end{aligned}$$

with $d(\mathbf{w},\partial\mathcal{E})$ the distance between $\mathbf{w}$ and $\mathcal{E}$, taken negative if $\mathbf{w}\in\mathcal{E}$ and positive otherwise.

Let $\eta > 0$. We claim that one can choose $\beta = \beta_0 > 0$ sufficiently large such that:

$$\text{For all } T,\delta,\varepsilon_1,\varepsilon_2 > 0,\ \mathbb{P}\left[\sup_{t\in[0,T]}|\hat{\mathbf{w}}_t^{\varepsilon} - \tilde{\mathbf{w}}_t^{\varepsilon,\beta_0}| > \delta\right] < \eta/2$$

where $\hat{\mathbf{w}}_t = w_{t\wedge\tau_\varepsilon}$ and

$$\tau_\varepsilon := \inf\{t \geq 0;\ \mathbf{w}_t^{\varepsilon} \notin \mathcal{E}\}$$

As assumptions B.3.2-B.3.3 are satisfied for the auxiliary system $(\tilde{\mathbf{v}}^{\varepsilon,\beta}, \tilde{\mathbf{w}}^{\varepsilon,\beta})$ (by assumption 1.), one can apply Theorem B.3.6: for all $\delta, T > 0$,

$$\overset{\mu}{\lim_{\varepsilon\to 0}}\mathbb{P}\left[\sup_{t\in[0,T]}|\tilde{\mathbf{w}}_t^{\varepsilon,\beta_0} - \bar{\mathbf{w}}_t| > \delta\right] = 0$$

where $\bar{\mathbf{w}}$ is defined by (B.7). As a consequence, there exists ε_0 such that for all $\varepsilon < \varepsilon_0$:

$$\mathbb{P}\left[\sup_{t\in[0,T]}|\tilde{\mathbf{w}}_t^{\varepsilon,\beta_0} - \bar{\mathbf{w}}_t| > \delta\right] < \eta/2$$

Then, as $|\hat{\mathbf{w}}_t^{\varepsilon} - \bar{\mathbf{w}}_t| \leq |\hat{\mathbf{w}}_t^{\varepsilon} - \tilde{\mathbf{w}}_t^{\varepsilon,\beta_0}| + |\tilde{\mathbf{w}}_t^{\varepsilon,\beta_0} - \bar{\mathbf{w}}_t|$, one deduces that for all $\varepsilon < \varepsilon_0$:

$$\mathbb{P}\left[\sup_{t\in[0,T]}|\hat{\mathbf{w}}_t^{\varepsilon} - \bar{\mathbf{w}}_t| > \delta\right] < \eta/2$$

that is to say:

$$\overset{\mu}{\lim_{\varepsilon\to 0}}\mathbb{P}\left[\sup_{t\in[0,T]}|\hat{\mathbf{w}}_t^{\varepsilon} - \bar{\mathbf{w}}_t| > \delta\right] = 0$$

We know by assumption 2., for all $t \geq 0$, $\bar{\mathbf{w}}_t \in \mathcal{E}$, so we conclude the proof by observing that for all $T > 0$,

$$\lim_{\varepsilon \to 0} \mathbb{P}\left[\tau_\varepsilon \geq T\right] = 1$$

□

Appendix C
Proofs of chapter 2

Contents

C.1 Proofs for slow inputs section 2.3

C.1.1 Proof of theorem 2.3.3

Consider the following Lyapunov function (see equation (2.18))

$$E(\mathbf{X},\mathbf{J}) = -\frac{1}{2}\langle \mathbf{X}, \mathbf{J}\cdot\mathbf{X}\rangle - \langle \mathbf{U},\mathbf{X}\rangle + \langle 1, \overline{S^{-1}}(\mathbf{X})\rangle + \frac{\tilde{\kappa}}{2}\|\mathbf{J}\|^2, \tag{C.1}$$

where $\tilde{\kappa} = \kappa m$, such that if $\mathbf{J} = \mathbf{J}_S + \mathbf{J}_A$, where $\mathbf{J}_S$ is symmetric and $\mathbf{J}_A$ is anti-symmetric.

$$-\nabla E(\mathbf{X},\mathbf{J}) = \begin{pmatrix} \mathbf{J}_S \cdot \mathbf{X} + \mathbf{U} - S^{-1}(\mathbf{X}) \\ \mathbf{X}\cdot\mathbf{X}' - \kappa\mathbf{J} \end{pmatrix} \tag{C.2}$$

Therefore, writing the system Σ', equation (2.17), as

$$\frac{d\mathcal{Y}}{dt} = \gamma \begin{pmatrix} \mathbf{J}_S \cdot S(\mathbf{V}) + \mathbf{U} - S^{-1}\big(S(\mathbf{V})\big) \\ S(\mathbf{V})\cdot S(\mathbf{V})' - \tilde{\kappa}\mathbf{J} \end{pmatrix} + \gamma \begin{pmatrix} \mathbf{J}_A . S(\mathbf{V}) \\ 0 \end{pmatrix},$$

where $\mathcal{Y} = (\mathbf{V},\mathbf{J})'$, we see that

$$\frac{d\mathcal{Y}}{dt} = -\gamma\Big(\nabla E\big(\sigma(\mathbf{V},\mathbf{J})\big)\Big) + \Gamma(t) \tag{C.3}$$

where $\gamma(\mathbf{V},\mathbf{J})' = (\mathbf{V}, \varepsilon\mathbf{J}/m)'$, $\sigma(\mathbf{V},\mathbf{J}) = (S(\mathbf{V}),\mathbf{J})$ and $\Gamma : \mathbb{R}_+ \to \mathcal{H}$ such that $\|\Gamma\| \underset{t\to+\infty}{\longrightarrow} 0$ exponentially (because the system converges to $\mathcal{A}$). It follows that the time derivative of $\tilde{E} = E\circ\sigma$ along trajectories is given by:

$$\frac{d\tilde{E}}{dt} = \left\langle \nabla\tilde{E}, \frac{d\mathcal{Y}}{dt}\right\rangle = \left\langle \nabla_{\mathbf{V}}\tilde{E}, \frac{d\mathbf{V}}{dt}\right\rangle + \left\langle \nabla_{\mathbf{J}}\tilde{E}, \frac{d\mathbf{J}}{dt}\right\rangle. \tag{C.4}$$

Substituting equation (C.3) then yields

$$\begin{aligned} \frac{d\tilde{E}}{dt} &= -\Big\langle \nabla\tilde{E}, \gamma\big(\nabla E\circ\sigma\big)\Big\rangle + \underbrace{\Big\langle \nabla\tilde{E}, \Gamma(t)\Big\rangle}_{\tilde{\Gamma}(t)} \qquad (C.5) \\ &= -\Big\langle S'(\mathbf{V})\nabla_{\mathbf{X}}E\circ\sigma, \nabla_{\mathbf{X}}E\circ\sigma\Big\rangle - \frac{\varepsilon}{m}\Big\langle \nabla_{\mathbf{J}}E\circ\sigma, \nabla_{\mathbf{J}}E\circ\sigma\Big\rangle + \tilde{\Gamma}(t). \end{aligned}$$

We have used the chain–rule of differentiation, whereby

$$\nabla_{\mathbf{V}}(\tilde{E}) = \nabla_{\mathbf{V}}(E\circ\sigma) = S'(\mathbf{V})\nabla_{\mathbf{X}}E\circ\sigma,$$

and $S'(\mathbf{V})\nabla_{\mathbf{X}}E$ (without dots) denotes the Hadamard (term by term) product, that is,

$$[S'(\mathbf{V})\nabla_{\mathbf{X}}E]_{ia} = S'(\mathbf{V}_i^{(a)})\frac{\partial E}{\partial \mathbf{X}_i^{(a)}}$$

Note that $|\tilde{\Gamma}| \underset{t\to+\infty}{\to} 0$ exponentially because $\nabla\tilde{E}$ is bounded, and $S'(\mathbf{V}) > 0$ because the trajectories are bounded. Thus, there exists $t_1 \in \mathbb{R}_+$ such that $\forall t > t_1$, $\exists k \in \mathbb{R}_+^*$ such that

$$\frac{d\tilde{E}}{dt} \leq -k\|\nabla E \circ \sigma\|^2 \leq 0. \tag{C.6}$$

As in [Cohen and Grossberg 1983] and [Dong and Hopfield 1992], we apply the Krasovskii-LaSalle invariance principle detailed in [Khalil and Grizzle 1996]. We check that:

- $\tilde{E}$ is lower bounded. Indeed, $\mathbf{V}$ and $\mathbf{J}$ are bounded. Given that $\mathbf{U}$ and S are also bounded it is clear that $\tilde{E}$ is bounded.
- $\frac{d\tilde{E}}{dt}$ is negative semidefinite on the trajectories as shown in equation (C.6).

Then the invariance principle tells us that the solutions of the system Σ' approach the set $M = \Big\{\mathcal{Y} \in \mathcal{H} : \frac{d\tilde{E}}{dt}(\mathcal{Y}) = 0\Big\}$. Equation (C.6) implies that $M = \Big\{Y \in \mathcal{H} : \nabla E \circ \sigma = 0\Big\}$. Since $\frac{d\mathcal{Y}}{dt} = -\gamma\Big(\nabla E \circ \sigma\Big)$ and $\gamma \neq 0$ everywhere, M consists of the equilibrium points of the system. This completes the proof.

C.1.2 Proof of theorem 2.3.4

Denote the right–hand side of system Σ', equation (2.17) by

$$F(\mathbf{V},\mathbf{J}) = \left\{ \begin{array}{c} -\mathbf{V} + \mathbf{J}\cdot S\big(\mathbf{V}\big) + \mathbf{U} \\ \frac{\varepsilon}{m}\big(S(\mathbf{V}).S(\mathbf{V})' - \kappa m \mathbf{J}\big) \end{array} \right.$$

The fixed points satisfy the condition $F(\mathbf{V},\mathbf{J}) = 0$ which immediately leads to equations (2.22). Let us now check the linear stability of this system. The differential of F at $\mathbf{V}^*, \mathbf{J}^*$ is

$$dF_{(\mathbf{V}^*,\mathbf{J}^*)}(\mathbf{X},\mathbf{W}) = \left(\begin{array}{c} -\mathbf{X} + \mathbf{J}^*\cdot\big(S'(\mathbf{V}^*)\mathbf{X}\big) + \mathbf{W}\cdot S(\mathbf{V}^*) \\ \frac{\varepsilon}{m}\Big(\big(S'(\mathbf{V}^*)\mathbf{X}\big)\cdot S(\mathbf{V}^*)' + S(\mathbf{V}^*)\cdot\big(S'(\mathbf{V}^*)\mathbf{X}\big)' - \kappa m \mathbf{W}\Big), \end{array} \right)$$

where $S'(\mathbf{V}^*)\mathbf{X}$ denotes a Hadamard product, that is, $[S'(\mathbf{V}^*)\mathbf{X}]_{ia} = S'(V_i^{*(a)})\mathbf{X}_i^{(a)}$. Assume that there exist $\lambda \in \mathbb{C}^*$, $(\mathbf{X}, \mathbf{W}) \in \mathcal{H}$ such that $dF_{(V^*,\mathbf{J}^*)}\begin{pmatrix}\mathbf{X}\\ \mathbf{W}\end{pmatrix} = \lambda\begin{pmatrix}\mathbf{X}\\ \mathbf{W}\end{pmatrix}$. Taking the second component of this equation and computing the dot product with $S(\mathbf{V}^*)$ leads to

$$(\lambda + \varepsilon\kappa)\mathbf{W}\cdot S = \frac{\varepsilon}{m}\left((S'\mathbf{X})\cdot S'\cdot S + S\cdot(S'\mathbf{X})'\cdot S\right)$$

where $\mathbf{S} = S(\mathbf{V}^*)$, $\mathbf{S}' = S'(\mathbf{V}^*)$ (and therefore use the explicit notation $.^T$ for the transpose operator in the rest of this section). Substituting this expression in the first equation leads to

$$m(\lambda+\varepsilon\kappa)(\lambda+1)\mathbf{X} = (\frac{\lambda}{\kappa}+\varepsilon)\mathbf{S}\cdot\mathbf{S}^T\cdot(\mathbf{S}'\mathbf{X})+\varepsilon(\mathbf{S}'\mathbf{X})\cdot\mathbf{S}^T\cdot\mathbf{S}+\varepsilon\mathbf{S}\cdot(\mathbf{S}'\mathbf{X})^T\cdot\mathbf{S} \tag{C.7}$$

Observe that setting $\varepsilon = 0$ in the previous equation leads to an eigenvalue equation for the membrane potential only:

$$(\lambda+1)\mathbf{X} = \frac{1}{\kappa m}\mathbf{S}\cdot\mathbf{S}^T\cdot(\mathbf{S}'\mathbf{X}).$$

Since $\mathbf{J}^* = \frac{1}{\kappa m}\big(\mathbf{S}\cdot\mathbf{S}^T\big)$, this equation implies that $\lambda+1$ is an eigenvalue of the operator $\mathbf{X} \mapsto \mathbf{J}^*.(\mathbf{S}'\mathbf{X})$. The magnitudes of the eigenvalues are always smaller than the norm of the operator. Therefore, we can say that if $1 > \|\mathbf{J}^*\|s'_m$ then all the possible eigenvalues λ must have a negative real part. This sufficient condition for stability is the same as in [Faugeras et al. 2008]. It says that fixed points sufficiently close to the origin are always stable.

Let us now consider the case $\varepsilon \neq 0$. Recall that $\mathbf{X}$ is a matrix. We now "flatten" $\mathbf{X}$ by storing its rows in a vector called $\mathbf{X}_{row}$. We use the following result in [Brewer 1978]: the matrix notation of operator $\mathbf{X} \mapsto \mathbf{A}\cdot\mathbf{X}\cdot\mathbf{B}$ is $\mathbf{A}\otimes\mathbf{B}^T$, where $\otimes$ is the Kronecker product. In this formalism the previous equation becomes

$$m(\lambda+\varepsilon\kappa)(\lambda+l)\mathbf{X}_{row} = \left((\frac{\lambda}{\kappa}+\varepsilon)\mathbf{S}\cdot\mathbf{S}^T\otimes I_d+\varepsilon I_d\otimes\mathbf{S}^T\cdot\mathbf{S}+\varepsilon\mathbf{S}\otimes\mathbf{S}^T\right)\cdot(\mathbf{S}'\mathbf{X})_{\text{row}} \tag{C.8}$$

where we assume that the Kronecker product has the priority over the dot product. We focus on the linear operator $\mathcal{O}$ defined by the right hand side and bound its norm. Note that we use the following norm $\|\mathbf{J}\|_\infty = \sup_\mathbf{X} \frac{\|\mathbf{J}.\mathbf{X}\|}{\|\mathbf{X}\|}$ which is equal to the largest magnitude of the eigenvalues of $\mathbf{J}$.

$$\|\mathcal{O}\|_\infty \leq s'_m\Bigg(|\frac{\lambda}{\kappa}|\|\mathbf{S}\cdot\mathbf{S}^T\otimes I_d\|_\infty + \varepsilon\|\mathbf{S}\cdot\mathbf{S}^T\otimes I_d\|_\infty + \varepsilon\|I_d\otimes\mathbf{S}^T\cdot\mathbf{S}\|_\infty + \varepsilon\|\mathbf{S}\otimes\mathbf{S}^T\|_\infty\Bigg). \tag{C.9}$$

Define, ν_m to be the magnitude of the largest eigenvalue of $\mathbf{J}^* = \frac{1}{\kappa m}(\mathbf{S}\cdot\mathbf{S}^T)$. First, note that $\mathbf{S}\cdot\mathbf{S}^T$ and $\mathbf{S}^T\cdot\mathbf{S}$ have the same eigenvalues $(\kappa m)\nu_i$ but different eigenvectors denoted by $\mathbf{u}_i$ for $\mathbf{S}\cdot\mathbf{S}^T$ and $\mathbf{v}_i$ for $\mathbf{S}^T\cdot\mathbf{S}$. In the basis set spanned by the $\mathbf{u}_i\otimes\mathbf{v}_j$, we find that $\mathbf{S}\cdot\mathbf{S}^T\otimes I_d$ and $I_d\otimes\mathbf{S}^T\cdot\mathbf{S}$ are diagonal with $(\kappa m)\nu_i$ as eigenvalues. Therefore, $\|\mathbf{S}\cdot\mathbf{S}^T\otimes I_d\|_\infty = (\kappa m)\nu_m$ and $\|I_d\otimes\mathbf{S}^T\cdot\mathbf{S}\|_\infty = (\kappa m)\nu_m$. Moreover, observe that

$$(\mathbf{S}^T\otimes\mathbf{S})^T\cdot(\mathbf{S}^T\otimes\mathbf{S})\cdot(\mathbf{u}_i\otimes\mathbf{v}_j) = (\mathbf{S}\cdot\mathbf{S}^T\cdot\mu_i)\otimes(\mathbf{S}^T\cdot\mathbf{S}\cdot\mathbf{v}_j) = (\kappa m)^2\nu_i\nu_j\ \mathbf{u}_i\otimes\mathbf{v}_j \tag{C.10}$$

Therefore, $(\mathbf{S}^T\otimes\mathbf{S})^T\cdot(\mathbf{S}^T\otimes\mathbf{S}) = (\kappa m)^2\mathrm{diag}(\nu_i\nu_j)$. In other words, $\mathbf{S}^T\otimes\mathbf{S}$ is the composition of an orthogonal operator (i.e. an isometry) and a diagonal matrix. Immediately, it follows that $\|\mathbf{S}^T\otimes\mathbf{S}\| \leq (\kappa m)\nu_m$.

Compute the norm of equation (C.8)

$$|(\lambda+\varepsilon\kappa)(\lambda+1)| \leq s'_m(|\lambda|+3\varepsilon\kappa)\nu_m. \tag{C.11}$$

Define $f_\varepsilon : \mathbb{C}\to\mathbb{R}$ such that $f_\varepsilon(\lambda) = |(\lambda+\varepsilon\kappa)||(\lambda+1)| - (|\lambda|+3\varepsilon\kappa)s'_m\nu_m$. We want to find a condition such that $f_\varepsilon(\mathbb{C}_+) > 0$, where $\mathbb{C}_+$ is the right half complex plane. This condition on ε, κ, ν_m, and s'_m will be a sufficient condition for linear stability. Indeed, under this condition we can show that only eigenvalues with a negative real part can meet the necessary condition (C.11). Complex number of the right half plane cannot be eigenvalues and thus the system is stable. The case $\varepsilon = 0$ tells us that $f_0(\mathbb{C}_+) > 0$ if $1 > s'_m\nu_m$, compute

$$\frac{\partial f_\varepsilon}{\partial\varepsilon}(\lambda) = \kappa(\Re(\lambda)+\kappa\varepsilon)\frac{|(\lambda+1)|}{|(\lambda+\varepsilon\kappa)|} - 3\kappa s'_m\nu_m$$

If $1 \geq \varepsilon\kappa$, which is most probably true given that $\varepsilon << 1$, then $\frac{|(\lambda+1)|}{|(\lambda+\varepsilon\kappa)|} \geq 1$. Assuming $\lambda\in\mathbb{C}_+$ leads to:

$$\frac{\partial f_\varepsilon}{\partial\varepsilon}(\lambda) \geq \kappa(\kappa\varepsilon - 3s'_m\nu_m) \geq \kappa(1-3s'_m\nu_m)$$

Therefore, the condition $3s'_m\nu_m < 1$, which implies $s'_m\nu_m < 1$, and leads to $f_\varepsilon(\mathbb{C}_+) > 0$.

C.2 Proofs for fast inputs section 2.4

C.2.1 Special notations

The computations involve a lot of convolutions and, for readability of the appendix, we introduce some new notations. Indeed, we rewrite the time-convolution between $\mathbf{u}$ and g a integrable function on $\mathbb{R}$:

$$\mathbf{u} * g = \mathbf{u}.\mathcal{G}$$

This suggests one should think of $\mathbf{v}$ as a semi-continuous matrix of $\mathbb{R}^{n\times\mathbb{R}}$ and $\mathcal{G}_{1/\gamma}$ as a continuous matrix of $\mathbb{R}^{\mathbb{R}\times\mathbb{R}}$, such that $\mathbf{u}_{it} = u_i(t)$ and $\mathcal{G}_{st} = g(t-s)$. Indeed, in this framework the convolution with g is nothing but the continuous matrix multiplication between $\mathbf{v}$ and a continuous Toeplitz matrix generated row by row by g. Hence, the operator ".“ can be though of as a matrix multiplication.

Therefore, it is natural to define $(\mathbf{u}*g)' = (\mathbf{u}.\mathcal{G})' = \mathcal{G}'.\mathbf{u}'$ where $\mathcal{G}' \in \mathbb{R}^{\mathbb{R}\times\mathbb{R}}$ is the transpose of $\mathcal{G}$, i.e. the continuous Toeplitz matrix generated row by row by $g(-.) : t \mapsto g(-t)$ and $\mathbf{u}' \in \mathbb{R}^{\mathbb{R}\times n}$. Thus, for g and h two integrable functions on $\mathbb{R}$, we can rewrite

$$(\mathbf{x} * g).(\mathbf{y} * h)' = \mathbf{x}.\mathcal{G}.\mathcal{H}'.\mathbf{y}'$$

where $\mathcal{G}$ and $\mathcal{H}$ are their associated continuous matrices. More generally, the bold curved letters $\mathcal{G}_c$, $\mathcal{V}$, $\mathcal{W}$ represent these continuous Toeplitz matrices which are well-defined through their action as convolution operators with g_c, v and w. The previous formulation naturally expresses the symmetry of relation (2.30).

With these notations we an alternative definition to the correlation matrices at the heart of our computations $\mathbf{C}^{k,q}$, $\tilde{\mathbf{C}}^{k,q}$ and $\mathbf{D}^{k,q}$,

$$\mathbf{C}^{k,q} \overset{def}{=} \frac{1}{u_m^2\tau}\mathbf{u}\cdot\mathcal{G}_{\mu/l}^{k+1}\cdot{\mathcal{G}'_{\mu/l}}^{q+1}\cdot\mathbf{u}' \tag{C.12}$$

$$\tilde{\mathbf{C}}^{k,q} \overset{def}{=} \frac{1}{u_m^2\tau\|v\|_1^{k+q+2}}\mathbf{u}\cdot\mathcal{V}^{k+1}\cdot\mathcal{V}'^{q+1}\cdot\mathbf{u} \tag{C.13}$$

$$\mathbf{D}^{k,q} \overset{def}{=} \frac{1}{u_m^2\tau\big(|a_+|+|a_-|\big)}\mathbf{u}\cdot\mathcal{G}_{\mu/l}^{k+1}\cdot\big(a_+\mathcal{G}'_{1/\gamma} - a_-\mathcal{G}_{1/\gamma}\big)\cdot{\mathcal{G}'_{\mu/l}}^{q+1}\cdot\mathbf{u} \tag{C.14}$$

C.2.2 Hebbian learning with linear activity

In this part, we consider system (2.28).

C.2.2.1 Application of temporal averaging theory

Theorem C.2.1.*If assumption 2.4.4 is verified for $p \in]0,1[$, then system (2.28) is asymptotically well-posed in probability and the connectivity matrix $\mathbf{J}^\varepsilon$ solution of system (2.28) converges to $\bar{\mathbf{J}}$, in the sense that for all $\delta, T > 0$,*

$$\lim_{\varepsilon \to 0}^{\mu} \mathbb{P}\left[\sup_{t\in[0,T]} |\mathbf{J}^\varepsilon_t - \bar{\mathbf{J}}_t|^2 > \delta\right] = 0$$

where $\bar{\mathbf{J}}$ is the deterministic solution of:

$$\frac{d\bar{\mathbf{J}}_{ij}}{dt} = \bar{G}(\bar{\mathbf{J}})_{ij} = \underbrace{-\kappa\bar{\mathbf{J}}_{ij}}_{decay} + \frac{\mu}{\tau}\underbrace{\int_0^{\frac{\tau}{\mu}} \mathbf{v}_i(s)\mathbf{v}_j(s)\ ds}_{correlation} + \underbrace{\frac{\sigma^2}{2}(\mathbf{L}-\bar{\mathbf{J}})^{-1}_{ij}}_{noise}$$

where $\mathbf{v}(t)$ is the $\frac{\tau}{\mu}$-periodic attractor of $\frac{d\mathbf{v}}{dt} = (\bar{\mathbf{J}} - \mathbf{L}).\mathbf{v} + \mathbf{u}(\mu t)$, where $\mathbf{J} \in \mathbb{R}^{n\times n}$ is supposed to be fixed.

Proof. We are going to apply Prop. 2.4.3. For $p \in]0,1[$, one introduces the space

$$E_p = \left\{\mathbf{J} \in \mathbb{R}^{n\times n} :\ \mathbf{J} \text{ is symmetric, } \mathbf{J} \geq 0 \text{ and } |||\mathbf{J}||| < lp\right\}$$

First, since $\mathbf{L}-\mathbf{J}$ is strictly positive in E_p, assumptions B.3.2-B.3.3 are satisfied on $\mathbb{R}^n \times E_p$. Then we will compute the averaged vector field $\bar{G}$ and show that E_p is invariant under the flow of $\bar{G}$.

