Complementary approaches to Synaptic Plasticity: from Objective Functions to Biophysics

DISSERTATION
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Physik der
Goethe-Universität Frankfurt
in Frankfurt am Main

von
Rodrigo S. Echeveste
aus Rosario, Argentinien

Frankfurt am Main, März 2016

vom Fachbereich Physik der

Goethe Universität Frankfurt

als Dissertation angenommen.

Dekan: Prof. Dr. Rene Reifarth

Gutachter: Prof. Dr. Claudius Gros
Prof. Dr. Christoph von der Malsburg

Datum der Disputation: Mai 2016

To those who made me feel at home, far away from home.

Contents

Deutsche Zusammenfassung

Für die Modellierung des Gehirns kann auf unterschiedliche Ansätze zurückgegriffen werden. Zum einen können für die Modellbildung die konstituierenden chemischen und biologischen Bausteine herangezogen werden. Man erhält dann, zusammen mit den grundlegenden physikalischen und chemischen Wechselwirkungen, eine detaillierte mikroskopische Beschreibung. Zum anderen ist es möglich Gehirnfunktionen auf einer makroskopischen Ebene zu beschreiben, welche ihrerseits den jeweiligen Zielsetzungen angepasst ist. Ziel könnte es z.B. sein die Stoffwechselkosten zu untersuchen oder die Stabilität und Robustheit, sowie die Frage nach der rechnerischen Effizienz. Das trifft auch für die synaptische Plastizität zu, das heißt für die zeitliche Adaption der interneuronalen Verbindungsstärken, welche wir in der vorliegenden Arbeit untersuchen.

Insbesondere formulieren und untersuchen wir zwei unterschiedliche Modelle, beruhend auf komplementären Methoden, für synaptische Plastizität: Mit einem "top-down" Ansatz, bei dem eine Lernregel aus einem erzeugenden Prinzip für frequenzkodierende Neuronen abgeleitet wird, und einer "bottom-up" Methode, bei der eine einfache, aber biophysikalische Regel für zeitabhängige Plastizität aufgestellt wird.

Obwohl unterschiedliche Wege beschritten werden, ist ein gemeinsames Thema in dieser Arbeit vorhanden: die Suche nach Einfachheit. Wir sind an Minimal-Modellen interessiert, die die Essenz der Prozesse einfangen. Diesem liegt die Überzeugung zugrunde, dass Einfachheit und die Reduktion auf das Wesentliche eines Phänomens helfen kann, die Rolle der verschiedenen Komponenten in komplexen Systemen besser zu verstehen.

In Kapitel 1 beginnen wir diese These mit einer Diskussion über die gegenwärtigen Herausforderungen der Computational Neuroscience, sowie die Rolle des Physikers in diesem Forschungszweig. Außerdem präsentieren wir einen allgemeinen Überblick über die Eigenschaften von Neuronen und ihren Verbindungen, die Bausteine unserer Modelle, welche zudem die Einschränkungen für ihre For-

mulierung bestimmen. Insbesondere präsentieren wir zwei Modelle für neuronale Dynamik: ein frequenzkodierendes und ein spikekodierendes Modell, für die wir später die jeweiligen Plastizitäts-Regeln entwickeln. Darüber hinaus führen wir hier die Bezeichnungen und den Jargon des Feldes ein.

In Kapitel 2 entwickeln und untersuchen wir für die synaptischen Gewichte eine lokale Plastizitäts-Regel, welche Hebb'sch ist, online und selbstlimitierend. Sie beruht auf dem oben erwähnten "top-down" Ansatz. Zuerst formulieren wir das Prinzip der Stationarität beim statistischen Lernen, die besagt, dass, wenn das Neuron die relevanten Merkmale einer stationären Eingangsverteilung gelernt hat, die Ausgangsverteilung auch stationär werden sollte. Wir argumentieren dann, dass eine notwendige Bedingung für die Stationarität in einer Umgebung mit Rauschen, die Stabilität der Lösung ist, die das Neuron in dem Raum der synaptischen Gewichte findet. Das bedeutet, dass diese Lösung lokal unempfindlich für weitere Änderungen der gefundenen synaptischen Gewichtungen sein sollte.

Um diese lokale Unempfindlichkeits-Bedingung auszudrücken, greifen wir in Abschnitt 2.2.1 auf die Fisher-Information zurück, ein Maß für die durchschnittliche Empfindlichkeit einer Wahrscheinlichkeitsverteilung auf einen gegebenen Parameter. In diesem Fall nutzen wir die Fisher-Information der Ausgangswahrscheinlichkeitsverteilung in Abhängigkeit der synaptischen Gewichte. Um sicherzustellen, dass die Lernregeln als Funktion lokaler Information (an einer Synapse) formuliert werden, nutzen wir die "local synapse extension" der eindimensionalen Fisher-Information. Sobald die Zielfunktion definiert wurde, leiten wir dann in Abschnitt 2.2.2 eine online Regel der synaptische-Plastizität über das stochastische Gradientenverfahren her.

Die daraus resultierende Lernregel besteht aus zwei Faktoren: eine Hebb'sche Funktion (proportional zum Produkt von prä- und postsynaptischen Aktivitäten) und eine selbstlimitierende Funktion, die das Vorzeichen des Lernens umkehrt, wenn die neurale Aktivität zu hoch oder zu niedrig ist. Bei diesem Vorgehen werden sowohl das neuronale Aktivitätsniveau als auch die synaptischen Gewichte reguliert. Ein expliziter Gewicht abklingender Ausdruck ist in dieser Weise nicht notwendig.

Um die Rechenkapazität eines Neurons zu testen, das sich nach diesen Regeln entwickelt (in Verbindung mit einer bereits vorhandenen intrinsischen Plastizitäts-Regel), führen wir in Abschnitt 2.3 eine Reihe von numerischen Experimenten durch, in denen wir das Neuron mit verschiedenen Eingabeverteilungen trainieren. Wir beobachten, dass für Eingabeverteilungen, die stark einer multivariaten Normalverteilung ähneln, das Neuron zuverlässig die erste Hauptkomponente der Verteilung auswählt. Das Neuron zeigt, ansonsten, eine starke Präferenz für Richtungen mit großer negativer Exzess Kurtosis. Insbesondere finden wir, dass das Neuron zu bimodalen Richtungen selektiv ist und geeignet für binäre Klassifizie-

rung. Darüber hinaus zeigen wir in Abschnitt 2.3.3, wie unsere Regel eine deutliche "Fading Memory" Funktion zeigt, mit sehr unterschiedlichen Zeitskalen für das Lernen und Verlernen, und eine besondere Robustheit gegen Rauschen.

In Kapitel 3 untersuchen wir die Zuverlässigkeit der Lernregel, die wir in Kapitel 2 abgeleitet haben, in Bezug auf Veränderungen in der neuronalen Modell-Übertragungsfunktion. Insbesondere finden wir eine äquivalente kubische Form der Regel, die es aufgrund ihrer funktionalen Einfachheit ermöglicht analytisch die Attraktoren (stationäre Lösungen) aus dem Lernverfahren in Abhängigkeit der statistischen Momente der Eingangsverteilung zu berechnen. Auf diese Weise ist es uns möglich in Abschnitt 3.2.1, die numerischen Ergebnisse aus Kapitel 2 analytisch zu erklären. Zudem sind wir in der Lage, die Stabilität dieser Attraktoren zu bewerten und die Eigenwerte der Jacobi-Matrix auf die statistischen Momente der Eingangsverteilung zu beziehen.

Diese Ergebnisse ermöglichen es uns, eine Vorhersage zu formulieren: Wenn das Neuron zu Nicht-Gauß-Eingangsrichtungen selektiv ist, sollte es für Independent Component Analysis (ICA) geeignet sein. Am Ende des Kapitels 2 testen wir diese Vorhersage, indem wir unsere Lernregel auf das "non-linear bars problem" anwenden. In dieser Aufgabe stellen die Eingänge zu dem Neuron Pixel von einem quadratischen Bild dar, die zwei Werte annehmen können: hell oder dunkel. Das Bild besteht aus einer Reihe von horizontalen und vertikalen Streifen, in dem ein Streifen eine vollständige Reihe oder Spalte von dunklen Pixeln ist. Am Schnittpunkt eines horizontalen und eines vertikalen Streifen hat das Pixel den gleichen Dunkelwert wie im Rest des Streifens (es ist nicht die Summe der Intensitäten), was das Problem nicht-linear macht. Wir trainieren zunächst das Neuron mit einem Trainingssatz, in dem jeder Streifen unabhängig zufällig gezogen wird, mit einer konstanten Wahrscheinlichkeit. Wir testen dann den Fall, in dem in jedem Bild mindestens eine horizontale und eine vertikale Leiste vorhanden ist (permanente partiale Verdeckung). In beiden Fällen finden wir, dass das Neuron in der Lage ist, die einzelnen Streifen des Trainingssatzes zu lernen, obwohl diese Streifen dem Neuron nie isoliert präsentiert wurden.

Die Relevanz dieser Ergebnisse liegt in ihrer Allgemeinheit: Will man eine Lernregel, die Hebb'sch für einen bestimmten Bereich von Aktivitäten ist, aber dann seine Steigung und schließlich sein Vorzeichen umkehrt, wenn die neuronale Aktivität zu groß oder zu klein wird, ist die kubische Form (hier im weitesten Sinne des Wortes Form) die minimale Konstruktion, die man sich vorstellen kann. Was wir hier zeigen ist, dass eine solche minimale Konstruktion (und Äquivalente, solange die allgemeine Form beibehalten wird) rechnerisch bereits sehr mächtig ist.

In Kapitel 4 folgen wir dem entgegengesetzten Weg, indem wir ein einfaches biophysikalisches Modell für zeitabhängige Plastizität (STDP) entwickeln. Eine phänomenologische paarweise Lernregel ist nicht genug um STDP zu erklären.

Dies wird deutlich, wenn man die Nichtlinearitäten in Triplett-Ergebnissen betrachtet. Deshalb benötigt man ein Modell, dass das Zusammenwirken mehrerer Spikes beinhalten kann.

Das Modell, das wir hier entwickeln, ist in Bezug auf zwei abklingende Spuren formuliert, die in der Synapse vorhanden sind: Zum einen der Anteil der aktivierten NMDA-Rezeptoren und zum anderen die Calciumkonzentration. Diese Spuren dienen als Uhren, die die Zeit der prä- und postsynaptischen Spikes messen. Trotz der Tatsache, dass wir das Modell in Bezug auf die biologischen Schlüsselelemente konstruieren, die an dem Prozess beteiligt sind, haben wir die funktionalen Abhängigkeiten der Variablen so einfach wie möglich gehalten, um eine analytische Lösung zu ermöglichen.

Wir behaupten nicht, mit diesem Modell die volle biologische Komplexität des Prozesses zu erfassen. Das dargestellte Modell ist ein effektives Modell für STDP, in dem die Effekte von einer großen Anzahl biologischer Komponenten innerhalb einiger Variablen zusammengelegt werden. Wir sind davon überzeugt, dass diese Vereinfachung, die wir brauchen um die Regel analytisch zu untersuchen, auch im Hinblick auf ein einfacheres Verständnis der allgemeinen beteiligten Regeln ein Vorteil ist.

Wir zeigen zuerst, dass trotz seiner Einfachheit das Modell mehrere experimentelle Ergebnisse reproduzieren kann. In Abschnitt 4.3.1 zeigen wir analytisch, dass für ein Paar Spikes (einen prä- und einen postsynaptischen Spike), das Modell in der Lage ist, die typische paarweise STDP Form nachzubilden. Das heißt, Paare in einer kausalen Ordnung induzieren Potenzierung, während Paare in einer anti-kausalen Ordnung Depression des synaptischen Gewichts erzeugen, mit einem reduzierten Effekt für längere Intervalle zwischen den Spikes. Darüber hinaus berechnen wir in Abschnitt 4.3.2 die analytischen Vorhersagen des Modells für Spike-Tripletts, in entweder PräPostPrä- oder PostPräPost-Ordnung. In Abschnitt 4.4 vergleichen wir Experimentelle- und Modellergebnisse, sowohl in einer Kultur von Nervenzellen aus dem Hippocampus als auch in L 2/3 kortikalen Neuronen. Dank der funktionalen Einfachheit des Modells sind wir in der Lage diese Ergebnisse analytisch zu berechnen und eine direkte und transparente Verbindung zwischen den internen Parametern des Modells und der qualitativen Merkmale der Ergebnisse zu etablieren.

In diesem Sinne beobachten wir, dass während Spurenakkumulation für Triplett-Nichtlinearitäten in Hippocampus-Neuronen verantwortlich zu sein scheint, dominieren starke Sättigungseffekte in kortikalen Neuronen. Dies steht im Einklang mit den Ergebnissen früherer phänomenologischer Regeln, die durch eine verminderte Wirksamkeit von zukünftigen Spikes, Triplett-Ergebnisse in kortikalen Neuronen erklärt.

Zum Schluss, um eine Verbindung zu der synaptischen Plastizität für frequenzkodierende neuronale Modelle herzustellen, trainieren wir in Abschnitt 4.5 die Synapse mit Poisson unkorrelierten prä- und postsynaptischen Pulszügen und berechnen die erwartete synaptische Gewichtsänderung in Abhängigkeit der Frequenzen dieser Pulszüge.

Interessant ist, dass ein Hebb'sches (im frequenzkodierenden Sinne des Wortes), BCM-ähnliches Verhalten für Hippocampus-Neuronen in diesem Setup beobachtet wird: Wir finden, dass die resultierende Kraft-Modifikation an einer bestimmten Schwellenfrequenz von Depression zu Potenzierung übergeht, wobei diese Schwelle eine monoton zunehmende Funktion der präsynaptischen Frequenz ist. Darüber hinaus kann der Wert des Schwellenwerts geregelt werden, während immer noch Paarweise- und Triplett-Ergebnisse reproduziert werden können.

Andererseits scheint dominierende Depression unvermeidlich zu sein für Parameterkonfigurationen, die experimentell Triplett-Nichtlinearitäten in der L 2/3 kortikalen Neuronen reproduzieren können. Potenzierung kann jedoch in diesen Neuronen wiederhergestellt werden, wenn Korrelationen zwischen prä- und postsynaptischen Spikes vorhanden sind. Wir weisen an dieser Stelle darauf hin, dass gleichzeitige Entkorrelation der neuronalen Aktivität und Depression im Kortex in sensorischen Deprivations-Experimenten gefunden wird.

Wir beenden Kapitel 4 mit einer Diskussion in Abschnitt 4.6 über das Verhältnis dieser Ergebnisse zu bestehenden experimentellen Ergebnissen und wir formulieren offene Fragen und Vorhersagen für zukünftige Experimente.

Übersichtskarten der Modelle, zusammen mit Auflistungen der relevanten Variablen und Parameter, werden für einen leichteren Zugang und permanente Referenz für den Leser, am Ende der Doktorarbeit präsentiert.

Abstract

Different approaches are possible when it comes to modeling the brain. Given its biological nature, models can be constructed out of the chemical and biological building blocks known to be at play in the brain, formulating a given mechanism in terms of the basic interactions underlying it. On the other hand, the functions of the brain can be described in a more general or macroscopic way, in terms of desirable goals. This goals may include reducing metabolic costs, being stable or robust, or being efficient in computational terms. Synaptic plasticity, that is, the study of how the connections between neurons evolve in time, is no exception to this.

In the following work we formulate (and study the properties of) synaptic plasticity models, employing two complementary approaches: a top-down approach, deriving a learning rule from a guiding principle for rate-encoding neurons, and a bottom-up approach, where a simple yet biophysical rule for time-dependent plasticity is constructed.

We begin this thesis with a general overview, in **Chapter 1**, of the properties of neurons and their connections, clarifying notations and the jargon of the field. These will be our building blocks and will also determine the constrains we need to respect when formulating our models. We will discuss the present challenges of computational neuroscience, as well as the role of physicists in this line of research.

In **Chapters 2** and **3**, we develop and study a local online Hebbian self-limiting synaptic plasticity rule, employing the mentioned top-down approach. Firstly, in **Chapter 2** we formulate the stationarity principle of statistical learning, in terms of the Fisher information of the output probability distribution with respect to the synaptic weights. To ensure that the learning rules are formulated in terms of information locally available to a synapse, we employ the local synapse extension to the one dimensional Fisher information. Once the objective function has been defined, we derive an online synaptic plasticity rule via stochastic gradient descent.

In order to test the computational capabilities of a neuron evolving according to this rule (combined with a preexisting intrinsic plasticity rule), we perform a series of numerical experiments, training the neuron with different input distributions. We observe that, for input distributions closely resembling a multivariate normal distribution, the neuron robustly selects the first principal component of the

distribution, showing otherwise a strong preference for directions of large negative excess kurtosis.

In **Chapter 3** we study the robustness of the learning rule derived in **Chapter 2** with respect to variations in the neural model's transfer function. In particular, we find an equivalent cubic form of the rule which, given its functional simplicity, permits to analytically compute the attractors (stationary solutions) of the learning procedure, as a function of the statistical moments of the input distribution. In this way, we manage to explain the numerical findings of **Chapter 2** analytically, and formulate a prediction: if the neuron is selective to non-Gaussian input directions, it should be suitable for applications to independent component analysis. We close this section by showing how indeed, a neuron operating under these rules can learn the independent components in the non-linear bars problem.

A simple biophysical model for time-dependent plasticity (STDP) is developed in **Chapter 4**. The model is formulated in terms of two decaying traces present in the synapse, namely the fraction of activated NMDA receptors and the calcium concentration, which serve as clocks, measuring the time of pre- and postsynaptic spikes. While constructed in terms of the key biological elements thought to be involved in the process, we have kept the functional dependencies of the variables as simple as possible to allow for analytic tractability. Despite its simplicity, the model is able to reproduce several experimental results, including the typical pairwise STDP curve and triplet results, in both hippocampal culture and layer 2/3 cortical neurons. Thanks to the model's functional simplicity, we are able to compute these results analytically, establishing a direct and transparent connection between the model's internal parameters and the qualitative features of the results.

Finally, in order to make a connection to synaptic plasticity for rate encoding neural models, we train the synapse with Poisson uncorrelated pre- and postsynaptic spike trains and compute the expected synaptic weight change as a function of the frequencies of these spike trains. Interestingly, a Hebbian (in the rate encoding sense of the word) BCM-like behavior is recovered in this setup for hippocampal neurons, while dominating depression seems unavoidable for parameter configurations reproducing experimentally observed triplet nonlinearities in layer 2/3 cortical neurons. Potentiation can however be recovered in these neurons when correlations between pre- and postsynaptic spikes are present. We end this chapter by discussing the relation to existing experimental results, leaving open questions and predictions for future experiments.

A set of summary cards of the models employed, together with listings of the relevant variables and parameters, are presented at the end of the thesis, for easier access and permanent reference for the reader.

Chapter 1

General Background

"Mind" can only be regarded, for scientific purposes, as the activity of the brain, and this should be mystery enough for anyone...
Donald Hebb. The Organization of Behavior.

In this chapter, a brief description of the goals and challenges of modern Neuroscience, at the crossroads of multiple disciplines is presented, with a focus on the role of physics in this grand scheme. Fundamental concepts such as the biological and computational aspects of the brain are introduced, and the specific jargon of the field is clarified.

1.1 Scope and challenges of computational neuroscience

Understanding how the brain works is surely one of the greatest challenges for present day science. Containing on the order of a hundred billion neurons [6], with several thousand connections each [30], the human brain constitutes an incredibly complex system, rendering its study (at least in a systematic quantitative way) practically intractable up to the second half of the twentieth century. The need for a greater comprehension of the brain's functioning, is however urgent, with neurological disorders constituting a major source of impairment and accounting for 12% of total deaths globally, according to the *World Health Organization* [88].

In the past fifty years, development of a wide range of experimental techniques, together with the availability of more potent computers, have enabled scientists to begin to shed light on the problem of understanding the brain. As an illustration of this, the number of articles containing the word *neuron*, listed in PubMed

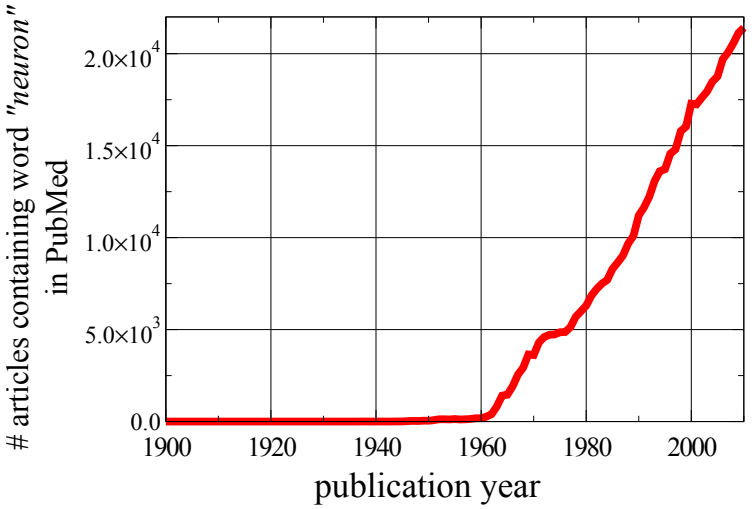


Figure 1.1: *Number of articles containing the word "neuron", per publication year, in the PubMed database (`http://www.pubmed.gov`)*

database [1], per year of publication, is presented in Fig. 1.1. A steady increase in the number of publications in the field, starting in the sixties, is evident in the plot.

The current relevance of the field has recently been made evident by the European Union's decision to grant *1 billion Euro* [1] over the course of ten years to *The Human Brain Project* [2], with a similar initiative, known as the *BRAIN Initiative* announced in the United States by the Obama administration in 2012 [3]).

The term *Neuroscience* refers to the scientific study of the brain, or more generally the nervous system. Given the very nature of its object of study (a biological system, in charge of acquiring, processing, and storing information, with the purpose of performing cognitive and behavioral tasks), neuroscience is an intrinsically interdisciplinary field, attracting the attention of biologists, physiologists, medical doctors, psychologists, computer scientists, mathematicians, engineers, and physicists, among others. Each field, contributing to the general understanding of the brain by bringing its own tools, methodologies, mind-frameworks, and sometimes biases, into the field.

In this context, physics has an important role to play. Devoted to understanding the fundamental laws of nature, physics already counts in many cases with the appropriate mathematical formulation to describe and predict the behavior of complex systems. With a different interpretation of the variables, two problems from completely different fields, may follow the exact same mathematical equations. Even when that is not the case, the analysis and modeling skills proper to physics can

[1] `http://www.pubmed.gov`
[2] `https://www.humanbrainproject.eu`
[3] `http://www.braininitiative.nih.gov`

be applied to all sorts of other systems, once the key ingredients have been identified by those with empiric knowledge of a particular system. This is usually the approach employed in the branch of Complex Systems in general, and in Computational Neuroscience in particular. A usual goal of physicists working in the field of Neuroscience is to construct or find general principles, operating behind the phenomenological rules known to be at play in neural systems. In this sense, a major question regarding the brain is how information is computed and how learning is achieved. Precisely how neurons interconnect, and specifically which principles guide the creation and modification of these connections (a process known as synaptic plasticity) remains, in many respects, an open question. Prototypical examples of the the contribution of physicists to the tackling these questions include:

- John Hopfield's content addressable memory [58]; an artificial neural network in which the evolution of the system is guided by an Ising model-type of Energy function [4], whose minima correspond to the stored memories, providing one of the first models for understanding human memory, using an artificial neural network.
- Leon Cooper's [4] and Paul Munro's contribution to the BCM theory of synaptic plasticity [14].
- Christoph von der Malsburg's contribution to the theory of temporal binding in the brain [114].
- Karl Friston's theory of perception or active inference based on the Free Energy Principle [41].
- Laurenz Wiskott's Slow Feature Analysis (SFA) learning algorithm [116], for extraction of slowly varying features in an input signal, allowing to obtain self-organized receptive fields.

In the present thesis, we will also bring from physics useful tools to study the interaction of neural activity and synaptic plasticity. In particular, this interaction will be analyzed from the perspective of Dynamical System's Theory and Information Theory. But before plunging into the specifics of the problem, a general overview of neural systems, together with the terminology employed in the field, are presented in the following subsections of this chapter.

1.2 Neurons, synapses, and spikes

Constituting roughly half of the cells in the brain [6], neurons are very specialized cells, capable of integration and transmission of electrical and chemical signals from and to other cells (neurons, muscle fibers, among others) [25]. As it can be observed in Fig. 1.2, neurons can be functionally divided into three distinct parts:

[4] Nobel Prize laureate in Physics for the BCS-Theory of Superconductivity

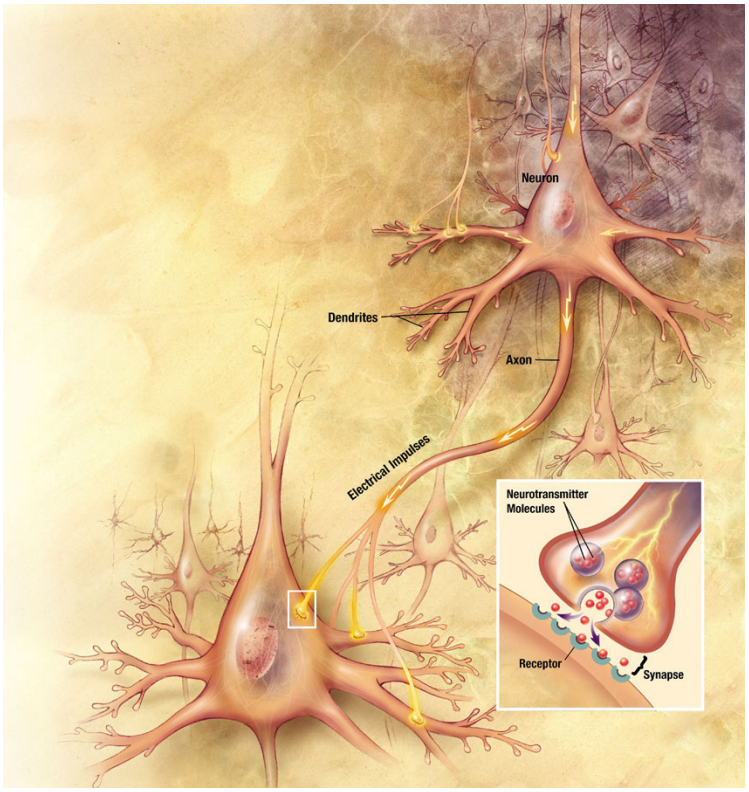


Figure 1.2: *Illustration of the main parts of a neuron, and the chemical communication mechanism between two cells. The three main parts of the neuron are: the* soma *(or body of the neuron), the* dendrites*, which receive the incoming signals from other cells, and the* axon*, which propagates the output of the neuron. In the sketch, one neuron (the presynaptic neuron) emits an electric impulse (after integration of previous inputs), which travels along its neural axon up to the synaptic terminals (see inset), where a chemical substance (termed neuro-transmitter) is released and later detected by another neuron (the postsynaptic neuron), which in turn starts a new integration process. Source: Wikimedia Commons*

the *soma* (or body of the neuron), the *dendrites*, which receive the incoming signals from other cells, and the *axon*, which propagate the output of the neuron to deliver it at the dendritic terminals of other cells.

The connection between two neurons is called *synapse*. These connections can be either electrical (also called *gap junctions*), or chemical. In the first case, an electric current flows directly from one cell to the next. In a chemical type of synapse, on the other hand, one neuron releases a chemical signal (termed *neurotransmitter*), which in turn opens a channel in the target neuron (by binding to the channel's neurotransmitter receptor), letting current flow in or out of the second neuron (see *inset* in Fig. 1.2). In this scheme, the sender of the signal is called *presynaptic* neuron, and the receiver *postsynaptic* neuron.

Both the intracellular and the extracellular medium, consist of a solution with different concentration of ions (Na^+, K^+, Ca^{2+}, Cl^-). The difference in ion concentrations inside and outside the cell produce a voltage difference across the neuron's membrane, termed *membrane potential* [25]. In absence of incoming signals, the membrane potential remains at a constant value, termed *resting potential*, which is usually between $-60\mathrm{mV}$ and $-70\mathrm{mV}$ [44]. Incoming signals via synapses can either hyper-polarize or depolarize the neuron. If a neuron is depolarized enough (reaching a so called *threshold potential*) a chain reaction occurs in the cell, producing a stereotypical voltage excursion, denoted *action potential*, or *spike* (see Fig. 1.3). When this happens, the neuron is said to have *spiked* or *fired*. This strong voltage perturbation, is able to travel along the neuron's axon, and trigger the release of neurotransmitter, and therefore communicate to other neurons that the neuron has fired.

Signals inducing the neuron to depolarize, drive the membrane potential closer to threshold and therefore facilitate firing. Synapses producing this effect are therefore termed *excitatory*. On the contrary, if a signal hyperpolarizes the neuron, the synapse is said to be *inhibitory*. In real physiological neurons one particular connection can only be either excitatory or inhibitory, depending on the particular type of neurotransmitter it employs. Moreover, all the outgoing synapses from one particular neuron are always found to be either excitatory or inhibitory, which allows to classify the whole neuron as either excitatory or inhibitory. This empirical finding is called *Dale's Law* [44]. In modeling of neural networks this condition is sometimes relaxed, building networks of a single type of neuron whose synapses can be either positive or negative, summarizing in a single connection between two neurons the effect that would other wise need to be mediated by a third neuron of the appropriate type.

In section 4.2.1 of **Chapter 4**, we will discuss in more detail the case of excitatory synapses, and in particular the case of synapses that use glutamate as neurotransmitter.

1.3 Of times and rates

In the previous section we described the generation process of an action potential, and how neurons signal to each other that they have fired. In this section, we present two complementary views about how the activity of a neuron can be quantified, and we then present two simplified models that allow to simulate the dynamics of a neuron or network of neurons, in these two paradigms.

The two mentioned paradigms differ in whether one quantifies the activity of neurons by their firing-rate, that is to say the frequency with which a neuron emits a spike; or by the precise timing of each spike. These two views imply two different

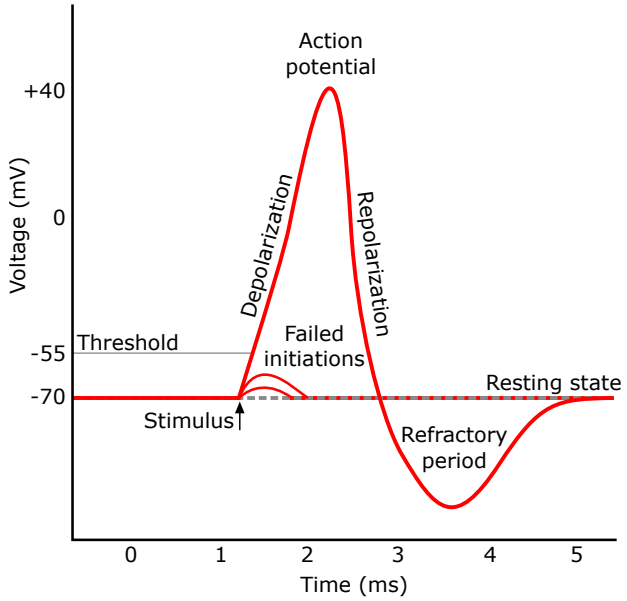


Figure 1.3: *Sketch of an action potential. In absence of incoming signals, the membrane potential remains at the* resting potential. *If a stimulus is strong enough to depolarize the neuron up to the* threshold potential*, a chain reaction occurs in the cell, producing a stereotypical voltage excursion denoted* action potential*, or* spike. *After each spike, the neuron becomes hyperpolarized for a time period denoted* refractory period. *Source: Wikimedia Commons*

possible neural codes. As we will later show throughout the text, both views have their own merit, and allow to explain different experimental phenomena. Depending on whether we work within one paradigm or the other, we will talk about *rate-encoding* or *time-encoding* neurons, respectively.

A second classification, when it comes to neural models is whether one considers the particular spacial structure of neurons when computing the neural activity, or not. In the first case, one talks about *multiple-compartment* models, in which dendrites, the soma, and axon, are considered different compartments with physical extensions that influence the dynamics; and where the membrane potential of the neuron can be defined and evaluated at each compartment. At the other end, as an extension of a typical physical simplification method, one can consider *point neurons*, in which a single scalar quantity defines the neural activity, and where a reduced set of scalar parameters define the whole configuration of the neuron. In this thesis, we will consider this second type of models, sacrificing a more detailed description for analytical tractability. As we will see, this approach allows however to reproduce a wide range of experimental findings.

