

Title	Dynamics of adaptive recurrent neural networks
Authors	Fox, David
Publication date	2023
Original Citation	Fox, D. 2023. Dynamics of adaptive recurrent neural networks. MSc Thesis, University College Cork.
Type of publication	Masters thesis (Research)
Rights	© 2023, David Fox. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-10-30 07:12:25
Item downloaded from	https://hdl.handle.net/10468/16503

Dynamics of Adaptive Recurrent Neural Networks

David Fox, BSc

Thesis submitted for the degree of
Master of Science



University College Cork
School of Mathematical Sciences
Discipline: Applied Mathematics

Supervisor: Dr. Andreas Amann

Advisor: Dr. Andrew Keane

Head of School: Dr. Kevin Hayes

Head of Discipline: Prof. Sebastian Wieczorek

September 2023

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

David Fox

September 2023

Acknowledgements

I would like to express my sincerest gratitude to my supervisor Andreas Amann, for igniting and maintaining my interest in the study of applied mathematics over the last five years. He has always been incredibly generous with his time, and has given me the guidance and encouragement I needed to get to this point in my studies.

I also wish to thank Andrew Keane, for acting as my advisor, and providing mentorship in the early stages of my transition from undergraduate to postgraduate research.

I would also like to thank the school of Mathematical Sciences at UCC, who graciously paid the fees for this MSc. Particularly Kevin Hayes and David O’Leary who provided me with so many amazing opportunities throughout my time at UCC. Thank you also to Tom Carroll for being such an excellent mentor and inspiring teacher.

Getting this far in my education would have been impossible without the support of many people who I am so grateful to have in my life; my parents, without the support of which I would have never been able to pursue my passion for mathematics, my siblings Catherine and Stephen, my lifelong friends Eamonn, Éoin and Rob, and the members of the journal club: Jordan, Ruaraí, Frank, Hassan, George and Andrew.

Finally, thank you to my partner Erin, who completes my life so perfectly. I’ve cherished every second of the last four years we have spent together and look forward to many more in future.

Abstract

In this thesis a simple, phenomenological model of a neural network with plasticity is presented in the form of a slow-fast adaptive dynamical recurrent neural network. The plasticity rule is chosen from the class of Hebbian learning rules, in which the synaptic connection between two neurons evolves continuously as a function of their correlation in the recent past. Initially an analysis of networks of two neurons is presented, which exhibit relaxation oscillations in which one neuron switches between an 'off' state, where it takes a negative value, and an 'on' state, where it takes a positive value, while the other neuron stays in one on/off state. Then, by means of an example with a nine neuron network, the system is shown to exhibit both stable frequency cluster synchronization and transient frequency cluster synchronization.

Contents

1	Introduction	11
1.1	Neuroscience Background	13
1.1.1	Neuronal Structure	13
1.1.2	Action Potentials	14
1.1.3	Connectivity Structure	14
1.1.4	Plasticity	15
1.1.5	Brain Inspired Models	15
1.2	Model Summary	16
2	Model	19
2.1	Single Neuron Dynamics	19
2.1.1	Bifurcations when varying m for fixed I	20
2.1.2	Bifurcations when varying I for fixed m	20
2.1.3	Time Varying Inputs	25
2.2	N dimensional CTRNNs	29
2.3	Learning Rule	30
2.3.1	The Learning Rate η	31
2.3.2	Hebbian and Anti-Hebbian Learning	32
2.3.3	Pairwise Interactions	32
2.4	Timescale Separation	33
3	Networks of Two Neurons	37
3.1	Hebbian Synapses	37
3.1.1	Symmetries and Equivariance	38
3.1.2	Fast Subsystem	39
3.1.3	Critical Manifold	41
3.2	Anti-Hebbian Synapses	42
3.3	H and AH Synapses	45
3.3.1	Symmetries and Equivariance	48
3.3.2	Fast Subsystem	49

3.3.3 Full System Dynamics	51
4 Cluster Synchronization	57
5 Conclusion	63

Chapter 1

Introduction

The brain is an incredibly powerful information processing system, which carries out extremely complex tasks in parallel while fulfilling its role as our bodies' control centre, allowing us to adapt to and navigate the world [1]. For example, while reading this text, your brain is carrying out visual processing of individual letters and words, accessing long term-memory to comprehend their meaning and storing recently read words in your short-term working memory, enabling you to combine them and understand the sentence as a whole [2]. Your brain is also responsible for coordinating conscious movement, reflexes, breathing and appetite, among other functions necessary to sustain life [1]. Our brains are not just capable of carrying out such tasks, but also learning to carry out new ones through experience [3]. A major feature of the brain which differentiates it from other computational structures, such as a CPU in a computer, is that these capabilities are not explicitly programmed into it. Rather, the complex behaviours of the nervous system emerge from comparatively simple, local interactions between its basic elements, the neurons [4]. Neurons are specialised cells of the nervous system, which pass electrical signals throughout complex neural networks via connections known as synapses [5]. Another major difference between the brain and traditional computers is that the brain can learn to carry out new tasks and store memories without being explicitly programmed to do so. This ability is widely attributed to neural plasticity, which is the brain's ability to alter its activity in response to intrinsic or extrinsic stimuli by the reorganization of its structure [6].

Given these impressive emergent computational abilities, one question immediately comes to mind: can we use our knowledge of neuroscience as a basis for the design of models and algorithms, which can mimic in some sense, the workings of the brain? This question has been a major driving force behind machine learning (ML) research, which uses computational models called artificial neural networks (ANNs) inspired

by brain structure and functionality to design algorithms for complex tasks such as image recognition, natural language processing and time series prediction [7]. The class of ANNs is very broad, but they typically consist of neuron like dynamical units which are coupled to form a (possibly complex) directed, weighted network topology. This network topology is sometimes capable of changing in accordance with some pre-specified rule, in analogy with neural plasticity [8].

Another field that takes inspiration from real neural systems is neuromorphic computing (NC), which seeks to design physical computing devices whose functionality and structure is inspired by the brain. This is in contrast with the vast majority of computers today which are based on the biologically implausible von Neumann architecture, in which a stored list of instructions is executed sequentially by a central processing unit, storing the results of computations in a memory location [9]. The overarching goal of building biologically inspired devices is to design a computer capable of carrying out complex calculations in a way which is more time and energy efficient than can be done with a von Neumann architecture. Additionally, neuromorphic architectures appear to be the more natural choice for implementing similarly brain-inspired ML algorithms in physical systems [10].

Closely related to ML and NC is the field of computational neuroscience (CN), which attempts to better understand the function of the nervous system by the construction and analysis of models of the various systems found in the nervous system [11]. The scale and detail included in these models is quite variable. On one end of the spectrum, there are detailed electro-physiological models of individual neurons which account for dynamics of molecules, such as the Hodgkin-Huxley equations [12]. On the other, there are models of entire subsystems of the brain based on ANNs, which account for the interaction of neurons through a complex connectivity structure, while suppressing many other biological details which may be included in single cell models. Such models based on ANNs have been used to gain a better understanding of various features of the nervous system that arise as emergent properties [11].

Given their importance to the fields of CN, NC and ML, the mathematical study of ANNs has gained significant popularity over the last several decades. In particular, dynamical systems (DS) theory has been used to study the properties of ANNs [13]. Much of this study has considered ANNs whose underlying network structure is considered fixed, or only changes as part of an offline learning rule. However, this does not accurately reflect the continuously evolving nature of plasticity observed in real neural systems. To date, the majority of ANNs which do account for plasticity as a process which co-evolves with neurons do so via comparatively complicated spike timing dependent plasticity rules. While such biologically plausible systems are certainly useful objects of study, it would still be beneficial to develop simpler models of

ANNs with dynamic plasticity for a number of reasons. Firstly, finding a minimalistic model capable of replicating certain features of brain dynamics is instructive from a neuroscientific perspective as it can help elucidate the crucial conditions necessary to generate such emergent properties. Secondly, improving our understanding of such an ANN architecture has the potential to motivate new machine learning algorithms. Lastly, a simplified DS model capable of replicating features of biological neural networks (BNNs) can act as a template for implementation of NC devices. To this end, in this thesis we present a model of an ANN with plasticity as an adaptive dynamical network [14], by considering the synaptic connections as dynamical variables in their own right, which co-evolve with the state of the neurons according to an ordinary differential equation (ODE). Our main interest will be investigating whether such a model can exhibit features which are observed in BNNs and what is the mechanism by which they do so? Investigating such fundamental properties is an important first step in using such a model as a basis for ML algorithms, NC devices or potentially understanding BNNs with plasticity.

The remainder of this chapter will be spent motivating and presenting the model used in the remainder of this thesis. Initially, some biological details of the nervous system are explained in order to fully appreciate how the model reflects a real neural system, and how it differs from it. This information can be found in any introductory biology textbook, such as [15][5]. Next we give a description of the underlying ANN system which is used, known as a recurrent neural network (RNN). Modelling plasticity in ANNs is then discussed and a method of augmenting the RNN to account for plasticity is presented.

1.1 Neuroscience Background

1.1.1 Neuronal Structure

The fundamental components of the nervous system are cells which specialize in information transmission, called neurons. This special role they fulfill is reflected in some structural differences they have compared to other types of somatic cells. It should be noted that there is a vast array of types of neurons, with different physical forms and features reflecting their specific purpose. The neuron described here is a 'generic' neuron which possesses the fundamental features common to most neurons. The generic neuron most closely resembles a pyramidal cell, a type of neuron found in the neocortex. A unique feature of neurons which is important for their function is their shape. Neurons have a cell body, called a soma, which houses the nucleus and other organelles such as ribosomes and mitochondria. Extending from one end of the soma are root like extensions called dendrites, which receive signals from

other neurons. Extending from the other end is the axon, a single cord coming from the soma and terminating at multiple distinct sites called axon terminals, where signals are passed to other neurons. An axon terminal of one neuron connects to a dendrite of another at a site called a synapse. At a synapse, the neuron with an axon terminal involved in the connection is called the presynaptic neuron, the other neuron is called the postsynaptic neuron.

1.1.2 Action Potentials

There are many active and passive channels embedded in the membrane of a neuron which facilitate influx and efflux of ions, thereby controlling its membrane potential. In the absence of an external electrical input, the membrane potential of a neuron rests in a stable equilibrium state at -70mV . If the neuron receives an external electrical input which increases its membrane potential to less than -55mV , the membrane potential will gradually decrease back -70mV . However, if the signal is large enough to increase the membrane potential to at least -55mV , the neuron generates an action potential. An action potential is an increase in membrane voltage up to 90mV which propagates along the axon, from the soma to the axon terminals. Once this change in voltage has reached the axon terminals, it causes neurotransmitters to be released into the synapse, which then diffuse across the synaptic cleft to bind to receptor sites embedded in the dendrite of the postsynaptic neuron. This causes ion channels to open in the cell membrane of the postsynaptic neuron, leading to an influx of current which changes the neuron's membrane potential. Thus, action potentials or spikes are the principle means by which neurons can send information to each other. Notably, this flow of information is unidirectional, always going from presynaptic to postsynaptic neuron.

1.1.3 Connectivity Structure

Cognitive functions such as memory or coordination of movement are not carried out by individual neurons, but are an emergent property of complex networks of neurons which are specialised to carry out such tasks [11]. Extensive synaptic connectivity is a pervasive feature of biological neural networks. The number of neurons in the human central nervous system is on the order of 10^{11} , with many neurons being connected to as many as 10^4 other neurons via synaptic connections [16]. Consider the human brain's outermost layer, the neocortex as an example. Neurons of the neocortex are organised into 6 vertical layers of columns. It has been suggested that these cortical columns are basic functional units for information processing in the brain. Across the cortex, there are numerous, repeated stereotypical patterns of connections within neurons in a column and between neurons in different columns

[17]. Indeed, a neuron in the neocortex will typically be connected to thousands of presynaptic neurons. [11].

1.1.4 Plasticity

Plasticity refers to the brain's ability to alter its structure. There are numerous types of plasticity mechanisms by which the brain can reorganize itself. An important form of plasticity, due to its relevance to learning and memory, is synaptic plasticity. Synaptic plasticity refers to the breaking or alteration of strength of pre-existing synapses or the formation of new synapses. The first guiding principle for synaptic plasticity in the brain was proposed by Donald Hebb in 1949 [18]. Hebb's postulate states that synaptic connections in the nervous system form dynamically in response to external stimuli such that the connection between a pre-synaptic neuron and a post-synaptic neuron is strengthened if the pre-synaptic neuron firing tends to also closely precede the post-synaptic neuron firing. This mechanism is summarised by the popular phrase "neurons that fire together, wire together". Today this famous postulate is often rephrased, to state that modifications in the synaptic transmission efficacy are driven by correlations in the firing activity of the pre- and postsynaptic neurons [19].

1.1.5 Brain Inspired Models

The combination of these three features of neural systems; signal transmission between cells, complex network structure and the ability to alter this structure via plasticity, are crucial requirements for enabling the nervous system to learn to carry out complex tasks such as language comprehension, coordination of motion and the formation of memories [6] [20]. Thus, a better understanding of the role these features play in the function and capabilities of biological neural networks is an essential part of developing the theory of brain function. These features are also the motivations behind the design of ANNs used in ML as well as circuits designed in NC. The big difference between ANNs used in the context of CN and those used in NC and ML is how realistically the features of the ANN reflect the reality of a BNN. In CN, the principal goal is to build models of neural systems to better understand them and to motivate new experiments in lab settings, as such, biological plausibility is quite important. In contrast to CN, the goal of ML and NC is to build algorithms or devices which can be used to solve a particular task, using the structure and function of the brain as inspiration. Because of this lack of necessity for biological plausibility, ANNs used in ML and NC often include features which are drastically biologically implausible, if the inclusion of these features leads to a performance advantage or simplifies the resulting model, algorithm or device. The

ANN model presented next, known as a continuous time recurrent neural network (CTRNN) [21], is an example of this type of ANN studied often in the context of ML.

1.2 Model Summary

The CTRNN consists of $N \in \mathbb{N}$ dynamical units $\{r_i\}_{i=1}^N$ representing neurons, whose individual dynamics in the absence of coupling are given by

$$\epsilon \dot{r}_i = -r_i + \tanh(mr_i + I_i). \quad (1.1)$$

The variable r_i represents the activity of the i^{th} neuron. With this choice of local dynamics given by the *tanh* activation function, r_i will converge to one of two fixed points based on the value of the input parameter $I_i \in \mathbb{R}$, provided that the parameter $m > 1$. These fixed points represent the neuron being in either a firing state or a non-firing state. The parameter $\epsilon > 0$ controls the timescale of the neuron dynamics. A detailed analysis establishing this behaviour is conducted in chapter 2.

Interaction between neurons is introduced by giving each neuron a weighted sum of the other neuron states as input, so the equation for the CTRNN is given by

$$\epsilon \dot{r}_i = -r_i + \tanh\left(mr_i + \sum_{j \neq i} m_{ij}r_j + I_i\right). \quad (1.2)$$

This choice of coupling is inspired by synaptic transmission of signals between neurons. A post-synaptic neuron receives input from a pre-synaptic neuron when the pre-synaptic neuron spikes, causing released neurotransmitters to allow an influx of current at a receptor site, thus altering the membrane potential of the post-synaptic neuron. The resulting change in state of the post-synaptic neuron as a result of activity in the pre-synaptic neuron depends on the strength of the synapse connecting them. To reflect this, each neuron state r_j is multiplied by a synaptic weight m_{ij} in the equation for $\dot{r}_i$. We assume no nonlinear interactions between the individual synaptic inputs to a neuron, so the total synaptic input into r_i is simply the sum of individual synaptic inputs $\sum_{j \neq i}^N m_{ij}r_j$. In general we can allow for an external input into each neuron, which for example, could be from neurons which are not explicitly included in the model whose activity is specified at the outset. The parameter I_i represents such an input to neuron r_i .

With such a model, which is inspired by the structure and function of a biological NN, it is important to also state ways that it differs from biological NNs. One

major difference is that the individual units in this model, which are the analogues of neurons, have much simpler dynamics than actual neurons, whose complex dynamics can be more accurately modelled by the Hodgkin-Huxley equations for example [22]. Another key difference is that a neuron effects the state of another neuron only when it spikes, as opposed to this model, where a neuron has a continuous influence on any neuron it is connected to. Even with these differences, a model such as (1.2) is still a useful object to study as it is a particularly simple model of a NN with the type of structure and information transmission which occurs in biological NNs.

CTRNNs such as (1.2), which do not consider the connectivity structure of the network as being dynamic have been used extensively in computational neuroscience to model biological NNs. They are also the basis for many machine learning methods for sequence tasks such as reservoir computers [23]. However, it is still important to acknowledge that in reality, biological NN connectivity structure given by synapses is not fixed, but changes dynamically in response to stimulus via plasticity. Therefore, in order to have the RNN model more accurately reflect a biological NN, we must augment the RNN model to include plasticity.

Choosing how to include plasticity into a NN model is a task which requires careful consideration as there are many possible ways to proceed. The approach taken in this thesis is to model a NN with plasticity as a so called adaptive dynamical network, which can be done by augmenting equation (1.2) to include dynamic equations for the connection weights m_{ij} ,

$$\begin{aligned} \epsilon \dot{r}_i &= -r_i + \tanh \left(m r_i + \sum_{j \neq i} m_{ij} r_j + I_i \right), \\ \dot{m}_{ij} &= F(\mathbf{r}, \mathbf{M}), \quad i \neq j \end{aligned}$$

where $F(\mathbf{r}, \mathbf{M})$ is some suitably chosen scalar function, which we refer to as the learning rule, the vector $\mathbf{M}$ consists of the m_{ij} values and the vector $\mathbf{r}$ consists of the r_i values. Our goal when choosing how to model dynamic plasticity in such a network will be to create a simple, phenomenological model which captures certain features of synaptic neural plasticity and discards others. The learning rule we choose will be from the class of Hebbian learning rules[19], which share the characteristic that the synaptic connection strength between two neurons evolves according to the correlation of their activity within the recent past [6]. The learning rule is given by

$$\dot{m}_{ij} = \alpha_{ij}\eta r_i r_j - m_{ij}, \quad (1.3)$$

where $\eta \in \mathbb{R}^+$ is a parameter called the learning rate and $\alpha_{ij} = \pm 1$ are parameters which determine whether a synapse is Hebbian ($\alpha_{ij} = 1$) or anti-Hebbian ($\alpha_{ij} = -1$). This is an example of a local learning rule, as the change in strength of a synaptic weight is based only on information available at the synapse, such as the current synaptic strength, and the activity of the pre- and post-synaptic neurons [19].