1. **Computation of the averaged vector field $\bar{G}$**
   The fast variable is linear, the averaged vector field is given by (B.9). This reads

$$\bar{G}(\mathbf{J}) = \left(\frac{\tau}{\mu}\right)^{-1} \int_0^{\frac{\tau}{\mu}} \int_{\mathbf{x}\in\mathbb{R}^n} G(\mathbf{v}(t) + \mathbf{x}, \mathbf{J})\varphi_{0,\mathbf{Q}}(d\mathbf{x})dt$$

   where $\varphi_{\mathbf{v},\mathbf{Q}}$ is the probability density function of the Gaussian law with mean $\mathbf{v}$ and covariance $\mathbf{Q}$. And $\mathbf{Q}$ is the unique solution of (2.27), with $\mathbf{\Sigma} = \sigma Id$. This leads to $\mathbf{Q} = \frac{\sigma^2}{2}(\mathbf{L} - \mathbf{J})^{-1}$.

Therefore,

$$
\begin{aligned}
\bar{G}(\mathbf{J}) &= -\kappa\mathbf{J} + \frac{\mu}{\tau}\int_0^{\frac{\tau}{\mu}}\left(\int_{\mathbf{v}\in\mathbb{R}^n}(\mathbf{v}(t)+\mathbf{v})\otimes(\mathbf{v}(t)+\mathbf{v})\varphi_{0,\mathbf{Q}}(d\mathbf{v})\right)dt \\
&= -\kappa\mathbf{J} + \frac{\mu}{\tau}\int_0^{\frac{\tau}{\mu}}\mathbf{v}(t)\otimes\mathbf{v}(t)dt \\
&+\frac{\mu}{\tau}\int_0^{\frac{\tau}{\mu}}\left(\mathbf{v}(t)\otimes\underbrace{\int_{\mathbf{v}\in\mathbb{R}^n}\mathbf{v}\varphi_{0,\mathbf{Q}}(d\mathbf{v})}_{\text{Expectation of } \mathcal{N}(0,Q)=0}\right)dt+\frac{\mu}{\tau}\int_0^{\frac{\tau}{\mu}}\left(\underbrace{\left(\int_{\mathbf{v}\in\mathbb{R}^n}\mathbf{v}\varphi_{0,\mathbf{Q}}(d\mathbf{v})\right.}_{\text{Expectation of } \mathcal{N}(0,Q)=0}\otimes\mathbf{v}(t)\right)dt \\
&+\frac{\mu}{\tau}\underbrace{\int_0^{\frac{\tau}{\mu}}\left(\left(\int_{\mathbf{v}\in\mathbb{R}^n}\mathbf{v}\otimes\mathbf{v}\varphi_{0,\mathbf{Q}}(d\mathbf{v})\right)\right.}_{\text{Covariance of } \mathcal{N}(0,Q)=Q}dt \\
&= -\kappa\mathbf{J} + \frac{\mu}{\tau}\int_0^{\frac{\tau}{\mu}}\mathbf{v}(t)\otimes\mathbf{v}(t)\ dt + \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{J})^{-1}
\end{aligned}
$$

The integral term in the equation above is the correlation matrix of the $\frac{\tau}{\mu}$-periodic function $\mathbf{v}$. To rewrite this term, we define $\mathbf{v} \in \mathbb{R}^{n\times[0,\frac{\tau}{\mu}[}$ such that $\mathbf{v}(i,t) = \mathbf{v}(t)_i$. $\mathbf{v}$ can be seen as a matrix gathering the history of $\mathbf{v}$, i.e. each column of $\mathbf{v}$ corresponds to the vector $\mathbf{v}(t)$ for a given $t \in [0, \frac{\tau}{\mu}[$. It turns out

$$\int_0^{\frac{\tau}{\mu}}\mathbf{v}(t)\otimes\mathbf{v}(t)\ dt = \mathbf{v}.\mathbf{v}'$$

Therefore,

$$\bar{G}(\mathbf{J}) = -\kappa\mathbf{J} + \frac{\mu}{\tau}\mathbf{v}.\mathbf{v}' + \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{J})^{-1}$$

According to the results in section 2.4.1, the solutions of this equation are close to that of the initial system (C.2.4). Hence, we focus exclusively on it and try to unveil the properties of its solutions which will be retrospectively extended to those of the initial system (2.28).

2. **Invariance of E_p under the flow of** (2.29):
Here we assume that $\mathbf{J}(0) \in E_p$ and we want to prove that the trajectory of $\mathbf{J}$ is in E_p too.

   (a) Symmetry:
   It is clear that each term in $\bar{G}$ is symmetric. Their sum is therefore symmetric and so is $\mathbf{J}(t)$.

(b) Inequality $\mathbf{J} \geq 0$:

The correlation term $\mathbf{v}.\mathbf{v}'$ is a Gramian matrix and is therefore positive. Because, $\mathbf{L}-\mathbf{J}$ is assumed to be positive, so is its inverse. Therefore, if $\mathbf{e}_i$ is an eigenvector of $\mathbf{J} \geq 0$ associated with a null eigenvalue, then $\mathbf{e}_i'.\bar{G}(\mathbf{J}).\mathbf{e}_i \geq 0$. Thus, the trajectories of (2.29) remain positive.

(c) Inequality $|||\mathbf{J}||| < lp$:

For all $\mathbf{x} \in \mathbb{C}^n$ such that $\|\mathbf{x}\| = 1$, define a family of positive numbers $(\alpha_{\mathbf{x}})$ whose supremum is written α^* and a family of functions $(g^{\mathbf{x}})$ such that

$$g^{\mathbf{x}} : \mathbf{W} \to \|\mathbf{W}.\mathbf{x}\|^2 - \alpha_{\mathbf{x}}^2$$

Observe that $dg^{\mathbf{x}}_{\mathbf{J}}(\mathbf{W}) = \frac{1}{2}\langle \mathbf{J}.\mathbf{x}, \mathbf{W}.\mathbf{x}\rangle$. For $\mathbf{J} \in {g^{\mathbf{x}}}^{-1}(0)$, i.e. $\|\mathbf{J}.\mathbf{x}\| = \alpha_{\mathbf{x}}$, compute

$$2dg^{\mathbf{x}}_{\mathbf{J}}\big(\bar{G}(\mathbf{J})\big) = -\kappa \underbrace{\langle \mathbf{J}.\mathbf{x}, \mathbf{J}.\mathbf{x}\rangle}_{=\alpha_{\mathbf{x}}^2} + \frac{\mu}{\tau} \underbrace{\langle \mathbf{J}.\mathbf{x}, \mathbf{v}.\mathbf{v}'.\mathbf{x}\rangle}_{=A} + \underbrace{\langle \mathbf{J}.\mathbf{x}, \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{J})^{-1}.\mathbf{x}\rangle}_{=B}$$

- Majoration of A:

  Applying Cauchy-Schwarz leads to

$$\begin{aligned} |A| \leq \|\mathbf{J}.\mathbf{x}\| \big\|\mathbf{v}.\mathbf{v}'.\mathbf{x}\big\| &\leq \alpha_{\mathbf{x}} \int_0^{\frac{\tau}{\mu}} \Big\|\mathbf{v}(s) \otimes \mathbf{v}(s).\mathbf{x}\Big\| ds \\ &\leq \alpha_{\mathbf{x}} \int_0^{\frac{\tau}{\mu}} \big|\langle \mathbf{v}(s), \mathbf{x}\rangle\big| \big\|\mathbf{v}(s)\big\| ds \leq \alpha_{\mathbf{x}} \int_0^{\frac{\tau}{\mu}} \big\|\mathbf{v}(s)\big\|^2 ds \end{aligned}$$

  However, for $t \geq 0$

$$\begin{aligned} \big\|\mathbf{v}(t)\big\| \leq \int_{-\infty}^{t} \big\|e^{(\mathbf{J}-\mathbf{L})(t-s)}.\mathbf{u}(\mu s)\big\| ds &\leq u_m \int_{-\infty}^{t} e^{(\alpha^*-l)(t-s)} ds \\ &\leq u_m e^{(\alpha^*-l)t} \left[\frac{e^{-(\alpha^*-l)s}}{l-\alpha^*}\right]_{-\infty}^{t} = \frac{u_m}{l-\alpha^*} \end{aligned}$$

  Therefore, $A \leq \frac{\alpha_{\mathbf{x}} \tau u_m^2}{\mu(l-\alpha)^2}$.

- Majoration of B:

  Observe that for $\mathbf{W}$ a positive definite matrix whose eigenvalues are the λ_i, then the spectrum of $\mathbf{W}^{-1}$ is $\{\frac{1}{\lambda_i}\}$. Therefore, $|||\mathbf{W}^{-1}||| = \frac{1}{min(\lambda_i)}$. Therefore, if $\mathbf{W} = \mathbf{L}-\mathbf{J}$, then $|||\mathbf{W}^{-1}||| \leq \frac{1}{l-\alpha^*}$.

Using the previous observation and Cauchy-Schwarz lead to

$$|B| \leq \alpha_{\mathbf{x}} \frac{\sigma^2}{2} \,|||(\mathbf{L}-\mathbf{J})^{-1}||| \leq \frac{\alpha_{\mathbf{x}}\sigma^2}{2(l-\alpha^*)}$$

Therefore, for $\alpha^* < l$

$$\frac{2\; dg^{\mathbf{x}}_{\mathbf{J}}\big(\bar{G}(\mathbf{J})\big)}{\alpha_{\mathbf{x}}} \leq -\kappa\alpha_{\mathbf{x}} + \frac{u_m^2}{(l-\alpha^*)^2} + \frac{\sigma^2}{2(l-\alpha^*)} = \frac{1}{(l-\alpha^*)^2}P(\alpha^*) + \kappa(\alpha^* - \alpha_{\mathbf{x}})$$

where

$$P(\alpha) = -\kappa\alpha^3 + 2\kappa l \alpha^2 - (\kappa l^2 + \frac{\sigma^2}{2})\alpha + (u_m^2 + \frac{l\sigma^2}{2})$$

Now write $\alpha^* = pl$ with $p \in]0,1[$. Equation (2c) becomes

$$P(p) = -\kappa l^3 p(1-p)^2 + \frac{l\sigma^2}{2}(1-p) + u_m^2$$

When there exists p such that $P(p) < 0$ (which corresponds to assumption 2.4.4), then their exist a ball of radius pl on which the dynamics is pointing inward. It means any matrix $\mathbf{J}$ whose maximal eigenvalue is $\alpha^* = pl$ will see this eigenvalue (and those which are sufficiently close to it, i.e. for which $\alpha^* - \alpha_{\mathbf{x}} > 0$ is sufficiently small) decreasing along the trajectories of the system. Therefore, the space E_p is invariant by the flow of the system iff assumption 2.4.4 is satisfied.

The trajectories of system (2.29) with the initial condition in E_p are defined on $\mathbb{R}_+$ and remain bounded. Indeed, if $\mathbf{J}(0) \in E_p$, the connectivity will stay in E_p, in particular $0 < \mathbf{L} - \mathbf{J} \leq \mathbf{L}$ along the trajectories, more precisely $\mathbf{L} - \mathbf{J} >$ str. postitive constant since $p \in]0,1[$. Because $\mathbf{v}$ is also bounded by $\frac{u_m}{l(1-p)}$, $\mathbf{v}.\mathbf{v}' + \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{J})^{-1}$ is bounded. The right hand side of system (2.29) is the sum a bounded term and a linear term multiplied by a negative constant, therefore, the system remains bounded and it can not explode in finite time: it is defined on $\mathbb{R}_+$. □

C.2.2.2 An expansion for the correlation term

We first write a lemma proved in appendix E.

Lemma C.2.2.*If* $\mathbf{v}$ *is the solution of* $\frac{d\mathbf{v}}{dt} = (\mathbf{J}-\mathbf{L}).\mathbf{v}+\mathbf{u}(t)$*, it can be written by the sum below which converges if* $\mathbf{J}$ *is in* E_p *for* $p \in]0,1[$*.*

$$\mathbf{v} = \sum_{k=0}^{+\infty} \frac{\mathbf{J}^k}{l^{k+1}} \cdot \mathbf{u} * g_{1/l}^{(k+1)}$$

where $g_{1/l} : t \mapsto le^{-lt}H(t)$*.*

Proof. See the first example of the appendix E. □

This is useful to find the next result

Proposition C.2.3. *The correlation term can be written*

$$\frac{\mu}{\tau}\mathbf{v}.\mathbf{v}' = \frac{u_m^2}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{J}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{J}'^q}{l^q}$$

where $\mathbf{C}^{k,q}$ *is defined in* (2.4.2.2) *or* (C.12)*.*

Proof. We can use lemma 2.4.6 with $\mu \neq 1$ and compute the cross product $\mathbf{v}.\mathbf{v}'$.

Therefore, consider $\mathbf{u}(\mu.) : t \mapsto \mathbf{u}(\mu t)$ instead of $\mathbf{u}$. A change of variable shows that $\big(\mathbf{u}(\mu.) * g_{1/l}^{(k)}\big)(t) = \frac{1}{\mu}\big(\mathbf{u} * g_{1/l}^{(k)}(\frac{\cdot}{\mu})\big)(\mu t)$. Therefore,

$$\frac{\mu}{\tau}\{\mathbf{v}.\mathbf{v}'\}_{ij} = \frac{\mu}{\tau}\int_0^{\frac{\tau}{\mu}} \mathbf{v}_i(t)\mathbf{v}_j(t)dt = \frac{1}{\tau}\int_0^{\tau} \mathbf{v}_i(\frac{s}{\mu})\mathbf{v}_j(\frac{s}{\mu})ds$$

$$= \frac{1}{\tau}\int_0^{\tau} \Big(\sum_{k=0}^{+\infty} \frac{\mathbf{J}^k}{l^{k+1}} \cdot \big(\mathbf{u}(\mu.) * g_{1/l}^{(k+1)}\big)(\frac{s}{\mu})\Big)_i \Big(\sum_{q=0}^{+\infty} \frac{\mathbf{J}^q}{l^{q+1}} \cdot \big(\mathbf{u}(\mu.) * g_{1/l}^{(q+1)}\big)(\frac{s}{\mu})\Big)_j ds$$

$$= \frac{1}{\tau}\int_0^{\tau} \Big(\sum_{k=0}^{+\infty} \frac{\mathbf{J}^k}{l^{k+1}} \cdot \big(\mathbf{u} * \frac{g_{1/l}^{(k+1)}(./\mu)}{\mu}\big)(s)\Big)_i \Big(\sum_{q=0}^{+\infty} \frac{\mathbf{J}^q}{l^{q+1}} \cdot \big(\mathbf{u} * \frac{g_{1/l}^{(q+1)}(./\mu)}{\mu}\big)(s)\Big)_j ds$$

$$= \Big\{\frac{u_m^2}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{J}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{J}'^q}{l^q}\Big\}_{ij}$$

□

C.2.2.3 Global stability of the single equilibrium point

Theorem C.2.4. *If assumption 2.4.4 is verified for* $p \leq \frac{1}{3}$ *then there is a unique equilibrium point in the invariant subset* E_p *which is globally, asymptotically stable.*

Proof. For this proof define $F(\mathbf{J}) = \frac{u_m^2}{l^2}\sum_{k,q=0}^{+\infty} \frac{\mathbf{J}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{J}'^q}{l^q} + \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{J})^{-1}$.

First, we compute the differential of F and show it is a bounded operator. Second, we show it implies the existence and uniqueness of an equilibrium point under some condition. Then we find an energy for the system which says the fixed point is a global attractor. Finally, we show the stability condition is the same as assumption 2.4.4 for $p \leq \frac{1}{3}$.

1. We compute the differential of each term in F:

   - Formally write the second term $\mathbf{v}.\mathbf{v}'(\mathbf{J}) = \sum_{k,q=0}^{+\infty} \frac{\mathbf{J}^k}{l^k} \cdot \mathbf{C}^{k,q} \cdot \frac{\mathbf{J}'^q}{l^q}$. To find its differential compute $\mathbf{v}.\mathbf{v}'(\mathbf{J}+\mathbf{W}) - \mathbf{v}.\mathbf{v}'(\mathbf{J})$ and keep the terms at the first order in $\mathbf{W}$. Before computing the whole sum observe that

     $$(\mathbf{J}+\mathbf{W})^k \cdot \mathbf{C}^{k,q} \cdot (\mathbf{J}+\mathbf{W})'^q - \mathbf{J}^k \cdot \mathbf{C}^{k,q} \cdot \mathbf{J}'^q$$
     $$= \sum_{m=0}^{k-1} \mathbf{J}^m \cdot \mathbf{W} \cdot \mathbf{J}^{k-1-m}.\mathbf{C}^{k,q}.\mathbf{J}'^q + \sum_{m=0}^{q-1} \mathbf{J}^k.\mathbf{C}^{k,q}.\mathbf{J}'^m \cdot \mathbf{W}' \cdot \mathbf{J}'^{q-1-m} + \mathcal{O}(\|\mathbf{W}\|^2)$$

     This leads to

     $$d\mathbf{v}.\mathbf{v}'_{\mathbf{J}}(\mathbf{W}) = \frac{1}{l} \sum_{k,q=0}^{+\infty} \Big(\sum_{m=0}^{k-1} \frac{\mathbf{J}^m}{l^m} \cdot \mathbf{W} \cdot \frac{\mathbf{J}^{k-1-m}}{l^{k-1-m}}.\mathbf{C}^{k,q}.\frac{\mathbf{J}'^q}{l^q} + \sum_{l=0}^{q-1} \frac{\mathbf{J}^k}{l^k}.\mathbf{C}^{k,q}.\frac{\mathbf{J}'^m}{l^m} \cdot \mathbf{W}' \cdot \frac{\mathbf{J}'^{q-1-m}}{l^{q-1-m}} \Big)$$

   - Write $Q : \mathbf{J} \mapsto (\mathbf{L} - \mathbf{J})^{-1}$. We can write $(\mathbf{L} - \mathbf{J}).Q(\mathbf{J}) = Id$ and use the chain rule to compute the differential of Q at $\mathbf{J}$ which gives $-\mathbf{W}.Q(\mathbf{J}) + (\mathbf{L} - \mathbf{J}).dQ_{\mathbf{J}}(\mathbf{W}) = 0$. Therefore,

     $$dQ_{\mathbf{J}}(\mathbf{W}) = (\mathbf{L} - \mathbf{J})^{-1}.\mathbf{W}.(\mathbf{L} - \mathbf{J})^{-1}$$

   The differential of F at $\mathbf{J}$ is the sum of these 2 terms.

2. We want to compute the norm of $\|dF_{\mathbf{J}}(\mathbf{W})\|_2$ for $\|J\|_2 = 1$. First, observe that for 3 square matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$,

   $$\|\mathbf{A}.\mathbf{B}.\mathbf{C}\|_2^2 = \sum_{i,j=1}^{n} B_{ij}^2 \|\mathbf{A}.(\mathbf{e}_i \otimes \mathbf{e}_j).\mathbf{C}\|_2^2 \leq \sum_{i,j=1}^{n} B_{ij}^2 \|\mathbf{A}.\mathbf{e}_i\|_2^2 \|\mathbf{C}.\mathbf{e}_j\|_2^2 \leq \sum_{i,j=1}^{n} B_{ij}^2 \, |||\mathbf{A}|||^2 \, |||\mathbf{C}|||^2$$

   for $\mathbf{e}_i$ the vectors of the canonical basis of $\mathbb{R}^n$. This leads to $\|\mathbf{A}.\mathbf{B}.\mathbf{C}\|_2 \leq \|\mathbf{B}\|_2 \, |||\mathbf{A}||| \, |||\mathbf{C}|||$. Therefore, because $|||\mathbf{A}||| \leq \|\mathbf{A}\|_2$

   $$\Big\| \frac{\mathbf{J}^m}{l^m} \cdot \mathbf{W} \cdot \frac{\mathbf{J}^{k-1-m}}{l^{k-1-m}}.\mathbf{C}^{k,q}.\frac{\mathbf{J}'^q}{l^q} \Big\|_2 \leq \frac{|||\mathbf{J}|||^m}{l^m} \Big\| \frac{\mathbf{J}^{k-1-m}}{l^{k-1-m}}.\mathbf{C}^{k,q}.\frac{\mathbf{J}'^q}{l^q} \Big\|_2$$
   $$\leq u_m^2 \frac{|||\mathbf{J}|||^{k-1}}{l^{k-1}} \frac{|||\mathbf{J}|||^q}{l^q}$$

Therefore,

$$\|dF_{\mathbf{J}}(\mathbf{W})\|_2 \leq \frac{u_m^2}{l^3}\sum_{k,q=0}^{+\infty}\Big(k\frac{|||\mathbf{J}|||^{k-1}}{l^{k-1}}\frac{|||\mathbf{J}|||^{q}}{l^{q}}+q\frac{|||\mathbf{J}|||^{k}}{l^{k}}\frac{|||\mathbf{J}|||^{q-1}}{l^{q-1}}\Big)+\frac{\sigma^2}{2}|||(\mathbf{L}-\mathbf{J})^{-1}|||^2$$

$$\leq \frac{2u_m^2}{l^3}\Big(\sum_{k=0}^{+\infty}kp^{k-1}\Big)\Big(\sum_{q=0}^{+\infty}p^q\Big)+\frac{\sigma^2}{2}|||(\mathbf{L}-\mathbf{J})^{-1}|||^2 \leq \frac{2u_m^2}{l^3(1-p)^3}+\frac{\sigma^2}{2l^2(1-p)^2}$$

This inequality is true for all $\mathbf{W}$ with $\|\mathbf{W}\|_2 = 1$, therefore it is also true for the operator norm:

$$|||dF_{\mathbf{J}}||| \leq \frac{2u_m^2}{l^3(1-p)^3}+\frac{\sigma^2}{2l^2(1-p)^2}$$

Therefore, F is a k-Lipschitz operator where $k = \frac{2u_m^2}{l^3(1-p)^3}+\frac{\sigma^2}{2l^2(1-p)^2}$. This means $\|F(\mathbf{J}) - F(\mathbf{W})\|_2 \leq k\|\mathbf{J} - \mathbf{W}\|_2$

3. The equilibrium points of system (2.31) necessarily verify the equation $\mathbf{J} = \frac{1}{\kappa}F(\mathbf{J})$. If
$$\frac{2u_m^2}{(1-p)^3}+\frac{l\sigma^2}{2(1-p)^2} < \kappa l^3 \tag{C.15}$$
then $\frac{1}{\kappa}F$ is a contraction map from E_p to itself. Therefore, the Banach fixed point theorem says that there is a unique fixed point which we write $\mathbf{J}^*$.

4. We now show that, under assumption (C.15), $\mathbf{J} \mapsto \|\mathbf{J}-\mathbf{J}^*\|^2$ is an energy function for system $\frac{d\mathbf{J}}{dt'} = -\mathbf{J} + \frac{1}{\kappa}F(\mathbf{J})$ (which is a rescaled version of system (2.31)).

   Indeed, compute the derivative of this energy along the trajectories of the system

$$\frac{d}{dt}\|\mathbf{J}(t)-\mathbf{J}^*\|^2 = \frac{1}{2}\langle\mathbf{J}-\mathbf{J}^*, -\mathbf{J}+\frac{1}{\kappa}F(\mathbf{J})\rangle = -\langle\mathbf{J}-\mathbf{J}^*,\mathbf{J}-\mathbf{J}^*\rangle+\langle\mathbf{J}-\mathbf{J}^*,\frac{1}{\kappa}F(\mathbf{J})-\mathbf{J}^*\rangle$$
$$= -\|\mathbf{J}-\mathbf{J}^*\|^2+\langle\mathbf{J}-\mathbf{J}^*,\frac{1}{\kappa}F(\mathbf{J})-\frac{1}{\kappa}F(\mathbf{J}^*)\rangle \leq -\|\mathbf{J}-\mathbf{J}^*\|^2+\|\mathbf{J}-\mathbf{J}^*\|\|\frac{1}{\kappa}F(\mathbf{J})-\frac{1}{\kappa}F(\mathbf{J}^*)\|$$
$$\leq \frac{1}{\kappa l^3}\Big(\frac{2u_m^2}{(1-p)^3}+\frac{l\sigma^2}{2(1-p)^2}-\kappa l^3\Big)\|\mathbf{J}-\mathbf{J}^*\|^2 \leq 0$$

   The energy is lower-bounded, takes its minimum for $\mathbf{J} = \mathbf{J}^*$ and the decreases along the trajectories of the system. Therefore, $\mathbf{J}^*$ is globally asymptotically stable if assumption (C.15) is verified.

5. Observe that if assumption 2.4.4 is verified for $p \leq \frac{1}{3}$, then $\frac{1}{1-p} < \frac{2}{1-p} \leq \frac{1}{p}$. Therefore, assumption 2.4.4 implies that (C.15) is also true. This concludes the proof.

□

C.2.2.4 Explicit expansion of the equilibrium point

Recall the notations $\tilde{p} = \frac{u_m^2}{\kappa l^3} + \frac{\sigma^2}{2\kappa l^2}$ and $\lambda = \frac{\sigma^2 l}{2u_m^2}$.