1.3.1 A rate-encoding neural model

A simple neural model for rate-encoding neurons we will later employ in this thesis, is the linear non-linear model [55]. We consider in this case instantaneous point neurons, in which the activity level y of the neuron is given by:

$$y = g(x) = \sigma(x - b), \qquad \sigma(z) = \frac{1}{1 + \mathrm{e}^{-z}}, \qquad x = \sum_{j=1}^{N_w} w_j (y_j - \bar{y}_j) \,, \tag{1.1}$$

representing the rescaled average firing rate of the neuron (by mapping the neuron's activity range to $[0, 1]$). Function g relates the total integrated input x to the output y of the neuron, and is termed *transfer function* (also sometimes called *activation function*). Here we have chosen for g a monotonically increasing sigmoidal function $\sigma(z)$, which maps any input also to the range $[0, 1]$. In future sections we will discuss several options for g, and their properties. N_w stands for the number of inputs y_j, each of which represents either an external input or the activity of another neuron in the network. w_j represents the connection strength between presynaptic neuron i and the postsynaptic neuron, and b is a bias in the neuron's sensitivity determining how high the overall activity has to be to produce a significant output activity. Finally, the $\bar{y}_j$ represent a trailing average of the input activity, so that only deviations from this average contribute to the integrated input.

In this model, the output of each neuron is then simply a non-linear function of a weighted average of the outputs of the neurons connecting to it.

We have described this model as instantaneous, since the output activity of each neuron is a function of the present activities of the other neurons; no differential equation has been involved so far. While this approach proves useful in simple scenarios, for extended neural networks with recurrent connections, the problem of how to simultaneously update the activities of the neurons in the network arises. To avoid this problem, and a simple extension in the form:

$$\tau \dot{y} = \sigma(x - b) - y, \qquad \sigma(z) = \frac{1}{1 + \mathrm{e}^{-z}}, \qquad x = \sum_{j=1}^{N_w} w_j (y_j - \bar{y}_j) \,, \tag{1.2}$$

can be employed when necessary. Eq. 1.1, becomes in this case the stationary solution to Eq. 1.2, for a constant input.

1.3.2 A spiking neural model

For completion, we present here a model for input integration in spiking neurons. In the present work, we will however study time-dependent plasticity (termed STDP), at the single synapse level only. We will nonetheless need to keep this type of model for neural integration in mind to express the variables in our model

in a language compatible with spike integrations models, for extensions of our work.

The term *Integrate and fire models* refers to a family of models, employed in a wide range of fields, to describe systems of coupled pulsating elements; such as networks of pacemaker cells in the heart [90], flashing extended populations of fireflies [17, 53], and, as in this case, neural networks. In these models, evolution (in between pulses, flashes, or spikes) of a continuous internal state variable V is governed by an equation of the shape:

$$\tau \dot{V} \;=\; f(V) + I. \tag{1.3}$$

τ is here the characteristic adaptation timescale of V, with f condensing the intrinsic dynamics of each unit. I represents the overall input to the unit (both from other units and from external stimuli). Whenever V reaches a threshold value V_θ, a pulse is emitted (the only information carried to other units) and the internal variable is reset to V_{rest}.

In the particular case of neurons, the continuous state variable V in Eq. (1.3), represents the neuron's membrane potential. When this voltage reaches and a threshold value V_θ, a spike is emitted, and the neuron's membrane potential is returned to its resting state, simulating the dynamics of physiological action potentials (see Fig. 1.3). A discrete state variable y indicates whether the neuron has fired a spike ($y = 1$) or not ($y = 0$) at a particular point in time. Consistently with this type of model, in **Chapter 4**, we represent spikes as delta functions, which are numerically implemented as a brief pulse of height 1.

So far, we have not specified the shape of function f. We present here as an option the conductance-based integrate-and-fire (COBA) model from [113], in which the evolution of each neuron i in the network is described by:

$$\tau \dot{V}_i \;=\; (V_{rest} - V_i) + g_i^{ex} \left(E_{ex} - V_i\right) + g_i^{inh} \left(E_{inh} - V_i\right), \tag{1.4}$$

where E_{ex} and E_{inh} represent the excitatory and inhibitory reversal potentials, and τ is the membrane time constant. The conductances $g_i^{ex/inh}$ in (1.4) have the mediate the effect of presynaptic spikes on the postsynaptic neuron. decaying on the other side in absence of inputs:

$$\tau_{ex/inh}\, \dot{g}_i^{ex/inh} \;=\; -g_i^{ex/inh}, \tag{1.5}$$

where $\tau_{ex/inh}$ are the conductance time constants and where incoming spikes from other neurons, on the other hand, produce an increase in the conductances: $g_i^{ex/inh} \longrightarrow g_i^{ex/inh} + \Delta g_i^{ex/inh}$. The size of this change is proportional to the connection strength w between the neurons. Both in the context of rate- and time-dependent plasticity the strength of interaction between two neurons is quantified by the value of w. In the following section we deal with how w is modified and how it evolves in time.

1.4 Synaptic Plasticity and Learning

As it was previously mentioned, a major question regarding the brain is how information is acquired, processed, and stored, and how learning is achieved. Precisely how neurons interconnect, and specifically how these connections evolve according to the activity in the network is still, in many respects, an open question. In section 1.2, we introduced the concept of a synapse: the connection between two neurons. These connections are however not static, they are constantly changing: which neuron is connected to which other neuron, but also, how effective is the connection. The *synaptic efficacy* between a pair of neurons i and j in the network (usually denoted w_{ij}) is a measure of how much impact the activity in neuron j has on the activity of neuron i. The process of modifying the value of a neural parameter in the model is termed in this context *plasticity*. In particular, the action of changing the synaptic efficacy w_{ij} is known as *synaptic plasticity*, which, combined with the adaption of other intrinsic parameters in neurons (*intrinsic plasticity*), is believed to be at the basis of long lasting learning and memory [44].

One of the first theories put forward in this sense was Hebb's rule [56], which could be roughly summarized as: *neurons that fire together, wire together*. But it was only in the 70s that the first precise mathematical formulations proposing how that wiring takes place were formulated, being Oja's rule [86] and BCM [14] among the ones with the greatest impact. These learning rules, in the form of differential equations, were based on experimental evidence indicating that synapses in the cerebral cortex are bidirectionally modified by sensory experience [61, 62]. These first rules considered the firing-rate (the frequency at which a neuron fires), as the fundamental unit of information transmitted between neurons. These neuronal models are thus referred to as *rate-encoding neurons*. In more recent years, experiments were performed that could only be explained by taking into account the specific timing of pairs of pre- and post-synaptic spikes, a form of plasticity known as *spike-timing dependent plasticity (STDP)* [43], generating thus two major lines of research in Synaptic Plasticity, namely rate-encoding vs. time-encoding modeling. How exactly this two forms of plasticity interact in different neurons, remains unclear and currently new rules of plasticity are still regularly formulated that consider more and more detailed features in the time structure of spike patterns such as triplets [91] or even quadruplets of spikes [115].

While successful at explaining experimental findings to which they were tailored, many of these rules are circumscribed to the specific setting at hand, without incorporating the results into a broader theory in the form of higher principles governing learning processes in general in the brain. As counter-side of this, other learning rules, derived from higher principles, do not yet count with a possible biological implementation. Ideally, one would wish to find a reduced set of governing principles determining the evolution of all the relevant variables in the brain. These principles should be expressed in a clear mathematical formulation allowing to

derive the equations of evolution for these variables. Finally, the results should also be independent of the fine tuning of internal parameters of the model; any result should be robust enough to be feasible in a biological context of great variability. With these requisites in mind, the concept of *Self-Organized Criticality (SOC)* [8] seems to provide an adequate framework in which to formulate these governing principles. Dynamical systems presenting SOC are those where a critical state attracts the dynamics of the system and this behavior is independent of the tunning of the model's parameters. The system self-regulates to reach equilibrium.

An example of such self-regulating mechanisms is that of *Homeostasis*, ubiquitous in the bodies of living organisms. By this process, a certain property is regulated to keep a constant value. Regulation of body temperature, for instance, allows an animal to survive in a wide range of exterior temperatures. In an analogous way, homeostasis permits a neuron to function in a wide range of conditions [110]. These types of mechanisms are then good candidates to express the higher principles we are looking for, counterparts of low level formulations that build up on local properties.

The challenge is then how to bridge these two worlds; how to formulate a principle such as that of homeostasis in a precise and useful mathematical way, and how then to derive the local plasticity mechanisms. Procedures in other fields of Physics, such as classical mechanics already exist that permit to determine the evolution of every variable in a system by minimizing a certain value over the all the possible paths in configuration space. Such a procedure can be extended by defining *Objective functions* for neural systems [64]. Furthermore, by use of informational theoretical measures, such as *Information Entropy* [77], one can express several Objective functions in terms of neurons' firing statistics and study how different principles, present simultaneously in the brain, could interact [108]. Moreover, the concept of homeostasis for a single scalar value can be extended by use of this formalism to regulate the system's evolution in terms of the whole distribution of states in a procedure known as *polyhomeostasis* [78].

1.5 Complementary approaches to Plasticity

At least two complimentary approaches exist when developing a synaptic plasticity rule: one may either employ a so called bottom-up approach, building up rules which reproduce certain aspects of experimental observations, in terms of the biological elements known to be involved in the process (as we will later do in **Chapter 4**); or a top-down approach, where synaptic plasticity rules are derived from a guiding principle, expressed in terms of desirable goals for the brain (metabolic, computational, or related to the stability of the system). This is the approach we follow in **Chapters 2** and **3**.

Examples of bottom-up methods include the formulation long-term potentiation (LTP) and long-term depression (LTD) in terms of the chemical and configurational effects of spikes in pre- and postsynaptic neurons [48, 66, 97, 100, 111]. An alternative approach is to build the model as a purely phenomenological rule, without trying to establish a connection to the particular elements that the variables involved might represent [7,42]. A wide spectrum of possibilities then exist between simplistic phenomenological rules, and highly detailed biophysical models. In **Chapter 4** we will show how a compromise can be made, by formulating a plasticity rule in terms of the key biological ingredients thought to be involved, albeit with a highly simplified mathematical expression.

A useful concept when following a top-down approach, is that of *objective functions* (also termed generating functionals within dynamical system theory [51, 74]), which allow for a wide theoretical perspective of synaptic plasticity (and learning in general), in the context of dynamical systems. These objective functions allow us to express general principles in the shape of a concrete mathematical expression, from which then the equations for the evolution of a system (in our case the plasticity rules) can be derived [9, 64]. In the following section we discuss a particular family of tools, originating from the field of information theory, which are helpful when dealing with objective functions in the context of probabilistic systems.

1.6 Use of information-theoretical quantities

As previously mentioned, when one deals with probabilistic systems whose goal is to process information, tools from the field of information theory come in handy. In particular, a wide variety of information measures stem from Shannon's information entropy [77], defined for a probability distribution $P(x)$ as:

$$H \;=\; -\int p(x)\log\left(p(x)\right)dx\;. \tag{1.6}$$

A family of functions can then be defined out of the information entropy, such as: the joint entropy between two processes, the conditional entropy, or the mutual information between input and output of a process [77]. These measures allow to quantify how much information is transmitted or lost in a given process, and are then helpful to express computational guiding principles in the form of objective functions, from which plasticity rules can be derived [64, 67, 79, 92, 93, 106, 108]. In the past, ideas such as as maximizing the output entropy of a system giving certain constraints have been used as a way of finding parameter configurations that maximize the representational capabilities of the neural code [108]. In other work, the mutual information between input and output of a system (or the transmitted information) is maximized, for instance for signal separation and deconvolution in networks [9], or to derive receptive fields in primary visual cortex [25]. This type of approach brings useful information to the discussion about to what extent is

the brain optimal, by comparing the predictions of such computational optimality principles to the connections actually observed in the brain.

Another associated measure, and one we will employ in this work, is the Fisher Information, defined as:

$$F_\theta \;=\; \int p(y) \left(\frac{\partial}{\partial \theta} \ln \big(p(y)\big) \right)^2 dy \;, \tag{1.7}$$

which is a measure of the average sensitivity of the probability distribution $p(y)$ with respect to parameter θ. This quantity will become useful for the formulation of our guiding principle and we will come back to this point in section 2.2.1, when we present our objective function.

To conclude this section, we present other past uses of the Fisher Information as a criterion for optimality, in information transmission and parameter estimation.

The Fisher Information can be related to the mutual information between input and output of a probabilistic system [16]. If one considers θ not as a parameter, but as a variable, which we will call x to avoid confusion, where x represents the input to the system, and $p(y|x)$ is the output probability for a fixed input x, (1.7) can be rewritten as

$$F_x \;=\; \int p(y|x) \left(\frac{\partial}{\partial \theta} \ln \big(p(y|x)\big) \right)^2 dy \;, \tag{1.8}$$

which then represents the sensitivity of the output with respect to the input for a particular input. If one finally averages (1.8) over the input probability distribution $p(x)$, one obtains an alternative measure of the information that the output conveys about the input [16].

Finally, the Fisher Information, or rather its inverse, is commonly employed, via the Cramer-Rao theory [52, 89, 99], as a lower bound for the variance when estimating an external parameter. In this case, the external parameter can be better estimated when the Fisher information is larger, that is when the distribution considered is highly sensitive to the parameter of interest. In this context one *maximizes* the Fisher Information with respect to an external parameter. In our work, however, we are not interested in the estimation of an external parameter, but rather in the adaption of internal network parameters, namely the strength of the synaptic weights. For this reason, in the following chapter we will follow a very different approach, actually *minimizing* the Fisher information with respect to internal parameters of our neural system.

Chapter 2

An objective function for self-limiting Hebbian learning rules

Echeveste, R., & Gros, C. (2014). *Generating Functionals for Computational Intelligence: The Fisher Information as an Objective Function for Self-Limiting Hebbian Learning Rules.* Frontiers in Robotics and AI, 1, 1.

In section 1.4, the need for a set of principles guiding plasticity, and learning in general, was introduced. In this chapter, we discuss how generating functionals can serve to guide the evolution of dynamical systems and, in particular, constitute a useful formalism in which to frame synaptic plasticity. Working within the framework of rate-encoding neurons, we propose and examine here a novel objective function from which plasticity rules for the afferent synaptic weights are then derived. These adaption rules are Hebbian and self-limiting.

The behavior of the new learning rules is then examined via a series of numerical simulations in various scenarios, observing that the synaptic weight vector aligns with the direction of the first principal component when the input distribution closely resembles a multivariate normal distribution. We will show however that when two or more input directions have the same standard deviation, but differ in their higher statistical moments, directions characterized by a high negative excess kurtosis, are preferentially selected. In particular, the rule tends to perform binary classification when the input distribution is bimodal in at least one direction.

Finally, we test the robustness in performance and show how a full homeostatic adaption of the synaptic weights results as a by-product of the objective function minimization. This self-stabilizing character makes stable online learning possible for arbitrary durations. The neuron is however able to acquire new information if the input statistics are modified at a certain point of the simulation; showing however distinct timescales for learning and unlearning.

2.1 Introduction

As presented in section 1.4, synaptic plasticity refers to the modification of the strength of synaptic connections as a function of pre- and postsynaptic neural activity. In this chapter we are interested in developing a synaptic plasticity rule within the framework of rate encoding neurons (see section 1.3). A minimal requirement for such rules, if one wants them to reproduce plasticity in the brain, is for them to be Hebbian, as discussed in **Chapter 1**. This is however not enough for stable learning. The principle of Hebbian learning on its own [57], is not stable; in the sense that strong synapses -which induce correlations in neural activities- are in turn made even stronger, leading (without an additional homeostatic regulative processes [110], such as synaptic scaling [2]) to runaway synaptic growth. Namely, it is not enough to state when should weights grow, but it is also necessary to define when should that growth stop, or even be reversed.

From a computational point of view, people have tackled this problem in different ways. Either by re-normalizing the synaptic connectivity matrix every certain number of learning steps, or by adding an explicit weight decay term to their learning rules (see for instance [14, 38, 46, 86]). What we will show here is how, from one single principle, one can obtain a learning rule that is -"out of the box"- Hebbian and self-limiting. Moreover, this will not achieved, as we will show, by an explicit constraint on the synaptic weights, but will result from a constraint on the desirable activity range of the neuron.

It has been shown in the past that Hebbian learning, inducing synaptic competition, tends to result in *principal component analysis* (PCA) [82, 86], in the sense that, after learning, the neuron becomes sensitive to input directions of high variance, by means of aligning its synaptic weight vector to the input direction having the highest variance, usually denoted the *first principal component* (FPC). In this way, neurons would select information coming from directions potentially less affected by noise. One of the first things we will test is whether this feature is present with our rule and in which context.

Now, an interesting question is, beyond this tendency to produce PCA, what computational capabilities will a neuron have, depending on the details of the learning rule. Concretely, to what features will it become selective if we make the input covariance matrix close to unity, but allow for different higher moments of the input distribution. This is a highly relevant question for biological and artificial applications since deviations from Gaussian statistics -given by higher moments of the input distribution- may contain relevant information, as observed, for instance in natural image statistics [101, 102]. For this reason, we will perform a series of numerical tests in this chapter to try to determine to what features is the learning rule selective.

A normal distribution is characterized only by its mean and standard deviation. So one could choose any of the higher moments of an input distribution to quantify deviations from normality. For symmetric distributions, the first non-vanishing higher moment is the excess kurtosis [26] (or forth moment of an input distribution). This quantity is, by construction, zero for normal distributions. A neuron may then be selective to either large positive or negative excess kurtosis, and indeed, examples exist of neural models which are selective for directions with heavy tails (corresponding to a large positive excess kurtosis) [108], as a way of detecting non-normal directions.

In this chapter, on the other hand, we study a rule which allows the neuron to discover directions having large *negative* excess kurtosis, an example of which are bimodal directions. Performing a binary classification, by linear discrimination of objects in the input data stream, has been proposed as a central aspect of unsupervised object recognition, for instance while performing slow feature analysis [28, 117]. Using supervised learning to train a neuron to binary classify a set of linearly separable data is already a well documented process [68]. What we will show here, however, is how a single neuron, guided by an objective function favoring input directions with large negative excess kurtosis, is able to perform this task unsupervised.

A second application for the preference for non-Gaussian input directions in general, and of our rule in particular, will be later discussed in **Chapter 3**, section 3.4.

As it was mentioned in section 1.5, one can think of at least two complimentary approaches when developing a synaptic plasticity rule. One may either employ a bottom-up approach, building rules which reproduce certain aspects of experimental observations, in terms of the biological elements involved (as we will later do in **Chapter 4**); or a top-down approach, where an objective function is constructed in terms of general principles, from which the plasticity rules are then derived [9, 64]. Objective functions (also termed generating functionals in the context of dynamical system theory [51, 74]), allow for a wide theoretical perspective, and have been used for instance to perform a stability analysis of Hebbian-type learning in autonomously active neural networks [29].

As discussed in section 1.6, the Fisher information [16] is a measure of the sensitivity of a probability distribution with respect to a given parameter. While usually associated with the task of parameter estimation via the Cramér-Rao bound [52, 89, 99], it is its property as a sensitivity encoder, which makes it a useful tool, both in the context of optimal population coding [10, 36, 70], or as in the present work, for the formulation of objective functions. Indeed, this procedure has been successfully employed in the past in Physics, to derive, for instance, the

Schrödinger Equation in Quantum Mechanics [96].

The proposal in the present work, is that a self-limiting learning rule can be expressed in terms of a principle we denote *the stationarity principle of statistical learning*, stating that:

"Once a neuron has extracted the relevant features of a stationary input distribution, the output distribution should also be stationary."

For this to be possible in the context of a noisy environment, we require this final state to be stable, in the sense that the output probability distribution should be locally insensitive to changes in the synaptic weights. This is where the Fisher Information comes into play, allowing us to formalize this condition of minimal sensitivity, as a minimal Fisher Information condition. In a multidimensional parameter space (as is the case of the synaptic weights) a particular generalization of the one-dimensional Fisher Information will be chosen to ensure local learning rules. Indeed, as we will later show in section 2.3, the synaptic plasticity rules obtained in this way have a set of attractive features; being Hebbian, local, and at the same time, self-limiting.

As mentioned in section 1.6, this is not the first initiative to use tools from information theory to derive plasticity rules, other examples include the use of the transfer entropy [112], or the Kullback-Leibler divergence to a target distribution, which one may use for instance to adapt intrinsic neural parameters [79, 108]. We will indeed use such an intrinsic plasticity rule in our work, which will complement the synaptic plasticity rule we derive. With the right choice of target distribution, minimizing the Kullback-Leibler divergence can be equated to maximizing Shannon's information content or output entropy of the neuron's firing statistics [108]. Interestingly, the combination of both rules, will result in an effective sliding threshold of the synaptic activity, similar in a broad sense to that of the BCM rule [14]. In **Chapter 3**, we will be able to expand on these ideas, once we are able to study the attractors of the learning rule analytically.

2.2 Theory

In this chapter we consider rate-encoding (point) neurons, as presented in 1.3. Namely, we will quantify the activity level of the neurons by their rescaled output firing rate $y \in [0, 1]$, where for each neuron the output y is a monotonic function g of its integrated input x (also usually denoted in this context as the membrane potential) computed as:

$$y = g(x), \qquad x = \sum_{j=1}^{N_w} w_j (y_j - \bar{y}_j) \; . \tag{2.1}$$

N_w is the dimension of the input space, that is the number of input synapses. w_j and y_j are respectively the synaptic weights and input firing rates. The values $\bar{y}_j$ represent the trailing averages of y_j,

$$\frac{d}{dt}\bar{y}_j = \frac{y_j - \bar{y}_j}{T_y} \,, \tag{2.2}$$

with T_y the averaging time-scale, so that only deviations from the average firing rate value influence postsynaptic activity. This is a frequent assumption for synaptic plasticity, usually implemented by either trailing averages or by preprocessing of the input data. The synaptic weights w may take, for simplified rate encoding neurons, both positive and negative values. This is a simplification. As we commented in section 1.2, real neurons respect *Dale's Law*: they are either excitatory or inhibitory.

So far we haven't specified the functional form of g, and since the exact shape of the learning rules will depend on g, throughout this work we will explore several possibilities for it. We begin in this section by considering the sigmoidal transfer function g we presented in (1.1), which we here recall:

$$y = g(x) = \sigma(x - b), \qquad \sigma(z) = \frac{1}{1 + \mathrm{e}^{-z}}, \tag{2.3}$$

where $\sigma(z)$ (usually denoted Fermi function in Physics), has a fixed gain or slope. The neuron has therefore a single intrinsic parameter, namely the bias b. We have not included an explicit gain parameter (as in [108]), acting on x since any multiplicative constant can in our case be absorbed into the $w_j s$.

In the following sections we will first define a guiding principle (section 2.2.1), and then derive synaptic plasticity rules for the synaptic weights (section 2.2.2), which we will analyze in conjunction with an intrinsic plasticity rule for b.

2.2.1 Motivation in terms of the Fisher Information

In section 1.6, we introduced the Fisher information:

$$F_\theta \;=\; \int p(y) \left(\frac{\partial}{\partial \theta} \ln \big(p(y)\big) \right)^2 dy \,, \tag{2.4}$$

which encodes the average sensitivity of a given probability distribution $p(y)$ with respect to a certain parameter θ. As mentioned in section 2.1, the proposal of the present work is to derive a synaptic plasticity rule from the stationarity principle of statistical learning, stating that, for a stationary input distribution, once the extraction of the relevant features has been completed, the output probability distribution should also be stationary. In terms of the stability of such a final state, we should expect this output distribution to be stable, and therefore minimally sensitive, to

local changes of the weight vector. The Fisher Information, being a measure of sensitivity for probability distributions, will therefore be a useful tool in this regard.

If we only had one incoming synapse y_1, integrated by a single parameter w_1, we could simply consider:

$$\mathcal{F}^{syn}_{N_w=1} \;=\; \int \left(w_1 \frac{\partial}{\partial w_1} \ln \left(p(y(y_1)) \right) \right)^2 p(y_1) dy_1, \tag{2.5}$$

where $w_1 \partial / \partial w_1$ is a dimensionless differential operator corresponding to the log-derivative of the synaptic weight. Since we are interested in a stochastic learning rule for the synaptic weights, with one input instance per time-step, we have defined the average sensitivity in terms of the input probability distribution, with the output probability distribution being a function of the input distribution:

$$p(y(y_1))dy = p(y_1)dy_1, \qquad p(y(y_1)) = \frac{p(y_1)}{\partial y / \partial y_1} \,. \tag{2.6}$$

In this way, we can rewrite 2.5, exclusively in terms of y_1 and w_1 as:

$$\mathcal{F}_{N_w=1} \;=\; \int \left(w_1 \frac{\partial}{\partial w_1} \ln \left(\frac{p(y_1)}{\partial y / \partial y_1} \right) \right)^2 p(y_1) dy_1 \,. \tag{2.7}$$

The problem arises when one tries to extend this concept to a multidimensional input space with a weight vector of size N_w. One possible approach would be to consider the full Fisher Information matrix, defined by all the partial derivatives with respect to every synaptic weight. This approach has the serious problem of being non-local, in the sense that the cross terms produce a learning rule for each synapse that explicitly depends on the value of every other synapse. Here we are interested, however, in local learning rules, in which each synapse is only allowed to "know" its own value, plus neuron-wide variables x and y. Defining with $\mathbf{y} = (y_1, \ldots, y_{N_w})$ the vector of afferent synaptic weights and with $p(\mathbf{y})$ the corresponding probability distribution, we propose the following extension, which we denote the *Local Synapse Extension*:

$$\mathcal{F}^{syn} \;=\; \int \left(\sum_{j=1}^{N_w} w_j \frac{\partial}{\partial w_j} \ln \left(\frac{p(y_j)}{\partial y / \partial y_j} \right) \right)^2 p(\mathbf{y}) d\mathbf{y} \,. \tag{2.8}$$

This form of extension will, as will we see, yield local learning rules, avoiding explicit *cross-talk* between the synapses.

The integral in 2.8 is weighted by the input probability distribution, and can therefore be expressed as an expectation value:

$$\mathcal{F}^{syn} \;=\; E\left[f^{syn} \right], \qquad f^{syn} = \left(\sum_{j=1}^{N_w} w_j \frac{\partial}{\partial w_j} \ln \left(\frac{p(y_j)}{\partial y / \partial y_j} \right) \right)^2 \,. \tag{2.9}$$

Indeed, we begun by considering the Fisher Information as the average sensitivity (which is nothing more than the expected value of the derivative of the probability distribution) with respect to the parameter in consideration. Now, by rewriting the derivatives in the previous expression in terms of x, and noting that the input distribution $p(y_j)$ does not depend explicitly on w_j, f^{syn} can be expressed as:

$$f^{syn} \;=\; \left(N + \frac{xy''}{y'}\right)^2 \;, \tag{2.10}$$

where y' and y'' represent the first and second derivatives of the transfer function $y = g(x)$ with respect to x. The constant N, originally comes out as an N_w in the derivation, which we chose to handle as an independent parameter for the learning rule. This last point will be discussed in more detail in the following chapter, when the properties of the learning rule are studied analytically.

In the form of equation 2.10, the dependence of the objective function on the exact choice of the transfer function $y = g(x)$ has become explicit. In the following chapter the effect of this particular choice will be studied in detail. In this chapter, however, we will continue to work with the already presented sigmoidal 2.3. With this particular choice, the objective function takes the form:

$$\mathcal{F}^{syn} \;=\; E\left[f^{syn}\right] \;=\; E\left[\left(N + x(1-2y)\right)^2\right]. \tag{2.11}$$

It is from this final form that we will derive the learning rules in the following section.

2.2.2 Derivation of the learning rules

As previously stated, we are interested in deriving local, instantaneous plasticity rules, defined in terms of the pre- and postsynaptic firing rates y_j and y. We hence proceed by performing a *stochastic (or online) gradient descent* [15] on the objective function (2.11). That is, instead of taking the gradient of the full expectation value of the objective function $\mathcal{F}^{syn}$ (denoted as *batch gradient descent*), which would require individual neurons to count with (at least) an estimate of the probability distributions, we compute the gradient of the inner function f^{syn}, and take one learning step per input instance:

$$\dot{w}_j \;\propto\; -\frac{\partial}{\partial w_j} f^{syn} \;=\; -\frac{\partial}{\partial w_j}\left[\left(N + x(1-2y)\right)^2\right]. \tag{2.12}$$

Making use of the fact that $\partial x/\partial w_j = (y_j - \bar{y}_j)$, and that, for the Fermi sigmoidal function (2.3), $\partial y/\partial w_j = (y_j - \bar{y}_j)(1-y)y$. We can finally write:

$$\dot{w}_j \;=\; \epsilon_w G(x) H(x)(y_j - \bar{y}_j), \tag{2.13}$$

with

$$\begin{aligned} G(x) &= N + x(1 - 2y(x)), \\ H(x) &= (2y(x) - 1) + 2x(1 - y(x))y(x)\;. \end{aligned} \tag{2.14}$$

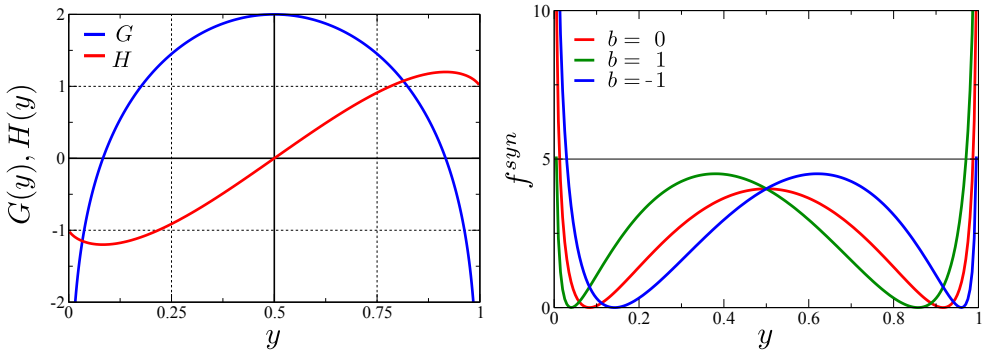


Figure 2.1: *The roots of the adaption factors. Left: Plasticity functions G and H, as defined by (2.14), here expressed entirely in terms of the output activity $y \in [0, 1]$, for clarity. As an illustration, parameters $b = 0$ and $N = 2$, have been used. H (when multiplied by $(y_j - \bar{y}_j)$) constitutes the Hebbian part of the rule, while G acts as a limiting factor, reverting the sign of (2.13) when the neural activity comes too close its extreme values of $0/1$. Right: Effect of the bias b on the objective function f^{syn} of Eq. (2.11).*

Since y is a monotonic function of x, synaptic functions G and H can also be entirely expressed either in terms of x or y, as shown in Fig. 2.1. As it can be observed in the plot, $H(y)$ is an essentially linear function with positive slope within most of the activity range of the neuron, only saturating for $y \to 1/0$. Therefore, the product $H(y)(y_j - \bar{y}_j)$ constitutes the Hebbian part of the synaptic plasticity rule (2.13), increasing the size of the synaptic weight w_j whenever the input y_j and the output y are correlated.

Function $G(y)$, on the other hand, serves as a limiting factor, reverting the sign of the learning rule if the neural activity approaches the extremes, $y \to 1/0$, keeping the membrane potential x close to the roots of $G(x)$. For this reason, the synaptic weight will also remain finite, making the adaption rules in (2.13) self-limiting.

As a side note, for $N = 2$ the two synaptic functions are proportional to each other's derivatives: $H(x) = -G'(x)$, and $G(x) = 2y(1 - y)H'(x)$, and the reversal of the learning rate takes place when the Hebbian factor is maximal/minimal [33]. For this reason, we will often employ $N = 2$ for our simulations. The effect of parameter N on the learning rule is discussed in detail in **Chapter 3**.

Although the learning rule presents no explicit cross-terms, as avoided by the local synapse approximation, it is important to note that synaptic competition is present implicitly in the rule via the membrane potential x, which integrates all individual contributions.

Finally, regarding the adaption rate ϵ_w in Eq. (2.13), we have tested that the Plasticity rule (2.13) works robustly for a wide range of parameter values. For all simulations we will present in this chapter we have used $\epsilon_w = 0.01$.

So far we have focused only in the adaption of the synaptic weights. We also count, however, with an intrinsic parameter in the model, namely the bias b, which could be regulated. In section 2.2.3, we study how this parameter can be adapted to scale the average activity level up or down, and how both rules (synaptic and intrinsic) interact.

2.2.3 Intrinsic plasticity rule

Parameter b was introduced in (2.3) as a bias, shifting the response curve of the neuron. Its adaption, via an intrinsic plasticity rule, will then determine the overall neural activity level.

There is, however, a second effect of b on the neuron's dynamics. Since the synaptic plasticity rules depend on the output y, which depends in turn on b, the bias also shifts our objective functions. As an illustration, on the right-hand side of Fig. 2.1, f^{syn} as a function of y is presented for several values of b.