Now that the ODE for link dynamics has been chosen, we can present the equation for the adaptive CTRNN (ARNN)

$$\begin{aligned} \epsilon \dot{r}_i &= -r_i + \tanh \left(m r_i + \sum_{j \neq i} m_{ij} r_j + I_i \right), \\ \dot{m}_{ij} &= \alpha_{ij} \eta r_i r_j - m_{ij}, \quad i \neq j. \end{aligned}$$

It is desirable to have the synaptic weights m_{ij} evolve on a slower timescale than the neurons r_i , so that when the synapse updates based on correlated activity, it is doing so over a time period of sufficient length. Mathematically, this can be introduced into the model by assuming $0 < \epsilon \ll 1$, making the ARNN a fast-slow system [14].

Despite the emphasis on simplicity when constructing this model, it is still a very difficult object to study from an analytical perspective. Typically, a first step towards better understanding such a model's behaviour is analysing certain network motifs of small size. Insights gained for these simpler examples can then sometimes be used to explain certain behaviour for higher dimensions. To this end, the remainder of this thesis is structured as follows; chapter two provides a more detailed analysis of the elements of the model, chapter three then studies examples of networks with two neurons, finally chapter four will give numerical results for networks of higher dimension.

Chapter 2

Model

The model considered in this thesis consists of individual dynamical units representing simplified neurons which are coupled together in a way that is permitted to change, in analogy with synaptic plasticity. In order to describe the model we need to specify three key ingredients, the individual neuron dynamics, the nature in which they are coupled and the way this coupling is permitted to change based on the state of the neurons. As such, this chapter is divided into three sections describing each of these ingredients. The first section introduces the dynamics of an individual neuron, which is a one dimensional system whose behaviour can be interpreted as a type of dynamic switch that is either in an "off" state where the neuron is not firing or an "on" state where the neuron is firing. Which state the neuron is in depends on the value of a time varying input. Next, we describe how the dynamical network is formed by taking $N \in \mathbb{N}$ identical neurons and augmenting each of their dynamical equations so that the input to each neuron is a weighted sum of the other neuron states. Through this coupling, the switching behaviour of each neuron is dependent on that of the other neurons. Finally, we introduce the mechanism by which this coupling adapts to the neuron dynamics, which is an example of a Hebbian learning rule, in which each weight of the network evolves based on the correlation of the activity in the recent past of the neurons connects [19].

2.1 Single Neuron Dynamics

In the absence of coupling, the dynamics of each neuron in the network is given by

$$\epsilon \dot{r} = -r + \tanh(mr + I), \quad (2.1)$$

where $r \in \mathbb{R}$ gives the node state and $\gamma, m \in \mathbb{R}^+$ and $I \in \mathbb{R}$ are parameters. I

represents some external input to the neuron which can be dependent on time in general, but for now we make the simplifying assumption that it is constant. Since I is constant, γ only affects the speed of the dynamics, not the qualitative features such as number of fixed points and their stability. So without loss of generality we set $\gamma = 1$ for now. There are then two parameters which affect the qualitative behaviour of the system, so the bifurcations which are possible upon varying one depend on the current value of the other. Next, we consider these bifurcations in one parameter when the other is fixed, to better understand the role they play in the neuron dynamics. Specifically, a combination of computational and analytical work reveals a cusp bifurcation with respect to the m and I parameters.

2.1.1 Bifurcations when varying m for fixed I

following results are derived computationally, by observing phase portraits for different parameter values. Figure 2.1 displays phase portraits which show the effect of increasing m when $I = 0$. If $0 < m < 1$, then the system has a single globally stable fixed point at the origin (fig 2.1 (a)). For $m = 1$, the origin undergoes a supercritical pitchfork bifurcation (fig 2.1 (b)), causing it to lose stability and give rise to two stable fixed points. One fixed point is positive, and represents the firing state and the other is negative and represents the non-firing state. As m increases further, the origin remains an unstable fixed point and the two fixed points created in the bifurcation converge to 1 and -1 (fig 2.1 (c) and (d)).

Figure 2.2 displays phase portraits for increasing m while $I > 0$. For sufficiently small m , the system has a single stable fixed point $r_0^* > 0$ (fig 2.2 (a)). A saddle node bifurcation occurs for a value of m which depends on I (fig 2.2 (b)), giving rise to fixed points $r_+^* \geq r_-^*$ both to the left of r_0^* . As m increases, $r_0^* \rightarrow 1$, $r_+^* \rightarrow 0$ and $r_-^* \rightarrow -1$ (fig 2.2 (c) and (d)). The case where $I < 0$ is essentially the same. For sufficiently small m , there is a globally stable fixed point $r_0^* < 0$. Then a saddle node bifurcation occurs for a value of m which depends on I giving rise to fixed points r_+^* and r_-^* , both to the right of r_0^* . As m increases, $r_0^* \rightarrow -1$, $r_+^* \rightarrow 1$ and $r_-^* \rightarrow 0$.

2.1.2 Bifurcations when varying I for fixed m

For a fixed value of m , increasing I has the effect of sliding the $\dot{r}$ curve, given by the right hand side of (2.1), up and to the left along the curve $f(r) = -r$, whereas decreasing I slides the nullcline down and to the right of this line. Examples of this can be seen for $m = 2$ in Fig. 2.4. In this case, we can also find a continuous, closed form expression for the I bifurcation curve by setting the left side of (2.1) to zero

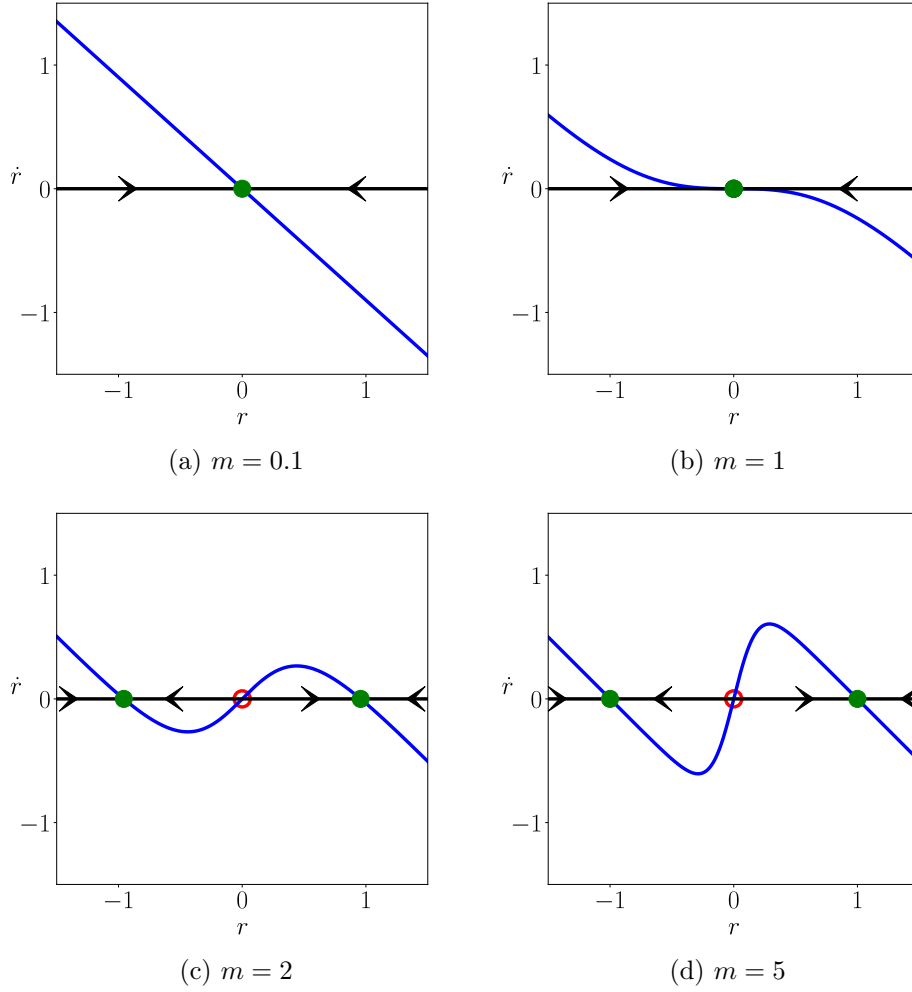


Figure 2.1: Single neuron phase portraits for $I = 0$ and $m = 0.1, 1, 2, 5$ in panels (a), (b), (c), and (d) respectively. Filled green and empty red circles represent stable and unstable fixed points respectively, arrows indicate flow of the r variable, the curve for $\dot{r}(r)$ given by the right hand side of (2.1) is shown in blue. Horizontal black line shows the one-dimensional phase space for the r variable.

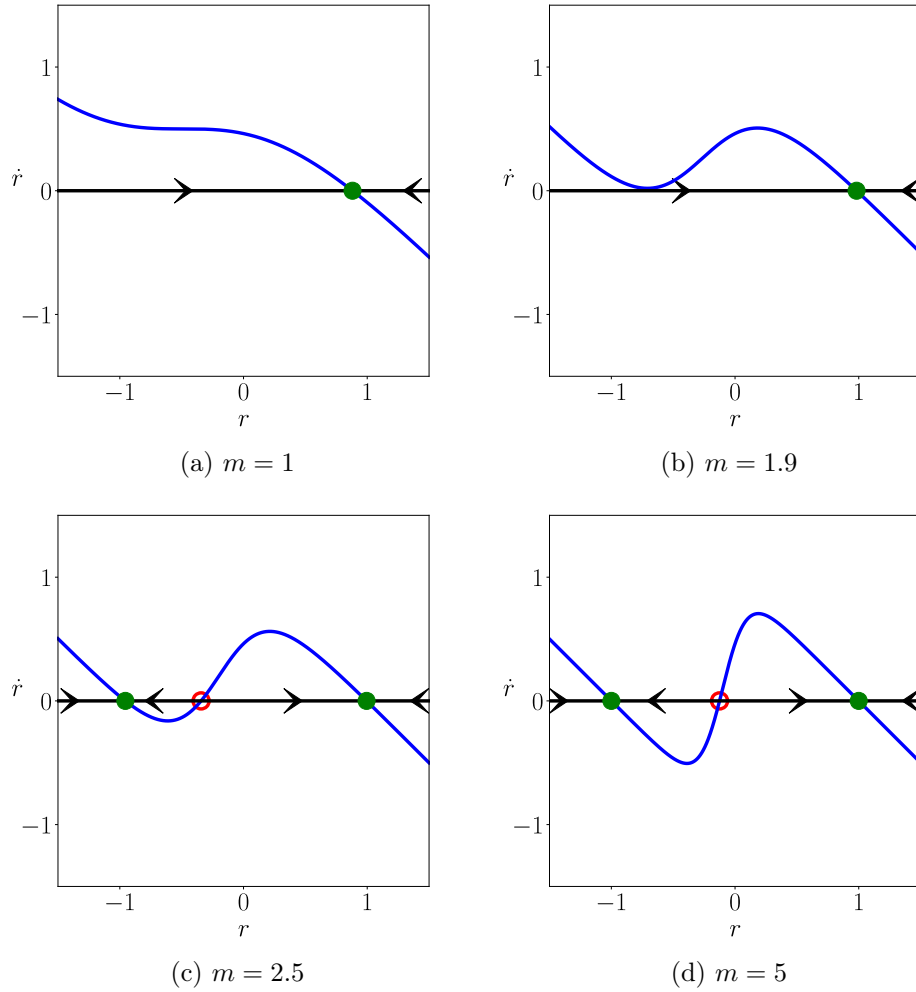


Figure 2.2: Single neuron phase portraits for $I=0.5$ and $m = 1, 1.9, 2.5, 5$ in panels (a), (b), (c), and (d) respectively. Filled green and empty red circles represent stable and unstable fixed points respectively, arrows indicate flow of the r variable, the curve for $\dot{r}(r)$ given by the right hand side of (2.1) is shown in blue. Horizontal black line shows the one-dimensional phase space for the r variable.

and solving for I ,

$$\begin{aligned} 0 &= -r + \tanh(mr + I), \\ \implies I &= \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) - mr, \quad r \in (-1, 1). \end{aligned} \quad (2.2)$$

Note that the expression for the I bifurcation curve means that its exact form depends on the fixed value of m . This can be seen by comparing the I bifurcation curves for $m = 0.5$ in Fig. 2.3, panel (a) and $m = 2$ in Fig. 2.4, panel (a). If $0 < m \leq 1$, for all values of I , the system has a single globally stable fixed point between -1 and 1 . The bifurcation diagram for $m = 0.5$ and some corresponding phase portraits for different I values are shown in figure 2.3.

If $m > 1$, there are two saddle node bifurcations which occur at $I_- < 0 < I_+$, as shown for the case $m = 2$ in the bifurcation diagram of figure 2.4 (a). For $I \in (I_-, I_+)$, there are two stable fixed points whose basins of attraction are separated by an unstable fixed point between them (figure 2.4 (d)). Outside this region of I values, there is a single globally stable fixed point close to either 1 or -1 depending on whether I is positive or negative respectively (figure 2.4 (b) and (f)). The I values where saddle node bifurcations occur are computed by finding the local optima of the I bifurcation curve given by equation (2.2),

$$\begin{aligned} \frac{dI}{dr} &= 0 \\ \implies \frac{1}{(1+r)(1-r)} - m &= 0 \\ \implies r_{\pm}(m) &= \pm \sqrt{\frac{m-1}{m}}, \quad \text{for } m \geq 1. \end{aligned}$$

The values of I where saddle node bifurcations occur are then,

$$I_{\pm}(m) := I(r_{\pm}(m)) = \frac{1}{2} \ln \left(\frac{1+r_{\pm}(m)}{1-r_{\pm}(m)} \right) - mr_{\pm}(m).$$

This saddle-node bifurcation curve in the $m - I$ plane, which is plotted in fig 2.5, consists of two branches symmetric about $I = 0$, which meet at the cusp point $I = 0, m = 1$. The horizontal distance between the two branches for each value of m is $2I(r_{\pm}(m)) \rightarrow \infty$ as $m \rightarrow \infty$. The curve divides the $m - I$ parameter space into two regions, where the system has one globally stable fixed point (region a in figure 2.5) and three fixed points, one unstable and two stable (region b). The

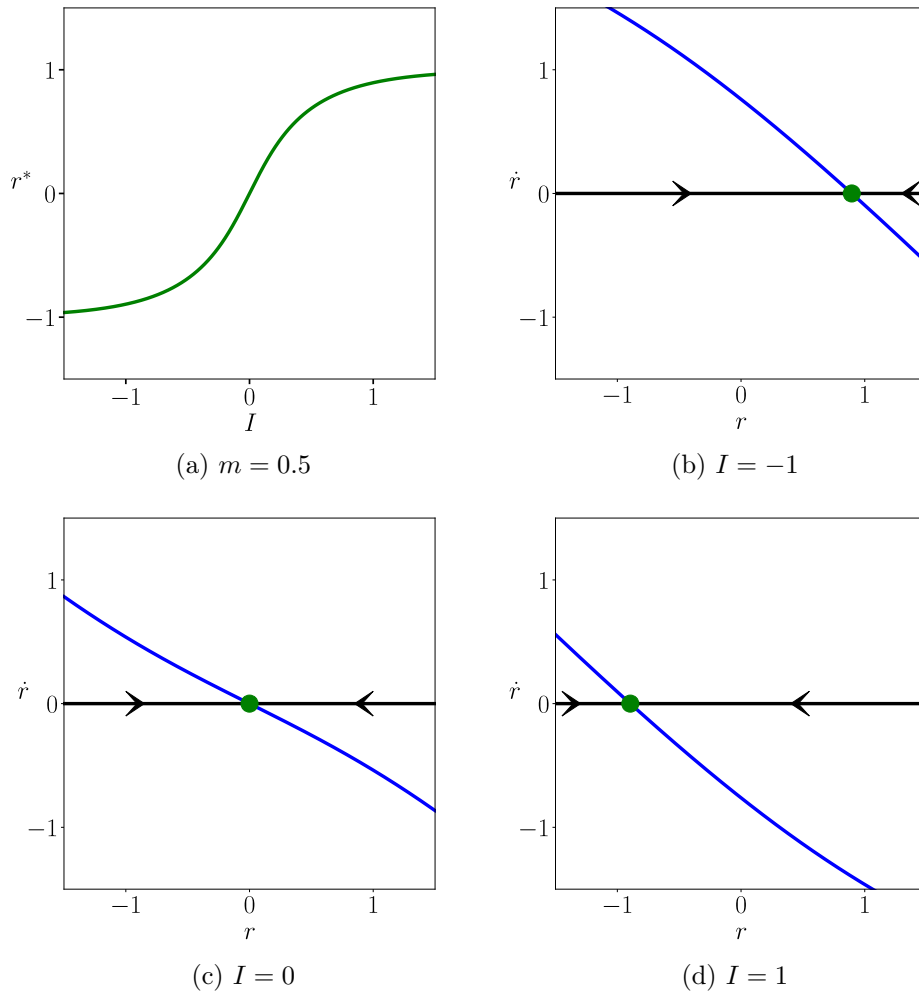


Figure 2.3: Panel (a) I shows a bifurcation diagram for $m = 0.5$, green solid curve indicates branch of stable equilibria. Panels (b), (c), (d) show single neuron phase portraits for $m = 0.5$ and $I = -1, 0, 1$ respectively. Filled green and empty red circles represent stable and unstable fixed points respectively, arrows indicate flow of the r variable, the curve for $\dot{r}(r)$ given by the right hand side of (2.1) is shown in blue. Horizontal black line shows the one-dimensional phase space for the r variable.

system has two fixed points for any pair of parameters on a branch of the curve except at the cusp point $(m, I) = (1, 0)$, where there is a single globally stable fixed point. This phenomenon is an example of a co-dimension two bifurcation known as a cusp bifurcation [24]. Systems with a cusp bifurcation are common choices for phenomenological models of state dependent tipping phenomena, where the system can transition between two qualitatively different states when some input crosses certain threshold values. In our case we choose such a system to mimic the transition between states of high density firing and low density firing in a single neuron, or populations of similar neurons, in response to a time varying external electrical input. For this reason we assume from this point onwards that $m > 1$, so that the single neuron has two stable fixed points, one positive and one negative. Additionally, we refer to the neuron as being in the "on" state if $r > 0$ and in the "off" state if $r < 0$. In the next section, we discuss some features of the dynamics of the single neuron when given a time varying input, and how they can be understood in terms of the underlying dynamics for fixed input.

2.1.3 Time Varying Inputs

The previous analysis on the single neuron dynamics was conducted for the autonomous case of a fixed input. However, it can still be used to guide our understanding of the non-autonomous case when I is a function of time, if we think of $r(t)$ at each instant in time as moving toward a fixed point which is present in the autonomous system when $I = I(t)$. The range of dynamical behaviour which is possible for a general input $I(t)$ is extremely broad and depends on many factors such as the timescales of the r dynamics and I as well as the range of values which I can take. Our goal in this section is to use this way of thinking to gain general insight into the behaviour of a neuron when given a time varying input.