Theorem C.2.5.

$$\mathbf{J}^* = \frac{\tilde{p}l}{1+\lambda}(\lambda + \mathbf{C}^{0,0})$$
$$+ \frac{\tilde{p}^2 l}{(1+\lambda)^2}\Big(\lambda^2 + \lambda(\mathbf{C}^{0,0} + \mathbf{C}^{1,0} + \mathbf{C}^{0,1}) + \mathbf{C}^{0,0}.\mathbf{C}^{1,0} + \mathbf{C}^{0,1}.\mathbf{C}^{0,0}\Big) \; + \mathcal{O}(\tilde{p}^3)$$

Actually , it is possible to compute recursively the nth term of the expansion above, although their complexity explodes.

Proof. Define p^* the smallest value in $]0,1[$ such that assumption 2.4.4 is valid. This implies

$$p^*\Big((1-p^*)^2 + \frac{\sigma^2}{2\kappa l^2}\Big) = \frac{u_m^2}{\kappa l^3} + \frac{\sigma^2}{2\kappa l^2}$$

The weak connectivity index $\tilde{p}$ controls the ratio of the connection over the strength of the intrinsic dynamics. Indeed, these 2 variable are of the same order, because

$$\frac{p^*}{\tilde{p}} = \frac{1}{(1-p^*)^2 + \frac{\sigma^2}{2\kappa l^2}} = \mathcal{O}_{\tilde{p}\to 0}(1)$$

We want to approximate the equilibrium $\mathbf{J}^*$, i.e. the solution of $\bar{G}(\mathbf{J}^*) = 0$, in the regime $\tilde{p} \ll 1$. Define $\mathbf{\Omega} = \frac{\mathbf{J}}{\tilde{p}l}$ such that $|||\mathbf{\Omega}||| = \mathcal{O}(1)$. We abusively write $\bar{G}(\mathbf{\Omega}) = \bar{G}(\tilde{p}l\mathbf{J})$ such that

$$\bar{G}(\mathbf{\Omega}) = -\tilde{p}l\kappa\mathbf{\Omega} + \frac{u_m^2}{l^2}\sum_{k,q=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k \cdot \mathbf{C}^{k,q} \cdot (\tilde{p}\mathbf{\Omega})^q + \frac{\sigma^2}{2l}\sum_{k=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k$$

Recalling $\lambda = \frac{\sigma^2 l}{2u_m^2}$ leads to

$$\bar{G}(\mathbf{\Omega}) = (\frac{u_m^2}{l^2} + \frac{\sigma^2}{2l})\Big(-\mathbf{\Omega} + \frac{1}{1+\lambda}\sum_{k,q=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k \cdot \mathbf{C}^{k,q} \cdot (\tilde{p}\mathbf{\Omega})^q + \frac{\lambda}{1+\lambda}\sum_{k=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k\Big)$$

Now we write a candidate $\boldsymbol{\Omega}^{(m)} = \sum_{a=0}^{m} \tilde{p}^a \boldsymbol{\Omega}_a$, then we chose the terms $\boldsymbol{\Omega}_a = \mathcal{O}(1)$ so that the first m-th orders in $\bar{G}(\boldsymbol{\Omega}^{(m)})$ vanish. This implies that $\|\bar{G}(\boldsymbol{\Omega}^*) - \bar{G}(\boldsymbol{\Omega}^{(m)})\| = \mathcal{O}(\tilde{p}^{m+1})$ where $\boldsymbol{\Omega}^* = \frac{\mathbf{J}^*}{\tilde{p}l}$. Then, we use the fact that the minimal absolute value of the eigenvalues of $\bar{G}$ is larger than $\kappa - \left(\frac{2u_m^2}{l^3(1-p)^3} + \frac{\sigma^2}{2l^2(1-p)^2}\right) > 0$. Indeed, it means

$$\|\mathbf{J}^* - \mathbf{J}^{(m)}\| < \frac{1}{\kappa - \left(\frac{2u_m^2}{l^3(1-p)^3} + \frac{\sigma^2}{2l^2(1-p)^2}\right)} \mathcal{O}(p^{m+1}) < \frac{1}{\kappa - \left(\frac{2u_m^2}{l^3} + \frac{\sigma^2}{2l^2}\right)} \mathcal{O}(p^{m+1})$$

i.e. $\boldsymbol{\Omega}^{(m)} = \boldsymbol{\Omega}^* + \mathcal{O}(\tilde{p}^{m+1})$

Thus, we need to find the $\boldsymbol{\Omega}_a$ such that the first m-th orders in $\bar{G}(\boldsymbol{\Omega}^{(m)})$ vanish. Therefore, we need to expand all the terms in $\bar{G}(\boldsymbol{\Omega})$. The first term is obvious. In the following, we write the second term $F(\boldsymbol{\Omega})$ associated to the correlations and look for an explicit expression of the $\mathbf{F}_a$ such that $F(\boldsymbol{\Omega}) = \sum_{a=0}^{+\infty} \tilde{p}^a \mathbf{F}_a$. Second, we write the third term $Q(\boldsymbol{\Omega})$ associated to the noise and look for an explicit expression of the $\mathbf{Q}_a$ such that $Q(\boldsymbol{\Omega}) = \sum_{a=0}^{+\infty} \tilde{p}^a \mathbf{Q}_a$.

- Finding the $\mathbf{F}_a$:
  First, observe that

  $$\boldsymbol{\Omega}^q = \sum_{i=0}^{+\infty} \tilde{p}^i \sum_{\eta \in \mathbb{N}^q,\ \sum_k \eta_k = i} \boldsymbol{\Omega}_{\eta_1}.\boldsymbol{\Omega}_{\eta_2}.\cdots.\boldsymbol{\Omega}_{\eta_q}$$

  This leads to

  $$F(\boldsymbol{\Omega}) = \frac{1}{1+\lambda} \sum_{k,q=0}^{+\infty} \sum_{\substack{i,j=0 \\ j \leq i}}^{+\infty} \tilde{p}^{i+k+q} \sum_{\substack{\eta \in \mathbb{N}^j,\ \sum_n \eta_n = k \\ \theta \in \mathbb{N}^{i-j},\ \sum_n \theta_n = q}} \boldsymbol{\Omega}_{\eta_1}.\cdots.\boldsymbol{\Omega}_{\eta_j}.\mathbf{C}^{j,i-j}.\boldsymbol{\Omega}'_{\theta_1}.\cdots.\boldsymbol{\Omega}'_{\theta_i}$$

  The a-th term in the power expansion in $\tilde{p}$ verifies $a = i + k + q$. More precisely, this reads

  $$\mathbf{F}_a = \frac{1}{1+\lambda} \sum_{\substack{k,q,i=0 \\ a=i+k+q}}^{+\infty} \sum_{j=0}^{i} \sum_{\substack{\eta \in \mathbb{N}^j,\ \sum_n \eta_n = k \\ \theta \in \mathbb{N}^{i-j},\ \sum_n \theta_n = q}} \boldsymbol{\Omega}_{\eta_1}.\cdots.\boldsymbol{\Omega}_{\eta_j}.\mathbf{C}^{j,i-j}.\boldsymbol{\Omega}'_{\theta_1}.\cdots.\boldsymbol{\Omega}'_{\theta_{i-j}}$$

  This equation is scary but it reduces to simple expressions for small $a \in \mathbb{N}$.

- Finding the $\mathbf{Q}_a$:
  Using equation (C.2.2.4) leads to

$$Q(\mathbf{\Omega}) = \frac{\lambda}{1+\lambda} \sum_{i,q=0}^{+\infty} \tilde{p}^{i+q} \sum_{\eta \in \mathbb{N}^q,\ \sum_k \eta_k = i} \mathbf{\Omega}_{\eta_1}.\mathbf{\Omega}_{\eta_2}.\cdots.\mathbf{\Omega}_{\eta_q}$$

  The a-th term in the power expansion in $\tilde{p}$ verifies $a = i + q$. More precisely, this reads

$$\mathbf{Q}_a = \frac{\lambda}{1+\lambda} \sum_{\substack{q,i=0 \\ a=i+q}}^{+\infty} \tilde{p}^{i+q} \sum_{\eta \in \mathbb{N}^q,\ \sum_k \eta_k = i} \mathbf{\Omega}_{\eta_1}.\mathbf{\Omega}_{\eta_2}.\cdots.\mathbf{\Omega}_{\eta_q}$$

Therefore,

a	$(1+\frac{1}{\lambda})\mathbf{Q}_a$	$(1+\lambda)\mathbf{F}_a$
0	Id	$\mathbf{C}^{0,0}$
1	$\mathbf{\Omega}_0$	$\mathbf{\Omega}_0.\mathbf{C}^{1,0} + \mathbf{C}^{0,1}.\mathbf{\Omega}_0$
2	$\mathbf{\Omega}_0^2 + \mathbf{\Omega}_1$	$\mathbf{\Omega}_0^2.\mathbf{C}^{2,0} + \mathbf{C}^{0,2}.\mathbf{\Omega}_0^2 + \mathbf{\Omega}_0.\mathbf{C}^{1,1}.\mathbf{\Omega}_0 + \mathbf{\Omega}_1.\mathbf{C}^{1,0} + \mathbf{C}^{0,1}.\mathbf{\Omega}_1$

Therefore, it is easy to compute $\mathbf{\Omega}_a = \mathbf{F}_a + \mathbf{Q}_a$ for $a \in \mathbb{N}$. By definition $\mathbf{J} = \tilde{p}l\mathbf{\Omega} = \tilde{p}l(\mathbf{F}+\mathbf{Q})$ which leads to the result. □

C.2.3 Trace learning with damped oscillators and dynamic synapses

Theorem C.2.6.*If assumption 2.4.4 is verified for $p \in]0,1[$, then system* (2.32) *is asymptotically well-posed in probability and the connectivity matrix $\mathbf{J}^\varepsilon$ solution of system* (2.32) *converges to $\bar{\mathbf{J}}$, in the sense that for all $\delta, T > 0$,*

$$\overset{\mu}{\lim_{\varepsilon \to 0}} \mathbb{P}\left[\sup_{t\in[0,T]} |\mathbf{J}_t^\varepsilon - \bar{\mathbf{J}}_t|^2 > \delta\right] = 0$$

where $\bar{\mathbf{J}}$ is the deterministic solution of:

$$\frac{d\bar{\mathbf{J}}_{ij}}{dt} = \bar{G}(\bar{\mathbf{J}})_{ij} = \underbrace{-\kappa\bar{\mathbf{J}}_{ij}}_{decay} + \frac{\mu}{\tau}\underbrace{\int_0^{\frac{\tau}{\mu}} (\mathbf{v}_i * g_{1/\beta})(s)(\mathbf{v}_j * g_{1/\beta})(s)\ ds}_{correlation} + \underbrace{\mathbf{Q}_{22}}_{noise}$$

where $\mathbf{v}(t)$ *is the* $\frac{\tau}{\mu}$*-periodic attractor of* $\frac{d\mathbf{v}}{dt} = (\bar{\mathbf{J}} - \mathbf{L}).\mathbf{v} * g_{1/\beta} + \mathbf{u}(\mu t)$*, where* $\mathbf{J} \in \mathbb{R}^{n\times n}$ *is supposed to be fixed. And* $\mathbf{Q}_{22}$ *is a noise matrix described below.*

Proof. First, it is useful to observe that one can introduce and adaptation variable $\mathbf{z}(t) \in \mathbb{R}^n$ such that system (2.32) is equivalent to

$$\begin{cases} d\begin{pmatrix}\mathbf{v}\\ \mathbf{z}\end{pmatrix} & = \frac{1}{\varepsilon_1}\left[\begin{pmatrix}0 & \mathbf{J}-\mathbf{L}\\ \beta & -\beta\end{pmatrix}\begin{pmatrix}\mathbf{v}\\ \mathbf{z}\end{pmatrix} + \begin{pmatrix}\mathbf{u}(\frac{t}{\varepsilon_2})\\ 0\end{pmatrix}\right]dt + \begin{pmatrix}\frac{\sigma}{\sqrt{\varepsilon_1}}dB(t)\\ 0\end{pmatrix} \\ \frac{d\mathbf{J}}{dt} & = -\kappa\mathbf{J} + \mathbf{z}\otimes\mathbf{z}\end{cases}$$

The structure of the proof of theorem C.2.1 remains unchanged. Indeed, The decay term of the averaged system does not change.

The correlation term is to be replaced by $\frac{\mu}{\tau}\mathbf{v}.\mathcal{G}_{1/\beta}.\mathcal{G}'_{1/\beta}.\mathbf{v}'$.

The noise term we are looking for is $\mathbf{Q}_{22}$ in the Lyapunov equation (see (2.27)) below

$$\begin{pmatrix}0 & \mathbf{J}-\mathbf{L}\\ \beta & -\beta\end{pmatrix}\cdot\begin{pmatrix}\mathbf{Q}_{11} & \mathbf{Q}'_{12}\\ \mathbf{Q}_{12} & \mathbf{Q}_{22}\end{pmatrix} + \begin{pmatrix}\mathbf{Q}_{11} & \mathbf{Q}'_{12}\\ \mathbf{Q}_{12} & \mathbf{Q}_{22}\end{pmatrix}\cdot\begin{pmatrix}0 & \beta\\ \mathbf{J}'-\mathbf{L} & -\beta\end{pmatrix} + \begin{pmatrix}\sigma^2 & 0\\ 0 & 0\end{pmatrix} = 0$$

Because the learning rule is symmetric then the space of symmetric matrices is invariant and we can restrict this section to the symmetric case. It is easy to show that this Lyapunov equation has a unique solution, because the sum of two eigenvalues of the drift matrix is never null (provided $\mathbf{J}$ stays in E_p). This leads to the system

$$\begin{cases} (\mathbf{J}-\mathbf{L}).\mathbf{Q}_{12} + \mathbf{Q}'_{12}.(\mathbf{J}-\mathbf{L}) + \sigma^2 = 0 & (a)\\ \beta(\mathbf{Q}_{11} - \mathbf{Q}_{12}) + \mathbf{Q}_{22}.(\mathbf{J}-\mathbf{L}) = 0 & (b)\\ \mathbf{Q}_{22} = \frac{\mathbf{Q}_{12}+\mathbf{Q}'_{12}}{2} & (c)\end{cases}$$

One solution of equation (a) is $\mathbf{Q}_{12} = \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{J})^{-1}$. Equation (c) defines $\mathbf{Q}_{22}$ properly. Indeed, because $\mathbf{J}$ is symmetric, so is $\mathbf{Q}_{12}$ and $\mathbf{Q}_{22} = \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{J})^{-1}$. Similarly, equation (b) defines $\mathbf{Q}_{11}$ but it remains to be checked that this definition is that of symmetric matrix. In fact it works because $\mathbf{J}$ is assumed symmetric and the noise is proportional to the identity matrix. Indeed, in this case $\mathbf{Q}_{11} = \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{J})^{-1} + \frac{\sigma^2}{2\beta}$. This solution is thus the unique solution of the Lyapunov equation.

Therefore,

$$\bar{G}(\mathbf{J}) = -\kappa\mathbf{J} + \frac{\mu}{\tau}\mathbf{v}\cdot\mathcal{G}_{1/\beta}\cdot\mathcal{G}'_{1/\beta}\cdot\mathbf{v}' + \frac{\sigma^2}{2}(\mathbf{L}-\mathbf{J})^{-1}$$

In the derivation of the condition under which $|||\mathbf{J}|||$ remain smaller than lp, the majoration of the term A changes as follow. Define $M \in \mathbb{R}_+$ so that $\|\mathbf{v}(t)\| \leq M$ for all $t > 0$. Because we assume $\mathbf{v}(\mathbb{R}_-) = 0$, the variation of constant formula for linear retarded differential equations with constant coefficients (see chapter 6 of [Hale and Lunel 1993]) reads $\mathbf{v}(t) = \int_0^t \mathbf{U}(t-s).\mathbf{u}(\mu s)ds$ where the resolvent $\mathbf{U}$ is the solution of $\dot{\mathbf{U}} = (\mathbf{J} - \mathbf{L}).(\mathbf{U} * g)$. We use corollary 1.1 of the chapter 6 of [Hale and Lunel 1993] which is based on Grönwall's lemma, to claim that $\|\mathbf{U}(t-s)\| \leq e^{(t-s)(\alpha^*-l)}$. Therefore,

$$\|\mathbf{v}(t)\| \leq \int_{-\infty}^{t} \|\mathbf{U}(t-s)\| \|u(\mu s)\| ds \leq u_m \left[\frac{e^{(t-s)(\alpha^*-l)}}{l-\alpha^*} \right]_{-\infty}^{t} \leq \frac{u_m}{l-\alpha^*} = M$$

Then, we used Young's inequality for convolution to get $\|(\mathbf{v} * g)(t)\|_2 \leq \|\mathbf{v}\|_2 \|g\|_1 = \|\mathbf{v}\|_2$.

Therefore, the majoration of A remain unchanged.

Therefore, the polynomial P remains the same and assumption 2.4.4 is still relevant to this problem. □

Lemma C.2.7.

$$\mathbf{v} = \sum_{k=0}^{+\infty} \frac{\mathbf{J}^k}{l^{k+1}} \cdot I \cdot \tilde{\mathcal{W}} \cdot \tilde{\mathcal{V}}^k$$

where $\mathbf{u}$ *and* $\mathcal{V}$ *are convolution operators respectively generated by the functions* $\tilde{u}$ *and* $\tilde{v}$ *detailed below*

$$\tilde{w} : t \mapsto \tfrac{l}{2\Delta}\Big((1+\Delta)e^{-\frac{\beta}{2}(1-\Delta)t} - (1-\Delta)e^{-\frac{\beta}{2}(1+\Delta)t}\Big)H(t)$$
$$\tilde{v} : t \mapsto \tfrac{l}{\Delta}\Big(e^{-\frac{\beta}{2}(1-\Delta)t} - e^{-\frac{\beta}{2}(1+\Delta)t}\Big)H(t)$$

where H *is the Heaviside function,* $\Delta = \sqrt{1 - \frac{4l}{\beta}}$. *If* Δ *is a pure imaginary number the expression above still holds with the hyperbolic functions sh and ch being turned into classical trigonometric functions* sin *and* cos *and* Δ *being replaced by its modulus.*

If $\mathbf{J}$ *is in* E_p *for* $p \in]0,1[$ *then this expansion converges.*

Proof. See the second example of the appendix E. □

Using lemma C.3.3, on can define redefine

$$\mathbf{C}^{k,q} = \frac{1}{u_m^2 \tau \|v\|_1^{k+q+2}} \mathbf{u} \cdot \mathcal{V}^{k+1} \cdot (\mathbf{u} \cdot \mathcal{V}^{q+1})'$$

where $\mathcal{V}$ is the convolution operator generated by $v(t) = \frac{l}{\mu\Delta}\Big(e^{-\frac{\beta}{2\mu}(1-\Delta)t} - e^{-\frac{\beta}{2\mu}(1+\Delta)t}\Big)H(t)$ (see section C.3 for details). Observe that applying Young's inequality for convolutions leads to $\|\mathbf{C}^{k,q}\|_2 \leq 1$.

Therefore, we can rewrite theorem 2.4.7 into

Theorem C.2.8.

$$\frac{\mu}{\tau}\mathbf{v}.\mathcal{G}_{1/\beta}.\mathcal{G}'_{1/\beta}.\mathbf{v}' = \frac{u_m^2\|v\|_1^2}{l^2}\sum_{k,q=0}^{+\infty}\frac{\mathbf{J}^k}{(l/\|v\|_1)^k}\cdot\mathbf{C}^{k,q}\cdot\frac{\mathbf{J}'^q}{(l/\|v\|_1)^q}$$

Proof. Similar to that of theorem 2.4.7. □

Theorem C.2.9.*If assumption 2.4.4 is verified for* $p \leq \frac{1}{3}$ *there is a unique equilibrium point which is globally, asymptotically stable.*

Proof. Similar to the previous case. □

With the same definitions for $\tilde{p} = \frac{u_m^2}{\kappa l^3} + \frac{\sigma^2}{2\kappa l^2}$ and $\lambda = \frac{\sigma^2 l}{2u_m^2}$, we can show

Theorem C.2.10.

$$\mathbf{J} = \frac{\tilde{p}l}{1+\lambda}(\lambda + \|v\|_0^2\tilde{\mathbf{C}}^{0,0})$$

$$+\frac{\tilde{p}^2\|v\|_1 l}{(1+\lambda)^2}\Big(\frac{\lambda^2}{\|v\|_1}+\lambda(\|v\|_1\tilde{\mathbf{C}}^{0,0}+\|v\|_1^2\tilde{\mathbf{C}}^{1,0}+\|v\|_1^2\tilde{\mathbf{C}}^{0,1})+\|v\|_1^4\tilde{\mathbf{C}}^{0,0}.\tilde{\mathbf{C}}^{1,0}+\|v\|_1^4\tilde{\mathbf{C}}^{0,1}.\tilde{\mathbf{C}}^{0,0}\Big)+\mathcal{O}(\tilde{p}^3\|v\|_1^2$$

Proof. Define $\mathbf{\Omega} = \frac{\mathbf{J}}{\tilde{p}l}$ so that

$$\bar{G}(\mathbf{\Omega}) = (\frac{u_m^2}{l^2}+\frac{\sigma^2}{2l})\Big(-\mathbf{\Omega}+\frac{\|v\|_1^2}{1+\lambda}\sum_{k,q=0}^{+\infty}(\tilde{p}\|v\|_1\mathbf{\Omega})^k\cdot\tilde{\mathbf{C}}^{k,q}\cdot(\tilde{p}\|v\|_1\mathbf{\Omega})^q+\frac{\lambda}{1+\lambda}\sum_{k=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k\Big)$$

So the expansion will be in orders of $\tilde{p}\|v\|_1$ with $\|v\|_1 \geq 1$.

Therefore,

a	$(1+\frac{1}{\lambda})\mathbf{Q}_a$	$\frac{1+\lambda}{\|v\|_1^2}\mathbf{F}_a$
0	Id	$\tilde{\mathbf{C}}^{0,0}$
1	$\frac{\mathbf{\Omega}_0}{\|v\|_1}$	$\mathbf{\Omega}_0.\tilde{\mathbf{C}}^{1,0}+\tilde{\mathbf{C}}^{0,1}.\mathbf{\Omega}_0$
2	$\frac{\mathbf{\Omega}_0^2+\mathbf{\Omega}_1}{\|v\|_1^2}$	$\mathbf{\Omega}_0^2.\tilde{\mathbf{C}}^{2,0}+\tilde{\mathbf{C}}^{0,2}.\mathbf{\Omega}_0^2+\mathbf{\Omega}_0.\tilde{\mathbf{C}}^{1,1}.\mathbf{\Omega}_0+\mathbf{\Omega}_1.\tilde{\mathbf{C}}^{1,0}+\tilde{\mathbf{C}}^{0,1}.\mathbf{\Omega}_1$

Actually , it is possible to compute recursively the nth terms, although their complexity explodes. Therefore, it is easy to compute $\mathbf{\Omega}_a = \mathbf{F}_a + \mathbf{Q}_a$ for $a \in \mathbb{N}$. By definition $\mathbf{J} = \tilde{p}l\mathbf{\Omega} = \tilde{p}l(\mathbf{F}+\mathbf{Q})$, which leads to the result. □

C.2.4 STDP learning with linear neurons and correlated noise

Consider the following n-dimensional stochastic differential system

$$\begin{cases} d\mathbf{v} & = \frac{1}{\varepsilon_1}\Big(-\mathbf{L}\mathbf{v}+\mathbf{J}.\mathbf{v}+\mathbf{u}(\frac{t}{\varepsilon_2})\Big)dt+\frac{1}{\sqrt{\varepsilon_1}}\mathbf{\Sigma}.dB(t) \\ \frac{d\mathbf{J}}{dt} & = G(\mathbf{v},\mathbf{J}) = -\kappa\mathbf{J}+a_+\mathbf{v}\otimes(\mathbf{v}*g_{1/\gamma})-a_-(\mathbf{v}*g_{1/\gamma})\otimes\mathbf{v} \end{cases}$$

where $\mathbf{u}$ is a continuous input in $\mathbb{R}^n$, $l, \varepsilon_1, \varepsilon_2, \kappa \in \mathbb{R}_+$, $a_+, a_- \in \mathbb{R}$, $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$ and $B(t)$ is a n-dimensional Brownian noise and for all $\gamma > 0$, $g_{1/\gamma} : t \mapsto \gamma e^{-\gamma t} H(t)$ where H is the Heaviside function. Recall the well-posedness assumption 2.4.10

Assumptions C.2.11. *There exists $p \in]0,1[$ such that*

$$\frac{|a_+| + |a_-|}{p(1-p)} \Big(\frac{d\gamma}{2(1+\gamma/l - p)} + \frac{u_m^2}{(1-p)} \Big) < \kappa l^3$$

Theorem C.2.12. *If assumption 2.4.10 is verified for $p \in]0,1[$, then system (2.33) is asymptotically well-posed in probability and the connectivity matrix $\mathbf{J}^\varepsilon$ solution of system (2.33) converges to $\bar{\mathbf{J}}$, in the sense that for all $\delta, T > 0$,*

$$\overset{\mu}{\lim_{\varepsilon \to 0}} \mathbb{P}\left[\sup_{t\in[0,T]} |\mathbf{J}_t^\varepsilon - \bar{\mathbf{J}}_t|^2 > \delta \right] = 0$$

where $\bar{\mathbf{J}}$ is the deterministic solution of:

$$\frac{d\bar{\mathbf{J}}_{ij}}{dt} = \bar{G}(\bar{\mathbf{J}})_{ij} = \underbrace{-\kappa \bar{\mathbf{J}}_{ij}}_{decay}$$

$$+ \underbrace{\frac{\mu}{\tau} \int_0^{\frac{\tau}{\mu}} a_+ \mathbf{v}_i(s)(\mathbf{v}_j * g_{1/\gamma})(s) - a_-(\mathbf{v}_i * g_{1/\gamma})(s)\mathbf{v}_j(s)\ ds}_{correlation} + \underbrace{\mathbf{Q}_{12}}_{noise}$$

where $\mathbf{v}(t)$ is the $\frac{\tau}{\mu}$-periodic attractor of $\frac{d\mathbf{v}}{dt} = (\bar{\mathbf{J}} - \mathbf{L}).\mathbf{v} + \mathbf{u}(\mu t)$, where $\mathbf{J} \in \mathbb{R}^{n\times n}$ is supposed to be fixed. And $\mathbf{Q}_{12}$ is described below.