To quantify this effect, we first invert g via $x = b - \log\big((1-y)/y\big)$ and express the synaptic function H solely in terms of y:

$$H(y) = (2y-1) \, + \, 2y(1-y)\left[b - \log\big((1-y)/y\big)\right] \, . \tag{2.15}$$

For $b = 0$, we have $H(1/2) = 0$, which, as we can observe in Fig. 2.1, is the only root of H. We call this root y_H^*. The bias b then regulates the position of y_H^*, representing the crossing point from anti-Hebbian (for low neural activity $y < y_H^*$) to Hebbian learning (for large firing rates $y > y_H^*$). The dependence of y_H^* on b, together with that of the roots of G (which will be discussed in section 2.2.4), are shown in Fig. 2.2. The monotonic relation between the Hebbian turning point and b, allows us to consider the bias also as a sliding threshold, analogous to the one present in the BCM theory [14], which regulates the crossover from anti-Hebbian to Hebbian learning with increasing output activity and which is adapted in order to keep the output activity within a given working regime. In BCM theory, the threshold is computed explicitly as a function of the mean activity. Here, however, we will adapt the bias by use of an additional objective function, as proposed in [108].

The Kullback-Leibler divergence

$$D_{KL} = \int dy \, p(y) \log\left(\frac{p(y)}{q(y)}\right), \tag{2.16}$$

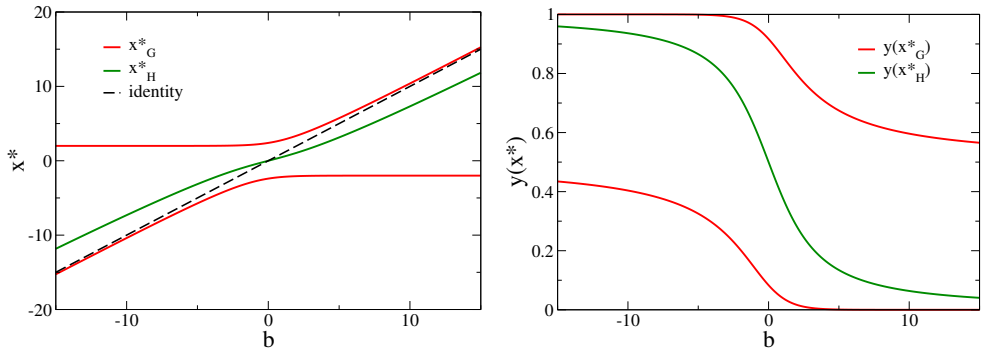


Figure 2.2: *Roots of the Hebbian and limiting factor in the plasticity rule. Left: The roots $G(x^*_{0,1}) = 0$ and $H(x^*) = 0$ (see Eq. (2.14)), respectively, as a function of the bias b. Right: The corresponding values $y(x^*)$ for the output activity. Note: the roots never cross.*

is a measure of the distance between two probability distributions $p(y)$ and $q(y)$ [77]. In particular, it can be used to quantify the distance between the actual firing-rate distribution $p(y)$ and a given target distribution $p_\lambda(y)$:

$$D_{KL} = \int dy\, p(y) \log\left(\frac{p(y)}{p_\lambda(y)}\right), \qquad p_\lambda(y) = \frac{\mathrm{e}^{\lambda y}}{N_\lambda}, \tag{2.17}$$

where the exponential distribution chosen here, maximizes the entropy, and therefore the information content, of the neural activity given the constraint of a fixed mean μ. D_{KL} will be minimal if $p_\lambda(y)$ is approximated as well as possible by $p(y)$, maximizing the output entropy. We observe that for $\lambda \to 0$ a uniform target distribution is obtained with a target mean activity $\mu \to 0.5$.

The bias b can therefore be adapted in order to minimize D_{KL}. The procedure, involving a very similar derivation via stochastic gradient descent to the one here presented for the synaptic weights has already been developed in the past (see [75, 108]) and will not be presented here. We simply present the final expression for the intrinsic adaption that we will use in this chapter for the numerical simulations:

$$\dot{b} \;=\; -\epsilon_b \big(1 - 2y + y(1-y)\lambda\big), \tag{2.18}$$

where ϵ_b is the adaption rate for the bias. For the simulations carried out in this chapter, we have used $\epsilon_b = 0.1$. We observed the actual value of the adaption rate had only a marginal influence on the overall behavior of the adaption processes.

As a final remark, we note that minimizing the Kullback-Leibler divergence and the Fisher information are examples of polyhomeostatic optimization [78, 79], as one optimizes an entire probability distribution function, here $p(y)$, instead of a single scalar quantity. This concept is an extension of the concept of basic homeostatic control, which aims to regulate a single scalar quantity, such as the mean firing rate.

2.2.4 Roots of the limiting function G

In the previous section we discussed the behavior of y_H^*: the root of the Hebbian part of the synaptic plasticity function. The limiting factor $G(x)$, on the other hand, has two roots $x_G^*(1)$ and $x_G^*(2)$, (see Fig. 2.2). For $b = 0$ the roots are symmetric around 0, and one finds $x_G^* \approx \pm 2.4$, which corresponds to firing-rates $y_G^* = 0.083$ and $y_G^* = 0.917$ respectively. The self-limiting feature of the synaptic plasticity rules (2.13) results from the two roots of $G(x)$, as both too large and too low activity levels will reverse sign of the learning rule.

Moreover, as one can see in Fig. 2.1, the roots of G correspond to the minima of f^{syn}. Since the learning rule is looking to minimize f^{syn}, the fact that G has two roots will generate a tendency to perform a binary classification, setting it apart from other objective functions for synaptic plasticity rules [64]. This feature will become evident in the numerical simulations of section 2.3, and will be discussed in detail in the following chapter, when the attractors of the learning rules are computed analytically. As a first illustration of this behavior we consider the case of a set of discrete input patterns

$$\mathbf{y}^\eta, \qquad \eta = 1, \dots, N_{patt} \ , \tag{2.19}$$

where the number of input patterns N_{patt} is smaller than the number of afferent neurons ($N_{patt} \leq N_w$) and where at each time-step we randomly select the inputs $(y_1, \dots, y_{N_w}) = \mathbf{y}$ out of the set (2.19). Under this conditions, we find numerically that the learning rules result in a synaptic vector $\mathbf{w}$ dividing the space of input patterns into two groups:

$$\begin{array}{rcll} \mathbf{w} \cdot (\mathbf{y}^\eta - \bar{\mathbf{y}}) & = & x_G^*(1) & \text{for} \quad \gamma N_{patt} \quad \text{states} \quad \mathbf{y}^\eta \\ \mathbf{w} \cdot (\mathbf{y}^\eta - \bar{\mathbf{y}}) & = & x_G^*(2) & \text{for} \quad (1-\gamma) N_{patt} \quad \text{states} \quad \mathbf{y}^\eta \end{array} \ , \tag{2.20}$$

which is a solvable set of N_{patt} equations for N_w variables $(w_1, w_2, \dots)$. Here $\bar{\mathbf{y}} = \left(\sum_\eta \mathbf{y}^\eta \right) / N_{patt}$ is the mean input activity, and γ and $(1-\gamma)$ are the fractions of patterns mapped to $x_G^*(1)$ and $x_G^*(2)$, respectively. Since γ determines the amount of times the output y will be low or high, and therefore the mean firing rate of the neuron, it is determined self-consistently via the adaption of the bias b, trying to approximate as closely as possible the target firing-rate distribution $\propto \exp(\lambda y)$, see Eq. (2.17).

Since the membrane potential $\mathbf{x} = \mathbf{w} \cdot \mathbf{y}$ takes in the end two values for all inputs $\mathbf{y}$ drawn from the the set of input patterns, the result of this learning procedure is the binary classification of the N_{patt} vectors.

2.3 Numerical results for continuous input distributions

In the previous section we had a first flavor of what the learning rule can perform, with a computational example consisting of the binary separation of a discreet input set. To test the learning behavior of the neuron described by rules (2.13 and 2.18), when presented with different continuous stationary input distributions, a series of numerical simulations have been performed, aimed at assessing the neuron's capabilities for principal component extraction and linear discrimination. In section 2.3.3, these results are then extended to a scenario of time-varying input distributions, where a fading memory effect found.

2.3.1 Principal component extraction

We begin by considering the case of N_w input neurons with Gaussian activity distributions $p(y_j)$ (truncated so that $y \in [0, 1]$), where a single direction has larger standard deviation σ, and all other $N_w - 1$ directions have a smaller standard deviation of $\sigma/2$, as illustrated in Fig. 2.3(A). We have selected, without loss of generality, y_1 as the direction of the principal component, since it will make the visual interpretation of the evolution of the synaptic weights simpler. It is important to note that, since the input x is computed as the scalar product of the input vector and the synaptic weight vector, it is rotational invariant, and so are the plasticity rules. We have nonetheless verified this independence by running simulations with dominant components selected randomly in the space of input activities.

In Fig. 2.3 we present the numerical results for a neuron with $N_w = 100$ input directions, where we have taken $\lambda = -2.5$ for the target distribution $p_\lambda(y)$ (see Eq. (2.17) in section 2.2). For the initial conditions, the synaptic weights $\{w_j\}$ were chosen initially small (randomly drawn from $[-0.005 : 0.005]$), so that the learning rule is initially exclusively Hebbian, that is to say the membrane potential x is substantially smaller than the roots x_G^* of the limiting factor $G(x)$ (see Fig. 2.3(B) where x and the roots x_G^* are given by the blue/red dots).

We observe that Hebbian learning leads to larger weights, with the weight along the first principal component (here w_1, red line in Fig. 2.3(F)) taking the largest value. As the membrane potential x grows, it starts to cross the roots x_G^* of the limiting factor $G(x)$, and a stationary state results. This is evidenced by the way the weight along the principal component saturates. The smaller weights, corresponding to the directions of small standard deviation, seem to perform a random walk around 0. We find that this stationary state, with continuously ongoing online learning, remains stable for arbitrary long simulation times. In **Chapter 3** we will be able to explain this solution and its stability analytically.

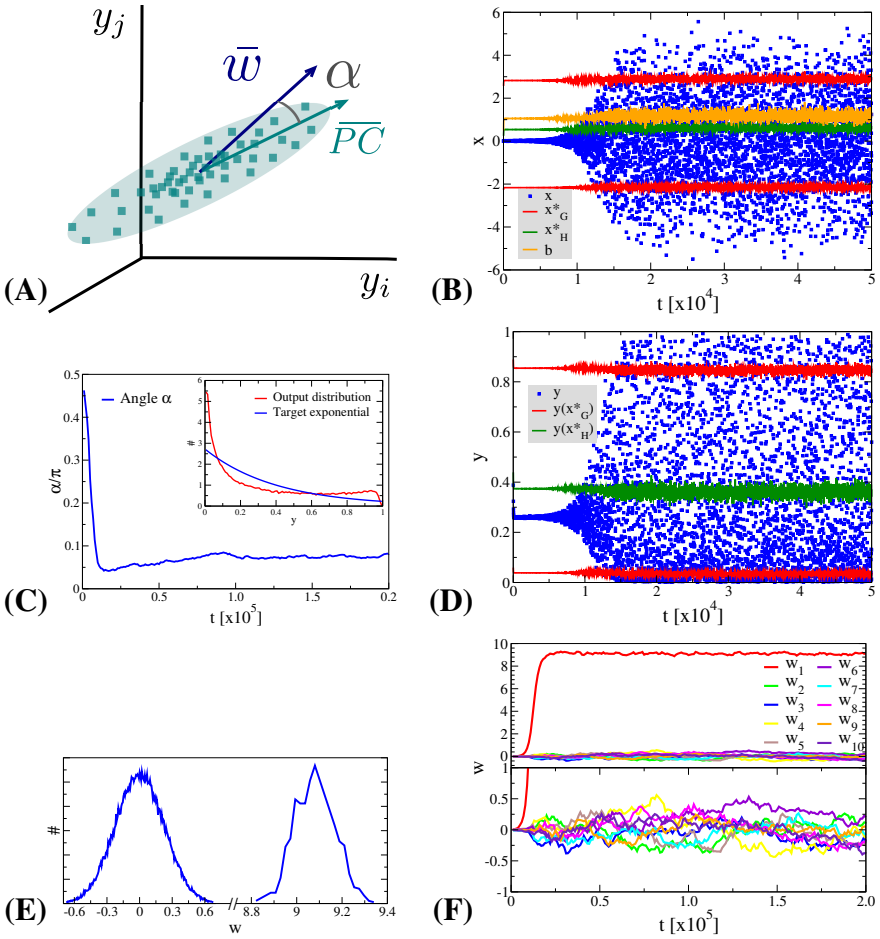


Figure 2.3: *Alignment of the synaptic weight vector to the direction of the first principal component. Numerical simulation results for a neuron with $N_w = 100$ input directions, each with Gaussian input distributions, where one direction (the first principal component) has twice the standard deviation of the other $N_w - 1$ directions.* ***(A)*** *Sketch of the input distribution density $p(y_1, y_2, \dots)$, with α the angle between the direction of the first principal component ($\overline{PC}$) and $\overline{w}$, the synaptic weight vector.* ***(B)*** *Time series of the membrane potential x (blue), the bias b (yellow), the roots x^*_G of the limiting factor $G(x)$ (red) and the root x^*_H of the Hebbian factor $H(x)$ (green).* ***(C)*** *The evolution of the angle α of the synaptic weight vector* **w** *with respect to the direction of the first principal component and (inset) the output distribution $p(y)$ (red) compared to the target exponential distribution(blue).* ***(D)*** *Time series of the output y (blue) and of the roots y^*_G of the limiting factor $G(y)$ (red) and the root y^*_H of the Hebbian factor $H(y)$ (green).* ***(E)*** *Distribution of synaptic weights $p(w)$ in the stationary state for large times.* ***(F)*** *Time evolution of the first ten synaptic weights $\{w_j\}$, separately focusing on the first principal component (upper panel) and on the nine other orthogonal directions (lower panel).*

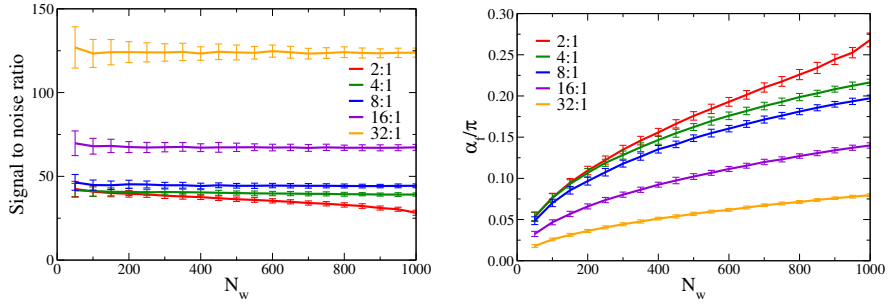


Figure 2.4: *Scaling behavior of the adaption rules with the number of afferent directions. Left: The signal to noise ratio $S_w = |w_1|/\sigma_{w_\perp}$, where w_1 is the strength of the synaptic weight along the FPC and $\sigma_{w_\perp}$ is the standard deviation of the synaptic weights in the orthogonal directions (see Eq. (2.23) in section 2.3.1.1). Right: the mean angle between weight vector and the first principal component. Shown are results for incoming signal-to-noise ratios of: 2:1, 4:1, 8:1, 16:1 and 32:1, defined as the ratio of the standard deviations between the large and the small components of the input distributions $p(y_j)$. While the output signal-to-noise ratio $S_w = |w_1|/\sigma_{w_\perp}$ remains essentially flat as a function of N_w; the increase observed for the average angle α is a statistical effect, produced by the increasingly large dimensionality of the input and synaptic weight space, as discussed in section 2.3.1.1: while all individually small in magnitude, the influence of the orthogonal weights becomes large for an increasing N_w.*

In Fig. 2.3(C), we observe how the final firing rate $y(t)$ covers the whole available interval $[0, 1]$, and the output distribution (see inset in panel (C)) tries to mimic the target exponential distribution $\propto \exp(\lambda y)$ in the Kullback-Leibler divergence of Eq. (2.17). The sliding threshold y_H^*, determined by b, as described in the previous sections, settles in this case to $y_H^* \simeq 0.4$, (green dots in Fig. 2.3(D)).

In panel (C) of Fig. 2.3 we observe how the angle α between the synaptic weight vector $\mathbf{w}$ and the direction of the first principal component of the input distribution is initially large, close to $\pi/2$ (the random value in large dimensional spaces), descending rapidly as synaptic adaption progresses, indicating that the neuron aligns its weight vector to the direction of the first principal component. Finally, in panel (E) we plot the distribution of the w_j obtained from several simulations, with a separate scale for the principal component, here $w_1 \approx 9.1$ (as averaged over 100 runs). The small components are found to be normally distributed around zero with a standard deviation of $\sigma_w^{(non)} \approx 0.23$, resulting in a a large signal-to-noise ratio $S_w = |w_1|/\sigma_w^{(non)} \approx 9.1/0.23 \approx 40$. In the following section, the interpretation of this ratio, compared to the angle α is discussed.

2.3.1.1 Signal-to-noise scaling

Biological neurons in the neocortex of mammals possess in the order of tens of thousands of synapses each [27], for this reason, for any synaptic adaption rule to be biologically meaningful, even if simplified in its details, it needs to be able to scale up for large numbers of input dimensions. In our case, this means making sure that the rule shows a stable performance even for large N_w, without the need for fine-tuning of the parameters. Fortunately, this is the case for our plasticity rules, as we will show in this section.

We will consider two possible performance measures to asses the scaling behavior of the learning rule: the angle α between the weight vector and the direction of the first principal component, and the signal-to-noise ratio of the output (defined as $S_w = |w_1|/\sigma_{w_\perp}$ where w_1 is the synaptic weight along the principal component and $\sigma_{w_\perp}$ the standard deviation of the remaining synaptic weights), each as a function of N_w.

We have considered both a large range for the number N_w of input directions and an extended range for the incoming signal-to-noise ratio (S_i), defined as the ratio between the standard deviation along the FPC (σ_1) and that one along the perpendicular directions ($\sigma_\perp$): $S_i = \sigma_1/\sigma_\perp$. As in the previous section, the input activity distributions $p(y_j)$ are Gaussians with standard deviations $\sigma_j = \sigma_\perp$ for ($j = 2, \ldots, N_w$). We consider here values for S_i of 2:1, 4:1, 8:1, 16:1 and 32:1. In Fig. 2.4 we present both performance measures as a function of N_w, for each value of S_i. All simulation parameters are kept otherwise constant.

Although the two measures would a priori quantify the same property (how strongly the weight vector signals the FPC), we observe a very different behavior of these two measures with increasing N_w. While the outgoing signal-to-noise ratio is remarkably independent of the actual number N_w of afferent neurons, the performance deteriorates in terms of the angle α, which increases steadily with N_w.

This discrepancy is, as we show in what follows, purely a statistical effect, originating from the dependence of angles on the dimensionality of the space.

In our simulations with N_w Gaussian input distributions $p(y_j)$ the synaptic weight vector adapts to

$$\mathbf{w} = (w_1, w_2, \ldots, w_{N_w}), \qquad w_1 \gg w_k \qquad (k \geq 2)\,, \tag{2.21}$$

where $p(y_1)$ has the largest standard deviation σ_1, with all other $p(y_k)$, for $k = 2, \ldots, N_w$ having a smaller standard deviation $\sigma_\perp$. The angle α between the synaptic weight vector $\mathbf{w}$ and the direction $\hat{\mathbf{e}}_\mathbf{1} = (1, 0, \ldots, 0)$ of the FPC is hence given by:

$$\cos(\alpha) \;=\; \frac{|\mathbf{w} \cdot \hat{\mathbf{e}}_\mathbf{1}|}{\|\mathbf{w}\|\|\hat{\mathbf{e}}_\mathbf{1}\|} \;=\; \frac{|w_1|}{\sqrt{w_1^2 + \sum_{k>1} w_k^2}}. \tag{2.22}$$

If we now denote with:

$$\sigma^2_{w_\perp} \;=\; \left(\sum_{k>1} w_k^2\right)/(N_w-1), \tag{2.23}$$

the standard deviation of the non-principal components (which have a vanishing mean), we can rewrite (2.22) as:

$$\cos(\alpha) \;=\; \frac{|w_1|}{\sqrt{w_1^2+(N_w-1)\sigma^2_{w_\perp}}} \approx \frac{1}{\sqrt{N_w}}\frac{|w_1|}{\sigma_{w_\perp}} \;=\; \frac{1}{\sqrt{N_w}} S_w \;, \tag{2.24}$$

This means that for a constant signal-to-noise ratio S_w (like the one we find in our simulation), α will grow with N_w. Moreover, for any arbitrarily small signal to noise ratio, the angle will always approach $\pi/2$ in the limit of large N_w, making the angle between the FPC and the weight vector a poor performance estimator. Moreover, from the point of view of biological neurons, the S_w seems like a more adequate measure than an angle which would require non-local information to be computed.

2.3.1.2 Comparison to other learning rules

At this point, we would like to compare the results of our synaptic plasticity rule to other well established models.

Oja's rule [87], introduces an additive weight decaying term to the purely Hebbian rule $y(y_j-\bar{y}_j)$:

$$\dot{w}_j \;=\; \epsilon_{oja}\left[y(y_j-\bar{y}_j) - \alpha\, y^2 w_j\right] \;. \tag{2.25}$$

The original formulation, designed for linear neurons, was proposed with $\alpha = 1$ for the relative weighting of the decay term in (2.25). In the case of non-linear neurons as the ones considered here (see Eq. (2.1)), we numerically found that Oja's rule does not converge for $\alpha \gtrsim 0.1$. In Fig. 2.5 we present the results obtained when adapting the bias using (2.18), and comparing the results for Oja's rule (2.25) and our plasticity rule (2.13), with the same training set from section 2.3.1. The parameter ϵ_{oja} was chosen in order to match the learning times (or the number of input patterns) needed for convergence for both rules (in this case ϵ_{oja} = 0.1).

Since in Oja's rule one can manually set the importance of the weight decaying term, and therefore set the size of the weights, by setting $\alpha \to 0$, one can achieve arbitrarily large S_w. When that happens, the resulting $p(y)$ becomes binary (see Fig. 2.5) because of the sigmoidal transfer function saturating for both large positive and negative x, as expected. The price of high S_w is then large weights, and with it binary output distributions, determining a trade-off between signal-to-noise and

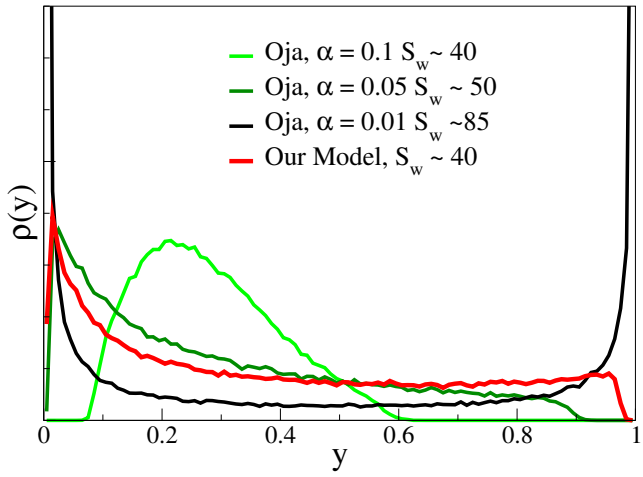


Figure 2.5: *Comparison of the final output distributions together with their signal-to-noise ratios S_w, when employing our learning rule (2.13) and Oja's rule (2.25), in this last case, for several weight decay parameters α. We have used here the same parameters and input distribution as in Fig. 2.3. Depending on the parameter α, controlling the strength of the weight decay term in (2.25), arbitrary large signal-to-noise ratios S_w can be achieved, making however the output distributions increasingly binary. The general shape of $p(y)$ is otherwise comparable, for similar signal-to-noise ratios, using both rules. Oja's rule tends to produce a more compact response however for small S_w.*

usage of the representational range of the firing rates. In other words, if one wishes to obtain smooth output firing-rate distributions $p(y)$, the level of S_w one can obtain is limited also in this case. As a note, Sanger's rule [98] reduces to Oja's rule for the case of a single neuron, as considered here.

Finally, we also attempted to make a comparison with the BCM learning rule [14,23,64], when applied to the same neural model as the one here employed. While the BCM update rule was also able to find the direction of the FPC, we always found runaway synaptic growth in the case of the type of neurons considered in our study (non-linear with bound $y \in [0, 1]$). We believe the reason for this is that the upper cut-off of the firing rate limits the value of the sliding threshold, so that it cannot raise to high enough values to induce sufficient synaptic weight decay. To sum up for the neural model here employed, and with the input distribution of section 2.3.1, runaway synaptic growth could not be avoided for the BCM rule.

2.3.2 Learning in terms of higher moments of the input distribution

So far we have shown that, for (bound) multivariate normal input distributions, our learning rule is able to find the first principal component (FPC). This is hardly surprising since most Hebbian learning rules, such as Oja's [86] and BCM [14], do precisely this. Our rule is however non-linear for large and low activity levels, and it is precisely in this way that it is able to revert the sign of the synaptic update rule to keep weights bound. This nonlinearity, inspires us to think that the rule could in fact be sensitive also to higher moments of the input distribution. To test this hypothesis, we perform a series of numerical experiments, in which input distributions with different characterestics in terms of their higher moments (but same mean and variance) are made to compete.

In Fig. 2.6 we present the results of our simulations, in which the neuron is presented with an input distribution in which the two dominant directions (without loss of generality we chose y_1 and y_2) have the same standard deviation $\sigma \approx 0.22$, with the remaining $N_w - 2$ directions having a smaller standard deviation $\sigma/4$.

As a first experiment, we chose the first direction, y_1, to be Gaussian, with the second direction, y_2 being bimodal (see Fig. 2.6(A)). The bimodal direction is composed of two superposed Gaussian distributions along y_2 having individual widths $\sigma/4$, with the distance between the two maxima adjusted so that the overall standard deviation along y_2 is also σ.

We observe that, in this situation, the learnt synaptic weight vector lies always in the plane of y_1 and y_2, as expected since the largest variance is still in this plane. Within this plane, it aligns for most randomly drawn starting ensembles $\{w_j\}$ to the direction of the bimodal distribution (y_2) (see Fig. 2.6(C)). Moreover, the system adjusts the synaptic the weights w and bias b so that the two peaks of the bimodal component are close to the two zeros $x_G^*(1/2)$ (see red symbols in Fig. 2.6(B)) corresponding to the minima of the objective function (compare Fig. 2.1). The system performs, therefore, a linear discrimination of the input, this time continuous (compare the analysis of section 2.2.4 for discreet inputs), having a bimodal output firing rate distribution (Fig. 2.6(D)).

Indeed the system seems to prefer bimodal to normal distributions. In order to provide a more general characterization, we will study the input distributions in terms of their higher moments. One way to characterize the non-normality of a given probability distribution, is its excess kurtosis [26], which we here denote with κ:

$$\kappa = \frac{Q_j}{\sigma_j^4} - 3, \qquad Q_j = \int (y_j - \bar{y}_j)^4 p(y_j) dy_j, \qquad \sigma_j^2 = \int (y_j - \bar{y}_j)^2 p(y_j) dy_j \ , \tag{2.26}$$

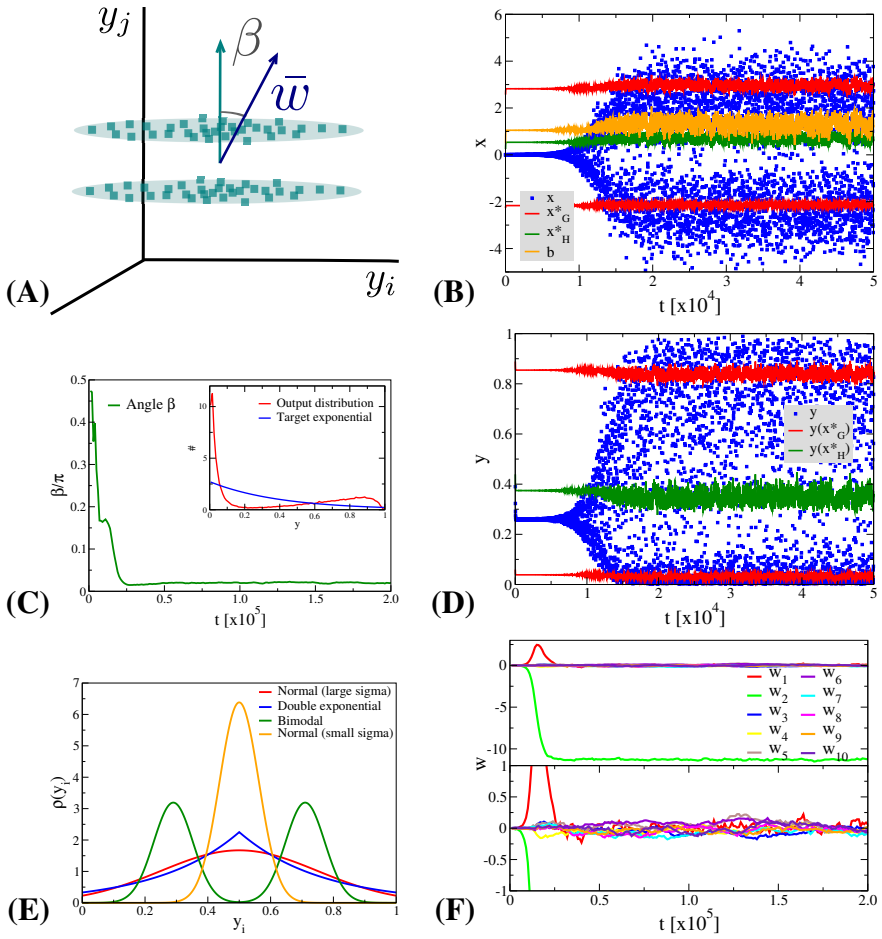


Figure 2.6: *Selectivity to higher moments of the input distribution. As in Fig.2.3, now for two input directions having the same variance, one Gaussian and one bimodal, and the other $N_w - 2$ directions having a standard deviation four times smaller.* ***(A)*** *Sketch of the input probability distribution $p(y_1, y_2, \dots)$.* ***(B)*** *Time evolution of the membrane potential x (blue), the bias b (yellow), the roots x^*_G of the limiting factor $G(x)$ (red) and the root x^*_H of the Hebbian factor $H(x)$ (green).* ***(C)*** *Evolution of the angle β between the synaptic weight vector* $\mathbf{w}$ *and the bimodal direction. (inset) The output distribution $p(y)$ (red) together with the target exponential p_λ (blue).* ***(D)*** *Time series of the output y (blue) and of the roots y^*_G of the limiting factor $G(y)$ (red) and the root y^*_H of the Hebbian factor $H(y)$ (green).* ***(E)*** *Probability distribution functions employed for the pairwise comparisons.* ***(F)*** *Evolution in time of the first ten synaptic weights $\{w_j\}$, separately for the principal component (upper panel) and for nine other orthogonal directions (lower panel).*

with the Gaussian distribution having, by construction, a value of $\kappa = 0$. Bimodal distributions, and in general distributions with most of their weight far from the mean, have a large negative excess kurtosis, reaching its minimum of -2 for the distribution composed of two δ-peaks [26]. On the other end, very peaked dis-

tributions, or distributions with heave tails, tend to have a large positive excess kurtosis. As a note, the excess kurtosis tends to be small or negative on a finite support $p_j \in [0, 1]$. Indeed, truncating distributions to fit within $[0, 1]$ (as we do here) generically produces distributions with negative excess kurtosis. This is also true for the truncated Gaussians used in our simulations.

Having now a measure of non-normality, we generalize the previous experiment by studying pairwise competition between two out of three distributions having all the same standard deviation σ, but different kurtosis, as depicted in Fig. 2.6(E), namely: a bimodal distribution with $\kappa = -1.69$, a truncated Gaussian distribution with $\kappa = -0.63$, and a unimodal double exponential with $\kappa = -0.43$.

Generating $n_{stat} = 1000$ instantiations of the simulation, with randomly drawn initial conditions, we found that the direction with lower κ (larger negative excess kurtosis) was chosen by the neuron 88.8% / 65.4% / 64.0% of the times, when the competing directions were bimodal vs. double exponential / Gaussian vs. double exponential / bimodal vs. Gaussian, respectively. We never found a case where both w_1 and w_2 were simultaneously large.

In **Chapter 3** we will present a systematic analytical study of the attractors of the learning rule in terms of the higher moments of the input distribution, which will justify these numerical findings. For the time being, we simply state that the reason for this relies on the the two minima of the objective function F_{ob} (2.10), to which for instance the maxima of the binary distribution were mapped.

Finally, we have repeated these simulations using the modified Oja's rule (2.25), with $\alpha = 0.1$ and $\epsilon_{oja} = 0.1$, finding the following relative selection rates: 97.0% / 99.8% / 42.1%, as before, when the competing directions were bimodal vs. double exponential / Gaussian vs. double exponential / bimodal vs. Gaussian. The last pair breaks the pattern and we then see no direct link between kurtosis and selection rate.

2.3.3 Continuous online learning - fading memory

Up to now we have studied the behavior of the learning rules when faced with a stationary distribution. It is relevant, however, to study the behavior of synaptic plasticity rules during continuous online learning, when the statistics of the inputs change in time, since for any real world application, a neuron should be able to adapt to changing statistics robustly. That is, adapting to new information without runaway growth of the synaptic weights.

Although permanent adaption is clearly necessary, the way and speed with which this should be achieved is still an open question and the optimal solution will probably depend on the application at hand. Concretely, whether the neuron

should adapt immediately, forgetting what was learnt, at a very short time scale, or whether it should show a certain resilience, to be safe from noise, will largely depend on the context of application. In this section we will study the behavior of the adaption rules when the input statistics changes, and we will compare the results to the behavior of Oja's rule in the same context.

In Fig. 2.7, we present the evolution of the synaptic weights (as in panel (F) of Figs. 2.3 and 2.6), when we now change the direction of the FPC, at certain points in time, or even make the input distribution completely spherical. Both learning rules recognize the new statistics autonomously, however, with very different timescales. While Oja's rule learns and unlearns the new statistics in the same timescale, our rule presents very different timescales, with a strong resilience to unlearn the previously acquired information about the input statistics. We call this feature of plasticity rules (2.13) a *fading memory*.

As in section 2.3.1, we used for the simulations $N_w = 100$ inputs or afferent neurons, with multivariate normal input distributions of standard deviation σ along the FPC (if any, see (a), (b) and (d) in Fig. 2.7) component and $\sigma/2$ for the remaining $N_w - 1$ directions. As a note, since the input distributions $p(y_j)$ are symmetric with respect to their means (here 0.5), the sign the synaptic weights are irrelevant. During the numerical experiment, the direction of the FPC changed several times, everything else remaining otherwise unchanged. As a test, we also include a period (c) in Fig. 2.7, with no principal component, that is during which all standard deviations are the same ($\sigma/2$). The initial values for the synaptic weights were drawn randomly from $[-0.005 : 0005]$. For Oja's rule (2.25), we used $\alpha = 0.1$ (which yields the same signal-to-noise ratio, compare Fig. 2.5), and $\epsilon_{oja} = 0.1$, so that the initial learning time (achieving 90% of the stationary value for the principal component) are the same for both updating rules.

We find the initial learning time $T_{initial} \approx 10^4$ steps, for $\epsilon_w = 0.01$ and $\epsilon_b = 0.1$ (same throughout this chapter). When a new direction is presented, the time the neuron takes to adapt to this new information is on the order of $T_{unlearn} \approx 10^6$, that is two orders of magnitude larger than $T_{initial}$. That is to say, once the neuron has learnt a given direction, it takes considerably longer to learn a new one. In particular, when the neuron is presented with pure noise (the no FPC case during phase (c) in Fig. 2.7), the neuron shows an even stronger resilience to forget the previously learnt direction, taking roughly $5 \cdot 10^7$ steps (5000 times longer) to go back to a fully random state of the synaptic weights. The synaptic plasticity rule (2.13) is therefore extremely robust to periods of noise, leading to what we call an *extended fading memory*.

As a note, we observe in Fig. 2.7 an initial overshoot of the larger synaptic weight whenever we change the direction of the FPC. We need to remember that the stability feature of our learning rule results from the neuron trying to keep

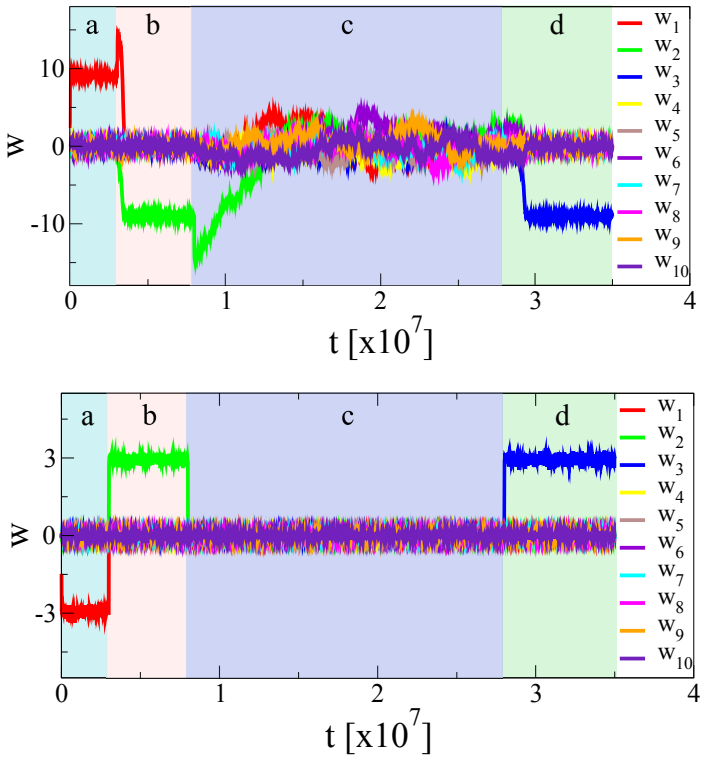


Figure 2.7: *Continuous online learning for changing input statistics and fading memory effect. Top: for our synaptic updating rule (2.13). Bottom: with Oja's rule (2.25). For both rules the same inputs are used. During time periods denoted as (a), (b) and (d), the FPC lies along y_1, y_2, and y_3, respectively, with a standard deviation σ for the FPC, and $\sigma/2$ for all other directions (same as in Fig. 2.3). At (c) there in no principal component, that is, the standard deviation along every direction is the same ($\sigma/2$). As in Figs. 2.3 and 2.6, we used $N_w = 100$ afferent neurons, with initial weights randomly drawn from $[-0.005 : 0005]$. In the figure we show the time evolution of a representative set of weights. In the case of our plasticity rule, we observe a longer timescale for forgetting than for learning a new input direction. In contrast, for Oja's rule, learning and unlearning happen at the same timescale.*

the membrane potential x (and not directly the weight) within its working regime. When the FPC is changed, the membrane potential is initially too low, since the big weight is now allocated to a direction of small variance around 0. The system then reacts by increasing the weights to try to bring the membrane potential up. This is purely a transient effect, until the new direction is learnt.

We begun the derivation of our learning rule by stating that we wanted to minimize the local sensitivity of the output distribution with respect to the synaptic weights. Once learning has been completed, the neuron is precisely at a very stable point, from where it is extremely hard to get out. This is true even once one

changes the statistics of the inputs as we here did. We will study the properties of this attractors in detail in the following chapter, when take an analytic approach and perform a dynamical system's study of the learning rules.

From the point of view of the biological relevance of the fading memory effect, it is yet not completely clear which form of unlearning is present in the brain, particularly at the level of individual neurons. On the one hand, studies carried out in pre-frontal cortex have shown full learning and unlearning of different categories in binary classification tasks (associated in this context to the concept of adaptive coding) [31]. On the other hand, there is a tendency in more complex behavioral responses to show slow or incomplete unlearning, as is the case for extinction of paired cue - response associations within the paradigm of Pavlovian conditioning [83, 95].

2.4 Discussion

In this chapter we have presented an objective function (2.10), motivated from the stationarity principle of statistical learning, from which a set of self-limiting, Hebbian plasticity rules (2.13) could be derived.

Beyond the particular objective function here presented, objective functions, or more generally generating functionals, can be used to derive equations of motion for variables or parameters in different fields. In the particular case of neuroscience, they can be used to derive adaption rules for both intrinsic and synaptic parameters. In particular, objective functions based on information theoretical principles in general, have played an important role in neuroscience [47, 64, 67, 71] as well as in applications to artificial intelligence and robotics [5, 105]. Many of these objective functions make use of Shannon's information content and related measures, such as the Kullback-Leibler distance [108], or the Mutual Information [69]. The objective function we present here can be either interpreted in terms of the Fisher Information, as we have done in section 2.2.1, or practically, in terms of its properties for self-limiting Hebbian learning.

We have already shown numerically in this chapter how the learning rule is sensitive to higher moments of the input distribution. In particular, we have shown that neurons evolving according to our rule show a strong preference for bimodal input directions. Such a preference may be particularly useful for modeling of neural systems in cortical areas. Transient bursting in the brain has been proposed as a mechanism for precise information transmission [76]. Examples of neurons showing this type of behavior are bursting pyramidal neurons in layer 5 of somatosensory and visual cortical areas [19]. This kind of neurons, switching between low (or quiet) activity states and bursting, would have bimodal rate distributions (and therefore a negative excess kurtosis). Neurons in higher cortical areas, using such a principle would be able to tune their intra-cortical receptive fields to be selective to

bursting neural populations. We have to date, no experimental evidence that these neurons are employing our proposed principle, and it would be interesting to test these predictions in future work, in collaboration with experimental groups.

The full application range of the rule will however become more clear in the following chapter, when we relate the selectivity for non normal directions to the task of independent component analysis. Neurons obeying our learning rules will turn out to be natural independent component seekers.

Finally, we have shown how our rules exhibit a distinct fading memory feature, with very different timescales for learning and unlearning, and a particular robustness vs noise.

From a wider perspective, the kind of adaptation rules here presented, where no fine tunning of the parameters is necessary to ensure the stability of the rules, falls into the framework of self-organizing processes governed by target objectives, or guiding principles [41, 50, 74, 94]. The difference with Oja's rule, in this regard, became evident when we discussed the dependence of convergence, signal-to-noise ratio, and final weight size, with the parameter α setting the strength of the weight decay term. The fact our rule is expressed in terms of the membrane potential x and not directly in terms of the weights, means that the overall size of the weights will automatically scale up or down with the size of the system or the activity level of the presynaptic neurons, to ensure the output activity is at its correct working range.

Here we have combined a novel objective function for synaptic plasticity with an existing rule of adaption for the intrinsic parameters. The interplay of different objective functions is important for several reasons [108]. From a mathematical point of view, any rule determined exclusively as the gradient of a single quantity can only show point attractors, not being able to capture the complexity of neural processes. The joint behavior of multiple generating functionals has been shown to exhibit highly nontrivial dynamics [51, 74].

Moreover, each objective function may represent different biological constraints, ranging from metabolic to computational objectives, and their interaction may only occur via the biological agent itself. In this case, it may not be possible to incorporate the multiple generating functionals into a single overarching objective function. In the case here presented, a synaptic plasticity rule motivated in terms of a stationarity principle, together with the intrinsic adaption rule, defined in terms of a computational principle (maximal information) and a metabolic constraint (fixed mean firing rate), has proven to be a viable solution to the problem of simultaneous synaptic and intrinsic adaption.

Chapter 3

Analytic study and applications of the Hebbian Self-Limiting learning rule

Echeveste, R., & Gros, C. *An objective function for self-limiting neural plasticity rules.* Proceedings of the 23th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), Bruges, Belgium, 22-24 April 2015.
Echeveste, R., Eckmann, S., & Gros, C. (2015). *The fisher information as a neural guiding principle for independent component analysis.* Entropy, 17(6), 3838-3856.

In **Chapter 2**, we presented a synaptic plasticity rule, motivated by the idea of stationarity of the output distribution, once learning of the relevant features of an also stationary input distribution is completed. Stationarity was expressed as a minimal local sensitivity condition, quantified by the Fisher information of the output distribution with respect to the synaptic weights. The resulting learning rule, was found to be Hebbian and self-limiting, avoiding in this way unbounded weight growth. Moreover, we found numerically that a neuron operating under these rules is able to find the first principal component (FPC) of multivariate Gaussian input distributions but, when presented with directions of high negative excess kurtosis, the neuron exhibits a strong preference for these non-normal directions.

While the numerical experiments provided a first intuition about the properties of the rules, questions remain about how precisely the neuron will respond in a more general case. When exactly, for instance, will the neuron switch from PCA, to higher moment sensitivity, in terms of the distribution's higher moments. In other words, what determines the working regime of the neuron, and the stability of the learnt directions. In this chapter we will study the learning rules from the point of view of Dynamical Systems, finding the attractors of the plasticity rule, that is to say, the states of the synaptic weights to which the learning rule can converge, and

their stability.

Another point that we will discuss is the generality of the rules. In **Chapter 2** we showed that the exact shape of the objective function (and therefore of the learning rule) depends on the particular choice of transfer function $y = g(x)$, relating the integrated input or membrane potential x, and the output firing rate y (see Eq. (2.10)). It is therefore important to understand how the results depend on this choice and whether they are robust to quantitative changes of g. In this chapter we will show how, for different qualitatively similar transfer functions (the family of sigmoidals, for instance), one obtains qualitatively equivalent learning rules. The robustness of the result makes an approximate functional implementation of the rule by a biological system, for instance, plausible, and it is therefore an important check.

We begin in section 3.1, with the robustness of the learning rule with respect to the particular choice of nonlinearity in the neural model. Once this robustness has been established, we will make a particular choice for g that will greatly simplify the analytic study of the attractors and their stability in later sections.

3.1 Robustness of the learning rule in terms of the chosen nonlinearity

In **Chapter 2**, we showed how, for a neuron defined by:

$$y = g(x), \qquad x = \sum_{j=1}^{N_w} w_j (y_j - \bar{y}_j) \,, \tag{3.1}$$

the stationarity principle, when formulated in terms of the Fisher information, resulted in objective function:

$$\mathcal{F}^{syn} \;=\; E\left[f^{syn}\right] \;=\; E\left[(N + A(x))^2\right], \qquad A(x) = \frac{xy''}{y'} \,, \tag{3.2}$$

where we defined for convenience an auxiliary function $A(x)$ containing the non constant part of f^{syn}. We note in (3.2), that A depends on the choice of g, since $y' = dg/dx$. For the exponential sigmoidal (or Fermi function) (2.3), employed in **Chapter 2**, we have, for instance:

$$A_{exp}(x) \;=\; x\,(1 - 2y(x)) \,. \tag{3.3}$$

With this particular choice, the Hebbian self-limiting plasticity rules (2.13) were obtained. We want to study now how these plasticity rules vary when other choices for g are considered. In particular, when g is chosen within the family of sigmoidal

transfer functions, a reasonable requirement to consider the rules as robust, is that the plasticity rules be still qualitatively similar to (2.13).

We begin by considering the alternative transfer function:

$$g_{tan}(x) \;=\; \frac{1}{\pi} arctan(x-b) + 1/2, \tag{3.4}$$

with the same limits $y \rightarrow 0/1$ when $x \rightarrow -\infty/\infty$. This choice of g, in turn determines the following version of $A(x)$ [34]:

$$A_{tan}(x) \;=\; -\frac{2x(x-b)}{1+(x-b)^2}. \tag{3.5}$$

We observe that the objective function (3.2) is by construction strictly positive ($f^{syn} \geq 0$), and the roots, determined by:

$$A_{exp/tan}(x) \;=\; -N, \tag{3.6}$$

therefore correspond to global minima, as shown in Fig. 3.1**(a)**. These minima can then be graphically found by finding the intersection of the plot of A with a horizontal at height $-N$ (see Fig. 3.1**(b)**).

Moreover, also in Fig. 3.1**(b)**, we show graphically that a mapping exists, so that, given the roots of the objective function obtained for one choice of g, the same roots can be selected for the other choice, by selecting an appropriate value of N. In other words a bijection $N_{tan}(N_{exp}, x^*)$ can be defined that preserves the roots of the objective function.

As a side note, for $A_{exp}(x)$ one finds global minima for all values of $N > 0$, whereas N needs to be within $[0, 2]$ for the case of $A_{tan}(x)$. However, in both cases, the exactly same values for the roots can be achieved. This is important since N is merely an internal parameter of the model, while the roots have an external observable correlate, namely the produced activity level.

It should be mentioned that, while the objective functions obtained by one or another choice of g present a similar behavior, they are not exactly identical. While A_{exp} (and therefore f^{syn}_{exp}) diverges for $x \longrightarrow \pm\infty$, keeping w tightly bound (regardless of the dispersion in the input distribution), A_{tan} has bound minima for $x \longrightarrow \pm\infty$, and therefore the maxima of f^{syn}_{exp} are of finite hight. This means that if the input dispersion is too strong, the weights may go on growing, making it in principle unstable to very noisy input distributions.

Finally, we tested the equivalence of the rules numerically. As in **Chapter 2**, we presented the neuron with a multivariate input distribution of N_w = 100 input dimensions (see Fig. 2.3), with one direction y_1 (without loss of generality since the rules are fully rotation invariant) having a larger standard deviation σ_1, and the

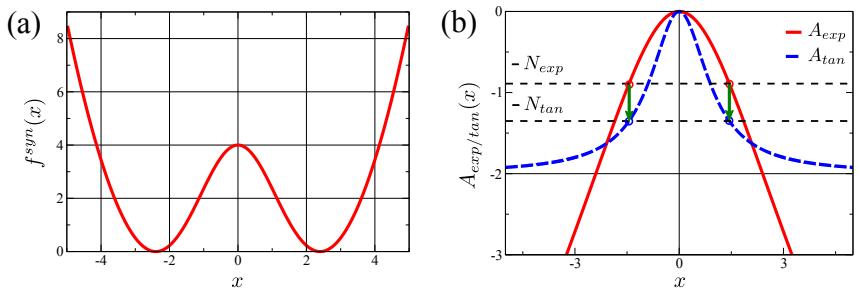


Figure 3.1: ***(a)*** *Plot of the objective function f^{syn} (3.2), as used in* ***Chapter 2****, for $A_{exp}(x) = x\,(1-2y(x))$, $b = 0$ and $N = 2$. The synaptic weights are adapted via (2.13), derived to minimize f^{syn}, so the membrane potential x is drawn towards the function's two minima.* ***(b)*** *Comparison of $A(x)$, as defined by Eqs. (3.2), (3.3), and (3.5), corresponding to the exponential and the tangential sigmoidal transfer functions g. As an illustration, we show the case $b = 0$. The roots of the objective function f^{syn} correspond to the solutions of $A(x)_{exp/tan} = -N$ (see Eq. (3.2)), which can be graphically solved by finding the intersection of the plot of A with a horizontal at height $-N$. We observe that a mapping exists, so that given the roots of the objective function for one choice of g, the same roots can be selected for the other, by selecting an appropriate value of N, as we show graphically. Namely a function $N_{tan}(N_{exp}, x^*)$ can be defined that preserves the roots of the objective function.*

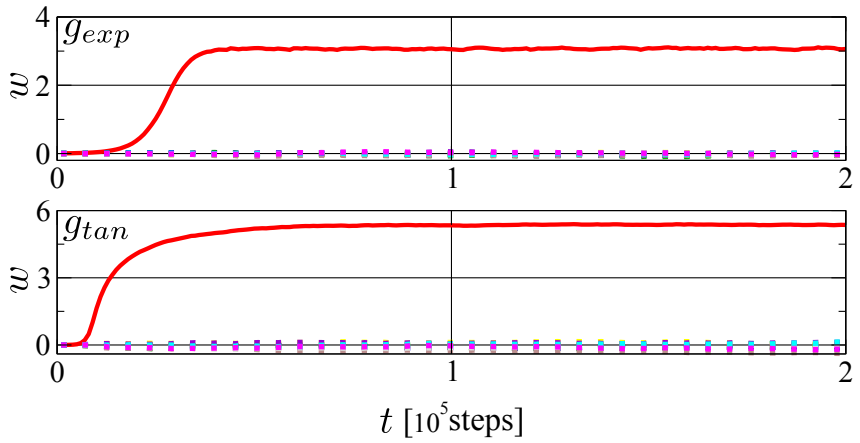


Figure 3.2: *Evolution during learning of the synaptic weights, when we employ either the exponential (2.3)(Top) or the tangential (3.4) (Bottom) transfer function. The continuous red line corresponds to w_1 (in the direction of the FPC). A representative subset of the remaining $N_w - 1 = 99$ weights plotted with dotted lines.*

other directions of $\sigma_\perp = \sigma_1/2$. We observe in Fig. 3.2, how for both the exponential or the tangential sigmoidal, the neuron is able to find the FPC.

The objective functions found so far, depend highly non-linearly on x, via $g(x)$ and non-linearities proper of the f^{syn}. This complicates further analytic treatment

of the rule. However, the robustness in terms of the general shape of the objective function encourages us to look for either an approximation or a more convenient choice of g, so that the learning rule becomes analytically tractable, without losing its properties. That is precisely what we do in section 3.2.

Note: An interesting case is that of an purely exponential transfer function:

$$g(x) = e^{x-b},$$

which corresponds to the exponential foot of the sigmoidal 2.3, and is employed when activities are low compared to the upper boundary, so that its effect can be neglected. In this case, $A(x)$ becomes equal to x (since $y'' = y'$) and the learning rule reads:

$$\dot{w}_j = \epsilon_w H(y)(y_j - \bar{y}_j), \qquad with \qquad H(y) = N + b + ln(y).$$

That is, if the upper activity bound is neglected for the neural model, it is also neglected by the learning rule, recovering a purely Hebbian rule, without reversal.

3.2 Analytic treatment of the learning rule: attractors and their stability

In **Chapter 2** we obtained a first understanding of the learning rule defined by (2.13), with the help of numerical simulations. These simulations showed that for bound multivariate normal distributions the rule was able to find the FPC (Fig.2.3). On the other hand, when two directions having the same standard deviation were made to compete (Fig.2.6), the neuron seemed to prefer directions of large negative excess kurtosis. What is not clear is why this is the case, and also what happens in between these two scenarios, that is when the input distributions have different standard deviations *and* different higher moments.

To have a better understanding of the rule, we would like to be able to analytically find the stationary solutions of (2.13), as a function of the different moments of the input probability distributions. As it was previously stated, this is very complex when the learning factors are non-polynomial in x. For this reason, we will consider a polynomial expansion in x, of Eq. (2.13), and we will do so with polynomials of the lowest possible degree, around the roots of functions H and G, that is, linear for H (which has a single root), and of degree 2 for G (which has 2 roots).

As a first step, we will consider $b = 0$. In this case, the two roots $x^*_G(1/2) = \pm x_0$ of the limiting function G are symmetric (see Fig. 2.1 **(a)**). H, the Hebbian

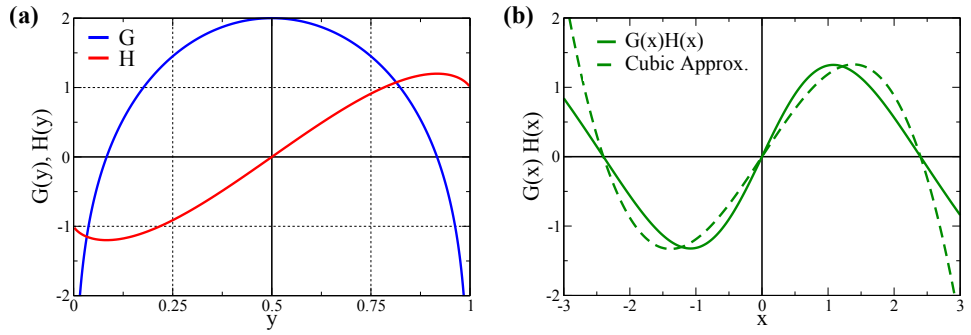


Figure 3.3: ***(a)*** *The plasticity functions G and H (2.14), which we here express solely in terms of the output activity $y \in [0,1]$ (by use of $x = g^{-1}(y)$). H is the Hebbian term, while G serves as a limiting factor, reverting the sign of the learning rule (2.13) when the output activity gets too close to its extreme values $(0/1)$.* ***(b)*** *Comparison of the original learning rule (2.13) and its cubic approximation (3.7), as a function of x for the same parameters as in* ***Chapter 2*** *($b = 0$ and $N = 2$).*

part of the learning rule, has a single root at $x = 0$ ($y = 0.5$). We therefore propose the following cubic approximation for (2.13):

$$\begin{aligned} \dot{w}_j &= \epsilon_w G(x) H(x)(y_j - \bar{y}_j) \approx -\epsilon_w x(x - x_0)(x + x_0)(y_j - \bar{y}_j)/N^2 \\ &= \epsilon_w x(x_0^2 - x^2)(y_j - \bar{y}_j)/N^2 \end{aligned} \tag{3.7}$$

where the scaling factor $1/N^2 > 0$ has been introduced to also reproduce the scaling of the height of the maximum in the plasticity function (found numerically). For a fixed N, however, this factor can simply be absorbed into ϵ_w.

For a graphic comparison of the original learning rule (2.13) to the cubic approximation (3.7) see Fig. 3.3 **(b)**.

Now, to simplify the notation, let us now denote $\gamma_j = (y_j - \bar{y}_j)$. To find the attractors (or steady state solutions of the learning rule), we are interested in computing the time-average weight change:

$$\langle \dot{w}_j \rangle = \epsilon_w \frac{1}{N^2} E\left[\gamma_j \left[\left(\sum_{i=1}^{N_w} w_i \gamma_i\right) x_0^2 - \left(\sum_{i=1}^{N_w} w_i \gamma_i\right)^3\right]\right] . \tag{3.8}$$

where, since we draw one input element of the input probability distribution per time-step of the learning rule, we have equated the time average with expectation value $E[\cdot]$, over the input distribution. As a simplification, we assume now uncorrelated and symmetric input distributions:

$$E[\gamma_i \gamma_j] = 0 = E[\gamma_i^k], \qquad k = 1, 3, 5, \ldots$$

which results in vanishing odd moments. We note here that, as previously mentioned, since the learning rule is rotational invariant, the result is independent of

the direction one chooses for the PCs, and we can choose, for instance the principal axes of the distribution to match the axes of reference, therefore eliminating the linear correlation terms.

Furthermore, by taking $\epsilon_w \to 0$ (in the limit of small adaption rates), we can assume the weights on the right hand side of (3.8) to be stationary, resulting in:

$$\langle \dot{w}_j \rangle \;=\; \epsilon_w \frac{1}{N^2} w_j \sigma_j^2 \left(x_0^2 - w_j^2 \sigma_j^2 K_j - 3\Phi \right), \tag{3.9}$$

where

$$\sigma_j^2 = E[\gamma_j^2], \qquad K_j = \frac{E[\gamma_j^4]}{\sigma_j^4} - 3, \qquad \Phi = \sum_j w_i^2 \sigma_j^2 \tag{3.10}$$

denoting the standard deviation (SD) σ_j along direction j, the excess kurtosis K_j, and a weighed average of input standard deviations that we have grouped for convenience under term Φ.

Finally, for the stationary solutions (or attractors) of (3.9), we have:

$$\langle \dot{w}_j \rangle \;=\; \epsilon_w \frac{1}{N^2} w_j \sigma_j^2 \left(x_0^2 - w_j^2 \sigma_j^2 K_j - 3\Phi \right) \;\stackrel{!}{=}\; 0, \qquad j = 1, 3, 5, \ldots, N_w \tag{3.11}$$

relating the attractors or stationary solutions w_j^* of the learning rule directly to the input moments. For each j in the set of equations (3.11), we have two possible solutions:

$$w_j^* = 0 \qquad \vee \qquad w_j^{*2} \sigma_j^2 K_j = x_0^2 - 3\Phi \,, \tag{3.12}$$

in line with numerical observations that for a fixed input distribution, the synaptic weights would always converge in different iterations of the numerical experiment to the same size. This is indeed a condition to obtain a fixed x, determined by the roots of the objective function.

In **Chapter 2** we trained the neuron with a multivariate normal distribution (truncated to $[0, 1]$ in every direction, for consistence with the neural outputs in our model), where one direction (the FPC) had a large SD σ_1, and all the perpendicular directions had a smaller SD, which was a given fraction of σ_1. We showed numerically how for this input distribution the neuron was able to find the FPC, by aligning its weight vector to the direction of the FPC. This meant that w_1 (corresponding in our simulations to the FPC direction) became large, with the remaining synaptic weights $w_j \approx 0 \,\forall\, j \neq 1$. The numerical solution we had previously obtained is therefore in line with the analytic solutions of (3.12), with the non-trivial condition in (3.12) predicting (within the employed approximations):

$$|w_1^{cub}| \;=\; \frac{x_0}{\sigma_1 \sqrt{K_1 + 3}} \,. \tag{3.13}$$

As already noted in 2.3.2, the negative excess kurtosis of any distribution has a minimum of -2 [26], which means $K_1 + 3 > 0$, and therefore the square root in (3.13) is well defined. We will perform a detailed comparison of (3.13), to the numerical results in Sect. 3.3.

3.2.1 Stability of the stationary solutions and sensitivity to the excess kurtosis

In the previous section we found the stationary solutions of the learning rule, under the cubic approximation. In this section we wish to study the stability of these fixpoints, that is to say, whether a perturbation to the synaptic weights in the vicinity of the solutions will tend to grow or decrease in time. A way to quantify this is by calculating the eigenvalues of the Jacobian Matrix at the fixpoints, a standard method in Dynamical Systems [50]. Furthermore, since in **Chapter 2** we observed a preference of the neuron for directions of large negative excess kurtosis, we want to relate the stability of the stationary solutions to the excess kurtosis.

Since the competition in the simulations of section 2.3.2 took place between the two directions with larger standard deviation (the smaller components seemed to be a source of noise only), we will study here for simplicity the case of two competing input directions with standard deviations and excess kurtosis σ_i and K_i, respectively, for $i = 1, 2$. We then have three different types of solutions for (3.12):

$$(0, 0), \qquad (w_1^* \neq 0, 0), \qquad (w_1^* \neq 0, w_2^* \neq 0)\ ,$$

with the additional solution $(0, w_2^* \neq 0)$ being simply an analog of $(w_1^* \neq 0, 0)$.

We now proceed to calculate the eigenvalues $\lambda_{1,2}$ for each type of stationary solution to then determine their stability (in Fig. 3.4 we present a sketch of these fixpoints). We have three cases:

- ◇ The trivial solution $(0, 0)$, which we found to be always unstable, given that it has positive eigenvalues:

$$\lambda_{1,2}(0, 0) = \epsilon_w \frac{x_0^2}{N^2} \left(\sigma_1^2\ ,\ \sigma_2^2\right)\ . \tag{3.14}$$

- ◇ The case $(w_1^* \neq 0, 0)$, for which we found the eigenvalues:

$$\lambda_{1,2}(w_1^* \neq 0, 0) = \epsilon_w \frac{x_0^2}{N^2} \left(-2\sigma_1^2\ ,\ \frac{\sigma_2^2 K_1}{K_1 + 3}\right)\ , \tag{3.15}$$

  where we see that λ_1 is always negative, and the sign of λ_2 depends only on K_1. The attractor $(w_1^* \neq 0, 0)$ is therefore stable (unstable) for negative (positive) excess kurtosis K_1. The same reasoning is of course valid for its analogous $(0, w_2^* \neq 0)$.

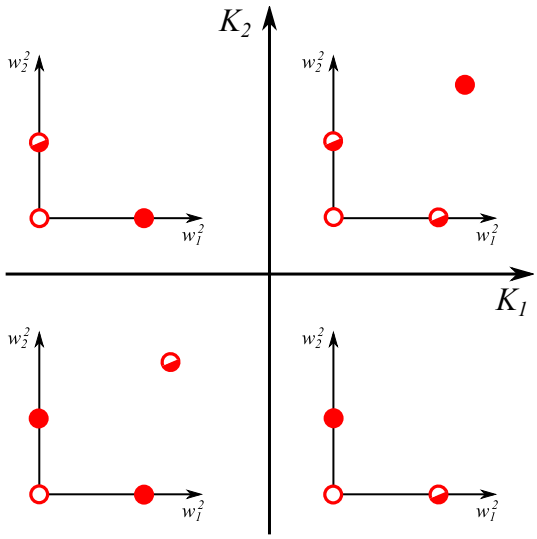


Figure 3.4: *Sketch of the stationary solutions of the learning rule after the cubic approximation (3.9), in the case of two competing directions of synaptic weights w_1 and w_2, as a function of the respective input excess kurtosis K_1 and K_2. Given that the solutions for the weights are determined only up to a sign change, we plot in the space of w_i^2. Visual representation of the solutions' stability: open circles (unstable), full circles (stable), half-full circles (saddles).*

- ◇ Finally, since the last term $(3\Phi - x_0^2)$ in (3.12) is the same for all synapses, the case $(w_1^* \neq 0, w_2^* \neq 0)$ is only possible if $K_1 K_2 \geq 0$ (K_1 and K_2 must have the same sign. Furthermore, it can be shown that this last case is only stable when both K_1 and K_2 are positive, simultaneously making the solutions (3.15) unstable. In Fig. 3.4, the stability exchange between the the attractors in the space of K_1 and K_2 is sketched.

This means that the only stable attractors in the case of distributions with negative excess kurtosis (as is the case for bound distributions covering a substantial portion of their range as discussed in **Chapter 2**), are of the type $(w_1^* \neq 0, 0)$.

Now that we have found an expression for the attractors of the learning rule, and their stability, we can try to explain the numerical results of **Chapter 2**.

Let us analyze the numerical finding that the update rules (2.13) perform a principal component analysis (PCA) for truncated multivariate normal input distributions. Firstly we note that the solution of the type $(w_1^* \neq 0, 0)$ (and its analogs) are stable for all directions (as long as they have a negative K, which is the case for truncated Gaussians). So why is the FPC selected if other directions are also stable? We can answer this question in two different ways. In the first place, as seen from (3.9), $\langle \dot{w}_j \rangle \sim \sigma_j^2$, and therefore the synaptic weights corresponding to

directions of large variance σ_j^2 will tend to grow faster and, since solutions with more than one non-zero weight are not stable, the neuron will find the FPC solution first and then stay since it is stable. Another way of seeing it, is to take into account the phase-space contraction factor $(\lambda_1^{(\alpha)} + \lambda_2^{(\alpha)})$ [50], for the alternative solutions $\mathbf{w}^{(\alpha=1)} = (w_1^* \neq 0, 0)$ and $\mathbf{w}^{(\alpha=2)} = (0, w_2^* \neq 0)$. Looking at (3.15) when $K_1 = K_2 < 0$, we observe a faster contraction rate in phase space close to $\mathbf{w}^{(1)}$ if $\sigma_1^2 > \sigma_2^2$ (and vice versa), indicating that the neuron will approach faster the attractors corresponding to directions with larger standard deviation.

3.2.2 Exact cubic learning rule

So far, we have studied the attractors and stability of the learning rule, via the cubic approximation, with the hope that the results actually hold for the full expression (2.13), given observed robustness of the learning rule as long as the roots are preserved (as shown in section 3.1).

We recall, however, expression (3.2):

$$\mathcal{F}^{syn} \;=\; E\left[f^{syn}\right] \;=\; E\left[(N + A(x))^2\right], \qquad A(x) = \frac{xy''}{y'},$$

in which the factor A is determined by the relation $y = g(x)$. Interestingly, as it was shown in [32, 37], it is possible to find a function g, so that the learning rule becomes exactly cubic, with the same roots as the original (2.13). This is the case for the rescaled error function:

$$\begin{aligned} y = g_{err}(x) \;&=\; \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{x-b}{s\sqrt{2}}\right) \\ &=\; \frac{1}{2} + \frac{1}{\sqrt{\pi}}\int_{-\infty}^{(x-b)/(s\sqrt{2})} \mathrm{e}^{-z^2} dz \\ &=\; \frac{1}{\sqrt{2\pi}s}\int_{-\infty}^{x-b} \mathrm{e}^{-\frac{z^2}{2s^2}} dz\,, \end{aligned} \tag{3.16}$$

defined as the integral of the normal distribution of variance s. Parameter s determines the slope of the transfer function and we only need to set $s = 4/\sqrt{2\pi}$ to even have the same slope as for the Fermi function (2.3). One can now easily compute the derivatives of (3.16), resulting in:

$$y' = \frac{1}{\sqrt{2\pi}s}\,\mathrm{e}^{-\frac{(x-b)^2}{2s^2}}, \qquad y'' = -\frac{(x-b)}{s^2}y'\,, \tag{3.17}$$

which in turn produce the following version of the objective function:

$$\mathcal{F}_{err}^{syn} \;=\; E\left[\left(N - \frac{x(x-b)}{s^2}\right)^2\right]. \tag{3.18}$$

Finally, we compute the corresponding synaptic plasticity rule via stochastic gradient descent (as we have done for (2.3), obtaining:

$$\begin{aligned} \dot{w}_j &= \epsilon_w (x - b/2) \left(Ns^2 - x(x-b) \right) (y_j - \bar{y}_j) \\ &= \epsilon_w (x - b/2) \left(x_0^2 - x(x-b) \right) (y_j - \bar{y}_j), \end{aligned} \tag{3.19}$$

where we have substituted Ns^2 by x_0^2 (the squared roots for the case $b = 0$). Apart from a scaling factor, Eq. (3.19) reduces exactly to previous cubic approximation (3.19) when we set $b = 0$:

$$\dot{w}_j = -\epsilon_w x (x - x_0)(x + x_0)(y_j - \bar{y}_j) = \epsilon_w x \left(x_0^2 - x^2 \right) (y_j - \bar{y}_j) . \tag{3.20}$$

For non-zero b, we can express (3.19), as:

$$\dot{w}_j = -\epsilon_w (x - b/2) \left(x - x^- \right) \left(x - x^+ \right) (y_j - \bar{y}_j) , \tag{3.21}$$

with

$$x^{\pm} = -\frac{b}{2} \pm \sqrt{b^2/4 + x_0^2} \approx -\frac{b}{2} \pm x_0 , \tag{3.22}$$

and where the approximation in (3.21) is valid for small b, simply shifting the entire learning rule (3.20) by a factor $b/2$:

$$\dot{w}_j = -\epsilon_w (x - b/2)(x - x_0 - b/2)(x + x_0 - b/2)(y_j - \bar{y}_j) . \tag{3.23}$$

To obtain an analogous expression to (2.14), in terms of the product of a Hebbian function H and a self-limiting one G, we rewrite (3.19) as:

$$\dot{w}_j = \epsilon_w H(x) G(x) (y_j - \bar{y}_j), \quad H(x) = (x - b/2) , \quad G(x) = \left(x_0^2 - x(x-b) \right) . \tag{3.24}$$

We plot these new versions of H and G in Fig. 3.5 **(a)** (compare with Figure 3.3 **(a)**).

This procedure allows us to compute the attractors of the learning rule (3.24) (as we did for (3.7)), also in the case $b \neq 0$, and for non-symmetric input distributions, resulting in:

$$\langle \dot{w}_j \rangle = \epsilon_w w_j \sigma_j^2 \left[\left(x_0^2 - \frac{b^2}{2} \right) + \frac{3b}{2} w_j \sigma_j S_j - w_j^2 \sigma_j^2 K_j - 3\Phi \right] , \tag{3.25}$$

where the additional parameter (compare (3.9)) S_j is the skewness of input distribution y_j, as defined by

$$S_j = \frac{E[\gamma_j^3]}{\sigma_j^3}. \tag{3.26}$$

The skewness, or third moment of an input distribution, is a measure of its asymmetry, and vanished therefore in the original computation, given the symmetry assumption. When it does not vanish, however, the second term of (3.25) generates an interaction of the intrinsic and synaptic plasticity through b. This is in line with the numerical finding in **Chapter 2**, that both rules interacted very weakly for the

symmetric input distributions ($S_j = 0$). In this case, small values of b ($b^2 < x_0^2/2$) merely produce a shift in the effective roots x_0.

As we mentioned before, for symmetric input distributions, the sign of the w_i is undetermined. This is however not the case for asymmetric distributions, where the skewness, plus the sign of b, will define the sign of the w_i.

As a final comment, we observe that, while the trivial solution $w_j = 0$ is predicted by this analysis to be stable for $x_0^2 - b^2/2 < 0$, we have not found this case numerically, where $b \approx 1$, for target activity levels $\langle y \rangle$ as low as 0.1, whereas x_0 is usually much larger (2.4 for $N = 2$).

3.3 Quantitative comparison of the numerical findings and analytic results from the cubic approximation

Let us recall that in section 3.2 we found an analytic expression for the stationary solution ($w_1 \neq 0, 0$), under the cubic approximation, resulting in the predicted value of:

$$|w_1^{cub}| \;=\; \frac{x_0}{\sigma_1 \sqrt{K_1 + 3}} \,.$$

We are interested now in performing two comparisons. Firstly, we will compare this analytic prediction to the numerical result produced by exact cubic learning rule obtained for the rescaled error transfer function (3.24), since this would serve as a check for the validity of the analytic procedure carried out while calculating the attractors. Secondly, we will compare these two values for w_1, with the numerical value of w_1 found for the original learning rule. The three values need to be computed for the same input distribution and, since the prediction of (3.13), determines a relation between w_1 and K_1, we choose for this purpose an input distribution, whose kurtosis along the FPC can be continuously adjusted, from unimodal to increasingly bimodal. We employ then $p(y_1)$:

$$\frac{1}{2}\left[N\left(x - \frac{1+2d}{2}, \sigma_s\right) + N\left(x - \frac{1-2d}{2}, \sigma_s\right)\right],$$

the sum of two normal distributions $N(x, \sigma_s)$, each with standard deviations σ_s, and peaks at $\pm d$ from the mean (0.5). We will vary d and, in order to keep the overall σ_1 constant, we adjust σ_s as a function of d. We therefore get for each value of d a different K_1, but always the same σ_1. For $d = 0$ we recover the normal distribution (bound to $y_1 \in [0, 1]$) with $K_1 \approx 0$, though slightly negative because of the bounds. At the other end, when $d \to \sigma_1$, $\sigma_s \to 0$, and the input distribution becomes the sum of two deltas (with $K_1 \to -2$). We can therefore test the resulting stationary

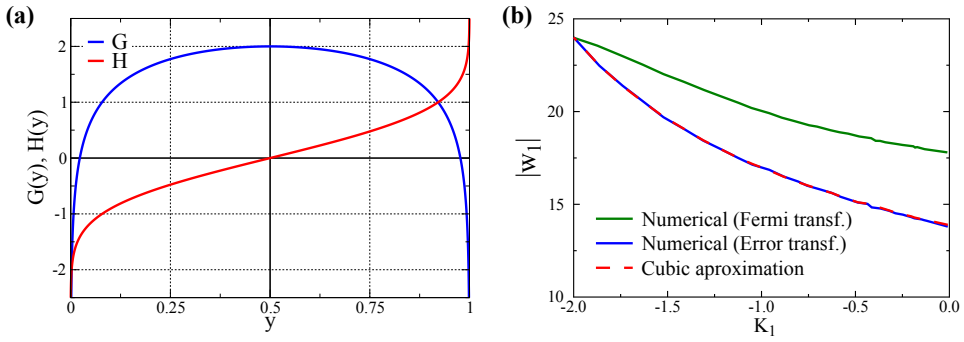


Figure 3.5: ***(a)*** *Synaptic plasticity functions G and H, as previously presented for the original rule in Fig. 3.3* ***(a)****, now for (3.24).* ***(b)*** *Value of w_1 (corresponding to the FPC) for the stationary solution, determined by either the original rule (2.13) (in green) or the exact cubic form from the error function (3.19) (in blue), together with the predicted value (3.13) of the cubic approximation, as a function of K_1. Parameters: $b = 0$, $\sigma_1 = 0.1$, $\sigma_{i\neq 1} = \sigma_1/2$, and $N_w = 100$. We find that the prediction is in excellent agreement with the numerical result for the error transfer function, while it is only qualitatively similar in the case of the original learning rule (2.13).*

value of w_1 for each $K_1 \in [-2, 0)$ and any given σ_1. Finally, for the remaining $N_w - 1$ input directions, we use bound Gaussians of $\sigma_i = \sigma_1/2$, as in **Chapter 2**.

The results are plotted in Fig. 3.5 **(b)**, where the numerical stationary value of w_1 (both using the original and the exact cubic learning rule) is presented as a function of K_1, and compared to the predicted value (3.13) of the cubic approximation. We employed in this case a constant $b = 0$, with $\sigma_1 = 0.1$, and $\sigma_i = \sigma_1/2 \,\forall\, i \neq 1$.

We observe in Fig. 3.5 **(b)** that the analytic prediction returns the same value as the numerical simulation for the exact cubic learning rule, validating thus our procedure. While the qualitative behavior of the numerical result of the original rule is preserved, we observe an increasing deviation for larger K_1. When the input distribution along y_1 is composed of two deltas, the three results match, since in this case each of the two possible input values is mapped to one of the two minima of the rules (which are the same in all three cases). As the input distribution becomes more disperse, it senses more of the rule away from the minima, and the numerical results reflect this. In other words, the quantitative deviation we observe between analytic prediction (or equivalently numerical result for the exact cubic rule) and the numerical result for the original rule, stems from the non-linearities away from the minima. As a further confirmation of this, we expect from visual inspection of Fig. 3.3 **(b)**, the cubic approximation to produce smaller values of w_1 than the original rule, since the cubic rule reverts sign faster than the original rule for large x. This is indeed the case in Fig. 3.5 **(b)**.

3.4 An application of the learning rule: Independent Component Analysis.

We can imagine the input distribution received by the neuron is actually a mixture of independent signals, which the neuron is trying to extract. The task of determining the independent signals from the mix is denoted Independent Component Analysis (ICA) [22].

The central limit theorem states that the sum of a large number of independent sources will approximate a Gaussian distribution [49]. For this reason, if one is interested in the independent components of a distribution, one can look for the input directions that are most non-normal. To assess how non-normal or non-Gaussian a probability distribution is, we can use its higher moments (for instance the excess kurtosis K or the skewness S), since they are 0 for a normal distribution. We will not go into detail here on all the different principles employed by typical ICA algorithms, but the interested reader can find an extensive and interesting review in [63].

The fact that our rule has proven to be selective to non-normal input directions, and furthermore we have been able to relate the stationary solutions of the rule precisely to the higher moments of the input distribution, encourages us to test whether a neuron functioning under these learning rules could serve as an independent component analyzer. If this were the case (which we will show it is), it is worth mentioning that this feature has not been the product of explicitly maximizing a given measure of non-Gaussianness (as carried out in [45, 72]), but rather it has been an side-effect of the stationarity principle.

In the line of our result, we note that it has been shown in the past how under certain conditions, a neural network employing a nonlinear principal component analysis learning rule, is also capable of performing ICA [87]. In this case, however, it is a single neuron that selects one of the independent components.

To test the hypothesis that our neurons are able to perform ICA, we will apply them to the non-linear bars problem proposed in [40]. In this example, the inputs to the neuron will represent pixels from a square picture, which can take only two values, corresponding to either a light or dark pixel. The picture will have a total size of $N_w = L \times L$ pixels, with L the length of the side of the square. The image is comprised of a number of horizontal and vertical bars, where a bar is a complete row or column of dark pixels (see Fig. 3.6). For each possible bar, the probability of it being actually drawn on the screen will be $p = 1/L$ (each bar is independently drawn from the other). At the intersection of a horizontal and a vertical bar the pixel takes the same dark value as in the rest of the bar (it is not the sum of the intensities), which makes the problem non-linear. As a note, classical ICA is, strictly speaking, defined for a linear mix of sources (in line with the central limit theorem), the task

employed in [40] and here corresponds therefore to a generalized or non-linear ICA.

In Fig. 3.6 we illustrate the task and show on the bottom-right examples of the learnt synaptic weights when employing the synaptic plasticity rules (2.13). We choose to show the weights graphically for clarity, where the darker the pixel, the larger the corresponding weight. For the adaption of the bias we have tested both using the intrinsic plasticity rule (2.18) derived from the Kullback-Leibler divergence, as in the previous chapter, or the more direct homeostatic rule $\dot{b} \propto (y - p)$, as employed in the original paper [40], finding no major differences. As we can see, the neuron becomes selective in different iterations to either individual bars, the independent components of the input patterns, or points. To make things even more interesting, we repeated the experiment taking out of the training set the inputs in which, by chance, only vertical or only horizontal bar were present (no occlusion), keeping sets with at least one of each type of bar. Interestingly, we found that the neuron is able to become selective to a given bar, even if that bar was never presented in isolation.

In future work, we would like to study an extended network of neurons of this type, in which every neuron receives the same input (the bars), plus recurrent connections from other neurons. In order to deal with the simultaneous update of neurons with recurrent connections, we would need to slightly modify the instantaneous model we have employed, via the time extension (1.2). It would be interesting to evaluate whether the network is able to represent a complete set of the independent components, or how this representation depends on the individual frequency of bars in the training set.

3.5 Discussion

In the present chapter we have presented an analysis of the stationary solutions to the learning rule of **Chapter 2**. We have shown the robustness of the rule to variations in the nonlinearity of the neural model, observing that the rules remain qualitatively equivalent, for the family of sigmoidal transfer functions. In particular, we have compared the original learning rule to two new versions, obtained for the tangential transfer function and for the rescaled error function. Importantly, we have presented a cubic version of the learning rule, which can be obtained either as an approximation of the original learning rule, as an expansion around the roots of functions H and G, or as an exact form, derived from the error transfer function.

The found cubic form has allowed us to obtain analytic expressions for the stationary solutions of the learning rules, and the stability of these attractors, in terms of the moments of the input distributions. In this way, we have been able to show that the solutions we had previously found numerically in **Chapter 2**, actually correspond to these attractors. Moreover, from the stability analysis, and

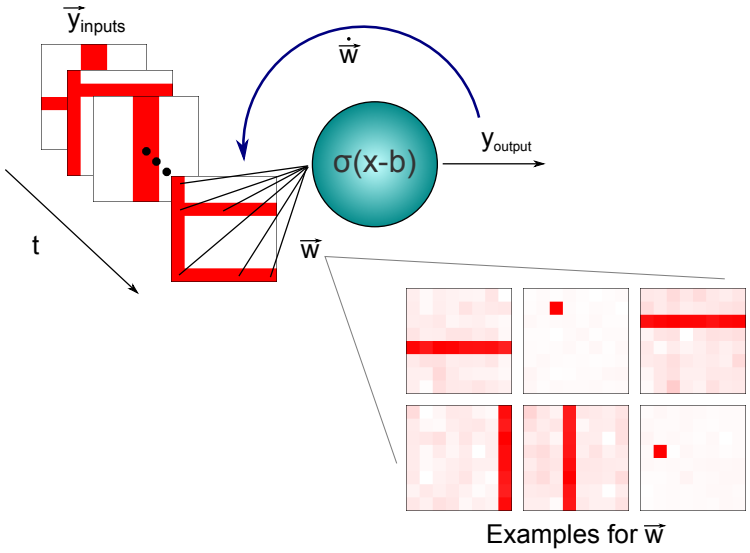


Figure 3.6: *Illustration of the non linear bars task (as performed in [40], now with our learning rule (2.13)) The input to the neuron is a set of images representing the non-linear superposition of a random set of bars. We found that the neuron is able to learns either single bars, or points, in different iterations of the simulations. On the bottom-right corner, examples of the learned weights are presented, where darker pixels correspond to larger weights. Moreover, we tested that the neuron was able to become selective to a given bar, even if this bar had never been presented in isolation in the input set. This experiment corroborates our hypothesis that our learning rule is suitable for ICA.*

particularly, from the dependence of the eigenvalues on the standard deviation of the input distribution, we have concluded that for a multivariate normal distribution, the rule should converge with higher probability to the FPC, as was indeed numerically observed. Furthermore, the analytic approach predicts the size of the learnt weights, which we have contrasted with the numerical simulations, finding a perfect agreement with the numerical simulations for the exact cubic rule (obtained for the rescaled error function), and finding a good qualitative agreement with original rule, being exact for delta distributions, and deteriorating for increased input variance, since the distribution is then able to feel more the surroundings of the minima, and not just the minima.

Finally, having established also analytically the predilection of the learning rule for non-Gaussian input directions, we have tested the use of the learning rule in the context of independent component analysis (ICA), finding that the rule is indeed able to find the independent components of the input set corresponding to the non-linear bars problem. In different iterations of the numerical experiment, a single neuron is able to become selective to a single bar, even when this bar was always partially occluded (never presented without at leas one other bar in the

opposite direction).

Chapter 4

A simplified biophysical model for STDP

Echeveste, R., & Gros, C. *Two-trace model for spike-timing-dependent synaptic plasticity*. Neural Comput. 2015, 27, 672â698.

So far, in **Chapters 2** and **3**, we have dealt with the problem of formulating a synaptic plasticity rule using a top down approach, that is to say, starting from a guiding principle which expresses some desired property of the system, we have derived our learning rules. While we have payed attention to satisfy basic biological constraints, such as the locality of the learning rules, the bounded minimal and maximal activity levels, and the homeostatic constraints for the average activity, we have not expressly related the resulting rule to the biological underpinnings, governing plasticity mechanisms in real cells. Moreover, we have so far only concerned ourselves with plasticity from a rate encoding perspective. As mentioned in Section 1.3, this is not the only available paradigm in neuroscience, with the timing of spikes playing an important role in certain protocols involving neural plasticity [12].

In this chapter, we will take the opposite approach, building a time-dependent online plasticity rule in a bottom-up fashion, using the key biological ingredients thought to be taking place in this process. Since we will be interested in future work to study the dynamical properties of systems employing this plasticity rule, we will keep the learning rule as functionally simple as possible, while still producing a reasonable fit of the experimental results.

The model we introduce here is an effective model (in the previously described sense of analytic simplicity) for timing-dependent synaptic plasticity, known as STDP. It is formulated in terms of two interacting traces (understood here as chemical or configurational signals left by neural activity that decay after a certain time), corresponding to: 1) the fraction of activated NMDA receptors, and 2) the Ca^{2+} concentration in the dendritic spine of the postsynaptic neuron.

With this proposed model, we intend to bridge the so far existing two classes of models: simplistic phenomenological rules, and highly detailed models. We believe that the properties of our model make it a very practical tool for studying the interplay between synaptic plasticity and neural activity neural networks.

4.1 Introduction

In past sections we had understood the principle of Hebbian learning (neurons that fire together wire together) from a rate encoding perspective, meaning that, if the average activity levels (over a long enough period to be able to talk about rates) of both pre- and postsynaptic neurons was high, then the synapse should be potentiated, and it should otherwise be depressed. Experiments have shown however how plasticity can depend on the precise timing of pre- and postsynaptic spikes [13, 42, 97]. In these cases, only if the postsynaptic neuron fires within a few milliseconds after the presynaptic one, implying causality, will the synaptic connection be potentiated. If the order is reversed (anti-causal order) the synapse will be depressed. Both effects show a decreased effect as the two spikes become more separated in time. This time-dependence suggests that time should somehow be encoded in the synapse or neuron. A "clock" is needed to determine the strength of synaptic modification as a function of the inter-spike time.

From the use in other fields of a decaying substance to measure time (such as the concentration of a radioactive isotope to date a fossil), we imagine that a possible mechanism to measure time in a cell could be the decaying concentration of a certain ion or molecule, which is reset with each spike. In this work we generically refer to such a substance as a *trace*. Traces will be here our time coders. While we have no direct experimental evidence that time is indeed coded in that way in the cell, we will show how candidates exist with reasonable timescales of decay, which have been previously shown to be involved in the process, and via which we will be able to formulate rules that closely reproduce experimental observations.

Our model is indeed not the first attempt to describe STDP. What makes it stand out, is rather the fact that it is an extremely simple though biophysical model, with strong explanatory power. A number of models have been proposed by other groups, formulating LTP and LTD (see 1.5), in terms of traces [7,48,66,97,100,111]. Many of them successfully explain experimental findings, such as: pairwise STDP, triplet (and higher order) nonlinearities. They differ from the here proposed model in that most of them require fitting of a large number of parameters for each experimental setup, and employ highly non-linear functions of the trace concentrations. While these models provide a possibly more realistic and detailed description of STPD, the analytic study of extended neural networks from a dynamical point of view results highly non-trivial. At the other end of the spectrum in terms of simplicity, as set of phenomenological rules have been presented [7, 42], which reproduce

some of the experimental findings but completely do away with the biological underpinnings, and are therefore hard to generalize to other settings. The triplet rule from [42], for instance, is built as a phenomenological rule to explain the interaction of three spikes in cortical plasticity, but cannot reproduce the qualitative triplet features of hippocampal STDP. At a similar level of complexity and explanation power as our model, we find [20], which is still a phenomenological model, since the authors do not establish a strong link between the employed variables and the biological underpinnings. Other phenomenological learning rules expressed in terms of abstract decaying markers or traces include [3] and [91].

The model we will present in this chapter, can be classified as a member of the the family of calcium-based spiking-neuron models (although we will actually employ two traces in our formulation, as explained in section 4.2). Previous examples of models formulating synaptic plasticity entirely in terms of the calcium levels include [48, 111]. These models are successful in reproducing a restricted set of experiments but show inconsistent results as soon as they are tested in more general contexts. The model proposed in [111], for instance, shows non-vanishing synaptic modification also in absence of presynaptic spikes. Similarly, plasticity is predicted by [48], in absence of either pre- or postsynaptic spikes. This happens because, having a single trace for both pre- and postsynaptic spikes, their contribution cannot be later distinguished. The resulting plasticity rules depend to a large extent on the fine tuning of the employed thresholds.

The model we propose in this chapter formulates STDP in terms of two interacting traces, namely the fraction of activated N-methyl-D-aspartate (NMDA) receptors and the concentration of intracellular Ca^{2+} at the postsynaptic spine. Being simple (employing polynomial expressions and a reduced number of parameters), and at the same time described in terms of the key biological elements thought to take place in the process, we believe the model builds an interesting bridge between the worlds of phenomenological rules and complex detailed biophysical models. In section 4.2 we first present the details of the model, and then show analytically the predictions of the model for pairs and triplets of spikes. We later fit the model's parameters to experimental results for STDP in hippocampal and cortical neurons, finding a good agreement between model and experiment.

4.2 The model

We are going to employ two traces in our model; two "clocks", coding the timing of pre- and postsynaptic spikes, respectively. We denote these two traces x and y, which represent the fraction of open-state NMDA receptors (or NMDARs) (x) and the Ca^{2+} concentration in the dendritic spine of the postsynaptic neuron (y). In order to understand this particular choice of variables, we will first briefly describe the synaptic transmission process and one of the candidate mechanisms thought to

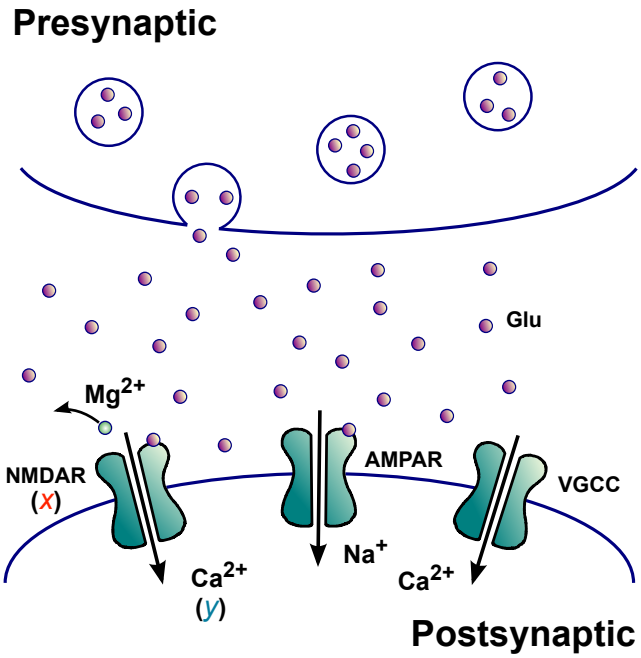


Figure 4.1: *Illustration of a glutamatergic synapse. The release of glutamate stored at vesicles in the presynaptic neuron during a spike, activates primary AMPA receptors (AMPAR) allowing the influx of a sodium current, depolarizing the postsynaptic cell. If several of these signals occur during a short time period, depolarizing the neuron enough, a chain reaction resulting in an action potential will result (see section 1.2). At the same time, NMDA receptors (NMDAR) are also activated by glutamate but only allow the influx of calcium if they are additionally unblocked by the back-propagating action potential, which needs to remove the blocking Mg^{2+} on the channel's pore. Additionally, current flows in through voltage gated calcium channels (VGCC), also triggered by the back-propagating action potential (without the need for glutamate).*

take play during STDP, for glutamatergic synapses (See Fig. 4.1). This is indeed not the only proposed mechanism for STDP in all cells in the mammalian brain, and in section 4.4.2, we will comment on a presynaptic mechanism thought to be involved in LTD in cortical neurons.

4.2.1 The biological mechanism

Within the framework of STDP, spikes have two roles in information transmission. On the one hand, the spike of one neuron influences the activity of other neurons connected to it. On the other hand, as previously mentioned, spike-timing signals whether a synapse should be strengthened or weakened. Signal transmission in a glutamatergic synapse is performed indirectly, via the release of the neurotransmitter, in this case glutamate (see 1.2), which binds to receptors on the postsynaptic

neuron, finally producing a current flow. Concretely, a presynaptic spike is signaled by the release of glutamate across the synaptic cleft, which activates different types of receptors on the postsynaptic spine. α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors (or AMPAR) [81], mainly serve the first mentioned function of influencing the activity of the postsynaptic neuron, by allowing an influx of Na^+ ions, which depolarize the cell, and (if the collective depolarization of several of these channels is strong enough) may lead to the firing of an action potential (see section 1.2). Glutamate can also induce a second cascade of events, this time mediated by NMDA receptors, related now to the second function of spikes in this paradigm: plasticity [81]. While NMDA receptors can allow for the flow of Ca^{2+} when activated by glutamate (an increase in x in our model), they are usually blocked by Mg^{2+} ions in the channel's pore [80]. They require a second event to permit a current flow, and that is the removal of the magnesium block by a back-propagating action potential (BAP) of the postsynaptic neuron. In this way, NMDA receptors act as natural coincidence detectors. The back-propagating action potential produces in this scenario two effects: it removes the mentioned Mg^{2+} block from NMDA receptors, and it activates the voltage-gated Ca^{2+} channels (VGCC). Both result in an influx of Ca^{2+} ions, producing in an increase of postsynaptic Ca^{2+} concentration y. In the case of the VGCCs, only a postsynaptic spike is required. In the following section we express these relations more concretely in terms of the model's variables.

4.2.2 Mathematical formulation: time evolution of the traces.

We begin by denoting the times of pre- and postsynaptic spikes as $\{t^\sigma_{pre}\}$ and $\{t^\sigma_{post}\}$, respectively. We will assume (and this is indeed a first simplification), that both traces in the model (the fraction x of open but blocked NMDA receptors and the concentration y of postsynaptic Ca^{2+}) decay regularly in the absence of spikes. We know that glutamate in the synaptic cleft is cleared both by passive diffusion and by glutamate transporters [60], and that Ca^{2+} concentration at the postsynaptic site is also regulated back to equilibrium in absence of postsynaptic spikes [18]. We propose [35]:

$$\begin{cases} \dot{x} = -\frac{x}{\tau_x} + E_x(x) \sum_\sigma \delta(t - t^\sigma_{pre}) \\ \dot{y} = -\frac{y}{\tau_y} + (x + y_c) E_y(y) \sum_\sigma \delta(t - t^\sigma_{post}) \end{cases} \tag{4.1}$$

where τ_x and τ_y represent the decay time constants for x and y. In (4.1), glutamate release from presynaptic spikes produces an increase in the fraction x of activated NMDA receptors, and postsynaptic spikes produce an increase in the Ca^{2+} concentration (y) through two terms: a constant term y_c (which represents the influx through VGCCs), and a term proportional to x (which represents the flow through the fraction of previously activated NMDARs). As a second simplification, we assume in our model that every NMDAR channel still open from the presynaptic spike is unblocked by the BAP, and contributes with the same amount of current, so that

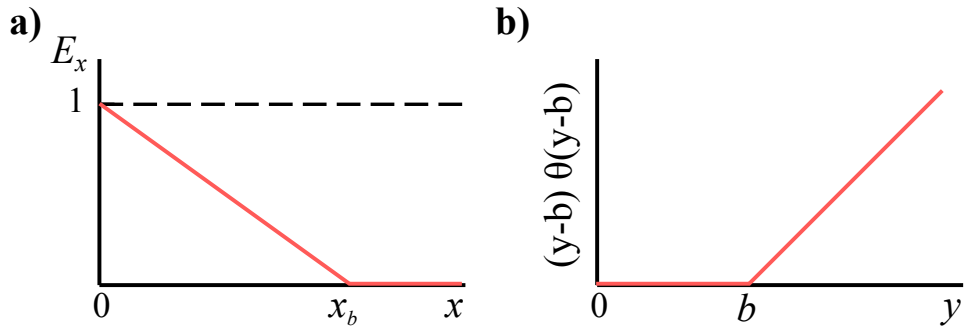


Figure 4.2: ***a)*** *Limiting factor E_x, as defined by (4.2), as a function of the trace concentration x (the case of E_y vs y is completely analogous).* ***b)*** *LTP threshold function $(y-b)\theta(y-b)$ of (4.3).*

that current is modeled as $\propto x$. As a note, in (4.1) we have absorbed in variables x and y all constants that could be rescaled, as discussed further in the **Appendix** (**A**). Two efficacy factors E_x and E_y are present in (4.1). They represent the limitation of future spikes in increasing the trace levels, given the saturation of the traces. We here chose the simple form:

$$E_z(z) = \theta(z_b - z)\left(1 - \frac{z}{z_b}\right), \qquad \theta(z) = \begin{cases} 0 & z \leq 0 \\ 1 & z > 0 \end{cases} \tag{4.2}$$

where z can represent either x or y. We see in (4.2) that the efficacy of a spike decreases linearly as a function of the existing concentration and, for trace levels above the respective reference values x_b and y_b, no additional increase is possible (see Fig. 4.2 **a)**). Therefore, once this level has been reached, trace concentrations can only decay exponentially (the solution to (4.2) for vanishing E_z), defining an effective refractory period (see the grated areas in Fig. 4.3). The length of this period is here a function of both the decay time constant of the trace and the size of the overshoot above the reference value. In absence of spikes, as the trace concentration diminishes, E grows back asymptotically to its max value of 1. We note here that similar mechanisms of reduced spike efficacy of future spikes by the effect of previous spikes have been already employed in other models of STDP [42].

We can already start to speculate on how several spikes might interact at the level of the traces. On the one hand, trace increments from several spikes accumulate via (4.1), on the other, the efficacy factors limit the effect of future spikes. Depending on the frequency of spikes and on how high or low the reference values x_b and y_b are, one or the other effect will dominate.

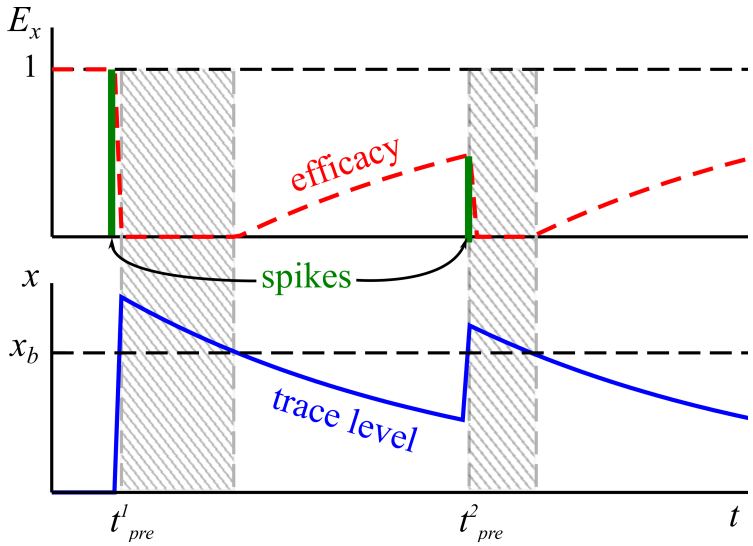


Figure 4.3: *Effect of the limiting factor on the trace dynamics. Above: Evolution of the efficacy factor E_x (dashed red line), according to Eq. (4.2) (see also Fig. 4.2), for an illustrative pair of spikes (green vertical bars) at times t^1_{pre} and t^2_{pre}. The height of the bars is proportional to E_x Below: Evolution of the trace dynamics (solid blue line), as defined by Eq. (4.1). As an example we present here the evolution of x (being analogous for y). In shaded gray areas we show the refractory periods, defined by $x > x_b$, where further input is ignored.*

4.2.3 Synaptic plasticity rule

The traces, whose evolution we presented in (4.1), code for the timing of the spikes, and therefore the next step is for us to relate the time evolution of the synaptic weight to the trace levels. Indeed, it has been found that the Ca^{2+} concentration (in our model denoted by y) is involved in both LTP and LTD [24, 84, 118], where high calcium concentrations induce LTP, and moderate and low levels result in LTD. Since two different enzymes mediate for LTP and LTD [21], we will assign one term to each of these two opposite mechanisms in the evolution equation for the synaptic weight w:

$$\dot{w} = \alpha x(y-b)\theta(y-b)\sum_{\sigma}\delta(t-t^{\sigma}_{post}) - \beta xy\sum_{\sigma}\delta(t-t^{\sigma}_{pre}), \qquad (4.3)$$

where θ is the previously defined step function, playing the role of a lower bound or threshold. Both LTP and LTD terms are here assumed to be triggered by pre- and postsynaptic spikes. Parameters $\alpha > 0$ and $\beta > 0$ represent the relative strengths of these two mechanisms. The first term in (4.3) corresponds to LTP and becomes active whenever a postsynaptic spike occurs, but produces a non-zero contribution only when $y > b$, where b represents an LTP threshold, in line with the previous statement that only high levels of calcium result in LTP (as illustrated in Fig. 4.2 **b**)). Such a threshold for LTP has indeed been observed experimentally [24], and, while

we here consider it as a constant, for simplicity, it has been shown to adapt to previous synaptic activity [59]. The second term in (4.3), on the other hand, represents the LTD contribution, and results in a reduction of the synaptic strength w. This term is active even for low calcium concentrations, and it is the balance of both LTP and LTD terms that will determine whether w grows or decreases. The fact that, in both terms we find the product of x and y, determines that in absence of either pre- or postsynaptic spikes no weight change takes place.

In our model, and as a simplification, it is the timing of pre- and postsynaptic spikes $t^{\sigma}_{pre/post}$ which marks the timing of the synaptic update. In reality this change is not instantaneous [12, 42]. Since we do not have an explicit dependence on w on the right-hand side of (4.3) (assuming the weights are far from saturation), we can ignore this lag. Finally we will consider in this chapter δ-like spikes. This means that spikes will produce small but discreet jumps in both the traces and the synaptic weights. Therefore the order for these two updates needs to be clarified: during the numerical simulations we update first the traces via (4.1) and then the weights via (4.3).

4.3 Analytic results

The mathematical simplicity of the model allows us to calculate analytically the effect of low frequency motifs of spikes on the synaptic weight, as employed by most experimental protocols. This is precisely what we sought after when formulating the model, since it will allow us to interpret the results in terms of the dynamics of the biological variables at play. In the following sections, we present these results for pairs and triplets of spikes.

4.3.1 Recovering the classic pairwise STDP rule

Several experimental protocols for STDP induction consist of low frequency ($\lesssim 1\text{Hz}$) stimulation of the pre- and postsynaptic neurons, forcing them to fire with a particular spike motif [12, 42, 115]. The decay timescale of the traces involved, however, is usually in the range of tens of milliseconds, as we will later discuss when we compare the fits of the model. This means that in between consecutive presentations of the motifs we can safely assume that the traces will have relaxed to equilibrium. We then only need to compute the evolution of the traces during the pattern length (no plasticity takes place in absence of spikes, as seen in Eq. (4.3)).

The most simple pattern or motif that can induce plasticity in this context is a pair of spikes, consisting of a presynaptic spike and a postsynaptic spike in either order. In this case, one can then measure or compute the increase in synaptic efficacy Δw as a function of the time Δt between the pre- and the postsynaptic spike (a positive value of Δt corresponds to a causal pre-post order and a negative

value to an anti-causal order). The reader might be familiar with the typical STDP plot from Fig. 4.4, whose shape is usually considered as the stereotypical STDP function, despite the fact that, as has been shown experimentally [42, 115], the pairwise plasticity plot is not enough to describe the changes in synaptic connectivity produced by more general spike patterns. However, clearly any model aiming at describing STDP should reduce to this plot for isolated spike pairs, and so we begin our analytic calculations by showing how the classic pairwise STDP rule is recovered by our model.

As previously mentioned, in the low-frequency regime, we can assume that each pair is isolated, and we can simply integrate Eqs. (4.1) and (4.3). We obtain:

$$\Delta w = \begin{cases} \alpha \mathrm{e}^{-|\Delta t|/\tau_x}\left(\mathrm{e}^{-|\Delta t|/\tau_x} + y_c - b\right) & \Delta t > 0 \\ -\beta y_c \mathrm{e}^{-|\Delta t|/\tau_y} & \Delta t < 0 \end{cases} \tag{4.4}$$

We observe then that if $y_c \geq b$, LTP occurs for $\Delta t > 0$ and LTD for $\Delta t < 0$, and, in particular, for $b = y_c$ the typical exponential shape for both LTP and LTD is recovered. As a note, the choice of formulating LTP and LTD as two separate terms in Eq. (4.3) requires in particular the LTP term to be always positive (otherwise it would not be potentiating). Interestingly, if one were to relax the threshold condition by removing the step function, and setting $b > y_c$, we would obtain a depression window on the $\Delta t > 0$ side, as observed experimentally in CA1 cells from rat hippocampal slices [85]. In the case $b < y_c$ there is a small deviation from the classical exponential shape, with a component of the LTP term decaying with a timescale τ_x and another one decaying with timescale $2\tau_x$.

While we have included in (4.3) an LTP threshold (b), we did not do so for the LTD term in the model. One could have had in principle employed an expression $(y - b_{LTD})\theta(y - b_{LTD})$ in the LTD term, analogous to the LTP one, which reduces to the here presented rule for $b_{LTD} = 0$ (note that $y \geq 0$). The reason for our choice is explained in what follows, by exploring the effect on the LTD term of setting $b_{LTD} \neq 0$. The LTD term is triggered at the time of the presynaptic spike, which means that, for an isolated pair or spikes, x will be 1 at this point. This means that the LTD threshold value would only represent a vertical shift by a constant factor βb_{LTD}, in the region of Δt where $y > \beta b_{LTD}$ (because of the step function). The sign of the shift will be given by the sign of b_{LTD}. For $b_{LTD} < 0$, y is always larger than b_{LTD}, which results in a constant downwards shift for all $\Delta t < 0$. This means depression would occur even for infinitely separated spikes. Experimental results [12, 42] suggest however that $\Delta w \to 0$ for $\Delta t \to \pm\infty$. On the other hand, for $b_{LTD} > 0$, the whole LTD plot is shifted upwards but, since the sign of the LTD term needs to be negative, the exponential tail where $y < \beta b_{LTD}$ is lost ($\dot{w}$ is set to 0 by the step function). From inspection of the experimental data (see Figs. 4.4 and 4.6), it is hard to establish whether the exponential tail should be kept or not (the data is indeed quite noisy). We decide here, given that no further resolution in the data is currently available, to respect the original exponential fits proposed in the

original experimental papers [11, 42], by setting $b = y_c$. Furthermore, by simply re-expressing the constants in the model as: $\alpha = A^+$, $\beta = A^-/y_c$, $\tau_x = 2\tau_+$, and $\tau_y = \tau_-$, we obtain:

$$\Delta w = \begin{cases} +A^+ \mathrm{e}^{-|\Delta t|/\tau_+} & \Delta t > 0 \\ -A^- \mathrm{e}^{-|\Delta t|/\tau_-} & \Delta t < 0 \end{cases} \tag{4.5}$$

namely the classical fit for pairwise STDP [11, 42], where A^+, A^-, τ_+ and τ_- represent the maximal intensities, and timescales of LTP and LTD for isolated spike pairs, respectively. We note here that the result in (4.5) does not depend on the values of y_c, x_b, and y_b. We will determine these parameters by comparison to experimental data coming from more complex spike motifs. From now on we will keep the new parameters related to the pairwise STDP function in our description, rewriting the plasticity rule (4.3) as:

$$\dot{w} = A^+ x(y - y_c)\theta(y - y_c) \sum_\sigma \delta(t - t^\sigma_{post}) - \frac{A^-}{y_c} xy \sum_\sigma \delta(t - t^\sigma_{pre}). \tag{4.6}$$

This final form of the synaptic plasticity rule is the one we will employ until the end of this chapter since it will allow us to interpret the results of more complex spike motifs as higher order contributions to the pairwise STDP rule.

It is important to note that the presented learning rule is a local *online* synaptic plasticity rule, as were the rules presented in **Chapters 2** and **3**, in the sense that the weights are updated on the fly without the need for an external memory of the timing of spikes. This is the advantage of working with the trace variables, which present a biologically implementable type of memory.

4.3.2 Triplets of spikes

As previously mentioned, being able to reproduce the effect of a pair of spikes is not the end of the story when it comes to STDP. It has been shown experimentally that additional spikes interfere non-linearly with the pairwise result, in a qualitatively different manner depending on the neural type [42, 115]. We will discuss this issue in detail in sections 4.4.1 and 4.4.2, when we compare the model's results with experimental data from hippocampal and cortical neurons.

In this section we will describe our model's prediction for the smallest higher order contribution to the pair, namely adding either an extra pre- or postsynaptic spike. This kind of motif is denoted as a triplet, and we will consider here triplets of the kind PrePostPre or PostPrePost. For compactness, we will identify a PostPrePost triplet as Δt_1PreΔt_2 and a PrePostPre one as Δt_1PostΔt_2, with Δt_1 and Δt_2 given in milliseconds.

Examples:

- 10Post5 *stands for a PrePostPre triplet defined by:*

$$\left\{t^{\sigma}_{pre}\right\} = \{-10, 5\}, \qquad \left\{t^{\sigma}_{post}\right\} = \{0\} \tag{4.7}$$

- 15Pre10 *stands for a PostPrePost triplet defined by:*

$$\left\{t^{\sigma}_{pre}\right\} = \{0\}, \qquad \left\{t^{\sigma}_{post}\right\} = \{-15, 10\} \tag{4.8}$$

Once again, for low frequency stimulation, we can consider the triplet as isolated and integrate (4.6), obtaining

$$\begin{aligned} \Delta w = \; & + \; A^{+} exp\left(-\tfrac{|\Delta t_1|}{\tau_+}\right) \\ & - \; A^{-} exp\left(-\tfrac{|\Delta t_2|}{\tau_-}\right) \left[1 + \tfrac{exp\left(-\frac{|\Delta t_1|}{\tau_x}\right)}{y_c}\right] \left[1 + exp\left(-\tfrac{|\Delta t_1|+|\Delta t_2|}{\tau_x}\right)\left(1 - \tfrac{1}{x_b}\right)\right] \end{aligned} \tag{4.9}$$

for PrePostPre triplets, and

$$\begin{aligned} \Delta w = \; & - \; A^{-} exp\left(-\tfrac{|\Delta t_1|}{\tau_-}\right) \\ & + \; A^{+} exp\left(-\tfrac{|\Delta t_2|}{\tau_+}\right) \left[1 + y_c exp\left(-\tfrac{|\Delta t_1|+|\Delta t_2|}{\tau_y} + \tfrac{|\Delta t_2|}{\tau_x}\right)\left(1 - \tfrac{exp(-|\Delta t_2|/\tau_x)+y_c}{y_b}\right)\right] \end{aligned} \tag{4.10}$$

for PostPrePost triplets. For the calculation of this first expression we have assumed that the traces are below their respective reference levels (x_b and y_b) at the time of the second spike (the case of the traces being above the reference levels when the second spike arrives is presented later in (4.11) and (4.12)).

Learning rule (4.6) is an online learning rule, and it therefore respects causality, future spikes do not influence the past. This is observed in the first terms of Eqs. (4.9) and (4.10), which give the contribution of the first pair, exactly as in (4.5). The second term of Eq. (4.9) is clearly a correction to the second pair predicted by Eq. (4.9), because of the presence of the additional spike. This correction results from the competition between two effects: the trace accumulation produced by successive spikes, and the suppression effect produced by the reduced efficacy of the second spike.

Now, if instead the third spike arrives during the effective refractory period when the traces are fully saturated (compare Fig. 4.3), Eqs. (4.9) and (4.10) become:

$$\begin{aligned} \Delta w = \; & + \; A^{+} exp\left(-\tfrac{|\Delta t_1|}{\tau_+}\right) \\ & - \; A^{-} exp\left(-\tfrac{|\Delta t_2|}{\tau_-}\right) \left(1 + \tfrac{exp\left(-\frac{|\Delta t_1|}{\tau_x}\right)}{y_c}\right) exp\left(-\tfrac{|\Delta t_1|+|\Delta t_2|}{\tau_x}\right) \end{aligned} \tag{4.11}$$

for PrePostPre triplets, and:

$$\Delta w = -A^{-} exp\left(-\frac{|\Delta t_1|}{\tau_-}\right) \tag{4.12}$$

for PostPrePost triplets. The second term in Eq. (4.12) has now completely vanished because of the LTP threshold. This last case is relevant only for high frequency stimulation, and does not occur for the low frequency pattern stimulations that we analyze here.

As a note, we have employed in our model sharp threshold functions. One could however extend the model by employing soft bounds instead. The disadvantage is an increase in the number of parameters, and potentially the loss of analytic tractability.

4.3.3 Biological implementation of the variables and parameters in the model

At this point it is important to remark that we are fully aware that the number of substances and biological elements involved in the process of STDP is far greater and the functional relations between them more complex than the simple polynomial expressions we have chosen. The goal of this work is to present an analytically tractable model which captures the essence of the biological process, and is expressed in terms of biologically plausible mechanisms. In this sense, our model can be categorized as an *effective model*. In this type of models, necessarily the collective effect of a large number of agents is condensed into a smaller number of variables.

Variable x, for instance, was defined in section 4.2.2, as the fraction of active (but not necessarily unblocked NMDAR channels). These channels become unblocked only when a postsynaptic spike arrives. When this happens, the influx of calcium through these channels is also modeled as x (since the proportionality factors have been absorbed in the model, as explained in **Appendix** (**A**)). In this way, x can also be associated with a transient calcium current. After fitting of the results to the pairwise STDP function, the decay timescale for LTP is related to τ_x by $\tau_x = 2\tau_+$ (as we showed in section 4.3.1). From these two facts we can associate τ_x to the decay time-constant of the transient calcium current. In [54], the authors argue that the narrow LTP window is a consequence of AMPA-EPSPs: fast excitatory postsynaptic potentials produced by AMPA channels located in the postsynaptic spine. As explained in the same work, the whole spine functions as an electrical amplifier, locally extending the duration of the depolarization at the spine. For this reason, one finds different time constants in different neurons (and even for different synapses belonging to the same neuron). We have not included an explicit computation of AMPA currents in our model, nor did we consider the morphology

of the spine, and simply condense all these effects into an effective τ_x. In the same way τ_y represents for us an effective decay time for the Ca^{2+} concentration (y) at postsynaptic site.

Finally, while it is clear that no more NMDAR channels can be unblocked than those present in the spine, and that calcium concentration needs to be bounded for the survival of the cell, the efficacy terms in (4.1), limiting further increase of traces x and y as a consequence of past spikes, effectively encompass saturation effects all along the cascade of events finally leading to LTP or LTD. A similar effective mechanism had already been proposed by [42], to explain triplet results in visual cortical neurons, which indeed evidence strong suppression effects.

4.4 Comparison to experimental results

In what follows we will compare the experimental results for pairs and triplets of spikes in hippocampal and cortical neurons, with the results produced by our model. To do so, we will adjust the seven parameters in Eqs. (4.1) and (4.6). As we saw in section 4.3.1, A^+, A^-, τ_+, and τ_-, come directly from the pairwise rule (4.5) and are experimentally determined. The remaining three parameters, namely y_c, x_b, and y_b, will be determined by the results of the triplet protocols in the same neurons.

4.4.1 Hippocampal neurons

Experimental data: from [11, 12] (pairs), and [115] (triplets).

Neural type: Cultured rat hippocampal neurons.

Training set: 60 repetitions of the pair/triplet motif at 1Hz.

In each case the protocol consisted of regular repetitive stimulation of the motif (pair or triplet), by dual clamp: each neuron has an electrode inserted that forces it to fire at a particular time.

In [11], it was found by fit of the experimental data produced by the same group in [12], for pairwise stimulation:

$$A^+ = 0.86/60, \qquad A^- = 0.25/60, \qquad \tau_+ = 19\,\text{ms}, \qquad \tau_- = 34\,\text{ms}. \tag{4.13}$$

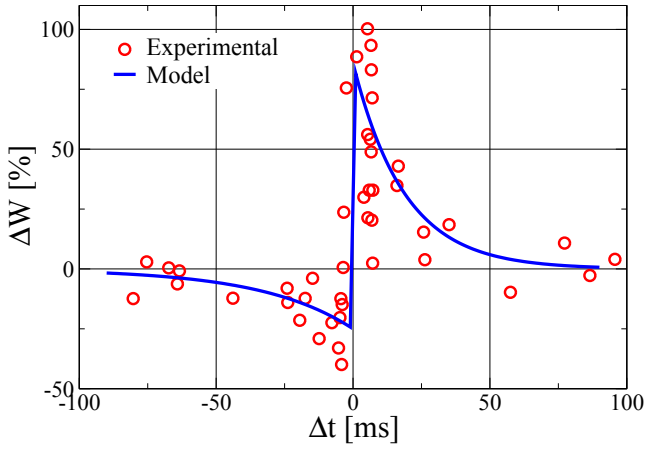


Figure 4.4: *Synaptic weight modification induced in hippocampal neurons by a train of 60 pairs at a frequency of* 1 Hz*, as a function of the time* Δt *between the two spikes in the pair. Red open circles correspond to the experimental data for cultured rat hippocampal neurons [11, 12]. The continuous blue line corresponds to model's results. Parameters:* $A^+ = 0.86/60$*,* $A^- = 0.25/60$*,* $\tau_+ = 19$ms*,* $\tau_- = 34$ms*, (fit of the experimental data, as presented in [11]).*

In Fig. 4.4 we compare these fits from experimental results with the ones produced by our model using the same protocol and parameters. We have tested the validity of our low-frequency analytic approach by producing a numerical simulation of Eqs. (4.1) and (4.6) for this protocol, finding no visible difference between the two.

As we showed in section 4.3.1, pairwise STDP does not depend in our model on y_c, x_b, and y_b, since traces cannot accumulate or saturate with only one pre- or postsynaptic spike present. To select these parameters we resort to higher order contributions, namely triplets. In Fig. 4.5 we compare experimental results [115] for the same type of neurons, now produced by triplet stimulation, with the results produced by our model (as described in section 4.3.2). Once again, the experimental protocol consists 60 triplet repetitions at a frequency of 1 Hz. Also in this case, we produced numerical simulations with the same protocol, without visible differences with the analytic low-frequency results. The pairwise STDP parameters (4.13) have been kept and only the remaining three free parameters y_c, x_b, and y_b have been fit by minimizing the standard deviation (SD) between the theoretical and the experimental results. We obtained in this way: $y_c = 0.28$, $y_b = 0.66$, and $x_b = 0.62$ (SD of the best fit 6.76). The comparison of the results indicates a good approximation of the experimental results by the model.

Regarding the fit of the parameters we encountered a smoothly varying SD with the parameters, indicating a robust result (no fine tuning is required). Moreover, we found that the three parameters can be adjusted with a certain flexibility, while still

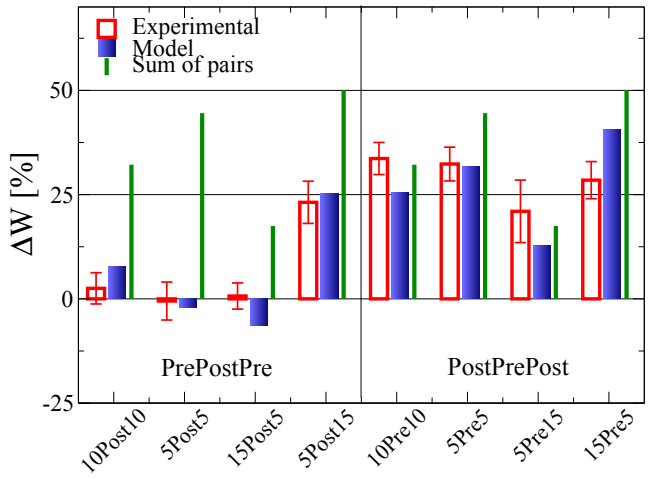


Figure 4.5: *Synaptic weight modification induced in hippocampal neurons by PrePostPre (see (4.7)) or PostPrePost (see (4.8)) triplets. The protocol consists of 60 repetitions of the triplet pattern at* $1\,\mathrm{Hz}$*. Blue bars represent the model's results. The experimental data from [115] is presented with empty red boxes. Green lines indicate the linear addition of the individual effect that each of the two pairs (PostPre and PrePost) included in a triplet would have. We observe how a triplet cannot be computed simply as the sum of the pairs. Model's parameters:* $A^+ = 0.86/60$*,* $A^- = 0.25/60$*,* $\tau_+ = 19\mathrm{ms}$*,* $\tau_- = 34\mathrm{ms}$*,* $y_c = 0.28$*,* $x_b = 0.62$*, and* $y_b = 0.66$*.*

obtaining a reasonable fit, which could be used to also reproduce other experimental effects. In Sect. 4.5 we will indeed use this fact to show how different results can be obtained for high frequency stimulation, while still producing a good pairwise and triplet fit.

We started this chapter by saying that knowing the effect of a pair of spikes on the synaptic weight is not enough, and that higher order contributions need to be considered. In Fig. 4.5 we included also in green lines the hypothetical weight change that the two pairs (PostPre and PrePost) that compose each triplet would produce if added up linearly. The result is clear; a triplet is not a sum of pairs, and this is particularly true in the case of PrePostPre triplets. Moreover, a spike suppression mechanism, as proposed in [42] from cortical neurons, also fails to explain the nonlinear deviations we find for hippocampal neurons, since suppression of the second presynaptic spike in PrePostPre triplets would only reduce depression, resulting in even more potentiation than the linear addition. The fact that our model is able to reproduce this effect is due to the ability of our traces to accumulate. Indeed, in PrePostPre triplets the LTD effect of the second pair is increased because the presynaptic trace is larger due to the combined effect of the two presynaptic spikes, explaining the observed deviation of our model's results from the linear addition, in line with the experiment.

4.4.2 Cortical neurons

Experimental data: Courtesy of Robert Froemke and Yang Dan, also partially in [42] (both pairs and triplets).

Neural type: Pyramidal neurons in layer 2/3 (L2/3) of rat visual cortical slices.

Training set: 60 repetitions of the pair/triplet motif at $0.2\,\mathrm{Hz}$.

In each case the protocol consisted of regular repetitive stimulation of the motif (pair or triplet). Presynaptic stimulation was done as a single extracellular pulse, whereas a brief depolarizing current was injected to the postsynaptic neuron that induced an action potential.

In this section we will repeat the analysis we carried out in the previous section, now for cortical neurons. It may be argued that both cases should not be treated with the same model since there is evidence of at least time dependent LTD involving presynaptic NMDARs and retro-cannabinoid signaling [103], which we have not included in our model. In this case, plasticity would take place on the presynaptic side, with the chemical signal traveling backwards and activating cannabinoid receptors on the presynaptic side. We will carry out the comparison nonetheless, in the spirit of an effective model, to see to what extent the model is able to capture the particular features of the results. It is possible that the functional dependence of the variables remains valid, although the biological implementation is not the same. In any case, for cortical neurons we should interpret these results with care.

We begin, as we did for hippocampal neurons, by setting the values of A^+, A^-, τ_+, and τ_- directly to the values resulting from the pairwise experiment, in this case in L2/3 cortical neurons, from [42]. It yielded:

$$A^+ = 1.03/60, \qquad A^- = 0.51/60, \qquad \tau_+ = 13.3\,\mathrm{ms}, \qquad \tau_- = 34.5\,\mathrm{ms}. \tag{4.14}$$

In Fig. 4.6 we present a comparison between model and experiment for pairwise stimulation in these cortical neurons.

As in 4.4.1, we determine y_c, x_b, and y_b from triplet experimental results [42]. The data is in this case differently structured from that in hippocampal neurons, where several measures were performed for a reduced number of triplet configurations. In the case of the here studied cortical neurons, a large number of triplet configurations covering the space of Δt_1-Δt_2 (see section 4.3.2), were presented, but each measured only once. We proceed nonetheless in the same way, fitting

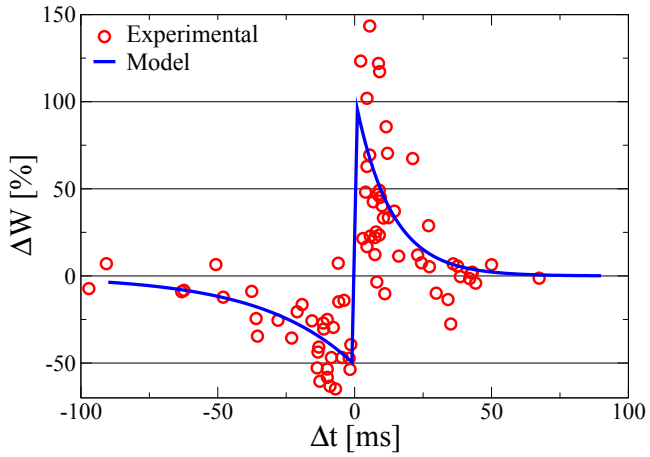


Figure 4.6: *As in Fig. 4.4, now for visual cortical neurons. The protocol consists of 60 repetitions at* $0.2\,\mathrm{Hz}$*, both for the model and the experiment [42]. The experimental data is presented with red open circles (courtesy of Robert Froemke and Yang Dan) the model's results are plotted with a continuous blue line. Parameters:* $A^+ = 1.03/60$*,* $A^- = 0.51/60$*,* $\tau_+ = 13.3\,\mathrm{ms}$*,* $\tau_- = 34.5\,ms$ *(fit of the experimental data presented in [42]).*

the complete set, also by minimizing the mean square error. In order to visually compare the results, and be able to present them in a similar way to the hippocampal triplets, we perform a smooth interpolation of the data, using Gaussian filters of width $5\mathrm{ms}$. This interpolation was *not* used for the fit, and only served a visualization purpose. We present this data, together with the model's results for the best fit in Fig. 4.5. The parameters obtained by the fit are $y_c = 11.6$, $y_b = 10.9$, and $x_b = 0.5$, with an SD of 37.4, much larger than the one found for hippocampus. It is important to note that the model captures the qualitative features of the triplets, as seen in Fig. 4.5. While not as good as fit as for hippocampal neurons, the large SD cannot be explained by these discrepancies alone. In order to understand the reason for this large SD, we look into the dataset, noticing a huge variability between neighboring points. Indeed, when we calculate the SD between data and the Gaussian filtered smoothing we used for visual representation we already get an SD of 32.5 (which explains already most of the SD between model and experiment). Therefore, any model which produces a smooth plasticity function in the Δt_1-Δt_2 space will have a high SD, simply because of the scattering of the data. In order to further clarify this issue, more experimental data, perhaps with a similar protocol used in hippocampal neurons, would be needed.

In the previous section we showed (see Fig. 4.5) how potentiation was heavily attenuated (when compared with the linear sum of pairs) for PrePostPre triplets in hippocampal neurons, whereas PostPrePost triplets were close to the linear model. This was explained by presynaptic trace accumulation in the PrePostPre triplets, which lead to increased depression. In Fig. 4.7, we observe a drastically different

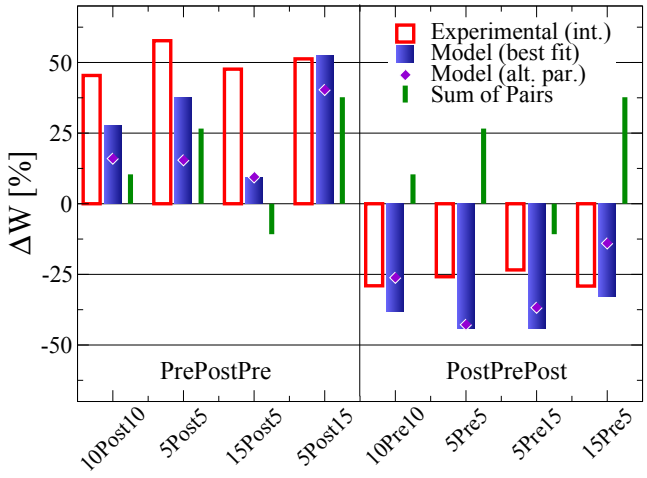


Figure 4.7: *As in Fig. 4.5, now for visual cortical neurons. The protocol consists of 60 repetitions at* $0.2\,\mathrm{Hz}$*, both for the model and the experiment [42]. Blue bars represent the model's results for the best fit. The experimental data is presented with empty red boxes. Green lines indicate the linear addition of the individual effect that each of the two pairs (PostPre and PrePost) included in a triplet would have. Diamonds represent the model's results for an alternative set of parameters. Despite the larger deviation of the fit, the model is still able to reproduce the distinct cortical triplet nonlinearities qualitatively (compare Fig. 4.5). Parameters:* $A^+ = 1.03/60$*,* $A^- = 0.51/60$*,* $\tau_+ = 13.3\,\mathrm{ms}$*,* $\tau_- = 34.5\,ms$*, Best fit:* $y_c = 11.6$*,* $y_b = 10.9$*, and* $x_b = 0.5$*. Diamond points:* $y_c = 1.0$*,* $y_b = 0.9$*, and* $x_b = 0.4$*. The experimental data is courtesy of Robert Froemke and Yang Dan.*

effect, with increased potentiation in PrePostPre triplets and very strong depression for PostPrePost triplets. While the SD of our model's fit is larger in this case, the model is still able to capture this very different behavior. We then look into the fitted parameters to shed light on what may be the reason for such qualitative differences between hippocampal and cortical triplet nonlinearities. Looking at the best fit parameters we find $x_b < 1$ and $y_b < y_c$. 1 is the rescaled contribution of one spike to x, and the contribution of one spike to y is equal or larger than y_c. This fits suggest that one spike is enough to saturate the traces, meaning a large NMDA activation and calcium flow for a single spike. Moreover, we tested with other parameter configurations -which would give the same qualitative tendency (though slightly poorer fits)- and we consistently found that, for this kind of triplets to be observed, the parameters forcefully needed to satisfy $x_b < 1$ and $y_b < y_c$ (we present with diamond symbols in Fig. 4.7 results from one of such configurations). This means that our model is only able to explain these results when the parameters are set to strong suppression. This is in line with the results in [42], where the authors propose a phenomenological suppression rule to explain their findings.

While the mechanism for LTD may indeed have a different biological implementation in these neurons, the prediction from both models seems to aim at the same conclusion, that strong suppression effects need to be in place to explain these measured nonlinearities.

At this point it is also important to note that while the protocol used in [12] and [115] for pairwise and triplet stimulation in hippocampal culture is symmetric; with both pre- and postsynaptic neurons being intracellularly stimulated. In [42], presynaptic stimulation is extracellular. Firstly, this induces an asymmetry between PrePostPre and PostPrePost triplets. Probably more than one cell is excited by extracellular stimulation, producing network effects, potentially inducing other forms of plasticity, such as local synaptic scaling [109]. Secondly, the larger degree of variability observed could indeed be due to the larger number of variables involved coming from the network. We point this out to try to understand the differences observed, and not as a criticism of the experimental procedure. It needs to be noted that these measurements are done in a cortical slice and not in culture as in the case of [12] and [115], with all the difficulties this involves.

As a final note, the best fit value of $y_c = 11.6$, is much larger than the one obtained for the hippocampal fit. The high level of noise in the data, creates a very wide minimum region for the fit and indeed, it is possible to obtain a reasonable fit to the data for $y_c = 1.0$, by setting $y_b = 0.9$, and $x_b = 0.4$ (corresponding to the diamond symbols in Fig. 4.7), which we present as a reference. If we had further information available, we could attempt a fit under the constraints of certain boundary values for the parameters. Since we do not count with this information, we simply show these possibilities, stressing that for the model to reproduce this particular type of nonlinearities the condition of strong saturation is unavoidable, regardless of the size of the selected y_c.

4.5 Frequency dependent plasticity: from spikes to rates

In the past sections of this chapter we have worked with highly structured low-frequency stimulation motifs. In this final section we will do the opposite, studying the implications of our STDP plasticity rule to rate-encoding plasticity, by evaluating the induced synaptic weight change produced by random uncorrelated spike trains at different frequencies.

Up to this point, we have closely followed with our theoretical analysis, experimental protocols involving low frequency pre- and postsynaptic stimulation of neurons from two different sources: hippocampal culture and cortical layer 2/3 neurons. Since, for each type of neuron, consistent experiments were available, where

pairwise and triplet results were obtained using the same stimulation technique, we were able to determine a set of parameters for each type of neuron and perform a direct comparison between experiment an model. We would like to now take one more step, evaluating the effect of high frequency stimulation, or the dependence of LTP and LTD on the stimulation frequency, in the same neurons, with the same parameters of our model. A quantitative comparison of this kind is unfortunately not possible.

While experiments testing frequency effects on LTP and LTD exist in the literature for other neural types and/or other stimulation protocols, such as in [104], where frequency dependent plasticity is studied in layer 5 neurons in visual cortex, we have found none using the same stimulation technique on the same neurons here studied. And vice-versa: in the case of the layer 5 cortical neurons used in [104], no triplet data is available and the pairwise results observed with their protocol are qualitatively very different from the one found in layer 2/3 by [42], making a quantitative comparison not possible.

The way we will proceed is then to perform numerical experiments with the two neural types we have been using so far, keeping for each their already fitted parameters. In each case, we will also perform the numerical experiments with at least one other set of parameters, for comparison. For each neural type, we will present the model's predictions and contrast these qualitatively to experimental results from other neurons and protocols.

We want to determine here the amount of synaptic modification induced by uncorrelated Poisson trains of pre- and postsynaptic spikes. We will work with two types of neurons, which we will call here as hippocampal-type and cortical-type. Hippocampal neurons will be for us those defined by the pairwise parameters (4.13), and non-saturated traces, which qualitatively reproduce the triplet nonlinearities from Fig. 4.5. We will consider two parameter configurations for this type, the best fit presented in Fig. 4.5 ($y_c = 0.28$, $x_b = 0.62$, and $y_b = 0.66$.), and a second set of parameters ($y_c = 0.8$, $x_b = 1.82$, and $y_b = 1.34$. SD 7.37, compared to the 6.76 of the best fit) for illustration of the effect of y_c. Analogously, cortical neurons will be for us those defined by the pairwise parameters (4.14), and now strongly saturated traces, which qualitatively reproduce the triplet nonlinearities from Fig. 4.7. We also consider two parameter configurations for this type: the best fit ($y_c = 11.6$, $x_b = 0.5$, and $y_b = 10.9$), and the parameters from the diamond symbols ($y_c = 1.0$, $x_b = 0.4$, and $y_b = 0.9$), both already presented in Fig. 4.7.

Each numerical simulation consists of a $1\mathrm{s}$ presentation of pre- and postsynaptic Poisson spike trains of frequencies f_{pre} and f_{post}, respectively. After training is completed, we evaluate the induced synaptic weight change ΔW. Since function $\Delta W\left(f_{pre}, f_{post}\right)$ depends on two variables, to represent it graphically, we will produce two types of plots:

- $\Delta W\left(f_{pre} = const,\ f_{post}\right)$ for several constant values of f_{pre},
- $\Delta W\left(f_{pre},\ f_{post} = f_{pre}\right)$.

The second kind could be considered as a case in which the output activity is a function g of the input activity, in the most simple case in which $f_{pre} = f_{post}$. We note that while the frequencies are the same, we keep the spike times uncorrelated. The effect of correlated spikes is discussed at the end of the section.

The numerical results for hippocampal-type neurons are presented in Fig. 4.8, and those for cortical-type neurons are plotted in Fig. 4.9.

Let us begin by analyzing the case of hippocampal neurons. In the plots for constant presynaptic frequency, we generically find a switch from depression to potentiation at a certain threshold frequency θ_H, where this threshold is a a monotonically increasing function of the presynaptic frequency $\theta_H = \theta_H\left(f_{pre}\right)$, with $\theta_H\left(f^1_{pre}\right) > \theta_H\left(f^2_{pre}\right)$ if $f^1_{pre} > f^2_{pre}$. For this reason we will refer to θ_H as a sliding threshold. Let us remember at this point that similar sliding thresholds are usually employed in rate-encoding plasticity rules, such as BCM [14]. In those cases, the threshold usually depends however on the time-averaged postsynaptic activity. We have not included in our model any slow-varying variable that could work as a memory of this kind, so our θ_H is a function of the present frequency only. The fact that our sliding threshold depends on the presynaptic activity, means that it is adjusted independently for each synapse of the neuron. At each synapse, the presynaptic activity determines via θ_H, the level of postsynaptic activity that needs to be considered as significant to trigger potentiation. We can qualitatively summarize the functional dependence of the weight change on the frequencies as: $\Delta W \propto f_{pre} \cdot (f_{post} - \theta_H(f_{pre}))$

By comparing the two sets of parameters, we observe that for small values of y_c (Fig. 4.8 **a)**), $\theta_H\left(f_{pre}\right) > f_{pre}$, and therefore depression dominates in the plot corresponding to $f_{post} = f_{pre} < \theta_H\left(f_{pre}\right)$. For larger values of y_c (Fig. 4.8 **b)**), $\theta_H\left(f_{pre}\right) < f_{pre}$, and potentiation dominates. In this way, if we counted with the experimental data, we could restrict the value of y_c during the triplet fit, to reproduce the frequency dependent plasticity results. As an example, if we wanted a similar frequency dependence to that found in [104] for L5 cortical neurons, where potentiation dominates for large frequencies, we would select parameters closer to those from Fig. 4.8 **b)**.

We should note that y_c plays two roles in the model: determining the contribution of VGCCs to the calcium current, and setting the value of the LTP threshold, which we fixed to y_c to recover the pure exponential shapes for pairwise STDP. It is therefore not trivial to associate the dependence of potentiation/depression on y_c to a single mechanism, in this case.

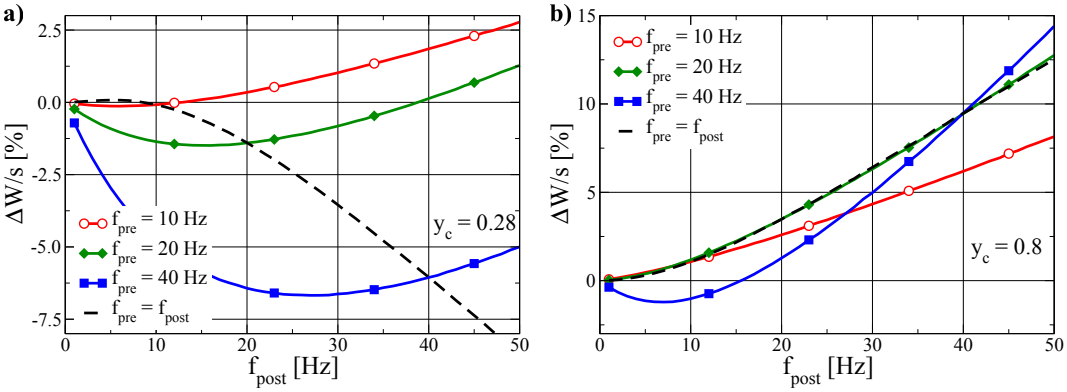


Figure 4.8: *Frequency dependent plasticity in hippocampal neurons induced by uncorrelated Poisson pre- and postsynaptic spike trains. Full lines: varying f_{post}, constant f_{pre}. Dashed lines: varying $f_{post} = f_{pre}$.*
Parameters: Pairwise STDP parameters from (4.13) plus:
a) $y_c = 0.28$, $y_b = 0.66$, *and* $x_b = 0.62$.
b) $y_c = 0.8$, $y_b = 1.34$, *and* $x_b = 1.82$.

In Fig. 4.9 we now present the the results of the numerical simulations for the cortical-type neurons. A stark difference can be observed, since in this case for both parameter configurations depression dominates for uncorrelated spike trains of all frequencies. Indeed this feature was present for all parameter configurations reproducing the cortical triplet nonlinearity of Fig. 4.7, which need to respect $y_b < y_c$, as discussed in section 4.4.2. As a test, keeping the other parameters unchanged, potentiation for large frequencies could be recovered by allowing $y_c < y_b$, losing at the same time the characteristic triplet nonlinearities. While this result might be controversial when compared, for instance, with the already mentioned frequency dependent results for L5 cortical neurons from [104], we show that they are directly related to the particular triplet results of L2/3 cortical neurons from [42]. Unfortunately, we do not count with triplet data for L5 cortical neurons as in [104], and we have already discussed the possible bias introduced by the asymmetric stimulation protocol employed by [42]. The model makes then interesting predictions for both neural types, and it would be interesting to count in the future with the missing experimental data.

The previous discussion is valid for high frequencies, where triplet interaction is strong. For low frequencies we also observe a small level of LTD for uncorrelated spikes. In the low frequency limit, triplets become less and less common in Poisson spike trains, and pairwise plasticity dominates. Uncorrelated spikes then randomly sample the pairwise STDP plot from Fig 4.6, and the resulting plasticity sign is then determined by the ratio r of the areas below LTP and the LTD sides of the curve:

$$r = \frac{A^+\tau_+}{A^-\tau_-}. \tag{4.15}$$

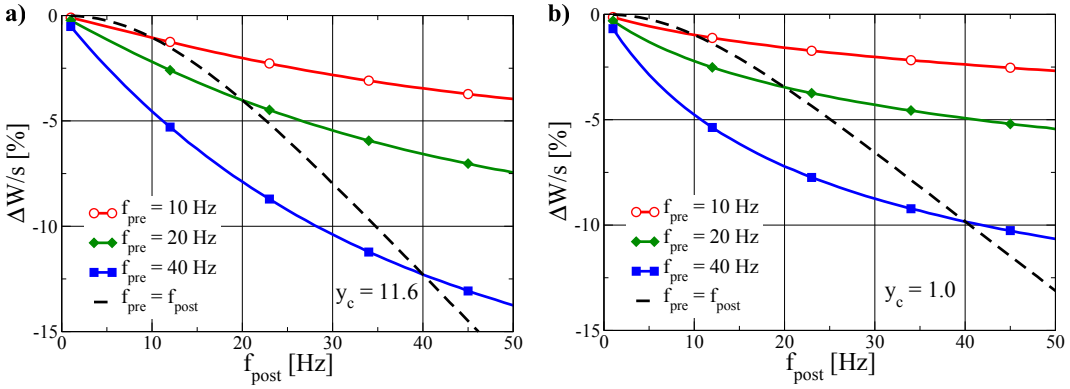


Figure 4.9: *As in Fig. 4.8, now for visual cortical neurons. Full lines: varying f_{post}, constant f_{pre}. Dashed lines: varying $f_{post} = f_{pre}$.*
Parameters: Pairwise STDP parameters from(4.14) plus:
a) $y_c = 11.6$, $y_b = 10.9$, *and* $x_b = 0.5$.
b) $y_c = 1.0$, $y_b = 0.9$, *and* $x_b = 0.4$.

We can easily compute this ratio for both cortical and hippocampal-type neurons, finding: for cortex $r = 0.77$ (LTD wins) and for hippocampus $r = 1.92$ (LTP wins). While hard to visually note from the plots, for both low f_{pre} and f_{post}, and uncorrelated Poisson spikes, LTP is small but prevailing in hippocampal neurons while LTD (also small) prevails for cortical-type neurons. The effects are small since, for low frequencies, uncorrelated random spikes tend to have a very large average Δt.

This last feature is in line with other model's findings, such as in [65], where it is shown how a straightforward application of the linear pairwise rule to every pair in the Poisson uncorrelated spike trains, always leads to depression for cortical parameters. This result changes however when the authors consider first neighbor pairs only, going from dominating LTD to BCM-like. It is however not clear how such a first-neighbor interaction could be implemented, since it would probably require hard reseting of the traces.

We have checked that potentiation is recovered in cortical-type neurons at high frequencies if the spikes are correlated. As an example, for a Poisson train of fixed PrePost pairs, with $\Delta t = 5\text{ms}$ at 10Hz, a 9% increase in the synaptic connection is found after 1s (corresponding to, on average, 10 pairs), or 54% after 60 pairs (as used in 4.3.1). Along these lines, it would be interesting in future work to study the effect from more realistic correlations generated by the neuron's own integration of inputs. To that end, we would need to add a neural model on top of our single-synapse model, and feed the neuron with an N_w dimensional spike time distribution, computing the corresponding output online, in a similar fashion

to what we did in **Chapter 2**, now for time-encoding neurons.

We finally mention that the result of prevailing depression for uncorrelated spike trains in our model cortical neurons, is in line with at least one interpretation of deprivation experiments. In certain cortical areas which usually present topological maps, deprivation of sensory input induces depression of the respective synaptic connections [39, 107]. Simultaneously, correlations in areas projecting to cortex are reduced after these procedures [73]. While simultaneity does not determine causality, one possible explanation of these results would be that decorrelation of the spike trains (produced by the lack of input) results in depression of the respective cortical connections.

4.6 Discussion

In this chapter we have presented a simplified biophysical model for STDP, which computes the evolution of the synaptic connection strength in terms of the calcium concentration in the postsynaptic spine and the fraction of open NMDA receptors. These traces, serve as two clocks, allowing the synapse to keep track of the timing of pre- and postsynaptic spikes. The resulting online synaptic plasticity rule is time-dependent and reproduces the experimentally observed plasticity results for pairwise and triplet stimulation protocols in hippocampal and cortical neurons.

The model, we believe, strikes an interesting balance, with its functional simplicity allowing us to analytically compute the induced synaptic change from several protocols, including pairwise and triplet stimulation, as we have shown here. Despite this simplicity, the model is able to capture most features of time-dependent LTP and LTD, and relate these changes to the evolution of the underlying traces.

Triplet non-linearities, as evidenced by triplet protocols in both hippocampal and cortical neurons, make it clear that a simple pairwise rule is insufficient to account for plasticity in more general scenarios, and nothing suggests that an expansion approach, by computing separately, the contribution of several orders: pairs, triplets, etc., would be any more successful. Moreover, such a model would not improve our understanding of the process. We believe then that it is important to count with an online learning rule, formulated in terms of biologically plausible variables, that can account for the observed plasticity results. The model we propose, does exactly that: it computes plasticity online, locally, and it does so with traces that have been experimentally found to be essential for LTP and LTD, namely the calcium concentration and NMDA receptors.

It should be clear here, that we do not claim the model to capture the full biological complexity of the process, and indeed, as mentioned several times throughout

the chapter, we consider our model to be an effective model for STDP, where the combined chain effect of a large number of biological elements is pooled together within a few variables. This simplification, needed for analytic tractability, is also, we believe, an advantage in terms of a simpler understanding of the general rules at play. An example of this is the case of our application of our model to cortical L2/3 neurons. While, as mentioned, other mechanisms seem to be at play -at least during LTD- in these neurons, as previously discussed, the model is nonetheless able to capture these neurons' particular triplet profile. In previous work [42], a phenomenological rule had been proposed for these neurons involving a suppression mechanism. With our model, we observe that only when the traces are strongly saturated and suppression is strong, can this type of triplet profile be recovered, supporting this previous intuition, now in terms of a trace model, therefore going one step further in the understanding of the mechanism. This is clearly not the end of the story, and further research, formulating plasticity in terms of a more accurate description for cortical neurons is still needed to provide confirmation of our predictions. Our model's predictions, however, provide a hint of what to look for. In a similar way, our model hints at trace accumulation as a source of triplet nonlinearities in hippocampal neurons, with our model producing a very good fit of the observed experimental results.

Moreover, as we discussed in the last section, the model still leaves room to adapt the parameters to reproduce different types of frequency dependencies, at least for hippocampal-type neurons. While we cannot directly compare our results with experiment, since they have not been carried out for the same types of neurons and with the same type of stimulation protocol, we show that both increasing potentiation or depression can be achieved for hippocampal-type neurons, by adjusting the model's parameters, while still reproducing pairwise and triplet results. In particular, the behavior of the observed sliding threshold is determined, as we have shown, by the value of y_c, resulting in predominant LTP or LTD. In future work simultaneous online adaptation of y_c could be employed to functionally play the role of an intrinsic plasticity mechanism [75, 79, 108], similar to that employed in **Chapters 2** and **3**.

The triplet nonlinearities of cortical neurons impose very heavy restrictions in terms of suppression, and at least with our model, depression for uncorrelated high frequency stimulation can only be avoided by adding correlations between pre- and postsynaptic spikes. As previously discussed, this might not be altogether unrealistic, as deprivation experiments show simultaneous decorrelation in the activity of neurons projecting to cortical areas, and depression in the respective connections. Further research is required along these lines, since many uncertainties remain. The large variability in the data we count with, the mentioned asymmetry in the the stimulation protocol employed in this case (using extracellular presynaptic stimulation), and the additional presynaptic mechanism thought to take place in

cortical LTD, all indicate that a long way still needs to be covered in this direction.

Finally, we believe that, at least for hippocampal neurons, the model has shown to reproduce a wide enough range of features for both low frequency stimulation and frequency dependent plasticity as to be tested in a wider range of applications. Future research along these lines should include a model for neural integration, such as that presented in 1.3.2, to go from the single synapse level to the whole cell level, being able to test the predictions of the learning rule for arbitrary input distributions, and output distributions produced by the neuron's response function, in a similar spirit as that of **Chapters 2** and **3**. We believe that our model's analytic simplicity makes it a good candidate for this kind of studies, where the complexity tends to increase very fast as one builds extended networks.

Chapter 5

Conclusions

In this work we have presented two complementary approaches to synaptic plasticity: a top-down approach, in which the learning rules are derived from a guiding principle, and a bottom-up approach, building the plasticity rule out of the key biological ingredients thought to take place in the process. Despite the different paths taken, a common theme has been present throughout this journey: the quest for simplicity. Finding minimal models that capture the essence of the processes at hand. Not because of an aesthetic fetish, but out of the conviction that simplicity and reduction to the essence of a phenomenon, can help better understand the role of different components in complex systems.

The first of the two paths that we took was that of developing a synaptic plasticity rule from a guiding principle, mathematically formulated as an objective function. The idea of guiding principles is to express a set of goals a system might have (computational, metabolic, homeostatic, among others), and to generate a set of adaption rules that guide the system towards the satisfaction of these goals. In the present work we have worked in particular with the stationarity principle of statistical learning, simply stating that, once the relevant features of a stationary input distribution have been learnt, the output distribution should also be stationary. For this to be possible in a noisy environment, we have argued that the solution found in the space of synaptic weights by the neuron needs to be stable, in the sense that it should be locally insensitive to further changes of the synaptic weights.

To express this local insensitivity condition, we have resorted to the Fisher Information, a measure of the average sensitivity of a probability distribution to a given parameter. We have extended this quantity to a multidimensional parameter space in a local way, termed the local synapse extension, and in this way derived a local, online, self-limiting, and Hebbian learning rule. While this model does not pretend to pass as a biophysically realistic model (it violates for instance Dale's law, by having synapses that can switch from excitatory to inhibitory in time), we believe that any rule from which we intend to learn something about the brain should be local and online. Contrary to an artificial neural network for machine

learning applications, a real synapse cannot have direct information about the value of every other synapse and also cannot store the whole history of a train of spikes for later processing. This would result in fundamentally different learning algorithms. For this reason, we have ensured with our local synapse approximation to keep the learning rules local, and the stochastic gradient descent procedure produces an online learning rule (in contrast to batch gradient descent, which computes an average over the whole probability distribution on every step).

We have shown the robustness of the learning rule with respect to the choice of transfer function in the neural model, finding in particular that the learning rule is factorized into a Hebbian function and a self-limiting function, for sigmoidal transfer functions. Altogether we have shown three examples of learning rules for transfer functions within this family: the Fermi function (or exponential sigmoidal), the arc-tangent function, and the re-scaled error function. All produce qualitatively equivalent results. This kind or robustness is important for biological plausibility. A learning rule requiring the system to reproduce a given function exactly would prove of little use in real noisy systems.

In our quest for simplicity and analytic tractability, we have kept stripping the learning rule of any degree of complexity that did not add to the computational capabilities of the neuron, arriving finally to the cubic learning rule, either as an approximation around the roots of any rule derived from a sigmoidal transfer function, or as the direct outcome of the objective function formulation when the error function is employed as transfer function. This procedure has allowed us to compute the attractors of the learning rule and their stability, in terms of the moments of the input distribution. In this way, we have been able to justify analytically our numerical findings that the rule picked up the FPC of input distributions resembling a multivariate normal distribution, and had a preference for directions of large negative kurtosis otherwise. Being able to compute this dependence analytically, allowed us to go one step further and predict that the rule would be suitable for ICA. We tested the prediction numerically by training a neuron operating under our rule, with the input set from the non-linear bars problem. As expected, the neuron was able to learn single bars (the independent components of the problem), even when these were never presented in isolation (permanent partial occlusion).

The importance of these results resides in their generality. If one wants a learning rule that is Hebbian for a certain range of activities, but then reverses its slope - and eventually its sign - when the neural activity gets too large or too small, the cubic shape (here in the broad sense of the word shape), is the minimal construction one can imagine. What we are showing here is that such a minimal construction is already very powerful computationally and that it is highly robust to quantitative deviations, as long as the general shape is maintained.

The second path we took was that of building a synaptic plasticity rule for STDP, out of the key elements thought to be involved in this process at glutamatergic synapses. Emphasis was once again made in constructing a local, online synaptic plasticity rule, formulated in this case in terms of two traces, each serving as a clock, timing the occurrence of pre- and postsynaptic spikes. The two traces, namely the fraction of activated NMDA receptors and the calcium concentration at the postsynaptic spine, interact, allowing for non linear spike effects.

Also in this procedure, simplicity, and capturing the essence of the problem, have been constant goals, keeping the equations linear or polynomial, with a minimal amount of variables and parameters. This approach has paid off, since in this way we have been able to analytically compute the model's predictions for standard low frequency experimental protocols involving pairs and triplets of spikes. Despite the model's simplicity, it is able to reproduce these results in both hippocampal and cortical neurons. The transparent link between the model's internal mechanisms and their outcome in the observed plasticity results, allows us to learn more about the underpinnings of the process and generate interesting predictions. While the model is simplistic, and in many ways an effective model (pooling together the effect of a large number of biological elements into a few variables), we believe it makes important general predictions, to be tested experimentally and with more detailed models.

One of these predictions is that trace accumulation is the main source of triplet nonlinearities in hippocampal neurons, while saturation is dominant in L2/3 cortical neurons. Previous phenomenological learning rules had already proposed explicit suppression as a way of explaining triplets in these cortical neurons. Here we come to this conclusion without forcing it, from a more general model. Parameter configurations that reproduce this kind of triplet nonlinearities present all strong trace saturation effects, and vice-versa. While, again, STDP may very well be implemented differently in these neurons, our model effectively predicts that strong suppression effects should be present to explain these results.

To close the loop we started in the first chapters, we mapped the time-dependent plasticity rule into a rate-dependent one, by evaluating the amount of synaptic modification induced by Poisson uncorrelated trains of pre- and postsynaptic spikes. This allowed us to determine the weight change predicted by the model, as a function of the pre- and postsynaptic firing rates. Interestingly for hippocampal-type neurons, a Hebbian (in the rate encoding interpretation of the word) rule emerges. The synaptic weight change results proportional to the product of the frequencies, with the postsynaptic frequency being affected by an LTP threshold that is a monotonic function of the level of presynaptic activity. This has a strong reminiscence to the BCM rule, except that our threshold is not a long-term average of the activity (no such timescales can be generated in our model), an its based on the level of

presynaptic activity, setting a different threshold for each synapse in the neuron.

In the case of L 2/3 cortical neurons, we observe with our model an unavoidable link between triplet results and frequency dependent results. Parameter configurations that reproduce cortical-type nonlinearities, also result in overall depression for uncorrelated spike trains. While simultaneous decorrelation of neural activity and depression is found in cortex during sensory deprivation experiments, we believe more data is required. We consider that the extra-cellular form of presynaptic stimulation in this experimental protocol, potentially resulting in strong network effects, might explain the large variability in the data and might have an influence on the observed triplet nonlinearities. While the experimental difficulties in conducting these experiments are clear, and this is in no way a criticism of the employed methods, we find it would be crucial for further understanding of the mechanism to count with a complete and consistent set of data for pairwise, triplet, and frequency dependent stimulation, all in the same type of neuron and with the same stimulation technique.

Our STDP rule is defined for one synapse; we have not included a model of the neuron's integration of inputs, since the experiments we have reproduced force the output activity and study one synapse at the time. A clear future line of work is to include a whole neuron model (such as that presented in section 1.3.2), where the output activity is computed directly as a function of the inputs, to then evaluate the resulting rate-encoding rule. In particular, we would like to study whether the rule is also self-limiting, or whether additional homeostatic mechanisms are needed to achieve this.

Finally, in order to understand how synaptic plasticity and neural activity interact and mutually condition each other, we would like in future work to study either of the rules here presented in extended neural networks. The complexity of the problem grows fast as one includes full network interactions, which means that having a tractable neural and plasticity model is vital. We believe that the here proposed rules, being minimal in their formulation are then ideal candidates for this task.

Appendix

A. Dimensionality reduction in the STDP model

In section 4.2, we chose a compact form for Eq. (4.1), in which we have absorbed several constants. In this section we clarify this point.

Let us begin by denoting here as x' and y' the fraction of NMDA receptors and the Ca^{2+} concentration, whose evolution evolves according to:

$$\begin{cases} \dot{x'} = -\frac{x'}{\tau_x} + c_1 E_x \sum_\sigma \delta(t - t^\sigma_{pre}) \\ \dot{y'} = -\frac{y'}{\tau_y} + (c_2 x' + y'_c) E_y \sum_\sigma \delta(t - t^\sigma_{post}) \end{cases} \tag{5.1}$$

where also here τ_y and τ_x represent the decay time constants of x' and y'. Comparing to (4.1) one sees that two extra parameters c_1 and c_2 are present. c_1 represents the increase in the value of x' caused by a single presynaptic spike (which we assume to be constant) and c_2 represents the increase of y' per amount of x'. Finally y'_c stands for the constant contribution to y' of each postsynaptic spike (via the VGCCs). Just as in section 4.2, a spike efficacy E is present in the model limiting trace concentrations (we compute E according to (4.2)).

It turns out that Eqs. (4.1) and (5.1), are completely equivalent, and one can go from one to the other by a change of variables:

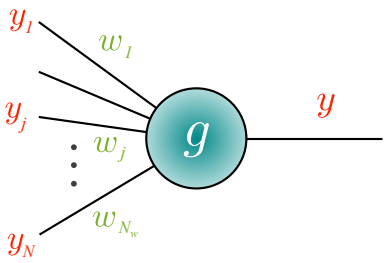
$$x = x'/c_1, \qquad x_b = x'_b/c_1, \qquad y = y'/c_1 c_2, \qquad y_c = y'_c/c_1 c_2, \qquad y_b = y'_b/c_1 c_2. \tag{5.2}$$

For this reason, we have chosen to present the functionally equivalent, but more compact form (4.1), in this work.

Model summary cards

As a help to the reader we briefly summarize here both synaptic plasticity models.

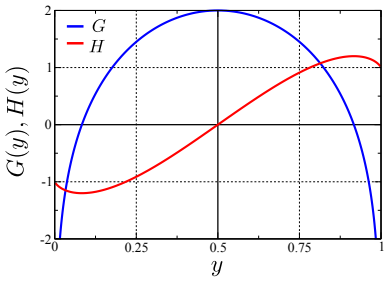
1) *Neural and plasticity models from* ***Chapters 2*** *and* ***3***:



$$y = g(x - b), \quad x = \sum_{j=1}^{N_w} w_j (y_j - \bar{y}_j)$$

$$\dot{w}_j = \epsilon_w G(x) H(x) (y_j - \bar{y}_j)$$

Options for the bias:

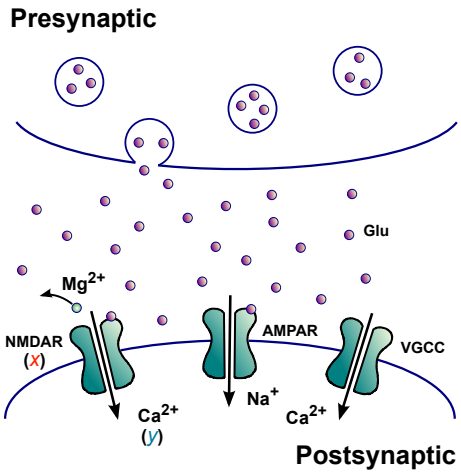


$$\begin{aligned} \dot{b}_i &= 0 \\ \dot{b}_i &= \epsilon_b * (y_i - p) \\ \dot{b}_i &= -\epsilon_b \nabla F_{KL}^{int} \\ &= -\epsilon_b \big(1 - 2y_i + y_i(1 - y_i)\lambda\big) \end{aligned}$$

Variables and parameters:

N_w : *number of inputs to the neuron*
y_j : *input* j
$\bar{y}_j$: *trailing average of input* j
x : *integrated input*
y : *neural output (input to other neurons)*
w_j : *synaptic weight strength* j
b : *bias*
ϵ_w and ϵ_b : *synaptic and intrinsic learning rates*
p : *target activity level for simple homeostatic adaption*
λ : *from the target exponential activity distribution* $p(y) \propto exp(-\lambda\, y)$

2) *STDP model from* ***Chapter 4***:



$$\begin{cases} \dot{x} = -\frac{x}{\tau_x} + E_x(x)\Sigma_\sigma \delta(t - t^\sigma_{pre}) \\ \dot{y} = -\frac{y}{\tau_y} + (x + y_c)E_y(y)\Sigma_\sigma \delta(t - t^\sigma_{post}) \end{cases}$$

$$E_z(z) = \theta(z_b - z)(1 - \frac{z}{z_b}), \quad z = x, y$$

$$\dot{w} = \alpha x(y - b)\theta(y - b)\,\Sigma_\sigma\,\delta(t - t^\sigma_{post}) - \beta x y\,\Sigma_\sigma\,\delta(t - t^\sigma_{pre})$$

Variables and parameters:

x : *presynaptic trace (fraction of activated NMDA receptors)*
y : *postsynaptic trace* (Ca^{2+} *concentration)*
y_c : *VGCCs contribution to the* Ca^{2+} *concentration*
τ_x and τ_y : *decay time constants of the traces*
x_b and y_b : *reference levels for the saturation functions* E_x *and* E_y
w : *synaptic weight strength*
α and β : *LTP and LTD strength constants*
b : *LTP threshold*

To obtain the classic pairwise STDP shape one sets:

$b = y_c$, $\alpha = A^+$, $\beta = A^-/y_c$, $\tau_x = 2\tau_+$, and $\tau_y = \tau_-$

A^+, A^-, τ_+ and τ_- : *maximal intensities and timescales of LTP and LTD for isolated spike pairs*

References

[1] Alison Abbott. Brain-simulation and graphene projects win billion-euro competition. *Nature News*, 2013.

[2] Larry F Abbott and Sacha B Nelson. Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–1183, 2000.

[3] Christian Albers, Joscha T Schmiedt, and Klaus R Pawelzik. Theta-specific susceptibility in a model of adaptive synaptic plasticity. *Front Comput Neurosci*, 7:170, 2013.

[4] Neil W Ashcroft and N David Mermin. Solid state physics (saunders college, philadelphia, 1976). *Appendix N*, 2010.

[5] Nihat Ay, Nils Bertschinger, Ralf Der, Frank Güttler, and Eckehard Olbrich. Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B*, 63(3):329–339, 2008.

[6] Frederico AC Azevedo, Ludmila RB Carvalho, Lea T Grinberg, José Marcelo Farfel, Renata EL Ferretti, Renata EP Leite, Roberto Lent, Suzana Herculano-Houzel, et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541, 2009.

[7] Mathilde Badoual, Quan Zou, Andrew P Davison, Michael Rudolph, Thierry Bal, Yves Frégnac, and Alain Destexhe. Biophysical and phenomenological models of multiple spike interactions in spike-timing dependent plasticity. *International journal of neural systems*, 16(02):79–97, 2006.

[8] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Physical review letters*, 59(4):381, 1987.

[9] Anthony J Bell and Terrence J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[10] M Bethge, D Rotermund, and K Pawelzik. Optimal neural rate coding leads to bimodal firing rate distributions. *Network: Computation in Neural Systems*, 14(2):303–319, 2003.

[11] Guo-Qiang Bi. Spatiotemporal specificity of synaptic plasticity: cellular rules and mechanisms. *Biological cybernetics*, 87(5-6):319–332, 2002.

[12] Guo-qiang Bi and Mu-ming Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of neuroscience*, 18(24):10464–10472, 1998.

[13] Guo-Qiang Bi and Jonathan Rubin. Timing in synaptic plasticity: from detection to integration. *TRENDS in Neurosciences*, 28(5):222–228, 2005.

[14] Elie L Bienenstock, Leon N Cooper, and Paul W Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience*, 2(1):32–48, 1982.

[15] Léon Bottou. Online learning and stochastic approximations. *On-line learning in neural networks*, 17(9):142, 1998.

[16] Nicolas Brunel and Jean-Pierre Nadal. Mutual information, fisher information, and population coding. *Neural Computation*, 10(7):1731–1757, 1998.

[17] John Buck. Synchronous rhythmic flashing of fireflies. ii. *Quarterly review of biology*, pages 265–289, 1988.

[18] Ernesto Carafoli. Intracellular calcium homeostasis. *Annual review of biochemistry*, 56(1):395–433, 1987.

[19] Yael Chagnac-Amitai, Heiko J Luhmann, and David A Prince. Burst generating and regular spiking layer 5 pyramidal neurons of rat neocortex have different morphological features. *Journal of Comparative Neurology*, 296(4):598–613, 1990.

[20] Claudia Clopath and Wulfram Gerstner. Voltage and spike timing interact in stdp–a unified model. *Spike-timing dependent plasticity*, page 294, 2010.

[21] Roger J Colbran. Protein phosphatases and calcium/calmodulin-dependent protein kinase ii-dependent synaptic plasticity. *The Journal of neuroscience*, 24(39):8404–8409, 2004.

[22] Pierre Comon. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314, 1994.

[23] Leon N Cooper and Mark F Bear. The bcm theory of synapse modification at 30: interaction of theory with experiment. *Nature Reviews Neuroscience*, 13(11):798–810, 2012.

[24] RJ Cormier, AC Greenwood, and JA Connor. Bidirectional synaptic plasticity correlated with the magnitude of dendritic calcium transients above a threshold. *Journal of Neurophysiology*, 85(1):399–406, 2001.

[25] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience*, volume 806. Cambridge, MA: MIT Press, 2001.

[26] Lawrence T DeCarlo. On the meaning and use of kurtosis. *Psychological Methods*, 2(3):292–307, 1997.

[27] Javier DeFelipe, Lidia Alonso-Nanclares, and Jon I Arellano. Microstructure of the neocortex: comparative aspects. *Journal of neurocytology*, 31(3-5):299–316, 2002.

[28] James J DiCarlo, Davide Zoccolan, and Nicole C Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, 2012.

[29] Dawei W Dong and John J Hopfield. Dynamic properties of neural networks with adapting synapses. *Network: Computation in Neural Systems*, 3(3):267–283, 1992.

[30] David A Drachman. Do we have brain to spare? *Neurology*, 64(12):2004–2005, 2005.

[31] John Duncan. An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience*, 2(11):820–829, 2001.

[32] Rodrigo Echeveste, Samuel Eckmann, and Claudius Gros. The fisher information as a neural guiding principle for independent component analysis. *Entropy*, 17(6):3838–3856, 2015.

[33] Rodrigo Echeveste and Claudius Gros. Generating functionals for computational intelligence: The fisher information as an objective function for self-limiting hebbian learning rules. *Frontiers in Robotics and AI*, 1:1, 2014.

[34] Rodrigo Echeveste and Claudius Gros. An objective function for self-limiting neural plasticity rules. *Proceedings of the 23th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, 2015.

[35] Rodrigo Echeveste and Claudius Gros. Two-trace model for spike-timing-dependent synaptic plasticity. *Neural computation*, 2015.

[36] Alexander S Ecker, Philipp Berens, Andreas S Tolias, and Matthias Bethge. The effect of noise correlations in populations of diversely tuned neurons. *The Journal of Neuroscience*, 31(40):14272–14283, 2011.

[37] Samuel Eckmann. Bachelor thesis: Cubic learning rules for unsupervised self-limiting hebbian learning in artificial neural networks. Master's thesis, Institute for Theoretical Physics Goethe University Frankfurt am Main, 2015.

[38] Terry Elliott. An analysis of synaptic normalization in a general class of hebbian models. *Neural Computation*, 15(4):937–963, 2003.

[39] Daniel E Feldman. Timing-based ltp and ltd at vertical inputs to layer ii/iii pyramidal cells in rat barrel cortex. *Neuron*, 27(1):45–56, 2000.

[40] Peter Földiak. Forming sparse representations by local anti-hebbian learning. *Biological cybernetics*, 64(2):165–170, 1990.

[41] Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010.

[42] Robert C Froemke and Yang Dan. Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature*, 416(6879):433–438, 2002.

[43] Wulfram Gerstner, Richard Kempter, J Leo van Hemmen, and Hermann Wagner. A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383(LCN-ARTICLE-1996-002):76–78, 1996.

[44] Wulfram Gerstner, Werner M Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

[45] Mark Girolami and Colin Fyfe. Negentropy and kurtosis as projection pursuit indices provide generalised ica algorithms. In *A. C, Back A (eds.), NIPS-96 Blind Signal Separation Workshop*, volume 8, 1996.

[46] Geoffrey J Goodhill and Harry G Barrow. The role of weight normalization in competitive learning. *Neural Computation*, 6(2):255–269, 1994.

[47] Geoffrey J Goodhill and Terrence J Sejnowski. A unifying objective function for topographic mappings. *Neural Computation*, 9(6):1291–1303, 1997.

[48] Michael Graupner and Nicolas Brunel. Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proceedings of the National Academy of Sciences*, 109(10):3991–3996, 2012.

[49] Charles Miller Grinstead and James Laurie Snell. *Introduction to probability*. American Mathematical Soc., 2012.

[50] C. Gros. *Complex and adaptive dynamical systems: A primer*. Springer Verlag, 2010.

[51] Claudius Gros. Generating functionals for guided self-organization. In M. Prokopenko, editor, *Guided Self-Organization: Inception*, pages 53–66. Springer, 2014.

[52] Diego A Gutnisky and Valentin Dragoi. Adaptive coding of visual information in neural populations. *Nature*, 452(7184):220–224, 2008.

[53] Frank E Hanson. Comparative studies of firefly pacemakers. In *Federation proceedings*, volume 37, pages 2158–2164, 1978.

[54] Jiang Hao and Thomas G Oertner. Depolarization gates spine calcium transients and spike-timing-dependent potentiation. *Current opinion in neurobiology*, 22(3):509–515, 2012.

[55] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, 1994.

[56] Donald Olding Hebb. *The Organization of Behavior*. Wiley, 1949.

[57] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2002.

[58] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[59] Yan-You Huang, Asuncion Colino, David K Selig, and Robert C Malenka. The influence of prior synaptic activity on the induction of long-term potentiation. *Science*, 255(5045):730–733, 1992.

[60] Yanhua H Huang and Dwight E Bergles. Glutamate transporters bring competition to the synapse. *Current opinion in neurobiology*, 14(3):346–352, 2004.

[61] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574–591, 1959.

[62] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106, 1962.

[63] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.

[64] Nathan Intrator and Leon N Cooper. Objective function formulation of the bcm theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5(1):3–17, 1992.

[65] Eugene M Izhikevich and Niraj S Desai. Relating stdp to bcm. *Neural computation*, 15(7):1511–1523, 2003.

[66] Uma R Karmarkar and Dean V Buonomano. A model of spike-timing dependent plasticity: one or two coincidence detectors? *Journal of Neurophysiology*, 88(1):507–513, 2002.

[67] Jim W Kay and WA Phillips. Coherent infomax as a computational goal for neural systems. *Bulletin of mathematical biology*, 73(2):344–372, 2011.

[68] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques, 2007.

[69] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.

[70] Petr Lansky and Priscilla E Greenwood. Optimal signal in sensory neurons under an extended rate coding concept. *BioSystems*, 89(1):10–15, 2007.

[71] Régis Lengellé and Thierry Denoeux. Training mlps layer by layer using an objective function for internal representations. *Neural Networks*, 9(1):83–97, 1996.

[72] Hualiang Li and Tülay Adali. A class of complex ica algorithms based on the kurtosis cost function. *Neural Networks, IEEE Transactions on*, 19(3):408–420, 2008.

[73] Monica L Linden, Arnold J Heynen, Robert H Haslinger, and Mark F Bear. Thalamic activity that drives visual cortical plasticity. *Nature neuroscience*, 12(4):390–392, 2009.

[74] Mathias Linkerhand and Claudius Gros. Generating functionals for autonomous latching dynamics in attractor relict networks. *Scientific Reports (in press)*, 2013.

[75] Mathias Linkerhand and Claudius Gros. Self-organized stochastic tipping in slow-fast dynamical systems. *Mathematics and Mechanics of Complex Systems*, 1-2:129, 2013.

[76] John E Lisman. Bursts as a unit of neural information: making unreliable synapses reliable. *Trends in neurosciences*, 20(1):38–43, 1997.

[77] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[78] Dimitrije Marković and Claudius Gros. Self-organized chaos through polyhomeostatic optimization. *Physical Review Letters*, 105(6):068702, 2010.

[79] Dimitrije Marković and Claudius Gros. Intrinsic adaptation in autonomous recurrent neural networks. *Neural Computation*, 24(2):523–540, 2012.

[80] Mark L Mayer, Gary L Westbrook, and Peter B Guthrie. Voltage-dependent block by mg2+ of nmda responses in spinal cord neurones. 1984.

[81] Brian S Meldrum. Glutamate as a neurotransmitter in the brain: review of physiology and pathology. *The Journal of nutrition*, 130(4):1007S–1015S, 2000.

[82] Kenneth D Miller and David JC MacKay. The role of constraints in hebbian learning. *Neural Computation*, 6(1):100–126, 1994.

[83] Karyn M Myers and Michael Davis. Behavioral and neural analysis of extinction. *Neuron*, 36(4):567–584, 2002.

[84] Dorine Neveu and Robert S Zucker. Postsynaptic levels of [ca 2+] i needed to trigger ltd and ltp. *Neuron*, 16(3):619–629, 1996.

[85] Makoto Nishiyama, Kyonsoo Hong, Katsuhiko Mikoshiba, Mu-Ming Poo, and Kunio Kato. Calcium stores regulate the polarity and input specificity of synaptic modification. *Nature*, 408(6812):584–588, 2000.

[86] Erkki Oja. Principal components, minor components, and linear neural networks. *Neural Networks*, 5(6):927–935, 1992.

[87] Erkki Oja. The nonlinear pca learning rule in independent component analysis. *Neurocomputing*, 17(1):25–45, 1997.

[88] World Health Organization. *Neurological disorders: public health challenges*. World Health Organization, 2006.

[89] MA Paradiso. A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological cybernetics*, 58(1):35–49, 1988.

[90] Charles S Peskin. *Mathematical aspects of heart physiology*. Courant Institute of Mathematical Sciences, New York University, 1975.

[91] Jean-Pascal Pfister and Wulfram Gerstner. Triplets of spikes in a model of spike timing-dependent plasticity. *The Journal of neuroscience*, 26(38):9673–9682, 2006.

[92] Daniel Polani. Information: currency of life? *HFSP journal*, 3(5):307–316, 2009.

[93] Daniel Polani, Mikhail Prokopenko, and Larry S Yaeger. Information and self-organization of behavior. *Advances in Complex Systems*, 16(02n03), 2013.

[94] M. Prokopenko. Guided self-organization. *HFSP Journal*, 3:287–289, 2009.

[95] Gregory J Quirk and Devin Mueller. Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology*, 33(1):56–72, 2007.

[96] Marcel Reginatto. Derivation of the equations of nonrelativistic quantum mechanics using the principle of minimum fisher information. *Physical Review A*, 58:1775–1778, 1998.

[97] Jonathan E Rubin, Richard C Gerkin, Guo-Qiang Bi, and Carson C Chow. Calcium time course as a signal for spike-timing–dependent plasticity. *Journal of neurophysiology*, 93(5):2600–2613, 2005.

[98] Terence D Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural networks*, 2(6):459–473, 1989.

[99] HS Seung and H Sompolinsky. Simple models for reading neuronal population codes. *Proceedings of the National Academy of Sciences*, 90(22):10749–10753, 1993.

[100] Harel Z Shouval, Mark F Bear, and Leon N Cooper. A unified model of nmda receptor-dependent bidirectional synaptic plasticity. *Proceedings of the National Academy of Sciences*, 99(16):10831–10836, 2002.

[101] Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–1216, 2001.

[102] Fabian Sinz and Matthias Bethge. Temporal adaptation enhances efficient contrast gain control on natural images. *PLoS computational biology*, 9(1):e1002889, 2013.

[103] Per Jesper Sjöström, Gina G 2, and Sacha B Nelson. Neocortical ltd via coincident activation of presynaptic nmda and cannabinoid receptors. *Neuron*, 39(4):641–654, 2003.

[104] Per Jesper Sjöström, Gina G Turrigiano, and Sacha B Nelson. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32(6):1149–1164, 2001.

[105] Olaf Sporns and Max Lungarella. Evolving coordinated behavior by maximizing information structure. In *Artificial life X: proceedings of the tenth international conference on the simulation and synthesis of living systems*, pages 323–329, 2006.

[106] Martin Stemmler and Christof Koch. How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate. *Nature neuroscience*, 2(6):521–527, 1999.

[107] Joshua T Trachtenberg, Christopher Trepel, and Michael P Stryker. Rapid extragranular plasticity in the absence of thalamocortical plasticity in the developing primary visual cortex. *Science*, 287(5460):2029–2032, 2000.

[108] Jochen Triesch. Synergies between intrinsic and synaptic plasticity mechanisms. *Neural Computation*, 19(4):885–909, 2007.

[109] Gina G Turrigiano. The self-tuning neuron: synaptic scaling of excitatory synapses. *Cell*, 135(3):422–435, 2008.

[110] Gina G Turrigiano and Sacha B Nelson. Hebb and homeostasis in neuronal plasticity. *Current opinion in neurobiology*, 10(3):358–364, 2000.

[111] Takumi Uramoto and Hiroyuki Torikai. A calcium-based simple model of multiple spike interactions in spike-timing-dependent plasticity. *Neural computation*, 25(7):1853–1869, 2013.

[112] Raul Vicente, Michael Wibral, Michael Lindner, and Gordon Pipa. Transfer entropyâa model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience*, 30(1):45–67, 2011.

[113] Tim P Vogels and Larry F Abbott. Signal propagation and logic gating in networks of integrate-and-fire neurons. *The Journal of neuroscience*, 25(46):10786–10795, 2005.

[114] Christoph Von Der Malsburg. Binding in models of perception and brain function. *Current opinion in neurobiology*, 5(4):520–526, 1995.

[115] Huai-Xing Wang, Richard C Gerkin, David W Nauen, and Guo-Qiang Bi. Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nature neuroscience*, 8(2):187–193, 2005.

[116] Laurenz Wiskott. Slow feature analysis: A theoretical analysis of optimal free responses. *Neural Computation*, 15(9):2147–2177, 2003.

[117] Laurenz Wiskott and Terrence J Sejnowski. Slow feature analysis: Unsupervised learning of invariances. *Neural Computation*, 14(4):715–770, 2002.

[118] Shao-Nian Yang, Yun-Gui Tang, and Robert S Zucker. Selective induction of ltp and ltd by postsynaptic [ca2+] i elevation. *Journal of neurophysiology*, 81(2):781–787, 1999.

Acknowledgements

Firstly, I would like to thank Prof. Claudius Gros for his guidance and trust. Also, many thanks to my office colleagues, for valuable discussions throughout these years: Guillermo Ludueña, Dimitrije Marcović, Bulcsú Sándor, Hendrik Wernecke, and Samuel Eckmann. A second thanks goes to Bulcsú for his input on gradient systems.

A huge thanks to Ms. Kolokotsa for the uncountable times she has helped me during these years.

I would also like to acknowledge Robert C. Froemke and Yang Dan for the experimental data from cortical visual neurons, used for fitting and comparison of the STDP model. Also many thanks to Máté Lengyel for interesting discussions on the properties of the self-limiting learning rule for other transfer functions.

A big thanks goes to Kira Riedl, Prof. Claudius Gros, and (again) to Ms. Kolokotsa, for their help with the German version of the abstract in the thesis.

Finally, I would like to thank the support of the German Science Foundation (DFG) and the German Academic Exchange Service (DAAD).

Rodrigo Echeveste

curriculum vitae

Address:	*Altenhöferalle 30,*
	60438 Frankfurt am Main,
	Germany
Birth Place:	*Rosario, Argentina*
Birth Date:	*20th January, 1987*
Nationality:	*Argentine*
Marital Status:	*Single*
E-mail:	*echeveste@itp.uni-frankfurt.de*
Phone number:	*+49-151-63400739*
Google Scholar:	*http://scholar.google.de/citations?user=AcT16qoAAAAJ&hl=en&oi=ao*

EDUCATION

PhD student in Physics. Topic: Plasticity Mechanisms in Neural Networks driven by Generating Functions. Supervisor: Prof. Dr. Claudius Gros. Institut für Theoretische Physik, Goethe Universität, Frankfurt am Main, Germany. (October 2012- present date)

Degree: "Magister en Ciencias Físicas" (Master's Degree Program in Physics). Instituto Balseiro, Universidad Nacional de Cuyo, Argentina, (January 2011 – December 2011). Thesis title: "Sensory Perception in Autistic Children". Supervisor: Dr. Inés Samengo. GPA: 9.71/10, obtaining the *Prize for best student in Physics of Instituto Balseiro* (December 2011).

Degree: "Licenciado en Física" (equivalent to a 5 year degree course in Physics). Instituto Balseiro, Universidad Nacional de Cuyo, Argentina, (July 2008 – December 2010). Thesis title: "Sensory Perception in Autistic Children". Supervisor: Dr. Inés Samengo. GPA: 8.4/10.

Three years of "Licenciatura en Física" (First three years* of a 5 year degree course in Physics), Universidad Nacional de Rosario, Argentina, (2005 – 2008). GPA: 9.5/10.

* *Instituto Balseiro demands all of it's students to complete at least 2 years in a related-field Undergraduate Course in another University before entering the Institute.*

LANGUAGES

Spanish (mother tongue), **English** (advanced), **French** (advanced), **German** (intermediate).

SCHOLARSHIPS

Research grant for Doctoral candidates from the German Academic Exchange Service (DAAD) with the goal of obtaining a PhD in Physics at the Goethe Universität, Frankfurt, Germany (May 2014 – present date).

Full Scholarship from "Fundación YPF" to pursue Bachelor and then Master studies at Instituto Balseiro, Universidad Nacional de Cuyo, Argentina (August 2008 – December 2011).

AWARDS

Prize for best student in Physics of Instituto Balseiro awarded by the Bariloche Chapter of the Argentine Association of Physics (AFA). This prize is awarded to the student with the best general average for the combined Bachelor and Master's Degree Studies in Physical Science at Instituto Balseiro (December 2011).

TRAVEL GRANTS

Competitive Travel Grants for Advanced PhD Students to present their work in the 2016 ELSC Annual Retreat. Kibbutz Ein Gedi, Israel (January 2016).

Mentorship Travel Grant Award to attend the Computational and Systems Neuroscience (COSYNE) meeting 2015. Salt Lake City, USA (March 2015).

RESEARCH POSITIONS

Part-time Research Assistant (Wissenschaftlicher Hilfskraft) at the Institut für Theoretische Physik, Goethe Universität, Frankfurt am Main, Germany (May 2014 - present date).

Research Assistant (Wissenschaftlicher Mitarbeiter) at the Institut für Theoretische Physik, Goethe Universität, Frankfurt am Main, Germany (October 2012- April 2014).

TEACHING EXPERIENCE

Tutor at the Goethe University of Frankfurt, Germany, for the subjects: *Introduction to Programming for Physicists* (October 2014 – February 2015), *Self-Organization: Theory and Simulations* (April 2014 – August 2014), *Electrodynamics* (October 2013 – February 2014, and October 2015 – present date), *Complex and Adaptive Dynamical Systems* (April 2013 – August 2013), *Programmierpraktikum (Java Programming Course)* (October 2012 – February 2013).

Ad Honorem Teaching Assistant at Instituto Balseiro, Universidad Nacional de Cuyo, Argentina for the course *Thermodynamics* (January 2011 – July 2011).

Teacher of the subject *Physics* at the Levelling Course for the Admission to the "Tecnicaturas Universitarias" (University Technician's Course), Instituto Politécnico Superior "General San Martín", Universidad Nacional de Rosario, Argentina (2007-2008).

Teaching Assistant at the Preparation Courses for the Physics Olympiads for Secondary School Students, Instituto Politécnico Superior "General San Martín", Universidad Nacional de Rosario, Argentina (2005-2007).

ELECTED POSITIONS

Member of the Academic Senate of Instituto Balseiro, elected by the body of Undergraduate Students (September 2009 – September 2010).

PUBLICATIONS

ARTICLES IN JOURNALS

Echeveste, R., Eckmann, S., & Gros, C. *The Fisher Information as a Neural Guiding Principle for Independent Component Analysis*. **Entropy** (2015), *17*(6), 3838-3856; doi:10.3390/e17063838.

Echeveste, R., & Gros, C. *Two-trace model for spike-timing dependent synaptic plasticity.* **Neural Computation** (2015), 27 (3), 672-698. doi:10.1162/NECO_a_00707

Echeveste, R., & Gros, C. *Generating functionals for computational intelligence: the Fisher information as an objective function for self-limiting Hebbian learning rules.* **Frontiers in Robotics and AI** (2014), 1:1. doi: 10.3389/frobt.2014.00001

ARTICLES IN PROCEEDINGS

Echeveste, R., & Gros, C. A*n objective function for self-limiting neural plasticity rules.* **ESANN 2015 Proceedings** (2015), ISBN 978-287587014-8.

EXTENDED ABSTRACTS

Gros, C., & Echeveste, R. *The Fisher information as a guiding principle for self-organizing processes.* **Workshop on Information Theoretic Incentives for Artificial Life** (2014). p.5

THESIS

Echeveste, R. Supervisor: Dr. Inés Samengo. Master's Thesis: *Sensory Perception in Autistic Children*. Instituto Balseiro, Universidad Nacional de Cuyo, Argentina (2011). Link (in Spanish with an English abstract): http://ricabib.cab.cnea.gov.ar/316/1/1Echeveste.pdf

POSTER AND ORAL PRESENTATIONS

TALKS

"Self-stabilizing Plasticity Rules derived from the Stationarity Principle of Statistical Learning" . ELSC Retreat. Kibbutz Ein Gedi, Israel (January 2016).

"Complementary approaches to Computational Neuroscience: Objective Functions and Biophysics". Condensed Matter Theory Seminar, ITP, Frankfurt University. Frankfurt, Germany (November 2015).

"Complementary approaches to Synaptic Plasticity: Objective Functions and Biophysics". Cambridge University, Engineering Department, CBL. Cambridge, UK (October 2015).

"Complementary approaches to Synaptic Plasticity: Objective Functions and Biophysics". NeuroBioTheory seminar, Frankfurt Institute for Advanced Studies (FIAS). Frankfurt, Germany (January 2015).

"Asymmetric two-trace model for STDP". DPG-Frühjahrstagung (German Physical Society Meeting) 2014, Condensed Matter Section. Dresden, Germany (April 2014).

"Física y Autismo: El Rol de la Física en Problemas Tradicionalmente Reservados a Otras Disciplinas". Asociación Latina de Programación Neurolingüística y Tecnologías Afines (A.La.P.N.L.). Paraná, Entre Ríos, Argentina, (July 2011).

POSTERS

Computational Neuroscience Society (CNS) meeting, “A simple effective model for STDP: from spike pairs and triplets to rate- encoding plasticity” Rodrigo Echeveste, and Claudius Gros. Prague, Czech Republic (July 2015).

Computational Neuroscience Society (CNS) meeting, “Should Hebbian learning be selective for negative excess kurtosis?” Claudius Gros, Samuel Eckmann, and Rodrigo Echeveste. Prague, Czech Republic (July 2015).

EITN Workshop on Learning and Plasticity, “An Objective Function for Hebbian self-stabilizing Plasticity Rules.”, Rodrigo Echeveste, Samuel Eckmann, and Claudius Gros. Paris, France (June 2015).

Osnabrück Computational Cognition Alliance Meeting (Occam) 2015, “From Stationarity to ICA: an Objective Function for Hebbian self-stabilizing Plasticity Rules.”, Rodrigo Echeveste, Samuel Eckmann, and Claudius Gros. Osnabrück, Germany (May 2015).

European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) 2015, “An objective function for self-limiting neural plasticity rules.” Rodrigo Echeveste and Claudius Gros. Bruges, Belgium (April 2015).

DPG-Frühjahrstagung (German Physical Society Meeting) 2015, Condensed Matter Section, “An objective function for Hebbian self-stabilizing neural plasticity rules”, Rodrigo Echeveste and Claudius Gros. Dresden, Germany (March 2015).

Computational and Systems Neuroscience (COSYNE) meeting 2015, “Deducing Hebbian Adaption Rules from the Stationarity Principle of Statistical Learning”, Claudius Gros and Rodrigo Echeveste. Salt Lake City, USA (March 2015).

Winter School in Quantitative Systems Biology. Topic: Systems Neuroscience. Abdus Salam International Centre for Theoretical Physics (ICTP), “Two-trace model for STDP” Rodrigo Echeveste and Claudius Gros. Trieste, Italy (December 2014).

ESI Systems Neuroscience Conference (ESI-SyNC) 2014, “Learning in Neural Models driven by Objective Functions”, Rodrigo Echeveste and Claudius Gros. Frankfurt, Germany (July 2014).

Osnabrück Computational Cognition Alliance Meeting (Occam) 2014, “Two-trace model for STDP”, Rodrigo Echeveste and Claudius Gros. Osnabrück, Germany (May 2014).

DPG-Frühjahrstagung (German Physical Society Meeting) 2014, Condensed Matter Section, “Self-stabilizing Learning Rules in Neural Models driven by Objective Functions”, Rodrigo Echeveste and Claudius Gros. Dresden, Germany (April 2014).

XXVIII Congreso de la Sociedad Argentina de Investigación en Neurociencias (SAN), “Visual-memory strategies employed by children in the autistic spectrum”, Melisa Maidana Capitán, Rodrigo Echeveste, Inés Samengo. Huerta Grande, Argentina (September 2013).

Bernstein Conference 2013, “Self-stabilizing Learning Rules in Neural Models driven by Objective Functions”, Rodrigo Echeveste and Claudius Gros. Tübingen, Germany (September 2013).

Osnabrück Computational Cognition Alliance Meeting (Occam) 2013, “Learning in Neural Models driven by Objective Functions”, Rodrigo Echeveste and Claudius Gros. Osnabrück, Germany (May 2013).

Taller Regional de Física Estadística y Aplicaciones a la Materia Condensada (Trefemac), “Estrategias de memoria visual en sujetos con diagnóstico del espectro autista”, Melisa Maidana Capitán, Rodrigo Echeveste and Inés Samengo . La Plata, Argentina (May 2013).

Third EUCogIII Members Conference, “Learning in Neural Models driven by Objective Functions”, Rodrigo Echeveste and Claudius Gros. Palma de Mallorca, Spain (April 2013).

97a Reunion Nacional de la Asociacion Física Argentina (AFA), “Estrategias de memoria visual en sujetos con diagnóstico del espectro autista”, Melisa Maidana Capitán, Rodrigo Echeveste and Inés Samengo. Villa Carlos Paz, Córdoba, Argentina (September 2012).

XXVI Reunión Anual de la Sociedad Argentina de Investigación en Neurociencias (SAN), “Sensory Stimulus Categorization in Autistic Children”, Rodrigo Echeveste and Inés Samengo. Huerta Grande, Argentina (October 2011).

II Reunión Conjunta de la Asociación de Física Argentina y la Sociedad Uruguaya de Física (AFA-SUF), “Categorización de Estímulos Sensoriales en Niños Autistas”, Rodrigo Echeveste and Inés Samengo. Montevideo, Uruguay (September 2011).

ATTENDED COURSES, SCHOOLS, RETREATS, AND CONFERENCES

ADDITIONAL UNIVERSITY COURSES

Machine Learning Course from the Computational Science Degree Course at Universidad Nacional de Rosario, Argentina (March - July 2012) Grade obtained: 10 out of 10.

SCHOOLS

INCF Summer School: Information Processing in Neural Systems: From Single Neurons to Large-Scale Models of Cognition, Osnabrück, Germany (May 2015).

Winter School in Quantitative Systems Biology. Topic: Systems Neuroscience. Abdus Salam International Centre for Theoretical Physics (ICTP), Trieste, Italy (December 2014).

Winter School “Escuela de Ciencias Informáticas (ECI) 2010” del Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina (July 2010). Courses taken: Reinforcement Learning, Natural Language Generation.

CONFERENCES, RETREATS, WORKSHOPS, AND MEETINGS

ELSC Retreat. Kibbutz Ein Gedi, Israel (January 2016).

Computational Neuroscience Society (CNS) meeting. Prague, Czech Republic (July 2015).

EITN Workshop on Learning and Plasticity, Paris, France (June 2015).

Osnabrück Computational Cognition Alliance Meeting (Occam) 2015, Osnabrück, Germany (May 2015).

European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN) 2015, Bruges, Belgium (April 2015).

DPG-Frühjahrstagung (German Physical Society Meeting) 2015, Condensed Matter Section, Dresden, Germany (March 2015).

Computational and Systems Neuroscience (COSYNE) meeting 2015, Salt Lake City, USA (March 2015).

ESI Systems Neuroscience Conference (ESI-SyNC) 2014, Frankfurt, Germany (July 2014).

Osnabrück Computational Cognition Alliance Meeting (Occam) 2014, Osnabrück, Germany (May 2014).

DPG-Frühjahrstagung (German Physical Society Meeting) 2014, Condensed Matter Section, Dresden, Germany (April 2014).

INS conference: “The Dynamic Brain”, Marseille, France (November 2013).

Bernstein Conference 2013, Tübingen, Germany (September 2013).

Osnabrück Computational Cognition Alliance Meeting (Occam) 2013, Osnabrück, Germany (May 2013).

Third EUCogIII Members Conference, Palma de Mallorca, Spain (April 2013).

XXVI Reunión Anual de la SAN, Huerta Grande, Córdoba, Argentina (October 2011).

II Reunión Conjunta de la Asociación de Física Argentina y la Sociedad Uruguaya de Física (AFA-SUF), Montevideo, Uruguay (September 2011).

Segunda Reunión Conjunta de Neurociencias (IIRCN): XXV Reunión Anual de la Sociedad Argentina de Investigación en Neurociencias (SAN) y XII Taller Argentino de Neurociencias (TAN), Huerta Grande, Córdoba, Argentina (October 2010).

Course "Physics and Neuroscience: heading towards quantitative biology", Huerta Grande, Córdoba, Argentina (October 2010).

95° Reunión de la Asociación de Física Argentina (AFA), Malargüe, Mendoza, Argentina (September 2010).

"XXIV International Conference on Photonic, Electronic and Atomic Collisions", Rosario, Santa Fe, Argentina (July 2005).

Frankfurt am Main, Germany, March 2016.