Suppose we have an input $I(t) \in (a, b) \subset \mathbb{R}$ to the single neuron with $m > 1$ and $\epsilon \in \mathbb{R}^+$. Using the corresponding I bifurcation curves, we can reach some conclusions about what types of motion which are possible for the variable r (fig 2.6). If $(a, b) \subset (I_-, I_+)$, then the possible trajectories are confined to the middle strip of the $I - r$ plane between the black dashed lines, displayed in figure 2.6 (a). If the initial condition is in either of the blue rectangles, then the trajectory will always approach the stable branch of equilibria which is in the corresponding rectangle. After a sufficiently long time, the motion will be confined to a small area around this branch regardless of the exact form of the input $I(t)$. If the initial condition is in the white rectangle, the type of motion depends on the input. For example, if the initial condition is below the red dashed curve, it will initially approach the lower stable branch, but it is possible for the input to change quickly enough to cause the

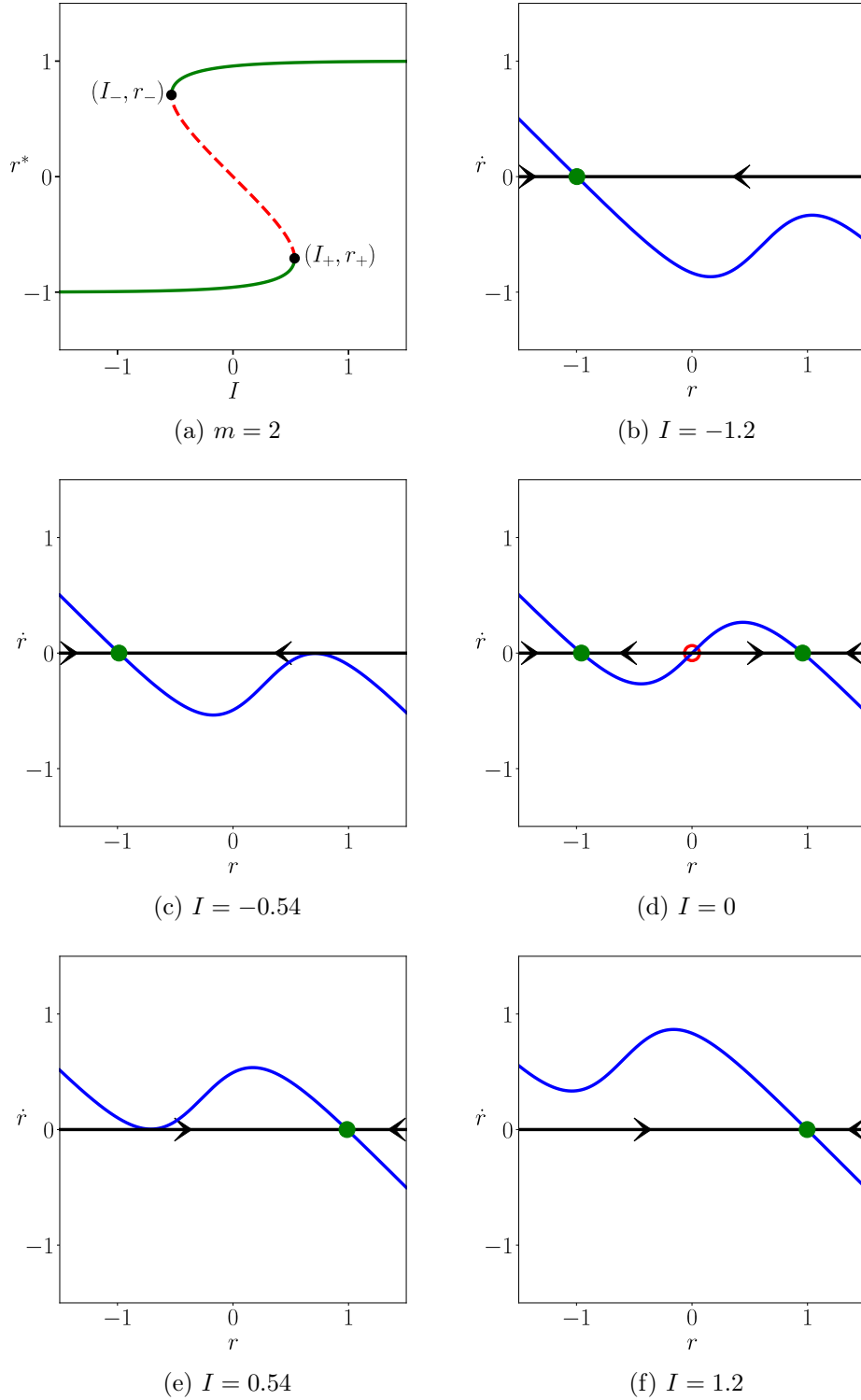


Figure 2.4: Panel (a) I shows a bifurcation diagram for $m = 2$, solve green curves and red dashed curve indicate stable and unstable equilibria respectively. Panels (b) - (f) show single neuron phase portraits for $m = 2$ and $I = -1.2, -0.54, 0, 0.54, 1.2$ respectively. Filled green and empty red circles represent stable and unstable fixed points respectively, arrows indicate flow of the r variable, the curve for $\dot{r}(r)$ given by the right hand side of (2.1) is shown in blue. Horizontal black line shows the one-dimensional phase space for the r variable.

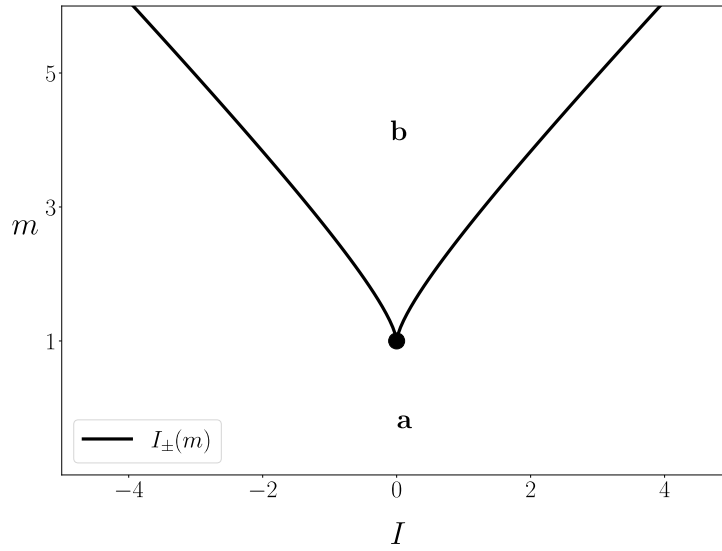


Figure 2.5: Cusp bifurcation diagram in the (I, m) plane: Curves of saddle node bifurcations $I_{\pm}(m)$ are shown in black, which meet at the cusp point at $(0, 1)$ shown as a black dot. $I_{-}(m)$ is on the left, $I_{+}(m)$ is on the right. **a**, **b** denote regions where the single neuron dynamics has one fixed point and three fixed points respectively.

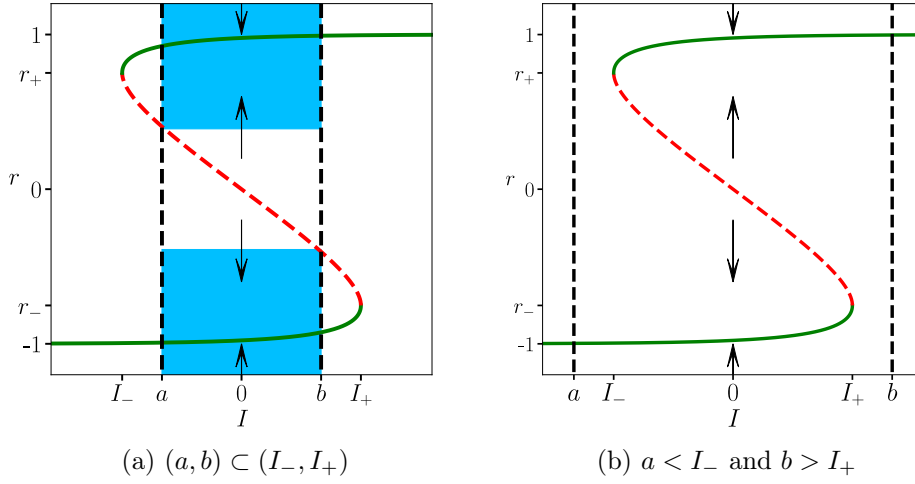


Figure 2.6: Tipping in the non-autonomous single neuron system: Black dashed lines indicate the possible range of inputs to the system I . Green and red dashed curves are the bifurcation diagram of the autonomous system for $m > 0$. $I_{\pm}$ are values of the parameter I where saddle node bifurcations occur. Arrows indicate direction of the flow of the autonomous system for a fixed value of I , depending on the value of r . Panel (a): Blue rectangles indicate regions which cannot be escaped once entered. Panel (b): Bifurcation induced tipping is possible, so there are no regions where the pair $(r(t), I(t))$ become trapped.

state to cross the red curve, so that r will begin approaching the upper branch. This is an example of rate induced tipping [25]. Note that this phenomenon is similar to basin instability [26], since this behaviour is caused by a change in the input which is fast enough to cause $r(t)$ to leave the basin of attraction of one stable branch and enter that of another. If this crossing of the dashed curve before entering a blue rectangle continues, then the motion will be confined to the white rectangle. However, if r ever crosses into a blue region, then the state will approach one of the stable branches and remain close to it for all time.

If we have either $a < I_-$ or $b > I_+$ (2.6 (b)), then it is possible for the input to pass a point where a saddle node bifurcation occurs, causing the trajectory to make a transition to tracking the other branch. As an example, consider the case where initially $I(t_0) \in (I_+, I_-)$ and $r(t_0) \in (r_+, 1)$, so that the starting position in the (I, r) plane is close to the upper branch of equilibria. If I crosses the threshold I_- at some time $t_1 > t_0$, then a saddle-node bifurcation in the autonomous neuron dynamics causes the current fixed on the upper branch of equilibria to disappear, so that r begins heading toward the lower branch of equilibria. If I crosses back over I_- before r crosses r_+ , then r will return to tracking the upper branch. If at some time t_2 we have $I(t_2) > I_+$ and $r(t_2) > r_+$, then in order for r to return to tracking the upper branch, I must increase enough for (r, I) to cross the red dashed curve. This

threshold I must cross increases gradually the closer r gets to r_- , until finally if we have $r > r_-$, then I must cross I_- in order to make r transition back to the upper branch. This scenario, where the input passes through a value where a bifurcation occurs in the autonomous system leads to a drastic change in the value of r , is known as bifurcation induced tipping [27]. In the scenario where $a < I_-$ and $b > I_+$, the system can switch between tracking one branch or the other, depending on the type of input. An important takeaway from this is that the parameter m determines how big the range of inputs to a neuron must be in order for switching to be possible, since the thresholds $I_{\pm}$ are both functions of m . A particularly important case is when $I(t)$ changes slowly relative to the timescale of the r dynamics, determined by the parameter ϵ . Then the behaviour of r is characterised by spending long periods of time close to a branch of equilibria and making fast transitions between branches as the input crosses the thresholds I_+, I_- .

2.2 N dimensional CTRNNS

To create an N dimensional CTRNN with $N > 1$, we take N individual neurons r_i and add a coupling term to each of their dynamical equations,

$$\epsilon \dot{r}_i = -r_i + \tanh \left(m r_i + \sum_{j \neq i} m_{ij} r_j \right). \quad (2.3)$$

Now the input to each node is a weighted sum of the other node states, weighted by synaptic efficacies $m_{ik} \in \mathbb{R}$. So the input controlling the switching behaviour of each neuron is determined by the state of the other neurons. The magnitude of m_{ik} determines the strength of the effect that r_k has on r_i , the sign of m_{ik} determines the nature of the effect. If $m_{ik} > 0$ and r_k is in the on state, then $m_{ik} r_k > 0$, so r_k makes a positive contribution to the total input to r_i , which pushes it towards the on state. Similarly, if $m_{ik} > 0$ and r_k is in the off state, then r_k is pushing r_i to the off state. This means if $m_{ik} > 0$, r_k always pulls r_i toward its current state. If $m_{ik} < 0$ and r_k is in the on state, then $m_{ik} r_k > 0$, so r_k makes a negative contribution to the total input to r_i , which pushes it towards the off state. Similarly, if $m_{ik} < 0$ and r_k is in the off state, then r_k is pushing r_i to the on state. So r_k always pushes r_i away from its current state if $m_{ik} < 0$. The sum of these forces acting on r_i , in addition to the current state of r_i determines whether the node makes a transition to another state. Equation (2.3) is an example of a dynamical network, with connection strengths given by the synaptic weights m_{ij} .

CTRNNs of the form (2.3), which do not consider the connectivity structure of the

network as being dynamic, have been used extensively in computational neuroscience to model biological NNs [17]. They are also the basis for many machine learning methods for sequence tasks such as reservoir computers [23]. However, it is still important to acknowledge that in reality, biological NN connectivity structure given by synapses is not fixed, but changes dynamically in response to network activity via plasticity. We are therefore motivated to augment the CTRNN to include a plasticity mechanism, so that it more accurately reflects a biological NN.

Choosing how to include plasticity into a NN model is a task which requires careful consideration as there are many possible way to proceed. The approach taken here is to model a NN with plasticity as a so called adaptive dynamical network [14], which can be done by augmenting equation (2.3) to include dynamic equations for the connection weights m_{ij} ,

$$\begin{aligned}\gamma \dot{r}_i &= -r_i + \tanh \left(\sum_{k=1}^N m_{ik} r_k + I_i(t) \right), \\ \dot{m}_{ij} &= F(\mathbf{r}, \mathbf{M}), i \neq j\end{aligned}$$

where $F(\mathbf{r}, \mathbf{M})$ is some suitably chosen scalar function of the neuron states $\mathbf{r} = (r_1, \dots, r_N)$ and the weight values $\mathbf{M}_{ij} = (m_{ij})$, called the learning rule.

2.3 Learning Rule

Our goal when choosing how to model dynamic plasticity in such a network will be to create a simple, phenomenological model which captures certain features of synaptic neural plasticity and discards others. The learning rule we choose is from the class of Hebbian learning rules, which share the feature that the evolution of each weight is based on the correlation of the past activity of the neurons it connects [19].

There are a couple of important aspects of Hebb's postulate which are taken into account when formulating a Hebbian learning rule. The first is locality, the change in a synaptic weight must be based only on information available at the synapse, such as the current synaptic weight, and the activity of the pre- and post-synaptic neurons. The next is co-operativity, the synaptic dynamics for a particular connection must be based on both the pre- and the post-synaptic neuron activity [19]. The learning rule we use here incorporates these two features is, and is given by

$$\dot{m}_{ij} = \alpha_{ij}\eta r_i r_j - m_{ij}, \quad (2.4)$$

where $\eta \in \mathbb{R}^+$ is a parameter called the learning rate and $\alpha_{ij} = \pm 1$ are parameters which determine if a synapse is Hebbian ($\alpha_{ij} = 1$) or anti-Hebbian ($\alpha_{ij} = -1$), concepts which are explained in section 2.3.2. Next we discuss some effects that these parameters have on the network dynamics.

2.3.1 The Learning Rate η

Integrating both sides of equation (2.4) gives

$$m_{ij}(t) = \alpha_{ij}\eta \int_0^t e^{\tau-t} r_i(\tau) r_j(\tau) d\tau + e^{-t} m_{ij}(0), \quad (2.5)$$

which shows that the influence of the initial condition $m_{ij}(0)$ on the value of $m_{ij}(t)$ vanishes exponentially fast. After a sufficiently long time we then have,

$$m_{ij}(t) \approx \alpha_{ij}\eta \int_0^t e^{\tau-t} r_i(\tau) r_j(\tau) d\tau, \quad (2.6)$$

which is simply the scaled average value of $r_i(\tau)r_j(\tau)$ on the interval $[0, t]$ taken using an exponential window function. The presence of the exponential window function means that only the recent correlated behaviour of r_i and r_j effects m_{ij} at any instant in time.

Equation (2.6) shows that η controls the speed of the m_{ij} dynamics, hence the name "learning rate". It also determines the magnitude which the weights can reach over the course of the system evolution, since on a solution we have

$$\begin{aligned} |m_{ij}(t)| &= |\alpha_{ij}\eta \int_0^t e^{\tau-t} r_i(\tau) r_j(\tau) d\tau + e^{-t} m_{ij}(0)|, \\ &\leq \eta \left| \int_0^t e^{\tau-t} r_i(\tau) r_j(\tau) d\tau \right| + |e^{-t} m_{ij}(0)|, \\ &\leq \eta \int_0^t e^{\tau-t} |r_i(\tau) r_j(\tau)| d\tau + e^{-t} |m_{ij}(0)|, \\ &\leq \eta e^{-t} \int_0^t e^{\tau} (1) d\tau + e^{-t} |m_{ij}(0)|, \\ \implies |m_{ij}(t)| &\leq \eta (1 + (|m_{ij}(0)|/\eta - 1) e^{-t}). \end{aligned}$$

From the last inequality it follows that if $|m_{ij}(0)| \leq \eta$, then $m_{ij}(t) \in [-\eta, \eta]$. If

$|m_{ij}(0)| > \eta$, then $\inf_{a \in [-\eta, \eta]} |m_{ij}(t) - a| \rightarrow 0$ exponentially fast as $t \rightarrow \infty$.

2.3.2 Hebbian and Anti-Hebbian Learning

One of the significant effects of adding adaptive coupling is that it makes the weights m_{ij} capable of changing sign based on the activity of the nodes it connects, and in doing so means that, for example, a node which was pulling another node toward its state can transition to pushing it away from its state based on their recent activity. The exact nature of the dependence of the sign of m_{ij} on the state of the neurons depends on the parameter α_{ij} .

If $\alpha_{ij} = 1$, then m_{ij} is a Hebbian synapse. At a given point in time t , if the two neurons tend to be simultaneously in the off or on state in the recent past, the weight will be positive. If they tend to be in opposite states simultaneously (i.e., $r_i = 1, r_j = -1$) then the weight will be negative. If no such relation exists, and they tend to be in opposite states as often as they are in the same state, the weight will be zero. So the Hebbian synapse causes r_j to attract r_i if they both tend to be in the same state, and repel r_i if they both tend to be in different states.

If $\alpha_{ij} = -1$, then m_{ij} is an anti-Hebbian synapse, as defined in the beginning of Sec. 2.3. At a given point in time t , if the two neurons tend to be simultaneously in the off or on state in the recent past, the weight will be negative. If they tend to be in opposite states simultaneously then the weight will be positive. If no such relation exists, and they tend to be in opposite states as often as they are in the same state, the weight will be zero. The Anti-Hebbian synapse causes r_j to repel r_i if they both tend to be in the same state, and attract r_i if they both tend to be in different states.