Proof. First, it is useful to observe that one can introduce and adaptation variable $\mathbf{z}(t) \in \mathbb{R}^n$ such that system (2.32) is equivalent to

$$\begin{cases} d\begin{pmatrix} \mathbf{v} \\ \mathbf{z} \end{pmatrix} & = \frac{1}{\varepsilon_1}\left[\begin{pmatrix} \mathbf{J} - \mathbf{L} & 0 \\ \gamma & -\gamma \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{z} \end{pmatrix} + \begin{pmatrix} \mathbf{u}(\frac{t}{\varepsilon_2}) \\ 0 \end{pmatrix} \right] dt + \begin{pmatrix} \frac{\sigma}{\sqrt{\varepsilon_1}} dB(t) \\ 0 \end{pmatrix} \\ \frac{d\mathbf{J}}{dt} & = -\kappa \mathbf{J} + a_+ \mathbf{v} \otimes \mathbf{z} - a_- \mathbf{z} \otimes \mathbf{v} \end{cases}$$

The structure of the proof of theorem C.2.1 remains unchanged. Indeed, The decay term of the averaged system does not change.

The correlation term is to be replaced by $\frac{\mu}{\tau}\big(a_+ \mathbf{v}.\mathcal{G}'_{1/\gamma}.\mathbf{v}' + a_- \mathbf{v}.\mathcal{G}_{1/\gamma}.\mathbf{v}'\big)$.

The noise term we are looking for is $\mathbf{Q}_{12}$ in the Lyapunov equation (see (2.27)) below

$$\begin{pmatrix} \mathbf{J} - \mathbf{L} & 0 \\ \gamma & -\gamma \end{pmatrix} . \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}'_{12} \\ \mathbf{Q}_{12} & \mathbf{Q}_{22} \end{pmatrix} + \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}'_{12} \\ \mathbf{Q}_{12} & \mathbf{Q}_{22} \end{pmatrix} . \begin{pmatrix} \mathbf{J}' - \mathbf{L} & \gamma \\ 0 & -\gamma \end{pmatrix} + \begin{pmatrix} \mathbf{\Sigma}.\mathbf{\Sigma}' & 0 \\ 0 & 0 \end{pmatrix} = 0$$

This leads to the system

$$\begin{cases} (\mathbf{J}-\mathbf{L}).\mathbf{Q}_{11}+\mathbf{Q}_{11}.(\mathbf{J}'-\mathbf{L})+\mathbf{\Sigma}.\mathbf{\Sigma}'=0 & (a) \\ \gamma(\mathbf{Q}_{11}-\mathbf{Q}_{12})+\mathbf{Q}_{12}.(\mathbf{J}'-\mathbf{L})=0 & (b) \\ \mathbf{Q}_{22}=\frac{\mathbf{Q}_{12}+\mathbf{Q}_{12}'}{2} & (c) \end{cases}$$

In fact, this system would be much more complicated if we had considered damped oscillators and dynamic synapses as in the previous section. In this case, it seems to be impossible to find $\mathbf{Q}_{12}$ explicitly.

In this simplistic case of linear neurons, $\mathbf{Q}_{11}$ is the solution of a Sylvester equation (see equation (a)). Lemma D.3.2 gives an explicit solution: $\mathbf{Q}_{11}=\sum_{k=0}^{+\infty}\mathbf{J}^k.\mathbf{\Sigma}.\mathbf{\Sigma}'.(2\mathbf{L}-\mathbf{J}')^{-(k+1)}$. Equation (b) leads to

$$\mathbf{Q}_{12}=\gamma\mathbf{Q}_{11}.(\mathbf{L}+\gamma-\mathbf{J}')^{-1}=\gamma\sum_{k=0}^{+\infty}\mathbf{J}^k.\mathbf{\Sigma}.\mathbf{\Sigma}'.(2\mathbf{L}-\mathbf{J}')^{-(k+1)}.(\mathbf{L}+\gamma-\mathbf{J}')^{-1}$$

Therefore,

$$\bar{G}(\mathbf{J})=-\kappa\mathbf{J}+\frac{\mu}{\tau}\left(a_+\mathbf{v}.\mathcal{G}_{1/\gamma}'.\mathbf{v}'-a_-\mathbf{v}.\mathcal{G}_{1/\gamma}.\mathbf{v}'\right)+a_+\mathbf{Q}_{12}'-a_-\mathbf{Q}_{12}$$

We show that for $\mathbf{J}$ already in E_p it will stay for ever in E_p:

1. Inequality $\mathbf{J}\geq 0$:
   Decomposing the connectivity as $\mathbf{J}=\mathbf{S}+i\mathbf{A}$ leads to $\langle\mathbf{x},\mathbf{J}.\mathbf{x}\rangle=\langle\mathbf{x},\mathbf{S}.\mathbf{x}\rangle+i\langle\mathbf{x},\mathbf{A}.\mathbf{x}\rangle$. By hermiticity of $\mathbf{S}$ and $\mathbf{A}$, $\langle\mathbf{x},\mathbf{S}.\mathbf{x}\rangle$ and $\langle\mathbf{x},\mathbf{A}.\mathbf{x}\rangle$ are real numbers. This means we only have to show that the eigenvalues of $\mathbf{S}$ remain positive along the dynamics. Taking the symmetric part of equation (C.2.4) leads to

   $$\frac{d\mathbf{S}}{dt}=-\kappa\mathbf{S}+\frac{\mu(a_+-a_-)}{2\tau}\mathbf{v}.(\mathcal{G}_{1/\gamma}+\mathcal{G}_{1/\gamma}').\mathbf{v}'+(a_+-a_-)\mathbf{Q}_{22}$$

   Suppose we take an initial condition $\mathbf{S}_0>0$. It is clear that, if $\mathbf{v}.(\mathcal{G}+\mathcal{G}').\mathbf{v}'$ and $\mathbf{Q}_{22}$ are always positive then $\mathbf{S}$ will remain positive. This would prove the result. Therefore, focus on

   - Proving $\mathbf{v}.(\mathcal{G}_{1/\gamma}+\mathcal{G}_{1/\gamma}').\mathbf{v}'\geq 0$:
     According to the first point of lemma C.3.1, $\mathcal{G}_{1/\gamma}+\mathcal{G}'=2\mathcal{G}_{1/\gamma}.\mathcal{G}_{1/\gamma}'$. Therefore, $\mathbf{v}.(\mathcal{G}_{1/\gamma}+\mathcal{G}_{1/\gamma}').\mathbf{v}'=2\mathbf{v}.\mathcal{G}_{1/\gamma}.(\mathbf{v}.\mathcal{G}_{1/\gamma})'$ is a Gramian matrix and therefore it is positive.

- Proving $\mathbf{Q}_{22} \geq 0$:
  $\mathbf{Q}_{22}$ is the covariance matrix of the random value $\mathbf{z}$, therefore, it is positive-semidefinite.

2. Inequality $|||\mathbf{J}||| < lp$:
   For all $\mathbf{x} \in \mathbb{C}^n$ such that $\|\mathbf{x}\| = 1$, define a family of positive numbers $(\alpha_{\mathbf{x}})$ whose supremum is written α^* and a family of functions $(g^{\mathbf{x}})$ such that

$$g^{\mathbf{x}} : \mathbf{J} \rightarrow \langle \mathbf{x}, \mathbf{J}.\mathbf{x}\rangle - \alpha_{\mathbf{x}}$$

Because g is linear, $dg^{\mathbf{x}}_{\mathbf{W}}(\mathbf{J}) = \langle \mathbf{x}, \mathbf{J}.\mathbf{x}\rangle$. For $\mathbf{W} \in g^{\mathbf{x}-1}(0)$, i.e. $\langle \mathbf{x}, \mathbf{W}.\mathbf{x}\rangle = \alpha_{\mathbf{x}}$, compute

$$dg^{\mathbf{x}}_{\mathbf{J}}\big(G^\mu(\mathbf{J})\big) = -\kappa \underbrace{\langle \mathbf{x}, \mathbf{J}.\mathbf{x}\rangle}_{=\alpha_{\mathbf{x}}} + \frac{\mu}{\tau} \underbrace{\langle \mathbf{x}, \mathbf{v}.(a_+\mathcal{G}_{1/\gamma} - a_-\mathcal{G}'_{1/\gamma}).\mathbf{v}'.\mathbf{x}\rangle}_{=A} + (|a_+|+|a_-|) \underbrace{\langle \mathbf{x}, \mathbf{Q}_{12}.\mathbf{x}\rangle}_{=B}$$

- Majoration of A:
  Cauchy Schwarz leads to

$$|A| \leq |a_+| \|\mathbf{v}.\mathcal{G}.\mathbf{v}'.\mathbf{x}\| + |a_-| \|\mathbf{v}.\mathcal{G}'.\mathbf{v}'.\mathbf{x}\|$$

As before we can find an upper bound of A which reads

$$A \leq \frac{\tau u_m^2(|a_+| + |a_-|)}{(l - \alpha^*)^2}$$

- Majoration of B:
  According to proposition 11.9.3 of [Bernstein 2009] the solution of the Lyapunov equation (a) in system (C.2.4) can be rewritten

$$\mathbf{Q}_{11} = \int_0^{+\infty} e^{-t(\mathbf{L}-\mathbf{J})}.\mathbf{\Sigma}.\mathbf{\Sigma}^t.e^{-t(\mathbf{L}-\mathbf{J}')}\, dt$$

because $(\mathbf{J} - \mathbf{L}) \oplus (\mathbf{J} - \mathbf{L})$ is not singular due to the fact $\mathbf{J} \in E_p$. Observe that for $\mathbf{A}$ a positive matrix whose eigenvalues are the λ_i, then the spectrum of $e^{-\mathbf{A}}$ is $\big\{e^{-\lambda_i} : i = 1..n\big\}$. Therefore, $|||e^{-\mathbf{A}}||| = e^{-min(|\lambda_i|)}$. Therefore, if $\mathbf{A} = \mathbf{L} - \mathbf{J}$, then $|||e^{-\mathbf{A}}||| \leq e^{\alpha^* - l}$. This leads to

$$|||\mathbf{Q}_{11}||| \leq d \int_0^{+\infty} e^{2(\alpha^*-l)t} dt = d\Big[\frac{e^{2(\alpha^*-l)t}}{2(\alpha^*-l)}\Big]_0^{+\infty} = \frac{d}{2(l-\alpha^*)}$$

Then we apply the same arguments to say that

$$|B| = |||\mathbf{Q}_{12}||| \leq |||\mathbf{Q}_{11}|||\ |||\gamma(\mathbf{L}+\gamma-\mathbf{J})^{-1}||| \leq \frac{d\gamma}{2(l-\alpha^*)(l+\gamma-\alpha^*)}$$

The rest of the proof is identical to the Hebbian case. Assumption 2.4.4 is changed to assumption 2.4.10 for E_p to be invariant by the flow $\bar{G}$. □

Define

$$\mathbf{D}^{k,q} = \frac{1}{u_m^2 \tau \big(|a_+| + |a_-|\big)} \mathbf{u} \cdot \mathcal{G}_{\mu/l}^{k+1} \cdot \big(a_+ \mathcal{G}'_{1/\gamma} - a_- \mathcal{G}_{1/\gamma}\big) \cdot \mathcal{G}'_{\mu/l}{}^{k+1} \cdot \mathbf{u}'$$

In this framework, on can prove

Theorem C.2.13. *The correlation term can be written*

$$\frac{\mu}{\tau}\big(a_+ \mathbf{v}.\mathcal{G}'_{1/\gamma}.\mathbf{v}' - a_- \mathbf{v}.\mathcal{G}_{1/\gamma}.\mathbf{v}'\big) = \frac{u_m^2\big(|a_+| + |a_-|\big)}{l^2} \sum_{k,q=0}^{+\infty} \frac{\mathbf{J}^k}{l^k} \cdot \mathbf{D}^{k,q} \cdot \frac{\mathbf{J}'^q}{l^q}$$

Proof. Similar to that of theorem 2.4.7. □

Theorem C.2.14. *If assumption 2.4.10 is verified for $p \leq \frac{1}{3}$ there is a unique equilibrium point which is globally, asymptotically stable.*

Proof. Similar to the previous case. □

Now, we proceed as before by defining

$$\tilde{p} = \frac{|a_+| + |a_-|}{\kappa l^3}\Big(\frac{d}{2(\frac{1}{l} + \frac{1}{\gamma})} + u_m^2\Big) \quad \text{and} \quad \lambda = \frac{d}{2u_m^2(\frac{1}{l} + \frac{1}{\gamma})}$$

Theorem C.2.15.

$$\begin{aligned}
\mathbf{J} = {} & \frac{\tilde{p} l}{1+\lambda}\big(\lambda(\alpha_+ - \alpha_-)\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d} + \mathbf{D}^{0,0}\big) \\
+ & \frac{\tilde{p}^2 l}{(1+\lambda)^2}\Bigg(\lambda^2(a_+ - a_-)^2(1+\frac{1}{1+\gamma/l})\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'^2}{d^2} + \lambda(\frac{a_+ - a_-}{2})\Big((\mathbf{D}^{0,0} + 2\mathbf{D}^{0,1}).\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d} + \frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d}.(\mathbf{D}^{0,0} + 2\mathbf{D}^{1} \\
& \quad + \frac{\lambda}{1+\gamma/l}\big(a_+ \mathbf{D}^{0,0}.\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d} - a_- \frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d}.\mathbf{D}^{0,0}\big) + \mathbf{D}^{0,0}.\mathbf{D}^{1,0} + \mathbf{D}^{1,0}.\mathbf{D}^{0,0}\Bigg) \\
& \qquad + \mathcal{O}(\tilde{p}^3)
\end{aligned}$$

Proof. First, we need to work on the noise term $\mathbf{Q} = a_+ \mathbf{Q}'_{12} + a_- \mathbf{Q}_{12}$. Recall $\mathbf{Q}_{11}$ is solution of the Lyapunov equation $(\mathbf{L} - \mathbf{J}).\mathbf{Q}_{11} + \mathbf{Q}_{11}.(\mathbf{L} - \mathbf{J})' + \mathbf{\Sigma}.\mathbf{\Sigma}' = 0$. Lemma D.3.2 says that

$$\mathbf{Q}_{11} = \sum_{k=0}^{+\infty} \mathbf{J}^k.\mathbf{\Sigma}.\mathbf{\Sigma}'.(2\mathbf{L} - \mathbf{J}')^{-(k+1)}$$

is a well-defined solution. We now use the fact that $(2\mathbf{L}-\mathbf{J}')^{-(k+1)} = \frac{1}{(2l)^{k+1}}\sum_{n=0}^{+\infty}\binom{n+k}{n}\frac{\mathbf{J}'^{n}}{(2l)^{n}}$ to show that

$$\mathbf{Q}_{11} = \sum_{k,n=0}^{+\infty}\frac{1}{(2l)^{k+n+1}}\binom{n+k}{n}\mathbf{J}^k.\mathbf{\Sigma}.\mathbf{\Sigma}'.\mathbf{J}'^n$$

and therefore

$$\mathbf{Q}_{12} = \frac{\gamma}{2l(l+\gamma)}\sum_{k,n,q=0}^{+\infty}\frac{1}{2^{k+n}(1+\gamma/l)^q}\binom{n+k}{n}\frac{\mathbf{J}^k}{l^k}.\mathbf{\Sigma}.\mathbf{\Sigma}'.\frac{\mathbf{J}'^{n+q}}{l^{n+q}}$$

Thus, writing $\alpha_\pm = \frac{a_\pm}{|a_+|+|a_-|}$ and $c_{k,n,q} = \frac{1}{2^{k+n}(1+\gamma/l)^q}\binom{n+k}{n}$, the noise term is

$$\mathbf{Q} = \frac{d\big(|a_+|+|a_-|\big)}{2l^2\big(\frac{1}{l}+\frac{1}{\gamma}\big)}\sum_{k,n,q=0}^{+\infty}c_{k,n,q}\Big(\alpha_+\frac{\mathbf{J}^{n+q}}{l^{n+q}}.\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d}.\frac{\mathbf{J}'^k}{l^k}-\alpha_-\frac{\mathbf{J}^k}{l^k}.\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d}.\frac{\mathbf{J}'^{n+q}}{l^{n+q}}\Big)$$

Define $\mathbf{\Omega} = \frac{\mathbf{J}}{\tilde{p}l}$ such that $|||\mathbf{\Omega}||| = \mathcal{O}(1)$. We abusively write $\bar{G}(\mathbf{\Omega}) = \bar{G}(\tilde{p}l\mathbf{J})$ such that

$$\bar{G}(\mathbf{\Omega}) = -\tilde{p}l\kappa\mathbf{\Omega} + \frac{u_m^2\big(|a_+|+|a_-|\big)}{l^2}\sum_{k,q=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k\cdot\mathbf{D}^{k,q}\cdot(\tilde{p}\mathbf{\Omega})^q$$

$$+\frac{d\big(|a_+|+|a_-|\big)}{2l^2\big(\frac{1}{l}+\frac{1}{\gamma}\big)}\sum_{k,n,q=0}^{+\infty}c_{k,n,q}\Big(\alpha_+(\tilde{p}\mathbf{\Omega})^{n+q}.\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d}.(\tilde{p}\mathbf{\Omega}')^k-\alpha_-(\tilde{p}\mathbf{\Omega})^k.\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d}.)(\tilde{p}\mathbf{\Omega}')^{n+q}\Big)$$

This leads to

$$\bar{G}(\mathbf{\Omega}) = \Big(\frac{u_m^2\big(|a_+|+|a_-|\big)}{l^2}+\frac{d\big(|a_+|+|a_-|\big)}{2l^2\big(\frac{1}{l}+\frac{1}{\gamma}\big)}\Big)\Big[-\mathbf{\Omega}+\frac{1}{1+\lambda}\overbrace{\sum_{k,q=0}^{+\infty}(\tilde{p}\mathbf{\Omega})^k\cdot\mathbf{D}^{k,q}\cdot(\tilde{p}\mathbf{\Omega})^q}^{\tilde{\mathbf{F}}}$$

$$+\frac{\lambda}{1+\lambda}\underbrace{\sum_{k,n,q=0}^{+\infty}c_{k,n,q}\Big(\alpha_+(\tilde{p}\mathbf{\Omega})^{n+q}.\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d}.(\tilde{p}\mathbf{\Omega}')^k-\alpha_-(\tilde{p}\mathbf{\Omega})^k.\frac{\mathbf{\Sigma}.\mathbf{\Sigma}'}{d}.)(\tilde{p}\mathbf{\Omega}')^{n+q}\Big)\Big]}_{\tilde{\mathbf{Q}}}$$

We are looking for $\mathbf{F}_a$ and $\mathbf{Q}_a$ in the following expansions $\tilde{\mathbf{F}} = \sum_{a=0}^{+\infty}\mathbf{F}_a\tilde{p}^a$ and $\tilde{\mathbf{Q}} = \sum_{a=0}^{+\infty}\mathbf{Q}_a\tilde{p}^a$. Recall

$$\mathbf{\Omega}^p = \sum_{i=0}^{+\infty}\tilde{p}^i\sum_{\eta\in\mathbb{N}^p,\ \sum_k\eta_k=i}\mathbf{\Omega}_{\eta_1}.\mathbf{\Omega}_{\eta_2}.\cdots.\mathbf{\Omega}_{\eta_p}$$

Therefore,

$$\tilde{\mathbf{Q}} = \sum_{k,n,q,i,j=0}^{+\infty} c_{k,n,q}\tilde{p}^{k+n+q+i+j}$$

$$\sum_{\substack{\eta \in \mathbb{N}^k,\ \sum_m \eta_m = i \\ \theta \in \mathbb{N}^{n+q},\ \sum_m \theta_m = j}} \alpha_+\boldsymbol{\Omega}_{\eta_1}.\cdots.\boldsymbol{\Omega}_{\eta_{n+q}}.\frac{\boldsymbol{\Sigma}.\boldsymbol{\Sigma}'}{d}.\boldsymbol{\Omega}'_{\theta_1}.\cdots.\boldsymbol{\Omega}'_{\theta_k} - \alpha_-\boldsymbol{\Omega}_{\eta_1}.\cdots.\boldsymbol{\Omega}_{\eta_k}.\frac{\boldsymbol{\Sigma}.\boldsymbol{\Sigma}'}{d}.\boldsymbol{\Omega}'_{\theta_1}.\cdots.\boldsymbol{\Omega}$$

Leading to

$$\mathbf{Q}_a = \sum_{\substack{k,n,i,j=0 \\ a=k+n+q+i+j}}^{+\infty} c_{k,n,q}\ \tilde{p}^{k+n+q+i+j}$$

$$\sum_{\substack{\eta \in \mathbb{N}^k,\ \sum_m \eta_m = i \\ \theta \in \mathbb{N}^{n+q},\ \sum_m \theta_m = j}} \alpha_+\boldsymbol{\Omega}_{\eta_1}.\cdots.\boldsymbol{\Omega}_{\eta_{n+q}}.\frac{\boldsymbol{\Sigma}.\boldsymbol{\Sigma}'}{d}.\boldsymbol{\Omega}'_{\theta_1}.\cdots.\boldsymbol{\Omega}'_{\theta_k} - \alpha_-\boldsymbol{\Omega}_{\eta_1}.\cdots.\boldsymbol{\Omega}_{\eta_k}.\frac{\boldsymbol{\Sigma}.\boldsymbol{\Sigma}'}{d}.\boldsymbol{\Omega}'_{\theta_1}.\cdots.\boldsymbol{\Omega}$$

This equation is scary but it reduces to simple expressions for small $a \in \mathbb{N}$.

a	$\mathbf{Q}_a$	$\mathbf{F}_a$
0	$(\alpha_+ - \alpha_-)\frac{\boldsymbol{\Sigma}.\boldsymbol{\Sigma}'}{d}$	$\mathbf{D}^{0,0}$
1	$\frac{\alpha_+ - \alpha_-}{2}\Big(\boldsymbol{\Omega}_0.\frac{\boldsymbol{\Sigma}.\boldsymbol{\Sigma}'}{d} + \frac{\boldsymbol{\Sigma}.\boldsymbol{\Sigma}'}{d}.\boldsymbol{\Omega}'_0\Big) + \frac{1}{1+\gamma/l}\Big(\alpha_+\boldsymbol{\Omega}_0.\frac{\boldsymbol{\Sigma}.\boldsymbol{\Sigma}'}{d} - \alpha_-\frac{\boldsymbol{\Sigma}.\boldsymbol{\Sigma}'}{d}.\boldsymbol{\Omega}'_0\Big)$	$\boldsymbol{\Omega}_0.\mathbf{D}^{1,0} + \mathbf{D}^{0,1}.\boldsymbol{\Omega}'_0$

Recall that $\mathbf{J} = \tilde{p}l\boldsymbol{\Omega} = \tilde{p}l(\frac{1}{1+\lambda}\tilde{\mathbf{F}} + \frac{\lambda}{1+\lambda}\tilde{\mathbf{Q}})$ to get the result. □

C.3 Properties of the convolution operators $\mathcal{G}_{1/\gamma}$, $\mathcal{W}$ and $\mathcal{V}$

Recall $\mathcal{G}_{1/\gamma}$, $\mathcal{W}$ and $\mathcal{V}$ are convolution operators respectively generated by $g_{1/\gamma}$, v and w defined in (A). Their Fourier transforms are respectively

$$\begin{aligned}
&\hat{g}_{1/\gamma} : \xi \mapsto \tfrac{\gamma}{\gamma+2i\pi\xi} \\
&\hat{v} : \xi \mapsto \tfrac{4\beta}{\big(\beta(1+\Delta)+4i\pi\mu\xi\big)\big(\beta(1-\Delta)+4i\pi\mu\xi\big)} \\
&\hat{w} : \xi \mapsto \tfrac{4\beta+8i\pi\mu\xi}{\big(\beta(1+\Delta)+4i\pi\mu\xi\big)\big(\beta(1-\Delta)+4i\pi\mu\xi\big)}
\end{aligned}$$

C.3.1 Algebraic properties

Lemma C.3.1.

$$\frac{\mathcal{G}_{1/\gamma}+\mathcal{G}'_{1/\gamma}}{2}=\mathcal{G}_{1/\gamma}.\mathcal{G}'_{1/\gamma}$$

Proof. Compute

$$\begin{aligned}(\mathcal{G}_{1/\gamma}.\mathcal{G}'_{1/\gamma})_{xy} &= \gamma^2\int_{-\infty}^{+\infty} e^{-\gamma(x-z)}H(x-z)e^{-\gamma(y-z)}H(y-z)\ dz\\ &= \gamma^2 e^{-\gamma(x+y)}\int_{-\infty}^{min(x,y)} e^{2\gamma z}\ dz = \gamma^2 e^{-\gamma(y+x)}\left[\frac{e^{2\gamma z}}{2\gamma}\right]_{-\infty}^{min(x,y)}\\ &= \frac{\gamma}{2}e^{-\gamma\left(y+x-2min(x,y)\right)}\end{aligned}$$

Therefore, if $y\geq x$ then $(\mathcal{G}_{1/\gamma}.\mathcal{G}'_{1/\gamma})_{xy}=\frac{\gamma}{2}e^{-\gamma(y-x)}$ and if $x\geq y$ then $(\mathcal{G}_{1/\gamma}.\mathcal{G}'_{1/\gamma})_{xy}=\frac{\gamma}{2}e^{-\gamma(x-y)}$. The result follows. □

Lemma C.3.2.

$$\mathcal{G}'_{1/\gamma}-\mathcal{G}_{1/\gamma}=\frac{1}{\gamma}\mathcal{D}.(\mathcal{G}'_{1/\gamma}+\mathcal{G}_{1/\gamma})$$

where $\mathcal{D}$ is the time-differentiation operator, i.e. $(\mathcal{X}.\mathcal{D})(t)=\frac{d\mathcal{X}}{dt}(t)$.

Proof. $\mathcal{G}_{1/\gamma}$ and $\mathcal{G}'_{1/\gamma}$ are a convolution operators respectively generated by $g_{1/\gamma}:t\mapsto\gamma e^{-\gamma t}H(t)$ and $g'_{1/\gamma}:t\mapsto\gamma e^{\gamma t}H(-t)$. The Fourier transform of $g_{1/\gamma}$ is $\hat{h}(\xi)=\frac{1}{\gamma+2i\pi\xi}$. Therefore, the Fourier transform of $g'_{1/\gamma}-g_{1/\gamma}$ is

$$\begin{aligned}\widehat{g'_{1/\gamma}-g_{1/\gamma}}(\xi) &= \frac{1}{\gamma-2i\pi\xi}-\frac{1}{\gamma+2i\pi\xi}=\frac{2i\pi\xi}{\gamma}\frac{2\gamma}{\gamma^2+4\pi^2\xi^2}\\ &= \frac{2i\pi\xi}{\gamma}\Big(\frac{1}{\gamma+2i\pi\xi}+\frac{1}{\gamma-2i\pi\xi}\Big)=\frac{2i\pi\xi}{\gamma}\big(\widehat{g'_{1/\gamma}+g_{1/\gamma}}(\xi)\big)\end{aligned}$$

Because $\hat{\frac{df}{dt}}(\xi)=2i\pi\xi\hat{f}$, taking the inverse Fourier transform of $\widehat{g'_{1/\gamma}-g_{1/\gamma}}(\xi)$ gives the result. □

Lemma C.3.3.

$$\mathcal{W}\cdot\mathcal{V}^k\cdot\mathcal{G}_{\mu/\beta}=\mathcal{V}^{k+1}$$

Besides, if $\Delta\in i\mathbb{R}$, $\mathcal{V}^k$ is a convolution operator generated by

$$v_k:t\mapsto\frac{\sqrt{\pi\beta}}{k!}e^{-\frac{\beta}{2}t}\Big(\frac{t}{|\Delta|}\Big)^{k+\frac{1}{2}}J_{k+\frac{1}{2}}\big(\frac{\beta|\Delta|}{2}t\big)H(t)$$

where $J_n(z)$ is the Bessel function of the first kind. If $\Delta\in\mathbb{R}$, the formula above holds if one replace $J_n(z)$ by $I_n(z)$ the modified Bessel function of the first kind.

Proof. We want to compute $\mathcal{W} \cdot \mathcal{V}^k \cdot \mathcal{G}_{\mu/\beta}$. Compute the Fourier transform of $w * v_k * g_{\mu/\beta}$, where v_k is the result of k convolutions of v with itself:

$$w \widehat{* v_k *} g_{\frac{\mu}{\beta}}(\xi) = \hat{w}(\xi)\hat{g}_{\frac{\mu}{\beta}}(\xi)\hat{v_k}(\xi) = \left(\frac{\beta}{\left(\frac{\beta(1+\Delta)}{2} + 2i\pi\mu\xi\right)\left(\frac{\beta(1-\Delta)}{2} + 2i\pi\mu\xi\right)}\right)^{k+1} = \hat{v}^{k+1}(\xi)$$

This proves the first result.

Then observe that

$$\begin{aligned} v_{k+1}(t) &= \beta^{k+1}\mathcal{F}^{-1}\Big(\xi \mapsto \frac{1}{\left(\frac{\beta(1+\Delta)}{2} + 2i\pi\mu\xi\right)^{k+1}}\Big) * \mathcal{F}^{-1}\Big(\xi \mapsto \frac{1}{\left(\frac{\beta(1-\Delta)}{2} + 2i\pi\mu\xi\right)^{k+1}}\Big)(t) \\ &= \beta^{k+1}\Big(s \mapsto \frac{s^k}{k!}e^{-\frac{\beta(1+\Delta)}{2}s}H(s)\Big) * \Big(s \mapsto \frac{s^k}{k!}e^{-\frac{\beta(1-\Delta)}{2}s}H(s)\Big)(t) \\ &= \frac{\beta^{k+1}}{k!^2}e^{-\frac{\beta(1-\Delta)}{2}t}\int_0^t s^k(t-s)^k e^{-\beta\Delta s}\, ds\ H(t) \end{aligned}$$

The last integral can be analytically computed with the help of Bessel functions. In fact, it gives different results depending on the nature of Δ (whether it is real or imaginary).

- If $\Delta \in \mathbb{R}$, then defining $I_n(z)$ the modified Bessel function of the first kind leads to
$$\int_0^t e^{-\beta\Delta s}s^k(t-s)^k ds = \sqrt{\pi}e^{-\frac{\beta\Delta}{2}t}k!\left(\frac{t}{\beta\Delta}\right)^{k+\frac{1}{2}}I_{k+\frac{1}{2}}\left(\frac{\beta\Delta}{2}t\right)$$
- If $\Delta \in i\mathbb{R}$, then defining $J_n(z)$ the Bessel function of the first kind leads to
$$\int_0^t e^{-\beta\Delta s}s^k(t-s)^k ds = \sqrt{\pi}e^{-\frac{\beta\Delta}{2}t}k!\left(\frac{t}{\beta|\Delta|}\right)^{k+\frac{1}{2}}J_{k+\frac{1}{2}}\left(\frac{\beta|\Delta|}{2}t\right)$$

This concludes the proof. □

C.3.2 Signed integral

1. $\int_{-\infty}^{+\infty} g_{1/\gamma}(t)dt = \gamma\frac{0-1}{-\gamma} = 1$.

2. For $\Delta = \sqrt{1 - \frac{4l}{\beta}} \in \mathbb{C}$ compute

$$\int_{-\infty}^{+\infty} v(t)dt = \frac{l}{\Delta\mu}\Big(\int_0^{+\infty} e^{-\frac{\beta}{2\mu}(1-\Delta)t}dt - \int_0^{+\infty} e^{-\frac{\beta}{2\mu}(1-\Delta)t}dt\Big)$$

$$= \frac{l}{\Delta\mu}\Big(\frac{0-1}{-\frac{\beta}{2\mu}(1-\Delta)} - \frac{0-1}{-\frac{\beta}{2\mu}(1+\Delta)}\Big) = \frac{2l}{\Delta\beta}\frac{1+\Delta-(1-\Delta)}{1-\Delta^2} = \frac{4l}{\beta}\frac{\beta}{4l} = 1$$

3. Similarly

$$\int_{-\infty}^{+\infty} w(t)dt = \frac{l}{2\Delta\mu}\Big((1{+}\Delta)\int_0^{+\infty} e^{-\frac{\beta}{2\mu}(1-\Delta)t}dt{-}(1{-}\Delta)\int_0^{+\infty} e^{-\frac{\beta}{2\mu}(1-\Delta)t}dt\Big)$$
$$= \frac{l}{\Delta\beta}\frac{(1+\Delta)^2-(1-\Delta)^2}{1-\Delta^2} = \frac{l}{\Delta\beta}\frac{4\Delta\beta}{4l} = 1$$

C.3.3 L1 norm

- For $4l \leq \beta$, i.e. $\Delta = \sqrt{1-\frac{4l}{\beta}} \in \mathbb{R}_+$, then

  1. $g_{1/\gamma}(t) > 0$ and $\|g_{1/\gamma}\|_1 = \int_{\mathbb{R}} g_{1/\gamma}(t)dt = 1$.
  2. $v(t) = \frac{2l}{\Delta\mu}e^{-\frac{\beta}{2\mu}t}sh(\frac{\beta\Delta}{2\mu}t)H(t) \geq 0$ and $\|v\|_1 = \int_{\mathbb{R}} v(t)dt = 1$.
  3. $w(t) = \frac{l}{\Delta\mu}e^{-\frac{\beta}{2\mu}t}\big(sh(\frac{\beta\Delta}{2\mu}t){+}\Delta ch(\frac{\beta\Delta}{2\mu}t)\big)H(t) \geq 0$ and $\|u\|_1 = \int_{\mathbb{R}} u(t)dt = 1$.

- For $4l > \beta$, i.e. Δ is a pure imaginary, we rewrite $\Delta = \sqrt{\frac{4l}{\beta}-1}$ and observe that

  1. $g_{1/\gamma}(t) > 0$ and $\|g_{1/\gamma}\|_1 = \int_{\mathbb{R}} g_{1/\gamma}(t)dt = 1$.
  2. $v(t) = \frac{2l}{\Delta\mu}e^{-\frac{\beta}{2\mu}t}\sin(\frac{\beta\Delta}{2\mu}t)H(t)$ which changes sign on $\mathbb{R}_+$. Therefore,

$$\|v\|_1 = \frac{2l}{\Delta\mu}\int_0^{+\infty} e^{-\frac{\beta}{2\mu}t}\big|\sin(\frac{\beta\Delta}{2\mu}t)\big|dt = \frac{4l}{\Delta^2\beta}\int_0^{+\infty} e^{-\frac{s}{\Delta}}\big|\sin(s)\big|ds$$
$$= \frac{4l}{\Delta^2\beta}\sum_{k=0}^{+\infty}(-1)^k\int_{k\pi}^{(k+1)\pi} e^{-\frac{s}{\Delta}}\sin(s)ds = \frac{4l}{\Delta^2\beta}\sum_{k=0}^{+\infty}(-1)^k\frac{(-1)^k e^{-\frac{k\pi}{\Delta}} - (-1)^{k+1}e^{-\frac{(k+1)\pi}{\Delta}}}{1+\frac{1}{\Delta^2}}$$
$$= (1+e^{-\frac{\pi}{\Delta}})\sum_{k=0}^{+\infty} e^{-\frac{k\pi}{\Delta}} = \frac{1+e^{-\frac{\pi}{\Delta}}}{1-e^{-\frac{\pi}{\Delta}}} = coth\Big(\frac{\pi}{2\Delta}\Big)$$

  3. $w(t) = \frac{l}{\Delta\mu}e^{-\frac{\beta}{2\mu}t}\big(\sin(\frac{\beta\Delta}{2\mu}t) + \Delta\cos(\frac{\beta\Delta}{2\mu}t)\big)H(t)$ which also changes sign on $\mathbb{R}_+$. We have not found a way to compute $\|w\|_1$ and write the result elegantly.

Appendix D

Appendix for Kronecker product

The Kronecker product between matrices (which is equal to the tensor product between matrices) is ubiquitous in the mathematical proofs of this thesis. Indeed, correlation-causation based learning in a neural network can be expressed using this operator which is written $\otimes$. It has interesting properties that we briefly describe below.

We refer to [Brewer 1978] and [Bernstein 2009] for an extensive introduction to the Kronecker product.

D.1 Kronecker product

D.1.1 Definition

Between matrices Let $\mathbf{A} \in \mathbb{R}^{n\times m}$ and $\mathbf{B} \in \mathbb{R}^{p\times q}$. The Kronecker product between $\mathbf{A}$ and $\mathbf{B}$ is a matrix of size $np \times mq$ defined by the partitioned matrix

$$\mathbf{A} \otimes \mathbf{B} \overset{def}{=} \begin{pmatrix} \mathbf{A}_{11}\mathbf{B} & \mathbf{A}_{12}\mathbf{B} & \dots & \mathbf{A}_{1m}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_{n1}\mathbf{B} & \mathbf{A}_{n2}\mathbf{B} & \dots & \mathbf{A}_{nm}\mathbf{B} \end{pmatrix}$$

Between vectors The definition above holds when $n = 1$, $m = 1$, $p = 1$ and/or $q = 1$ extending the definition to vectors. However, the most common way of considering the Kronecker product between two vectors assumes $m = 1$ and $p = 1$, i.e. $\mathbf{A}$ is a column and $\mathbf{B}$ is a row. In this case, $\mathbf{A} \otimes \mathbf{B} \in \mathbb{R}^{n\times q}$ and $\{\mathbf{A} \otimes \mathbf{B}\}_{ij} = \mathbf{A}_i\mathbf{B}_j$. It can also be reformulated as $\mathbf{A} \otimes \mathbf{B} = \mathbf{A} \cdot \mathbf{B}$ where $\cdot$ is the matrix multiplication.

Between operators We illustrate in appendix E, an extension of the Kronecker product to operators. Actually, it can be defined based on the caracteristic property D.1.8 of the Kronecker product.

D.1.2 Properties

D.1.2.1 Immediate properties

The Kronecker product is **associative**, **non-commutative**, **bilinear** and **distributive** with respect to the sum of matrices.

Proposition D.1.1.*If* $\mathbf{A} \in \mathbb{C}^{n\times m}, \mathbf{B} \in \mathbb{C}^{p\times q}$*then*

$$(\mathbf{A} \otimes \mathbf{B})^* = \mathbf{A}^* \otimes \mathbf{B}^*$$

Proposition D.1.2.*If* $\mathbf{A} \in \mathbb{R}^{n\times m}, \mathbf{B} \in \mathbb{R}^{p\times q}, \mathbf{C} \in \mathbb{R}^{m\times l}, \mathbf{D} \in \mathbb{R}^{q\times k}$ *, then*

$$(\mathbf{A} \otimes \mathbf{B}) \cdot (\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C}) \otimes (\mathbf{B} \cdot \mathbf{D})$$

Proposition D.1.3.*If* $\mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{B} \in \mathbb{R}^{m\times m}$ *are invertible, then* $\mathbf{A} \otimes \mathbf{B}$ *is invertible and*

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$$

Proposition D.1.4.*If* $\mathbf{A} \in \mathbb{R}^{n\times n}, \mathbf{B} \in \mathbb{R}^{m\times m}$ *are invertible, then* $\mathbf{A} \otimes \mathbf{B}$ *is invertible and*

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$$

Proposition D.1.5.*If* $\{\lambda_i \in \mathbb{C} : i = 1..n\}$ *(resp.* $\{\mu_j \in \mathbb{C} : j = 1..m\}$*) is the set of eigenvalues of* $\mathbf{A} \in \mathbb{R}^{n\times n}$ *(resp.* $\mathbf{B} \in \mathbb{R}^{m\times m}$*), then the eigenvalues of* $\mathbf{A} \otimes \mathbf{B}$ *are all the combinations* $\lambda_i\mu_j$*.*

D.1.2.2 Caracteristic property

It is well know that any linear operator on $\mathbb{R}^n$ can be described by a matrix (in fact several matrices due to the multiplicity of the basis of $\mathbb{R}^n$). Conversely, to any matrix can be associated a linear operator (in fact several operators due to the multiplicity of the basis of $\mathbb{R}^n$). Therefore, we might be interested by the operators associated to the Kronecker product $\mathbf{A} \otimes \mathbf{B}$. One of them turns out to be $\mathbf{X} \mapsto \mathbf{A}.\mathbf{X}.\mathbf{B}'$. It corresponds to the choice of a certain basis. Another basis leads to the operator $\mathbf{X} \mapsto \mathbf{B}'.\mathbf{X}.\mathbf{A}$. This is summarized the following propositions.

To get to this result it is convenient to introduce the "vectorize" operators.

Definition D.1.6.*The operator* $vec_r : \begin{array}{ccc} \mathbb{R}^{n\times m} & \rightarrow & \mathbb{R}^{nm} \\ \mathbf{A} & \mapsto & vec_r(\mathbf{A}) \end{array}$ *corresponds to the turning a matrix into a vector by concatenating the rows of the matrix in a single vector.*

Similarly we can define the operator vec_c *which corresponds to the concatenation of the columns of a matrix.*

Proposition D.1.7.

Lemma D.1.8.*Note* $\forall i \in \{1,..,4\},\ n_i \in \mathbb{N}^*$. *For* $\mathbf{A} \in \mathbb{R}^{n_1\times n_2}$ *and* $\mathbf{B} \in \mathbb{R}^{n_3\times n_4}$.

$$vec_r(\mathbf{A}.\mathbf{X}.\mathbf{B}) = (\mathbf{A}\otimes\mathbf{B}').vec_r(\mathbf{X})$$

$$vec_c(\mathbf{A}.\mathbf{X}.\mathbf{B}) = (\mathbf{B}'\otimes\mathbf{A}).vec_c(\mathbf{X})$$

Proof. To do so we introduce, the operator $F : \mathbf{X} \mapsto \mathbf{A}.\mathbf{X}.\mathbf{B}$. Then, substitute $\mathbf{X}$ by $\mathbf{E}_{ij} = (\delta_{i-k}\delta_{j-l})_{kl}$ one of the vectors of the canonical basis of $\mathbb{R}^{n_2\times n_3}$. Then observe that $F(\mathbf{E}_{ij}) = \mathbf{A}.\mathbf{E}_{ij}.\mathbf{B} = \mathbf{A}_{.i}\otimes\mathbf{B}_{j.}$ ($\mathbf{A}_{.i}$ is the ith columns of $\mathbf{A}$ and $\mathbf{B}_{j.}$ is the jth row of $\mathbf{B}$).

Consider the vector $\mathrm{vec}_c(\mathbf{X}) = \Big(\underbrace{\mathbf{X}_{11},\cdots,\mathbf{X}_{n_2 1}}_{\mathbf{X'}_{.1}},\underbrace{\mathbf{X}_{12},\cdots,\mathbf{X}_{n_2 2}}_{\mathbf{X'}_{.2}},\cdots,\underbrace{\cdots}_{\mathbf{X'}_{.n_3}}\Big)'$ made of the columns of $\mathbf{X}$. Similarly, define $\mathrm{vec}_r(\mathbf{X})$ by concatenating the rows of $\mathbf{X}$. Therefore,

$$\begin{aligned}
&\mathrm{vec}_c\big(F(\mathbf{X})\big) = \\
&= \Big(\underbrace{\mathrm{vec}_c(\mathbf{A}_{.1}\otimes\mathbf{B}_{1.}),\cdots,\mathrm{vec}_c(\mathbf{A}_{.n_2}\otimes\mathbf{B}_{1.})}_{\mathbf{B}'_{1.}\otimes\mathbf{A}},\cdots,\underbrace{\mathrm{vec}_c(\mathbf{A}_{.1}\otimes\mathbf{B}_{n_3.}),\cdots,\mathrm{vec}_c(\mathbf{A}_{.n_2}\otimes\mathbf{B}_{n_3.})}_{\mathbf{B}'_{n_3.}\otimes\mathbf{A}}\Big).\mathrm{vec}_c(\mathbf{X}) \\
&= (\mathbf{B}'\otimes\mathbf{A}).\mathrm{vec}_c(\mathbf{X}) \qquad (D.1)
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathrm{vec}_r\big(F(\mathbf{X})\big) = \\
&= \Big(\underbrace{\mathrm{vec}_r(\mathbf{A}_{.1}\otimes\mathbf{B}_{1.}),\cdots,\mathrm{vec}_r(\mathbf{A}_{.1}\otimes\mathbf{B}_{n_3.})}_{\mathbf{A}_{.1}\otimes\mathbf{B}'},\cdots,\underbrace{\mathrm{vec}_r(\mathbf{A}_{.n_2}\otimes\mathbf{B}_{1.}),\cdots,\mathrm{vec}_r(\mathbf{A}_{.n_2}\otimes\mathbf{B}_{n_3.})}_{\mathbf{A}_{.n_2}\otimes\mathbf{B}'}\Big).\mathrm{vec}_r(\mathbf{X}) \\
&= (\mathbf{A}\otimes\mathbf{B}').\mathrm{vec}_r(\mathbf{X}) \qquad (D.2)
\end{aligned}$$

□

Lemma D.1.9.*Note* $\forall i \in \{1,..,4\},\ n_i \in \mathbb{N}^*$. *For* $\mathbf{A} \in \mathbb{R}^{n_1\times n_3}$ *and* $\mathbf{B} \in \mathbb{R}^{n_2\times n_4}$..

$$vec_r(\mathbf{A}.\mathbf{X}'.\mathbf{B}) = (\mathbf{A}\otimes\mathbf{B}').vec_r(\mathbf{X})$$

$$vec_c(\mathbf{A}.\mathbf{X}'.\mathbf{B}) = (\mathbf{B}'\otimes\mathbf{A}).vec_c(\mathbf{X})$$

Proof. To do so we introduce, the operator $F : \mathbf{X} \mapsto \mathbf{A}.\mathbf{X}'.\mathbf{B}$. Then, substitute $\mathbf{X}$ by $\mathbf{E}_{ij} = (\delta_{i-k}\delta_{j-l})_{kl}$ one of the vectors of the canonical basis of $\mathbb{R}^{n_2 \times n_3}$. Then observe that $F(\mathbf{E}_{ij}) = \mathbf{A}.\mathbf{E}'_{ij}.\mathbf{B} = \mathbf{A}_{.j} \otimes \mathbf{B}_{i.}$ ($\mathbf{A}_{.j}$ is the jth columns of $\mathbf{A}$ and $\mathbf{B}_{i.}$ is the ith row of $\mathbf{B}$).

Consider the vector $\text{vec}_c(\mathbf{X}) = \Big(\underbrace{\mathbf{X}_{11}, \cdots, \mathbf{X}_{n_2 1}}_{\mathbf{X}'_{.1}}, \underbrace{\mathbf{X}_{12}, \cdots, \mathbf{X}_{n_2 2}}_{\mathbf{X}'_{.2}}, \cdots, \underbrace{\cdots}_{\mathbf{X}'_{.n_3}}\Big)'$ made of the columns of $\mathbf{X}$. Similarly, define $\text{vec}_r(\mathbf{X})$ by concatenating the rows of $\mathbf{X}$. Therefore,

$$\text{vec}_c\big(F(\mathbf{X})\big) =$$

$$= \Big(\underbrace{\text{vec}_c(\mathbf{A}_{.1} \otimes \mathbf{B}_{1.}), \cdots, \text{vec}_c(\mathbf{A}_{.1} \otimes \mathbf{B}_{n_2.})}_{\mathbf{B}'_{1.} \otimes \mathbf{A}}, \cdots, \underbrace{\text{vec}_c(\mathbf{A}_{.n_3} \otimes \mathbf{B}_{1.}), \cdots, \text{vec}_c(\mathbf{A}_{.n_3} \otimes \mathbf{B}_{n_2.})}_{\mathbf{B}'_{n_2.} \otimes \mathbf{A}}\Big).\text{vec}_c(\mathbf{X}$$

$$= (\mathbf{B}' \otimes \mathbf{A}).\text{vec}_c(\mathbf{X}) \quad \text{(D.3)}$$

Similarly,

$$\text{vec}_r\big(F(\mathbf{X})\big) =$$

$$= \Big(\underbrace{\text{vec}_r(\mathbf{A}_{.1} \otimes \mathbf{B}_{1.}), \cdots, \text{vec}_r(\mathbf{A}_{.n_3} \otimes \mathbf{B}_{1.})}_{\mathbf{A} \otimes \mathbf{B}'_{1.}}, \cdots, \underbrace{\text{vec}_r(\mathbf{A}_{.1} \otimes \mathbf{B}_{n_3.}), \cdots, \text{vec}_r(\mathbf{A}_{.n_3} \otimes \mathbf{B}_{n_2.})}_{\mathbf{A} \otimes \mathbf{B}'_{n_2.}}\Big).\text{vec}_r(\mathbf{X}$$

$$= (\mathbf{A} \otimes \mathbf{B}').\text{vec}_r(\mathbf{X}) \quad \text{(D.4)}$$

□

D.2 Kronecker sum

Definition D.2.1.*Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{m \times m}$, then the Kronecker sum between $\mathbf{A}$ and $\mathbf{B}$ is written $\mathbf{A} \oplus \mathbf{B}$ and is defined by*

$$\mathbf{A} \oplus \mathbf{B} \overset{def}{=} \mathbf{A} \otimes I_m + I_m \otimes \mathbf{B}$$

It is **associative** and its eigenvalues can be described as follow

Proposition D.2.2.*If $\{\lambda_i \in \mathbb{C} : i = 1..n\}$ (resp. $\{\mu_j \in \mathbb{C} : j = 1..m\}$) is the set of eigenvalues of $\mathbf{A} \in \mathbb{R}^{n \times n}$ (resp. $\mathbf{B} \in \mathbb{R}^{m \times m}$), then the eigenvalues of $\mathbf{A} \oplus \mathbf{B}$ are all the combinations $\lambda_i + \mu_j$.*

D.3 Sylvester equations

Sylvester equations are linear matrix equations. As opposed to the real case were the commutativity of the real product leads to a simple solution, these equations are much more difficult to solve in a matrix case.