2.3.3 Pairwise Interactions

A first step in better understanding the dynamics of the network as a whole, is to understand the nature of pairwise interactions between nodes r_i, r_j and their shared weights m_{ij}, m_{ji} . Due to the symmetric nature of the weight dynamics, after a sufficiently long time, any pair of weights m_{ij}, m_{ji} will satisfy $m_{ij} \approx \alpha_{ij}\alpha_{ji}m_{ji}$. To show this, first define $\Delta m_{ij} = m_{ji} - \alpha_{ij}\alpha_{ji}m_{ij}$. Then

$$\begin{aligned}
\frac{d\Delta m_{ij}}{dt} &= \dot{m}_{ji} - \alpha_{ij}\alpha_{ji}\dot{m}_{ij}, \\
&= \alpha_{ji}r_jr_i - m_{ji} - \alpha_{ij}^2\alpha_{ji}r_ir_j + \alpha_{ij}\alpha_{ji}m_{ij}, \\
&= \alpha_{ji}r_jr_i - (1)\alpha_{ji}r_ir_j - (m_{ji} - \alpha_{ij}\alpha_{ji}m_{ij}), \\
&= -\Delta m_{ij}.
\end{aligned}$$

So $\Delta m_{ij} \rightarrow 0$ as $t \rightarrow \infty$, which implies $m_{ij} \rightarrow \alpha_{ij}\alpha_{ji}m_{ji}$ as $t \rightarrow \infty$. The convergence to this state also happens exponentially fast. We consider the cases where both synapses are Hebbian, anti-Hebbian and where one synapse is Hebbian and the other anti-Hebbian.

If the neurons r_i and r_j are connected via a pair of Hebbian synapses or a pair of anti-Hebbian synapses, then $\alpha_{ij} = \alpha_{ji}$, so $m_{ij} \rightarrow m_{ji}$. This means the network quickly settles to a state where nodes r_i and r_j are either simultaneously attracting each other or repelling each other toward the state they are currently in. In the case of two Hebbian synapses, the neurons will eventually attract each other if they spend sufficiently long in the same state, since this means $\eta r_i r_j > 0$ for long enough to make $m_{ji} > 0$. Similarly, they eventually repel each other if they spend sufficiently long in different states, as this means $\eta r_i r_j < 0$ for long enough to make $m_{ij}, m_{ji} < 0$. If both synapses are Anti-Hebbian, the neurons will eventually repel each other if they spend sufficiently long in the same state, because that means $\eta r_i r_j > 0$ for long enough to make $m_{ij}, m_{ji} > 0$. Similarly, they eventually attract each other if they spend sufficiently long in different states, as this means $-\eta r_i r_j > 0$ for long enough to make $m_{ij}, m_{ji} < 0$.

If the neurons r_i and r_j are connected via a Hebbian synapse and an anti Hebbian synapse, then $\alpha_{ij} = -\alpha_{ji}$, so $m_{ij} \rightarrow -m_{ji}$. This means the network quickly settles to a state where one node is attracting and another node is repelling. Without loss of generality, assume $\alpha_{ij} = 1$ and $\alpha_{ji} = -1$, so the synapse from neuron j to neuron i is Hebbian and the synapse from neuron i to j is Anti-Hebbian. If the neurons spend sufficiently long in the same state, r_i will attract r_j and r_j will repel r_i . If they spend sufficiently long in different states, r_i will repel r_j and r_j will attract r_i .

2.4 Timescale Separation

$$\begin{aligned}
\epsilon \frac{dr_i}{dt} &= -r_i + \tanh \left(\sum_{k=1}^N m_{ik} r_k + I_i(t) \right), \\
\frac{dm_{ij}}{dt} &= \alpha_{ij} \eta r_i r_j - m_{ij},
\end{aligned} \tag{2.7}$$

The parameter $0 < \epsilon \ll 1$ represents the ratio of timescales between the fast neuron dynamics and the slow synaptic dynamics. These equations are written with respect to the slow timescale t , we can also write the equations with respect to the fast timescale $\tau = t/\epsilon$,

$$\begin{aligned}\frac{dr_i}{d\tau} &= -r_i + \tanh\left(\sum_{k=1}^N m_{ik}r_k + I_i(t)\right), \\ \frac{dm_{ij}}{d\tau} &= \epsilon(\alpha_{ij}\eta r_i r_j - m_{ij}).\end{aligned}\tag{2.8}$$

Suppose the system state is far from the critical manifold [28], which is the set of points C_0 , where $\frac{dr_i}{d\tau} = 0$

$$C_0 = \{(\mathbf{r}, \mathbf{M}) \in \mathbb{R}^{N^2} | f(\mathbf{r}, \mathbf{M}) = 0\}.$$

Then on the fast timescale, it is reasonable to approximate the dynamics of the full system by considering the slow variables m_{ij} as fixed parameters. This is equivalent to taking the limit $\epsilon \rightarrow 0$ in the system written with respect to the fast timescale, which gives the fast-subsystem

$$\begin{aligned}\frac{dr_i}{d\tau} &= -r_i + \tanh\left(\sum_{k=1}^N m_{ik}r_k + I_i(t)\right), \\ \frac{dm_{ij}}{d\tau} &= 0.\end{aligned}\tag{2.9}$$

Which is simply the equation (1.2) for the CTRNN with fixed weights. If the full system state is close to the critical manifold, it is no longer a good approximation to consider m_{ij} as parameters. This case is instead approximated by taking the limit $\epsilon \rightarrow 0$ in the slow timescale equation, giving

$$\begin{aligned}\frac{dm_{ij}}{dt} &= \alpha_{ij}\eta r_i r_j - m_{ij}, \\ 0 &= -r_i + \tanh\left(\sum_{k=1}^N m_{ik}r_k + I_i(t)\right),\end{aligned}$$

which is called the slow subsystem. This differential - algebraic equation can be used to define a flow on regions of the critical manifold where it is normally hyperbolic

[28]. $S \subset C_0$ is normally hyperbolic if $D_{\mathbf{r}}f(\mathbf{r}^*, \mathbf{M}^*)$ has eigenvalues with nonzero real part for all $(x^*, y^*) \in S$. This is equivalent to saying that for each point $(\mathbf{r}^*, \mathbf{M}^*) \in S$, $\mathbf{r}^*$ is a hyperbolic fixed point of the fast subsystem with parameter $\mathbf{M} = \mathbf{M}^*$. Furthermore, as stated in [28] a normally hyperbolic subset $S \subset C_0$ is called

1. Attracting: If $D_{\mathbf{r}}f(\mathbf{r}^*, \mathbf{M}^*)$ has eigenvalues with strictly negative real parts $\forall(\mathbf{r}^*, \mathbf{M}^*) \in S$,
2. Repelling: If $D_{\mathbf{r}}f(\mathbf{r}^*, \mathbf{M}^*)$ has eigenvalues with strictly positive real parts $\forall(\mathbf{r}^*, \mathbf{M}^*) \in S$,
3. Saddle – Type: If it is neither attracting or repelling.

On normally hyperbolic subsets of the critical manifold, we can define a slow-flow by implicitly differentiating the equation for the critical manifold [28], giving

$$\dot{\mathbf{r}} = (D_{\mathbf{r}}f)^{-1} (D_{\mathbf{M}}f)\mathbf{M}.$$

The flow on the critical manifold is related to that of the full system by the following theorem due to Fenichel [28],

Theorem 1 (Fenichel) *Assume that f and g are sufficiently smooth. Suppose S_0 is a compact, normally hyperbolic sub-manifold of the critical manifold C_0 . Then for $\epsilon > 0$ sufficiently small,*

- *There exists a locally invariant manifold S_ϵ , diffeomorphic to S_0 ,*
- *S_ϵ has a distance $O(\epsilon)$ from S_0 ,*
- *The flow on S_ϵ converges to the slow flow as $\epsilon \rightarrow 0$,*
- *S_ϵ is normally hyperbolic, with the same stability properties as S_0 with respect to the fast variables (i.e., attracting, repelling or saddle type),*
- *For fixed $\epsilon > 0$, S_ϵ is typically not unique, but all manifolds satisfying properties (1) to (5) lie at a Hausdorff distance $O(e^{-K/\epsilon})$ from each other, for some $K > 0$ with $K = O(1)$.*

Now that we have better understanding of some general properties of this model, we move on to studying the simplest possible examples, which are networks consisting of two neurons.

Chapter 3

Networks of Two Neurons

The system of two mutually coupled neurons has dynamics given by the following equations,

$$\begin{aligned}\epsilon \dot{r}_1 &= -r_1 + \tanh(mr_1 + m_{12}r_2), \\ \epsilon \dot{r}_2 &= -r_2 + \tanh(mr_2 + m_{21}r_1), \\ \dot{m}_{12} &= \alpha_{12}\eta r_1 r_2 - m_{12}, \\ \dot{m}_{21} &= \alpha_{21}\eta r_2 r_1 - m_{21}.\end{aligned}\tag{3.1}$$

From the previous chapter, we know that the weights converge exponentially fast to the state $m_{21} \approx \alpha_{12}\alpha_{21}m_{12}$. Because of this, we reduce the dimension of the system by making the approximation $m_{21} = \alpha_{12}\alpha_{21}m_{12}$ and relabelling $m_{21} \mapsto m_c$. Then after substituting the new variables the equations become,

$$\begin{aligned}\epsilon \dot{r}_1 &= -r_1 + \tanh(mr_1 + m_c r_2), \\ \epsilon \dot{r}_2 &= -r_2 + \tanh(mr_2 + \alpha_{12}\alpha_{21}m_c r_1), \\ \dot{m}_c &= \alpha_{12}\eta r_1 r_2 - m_c.\end{aligned}\tag{3.2}$$

We also fix $m = 2$ and $\epsilon = 0.01$ for the remainder of this chapter, and consider the effect of changing α_{12} and η .

3.1 Hebbian Synapses

When both synapses are Hebbian, $\alpha_{12} = \alpha_{21} = 1$, so the equations are,

$$\begin{aligned}
\epsilon \dot{r}_1 &= -r_1 + \tanh(mr_1 + m_c r_2), \\
\epsilon \dot{r}_2 &= -r_2 + \tanh(mr_2 + m_c r_1), \\
\dot{m}_c &= \eta r_1 r_2 - m_c.
\end{aligned} \tag{3.3}$$

3.1.1 Symmetries and Equivariance

The equations (3.3) are equivariant under the group of transformations Γ generated by

$$\begin{aligned}
\gamma_{--+} &: (r_1, r_2, m_c) \rightarrow (-r_1, -r_2, m_c), \\
\gamma_{-+-} &: (r_1, r_2, m_c) \rightarrow (-r_1, r_2, -m_c), \\
\gamma_{+--} &: (r_1, r_2, m_c) \rightarrow (r_1, -r_2, -m_c), \\
\gamma_{(12)} &: (r_1, r_2, m_c) \rightarrow (r_2, r_1, m_c),
\end{aligned}$$

which is an Abelian group consisting of eight elements,

$$\Gamma = \{I, \gamma_{--+}, \gamma_{-+-}, \gamma_{+--}, \gamma_{(12)}, \gamma_{(12)}\gamma_{--+}, \gamma_{(12)}\gamma_{-+-}, \gamma_{(12)}\gamma_{+--}\}. \tag{3.4}$$

Each element of Γ is its own inverse, so $\langle \gamma \rangle = \{I, \gamma\}$ for each $\gamma \in \Gamma \setminus \{I\}$. The unique fixed point subspaces given by these subgroups is

$$\begin{aligned}
\text{Fix}(\gamma_{--+}) &= \{(0, 0, m_c) : m_c \in \mathbb{R}\}, \\
\text{Fix}(\gamma_{-+-}) &= \{(0, r_2, 0) : r_2 \in \mathbb{R}\}, \\
\text{Fix}(\gamma_{+--}) &= \{(r_1, 0, 0) : r_1 \in \mathbb{R}\}, \\
\text{Fix}(\gamma_{(12)}) &= \{(r, r, m_c) : r, m_c \in \mathbb{R}\}, \\
\text{Fix}(\gamma_{(12)}\gamma_{--+}) &= \{(r, -r, m_c) : r, m_c \in \mathbb{R}\}.
\end{aligned}$$

These fixed point subspaces are invariant under the flow given by (3.3) [29]. We do not know their stability a priori, but they can still help us understand how the dynamics is organised. For example, by reducing the full system dynamics to the one dimensional fixed point subspaces, it is clear that the points $(0, 0, 0)$, $(\pm r^*, 0, 0)$ and $(0, \pm r^*, 0)$ are equilibria for all values of η , where $\pm r^*$ are the stable fixed points for

the single neuron dynamics for $m = 2, I = 0$ which correspond to the neuron being in the on or off state. We can conclude that these fixed points must be unstable, since they are not contained on the stable branches of the critical manifold, as will be shown in the next section. There are four other stable fixed points present in the system, labelled r_{++}, r_{+-}, r_{-+} and r_{--} . The subscripts of these fixed points encode the activity of the r_1 and r_2 neurons when the system is in the corresponding fixed point, for example r_{+-} has positive first coordinate and negative second coordinate, so represents a state where the r_1 neuron is active and the r_2 neuron is inactive. These fixed points are discussed further in the remainder of this section.

3.1.2 Fast Subsystem

The fast subsystem is given by

$$\begin{aligned}\dot{r}_1 &= -r_1 + \tanh(mr_1 + m_c r_2), \\ \dot{r}_2 &= -r_2 + \tanh(mr_2 + m_c r_1),\end{aligned}$$

This is a planar dynamical system, as such, many features of its dynamics can be determined by looking at it's nullclines.

$$\begin{aligned}\dot{r}_1 : 0 &= -r_1 + \tanh(mr_1 + m_c r_2), \\ \dot{r}_2 : 0 &= -r_2 + \tanh(mr_2 + \alpha_{12}\alpha_{21}m_c r_1).\end{aligned}$$

If $m_c = 0$, then the nullclines are

$$\begin{aligned}\dot{r}_1 : 0 &= -r_1 + \tanh(mr_1), \\ \dot{r}_2 : 0 &= -r_2 + \tanh(mr_2).\end{aligned}$$

The r_1 nullcline consists of three vertical lines through the points $(\pm r^*, 0)$ and $(0, 0)$. The r_2 nullcline consists of three horizontal lines through the points $(0, \pm r^*)$ and $(0, 0)$. The system then has nine fixed points, given by the possible ordered pairs from the set of three fixed points for the 1d neuron. The four fixed points $\pm(r^*, \pm r^*)$ corresponding to the different combinations of off and on states for the two neurons are stable nodes. The four fixed points which coincide

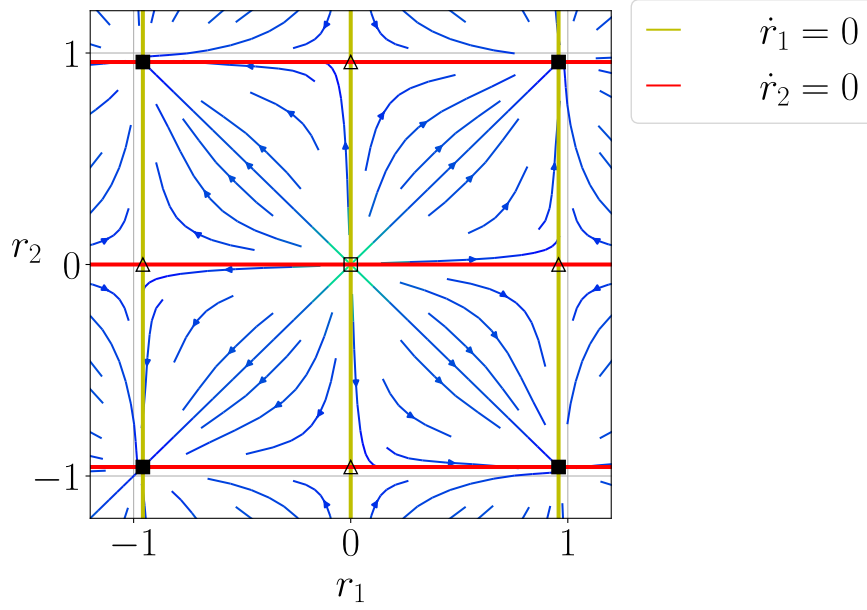


Figure 3.1: Fast subsystem phase portrait for $m = 2, m_c = 0$, squares are nodes, triangles are saddles, filled shapes are stable, unfilled shapes are unstable. Yellow and red curves are the r_1 and r_2 nullclines respectively. Curves with arrows are portions of solutions, indicating the flow of the vector field giving the dynamics.

with an axis are saddles, and the origin is an unstable node. The phase portrait for $m = 2$ and $m_c = 0$ is shown in figure 3.1.

If $m_c \neq 0$, the nullclines can be re-arranged to give,

$$\begin{aligned} \dot{r}_1 : r_2 &= \frac{1}{m_c} \left(\frac{1}{2} \ln \left(\frac{1+r_1}{1-r_1} \right) - mr_1 \right), \quad r_1 \in (-1, 1), \\ \dot{r}_2 : r_1 &= \frac{1}{m_c} \left(\frac{1}{2} \ln \left(\frac{1+r_2}{1-r_2} \right) - mr_2 \right), \quad r_2 \in (-1, 1). \end{aligned}$$

Recall that the bifurcation curve for a single neuron can be written as,

$$I(r, m) = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) - mr, \quad (3.5)$$

so the nullclines can be written as

$$\begin{aligned}\dot{r}_1 : r_2 &= \frac{1}{m_c} I(r_1, m), \\ \dot{r}_2 : r_1 &= \frac{1}{m_c} I(r_2, m).\end{aligned}$$

Thus, the r_2 nullcline is just the I bifurcation curve for the single neuron, scaled by $\frac{1}{m_c}$, meaning it has an S shape for $m_c > 0$ and a backwards S shape for $m_c < 0$. The r_1 nullcline is the curve we get if we swap the axes on the bifurcation curve, then scale by $\frac{1}{m_c}$, so it has an N shape for $m_c > 0$ and an upside down N shape for $m_c < 0$. The magnitude of the parameter m_c affects the r_1 and r_2 nullclines by scaling them in the vertical and horizontal directions respectively.