Definition D.3.1. *Given $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{m\times m}$ and $\mathbf{C} \in \mathbb{R}^{n\times m}$, the following equation is a* ***Sylvester equation*** *(where $\mathbf{X} \in \mathbb{R}^{n\times m}$ is the unknown variable)*

$$\mathbf{A}.\mathbf{X} + \mathbf{X}.\mathbf{B} + \mathbf{C} = 0 \tag{D.5}$$

If $\mathbf{C}$ is Hermitian and $\mathbf{B} = \mathbf{A}^*$ then this equation is known as the **continuous Lyapunov equation**.

D.3.1 Link with Kronecker sum

Actually, the explicit solution of a Sylvester equation can be written using Kronecker sums. Indeed, apply the vectorize operator (with respect to the rows for instance), defined in D.1.6, to the Sylvester equation. This gives

$$(\mathbf{A} \otimes I_m + I_n \otimes \mathbf{B}').\text{vec}_r(\mathbf{X}) = -\text{vec}_r(C)$$

Therefore, provided $\mathbf{A} \oplus \mathbf{B}'$ is invertible the solution of a Sylvester equation is

$$\mathbf{X}^* = -\text{vec}_r^{-1}\Big((\mathbf{A} \oplus \mathbf{B}').\text{vec}_r(\mathbf{X})\Big)$$

The main problem of this formula is that it may be computationally heavy to compute the inverse of the Kronecker sum.

D.3.2 A new way of writing the solutions

Here, we propose an original explicit solution for Sylvester equations D.5. Indeed, provided the sum converges, the solution can be written

$$\mathbf{X}^* = -\sum_{k=0}^{+\infty} (-\mathbf{A})^k.\mathbf{C}.\mathbf{B}^{-(k+1)}$$

Indeed, to prove this it is sufficient to observe that $\mathbf{A}.\mathbf{X}^* + \mathbf{X}^*.\mathbf{B} = -\mathbf{C}$ where $\mathbf{X}^*$ is defined by the equation above.

$$\mathbf{A}.\mathbf{X}^* + \mathbf{X}^*.\mathbf{B} = \sum_{k=0}^{+\infty} (-\mathbf{A})^{k+1}.\mathbf{C}.\mathbf{B}^{-(k+1)} - \sum_{k=0}^{+\infty} (-\mathbf{A})^k.\mathbf{C}.\mathbf{B}^{-k} = -\mathbf{C}$$

We guessed this formula from a deep inspection of the Kronecker sums and products defined above. Actually, the serie has to converge for $\mathbf{X}^*$ to be a solution and it not an easy task to have a subtle description of all the converging cases. Therefore, we introduce the following rigorous lemma which is sufficient for the computations done in this thesis.

We believe this result may interest theoreticians in different field and that it deserves to be more developped. Indeed, it corresponds to finding an explicit formula for the inverse of a Kronecker sum. Therefore, exploring this heuristic result is a technical perspective of this thesis.

Lemma D.3.2. *The solution of the following Sylvester equation*

$$(\mathbf{L}-\mathbf{W}).\mathbf{X}+\mathbf{X}.(\mathbf{L}-\mathbf{W})'+\mathbf{D}=0$$

where $\mathbf{L}=lId$ *and* $0<\mathbf{W}<\mathbf{L}$ *is*

$$\mathbf{X}=-\sum_{k=0}^{+\infty}\mathbf{W}^k.\mathbf{D}.(2\mathbf{L}-\mathbf{W}')^{-(k+1)}$$

Proof. First, observe that if $\{|\lambda| : \lambda \text{ eigenvalue of } \mathbf{W}\}\in]0,l[$ and $\mathbf{W}>0$ then $|||\mathbf{W}|||\ |||(2\mathbf{L}-\mathbf{W})^{-1}|||<1$. Therefore, $\mathbf{X}$ is well defined by equation (D.3.2).

Observe that $(2\mathbf{L}-\mathbf{W}')^{-1}.(\mathbf{L}-\mathbf{W})'=Id-L.(2\mathbf{L}-\mathbf{W}')^{-1}$. Then assuming $\mathbf{X}$ is defined by equation (D.3.2) then

$$\begin{aligned}&(\mathbf{L}-\mathbf{W}).\mathbf{X}+\mathbf{X}.(\mathbf{L}-\mathbf{W})'\\&\quad=-\Big(\mathbf{L}.\sum_{k=0}^{+\infty}\mathbf{W}^k.\mathbf{D}.(2\mathbf{L}-\mathbf{W}')^{-(k+1)}-\mathbf{W}.\sum_{k=0}^{+\infty}\mathbf{W}^k.\mathbf{D}.(2\mathbf{L}-\mathbf{W}')^{-(k+1)}\\&\qquad+\sum_{k=0}^{+\infty}\mathbf{W}^k.\mathbf{D}.(2\mathbf{L}-\mathbf{W}')^{-k}-\mathbf{L}.\sum_{k=0}^{+\infty}\mathbf{W}^k.\mathbf{D}.(2\mathbf{L}-\mathbf{W}')^{-(k+1)}\Big)\\&\quad=-\Big(\sum_{k=0}^{+\infty}\mathbf{W}^k.\mathbf{D}.(2\mathbf{L}-\mathbf{W}')^{-k}-\sum_{k=0}^{+\infty}\mathbf{W}^{k+1}.\mathbf{D}.(2\mathbf{L}-\mathbf{W}')^{-(k+1)}\Big)=-\mathbf{D}\end{aligned}$$

□

Appendix E

Appendix for the solutions of linear SPDEs with delays

Contents

abstract Based on the analysis of a certain class of linear operators on a Banach space, we provide a closed form expression for the solutions of certain linear partial differential equations with non-autonomous input, time delays and stochastic terms, which takes the form of an infinite series expansion.

Keywords: Linear, delay, stochastic, non-autonomous, partial differential equations, series expansion, Kronecker product.

Introduction

Linear differential systems are ubiquitous in pure and applied mathematics, either as models, approximations, but also because the stability of solutions of nonlinear differential systems reduces to the study of linear systems. Such systems might include stochastic terms (see [Mao 1997]), temporal delays (see [Hale and Lunel 1993]), and also encompass the case of partial differential equations. Apart from the simplest linear finite-dimensional differential equations, finding closed forms expressions for the solutions of general linear differential systems is very complex. In this paper, based on the treatment of evolution equations as algebraic equations in a suitable Banach space, we propose a closed-form expression for the solution of linear, non-autonomous, stochastic, time-delayed partial differential systems. Application of this framework to several classical examples such as the delayed Ornstein-Uhlenbeck process or the stochastic heat equation are developed in sections E.2.3 and E.2.4. This expression is especially useful to understand the dynamics of weakly connected linear learning neural networks, problem which motivated the development of this more general framework and which is the topic of this thesis.

E.1 Framework and General Result

The framework we develop here is based on extending notions of matrix calculus to infinite dimensional spaces. The linearity of the equation motivates to extend some finite-dimensional linear algebra and matrix concepts to infinite-dimensional spaces.

We consider in the manuscript linear equations in a Banach space $\mathcal{C}$ of real functions of time t and a variable $x \in E$, called *space* variable, where E can either be a finite set $\{1, \dots, N\}$ (in which case $\mathcal{C}$ is equivalent to the space of $\mathbb{R}^N$-valued functions), countable or continuous, typically $\mathbb{R}$, in which case $\mathcal{C}$ is a space of two-variables functions. The particular problem under consideration governs the choice of the space $\mathcal{C}$, in particular including regularity or integrability properties (typically $\mathcal{C}$ is a L^p or a Sobolev space). Similarly to a matrix notation, we denote the value of $X \in \mathcal{C}$ at $(x, t) \in E \times \mathbb{R}$ by X_{xt}.

Let $\mathcal{E}$ denote the space of bounded linear operators on $\mathcal{C}$. We are interested in solving equations of type $\mathcal{L}X = B$ where $\mathcal{L} \in \mathcal{E}$ (this operator may involve differentials in time and/or space) and $B \in \mathcal{C}$. We will restrict the study to a class of operators of a particular form we now detail. To this end, we introduce two kinds of linear operators on $\mathcal{C}$: the *space* operators L acting on the first (space) variable, i.e. linear operators on $\mathbb{R}^E$. If E is finite, this set is reduced to the matrices. If E is equal to $\mathbb{R}^d$, it contains all the linear operators acting on functions of the space variable, in particular, under suitable regularity conditions, integral or differential operators. The action of the space operators L on a function $X \in \mathcal{C}$ is denoted $L \cdot X$ (acting on the left). The *time* operators essentially act on the second (time) variable, and the transform might depend on the space variable x. In other words, these transforms $\mathcal{R}$ can be represented by a family of operators $(\mathcal{R}_x, x \in E)$ such that for any x, $\mathcal{R}_x$ is a linear operator on $L^2(\mathbb{R})$. The action of a time operator $\mathcal{R}$ on $X \in \mathcal{C}$ is written $X \cdot \mathcal{R}$ (acting on the right). In the paper, we will mainly be interested in diagonalizable time operators. Diagonal operators in the time domain are operators $\mathcal{R}$ whose action can be written in the form $(X \cdot \mathcal{R})_{xt} = r(x,t)X_{x,t}$. This class includes for instance all linear differential time operators, which are diagonalizable in the Fourier basis. Another class of time operators we will be considering is the class $\mathcal{C}$ of convolution operators with respect to time. Given a finite measure g of $\mathbb{R}$, the convolution operator $\mathcal{T}_g \in \mathcal{C}$ associated with g is defined as $\left(X \cdot \mathcal{T}_g\right)_{xt} = \int_{-\infty}^{\infty} X_{x(t-s)} dg(s)$. Such operators are generalizations of Toeplitz matrices generated by g, with, loosely speaking, infinitely many rows and columns. An important property of the convolution operators is that they are diagonal in the Fourier basis.

For L a space operator and $\mathcal{R}$ a time operator, we define the Kronecker product $L \otimes \mathcal{R}$ as the mixed operator of $\mathcal{E}$ such that $\big(L \otimes \mathcal{R}\big)(X) = L \cdot (X \cdot \mathcal{R})$. Note that the product becomes associative when $\mathcal{R}$ is a convolution operator which will be the case in section E.2. This definition extends the property of vectorization of the Kronecker product of matrices in linear algebra (see e.g. [Brewer 1978]).

The main technical result of the paper is given in the following:

Proposition E.1.1. *Let $\mathcal{L} = A\otimes\mathcal{B}+Id_{\mathcal{C}}\otimes\mathcal{D}$ be a linear operator, for A a space operator and $\mathcal{B}$, $\mathcal{D}$ co-diagonalizable time operators, with $\mathcal{B}$ invertible. For the sake of simplicity, we assume that they are diagonal in the natural time basis, and denote for $x \in E$, $\mathcal{B}_x = diag_{t\in\mathbb{R}}\Big(b(x,t)\Big)$ and $\mathcal{D}_x = diag_{t\in\mathbb{R}}\Big(d(x,t)\Big)$. We assume that $\inf_{x,t}|b(x,t)| > 0$ and the spectral condition:*

$$\exists\, l \in \mathbb{R}^* \text{ such that } \lambda \overset{def}{=} \frac{\|W\|}{\inf_{x,t}\Big\{\Big|l - \frac{d(x,t)}{b(x,t)}\Big|\Big\}} < 1, \tag{E.1}$$

where $W \overset{def}{=} l\, Id_{\mathbb{R}^E} + A$ and $\|W\| = \sup_{X\neq 0}\frac{\|W\cdot X\|}{\|X\|}$ is the operator norm. Then $A\otimes\mathcal{B}+Id_{\mathcal{C}}\otimes\mathcal{D}$ is invertible and its inverse reads:

$$\Big(A\otimes\mathcal{B}+Id_{\mathcal{C}}\otimes\mathcal{D}\Big)^{-1} = -\sum_{k=0}^{+\infty} W^k \otimes diag_{t\in\mathbb{R}}\Big(\frac{1}{b(x,t)(l-\frac{d(x,t)}{b(x,t)})^{k+1}}\Big). \tag{E.2}$$

Remark *The spectral condition is merely a technical sufficient condition for the convergence of the series. The relatively formal setting and assumptions will become clearer in the applications, section E.2.*

Proof. The direct introduction of the inverse can appear artificial at first sight. However, this formula is a natural extension of the discrete-time case where direct linear algebra and Kronecker products calculations quite simply provide a closely related expression the interested reader can readily derive.

In order to prove the proposition, we first need to prove that the operator indeed exists, and that it constitutes the inverse of $\mathcal{L}$. It is easy to show that under the assumption of the proposition that the sequence of operators in $\mathcal{E}$ defined by: $M_N \overset{def}{=} -\sum_{k=0}^{N} W^k \otimes diag_{t\in\mathbb{R}}\Big(\frac{1}{b(x,t)(l-\frac{d(x,t)}{b(x,t)})^{k+1}}\Big)$ constitutes a Cauchy sequence in $\mathcal{E}$. Since $\mathcal{C}$ is a Banach space, so is $\mathcal{E}$, and hence the sequence $(M_n)_n$ converges. The limit of this sequence is our inverse candidate, and is denoted as the infinite series (E.2).

In order to prove that this limit is indeed the inverse of $\mathcal{L}$, we compute the limit of $(M_N\circ\mathcal{L})X$ (or similarly $(\mathcal{L}\circ M_N)X$) for a given $X\in\mathcal{C}$. It is easy to show, developing the series, that we have:

$$((M_N\circ\mathcal{L})X)_{xt} = -\sum_{k=0}^{N} -W^k\cdot\frac{X_{.,t}}{\Big(l-\frac{d(.,t)}{b(.,t)}\Big)^k} + W^{k+1}\cdot\frac{X_{.,t}}{\Big(l-\frac{d(.,t)}{b(.,t)}\Big)^{k+1}} = X_{xt} - W^{N+1}\cdot\frac{X_{.,t}}{\Big(l-\frac{d(.,t)}{b(.,t)}\Big)^{N+1}}$$

where $Y_{.,t}$ for $Y\in\mathcal{C}$ denotes the application $E\mapsto\mathbb{R}$ such that $Y_{.,t}(x) = Y_{xt}$. Here again, the assumptions of the proposition ensures that the second term vanishes as N goes to infinity. □

E.2 Application to solving linear time-delayed Stochastic Partial Differential Equations

In this section we make explicit the use of the inversion formula (E.2) in the case of linear delayed, stochastic, partial differential equations. Several examples with different convolution operators will illustrate the main result of the section stated in theorem E.2.1.

E.2.1 General Result

Let $\mathcal{X}$ be a Hilbert space, typically $\mathbb{R}^n$ for $n \in \mathbb{N}$, $L^2(\mathbb{R}^n)$ or a Sobolev space of applications on $\mathbb{R}^n$. We consider a probability space $(\mathcal{O}mega, \mathcal{F}, \mathbb{P})$ satisfying the usual conditions and B a standard adapted $\mathcal{X}$-Brownian motion (for the existence and properties of this object in infinite-dimensional spaces, see [Da Prato and Zabczyk 1992, Chapter 4]). We aim at solving the non-autonomous time-delayed stochastic differential equation:

$$\begin{cases} dX = \big(A\cdot(X*g)+I\big)\,dt + \Sigma\cdot dB \\ \qquad X_{|\mathbb{R}_-} = \zeta_0 \in L^2_{\mathcal{X}}(\mathbb{R}^*_-) \end{cases} \tag{E.3}$$

with $\Sigma : \mathcal{X} \mapsto \mathcal{X}$ linear, $I \in C(\mathbb{R}_+, \mathcal{X})$ an external input, g a finite measure of the real line supported on $\mathbb{R}_+$, i.e. a causal measure, and $*$ denoting the convolution. Existence and uniqueness of weak solutions for such equations is ensured, see e.g. [Mao 1997, Da Prato and Zabczyk 1992]. We consider the case where the system has a unique strong solution. In the case where $\mathcal{X} = \mathbb{R}^n$, this occurs under the assumptions of the section, see e.g. [Mao 1997, Chapter 5], and in the infinite-dimensional case, we need to assume that B is a genuine Wiener process (i.e. the trace of the covariance matrix is finite, and the initial condition is in the domain of A, see [Da Prato and Zabczyk 1992]). The solution of this stochastic differential equation at time $t \in \mathbb{R}_+$ is defined by the integral equation $X(t) = \zeta_0(0) + \int_0^t \big(A\cdot(X*g)(s) + I(s)\big)\,ds + \int_0^t \Sigma \cdot dB$ and $X_{|\mathbb{R}_-} = \zeta_0$.

This problem can be set in the framework described in section E.1 using a transformation inspired by the classical Fourier transform of the solution in the time domain. To perform this transformation rigorously in our particular stochastic setting, we stop our processes at a finite time $\tau > 0$. We define $X_\tau : t \in \mathbb{R} \to \mathbb{1}_{[0,\tau]}(t)X(t)$ the restriction of X to the compact support $[0,\tau]$

and null elsewhere. Similarly, define $I_\tau = \mathbb{1}_{[0,\tau]} I$ and $dB_\tau = \mathbb{1}_{[0,\tau]} dB$. We have:

Theorem E.2.1. *For all $\tau \in \mathbb{R}_+$, choose $l \in \mathbb{R}^*$ and W a space operator such that $W = l\, Id_{\mathcal{C}} + A$. If the spectral condition (E.1) is satisfied, i.e. in the present case $\|W\| < \inf_\xi\{\left|l + \frac{2i\pi\xi}{\hat{g}(\xi)}\right|\}$, then the solution of equation (E.3) is given by*

$$X_\tau = \sum_{k=0}^{+\infty} W^k \cdot \Big(\zeta_0(0)\delta_0 + \widetilde{I}_\tau + \Sigma \cdot dB_\tau\Big) \cdot \mathcal{U} \cdot \mathcal{V}^k \tag{E.4}$$

where $\mathcal{U} = \mathcal{F} \cdot diag_{\xi\in\mathbb{R}}\Big(\frac{1}{l\hat{g}(\xi)+2i\pi\xi}\Big) \cdot \mathcal{F}^{-1}$, $\mathcal{V} = \mathcal{F} \cdot diag_{\xi\in\mathbb{R}}\Big(\frac{\hat{g}(\xi)}{l\hat{g}(\xi)+2i\pi\xi}\Big) \cdot \mathcal{F}^{-1}$ and $\widetilde{I}_\tau = I_\tau + A \cdot (\zeta_0 * g)$. The notation $\big(dB_\tau \cdot \mathcal{U}\big)_{xt}$ stands for the stochastic integral $\int_0^{min(t,\tau)} \mathcal{U}_{st} dB(s)$ which is square integrable on $[0, \tau[$. **Remark:** *The convergence of the series (E.4) occurs as soon as the spectral condition (E.1) is satisfied on the subspace spanned by $W^k \cdot (\zeta_0(0)\delta_0 + \widetilde{I}_\tau + \Sigma \cdot dB_\tau) \cdot \mathcal{U} \cdot \mathcal{V}^k$.*

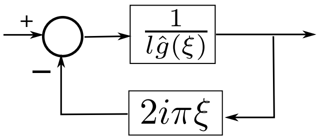
Proof. First, note that $A \cdot (X * g) = A \cdot (X_\tau * g) + A \cdot (\zeta_0 * g)$ yielding the equation on X_τ: $dX_\tau = \big(A \cdot (X_\tau * g) + \widetilde{I}_\tau\big)\, dt + \Sigma \cdot dB_\tau$. Thus, the initial condition on X acts as an external input on X_τ. In the deterministic finite-dimensional case, it is well known that differential operators are diagonal in the Fourier basis. Based on this result, we introduce the Fourier transform $\mathcal{F}$ of equation (E.3) for a fixed $\omega \in \mathcal{O}mega$. As mentioned, for almost all $\omega \in \mathcal{O}mega$, the processes involved are bounded, hence the function of time, on the compact interval $[0, \tau]$, is square integrable in time. Let $Z^\xi : t \in \mathbb{R} \to e^{-2i\pi t\xi} X_\tau(t)$ for $\xi \in \mathbb{R}$ the Fourier variable and X is the unique solution of equation E.3. Itô formula yields for $t < \tau$

$$dZ^\xi(t) = \Big(-2i\pi\xi Z^\xi(t) + A \cdot \big(Z^\xi * g\big)(t) + e^{-2i\pi t\xi} \widetilde{I}_\tau(t)\Big) dt + e^{-2i\pi t\xi} \Sigma \cdot dB_\tau(t) \tag{E.5}$$

Let us denote by $\hat{X}_\tau : \xi \in \mathbb{R} \to \int_0^\tau Z^\xi(s) ds$ the Fourier transform of X_τ and $\hat{I}_\tau : \xi \to \int_0^\tau e^{-2i\pi t\xi} \widetilde{I}_\tau(t) dt$. The process $\hat{B}_\tau$ is the well-defined stochastic integral $\int_0^\tau e^{-2i\pi t\xi} dB(t)$. The integral form of equation (E.5), using the fact that the convolution is diagonal in the Fourier basis, denoting $\hat{D} = diag_{\xi\in\mathbb{R}}\big(-2i\pi\xi\big)$ and $\hat{\mathcal{G}} = diag_{\xi\in\mathbb{R}}\big(\hat{g}(\xi)\big)$, leads to the functional equation:

$$Z^{\cdot}(\tau) - Z^{\cdot}(0) = A \cdot \hat{X}_\tau \cdot \hat{\mathcal{G}} + \hat{X}_\tau \cdot \hat{D} + \hat{I}_\tau + \Sigma \cdot \hat{B}_\tau$$

Applying proposition E.1.1 for a fixed $\omega \in \mathcal{O}mega$ where $\mathcal{C}$ is the set of square



integrable functions on $[0,\tau[$ which is a Banach space, we obtain:

$$\hat{X}_\tau = \sum_{k=0}^{+\infty} W^k \cdot \Big(-Z^{\cdot}(\tau) + Z^{\cdot}(0) + \hat{I}_\tau + \Sigma \cdot \hat{B}_\tau\Big) \cdot diag_{\xi\in\mathbb{R}}\Big(\frac{1}{\hat{g}(\xi)(l+\frac{2i\pi\xi}{\hat{g}(\xi)})^{k+1}}\Big) \quad \text{(E.6)}$$

We now take the inverse Fourier transform of this expression by applying the time operator $\mathcal{F}^{-1}$. First of all, we perform the inversion on the terms $\hat{I}_\tau = \widetilde{I}_\tau \cdot \mathcal{F}$. It is easy to show that $\hat{I}_\tau \cdot diag(\frac{1}{\hat{g}(\xi)(l+\frac{2i\pi\xi}{\hat{g}(\xi)})^{k+1}}) \cdot \mathcal{F}^{-1} = \widetilde{I}_\tau \cdot \mathcal{U} \cdot \mathcal{V}^k$. Similarly, the term $\Big(\hat{B}_\tau \cdot diag(\frac{1}{\hat{g}(\xi)(l+\frac{2i\pi\xi}{\hat{g}(\xi)})^{k+1}}) \cdot \mathcal{F}^{-1}\Big)_{.t}$ can be written $dB_\tau \cdot \mathcal{U} \cdot \mathcal{V}^k$.

Moreover, for $x \in \{0,\tau\}$ an easy computation shows that $\Big(Z^{\cdot}(x) \cdot diag(\frac{1}{\hat{g}(\xi)(l+\frac{2i\pi\xi}{\hat{g}(\xi)})^{k+1}}) \cdot \mathcal{F}^{-1}\Big)_{.t} = \big(X(x)\delta_x\big) \cdot \mathcal{U} \cdot \mathcal{V}^k$. Furthermore, the operators $\mathcal{U}$ and $\mathcal{V}$ are causal, i.e. if Y has a support $\subset [c,+\infty[$ then $Y \cdot \mathcal{U} \cdot \mathcal{V}^k$ also has a support $\subset [c,+\infty[$. Indeed, $\hat{u} : \xi \mapsto \frac{1}{l\hat{g}(\xi)+2i\pi\xi}$ corresponds to the transfer function of a closed loop filter shown on the right, and hence $\mathcal{U}$ is clearly causal since g is. $\mathcal{V}$ is also causal as the convolution of g and $\mathcal{U}$. This implies that the contribution of $Z(\tau)$ vanishes in equation (E.6) since it has its support in $[\tau,\infty]$.

□

E.2.2 Computational Remarks

Truncations of the formula (E.4) provides approximations of the solution of system (E.3). We observe that the smaller the difference between the spatial operator A and a multiple of the identity, the more accurate a truncation of this expansion. In other words, this expansion is particularly useful if A is a small perturbation of the (scaled) identity (e.g. the case of weakly connected linear neural networks).

This representation allows development of new numerical schemes for the simulations of the solutions of system (E.3). For simplicity, consider the case where $E = \{1,\cdots,n\}$. To approximate the solution over the interval $[0,\tau]$ define a time step Δt an number of points $T = \tau/\Delta t \in \mathbb{N}$ and replace $\mathcal{U}$ and $\mathcal{V}$ by the Toeplitz square matrices $\tilde{\mathcal{U}}$ and $\tilde{\mathcal{V}}$, generated by $i \in \{0,\cdots,T-1\} \mapsto \int_{i\Delta t}^{(i+1)\Delta t} u(s)ds$, where u is the function generating $\mathcal{U}$ (and similarly for $\mathcal{V}$). The number of operations needed is $\mathcal{O}((k+1)nT(n+\ln T))$ since the product with a Toeplitz matrix, as a convolution, has a cost $\mathcal{O}(T\ln T)$. This scheme has a first order accuracy, $\mathcal{O}(dt^{\gamma}+dx+\lambda^{k+1})$ where γ is equal to 1 for deterministic equations or $\frac{1}{2}$ if stochastic. In comparison, the Euler-Maruyama method has a complexity of $\mathcal{O}(T\big(n^2 + n\,\frac{\theta}{dt}ln(\frac{\theta}{dt})\big))$ where θ is the support of g and an accuracy of $\mathcal{O}\big(dt^{\gamma} + dx\big)$, comparable to the expansion method in both aspects.

Two interesting advantages of the expansion over Euler-like methods are that (i) it is parallelizable and (ii) it appears to be numerically very stable, i.e. large Δt do not lead to a diverging scheme.

E.2.3 Examples

Let us now treat some classical problems that are solved in the present framework.

- **Ornstein-Uhlenbeck process:** The simplest example is the Ornstein-Uhlenbeck process with no delays (i.e. $g = \delta_0$). In that case, $\hat{g} = 1$, and therefore, for all $l \in \mathbb{R}$, $\inf_{\xi}\{|l + 2i\pi\xi|\} = |l|$ and the expansion is valid if there exists $l \in \mathbb{R}^*$ such that $\|l + A\| < |l|$, i.e. for any operator A whose spectrum is bounded and entirely contained in the left or right half plane. For negative matrices A (i.e. $l > 0$) $\mathcal{T}_h = \mathcal{U} = \mathcal{V}$ is a Toeplitz operator generated by the function $h : z \to e^{-lz}H(z)$ with H the Heaviside function. Therefore, the solution of $\dot{X} = A\cdot X + I + \Sigma\cdot dB$ can be written as $X_\tau = \sum_{k=0}^{+\infty} W^k\cdot(X_0\delta_0 + I_\tau + \Sigma\cdot dB_\tau)\cdot\mathcal{T}_h^{k+1}$. If $A \propto Id_{\mathcal{C}}$ (e.g. in one dimension), the terms for $k > 0$ vanish in the above equation and we get the simple well-known expression $X_\tau = (X_0\delta_0 + I_\tau + dB_\tau) * h$.

- **Exponentionaly distributed delays:** Let us now treat the case $g : x \mapsto \beta e^{-\beta x}H(x)$. In this case, $\hat{g}(\xi) = \frac{\beta}{\beta+2i\pi\xi}$. Therefore, $\frac{2i\pi\xi}{\hat{g}(\xi)} = -2\pi(\frac{2\pi}{\beta}\xi^2 - i\xi)$ which corresponds to the red curve in the left picture of figure E.1. Operators A satisfying the spectral condition E.1 are the ones whose spectrum is contained in an open ball centered at $-l$ that does not intersect the red curve (blue disks of in figure E.1).

  When A is negative, the operators $\mathcal{U}$ and $\mathcal{V}$ can be made completely explicit. Indeed, observing that:
  $\frac{\hat{g}(\xi)}{l\hat{g}(\xi)+2i\pi\xi} = \frac{\beta}{(\frac{\beta}{2}+2i\pi\xi)^2 - \frac{\beta^2\Delta^2}{4}} = \frac{\beta}{(\beta\frac{1+\Delta}{2}+2i\pi\xi)(\beta\frac{1-\Delta}{2}+2i\pi\xi)}$ with $\Delta = \sqrt{1-4l/\beta}$,

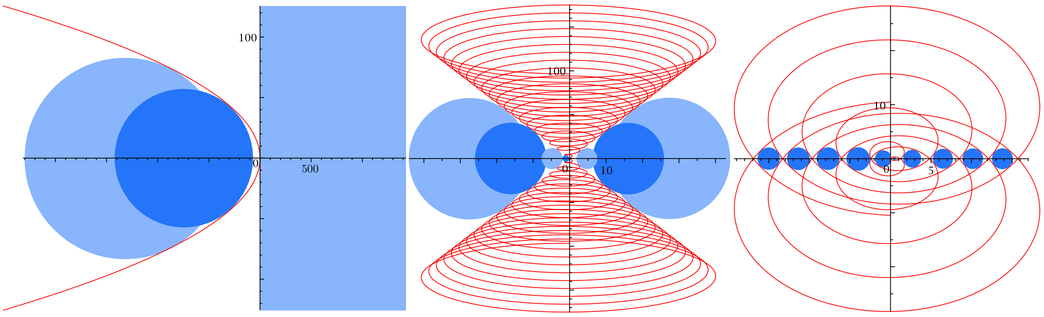


Figure E.1: The three different pictures correspond to different time-convolution kernels g. The red curves are the parametric plots of $\xi \in \mathbb{R} \mapsto \frac{2i\pi\xi}{\hat{g}(\xi)} \in \mathbb{C}$ and the blue balls are examples of sets within which the eigenvalues of the space operator A need to live for the expansion to be well-defined. The eigenvalues have to be contained in a single ball. The center of each ball is $-l$, for different $l \in \mathbb{R}$. To satisfy the spectral condition (E.1) the balls cannot intersect the red lines. (left) Exponentialy distributed delays with $\beta = 2\pi$ and $\xi \in [-20, 20]$. (middle) Single delay with $\alpha = 2$ and $\theta = 1$ and $\xi \in [-20, 20]$ (right) Single delay with $\alpha = 0.3$, $\theta = 1$ and $\xi \in [-5, 5]$

the operator $\mathcal{V}$ is the convolution operators generated by $\beta\big(h_- * h_+\big)$ with $h_\pm : t \mapsto e^{-\beta\frac{1\pm\Delta}{2}t}H(t)$. Similarly $\mathcal{U}$ is generated by $\beta\big(h_- * h_+ + \frac{1}{\beta}h'_- * h_+\big)$. Even more explicitly, for $\beta > 4l$, $\mathcal{V}$ is generated by $t \mapsto \frac{2}{\Delta}e^{-\frac{\beta}{2}t}sh\big(\frac{\beta\Delta}{2}t\big)$ and $\mathcal{U}$ by $t \mapsto \frac{1}{\Delta}e^{-\frac{\beta}{2}t}\Big(\, sh\big(\frac{\beta\Delta}{2}t\big) + \Delta ch\big(\frac{\beta\Delta}{2}t\big)\Big)$. When $4l > \beta$, a similar result holds replacing the hyperbolic functions ch and sh by cos and sin.

- **Single fixed delay:** For $g = \delta_0 + \alpha\delta_\theta$, we have $\hat{g}(\xi) = 1 + \alpha e^{2i\pi\theta\xi}$. The convergence domain of the expansion (condition E.1) is shown in the middle and right pictures of figure E.1, for two different $\alpha \in \mathbb{R}_+$. The red curve seems to be living on the 2-dimensional projection of a simple 3-dimensional cone of revolution whose section is a circle. In that case, it appears quite difficult to express $\mathcal{U}$ and $\mathcal{V}$ in a simple form, though their Fourier transform is explicit.

Remark: As illustrated in the previous example, a procedure to find the constant l such that the expansion converges consists in plotting on the same figure the complex eigenvalues of the spatial operator and the red curve $\xi \in \mathbb{R} \mapsto \frac{2i\pi\xi}{\hat{g}(\xi)} \in \mathbb{C}$ related to the time operators. If there exists a ball centered on the real axis which contains all the eigenvalues and that does not intersect the red curve, then choosing $-l$ as the value of its center ensures that the expansion will converge.

E.2.4 Stochastic heat equation

Let us now deal with a classical the stochastic heat equation on $\mathbb{S}^1$ (described as the interval $[0, 1]$ where 0 and 1 are identified) as a classical example of linear partial differential equations:

$$\frac{\partial u}{\partial t}(x,t) = \Delta u(x,t) + v(x,t) + \sigma\eta(x,t)$$

with periodic boundary conditions, where Δ is the Laplacian on $\mathbb{S}^1$, v is an external forcing and η is a multidimensional white noise. The input $v(x, t)$ is set to $\delta_{x=x_0}(x)$ and we take the initial condition $u(., t = 0) = 0$.

The Laplacian operator has eigenvalues $-4\pi^2k^2$ with $k \in \mathbb{N}$, corresponding to the eigenvectors $\cos(2k\pi x)$ and $\sin(2k\pi x)$. Since the eigenvalues are not bounded, it is not possible to find a suitable constant l to define the solution of the heat equation in our framework. However, the semi-discretized in space equation overcomes this problem by preventing the existence of very fast oscillations (corresponding to large eigenvalues of the Laplacian). We choose to discretize the space with N points regularly spaced, corresponding to a discretization step $dx = 1/N$. The resulting equation corresponds to (E.3) in dimension N, with $g = \delta(t)$ and $A \in \mathbb{R}^{N\times N}$ such that $A_{ii} = -2/dx^2$, $A_{ij} = 1/dx^2$ if $i = j \pm 1$, $A_{1n} = A_{n1} = 1/dx^2$ accounting for the periodicity of the medium, and $A_{ij} = 0$ otherwise . This matrix has eigenvalues in $[-4/dx^2, 0]$. This suggests the choice $l = 2/dx^2$ so that all the eigenvalues are in this ball a center $-l$ and radius l. This ball intersect the imaginary axis only in 0 (corresponding to spatially constant functions), so convergence issues might arise if one of the terms $W^k \cdot (\widetilde{I}_\tau + \Sigma \cdot dB_\tau) \cdot \mathcal{U} \cdot \mathcal{V}^k$ is spatially constant, which clearly never occurs in our case. Therefore, our expansion is well-posed and provides a numerical scheme to compute the solution as shown in figure E.2.d. In that figure, we exhibit the fact that the solution is well retrieved by the expansion, and the error compared to Euler's scheme with a time step $dt = 0.01$ (Fig. E.2.b) is more than two order of magnitude smaller than the solution. An interesting point of this method is that it works for any time step interval dt which is not the case for the Euler method which

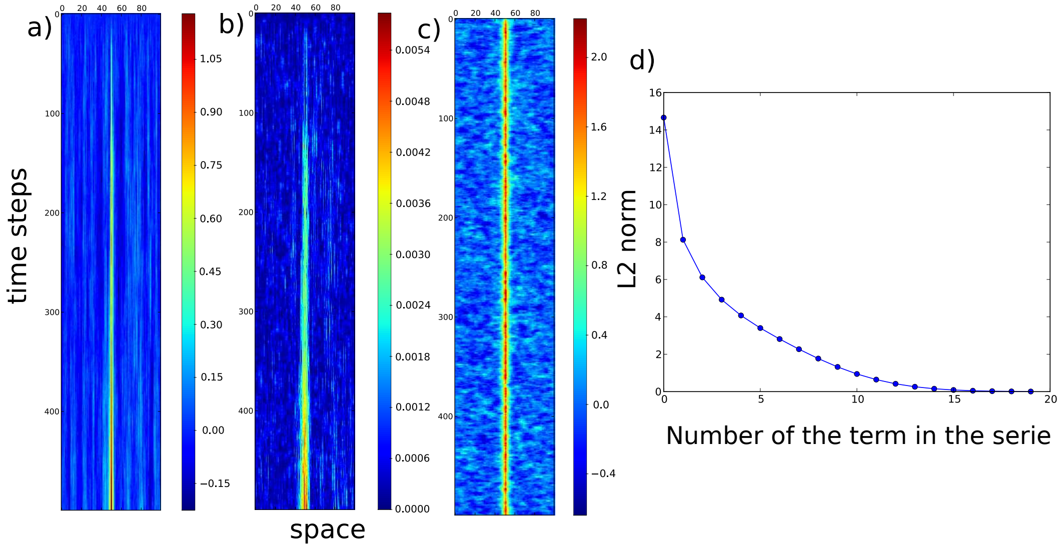


Figure E.2: Application of the expansion method to the stochastic heat equation on the circle with a Dirac source on the neuron in the middle. a) Space-time diagram of the solution given by the expansion method for $dt = 0.01$. b) Space-time diagram of the error between the solution in a) and the solution given by Euler's method. c) Space-time diagram of the solution given by the expansion method for $dt = 1$. d) L_2 norm of the terms in the expansion. The parameters for these simulations are $n = 100$, number of time steps $= 500$, $\sigma = 0.1$ and $l = 2$.

rapidly diverges as soon as the CFL condition is not satisfied for instance. Moreover, extending the approach to a delayed formalism $g \neq \delta$ is costless in our framework.

Bibliography

[Abbott et al. 2000] L.F. Abbott, S.B. Nelson*et al. Synaptic plasticity: taming the beast.* Nature neuroscience, vol. 3, pages 1178–1183, 2000. 75

[Ackley et al. 1985] D.H. Ackley, G.E. Hinton and T.J. Sejnowski. *A learning algorithm for Boltzmann machines.* Cognitive science, vol. 9, no. 1, pages 147–169, 1985. 67

[Amari et al. 1977] S.I. Amari, K. Yoshida and K.I. Kanatani. *A mathematical foundation for statistical neurodynamics.* SIAM Journal on Applied Mathematics, pages 95–126, 1977. 42

[Amari et al. 2002] S. Amari, K. Kurata and H. Nagaoka. *Information geometry of Boltzmann machines.* Neural Networks, IEEE Transactions on, vol. 3, no. 2, pages 260–271, 2002. 67

[Amari 1977] S. Amari. *Dynamics of pattern formation in lateral-inhibition type neural fields.* Biological Cybernetics, vol. 27, no. 2, pages 77–87, 1977. 66

[Amit and Brunel 1997] D.J. Amit and N. Brunel. *Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex.* Cerebral Cortex, vol. 7, no. 3, page 237, 1997. 37

[Amit and Tsodyks 1991] D.J. Amit and MV Tsodyks. *Quantitative study of attractor neural network retrieving at low spike rates: I. Substrate-spikes, rates and neuronal gain.* Network: Computation in neural systems, vol. 2, no. 3, pages 259–273, 1991. 60

[Arnold and Levi 1988] V.I. Arnold and M. Levi. Geometrical methods in the theory of ordinary differential equations, volume 250. Springer, 1988. 189

[Arous and Guionnet 1995] G.B. Arous and A. Guionnet. *Large deviations for Langevin spin glass dynamics.* Probability Theory and Related Fields, vol. 102, no. 4, pages 455–509, 1995. 42

[Atick and Redlich 1990] J.J. Atick and A.N. Redlich. *Towards a theory of early visual processing.* Neural Computation, vol. 2, no. 3, pages 308–320, 1990. 67

[Azari et al. 2001] N.P. Azari, J. Nickel, G. Wunderlich, M. Niedeggen, H. Hefter, L. Tellmann, H. Herzog, P. Stoerig, D. Birnbacher and R.J. Seitz. *Neural correlates of religious experience.* European Journal of Neuroscience, vol. 13, no. 8, pages 1649–1652, 2001. 2

[Baladron et al. 2011] J. Baladron, D. Fasoli, O. Faugeras and J. Touboul. *Mean Field description of and propagation of chaos in recurrent multipopulation networks of Hodgkin-Huxley and Fitzhugh-Nagumo neurons.* Arxiv preprint arXiv:1110.4294v1, 2011. 10, 41, 173

[Barlow 1989] H.B. Barlow. *Unsupervised learning.* Neural Computation, vol. 1, no. 3, pages 295–311, 1989. 67

[Barlow 2001] H. Barlow. *Redundancy reduction revisited.* Network: Computation in Neural Systems, vol. 12, no. 3, pages 241–253, 2001. 67

[Bartsch and Van Hemmen 2001] AP Bartsch and JL Van Hemmen. *Combined Hebbian development of geniculocortical and lateral connectivity in a model of primary visual cortex.* Biological Cybernetics, vol. 84, no. 1, pages 41–55, 2001. 130, 136

[Berkes et al. 2011] P. Berkes, G. Orbán, M. Lengyel and J. Fiser. *Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment.* Science, vol. 331, no. 6013, page 83, 2011. 119

[Bernstein 2009] D.S. Bernstein. Matrix mathematics: theory, facts, and formulas. Princeton Univ Pr, 2009. 226, 234

[Bi and Poo 1998] G. Bi and M. Poo. *Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type.* The Journal of Neuroscience, vol. 18, no. 24, page 10464, 1998. 29, 30, 110

[Bienenstock et al. 1982] E.L. Bienenstock, L.N. Cooper and P.W. Munro. *Theory for the development of neuron selectivity: orientation specificity* *and binocular interaction in visual cortex.* J Neurosci, vol. 2, pages 32–48, 1982. 66, 76, 77, 81, 125

[Blais and Cooper 2008] B. S. Blais and L. Cooper. *BCM theory.* Scholarpedia, vol. 3, no. 3, page 1570, 2008. 81, 125

[Borg and Groenen 2005] I. Borg and P.J.F. Groenen. Modern multidimensional scaling: Theory and applications. Springer Verlag, 2005. 138, 148, 164, 165, 166, 168

[Bosking et al. 1997] W.H. Bosking, Y. Zhang, B. Schofield and D. Fitzpatrick. *Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex.* Journal of neuroscience, vol. 17, no. 6, page 2112, 1997. 134, 156

[Bressloff and Cowan 2003] P. C. Bressloff and J. D. Cowan. *A spherical model for orientation and spatial frequency tuning in a cortical hypercolumn.* Philosophical Transactions of the Royal Society B, 2003. 135

[Bressloff et al. 2001] P.C. Bressloff, J.D. Cowan, M. Golubitsky, P.J. Thomas and M.C. Wiener. *Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex.* Phil. Trans. R. Soc. Lond. B, vol. 306, no. 1407, pages 299–330, March 2001. 133

[Bressloff 2005] P. Bressloff. *Spontaneous symmetry breaking in self–organizing neural fields.* Biological Cybernetics, vol. 93, no. 4, pages 256–274, October 2005. 67, 135, 150

[Bressloff 2009] P.C. Bressloff. *Stochastic neural field theory and the system-size expansion.* Not specified, 2009. 71

[Brette and Gerstner 2005] R. Brette and W. Gerstner. *Adaptive exponential integrate-and-fire model as an effective description of neuronal activity.* Journal of neurophysiology, vol. 94, no. 5, page 3637, 2005. 15

[Brewer 1978] J. Brewer. *Kronecker products and matrix calculus in system theory.* IEEE Transactions on Circuits and Systems, vol. 25, no. 9, pages 772–781, 1978. 208, 234, 243

[Brodmann 1909] K. Brodmann. *Vergleichende lokalisationslehre der grobhirnrinde.* Barth, Leipzig, 1909. 2

[Brunel et al. 2004] N. Brunel, V. Hakim, P. Isope, J.P. Nadal and B. Barbour. *Optimal Information Storage and the Distribution of Synaptic Weights:: Perceptron versus Purkinje Cell.* Neuron, vol. 43, no. 5, pages 745–757, 2004. 123

[Brunel 2000] N. Brunel. *Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons.* Journal of computational neuroscience, vol. 8, no. 3, pages 183–208, 2000. 16, 37

[Buice et al. 2010] M.A. Buice, J.D. Cowan and C.C. Chow. *Systematic fluctuation expansion for neural network activity equations.* Neural computation, vol. 22, no. 2, pages 377–426, 2010. 71

[Butler and Hodos 1996] A.B. Butler and W. Hodos. Comparative vertebrate neuroanatomy: evolution and adaptation. Wiley Online Library, 1996. 2

[Camera et al. 2004] G.L. Camera, A. Rauch, H.R. Lüscher, W. Senn and S. Fusi. *Minimal models of adapted neuronal response to in vivo-like input currents.* Neural computation, vol. 16, no. 10, pages 2101–2124, 2004. 60

[Caporale and Dan 2008] N. Caporale and Y. Dan. *Spike timing-dependent plasticity: a Hebbian learning rule.* Annu. Rev. Neurosci., vol. 31, pages 25–46, 2008. 4, 28, 31, 109

[Castellani et al. 1999] G.C. Castellani, N. Intrator, H. Shouval and L.N. Cooper. *Solutions of the BCM learning rule in a network of lateral interacting nonlinear neurons.* Network: Computation in Neural Systems, vol. 10, no. 2, pages 111–121, 1999. 81, 126

[Chapeau-Blondeau and Chambet 1995] F. Chapeau-Blondeau and N. Chambet. *Synapse models for neural networks: from ion channel kinetics to multiplicative coefficient w ij.* Neural computation, vol. 7, no. 4, pages 713–734, 1995. 26

[Chen et al. 2007] Z. Chen, S. Haykin and J.J. Eggermont. Correlative learning: a basis for brain and adaptive systems, volume 49. Wiley-Interscience, 2007. 67

[Chklovskii et al. 2002] D.B. Chklovskii, T. Schikorski and C.F. Stevens. *Wiring optimization in cortical circuits.* Neuron, vol. 34, no. 3, pages 341–347, 2002. 147

[Chossat and Faugeras 2009] P. Chossat and O. Faugeras. *Hyperbolic planforms in relation to visual edges and textures perception.* PLoS Computational Biology, vol. 5, no. 12, pages 367–375, 2009. 135

[Cohen and Grossberg 1983] M.A. Cohen and S. Grossberg. *Absolute stability of global pattern formation and parallel memory storage by competitive neural networks.* In IEEE Transactions on Systems, Man, and Cybernetics, SMC-13, pages 815–826, 1983. 207

[Coifman et al. 2005] R.R. Coifman, M. Maggioni, S.W. Zucker and I.G. Kevrekidis. *Geometric diffusions for the analysis of data from sensor networks.* Current opinion in neurobiology, vol. 15, no. 5, pages 576–584, 2005. 157

[Coombes 2001] S. Coombes. *Phase locking in networks of synaptically coupled McKean relaxation oscillators.* Physica D: Nonlinear Phenomena, vol. 160, no. 3-4, pages 173–188, 2001. 46

[Coombes 2005] S. Coombes. *Waves, bumps, and patterns in neural field theories.* Biological Cybernetics, vol. 93, no. 2, pages 91–108, 2005. 40, 62, 67, 159

[Cottet 1995] G.H. Cottet. *Neural networks: Continuous approach and applications to image processing.* Journal of Biological Systems, vol. 3, pages 1131–1139, 1995. 159

[Da Prato and Zabczyk 1992] G. Da Prato and J. Zabczyk. Stochastic equations in infinite dimensions, volume 45. Cambridge Univ Pr, 1992. 72, 245

[Dawson and Gartner 1987] D.A. Dawson and J. Gartner. *Large deviations from the McKean-Vlasov limit for weakly interacting diffusions.* Stochastics, vol. 20, no. 4, pages 247–308, 1987. 41

[Dayan and Abbott 2001] P. Dayan and L.F. Abbott. Theoretical neuroscience: Computational and mathematical modeling of neural systems. The MIT Press, 2001. 3, 67, 68, 123, 135, 141

[Degond and Mas-Gallic 1989] P. Degond and S. Mas-Gallic. *The Weighted Particle Method for Convection-Diffusion Equations. Part 1: The Case of an Isotropic Viscosity.* Mathematics of Computation, pages 485–507, 1989. 159

[Destexhe et al. 1998] A. Destexhe, Z.F. Mainen and T.J. Sejnowski. *Kinetic models of synaptic transmission.* Methods in neuronal modeling, pages 1–25, 1998. 26

[Dong and Hopfield 1992] D.W. Dong and J.J. Hopfield. *Dynamic properties of neural networks with adapting synapses.* Network: Computation in Neural Systems, vol. 3, no. 3, pages 267–283, 1992. 67, 88, 89, 207

[Durbin and Mitchison 1990] R. Durbin and G. Mitchison. *A dimension reduction framework for understanding cortical maps.* Nature, vol. 343, no. 6259, pages 644–647, 1990. 135

[Edwards 1996] R. Edwards. *Approximation of neural network dynamics by reaction-diffusion equations.* Mathematical methods in the applied sciences, vol. 19, no. 8, pages 651–677, 1996. 159

[Engel and Broeck 2001] A. Engel and C. Broeck. Statistical mechanics of learning. Cambridge Univ Pr, 2001. 123

[Ermentrout and Cowan 1980] GB Ermentrout and JD Cowan. *Large scale spatially organized activity in neural nets.* SIAM Journal on Applied Mathematics, pages 1–21, 1980. 38

[Ermentrout and Terman 2010] G.B. Ermentrout and D. Terman. Mathematical foundations of neuroscience, volume 35. Springer Verlag, 2010. 16, 19, 22, 26, 38

[Ermentrout 1994] B. Ermentrout. *Reduction of conductance-based models with slow synapses to neural nets.* Neural Computation, vol. 6, no. 4, pages 679–695, 1994. 60

[Erwin et al. 1995] E. Erwin, K. Obermayer and K. Schulten. *Models of orientation and ocular dominance columns in the visual cortex: A critical comparison.* Neural Computation, vol. 7, no. 3, pages 425–468, 1995. 135