The fast-subsystem undergoes six pitchfork bifurcations for varying m_c , the resulting diagram is depicted on the left in Fig. 3.2. In this figure, bifurcation points are indicated with black points and labelled $PB_1, \dots, PB_6$, while the green and red curves show branches of stable and unstable equilibria respectively. As m_c decreases from zero, the equilibrium close to $(-1, 1)$ collides with the two saddle equilibria beside it in a subcritical pitchfork bifurcation. Due to the symmetry in the equations, the same happens with the equilibrium $(1, -1)$ and the saddles beside it, these are denoted PB_2 and PB_3 . The remaining saddles in the upper left and bottom right quadrants collide with the origin in a supercritical pitchfork bifurcation (PB_1), the remaining fixed points after this are a saddle at the origin and two stable nodes close to $(1, 1)$ and $(-1, -1)$. As m_c increases from zero, the equilibria close to $(1, 1)$ and $(-1, -1)$ undergo pitchfork bifurcations labelled PB_6 and PB_5 respectively. The remaining saddles in the system go through another pitchfork with the origin (PB_4), so that for sufficiently negative m_c there are two stable nodes close to $(-1, 1)$ and $(1, -1)$, as well as a saddle at the origin. We denote the four branches of stable equilibria as S_{++}, S_{+-}, S_{-+} and S_{--} , where the subscripts denote which quadrant of the r_1, r_2 plane the branch is contained in. Note that the branches S_{++}, S_{--} are contained in the plane $r_1 = r_2$, the branches S_{+-}, S_{-+} are contained in the plane $r_1 = -r_2$. A two dimensional projection of portions of this bifurcation diagram onto the (m_c, r_2) plane is shown in Fig. 3.2 on the right.

3.1.3 Critical Manifold

The critical manifold is given by,

$$C_0 = \{(r_1, r_2, m_c) \in \mathbb{R}^3 \mid -r_1 + \tanh(mr_1 + m_cr_2) = -r_2 + \tanh(mr_2 + m_cr_1) = 0\},$$

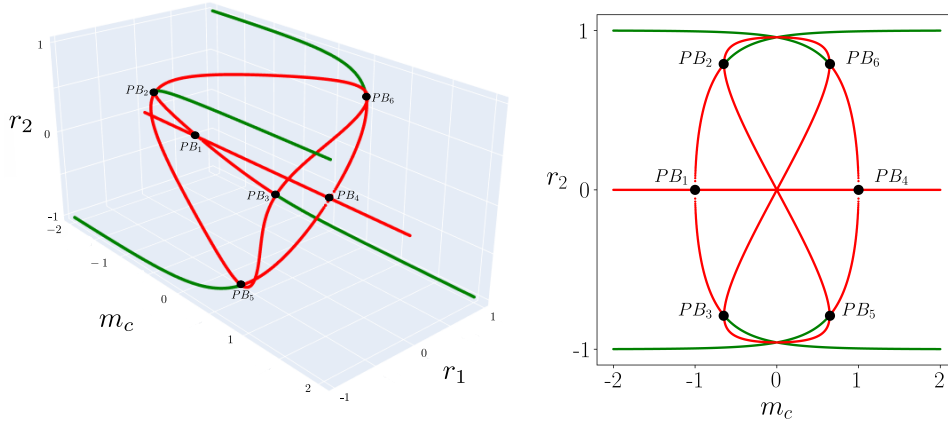


Figure 3.2: Left: Fast subsystem bifurcation diagram, green and red curves indicate branches of stable and unstable equilibria respectively. Black points labelled PB_1 to PB_6 indicate pitchfork bifurcations. Right: Two-dimensional projections onto the (m_c, r_2) plane of the bifurcation diagram

which is the m_c bifurcation curves of the equilibria for the fast-subsystem. The equilibria of the full system are the points where C_0 intersects the saddle-surface nullcline $m_c = \eta r_1 r_2$. For any $\eta > 0$, each of the branches S_{++}, S_{+-}, S_{-+} and S_{--} contains a single fixed point, r_{++}, r_{+-}, r_{-+} and r_{--} . This is because of the way the folds of the surface at the origin line up with the four stable branches of the critical manifold (see figure 3.3). These fixed points are stable in the slow flow. For any value of m_c , the fast subsystem converges to a stable fixed point, so the full system quickly approaches the critical manifold for any initial condition. It then proceeds to move along this stable branch to converge to the fixed point on it.

Because of the timescale separation, the nodes r_1, r_2 quickly approach a stable branch of the critical manifold, which corresponds to them being in a certain configuration of on-off states. Because of the Hebbian synapses connecting the neurons, as they spend more time in the same (different) on-off state, the weights connecting them adapt to reinforce them being on the same (different) state. The fixed points which attract all initial conditions, as well as the slow flow are depicted in Fig. 3.3.

3.2 Anti-Hebbian Synapses

When both synapses are Anti-Hebbian, $\alpha_{12} = \alpha_{21} = -1$, so the equations are,

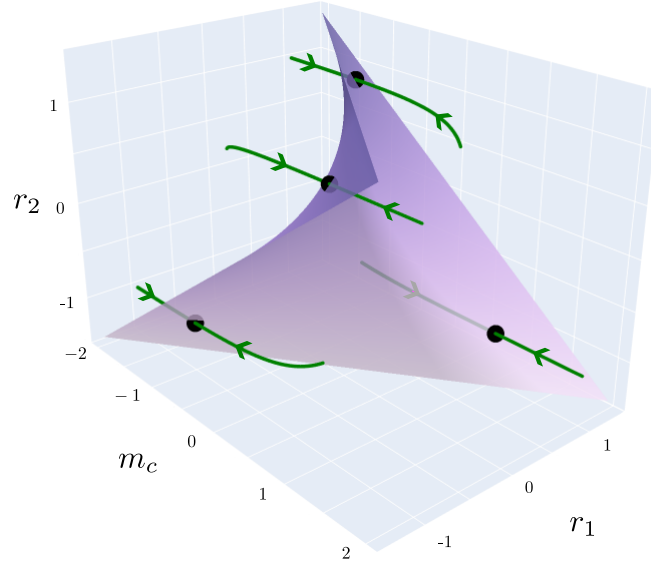


Figure 3.3: Stable fixed points of the full system with Hebbian synapses plotted as black points. Green curves are stable branches of the critical manifold with arrows indicating direction of the slow flow. Nullsurface $\dot{m}_c = 0$ in purple.

$$\begin{aligned}
 \epsilon \dot{r}_1 &= -r_1 + \tanh(mr_1 + m_cr_2), \\
 \epsilon \dot{r}_2 &= -r_2 + \tanh(mr_2 + m_cr_1), \\
 \dot{m}_c &= -\eta r_1 r_2 - m_c.
 \end{aligned} \tag{3.6}$$

The fast subsystem dynamics and critical manifold are the exact same as the case with only Hebbian synapses. The change in sign of the first term in the m_c equation has the effect of reflecting the m_c nullcline across the plane $m_c = 0$, so that a bifurcation occurs when increasing η that leads to a transition from stable fixed point dynamics, to relaxation oscillations. For sufficiently small η , the surface intersects the four stable branches of the critical manifold, leading to four stable fixed points r_{++}, r_{+-}, r_{-+} and r_{--} as before. As discussed in section 3.1, the system also has the fixed points $(0, 0, 0)$, $(\pm r^*, 0, 0)$ and $(0, \pm r^*, 0)$. As η increases, the surface becomes steeper, causing the stable fixed points to move along the stable branches, eventually simultaneously reaching the singular points where C_0 transitions to being unstable through pitchfork bifurcations in the fast subsystem. As this happens, the four fixed points each split into three saddles by undergoing separate pitchfork bifurcations. For the three saddles contained in each of these groups, one is contained in the same invariant half plane which contained the bifurcating fixed point and has one unstable direction, the other two are symmetric under reflections across this half plane and have two unstable directions. As η increases further, each pair of fixed points which

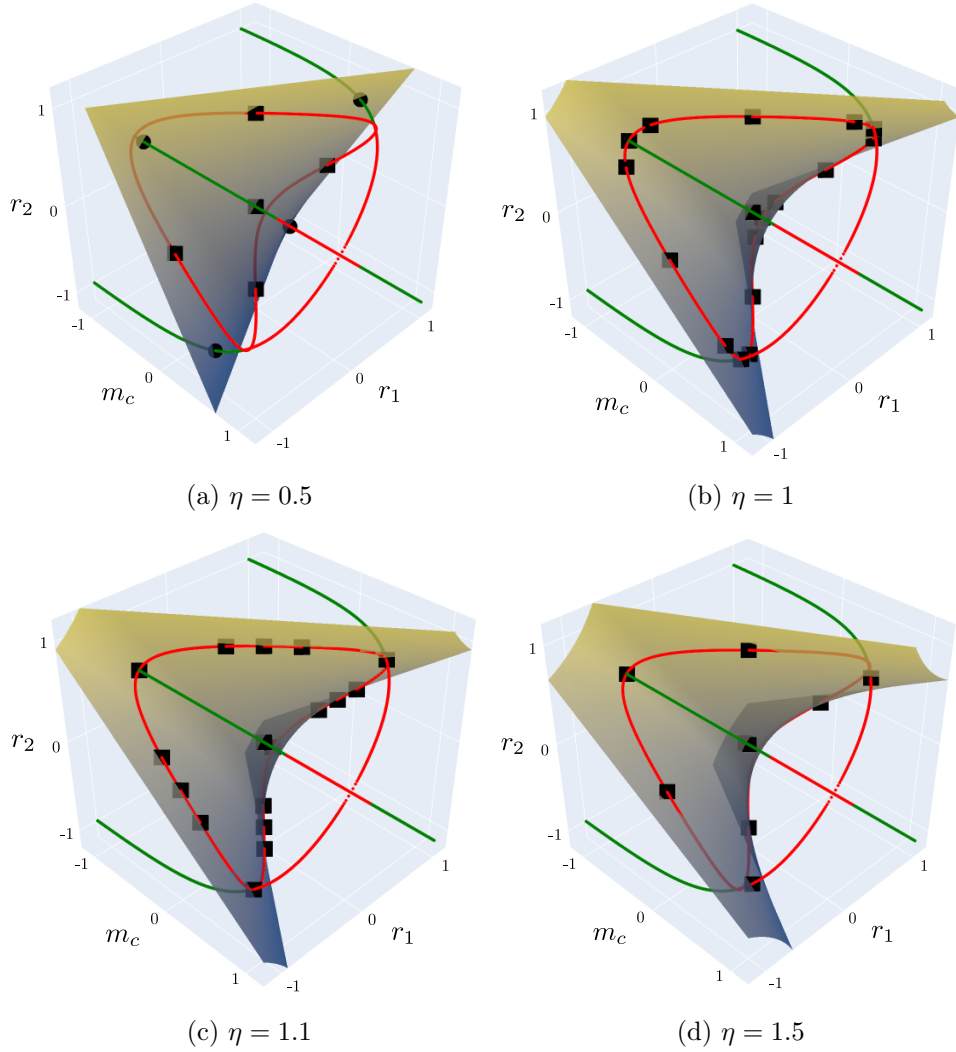


Figure 3.4: Pitchfork bifurcations in the full system when increasing η with $m = 2$. Green and red curves are stable and unstable branches of the critical manifold respectively. Surface is the nullsurface $\dot{m}_c = 0$, squares and circles are unstable and stable equilibria respectively.

is not on one of the half planes disappears as part of a pitchfork bifurcation with one of the fixed points $(\pm r^*, 0, 0)$ or $(0, \pm r^*)$. This sequence of bifurcations is shown in figure 3.4

Now when the system converges to the critical manifold and follows the slow flow, it travels past one of the singularities on C_0 , towards a saddle. When the trajectory is sufficiently close to the saddle, it makes a fast transition to another branch along the portion of the saddle's unstable manifold which is on the same side of the invariant planes as the trajectory. This process of jumping between two branches continues for all time, giving relaxation oscillations. Figure 3.8 panel (a) shows the skeleton for one such relaxation oscillation for $m = 2$, $\eta = 1.5$ and $\epsilon = 0.01$ where the system state switches between the branches S_{++} and S_{-+} along the unstable manifolds of

the saddles. In figures 3.5 and 3.6, each panel shows the projection of the evolution along this limit cycle onto the r_1, r_2 plane up to a certain time t . The fast subsystem phase portrait corresponding to $m_c(t)$, including r_1 and r_2 nullclines (in yellow and red respectively), is also plotted to show how the full system transitions from the slow flow to the fast flow, as a result of bifurcations in the fast subsystem caused by the slow variable.

Applying the three symmetries γ_{--+} (magenta), $\gamma_{(12)}$ (cyan) and $\gamma_{-+-}\gamma_{(12)}$ gives three more limit cycles which switch between the other possible pairs of stable branches of C_0 . The basins of attraction of these limit cycles are the four regions given by the complement of the union of the invariant planes $r_1 = r_2$ and $r_1 = -r_2$. Figure 3.8 shows all four limit cycles. The limit cycle which switches between S_{++} and S_{-+} is plotted again in blue. The limit cycles we get by applying the symmetries γ_{--+} , $\gamma_{(12)}$ and $\gamma_{-+-}\gamma_{(12)}$ are plotted in magenta, cyan and yellow respectively. After the points where the trajectories appear to 'enter' the critical manifold, they follow it closely on the slow timescale, without passing through the invariant plane containing it.

Because of the anti-Hebbian synapses connecting the neurons, as they spend time in the same on-off state, the weights connecting them adapt to cause them to repel each other from their current state. Eventually when this repelling force is strong enough, one of the nodes makes a transition to being in the other state. After enough time, the weights adapt to cause them to attract each other, causing the node which previously switched to transition back. In this way, the nodes alternate between being synchronised, and satisfying $r_1 = -r_2$.

3.3 Hebbian and Anti-Hebbian Synapses

The third case to consider is when one synapse is Hebbian and the other is anti-Hebbian. Without loss of generality, we consider the case where the synapse from r_1 to r_2 is Hebbian and the synapse from r_2 to r_1 is anti Hebbian. Then $\alpha_{12} = -1, \alpha_{21} = 1$, and the equations for the dynamics are

$$\begin{aligned}\epsilon \dot{r}_1 &= -r_1 + \tanh(mr_1 + m_c r_2), \\ \epsilon \dot{r}_2 &= -r_2 + \tanh(mr_2 - m_c r_1), \\ \dot{m}_c &= -\eta r_1 r_2 - m_c.\end{aligned}\tag{3.7}$$

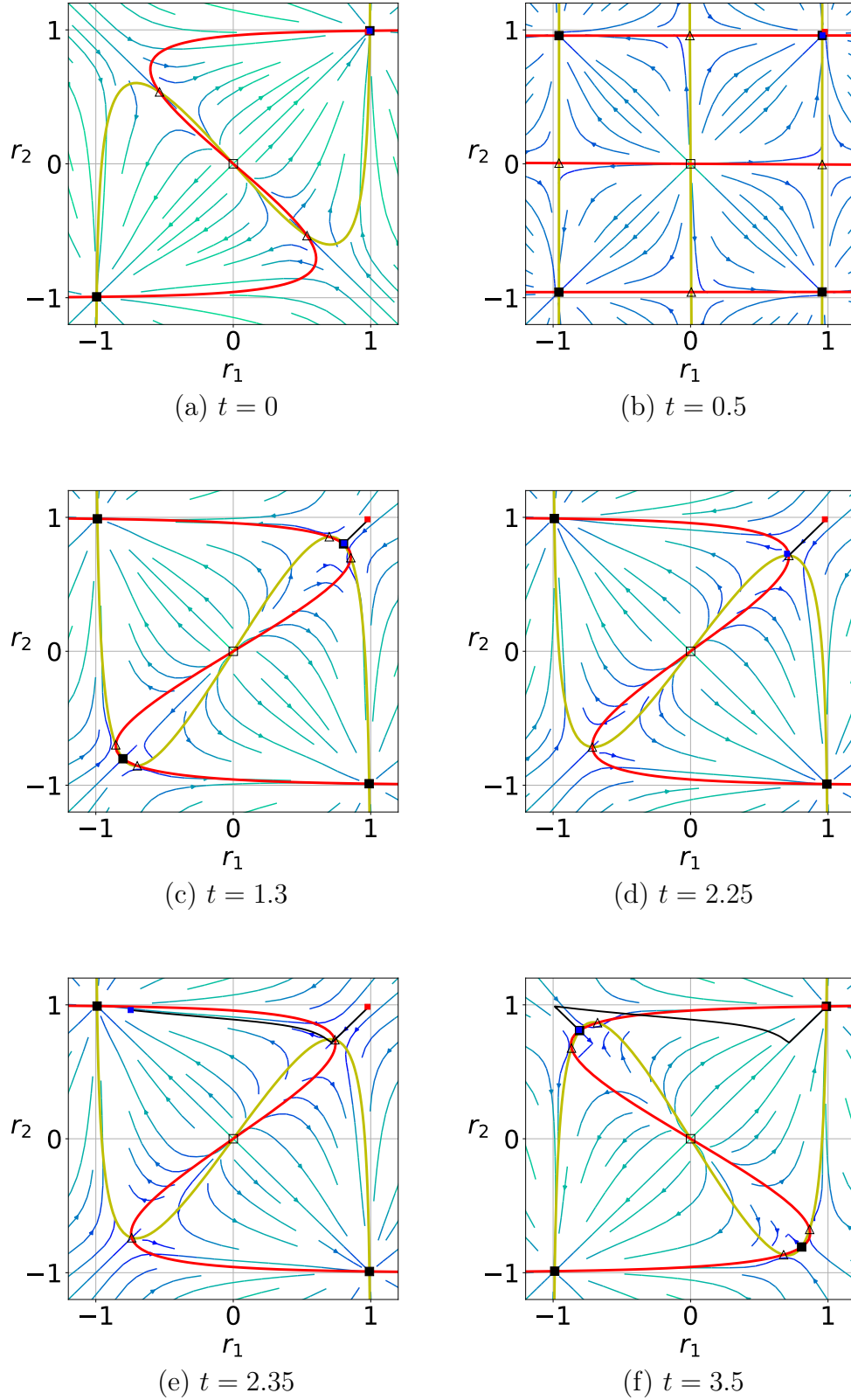


Figure 3.5: Each panel is a projection of the full system trajectory onto the r_1, r_2 plane for different values of time, including fast subsystem phase portrait, with nullclines r_1 and r_2 (in yellow and red respectively), corresponding to $m_c(t)$. Black squares and triangles are nodes and saddles respectively. Filled shapes are stable, unfilled are unstable. The smaller red square indicates the starting point of the trajectory, the small blue square indicates the location of the trajectory at the current time. Panels (a)-(f) correspond to $t = 0, 0.5, 1.3, 2.25, 2.35, 3.5$.

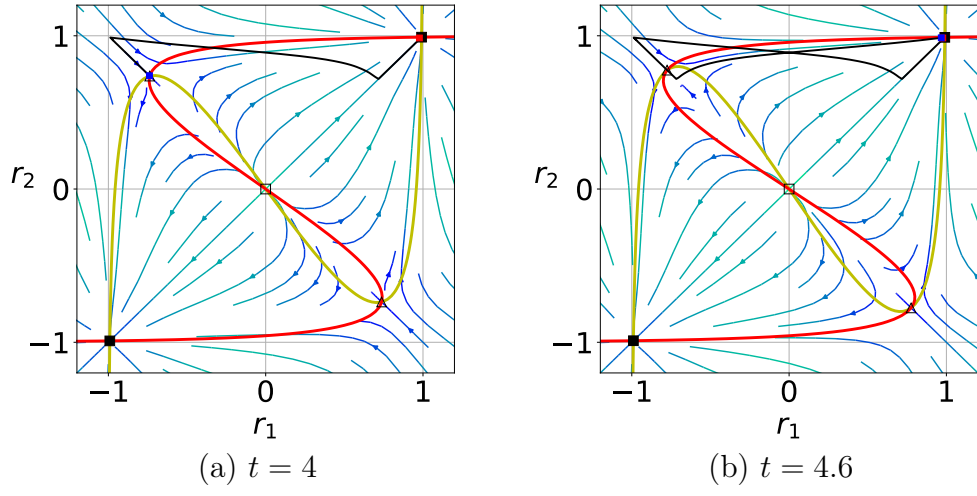


Figure 3.6: Figure 3.5 continued, including times $t = 4, 4.6$ in panels (a) and (b) respectively.

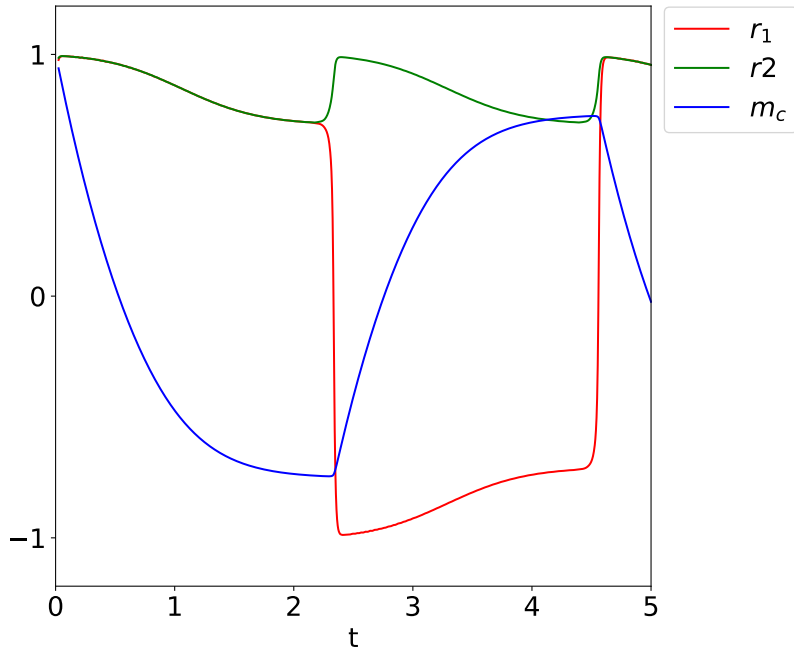


Figure 3.7: Time trace for the trajectory shown in figures 3.5 and 3.6.

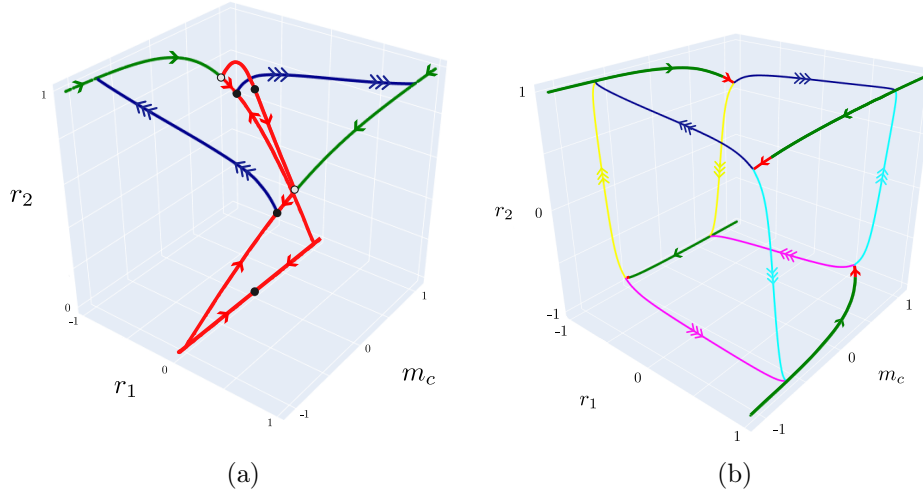


Figure 3.8: Panel (a) shows the region of critical manifold above and contained in the planes $r_1 = r_2$ and $r_1 = -r_2$, slow flow indicated by arrows, red and green curves are stable and saddle type respectively. Grey circles indicate points where normal hyperbolicity is lost. Black circles are saddle equilibria for the full system. Blue curves are portions of the unstable manifolds for saddles they are connected to, triple arrows indicate fast motion. Panel (b) regions of the critical manifold which are close to a limit cycle. Blue, cyan, yellow and magenta curves are portions of the limit cycles which follow unstable manifolds of saddles, triple arrows indicate speed of flow along these manifolds.

3.3.1 Symmetries and Equivariance

The equations (3.7) are not equivariant under $\gamma_{(12)}$, due to the difference in sign between the m_c term in the equations for r_1 and r_2 . However it is still equivariant with respect to $\Gamma = \{I, \gamma_{--+}, \gamma_{-+-}, \gamma_{+--}\}$. The invariant fixed point subspaces associated with the subgroups of this group are

$$\begin{aligned} \text{Fix}(\gamma_{--+}) &= \{(0, 0, m_c) : m_c \in \mathbb{R}\}, \\ \text{Fix}(\gamma_{-+-}) &= \{(0, r_2, 0) : r_2 \in \mathbb{R}\}, \\ \text{Fix}(\gamma_{+--}) &= \{(r_1, 0, 0) : r_1 \in \mathbb{R}\}. \end{aligned} \tag{3.8}$$

Like the last two examples, $(0, 0, 0)$, $(\pm r^*, 0, 0)$ and $(0, \pm r^*, 0)$ are equilibria of the full system for any parameter values. In contrast to the last examples, the half planes are no longer invariant subspaces.

3.3.2 Fast Subsystem

The fast sub-system dynamics are given by,

$$\begin{aligned}\dot{r}_1 &= -r_1 + \tanh(mr_1 + m_cr_2), \\ \dot{r}_2 &= -r_2 + \tanh(mr_2 + -m_cr_1).\end{aligned}$$

The nullclines for this system are

$$\begin{aligned}\dot{r}_1 : r_2 &= \frac{1}{m_c}I(r_1, m), \\ \dot{r}_2 : r_1 &= -\frac{1}{m_c}I(r_2, m),\end{aligned}$$

So for a given value of $m_c \neq 0$, the r_1 nullcline is the same as it is for the case with the same synapse type, e.g., the case covered in section 3.1. However in this case where we have an anti-Hebbian synapse together with a Hebbian synapse, the r_2 nullcline is the result of taking the r_2 nullcline of the fast-subsystem from section 3.1, and reflecting it across the r_2 axis. An illustration of the bifurcations which occur when increasing the m_c parameter from 0 is shown in 3.1. If $m_c = 0$, the phase portrait is the same as the case with the same synapses (panel (a)). For $m_c > 0$, the r_1 nullcline has an N shape and the r_2 has a backwards S shape panel (b). As m_c increases from zero, the saddles move towards the stable nodes in an anti clockwise orientation (panels (b) and (c)). Eventually, all fixed points other than the origin are destroyed in simultaneous saddle-node infinite period (SNIPER) bifurcations [30] (panel (d)). This gives rise to a clockwise oriented limit cycle which attracts all initial conditions except the origin, which is an unstable spiral. Due to the symmetry in the equations, the fixed points undergo a similar bifurcation upon decreasing m_c from 0. This time the saddles move toward the nodes in a clockwise orientation, and the limit cycle which appears after the SNIPER bifurcations has anti-clockwise orientation. The bifurcation diagram is shown in figure 3.10. Through approximately following these limit cycles, which exist in the fast-subsystem, the full system is able to switch between branches of the critical manifold. This is discussed in more detail in the next section.

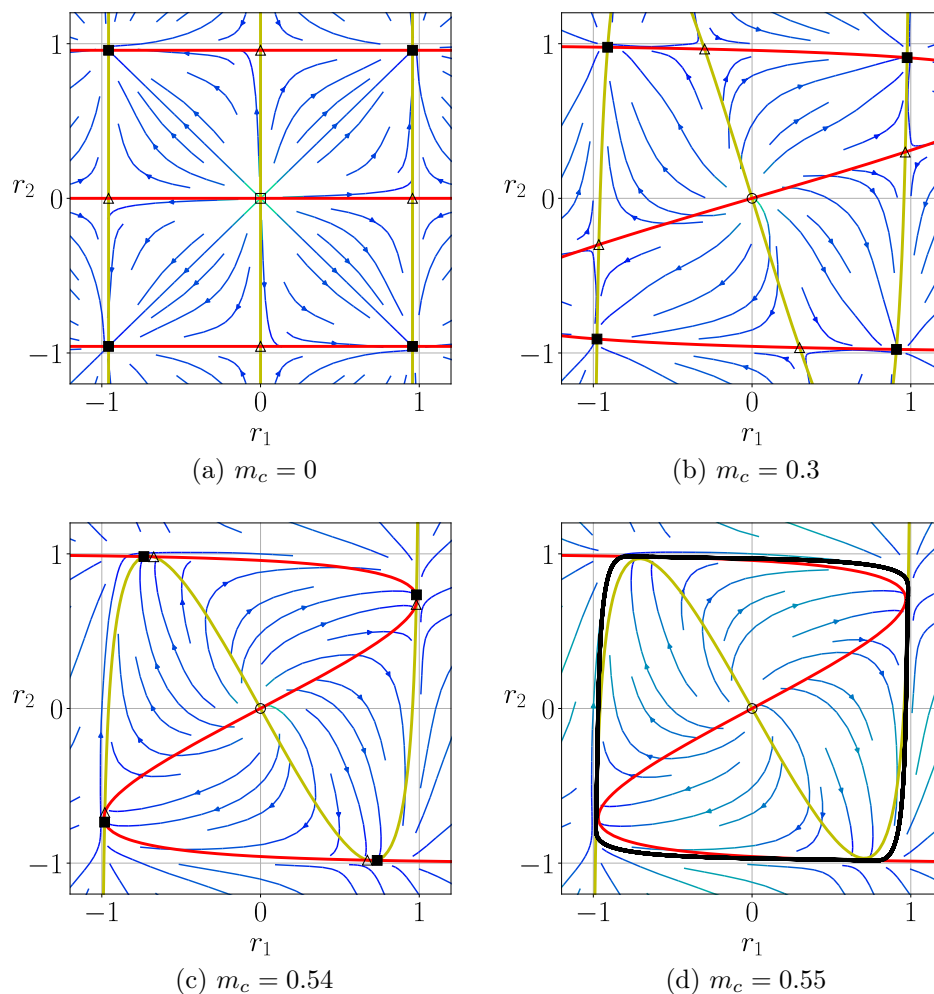


Figure 3.9: Phase portraits of the fast subsystem for $m = 2$, $\epsilon = 0.01$ and increasing values increasing m_c . Panels (a)-(d) are for $m_c = 0, 0.3, 0.54, 0.55$ respectively.

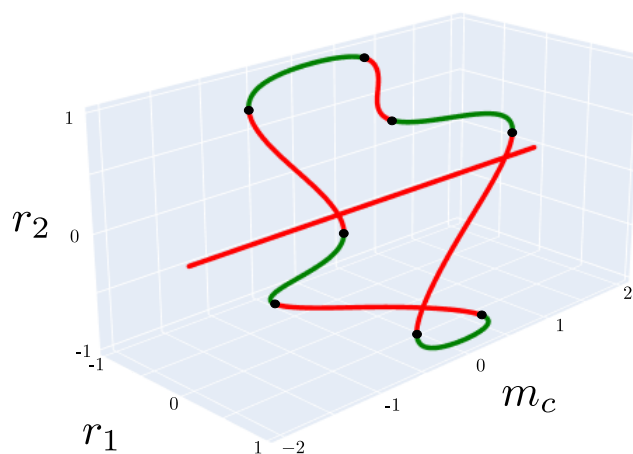


Figure 3.10: Bifurcation diagram for fixed points of the fast subsystem. Black points indicate SNIPER bifurcations.

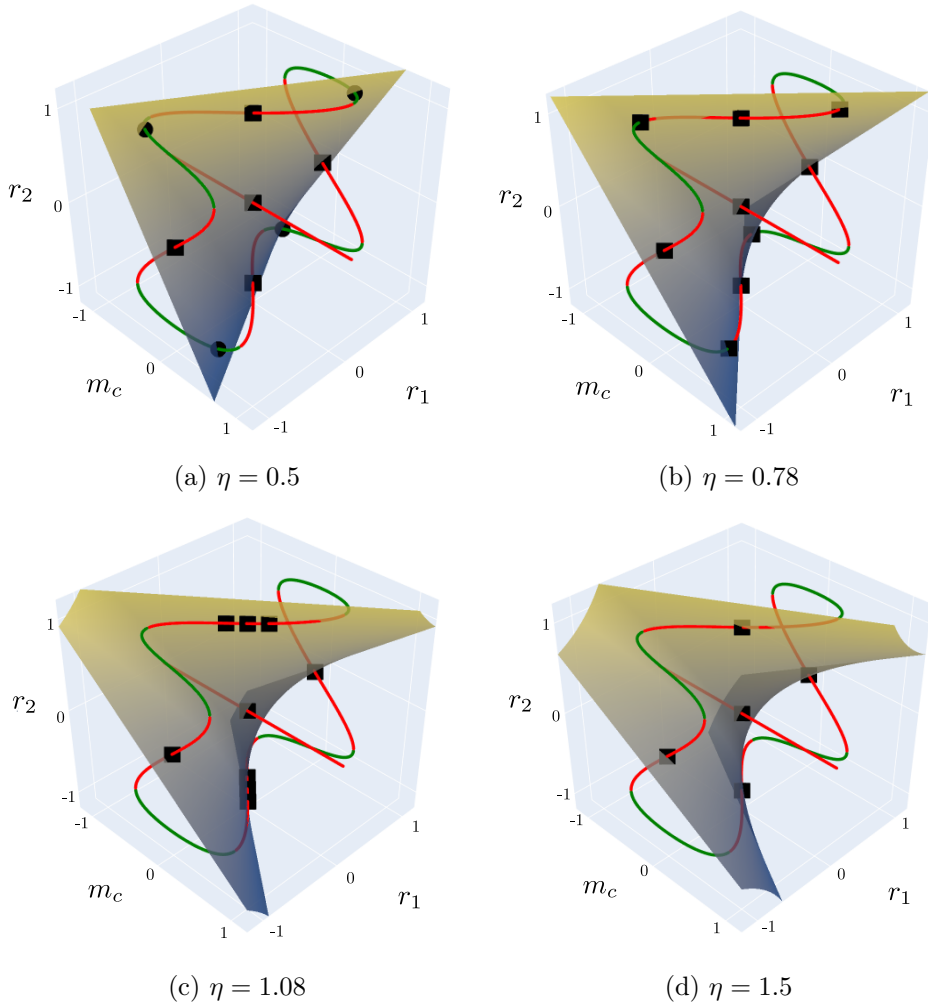


Figure 3.11: Fixed points in the full system for increasing values of η with $m = 2$, $\epsilon = 0.01$. Red and green curves are stable and unstable branches of the critical manifold respectively. Surface is the nullsurface $\dot{m}_c = 0$, squares and circles are unstable and stable equilibria respectively.

3.3.3 Full System Dynamics

For sufficiently small η , the full system has nine equilibria in total, consisting of the five points $(0, 0, 0)$, $(\pm r^*, 0, 0)$ and $(0, \pm r^*, 0)$, as well as four stable equilibria where the surface $m_c = -\eta r_1 r_2$ intersects the stable branches of the critical manifold. In this case, all initial conditions converge to one of these stable fixed points. As η increases, the stable equilibria move along the stable branches toward the fold points of the critical manifold, marked as black points in figure 3.10. Eventually, these fixed points cross over into the unstable branches of C_0 , each undergoing a subcritical hopf bifurcation as it does so. Eventually these fixed points merge with the equilibria $(0, \pm r^*, 0)$ in two simultaneous pitchfork bifurcations. Figure 3.11 shows the system fixed points for increasing values of η .

There is a seemingly sudden emergence of two large amplitude limit cycles. One travels along the upper branches, periodically making fast transitions between them at the fold points of the critical manifold by approximately following the limit cycle of the fast subsystem created after the fold. The other does the same but travels along the lower branches. Each panel of figures 3.12 and 3.13 shows the projection onto the r_1, r_2 plane of the evolution along the upper limit cycle up to a certain time t . The fast subsystem phase portrait corresponding to $m_c(t)$, including r_1 and r_2 nullclines (in yellow and red respectively), is also plotted to show how the full system transitions from the slow flow to the fast flow, as a result of bifurcations in the fast subsystem caused by the slow variable. In Fig 3.12 panel (a), the initial condition is close to the stable branch of the critical manifold. The trajectory travels along this branch as seen in panels (b) and (c), until in panel (d) it reaches a fold point in the critical manifold, which corresponds to a saddle node bifurcation in the fast subsystem. The trajectory then quickly follows the fast subsystem dynamics while m_c stays approximately constant, shown in panel (d). In panel (e), the trajectory of the full system has reached another branch of the critical manifold. In the projection this is seen from $(r_1(t), r_2(t))$ tracking the stable node in the fast subsystem dynamics. In panel (f), as well as panels (a) and (b) of figure 3.13, the trajectory travels back to the original branch of the critical manifold due to a saddle node bifurcation in the fast subsystem caused by the slow variable. The trajectory cycles through this periodic orbit for all future time. The time trace of this trajectory is shown in figure 3.14. Sharp changes in the values of r_1 and r_2 correspond to transitions between branches of the critical manifold, flat parts of the r_1 and r_2 curves correspond to periods of time spent close to the critical manifold. Figure 3.15 panel (a) shows how the upper limit cycle (plotted in magenta) approximately follows the slow manifold (plotted in green) and the upper portion of the limit cycles present in the fast subsystem for fixed m_c . Specifically, the upper portion of the limit cycles for $m_c = \pm 0.58$ are plotted in the planes $m_c = \pm 0.58$ as blue curves. Panel (b) shows both limit cycles present in the full system (upper in magenta, lower in cyan), as well as the entire fast subsystem limit cycles, alongside the stable regions of the critical manifold.

When the two neurons are connected with a Hebbian synapse from r_1 to r_2 and an anti Hebbian synapse from r_2 to r_1 , the result was also relaxation oscillations where one neuron switched between off and on states and the other stayed in a single, fixed on or off state. It is only possible for the neuron with the incoming anti-Hebbian synapse to switch. This is in contrast to the possible limit cycles in the case with only anti Hebbian synapses, where there existed limit cycles where either of the neurons switched. If the two neurons are in the same on/off state for sufficiently

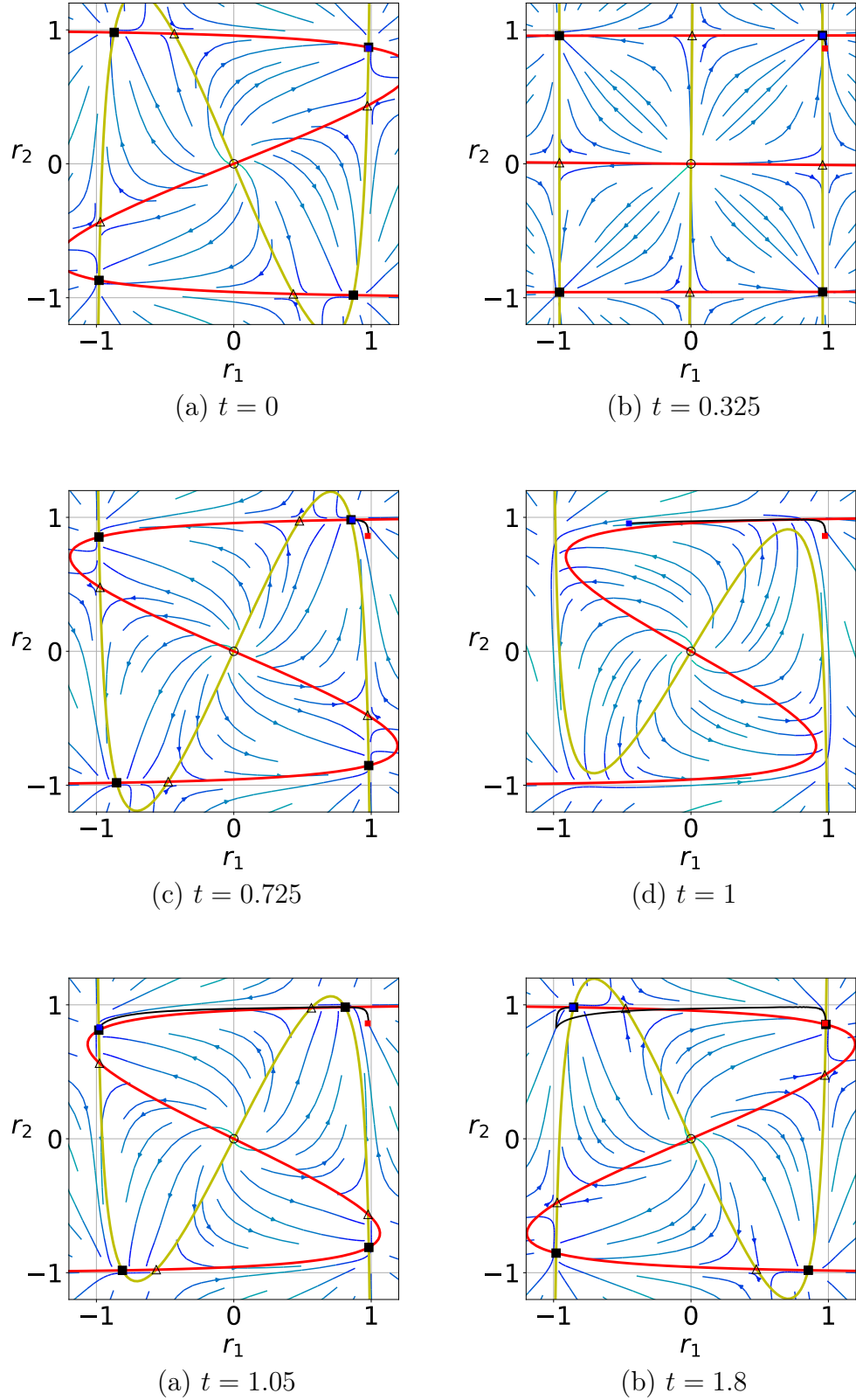


Figure 3.12: Each panel is a projection of the full system trajectory onto the r_1, r_2 plane for different values of time, including fast subsystem phase portrait, with nullclines r_1 and r_2 (in yellow and red respectively), corresponding to $m_c(t)$. Squares triangles and triangles are nodes and saddles. Filled shapes are stable, unfilled are unstable. Panels (a)-(f) correspond to $t = 0, 0.325, 0.725, 1, 1.05, 1.8$.

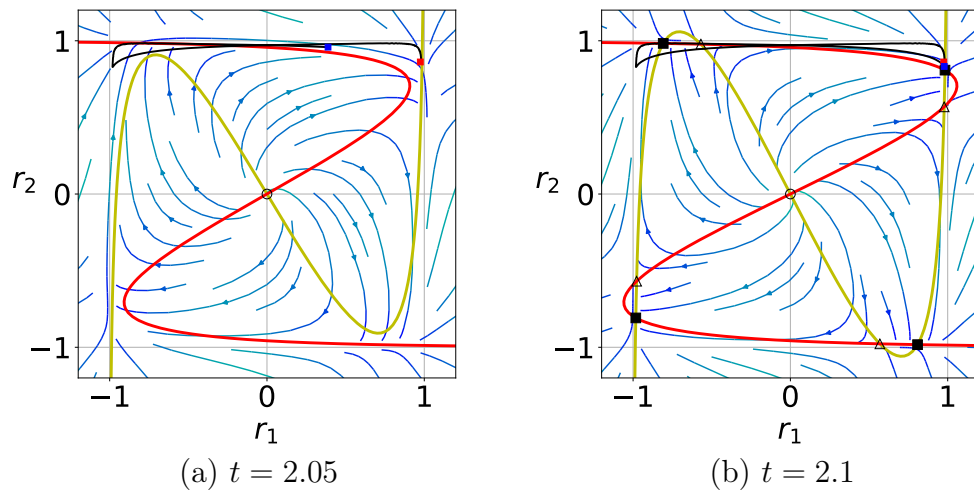


Figure 3.13: Figure 3.5 continued, including times $t = 2.05, 2.1$ in panels (a) and (b) respectively.

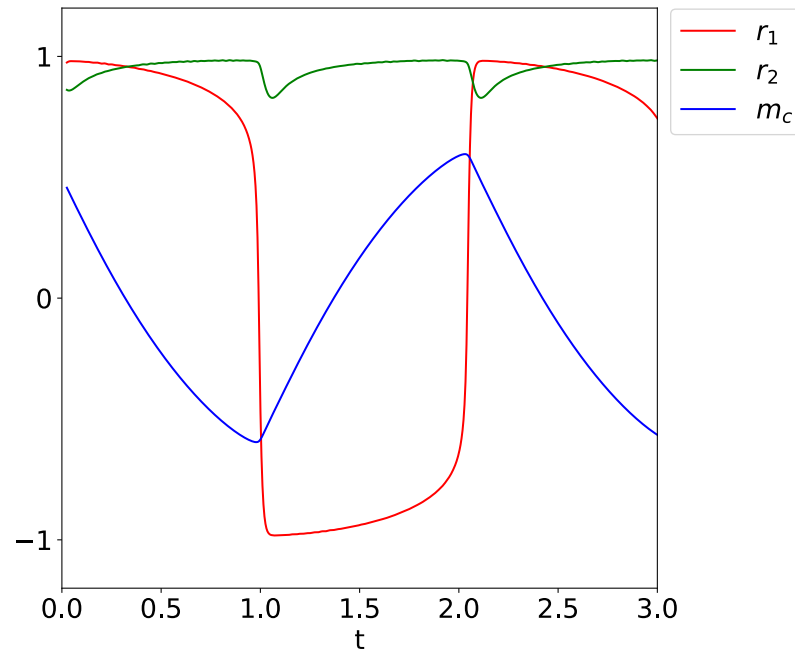


Figure 3.14: Time trace for the trajectory shown in figures 3.12 and 3.13

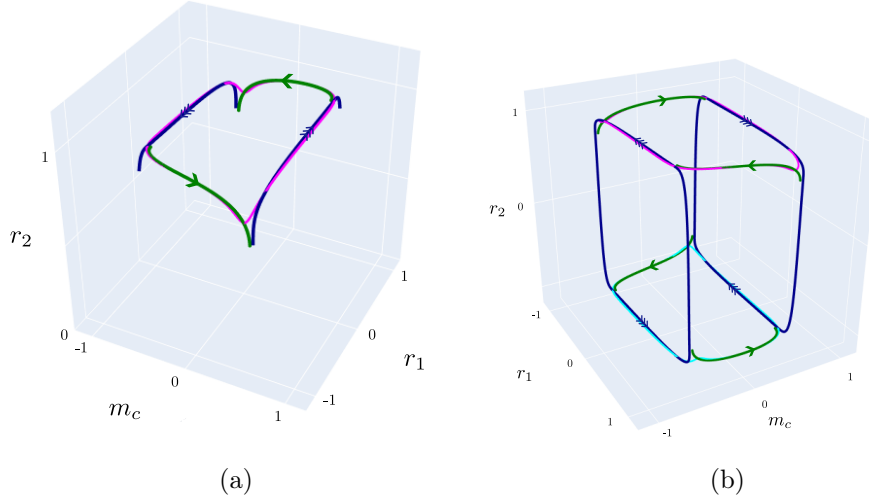


Figure 3.15: Panel (a) shows, for $m = 2, \epsilon = 0.01$ and $\eta = 1.5$, the stable branches of the critical manifold in green with slow flow indicated by arrows, portions of the limit cycles for $m_c = \pm 0.58$ are plotted in the planes $m_c = \pm 0.58$ as blue curves, triple arrows indicate motion on the fast timescale for nearby trajectories, upper limit cycle shown in purple. Panel (b) shows both limit cycles present in the full system in magenta and cyan, as well as the entire fast subsystem limit cycles and the stable regions of the critical manifold.

long, we will have $m_c < 0$, so r_2 begins pushing r_1 to the other state, while r_1 reinforces r_2 being on the current shared on/off state. If m_c decreases enough, r_1 will be pushed to make a fast transition to the other on/off state. When they are on different on/off states for sufficiently long, r_2 begins pulling r_1 back to the positive state while r_2 is still being pushed into its current on state. As m_c increases enough, r_1 transitions back to the on state. In this way, the two neurons transition between being in the same state and being in different states.

Chapter 4

Cluster Synchronization

In this chapter we consider an example where a nine dimensional adaptive network makes a transition from having stable cluster synchronization states to transient cluster synchronization as the proportion of neurons connected by mutual anti Hebbian synapses increases. In this example the initial conditions and parameter values η and m are fixed while the parameters α_{ij} are varied. Specifically, we let $\eta = 0.125$ and $m = 1.1$. The initial node states are chosen uniformly at random from the interval $(-1, 1)$ and the initial weights are chosen uniformly at random from $(-\eta, \eta)$. Then for different choices of α_{ij} , we numerically solve the adaptive network equations given by,

$$\begin{aligned} \epsilon \dot{r}_i &= -r_i + \tanh \left(mr_i + \sum_{j \neq i} m_{ij} r_j \right), \quad i = 1, \dots, 9 \\ \dot{m}_{ij} &= \alpha_{ij} \eta r_i r_j - m_{ij}, \quad i \neq j \end{aligned}$$

using the SciPy `solve_ivp` function over the time interval $[0, 200]$ with a step size 5×10^{-4} . We solve the system for ten different choices of the parameters α_{ij} . The first configuration has all Hebbian connections (i.e., $\alpha_{ij} = 1 \forall i, j$). Each configuration after this increases the number of nodes with all incoming anti-Hebbian synapses by one. So the parameters α_{ij} for the n^{th} configuration are given by

$$\alpha_{ij} = \begin{cases} 1 & \text{for } i \leq 9 - n, \\ -1 & \text{for } i > 9 - n, \end{cases}$$

for $n \in 0, \dots, 9$. Figures 4.1, 4.2, 4.3 show the results of these calculations. Each panel is a colour plot for the sign of the neuron state r_i vs time, with yellow and purple corresponding to negative (off state) and positive (on state) respectively. The

trajectories are only plotted for $t \geq 50$ to remove any transience.

Fig 4.1 panel (a) shows the case with all Hebbian synapses, in which all neurons settle to a state where no switching occurs, as expected from the two dimensional case which always converged to a fixed point. In panel (b), only r_9 has incoming anti Hebbian synapses and is the only neuron which periodically switches between on and off states. This is again similar to the two dimensional case, where only the neuron with an incoming anti Hebbian weight would switch. In panel (c), only the neurons r_9 and r_8 with all incoming anti Hebbian synapses switch between on and off states, but appear to do so in a phase-locked state. Specifically, r_8 appears to be shifted less than half a period off of r_9 , so that they alternate between being in the same on/off state and being in different on/off states. This is also similar to the behaviour of the two dimensional system with mutual anti Hebbian synapses, which alternated in and out of a state of complete synchrony. Panel (d) is similar to panel (c), where all neurons with incoming anti Hebbian connections r_7, r_8 and r_9 form a frequency cluster, where each pair of neurons are phase shifted.

Fig. 4.2 panel (a) where the neurons r_6, r_7, r_8 and r_9 have incoming anti Hebbian synapses, shows the emergence of two frequency clusters, one with a lower frequency consisting of r_7 and r_8 and the other with a higher frequency consisting of r_6 and r_9 . In panel (b), the neurons with anti-Hebbian synapses form two frequency clusters, r_6, r_5 and r_7, r_8, r_9 , in order of increasing frequency. In panel (c), neurons with anti hebbian synapses split into three frequency clusters r_4, r_5 , r_6, r_7 , and r_8, r_9 , in order of increasing frequency. In panel (d), there is a transition from neurons converging to stable frequency clustered states, to transient frequency clustering. For example, neurons r_7 and r_3 form a cluster at $t \approx 60$, which breaks down at $t \approx 85$, and neurons r_9, r_8 and r_6 form a cluster at $t \approx 80$ which breaks down at $t \approx 90$. This transient frequency clustering persists in the cases when all but neuron r_1 have incoming anti-Hebbian synapses, and when all neurons have incoming anti-Hebbian synapses. These cases are shown in Fig 4.3 panels (a) and (b) respectively.

Fig. 4.4 shows the trajectories of panels (a) and (b) over the longer time period $t \in [50, 200]$, showing the episodes of transient synchronisation continuing to reoccur.

The features of parameter dependent stable cluster synchronisation and transient cluster synchronisation shown in this example indicates the potential for this system's use as a simple phenomenological model of certain features observed in neural systems, such as coordinated task switching or synchronisation of large regions of the brain during epileptic seizures. Further study is required to better understand this model's dependence on the parameters α_{ij} for $N > 2$, which clearly has a significant impact on the network dynamics. In the two dimensional case, the choice of

α_{ij} gave rise to certain symmetries which was important for the attractors present in the system. In light of this, a natural next step would be investigating the connection between the symmetry of higher dimensional networks dictated by the choice of α_{ij} , and the resulting attractors of the system. It was also shown by Ashwin et. al. that networks of identical coupled phase oscillators with heteroclinic networks can exhibit transient synchronisation states [31]. Thus, it would be instructive to investigate whether this system has heteroclinic networks, and if this is the case, what their effect on the network dynamics is. A better understanding of these features could improve our understanding enough to allow the design of control methods for the system, so that it could be in neuromorphic computing devices.

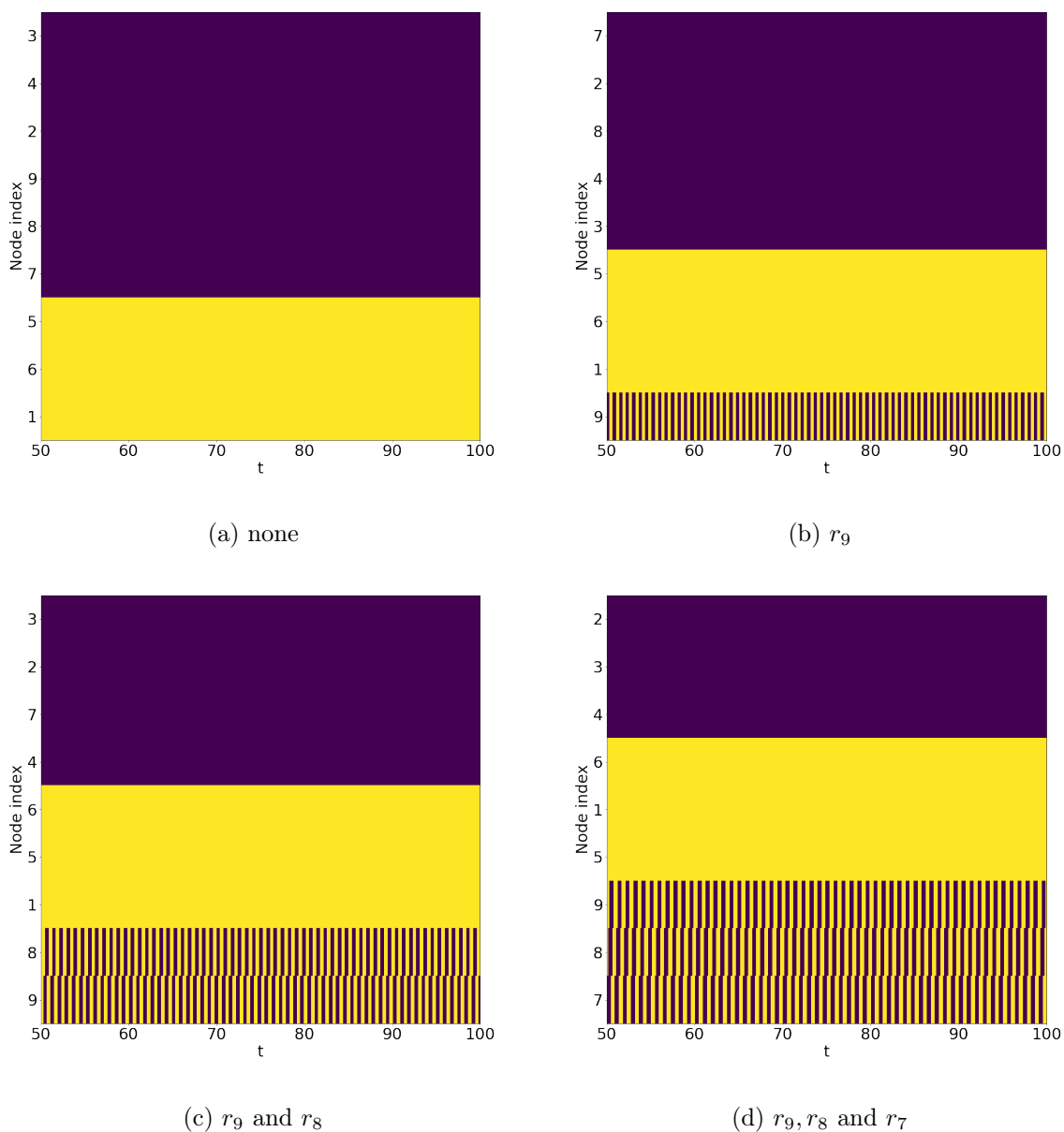


Figure 4.1: Colour plots for the sign of the neuron state r_i vs time for different configurations of Hebbian and anti Hebbian synapses. Yellow and purple correspond to negative (off state) and positive (on state) respectively. Panel (a) has all Hebbian synapses. In panels b-d, the neurons indicated in the panel captions have been chosen to have all incoming anti Hebbian synapses.

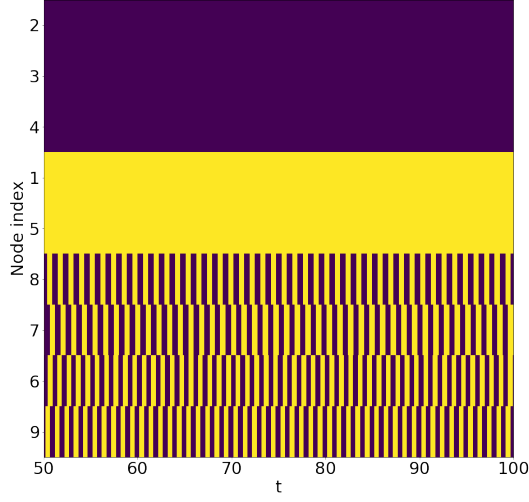
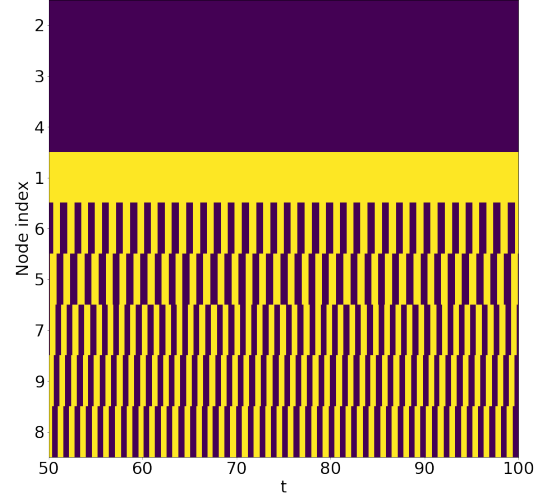
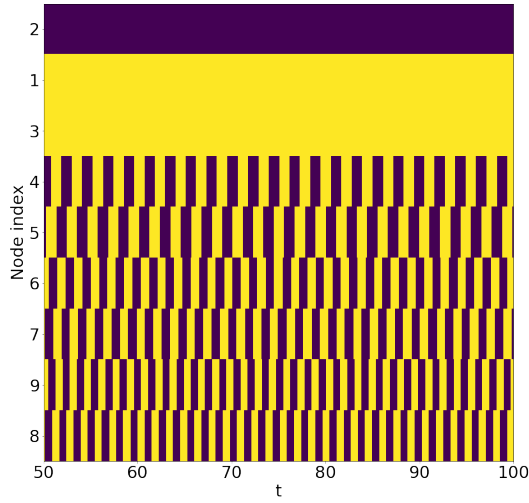
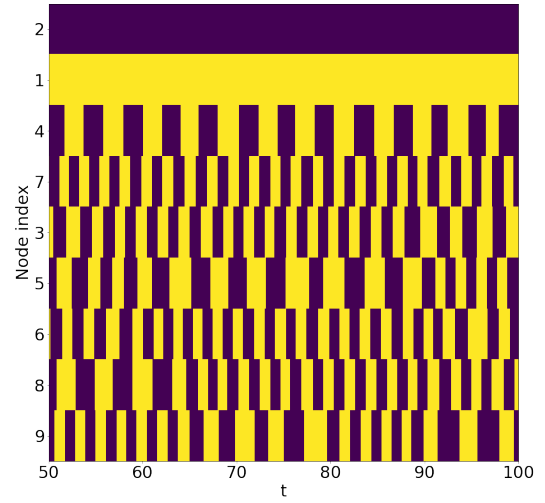

 (a) r_9, r_8, r_7 and r_6

 (b) r_9, r_8, r_7, r_6 and r_5

 (c) r_9, r_8, r_7, r_6, r_5 and r_4

 (d) $r_9, r_8, r_7, r_6, r_5, r_4$ and r_3

Figure 4.2: Colour plots for the sign of the neuron state r_i vs time for different configurations of Hebbian and anti Hebbian synapses. Yellow and purple correspond to negative (off state) and positive (on state) respectively. In panels each panel, the neurons indicated in the panel caption have been chosen to have all incoming anti Hebbian synapses.

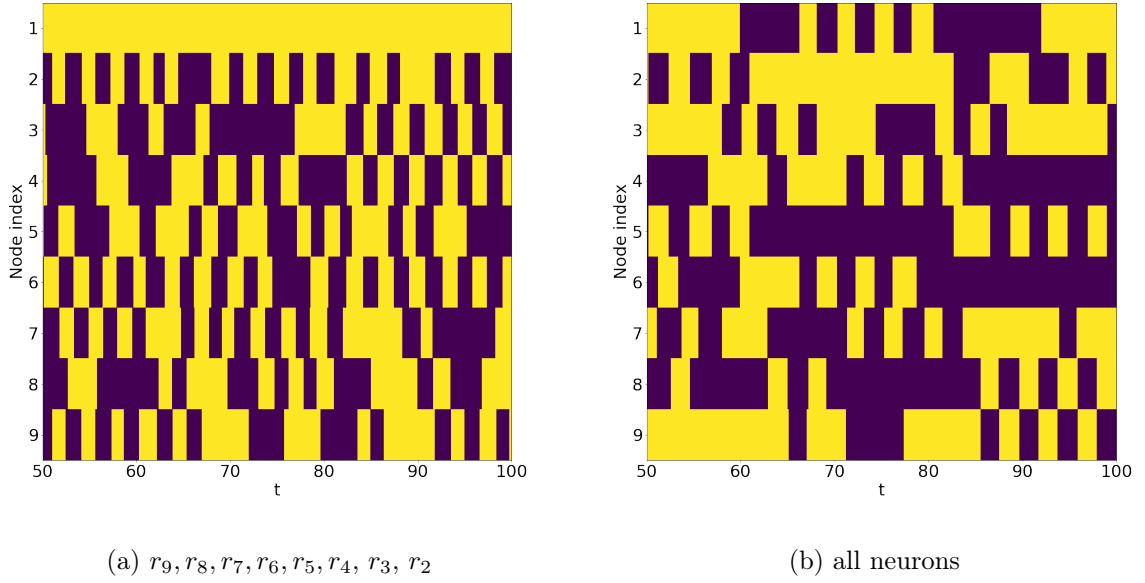


Figure 4.3: Colour plots for the sign of the neuron state r_i vs time for different configurations of Hebbian and anti Hebbian synapses. Yellow and purple correspond to negative (off state) and positive (on state) respectively. In panels each panel, the neurons indicated in the panel caption have been chosen to have all incoming anti Hebbian synapses.

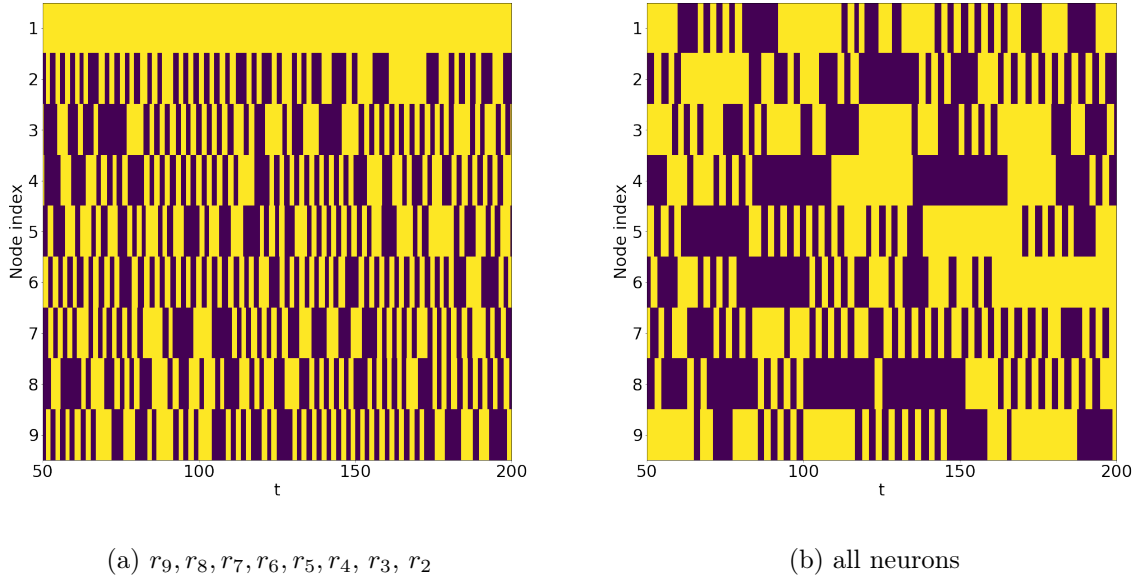


Figure 4.4: Colour plots for the sign of the neuron state r_i vs time over a longer time interval [50, 200], for different configurations of Hebbian and anti Hebbian synapses. Yellow and purple correspond to negative (off state) and positive (on state) respectively. In panels each panel, the neurons indicated in the panel caption have been chosen to have all incoming anti Hebbian synapses.

Chapter 5

Conclusion

In this thesis a simple, phenomenological model of a neural network with plasticity is presented. This was motivated by the desire for a minimalistic model capable of replicating certain features of brain dynamics, as such a model would not only be instructive from a neuroscientific perspective, but also has the potential to motivate new machine learning algorithms or act as a template for implementation of neuromorphic computing devices.

The model was chosen to account for three essential ingredients for information processing in the brain,

- signal transmission between cells,
- complex network structure,
- the ability to alter this structure in response to neuron activity via plasticity.

This was done by taking one of the simplest representations of a neural network which evolves in continuous time, the continuous time recurrent neural network, and augmenting it with a plasticity rule which changes the network connectivity structure. The plasticity rule was chosen from the class of Hebbian learning rules, in which the synaptic connection between two neurons evolves continuously as a function of their correlation in the recent past. We consider the case where the neurons evolve on a faster timescale than the synapses, leading to a slow-fast adaptive dynamical network.

As a first step in the analysis of this model, a bifurcation analysis of the individual neuron dynamics is conducted. This analysis shows that the individual neurons undergo a cusp bifurcation under variation of the parameters m and I . As a result of this, for fixed $m > 1$, the single neuron goes through two saddle node bifurcations by varying the input parameter I , and has two branches of stable equilibria, one

consisting of positive fixed points, and the other of negative fixed points. By varying the parameter I across threshold values determined by m , the system can switch from converging to a positive equilibrium or a negative equilibrium. This leads to the interpretation of its behaviour as a type of dynamic switch that is either in an 'off' state if negative or an 'on' state if positive, based on the value of a time varying input.

Next the coupling is introduced, so that the input to each neuron is a weighted sum of the other neuron states. Through the coupling, each neuron acts on the others by either attracting them to its on/off state or repelling them from it.

The Hebbian learning rule is then studied, in which the weights of the network evolve based on the correlated activity of the neurons in the recent past. Adding the Hebbian learning rule adds extra parameters to the system. The learning rate η controls the speed of synaptic dynamics as well as the magnitude which the weights can evolve to a time progresses. The other parameters α_{ij} , determine whether the synapse from neuron j to neuron i is Hebbian or anti-Hebbian. A Hebbian synapse will cause r_j to attract r_i if they both tend to be in the same state, and repel r_i if they both tend to be in different states. An anti Hebbian synapse will cause r_j to repel r_i if they both tend to be in the same state, and repel r_i if they both tend to be in different states.

As a first step in studying the network with plasticity, we begin with networks of two nodes with $m = 2$ and $\epsilon = 0.01$. Then consider the three possible configurations of synapse types; two Hebbian synapses, two anti Hebbian synapses and a Hebbian synapse and an anti Hebbian synapse. With two Hebbian synapses, the system converges to one of four fixed points which correspond to the neurons being in one of the four possible combinations of on/off states. With two anti Hebbian synapses, for sufficiently large η there are four symmetric relaxation oscillation type limit cycles. On each limit cycle, one neuron switches between on and off states and the other stays in a fixed on/off state. The final case, where the synapse from r_1 to r_2 is Hebbian and the synapse from r_2 to r_1 is anti Hebbian, there are two relaxation oscillation type limit cycles for sufficiently large η . On each limit cycle, r_2 stays in a fixed on/off state, while r_1 switches.

Finally, the effect of having different proportions of Hebbian and anti Hebbian synapses is shown for an example with nine neurons, with parameter values m, η, ϵ and initial conditions are fixed. Then the system is evolved using ten different configurations of Hebbian and anti Hebbian synapses. Initially, all synapses are Hebbian, after this increases the number of nodes with all incoming anti-Hebbian synapses by one until all synapses are Hebbian. We observe that there is a transition from stable

fixed point dynamics, to stable frequency cluster synchronization, to transient synchronization. The features of parameter dependent stable cluster synchronisation and transient cluster synchronisation shown in this example indicates the potential for this system's use as a simple phenomenological model of certain features observed in neural systems, such as coordinated task switching or synchronisation of large regions of the brain during epileptic seizures. Further study is required to better understand this model's dependence on the parameters α_{ij} for $N > 2$, which clearly has a significant impact on the network dynamics. In the two dimensional case, the choice of α_{ij} gave rise to certain symmetries which was important for the attractors present in the system. In light of this, a natural next step would be investigating the connection between the symmetry of higher dimensional networks dictated by the choice of α_{ij} , and the resulting attractors of the system. It was also shown by Ashwin et. al. that networks of identical coupled phase oscillators with heteroclinic networks can exhibit transient synchronisation states [31]. Thus, it would be instructive to investigate whether this system has heteroclinic networks, and if this is the case, what their effect on the network dynamics is. A better understanding of these features could improve our understanding enough to allow the design of control methods for the system, so that it could be in neuromorphic computing devices.

Bibliography

- [1] Mark Bear, Barry Connors, and Michael A. Paradiso. *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain, Enhanced Edition*. Jones & Bartlett Learning, Mar 2020.
- [2] Siusana Kweldju. Neurobiology research findings: How the brain works during reading. *Pasaa: Journal of language teaching and learning in Thailand*, 50:125–142, 2015.
- [3] Bogdan Feliks Kania, Danuta Wrońska, and Dorota Zięba. Introduction to neural plasticity mechanism. *Journal of Behavioral and Brain Science*, 7(2):41–49, 2017.
- [4] 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.
- [5] J.B. Reece, M.R. Taylor, N.A. Campbell, E.J. Simon, and J.L. Dickey. *Campbell Biology: Concepts & Connections*, chapter 28, pages 563–585. Cummings, 2012.
- [6] Pedro Mateos-Aparicio and Antonio Rodríguez-Moreno. The impact of studying brain plasticity. *Frontiers in cellular neuroscience*, 13:66, 2019.
- [7] Terrence J Sejnowski. The unreasonable effectiveness of deep learning in artificial intelligence. *Proceedings of the National Academy of Sciences*, 117(48):30033–30038, 2020.
- [8] A. Ceni. *Interpreting Multi-stable Behaviour in Input-driven Recurrent Neural Networks*, chapter 2, pages 11–26. University of Exeter, 2021.
- [9] II Arikpo, FU Ogban, and IE Eteng. Von neumann architecture and modern computers. *Global Journal of Mathematical Sciences*, 6(2):97–103, 2007.
- [10] Catherine D Schuman, Thomas E Potok, Robert M Patton, J Douglas Birdwell, Mark E Dean, Garrett S Rose, and James S Plank. A survey of neuromorphic

- computing and neural networks in hardware. *arXiv preprint arXiv:1705.06963*, 2017.
- [11] T.P. Trappenberg. *Fundamentals of Computational Neuroscience*, chapter 4, pages 3–31. Oxford University Press, 2023.
 - [12] T.P. Trappenberg. *Fundamentals of Computational Neuroscience*, chapter 4, pages 65–97. Oxford University Press, 2023.
 - [13] B. Cessac. A view of neural networks as dynamical systems. *International Journal of Bifurcation and Chaos*, 20(06):1585–1629, Jun 2010.
 - [14] Rico Berner, Thilo Gross, Christian Kuehn, Jürgen Kurths, and Serhiy Yanchuk. Adaptive dynamical networks. *arXiv preprint arXiv:2304.05652*, 2023.
 - [15] Neil A Campbell and Jane B Reece. *Biology*, chapter 48, pages 1047–1062. Pearson Education India, 2005.
 - [16] Anil Gulati. Understanding neurogenesis in the adult human brain, 2015.
 - [17] Peter Dayan, Laurence F Abbott, et al. *Theoretical neuroscience: computational and mathematical modeling of neural systems*, chapter 7, pages 231–278. MIT Press, 2005.
 - [18] D.O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Taylor & Francis, 2005.
 - [19] Wulfram Gerstner and Werner M Kistler. Mathematical formulations of hebbian learning. *Biological cybernetics*, 87(5):404–415, 2002.
 - [20] T. Trappenberg. *Fundamentals of Computational Neuroscience*. OUP Oxford, 2010.
 - [21] Randall D Beer. On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior*, 3(4):469–509, 1995.
 - [22] E.M. Izhikevich. *Dynamical Systems in Neuroscience*, chapter 2, pages 37–44. Computational neuroscience Dynamical systems in neuroscience. MIT Press, 2007.
 - [23] Zhixin Lu, Brian R Hunt, and Edward Ott. Attractor reconstruction by machine learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(6), 2018.

- [24] Y.A. Kuznetsov. *Elements of Applied Bifurcation Theory*, chapter 8, pages 327–439. Applied Mathematical Sciences. Springer International Publishing, 2023.
- [25] Peter Ashwin, Sebastian Wieczorek, Renato Vitolo, and Peter Cox. Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1962):1166–1184, 2012.
- [26] Paul E O’Keeffe and Sebastian Wieczorek. Tipping phenomena and points of no return in ecosystems: beyond classical bifurcations. *SIAM Journal on Applied Dynamical Systems*, 19(4):2371–2402, 2020.
- [27] Ann Kristin Klose, Volker Karle, Ricarda Winkelmann, and Jonathan F Donges. Emergence of cascading dynamics in interacting tipping elements of ecology and climate. *Royal Society open science*, 7(6):200599, 2020.
- [28] Christian Kuehn et al. *Multiple time scale dynamics*, volume 191. Springer, 2015.
- [29] M. Golubitsky and D.G. Schaeffer. *Singularities and Groups in Bifurcation Theory: Volume I*. Applied Mathematical Sciences. Springer New York, 2013.
- [30] S.H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. CRC Press, 2018.
- [31] Peter Ashwin, Gábor Orosz, John Wordsworth, and Stuart Townley. Dynamics on networks of cluster states for globally coupled phase oscillators. *SIAM Journal on Applied Dynamical Systems*, 6(4):728–758, 2007.