[Faugeras et al. 2008] O. Faugeras, F. Grimbert and J.-J. Slotine. *Abolute stability and complete synchronization in a class of neural fields models.* SIAM J. Appl. Math, vol. 61, no. 1, pages 205–250, September 2008. 86, 90, 208

[Faugeras et al. 2009a] O. Faugeras, J. Touboul and B. Cessac. *A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs.* Frontiers in computational neuroscience, vol. 3, 2009. 42

[Faugeras et al. 2009b] O. Faugeras, R. Veltz and F. Grimbert. *Persistent neural states: stationary localized activity patterns in nonlinear continuous n-population, q-dimensional neural networks.* Neural computation, vol. 21, no. 1, pages 147–187, 2009. 67

[Felleman and Van Essen 1991] D.J. Felleman and D.C. Van Essen. *Distributed hierarchical processing in the primate cerebral cortex.* Cerebral cortex, vol. 1, no. 1, 1991. 122

[Feng 2004] J. Feng. Computational neuroscience: A comprehensive approach. CRC press, 2004. 37, 38

[Fenichel 1979] N. Fenichel. *Geometric singular perturbation theory for ordinary differential equations.* J. Differential Equations, vol. 31, no. 1, pages 53–98, 1979. 189

[Fitzhugh 1961] R. Fitzhugh. *Impulses and physiological states in theoretical models of nerve membrane.* Biophysical Journal, vol. 1, no. 6, pages 445–466, 1961. 18

[Florence and Kaas 1992] S.L. Florence and J.H. Kaas. *Ocular dominance columns in area 17 of Old World macaque and talapoin monkeys: complete reconstructions and quantitative analyses.* Visual neuroscience, vol. 8, no. 05, pages 449–462, 1992. 133

[Földiák 1991] P. Földiák. *Learning invariance from transformation sequences.* Neural Computation, vol. 3, no. 2, pages 194–200, 1991. 66, 80, 106, 130

[Folias and Bressloff 2004] S.E. Folias and P.C. Bressloff. *Breathing pulses in an excitatory neural network.* SIAM J. Appl. Dyn. Syst, vol. 3, no. 3, pages 378–407, 2004. 67

[Fourcaud-Trocmé et al. 2003] N. Fourcaud-Trocmé, D. Hansel, C. Van Vreeswijk and N. Brunel. *How spike generation mechanisms determine the neuronal response to fluctuating inputs.* The Journal of neuroscience, vol. 23, no. 37, page 11628, 2003. 37

[Frey and Morris 1997] U. Frey and R.G.M. Morris. *Synaptic tagging and long-term potentiation.* Nature, vol. 385, no. 6616, pages 533–536, 1997. 28

[Funahashi and Nakamura 1993] K. Funahashi and Y. Nakamura. *Approximation of dynamical systems by continuous time recurrent neural networks.* Neural networks, vol. 6, no. 6, pages 801–806, 1993. 119

[Gardner and Derrida 1988] E. Gardner and B. Derrida. *Optimal storage properties of neural network models.* Journal of Physics A: Mathematical and General, vol. 21, page 271, 1988. 123

[Gerstner and Kistler 2002a] W. Gerstner and W.M. Kistler. *Mathematical formulations of Hebbian learning.* Biological cybernetics, vol. 87, no. 5, pages 404–415, 2002. 30, 67, 68, 99

[Gerstner and Kistler 2002b] W. Gerstner and W.M. Kistler. Spiking neuron models: Single neurons, populations, plasticity. Cambridge Univ Pr, 2002. 67, 68, 141

[Gerstner 1995] W. Gerstner. *Time structure of the activity in neural network models.* Physical Review E, vol. 51, no. 1, page 738, 1995. 16

[Goodhill and Willshaw 1990] GJ Goodhill and DJ Willshaw. *Application of the elastic net algorithm to the formation of ocular dominance stripes.* Network: Computation in Neural Systems, vol. 1, no. 1, pages 41–59, 1990. 135

[Gütig et al. 2003] R. Gütig, R. Aharonov, S. Rotter and H. Sompolinsky. *Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity.* The Journal of neuroscience, vol. 23, no. 9, page 3697, 2003. 79

[Hale and Lunel 1993] J.K. Hale and S.M.V. Lunel. Introduction to functional differential equations. Springer, 1993. 72, 222, 242

[Han et al. 2008] F. Han, N. Caporale and Y. Dan. *Reverberation of recent visual experience in spontaneous cortical waves.* Neuron, vol. 60, no. 2, pages 321–327, 2008. 119

[Hansel et al. 1995] D. Hansel, G. Mato and C. Meunier. *Synchrony in excitatory neural networks.* Neural Computation, vol. 7, no. 2, pages 307–337, 1995. 16

[Hebb 1949] D.O. Hebb. The organization of behavior: a neuropsychological theory. Wiley, NY, 1949. 74

[Hertz et al. 1991] J. Hertz, A. Krogh and R.G. Palmer. Introduction to the theory of neural computation, volume 1. Westview press, 1991. 66

[Hille 1992] B. Hille. Ionic channels of excitable membranes, volume 348. Sinauer Associates Sunderland, MA, 1992. 13

[Hodgkin and Huxley 1952] A.L. Hodgkin and A.F. Huxley. *A quantitative description of membrane current and its application to conduction and excitation in nerve.* The Journal of physiology, vol. 117, no. 4, page 500, 1952. 5, 16

[Hopfield 1982] J.J. Hopfield. *Neural networks and physical systems with emergent collective computational abilities.* Proceedings of the national academy of sciences, vol. 79, no. 8, page 2554, 1982. 66

[Hopfield 1984] J.J. Hopfield. *Neurons with graded response have collective computational properties like those of two-state neurons.* Proceedings of the National Academy of Sciences, vol. 81, no. 10, page 3088, 1984. 66

[Hopfield 2007] J. J. Hopfield. *Hopfield network.* Scholarpedia, vol. 2, no. 5, page 1977, 2007. 39, 67

[Hubel and Wiesel 1962] D.H. Hubel and T.N. Wiesel. *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.* The Journal of Physiology, vol. 160, no. 1, page 106, 1962. 3, 121

[Hubel and Wiesel 1977] D.H. Hubel and T.N. Wiesel. *Functional architecture of macaque monkey visual cortex.* Proc. Roy. Soc. B, vol. 198, pages 1–59, 1977. 134

[Intrator and Cooper 1992] N. Intrator and L.N. Cooper. *Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions*.* Neural Networks, vol. 5, no. 1, pages 3–17, 1992. 81, 125

[Izhikevich and Desai 2003] E.M. Izhikevich and N.S. Desai. *Relating stdp to bcm.* Neural Computation, vol. 15, no. 7, pages 1511–1523, 2003. 30, 81

[Izhikevich 2003] E.M. Izhikevich. *Simple model of spiking neurons.* Neural Networks, IEEE Transactions on, vol. 14, no. 6, pages 1569–1572, 2003. 15

[Izhikevich 2007] E.M. Izhikevich. Dynamical systems in neuroscience: The geometry of excitability and bursting. The MIT press, 2007. 13, 14, 16, 17

[Kandel et al. 1991] E.R. Kandel, J.H. Schwartz, T.M. Jessell, S.A. Siegelbaum and A.J. Hudspeth. Principles of neural science, volume 3. Elsevier New York, 1991. 132

[Kenet et al. 2003] T. Kenet, D. Bibitchkov, M. Tsodyks, A. Grinvald and A. Arieli. *Spontaneously emerging cortical representations of visual attributes.* Nature, vol. 425, no. 6961, pages 954–956, 2003. 119

[Khalil and Grizzle 1996] H.K. Khalil and JW Grizzle. Nonlinear systems. Prentice hall Upper Saddle River, NJ, 1996. 90, 207

[Khas' minskii 1968] R.Z. Khas' minskii. *The Averaging Principle for Stochastic Differential Equations.* Problemy Peredachi Informatsii, vol. 4, no. 2, pages 86–87, 1968. 190, 196

[Kifer 2009] Y. Kifer. Large deviations and adiabatic transitions for dynamical systems and markov processes in fully coupled averaging, volume 944. Amer Mathematical Society, 2009. 192

[Kohonen 1990] T. Kohonen. *The Self-Organizing Map.* Proceedings of the IEEE, vol. 78, no. 9, 1990. 135, 158

[Laing et al. 2002] C.R. Laing, W.C. Troy, B. Gutkin and G.B. Ermentrout. *Multiple bumps in a neuronal model of working memory.* SIAM Journal on Applied Mathematics, pages 62–97, 2002. 67

[Lee and Wolpoff 2003] S.H. Lee and M.H. Wolpoff. *The pattern of evolution in Pleistocene human brain size.* Paleobiology, vol. 29, no. 2, page 186, 2003. 2

[Linkser 1988] E. Linkser. *Self-organization in a perceptual network.* Computer, vol. March, pages 105–117, 1988. 67

[Linkser 1992] R. Linkser. *Local synaptic rule suffice to maximize mutual information in a linear network.* Neural Computation, vol. 4, pages 691–702, 1992. 67

[Linsker 1988] R. Linsker. *Self-organization in a perceptual network.* COMPUTER,, pages 105–117, 1988. 130

[Lorenzi et al. 2010] L. Lorenzi, A. Lunardi and A. Zamboni. *Asymptotic behavior in time periodic parabolic problems with unbounded coefficients.* Journal of Differential Equations, 2010. 194, 196, 198

[Malsburg and Cowan 1982] C. Malsburg and J.D. Cowan. *Outline of a theory for the ontogenesis of iso-orientation domains in visual cortex.* Biological cybernetics, vol. 45, no. 1, pages 49–56, 1982. 77

[Mao 1997] X. Mao. Stochastic differential equations and their applications. Horwood publishing, 1997. 72, 242, 245

[Masquelier et al. 2009] T. Masquelier, R. Guyonneau and S.J. Thorpe. *Competitive STDP-based spike pattern learning.* Neural computation, vol. 21, no. 5, pages 1259–1276, 2009. 30

[McCulloch and Pitts 1943] W.S. McCulloch and W. Pitts. *A logical calculus of the ideas immanent in nervous activity.* Bulletin of Mathematical Biology, vol. 5, no. 4, pages 115–133, 1943. 66

[McKean 1970] H. P. McKean. *Nagumo's equation.* Advances in Mathematics, vol. 4, pages 209–223, 1970. 20

[Miikkulainen et al. 2005] R. Miikkulainen, J.A. Bednar, Y. Choe and J. Sirosh. Computational maps in the visual cortex. Springer, New York, 2005. 130, 136, 141, 142, 152

[Miller and MacKay 1994] K.D. Miller and D.J.C. MacKay. *The role of constraints in Hebbian learning.* Neural Computation, vol. 6, no. 1, pages 100–126, 1994. 66, 77, 79, 124, 127, 135, 141

[Miller et al. 1989] K. D. Miller, J. B. Keller and M. P. Stryker. *Ocular dominance column development: analysis and simulation.* Science, vol. 245, pages 605–615, 1989. 158

[Miller 1996] K.D. Miller. *Synaptic economics: Competition and cooperation in correlation-based synaptic plasticity.* Neuron, vol. 17, pages 371–374, 1996. 77

[Morris and Lecar 1981] C. Morris and H. Lecar. *Voltage oscillations in the barnacle giant muscle fiber.* Biophysical journal, vol. 35, no. 1, pages 193–213, 1981. 18

[Muratov et al. 2005] C.B. Muratov, E. Vanden-Eijnden*et al.* *Self-induced stochastic resonance in excitable systems.* Physica D: Nonlinear Phenomena, vol. 210, no. 3-4, pages 227–240, 2005. 21, 22, 23, 47

[Nagumo et al. 1962] J. Nagumo, S. Arimoto and S. Yoshizawa. *An active pulse transmission line simulating nerve axon.* Proceedings of the IRE, vol. 50, no. 10, pages 2061–2070, 1962. 18

[Oja 1982] E. Oja. *Simplified neuron model as a principal component analyzer.* Journal of mathematical biology, vol. 15, no. 3, pages 267–273, 1982. 66, 77, 80, 124, 127

[Olshausen and Field 1996] B.A. Olshausen and D.J. Field. *Emergence of simple-cell receptive field properties by learning a sparse code for natural images.* Nature, vol. 381, pages 607–609, 1996. 67

[O'Malley 1991] R.E. O'Malley. Singular perturbation methods for ordinary differential equations. Springer New York, 1991. 189

[Ostojic et al. 2009] S. Ostojic, N. Brunel and V. Hakim. *How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains.* The Journal of Neuroscience, vol. 29, no. 33, page 10234, 2009. 60

[Petitot 2003] J. Petitot. *The neurogeometry of pinwheels as a sub-Riemannian contact structure.* Journal of Physiology-Paris, vol. 97, no. 2-3, pages 265–309, 2003. 134

[Pfister and Gerstner 2006] J.P. Pfister and W. Gerstner. *Triplets of spikes in a model of spike timing-dependent plasticity.* The Journal of neuroscience, vol. 26, no. 38, pages 9673–9682, 2006. 81

[Pfister et al. 2010] J.P. Pfister, P. Dayan and M. Lengyel. *Synapses with short-term plasticity are optimal estimators of presynaptic membrane potentials.* Nature Neuroscience, vol. 13, no. 10, pages 1271–1275, 2010. 27

[Pinto and Ermentrout 2001] D.J. Pinto and G.B. Ermentrout. *Spatially structured activity in synaptically coupled neuronal networks: I. Traveling fronts and pulses.* SIAM journal on Applied Mathematics, pages 206–225, 2001. 67

[Redondo and Morris 2010] R.L. Redondo and R.G.M. Morris. *Making memories last: the synaptic tagging and capture hypothesis.* Nature Reviews Neuroscience, vol. 12, no. 1, pages 17–30, 2010. 28

[Ricciardi and Smith 1977] L.M. Ricciardi and C.E. Smith. Diffusion processes and related topics in biology. Springer-Verlag Berlin, 1977. 37

[Rinzel and Frankel 1992] J. Rinzel and P. Frankel. *Activity patterns of a slow synapse network predicted by explicitly averaging spike dynamics.* Neural Computation, vol. 4, no. 4, pages 534–545, 1992. 60

[Risken 1996] H. Risken. The fokker-planck equation: Methods of solution and applications, volume 18. Springer Verlag, 1996. 198

[Rolls and Deco 2002] E.T. Rolls and G. Deco. Computational neuroscience of vision. Oxford University Press, 2002. 77, 106, 121

[Romani et al. 1982] G.L. Romani, S.J. Williamson and L. Kaufman. *Tonotopic organization of the human auditory cortex.* Science, vol. 216, no. 4552, page 1339, 1982. 112

[Rubner and Tavan 1989] J. Rubner and P. Tavan. *A self-organizing network for principal-component analysis.* Europhysics Letters, vol. 10, page 693, 1989. 125, 130

[Samuelides and Cessac 2007] M. Samuelides and B. Cessac. *Random recurrent neural networks.* European Physical Journal-Special Topics, vol. 142, pages 7–88, 2007. 42

[Sanders et al. 2007] J.A. Sanders, F. Verhulst and J.A. Murdock. Averaging methods in nonlinear dynamical systems, volume 59. Springer Verlag, 2007. 86

[Sejnowski and Tesauro 1989] T.J. Sejnowski and G. Tesauro. *The Hebb rule for synaptic plasticity: algorithms and implementations.* Neural models of plasticity, pages 94–103, 1989. 157

[Sejnowski et al. 1977] T.J. Sejnowski*et al. Statistical constraints on synaptic plasticity.* Journal of theoretical biology, vol. 69, no. 2, page 385, 1977. 81

[Serre 2005] T. Serre. *A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex.* Rapport technique, DTIC Document, 2005. 130

[Shriki et al. 2003] O. Shriki, D. Hansel and H. Sompolinsky. *Rate models for conductance-based cortical neuronal networks.* Neural computation, vol. 15, no. 8, pages 1809–1841, 2003. 60

[Sjöström and Gerstner 2010] J. Sjöström and W. Gerstner. *Spike-timing dependent plasticity.* Scholarpedia, vol. 5, no. 2, page 1362, 2010. 28, 30

[Sompolinsky et al. 1988] H. Sompolinsky, A. Crisanti and HJ Sommers. *Chaos in random neural networks.* Physical Review Letters, vol. 61, no. 3, pages 259–262, 1988. 42

[Song et al. 2000] S. Song, K.D. Miller, L.F. Abbott*et al.* *Competitive Hebbian learning through spike-timing-dependent synaptic plasticity.* nature neuroscience, vol. 3, pages 919–926, 2000. 30

[Stringer and Rolls 2002] S.M. Stringer and E.T. Rolls. *Invariant object recognition in the visual system with novel views of 3D objects.* Neural Computation, vol. 14, no. 11, pages 2585–2596, 2002. 130

[Sur et al. 1999] M. Sur, A. Angelucci, J. Sharma*et al.* *Rewiring cortex: The role of patterned activity in development and plasticity of neocortical circuits.* Journal of Neurobiology, vol. 41, no. 1, pages 33–43, 1999. 3

[Swindale 1996] NV Swindale. *The development of topography in the visual cortex: a review of models.* Network: Computation in Neural Systems, vol. 7, no. 2, pages 161–247, 1996. 119, 135, 147, 158

[Swindale 2000] N.V. Swindale. *How many maps are there in visual cortex?* Cerebral cortex, vol. 10, no. 7, page 633, 2000. 133, 135

[Sznitman 1991] A.S. Sznitman. *Topics in propagation of chaos.* Ecole d'Eté de Probabilités de Saint-Flour XIX–1989, pages 165–251, 1991. 41

[Takeuchi and Amari 1979] A. Takeuchi and S. Amari. *Formation of topographic maps and columnar microstructures in nerve fields.* Biological Cybernetics, vol. 35, no. 2, pages 63–72, 1979. 67, 135

[Tanaka 1996] K. Tanaka. *Inferotemporal cortex and object vision.* Annual review of neuroscience, vol. 19, no. 1, pages 109–139, 1996. 133

[Tikhonov 1952] AN Tikhonov. *Systems of differential equations with small parameters multiplying the derivatives.* Matem. sb, vol. 31, no. 3, pages 575–586, 1952. 85

[Tonnelier 2007] A. Tonnelier. *Mckean model.* Scholarpedia, vol. 2, no. 4, page 2795, 2007. 20

[Touboul 2009] J. Touboul. *Importance of the cutoff value in the quadratic adaptive integrate-and-fire model.* Neural computation, vol. 21, no. 8, pages 2114–2122, 2009. 15

[Touboul 2011] J. Touboul. *Mean-field equations for stochastic neural fields with spatio-temporal delays.* Arxiv preprint arXiv:1108.2414v1, 2011. 10, 41, 173

[Tsodyks and Sejnowski 1995] MV Tsodyks and T. Sejnowski. *Rapid state switching in balanced cortical network models.* Network: Computation in Neural Systems, vol. 6, no. 2, pages 111–124, 1995. 16

[Tsodyks et al. 1998] M. Tsodyks, K. Pawelzik and H. Markram. *Neural networks with dynamic synapses.* Neural Computation, vol. 10, no. 4, pages 821–835, 1998. 26

[Turrigiano and Nelson 2004] G.G. Turrigiano and S.B. Nelson. *Homeostatic plasticity in the developing nervous system.* Nature Reviews Neuroscience, vol. 5, no. 2, pages 97–107, 2004. 75

[Van Essen and Gallant 1994] D.C. Van Essen and J.L. Gallant. *Neural mechanisms of form and motion processing in the primate visual system.* Neuron, vol. 13, 1994. 122

[Van Rossum et al. 2000] M.C.W. Van Rossum, G.Q. Bi and G.G. Turrigiano. *Stable Hebbian learning from spike timing-dependent plasticity.* The Journal of Neuroscience, vol. 20, no. 23, page 8812, 2000. 30

[Varela et al. 1997] J.A. Varela, K. Sen, J. Gibson, J. Fost, LF Abbott and S.B. Nelson. *A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex.* The Journal of neuroscience, vol. 17, no. 20, page 7926, 1997. 27

[Veltz and Faugeras 2009] R. Veltz and O. Faugeras. *Local/global analysis of the stationary solutions of some neural field equations.* Arxiv preprint arXiv:0910.2247, 2009. 67

[Veltz and Faugeras 2011] R. Veltz and O. Faugeras. *Stability of the stationary solutions of neural field equations with propagation delays.* The Journal of Mathematical Neuroscience, vol. 1, no. 1, pages 1–25, 2011. 34, 91

[Verhulst 2007] F. Verhulst. *Singular perturbation methods for slow–fast dynamics.* Nonlinear Dynamics, vol. 50, no. 4, pages 747–753, 2007. 85

[Wainrib 2011] G. Wainrib. *Double averaging principle for periodically forced slow-fast stochastic systems.* submitted, 2011. 64, 95, 174, 188, 195

[Wallis and Baddeley 1997] G. Wallis and R. Baddeley. *Optimal, unsupervised learning in invariant object recognition.* Neural computation, vol. 9, no. 4, pages 883–894, 1997. 66

[Wang et al. 1998] G. Wang, M. Tanifuji and K. Tanaka. *Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging.* Neuroscience Research, vol. 32, no. 1, pages 33–46, 1998. 133

[Wilson and Cowan 1972] H.R. Wilson and J.D. Cowan. *Excitatory and inhibitory interactions in localized populations of model neurons.* Biophysical journal, vol. 12, no. 1, pages 1–24, 1972. 40, 66

[Wilson and Cowan 1973] H.R. Wilson and J.D. Cowan. *A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue.* Biological Cybernetics, vol. 13, no. 2, pages 55–80, 1973. 40, 67

[Wong 1999] R.O.L. Wong. *Retinal waves and visual system development.* Annual Review of Neuroscience, vol. 22, no. 1, pages 29–47, 1999. 142

[Xie and Seung 2000] X. Xie and H.S. Seung. *Spike-based learning rules and stabilization of persistent neural activity.* Advances in neural information processing systems, vol. 12, pages 199–205, 2000. 111

Une approche mathématique de l'apprentissage non supervisé dans les réseaux de neurones récurrents.

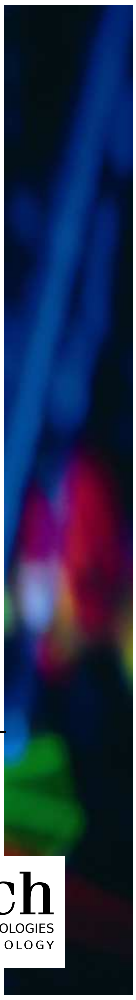
Résumé : Dans cette thèse nous tentons de donner un sens mathématique à la proposition : le néocortex se construit un modèle de son environnement.

Nous considérons que le néocortex est un réseau de neurones spikants dont la connectivité est soumise à une lente evolution appelée apprentissage. Dans le cas où le nombre de neurones est proche de l'infini, nous proposons une nouvelle methode de champ-moyen afin de trouver une équation decrivant l'évolution du taux de décharge de populations de neurones.

Nous étudions donc la dynamique de ce system moyénisé avec apprentissage. Dans le régime oú l'apprentissage est beaucoup plus lent que l'activité du réseau nous pouvons utiliser des outils de moyennisation temporelle pour les système lents/rapides. Dans ce cadre mathématique nous montrons que la connectivité du réseau converge toujours vers une unique valeur d'équilibre que nous pouvons calculer explicitement. Cette connectivité regroupe l'ensemble des connaissances du réseau à propos de son environnement.

Nous comparons cette connectivité à l'équilibre avec les stimuli du réseau. Considérant que l'environnement est solution d'un système dynamique quelconque, il est possible de montrer que le réseau encode la totalité de l'information nécessaire à la definition de ce systeème dynamique. En effet nous montrons que la partie symmétrique de la connectivité correspond à la variété sur laquelle est definie le système dynamique de l'environnement, alors que la partie anti-symmétrique de la connectivité correspond au champ de vecteur définissant le système dynqmique de l'environnement. Dans ce contexte il devient clair que le réseau agit comme un predicteur de son environnement.

Mots clés : systèmes dynamiques, réseaux de neurones spikants, champ moyen, réseaux de neurones à taux de décharge, apprentissage non-supervisé, basé sur les correlations , la causalité, apprentissage Hebbien.



A mathematical approach to unsupervised learning in recurrent neural networks

Abstract: In this thesis, we propose to give a mathematical sense to the claim: the neocortex builds itself a model of its environment.

We study the neocortex as a network of spiking neurons undergoing slow STDP learning. By considering that the number of neurons is close to infinity, we propose a new mean-field method to find the "smoother" equation describing the firing-rate of populations of these neurons.

Then, we study the dynamics of this averaged system with learning. By assuming the modification of the synapses' strength is very slow compared the activity of the network, it is possible to use tools from temporal averaging theory. They lead to showing that the connectivity of the network always converges towards a single equilibrium point which can be computed explicitely. This connectivity gathers the knowledge of the network about the world.

Finally, we analyze the equilibrium connectivity and compare it to the inputs. By seeing the inputs as the solution of a dynamical system, we are able to show that the connectivity embedded the entire information about this dynamical system. Indeed, we show that the symmetric part of the connectivity leads to finding the manifold over which the inputs dynamical system is defined, and that the anti-symmetric part of the connectivity corresponds to the vector field of the inputs dynamical system. In this context, the network acts as a predictor of the future events in its environment.

Keywords: Spiking networks, mean-field, rate-based models, unsupervised, correlation-based, causation-based, Hebbian learning, dynamical systems

MINES
ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY