
FIXED POINTS AND LINEAR STABILITY OF A WEIGHT-DEPENDENT BCM RULE

DISSERTATION SUBMITTED TO THE SCHOOL OF PSYCHOLOGY AT THE UNIVERSITY OF NOTTINGHAM IN
FULFILMENT OF THE REQUIREMENTS FOR THE MASTER OF SCIENCE IN COMPUTATIONAL
NEUROSCIENCE, COGNITION AND AI

AUTHORED BY

ALBERT ALBESA GONZÁLEZ

SUPERVISED BY

PROF. MARK VAN ROSSUM



UNITED KINGDOM • CHINA • MALAYSIA

Abstract

The Bienenstock-Cooper-Munro (BCM) rule is a model of synaptic plasticity that was established almost 40 years ago. The model proposes that inputs lead to potentiation or depression of synapses depending on their ability to surpass an activity threshold, which is dependent on time. In a recent *preprint* (Froc et al., 2021) a modification of this rule that is weight-dependent (weight-dependent BCM, wBCM) was presented to include the experimental observation that relative depression of synapses is independent of their strength, while this is not the case for potentiation.

Here, we study wBCM in 2 dimensions as a dynamical system. We find its *stationary states* (or Fixed Points, FP) and, in particular, how their stability is related to background conditions like inhibition and environment stimuli. This allows us to characterize, for example, how a neuron might respond to different stimuli after development (its *selectivity*). Our theoretical analysis confirm what had been previously observed in simulations in (Froc et al., 2021): background inhibition plays a very important role in determining the stability of the different types of fixed points, which in turn strongly characterize the selectivity of the neuron. We find that very low inhibition leads to non-selective weights after learning, while for strong inhibition the standard BCM fixed points, which are highly selective, are recovered.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	iv
List of Figures	iv
List of Abbreviations	iv
1 INTRODUCTION	1
1.1 Learning in Living Organisms	1
1.2 Origins of Learning Rules	1
1.3 The BCM Rule	2
1.4 Weight Dependence in Synaptic Plasticity	2
1.5 The wBCM Rule	2
1.6 Dissertation Overview	3
2 MODEL	4
2.1 Neuron Model	4
2.2 standard BCM (sBCM)	4
2.3 weight-dependent BCM (wBCM)	5
2.3.1 Inhibition	5
2.3.2 wBCM Dynamics	5
2.3.3 2D System	6
3 RESULTS	7
d splits w and y into 4 different <i>dynamical regimes</i>	7
wBCM preserves original fixed points and generates additional ones	8
FP Candidates <i>consistency</i> depends on system parameters	10
Parameter u determines the selectivity of the weights after training	12
The system is non-differentiable at the standard BCM fixed points	14
Different FPC <i>consistency</i> and <i>stability</i> regions are complementary in u, ϕ	14
4 DISCUSSION	18
References	19
Appendices	21
A Weights → Activities Mapping	21
B Derivations	22

List of Tables

1	Classification of the different types of FP Candidates	10
---	--	----

List of Figures

1	Neuron model without (A) and with (B) splitting into excitatory and inhibitory weights	4
2	Diagram depicting the activities y_1 and y_2 that come from the two input patterns (respectively)	6
3	Different <i>dynamical regimes</i> in \mathbf{w} (Left) and \mathbf{y} (Right) for $\phi = 0.3$, $u = 0.6$	7
4	Fixed Point Candidate solutions, in \mathbf{w} and \mathbf{y} , for $\phi = 0.3$	11
5	Fixed Point Candidate solutions in \mathbf{w} for increasing values of u	12
6	Evolution with u of dw_{ij} (Left) and \mathbf{w} (Right), after training, for $\phi = 0.3$	13
7	Simulations convergence in \mathbf{y} (Left) and final selectivity (Right) for increasing values of u . .	14
8	Heatmaps of maximum eigenvalue λ_{Max} of J at standard BCM fixed point $\mathbf{y} = (2, 0)$, taking the limit from different regions, as a function of u and ϕ	15
9	<i>Consistency and stability</i> regions in (u, ϕ) for every type of FPC	16
10	Critical FPC bifurcation curves	17
11	$\mathbf{y} \rightarrow \mathbf{w}$ (Top) and $\mathbf{w} \rightarrow \mathbf{y}$ (Bottom) transformation of a circle	21
12	Discontinuities of the Jacobian at the selective BCM fixed point $\mathbf{y} = (0, 2)$ across $\mathbf{y} = m(1, 0)$	23

List of Abbreviations

BCM Bienenstock-Cooper-Munro

CLO Cooper-Liberman-Oja

sBCM standard BCM

wBCM weight-dependent BCM

LTD Long-Term Depression

LTP Long-Term Potentiation

LGN Lateral Geniculate Nucleus

FP Fixed Point

FPC Fixed Point Candidate

i-FPC inhibition Fixed Point Candidate

w-FPC weight-dependent Fixed Point Candidate

s-FPC standard Fixed Point Candidate

LSA Linear Stability Analysis

1 INTRODUCTION

1.1 Learning in Living Organisms

All living beings, from *E.coli* to humans, have evolved to gather information from their environment, process it and act accordingly. At some point of life history, though, some cells started specialising as information processing units (Kristan Jr, 2016) and, eventually, became *neurons* (which in turn gave rise to *neural systems* -or *neural networks*). While there is little disagreement on the fact that neurons have a computational purpose, there is another function of neural networks the mechanism of which has not always been clear: *learning*. If one understands a neural system as a black box that generates outputs given certain inputs, learning can be described as the process of varying and adapting the response to these inputs -presumably in a way that is evolutionary advantageous. Current view is that animals brains adapt their responses to stimuli by means of *synaptic plasticity*, i.e. the changes in the efficacies with which action potentials are effectively propagated from one neuron to another. This way, synaptic plasticity is thought to underlie *memory* and *learning* (Abbott & Nelson, 2000).

1.2 Origins of Learning Rules

Learning rules (in neuroscience) are models that describe how synapses evolve in time. These learning rules are typically defined at each synapse as a function of local information (such as presynaptic and postsynaptic activity or the current state of the synapse) and are unsupervised (do not depend on a loss function that explicitly evaluates an error -although they can sometimes be defined as an optimization problem by means of an objective function, see Intrator and Cooper (1992)). A sensible candidate for doing this, thus, is modelling synapses as dynamical systems and it is common to establish correspondences between the terminology of the latter and biological concepts. In this sense, for example, the *time constants* of the equations that describe different plasticity mechanisms are understood as the *timescales* in which these processes take place in nature, and the initial connectome (be it simply the weight distribution at birth or the specific topology of the network when learning starts) is clearly related to the *initial conditions* of the system. Furthermore, another important concept that can be borrowed from dynamical systems to generate predictions of the learning rule is that of fixed point. In particular, one would expect that the stable fixed points of the system (which will correspond to a set of weights that is stable in time) are able to describe the synaptic structure of mature neural networks, once learning has stopped or is much slower.

An early learning rule was the one proposed by Donald Hebb (1949), in which the strength of a synapse is increased proportionally to the product of both pre and postsynaptic activity. This rule is compatible with local, unsupervised learning, and any type of learning based on this rule is often called *Hebbian* or *associative*. Its intuition is that, given a stimulus, initial random connections can lead to specific neural activity in the brain. If the stimulus is presented repeatedly, the connections that more strongly participate in shaping the response are strengthened, which leads to a faster and more stable response in the future. The problem with this rule is that it is necessarily *incomplete*. While this can be part of the story, one needs complementary mechanisms that reduce the strength of neurons (or saturate its growth) to avoid Hebbian learning to lead to divergence of the weights (Turriagano & Nelson, 2004; Zenke et al., 2017).

One of the candidates to solve this issue was presented by Cooper, Liberman, and Oja (1979). The model, known as the CLO rule, included an *activity threshold*, which directed the weights to potentiation when the postsynaptic activity was high enough (above threshold) and to depression in the contrary case (below threshold). One of the problems that presented Hebbian learning in its original formulation was, besides the unrealistic lack of bounds in synaptic efficacy, the non-specificity of a scenario where all the weights are always potentiated. The CLO rule made an important leap in this sense, by leading to *some selectivity* (Cooper & Bear, 2012). Selectivity, although will be formulated precisely later, is a very important property of weight states, and is related to the capacity of a neuron to respond strongly to a stimulus (or set of them) while being silent to others. One can see how this is important in terms of information, as if all neurons responded approximately the same to all inputs the neural state would be non-informative and it would be difficult to establish maps with external world states. This learning rule, however, is heavily dependent on the initial weight distribution, and can still lead to *runaway excitation* (weight divergence) if the threshold is much lower than typical neural activity and *quiescence* (all synapses having zero efficacy) if the threshold is too high.

1.3 The BCM Rule

These ideas and previous proposals were the substrate that led Bienenstock, Cooper and Munro to come up with a model that has now been accepted to describe -at least in some regions of mammal brains- how synapses change with time: the BCM rule (Bienenstock et al., 1982). This model incorporated the ideas of input-output correlation of Hebbian Learning and threshold-dependent potentiating/depressing activity of the CLO rule, while solving the problem of weight divergence that preceding models presented. What characterizes this particular model is the introduction of threshold dynamics that try to capture recent neural activity. In this sense, the BCM threshold is dependent on time and can be understood as a temporal kernel that integrates statistical information of past postsynaptic activity. This sliding threshold typically leads to easily potentiated synapses when there is little average activity. At the same time, when synapses grow (and so does postsynaptic activity) the threshold also increases and it becomes harder and harder to potentiate neural connections. These two opposite forces (potentiation with low activity, depression with high activity) intuitively allow a stable learning in a variety of internal and external contexts.

The BCM rule has been reasonably successful in describing preexisting data, as well as generating predictions that have later been tested experimentally. An example of this is the reinterpretation of the long-known effect of monocular deprivation in animals (Wiesel & Hubel, 1963; Mioche & Singer, 1989). BCM theory corrected an oversimplified explanation of synaptic depression in these experiments by proposing that the real source of connection weakening is uncorrelated (rather absent) activity. This was further tested in later experiments, where chemical inactivation of the Lateral Geniculate Nucleus (LGN) effectively was shown to prevent synaptic depression in connections to primary visual cortex (Rittenhouse et al., 1999; Frenkel & Bear, 2004).

1.4 Weight Dependence in Synaptic Plasticity

One of the motivations for the plasticity rule here presented and studied (the weight-dependent BCM rule) is to incorporate an experimentally observed asymmetry in potentiation and depression. In particular, the model tries to account for the explicit weight dependence that there is in synapse depression ($\Delta w_i \propto w_i$), that results in a relative change that is independent from the weight ($\Delta w_i / w_i \propto 1$) (Debanne et al., 1996; Bi & Poo, 1998). Experimental results shows that this is not the case for LTP, where the absolute synaptic change is independent of the current weight (Bi & Poo, 1998; Debanne et al., 1999; Montgomery et al., 2001; Loebel et al., 2013). These observations have been implemented in computational models of plasticity (Van Rossum et al., 2000; van Rossum & Turrigiano, 2001) and have led to the so-called *weight-dependent synaptic plasticity*.

1.5 The wBCM Rule

The standard BCM rule (sBCM) has been arguably successful in explaining and predicting experimental observations. There are several factors, though, that the theory does not take into account or that seem at odds with biology. A clear example is how weights are allowed to change sign, which would correspond to an excitatory synapse becoming inhibitory. In their original work, it is posited that synapses should be considered as *ideal* (Nass & Cooper, 1975), i.e. a net effect efficacy that *may be mediated by several interneurons, some of which are excitatory and others inhibitory* (Bienenstock et al., 1982). This view assumes that the BCM model is at most a high-level description of learning dynamics. Since its original formulation, though, the BCM rule has been related to more detailed models of synaptic plasticity that try to get closer to biophysical mechanisms, as can be spike-timing dependent plasticity (Izhikevich & Desai, 2003; Gjorgjieva et al., 2011; Graupner & Brunel, 2012). If one wants to use it as a model that describes dynamics taking place at real synapses, there are certain assumptions (as the weight sign change) that need to be revised. At the same time, it can be interesting to try to incorporate details that are known from experimental data, as can be the above explained weight-dependence asymmetry in potentiation and depression.

In (Froc et al., 2021) a model of weight-dependent BCM (wBCM) is presented. The model not only introduces weight-dependent depression, but also restricts dynamic weights to be exclusively excitatory. This is done in pair with a background inhibition that responds proportionally to the input. Early results in a 2-dimensional system show that these dynamics favour the evolution of the system to less selective fixed points. These simulations can be extended to an N -dimensional model that includes N neurons from the Lateral Geniculate Nucleus (LGN) projecting to a neuron in the primary visual cortex. When trained with

stimuli selected from natural images, sBCM leads to very selective and oriented RFs (Law & Cooper, 1994). On the contrary, wBCM, while still presents a clear orientation preference, has much less selective RF, which respond more uniformly to different stimuli.

1.6 Dissertation Overview

Simulations results from (Froc et al., 2021) show that, in general, weight-dependent BCM leads to less selective and arguably more realistic receptive fields. However, the question of to what extent and under what conditions this holds true, remains to be answered. The present study tries to shed some light on this issue by doing a more analytical treatment on the weight-dependent dynamics of BCM. To do that, the different types of fixed points that the system can give rise to, together with their stability, are found as a function of different parameters. For simplicity, we focus on a reduced system where a single postsynaptic neuron receives input from two different neurons. This initial approach is also usually followed as a first result in other theoretical sBCM studies, be it in the original work or in modern research on its emergent dynamics (Bienenstock et al., 1982; Udeigwe et al., 2017).

2 MODEL

2.1 Neuron Model

We use a rate-based model where a single neuron performs a weighted summation of the outputs of a series of presynaptic neurons. A weight w_i can be understood as the synaptic strength with which, upon a spike, the axon of a presynaptic neuron i propagates current to the postsynaptic neuron. Because this is a rate-based model, though, the weights indicate how much the firing rate (or *activity*) x_i of each *pre* neuron can affect the firing rate of the *post*. Given N presynaptic neurons, one then has a postsynaptic activity y :

$$y = \sum_i^N w_i x_i \quad (1)$$

Previous versions of the model consider also an activation function $y = \text{ReLU}(\sum_i^N w_i x_i)$. However, in (Froc et al., 2021) it is pointed out how the addition or not of the rectification did not affect results significantly. To simplify analytical treatment of the system, which is the goal of this work, we will thus consider the output to simply be a linear response to the inputs. This is depicted in figure 1A, where the diagram shows that the activity y corresponds to the weighted summation of presynaptic activities x_i .

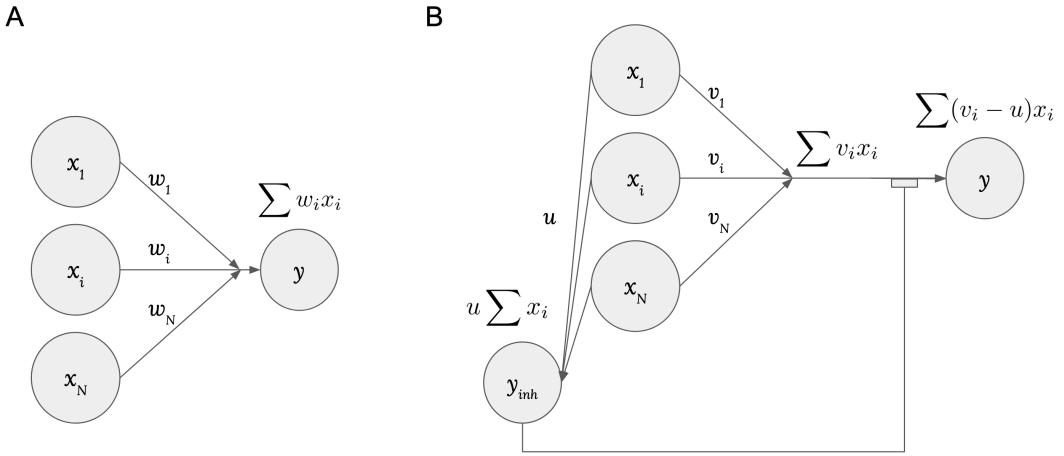


Figure 1: Neuron model without (A) and with (B) splitting into excitatory and inhibitory weights. Circles denote neurons, and arrows denote synapses, that can be *effective* (w_i), *excitatory* (v_i) or *inhibitory* (u). Paradigm B can be considered a special case of A with $w_i = v_i - u$, where u is fixed while v_i is restricted to be positive and is the only plastic component of the effective weight. Neuron y_{inh} represents an inhibitory neuron that affects negatively the output of the main postsynaptic neuron y .

2.2 standard BCM (sBCM)

The BCM rule gives weight dynamics in terms of presynaptic activity x_i , postsynaptic activity y and a threshold θ :

$$\tau_w \frac{dw_i}{dt} = x_i y (y - \theta) \quad (2)$$

where w_i is the weight that connects neuron i with the postsynaptic neuron and τ_w the weights time constant.

As already introduced, the key feature of this model is that θ is not static, and is otherwise defined to capture statistical information of postsynaptic activity in the recent past. Typically, and in the extension of BCM that we will consider here, the threshold integrates past squared activity y^2 convolved with an exponential kernel:

$$\theta(t) = \int_{-\infty}^t y^2 e^{-(t-t')/\tau_\theta} dt' \implies \tau_\theta \frac{d\theta}{dt} = y^2 - \theta \quad (3)$$

where τ_θ is the threshold time constant. The value of τ_θ defines the temporal window for which y^2 is effectively integrated, and plays an important role in the stability of the system.

2.3 weight-dependent BCM (wBCM)

2.3.1 Inhibition

With some exceptions (Owens et al., 1996), neurons are either excitatory or inhibitory, and it seems unrealistic that plasticity could change the sign of a synapse. In wBCM, this is taken into consideration by defining *excitatory weights* $v_i \geq 0$, together with an inhibition parameter u . Inhibition follows no plasticity rules and does not have a specific role in shaping network activity; rather, it can be considered background activity y_{inh} that responds proportionally to the sum of the input and is then subtracted to the excitatory response y_{ex} .

$$, \quad y_{inh} = u \sum_i x_i \quad y_{ex} = \sum_i v_i x_i \quad (4)$$

y_{in} then negatively affects the output of the neuron:

$$y = y_{ex} - y_{inh} = \sum_i v_i x_i - u \sum_i x_i = \sum_i (v_i - u) x_i \equiv \sum_i w_i x_i \quad (5)$$

where we have defined w_i (the *effective weights*) to be the excitatory weights minus the constant inhibition parameter u . Because excitatory weights can only contribute positively to postsynaptic activity, there is a hard-bound $v_i \geq 0$, i.e. $w_i = v_i - u \geq -u$. A diagram illustrating inhibitory-excitatory split of the effective weights is included in Figure 1B.

Although for u to be truly inhibitory it is required that it is positive, through most of this work it is, for completeness, just considered another parameter which can also take negative values. If one was to assign a biological correspondence to $u < 0$, it could be understood as background excitation that also responds linearly with the input.

2.3.2 wBCM Dynamics

As anticipated in the introduction, we want to include in sBCM a weight dependence that is sensitive to potentiation/depression. In particular, we would like the relative synaptic change $\Delta w_i / w_i$ to be independent of the weight for depression. We have seen how that can be obtained by having a total change (derivative) that is proportional to the plastic weight (v_i). Thus, when the synapse is being depressed, we propose the following dynamics:

$$\tau_v \frac{dv_i}{dt} = v_i x_i y (y - \theta) \quad (6)$$

But, when the synapse is being potentiated, sBCM is recovered:

$$\tau_v \frac{dv_i}{dt} = x_i y (y - \theta) \quad (7)$$

This piece-wise defined model can be compacted with the following expression

$$\tau_v \frac{dv_i}{dt} = [v_i]^d x_i y (y - \theta) \quad (8)$$

where d (for depression) is a dynamic variable that takes value 0 when the synapse is being potentiated (Long Term Potentiation, LTP *regime*) and 1 when the synapse is being depressed (Long Term Depression, LTD *regime*). Because the inputs are assumed to be positive, one can define d as follows:

$$d(y(t)) = \begin{cases} 0 & \text{if } y(y - \theta) \geq 0 \text{ (LTP)} \\ 1 & \text{if } y(y - \theta) < 0 \text{ (LTD)} \end{cases} \quad (9)$$

where the temporal dependence indicates that d is indeed dynamic. The arbitrary inclusion of 0 weight change in LTP (there is really no potentiation or depression taking place) facilitates the case-based analysis that will be done in the following sections.

To show the system dependence on inhibition, it can be convenient to now rewrite the dynamics in terms of the effective weights $w_i = v_i - u$. Using that $v_i = w_i + u$, and $\frac{dv_i}{dt} = \frac{d(w_i + u)}{dt} = \frac{dw_i}{dt}$, equation (8) can be rewritten as

$$\tau_w \frac{dw_i}{dt} = [w_i + u]^d x_i y (y - \theta) \quad (10)$$

As in sBCM, the threshold dynamics are kept to be a moving average of the second momentum of the output:

$$\tau_\theta \frac{d\theta}{dt} = y^2 - \theta \quad (11)$$

2.3.3 2D System

We will study the fixed points of wBCM for two input neurons and two different stimuli \mathbf{x}_1 and \mathbf{x}_2 . Following the proposal in (Udeigwe et al., 2017), the input is parametrized as:

$$\mathbf{x}_1 = (\cos \phi, \sin \phi) , \quad \mathbf{x}_2 = (\sin \phi, \cos \phi) \quad (12)$$

This particular choice reduces the input to one single parameter, which is in turn informative regarding the similarity of the presented stimuli ($\phi = 0 \implies \mathbf{x}_1 \cdot \mathbf{x}_2 = 0$, $\phi = \pi/4 \implies \mathbf{x}_1 \cdot \mathbf{x}_2 = 1$).

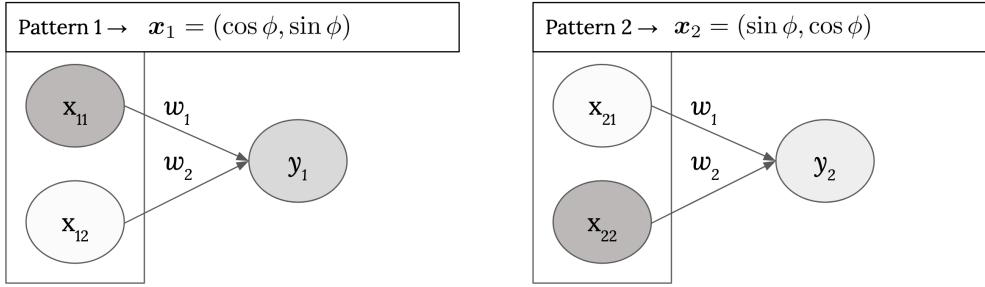


Figure 2: Diagram depicting the activities y_1 and y_2 that come from the two input patterns (respectively). The two inputs are presented in an alternated manner so that they are considered constant at each of the discretized timesteps.

Also following (Udeigwe et al., 2017), one can in this simplified scenario obtain the mean-field dynamics of the system. The mean-field approximation assumes that the pace of stimulus presentation is much faster than weight change, so that the *field* (dynamics) experienced by synapses is equivalent to that coming from the average of both stimuli. To do that, one first defines the activities that come from each pattern:

$$y_1 = w_1 x_{11} + w_2 x_{12} \quad (13)$$

$$y_2 = w_1 x_{21} + w_2 x_{22} \quad (14)$$

where x_{ij} is the j^{th} component of the i^{th} stimulus (i.e. $x_{11} = \cos \phi$). The same thing can be done in terms of the depression variable d , by defining a vector \mathbf{d} that follows equation (9) for each of the two inputs:

$$\mathbf{d} = (d_1, d_2) \equiv (d(y_1), d(y_2)) \quad (15)$$

This is, d_i takes value 1 if \mathbf{x}_i leads to depression and 0 otherwise. With these definitions, the mean-field dynamics of each weight are

$$2\tau_w \frac{dw_1}{dt} = [w_1 + u]^{d_1} x_{11} y_1 (y_1 - \theta) + [w_1 + u]^{d_2} x_{21} y_2 (y_2 - \theta) = 2(dw_{11} + dw_{12}) \quad (16)$$

$$2\tau_w \frac{dw_2}{dt} = [w_2 + u]^{d_1} x_{12} y_1 (y_1 - \theta) + [w_2 + u]^{d_2} x_{22} y_2 (y_2 - \theta) = 2(dw_{21} + dw_{22}) \quad (17)$$

where we have defined the field contribution at synapse i coming from pattern j to be dw_{ij} . The value of vector \mathbf{d} completely specifies the expressions that govern the dynamics of the system, and gives rise to the different *dynamical regimes*. These *regimes* will also be referred to as LTP-LTD, for example, when \mathbf{x}_1 leads to potentiation and \mathbf{x}_2 to depression (it would be the same as saying that $\mathbf{d} = (0, 1)$).

A final approximation is that, in the limit $\tau_\theta/\tau_w \rightarrow 0$, the threshold can be considered to capture fast enough the average of both square activities, and has otherwise no intrinsic dynamics:

$$\theta = \frac{y_1^2 + y_2^2}{2} \quad (18)$$

3 RESULTS

d splits w and y into 4 different *dynamical regimes*

This work focuses on understanding what are the stable fixed points of wBCM. The piece-wise definition of the dynamics, that only show explicit weight dependence upon depression, complicates this analytical treatment. The preliminary results here presented (Figure 3) show that both the weight $\mathbf{w} = (w_1, w_2)$ and activity $\mathbf{y} = (y_1, y_2)$ spaces can be separated into 4 different depression and potentiation *dynamical regimes*: LTD-LTD, LTD-LTP, LTP-LTD and LTP-LTP.

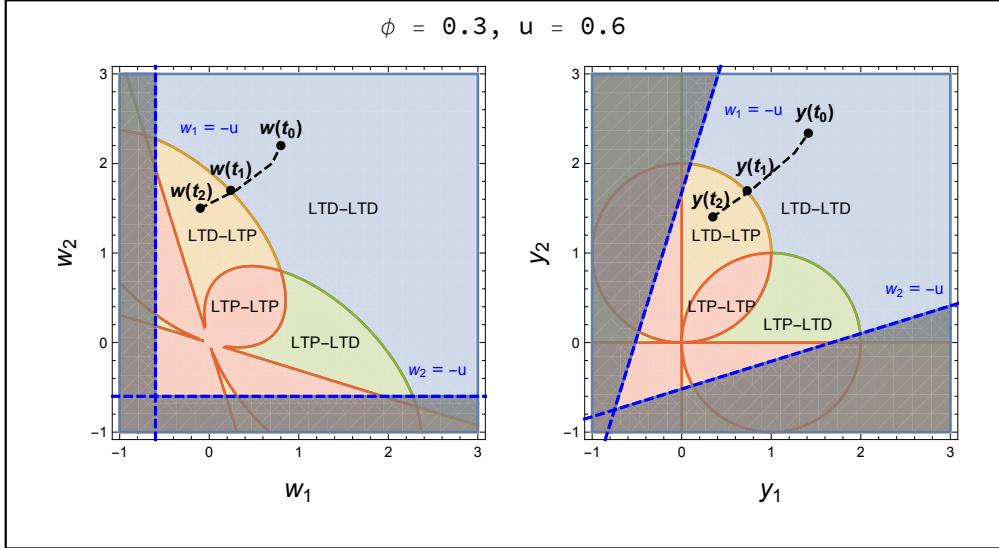


Figure 3: Different *dynamical regimes* in \mathbf{w} (Left) and \mathbf{y} (Right) for $\phi = 0.3$, $u = 0.6$. Red regions: $\mathbf{d} = (1, 1)$, LTP-LTP. Green regions: $\mathbf{d} = (0, 1)$, LTP-LTD. Orange regions: $\mathbf{d} = (1, 0)$, LTD-LTP. Blue regions: $\mathbf{d} = (1, 1)$, LTD-LTD. Dark regions: out of hard-bound ($w_i < 0$). Blue dashed lines: Limit of hard bounds ($w_i = -u$). Black dashed lines: Example trajectory in \mathbf{w} (Right) and its mapping in \mathbf{y} (Left) that transitions from LTD-LTD to LTD-LTP.

Figure 3 presents the *dynamical regimes* in terms of the weights and activities in the range $[-1, 3] \times [-1, 3]$. In general, if the inputs are not collinear ($\phi = \pi/4$), there is a one to one correspondence between \mathbf{w} and \mathbf{y} . Both state spaces will be used throughout depending on what needs to be highlighted. The details of how to map \mathbf{y} and \mathbf{w} can be found in the appendices (section A), but it will always correspond to a stretching/compression along directions $u_1 = \frac{1}{\sqrt{2}}(1, -1)$ and $u_2 = \frac{1}{\sqrt{2}}(1, 1)$. The \mathbf{y} plot in Figure 3 can be obtained following the definition in equation (9) for both activities y_1 and y_2 . The different *dynamical regimes* (coloured regions) can then be mapped to \mathbf{w} by following the transformations described in the appendices.

The purpose of this section is to show that, despite the asymmetry in weight-dependence makes the system non-continuous in general, it can be divided into regions where it does follow continuous dynamics (coloured regions in Figure 3). For example, one can consider the trajectory (only a sketch, not a real trajectory) indicated in the figure with a dashed black line ($\mathbf{w}(t)$ in \mathbf{w} and $\mathbf{y}(t)$ in \mathbf{w}). At $t = t_0$ the system is in LTD-LTD and the dynamic expressions come from substituting $(d_1, d_2) = (1, 1)$ in equations (16) and (17). The dynamics are continuous until $t = t_1$, where there is a transition from LTD-LTD to LTD-LTP. At this point, \mathbf{d} changes from $(d_1, d_2) = (1, 1)$ to $(d_1, d_2) = (1, 0)$, and so do the resulting dynamic expressions, that will again be continuous from $t = t_1$ to $t = t_2$.

These different regions where the dynamics are continuous will be used in the following section to perform a case-based fixed point analysis, that will first assume that a fixed point is found in a *regime* and then find its possible solutions and evaluate the consistency of such assumption.

wBCM preserves original fixed points and generates additional ones

As in previous section, and throughout all this study, we will focus on a 2-dimensional system, with 2 presynaptic neurons (2 inputs) and 2 different patterns \mathbf{x}_1 and \mathbf{x}_2 (presented in section 2.3.3). Now, we will in particular investigate what are the fixed points of wBCM. Characterising its stationary states, and understanding what is its correspondence with biology, is vital to both validate the model and make experimental predictions.

For better understanding the wBCM fixed point analysis, we first present the case of sBCM. The approach here presented is very similar to that in (Udeigwe et al., 2017), where the mean-field dynamics allow for exact solutions of the fixed points. Our case is, though, more simple, as the two inputs are not presented stochastically (they are alternated at every timestep). To construct the mean-field equations, this can be interpreted as a known probability of 1/2 for each input. In the original formulation of BCM:

$$\tau_w \frac{dw_i}{dt} = x_i y (y - \theta) \quad (19)$$

The mean-field equations are given by averaging over the two stimuli:

$$\tau_w \frac{dw_i}{dt} = \frac{1}{2} [x_1 y_1 (y_1 - \theta) + x_2 y_2 (y_2 - \theta)] \quad (20)$$

(see MODEL section for variable definitions)

Now one can change the dynamic variables w_i and use y_i instead by deriving at both sides in equations (13) and (14):

$$\dot{y}_i = \dot{w}_1 x_{i1} + \dot{w}_2 x_{i2}, \quad \dot{z} \equiv \frac{dz}{dt} \quad (21)$$

and substitute \dot{w}_i using equation (20) to eliminate weight dependence. Rearranging terms, the dynamics of the system in terms of y_i are given by:

$$2\tau_w \frac{dy_1}{dt} = [\mathbf{x}_1 \cdot \mathbf{x}_1 y_1 (y_1 - \theta) + \mathbf{x}_1 \cdot \mathbf{x}_2 y_2 (y_2 - \theta)] \quad (22)$$

$$2\tau_w \frac{dy_2}{dt} = [\mathbf{x}_2 \cdot \mathbf{x}_1 y_1 (y_1 - \theta) + \mathbf{x}_2 \cdot \mathbf{x}_2 y_2 (y_2 - \theta)] \quad (23)$$

The fixed points of \mathbf{y} are, by definition, the solutions to $\dot{y}_1 = \dot{y}_2 = 0$. Assuming that \mathbf{x}_1 and \mathbf{x}_2 are not collinear, that holds true if and only if

$$y_1(y_1 - \theta) = 0 \quad (24)$$

$$y_2(y_2 - \theta) = 0 \quad (25)$$

Which has 4 solutions that come from the different combinations of $y_i \in \{0, \theta\}$. A useful concept to characterize the fixed points (or any weight/activity state in general) is that of selectivity. Given different stimuli, the selectivity of a set of activities indicates how differently a neuron responds to each stimulus, and can be defined (in this particular case) as

$$Sel(y_1, y_2) = 1 - \frac{\text{Mean}(y_1, y_2)}{\text{Max}(y_1, y_2)} \quad (26)$$

As mentioned in the introduction, for neurons to encode relevant information about environment one needs them to be to some extent selective to the stimuli found in such environment. Although not shown here, in sBCM the solutions $\mathbf{y} = (0, 0)$ and $\mathbf{y} = (\theta, \theta)$, which have 0 selectivity, are *unstable* (and they are called *non-selective*), while solutions $\mathbf{y} = (0, \theta)$ and $\mathbf{y} = (\theta, 0)$, which are *maximally selective*, are *stable* (and they are called *selective*). Furthermore, the numerical values of these solutions can be found if one takes into account that

$$\theta = \frac{y_1^2 + y_2^2}{2} \quad (27)$$

which, substituting in equations (24) and (25), leads to:

$$y_1(y_1 - \frac{y_1^2 + y_2^2}{2}) = 0 \quad (28)$$

$$y_2(y_2 - \frac{y_1^2 + y_2^2}{2}) = 0 \quad (29)$$

That have solutions

$$(y_1, y_2, \theta) = \left\{ (0, 0, 0), (2, 0, 2), (0, 2, 2), (1, 1, 1) \right\} \quad (30)$$

One can substitute the sBCM FP activities in the selectivity definition and find that the selective FP have $Sel(y_1, y_2) = 1/2$, while the non-selective FP have $Sel(y_1, y_2) = 0$.

These results will facilitate the FP analysis of wBCM, as the sBCM FP will be found to be a subset of the possible wBCM fixed points. Nevertheless, the case-based formulation of its dynamics, that have a different analytical expression depending on whether the synapses are being depressed or potentiated, adds certain difficulties that need to be handled first. Here, we will solve this issue by following a self-consistency argument:

- First, we notice that, once in a FP, the system is stationary and can only be found in one potentiation/depression *regime* (the value of \mathbf{d} is unique and constant).
- Given a *regime*, there is an analytical expression for the temporal derivatives of the weights.
- Thus, one can obtain the exact fixed point equations for every *regime*.

This allows, for every *dynamical regime*, to obtain the formal fixed point equations. The equations governing every *regime* give rise to the *Fixed Point Candidates* (FPCs), which is the name given to its solutions. These will be *actual* Fixed Points when the solutions are *consistent* with the initial assumptions:

- i The solutions *exist*: There are solutions to the obtained fixed point equations.
- ii The solutions are *compatible*: The *regime* $\mathbf{d} = (d_1, d_2)$ associated with the solutions is the same as the *regime* that gave rise to the FP equations in the first place. For example, if the FPC comes from assuming LTD-LTD dynamics, the corresponding y_1, y_2 solutions also need to lead to $\mathbf{d} = (1, 1)$.
- iii The solutions are *accessible*: The solutions are within the hard-bounds of the system ($w_i \geq -u$).

We will say that *a FPC is consistent* (and therefore is an actual fixed point) whenever assumptions i, ii and iii hold true. Similarly, that *a FPC is not consistent* means that any of the assumptions is false (either because the FPC *does not exist* (condition i), *it is not compatible* (condition ii), or *it is not accessible* (condition iii)).

These preliminary arguments allow us to now do a case-based analysis of the Fixed Point Candidates that can emerge at each of the 4 *dynamical regimes* defined in the previous section, as well as the conditions for their *consistency*.

Case $\mathbf{d} = (0, 0) \implies \text{LTP-LTP}$

If the system is found in the LTP-LTP *regime*, then the equations that describe dynamics are those of standard BCM:

$$2\tau_w \frac{dw_1}{dt} = x_{11}y_1(y_1 - \theta) + x_{21}y_2(y_2 - \theta) \quad (31)$$

$$2\tau_w \frac{dw_2}{dt} = x_{12}y_1(y_1 - \theta) + x_{22}y_2(y_2 - \theta) \quad (32)$$

We have already seen how the solutions in this case come from $y_i(y_i - \theta) = 0$. These solutions always *exist* (we have found them in sBCM FP analysis). It is also clear, from equation (9), that the solutions will be *compatible* with the hypothesised *regime* (LTP-LTP). wBCM, thus, preserves (in principle) sBCM fixed points. The only reason why these FPC could be *not consistent* would be that they are not *accessible*: i.e., the hard-bound imposed by $w_i \geq -u$ is not compatible with the sBCM solutions (assumption iii is not true).

Because they are the same as the sBCM fixed points, these are called s-FPC (*standard Fixed Point Candidates*). The same labelling in terms of selectivity applies (selective s-FPC: $\mathbf{y} = (0, 2)$ and $\mathbf{y} = (2, 0)$, non-selective s-FPC: $\mathbf{y} = (0, 0)$ and $\mathbf{y} = (1, 1)$).

Case $\mathbf{d} = (1, 1) \implies \text{LTD-LTD}$

In this opposite case, both stimuli lead to depression and thus weight-dependent equations govern everywhere:

$$2\tau_w \frac{dw_1}{dt} = (w_1 + u)x_{11}y_1(y_1 - \theta) + (w_1 + u)x_{21}y_2(y_2 - \theta) \quad (33)$$

$$2\tau_w \frac{dw_2}{dt} = (w_2 + u)x_{12}y_1(y_1 - \theta) + (w_2 + u)x_{22}y_2(y_2 - \theta) \quad (34)$$

Because of the definition of LTD in equation (9), the standard BCM fixed points cannot give rise to this *regime*. The only possibility that the temporal derivatives are zero is that $(w_1 + u) = 0$ and $(w_2 + u) = 0$, so that every single term has zero value. Again, formal fixed point equations always have solutions, in this case $w_i = -u$. Given that the hard-bound is $w_i \geq -u$ it is also clear that they are also *compatible*. The only assumption that can sometimes fail to be true is assumption (ii): the solutions will not always lead to the hypothesised *regime* (in this case LTD-LTD, $\mathbf{d} = (1, 1)$).

If the system is found LTD-LTD, the only FPC is then $\mathbf{w} = (-u, -u)$. Because it depends exclusively on inhibition parameter u , it is called i-FPC (*inhibition Fixed Point Candidate*).

Case $\mathbf{d} = (1, 0) \implies \text{LTD-LTP (or } \mathbf{d} = (1, 0) \implies \text{LTP-LTD)}$

We will follow the equations governing for LTD-LTP, but the reasoning would be the same for LTP-LTD. For $\mathbf{d} = (1, 0)$, one has:

$$2\tau_w \frac{dw_1}{dt} = (w_1 + u)x_{11}y_1(y_1 - \theta) + x_{21}y_2(y_2 - \theta) \quad (35)$$

$$2\tau_w \frac{dw_2}{dt} = (w_2 + u)x_{12}y_1(y_1 - \theta) + x_{22}y_2(y_2 - \theta) \quad (36)$$

The only type of fixed point solutions that this *regime* accepts come from potentiation and depression terms being different than 0, of equal magnitude, and of opposite sign:

$$dw_{11} = -dw_{12}, \quad dw_{21} = -dw_{22}, \quad dw_{ij} \neq 0 \quad (37)$$

These are the formal equations of the *weight-dependent Fixed Point Candidates* (w-FPC). They have this name for being the wBCM fixed points that (as we will see) give rise to its characteristic *less selective* receptive fields.

In the appendices (section B), it is shown how some analytical results can be further obtained in this *regime*. In particular, w_1 and w_2 can be shown (B.1) to have a linear relation:

$$w_1 = kw_2 + (k-1)u, \quad k = \frac{x_{21}x_{12}}{x_{22}x_{11}} \quad (38)$$

However, substituting $w_1(w_2)$ and imposing a 0 derivative leads to a very complicated polynomial that can only be solved numerically. This makes it not possible to know beforehand (without u and ϕ) if it will have solutions (*exist*, assumption i), if they will be *compatible* (assumption ii), or if it will be *accessible* (assumption iii).

Table 1: Classification of the different types of FP Candidates

Name	Expression	Necessary <i>Regime</i>
i-FPC	$w_i = -u$	LTD-LTD
w-FPC	$dw_{i1} = -dw_{i2} \neq 0$	LTD-LTP / LTP-LTD
s-FPC	$y_i(y_i - \theta) = 0$	LTP-LTP

These results allow us to exhaustively map any possible *regime* to a set of fixed point candidates. At the same time, because any real fixed point will be found in one of these *regimes*, it is also always possible to classify an actual FP in terms of the FPC it comes from.

FP Candidates *consistency* depends on system parameters

So far we have seen that a point $\mathbf{y} = (y_1, y_2)$ (or its corresponding $\mathbf{w} = (w_1, w_2)$) has an associated *regime* $\mathbf{d} = (d_1, d_2)$ that limits the types of potential FPs that could emerge in that particular position. In turn, the *consistency* of such FPC depends on whether it meets certain requirements (*existence* (i), *compatibility* (ii) and *accessibility* (iii)). This section follows a more visual approach to locate the FPC solutions and visually inspect how they relate to each other, as well as how their *consistency* depends on system parameters.

Figure 4 shows, together with the different *dynamical regimes* in \mathbf{w} and \mathbf{y} , the FPCs for $\phi = 0.6$, $u = 0.6$. This plot is included as an example of conditions where all FPC have solutions (*exist*), although not all of them are *consistent*. The i-FPC (blue dot) is located at the intersection of hard-bounds (blue dashed lines, $w_1 = -u$ and $w_2 = -u$), as its solution is $\mathbf{w} = (-u, -u)$. In this example this FPC is *incompatible*, as the i-FPC comes from assuming LTD-LTD dynamics (blue region) and here it leads to LTP-LTP (red region). The w-FPCs (black dots) have one solution that comes from assuming LTD-LTP (black dot in orange region) and the other from assuming LTP-LTD (black dot in green region). The non-selective s-FPC (red empty dots) are *accessible* (within hard-bounds) and therefore *consistent*, while the selective s-FPC (red filled dots) are not because they are not *accessible* (fall within dark region).

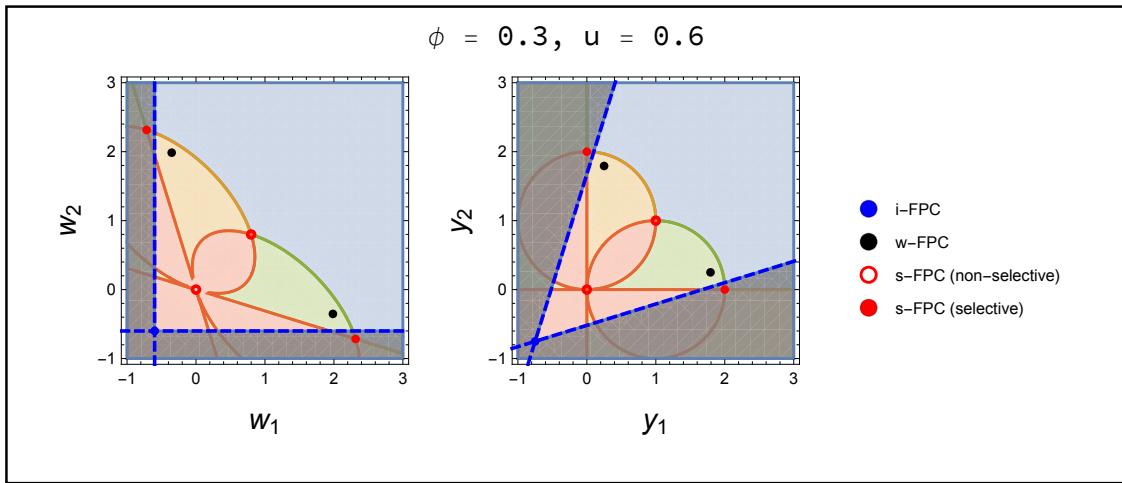


Figure 4: Fixed Point Candidate solutions, in \mathbf{w} and \mathbf{y} , for $\phi = 0.3$. *Dynamical regimes* and hard-bounds follow the same colour legend as in Figure 3.

In Figure 5, 4 plots (only \mathbf{w}) with increasing values of inhibition ($u = -2, -0.3, 0.85, 1.2$) are included. These plots give visual insight on the *consistency* and distribution of the different types of FPCs, as well as how this changes with u . For each type of FPC, the evolution with parameter u is as follows:

- i-FPC: Initially ($u = -2$), it is the only *accessible* FPC. As the hard-bounds (blue dashed lines) and its intersection (solution of i-FPC, blue dot) move with u , it eventually transitions from LTD-LTD (blue region), to LTP-LTP (red region), and thus stops being *compatible*.
- w-FPC: These is the only FPC for which there are not always solutions to formal fixed point equations. For this reason, only in the plots $u = -0.3$ and $u = 0.85$ there black dots are shown (while this is not the case for $u = -2$ and $u = 1.2$). Whenever it admits solutions, there appears one *consistent* FPC per possible *regime*: one in LTD-LTP (orange region) and one in LTP-LTD (green region). The solutions are always symmetric in both the weight and activity spaces.
- s-FPC: We have seen that this FPCs *consistency* depends exclusively on their *accessibility*. For $u = -2$, none of them is *accessible*. As u increases, the hard-bounds allow wider regions in both spaces. For $u = -0.3$ the non-selective FP $\mathbf{y} = (1, 1)$ is *accessible*, and for $u = 0.85$ and 1.2 all of the s-FPC are *accessible* (and thus *consistent*).

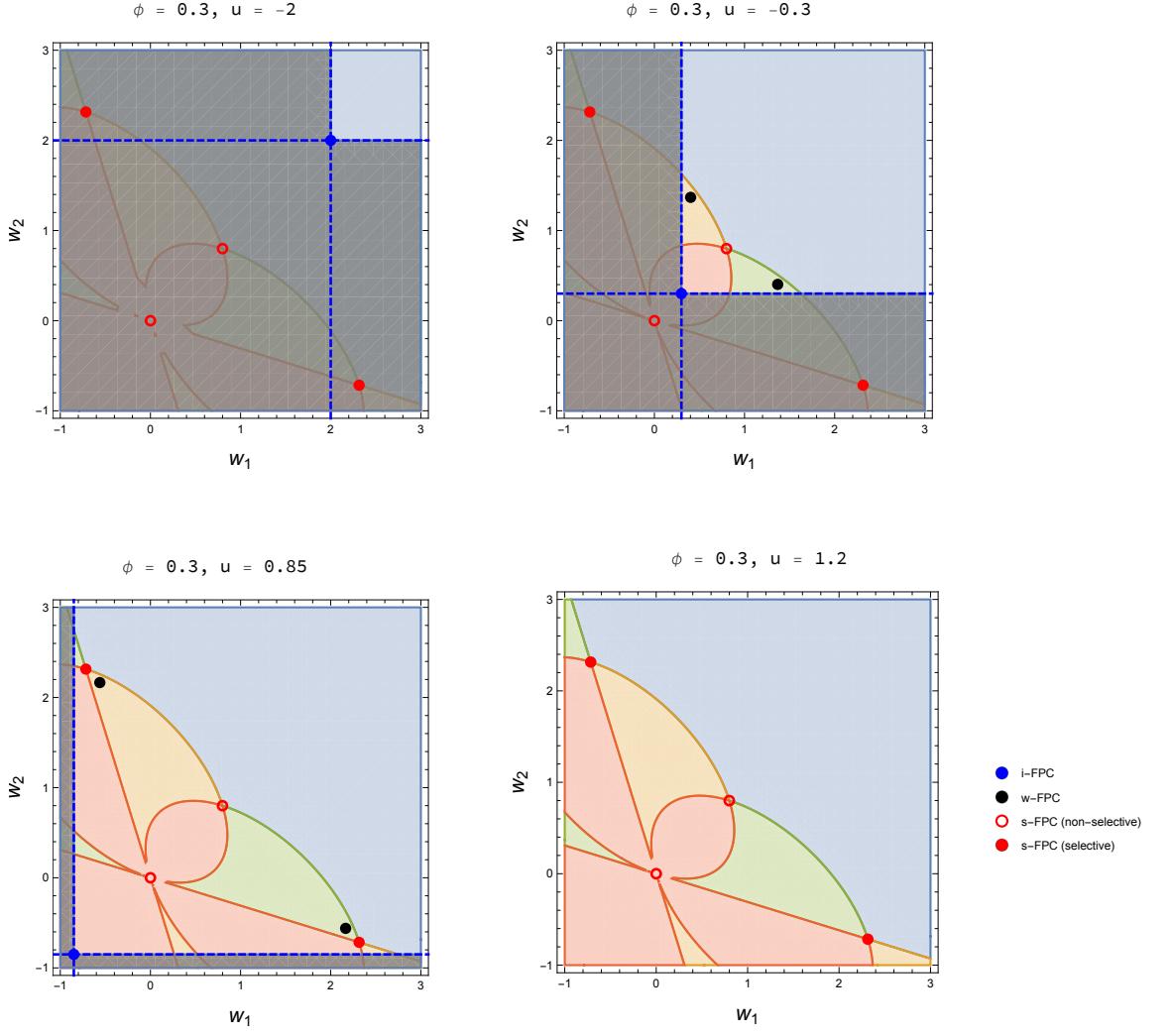


Figure 5: Fixed Point Candidate solutions in \mathbf{w} for increasing values of u . *Dynamical regimes* and hard-bounds follow the same colour legend as in Figure 3.

Parameter u determines the selectivity of the weights after training

In previous sections, we have discussed what are the types of Fixed Point Candidates that one can find in each *regime* \mathbf{d} . At the same time, we have also seen how the solutions (and *consistency*) of the FPCs depends on system parameters, specially on u . The next natural question to ask is what is the stability of these fixed points. This section uses simulations to see the effect of parameter u in the stability of the system. We train (present with stimuli) a neuron until the weights converge (the *converging solutions*), which presumably correspond to the stable fixed points of the system. We also analyze how the selectivity of such solutions changes with u and find that there is in fact a direct correspondence between parameter u and the selectivity of the weights after training.

Figure 6 plots different quantities of the system after training, for increasing values of u , while keeping ϕ fixed to 0.3. In the left panel we plot the values of dw_{ij} (weight changes coming from each input, defined in MODEL), and in the right panel the two synaptic weights w_1 and w_2 . These plots complement Figure 5 in that both show how the system changes with u , but allows for a more complete picture by showing all intermediate values of u .

The two plots in Figure 6 together inform about what could be the converged FPC for each value of u . For example, one can observe that for intermediate values of u (in the plot indicated between u_i^c and u_s^c) the weight changes coming from each stimulus are different than 0 and have same magnitude and opposite sign within the same synaptic weight ($dw_{i1} = dw_{i2} \neq 0$). We have seen how this condition is necessary for w-FPC, while both i-FPC and s-FPC have $dw_{ij} = 0$. Right panel, thus, tells us that for $u_i^c < u < u_s^c$ the system is converging towards the w-FPC, and these are probably the stable fixed points. For $u < u_i^c$ and $u > u_s^c$, the left panel can shed some light on what type of FPC the system converges to. In the case $u < u_i^c$ we see that both weights have the same value and decrease with u . We know that the i-FPC, which has solution $\mathbf{w} = (-u, -u)$, meets these requirements. On the other hand, the values of each synaptic weight w_1 and w_2 are different at $u > u_s^c$. Another important observation is that after crossing u_s^c the final synaptic weights are independent of u . The only FPCs that meet these requirements are the selective s-FPC (the non-selective s-FPC would still be independent of u but would satisfy $w_1 = w_2$).

Results in Figure 6 seem to indicate that the system stable fixed points smoothly transition from i-FPC, to w-FPC and finally remain in s-FPC when varying parameter u . For this reason, we have labelled the critical values of u as u_i^c (critical value for convergence towards i-FPC) and u_s^c (critical value for convergence towards s-FPC).

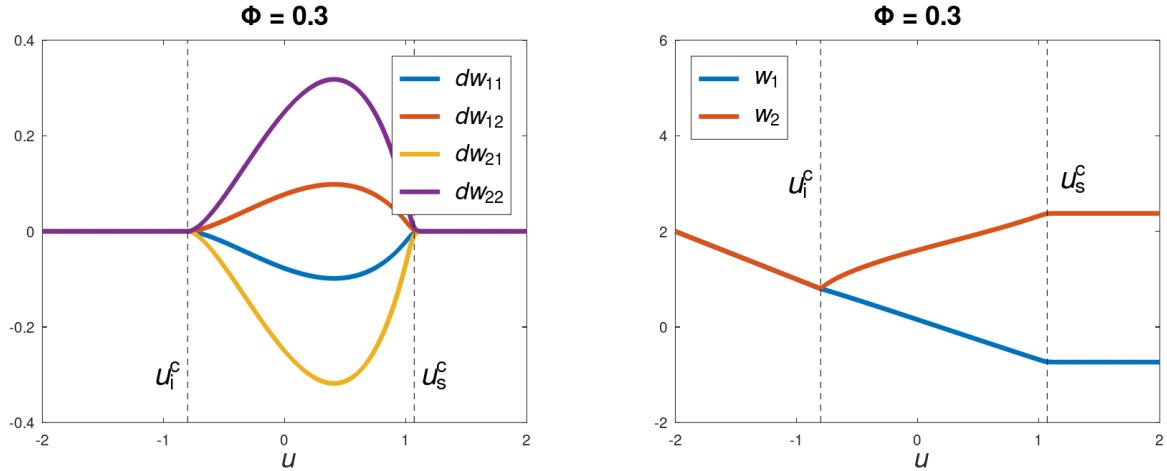


Figure 6: Evolution with u of dw_{ij} (Left) and \mathbf{w} (Right), after training, for $\phi = 0.3$. Final state after training is tightly related to the different type of Fixed Point Candidates, the *stability* of which seems to be disjointly distributed in u . This makes it natural to split u into the proposed regions: (i) $u > u_s^c$, (ii) $u_i^c < u < u_s^c$ and (iii) $u < u_i^c$.

An important consequence of this observed transition between FPC is in terms of the selectivity that the neuron presents towards the different stimuli once training has finished. For $u < u_i^c$ the neuron is completely non-selective, as both weights (and both activities) are the same, while for $u > u_s^c$ the neuron is maximally selective (we know this because for those values of u the weights converge towards the selective s-FPC). We now investigate how the solutions evolve with u (in the activity space) in relation to sBCM fixed points, as well as how the selectivity of the system after training is affected by inhibition.

In Figure 8 (Right) converging solutions in \mathbf{y} are continuously plotted as a function of parameter u . The color in the tick line indicates for which value of inhibition a point corresponds to activities after training. Points where the line is bright correspond to very negative values of u , while darker points correspond to values of u getting closer to 1. It is clear now how the converging solutions, when LTD-LTD (blue region) is the only *accessible regime*, approach the non-selective s-FPC $\mathbf{y} = (1, 1)$. Then, the two symmetric w-FPC solutions appear (this would correspond to $u > u_s^c$). As u keeps on increasing, the w-FPC solutions resemble more and more the selective s-FPC $\mathbf{y} = (2, 0)$ and $\mathbf{y} = (0, 2)$, until these become the converging state ($u > u_s^c$). The evolution in selectivity of the weights after training, according to equation 26, is presented in the right plot. What can be intuitively seen in the left panel with the transition from the non-selective to the selective s-FPC is now quantified: parameter u determines the selectivity of weights after training.

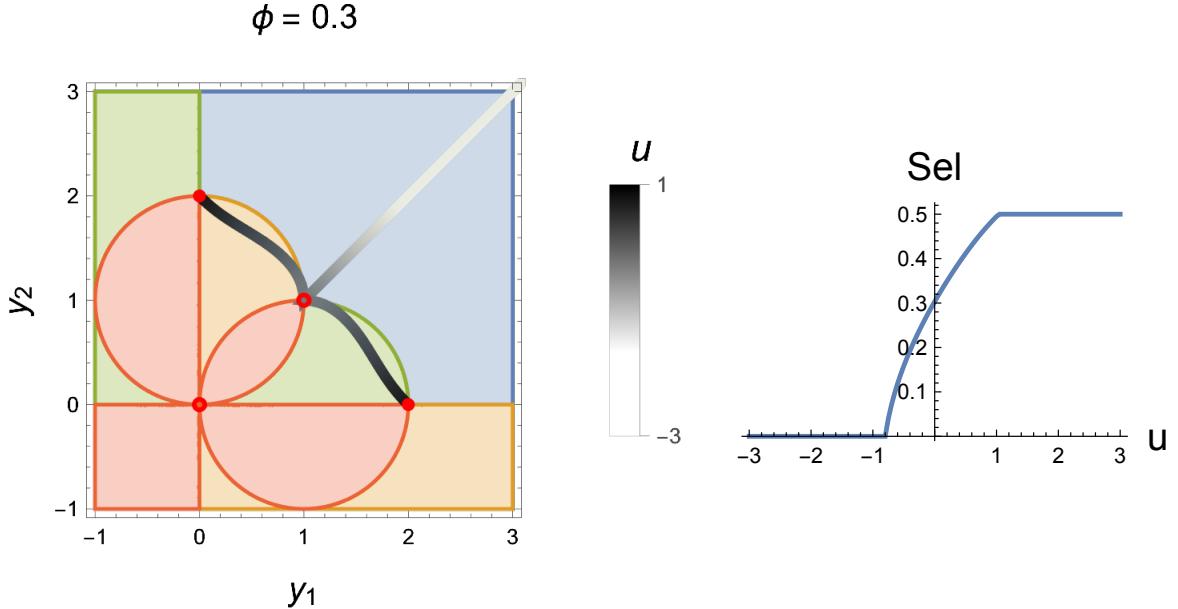


Figure 7: Simulations convergence in \mathbf{y} (Left) and final selectivity (Right) for increasing values of u . The system converges, for very negative values of u , to the i-FPC. As u increases, converging solutions bifurcate (the system might converge to one or another depending on initial conditions) into symmetric w-FPC. For even greater values of u , w-FPC solutions are eventually identical to the selective s-FPC (lying in $y_1 = 0$ and $y_2 = 0$) and, from there on, the system always converges to standard BCM selective fixed points.

The system is non-differentiable at the standard BCM fixed points

Previous section presents how parameter u plays an important role in stability, as it determines what is the converging state of the system in simulations. This motivates the study of linear stability as a function of the two parameters that characterize the system: u and ϕ . Before doing that, though, one must solve a difficulty that is presented at the the s-FPC. At the sBCM fixed points there is a non-smooth transition in the dynamics of the system, as they are found at the intersection between different LTD/LTP *regimes* (see figure 4 for example). In the appendices (section B.2), it is shown that this dynamical discontinuity results in a value of the Jacobian J that is not well-defined. In the context of *Linear Stability Analysis*, to see if a fixed point is stable, one investigates the sign of the eigenvalues (λ_i) of J . Then, one considers that a fixed point is *stable* if all eigenvalues are strictly negative .

The discontinuity of the Jacobian at the sBCM fixed points has therefore important implications in the LSA that follows. There is not *a Jacobian* that describes how the system evolves when perturbed at a FP, but *limits of the Jacobian* that describe how the system evolves when it is perturbed in the direction in which the limit is taken. To perform stability analysis at the s-FPC, we will thus require a stronger condition, which is that the system is stable (all the eigenvalues of the Jacobian are less than 0) for all the different types of dynamics (LTP-LTP, LTD-LTP, etc.) independently. We will say that a FPC with a discontinuity in J is *stable* when this condition is met. Then, one can assume that if the system is perturbed in any direction, the perturbation will always be damped regardless of the *regimes* that are visited.

Different FP Candidates *consistency* and *stability* regions are complementary in (u, ϕ)

So far, stability analysis has been done in terms of *converging states* in simulations. When trying to move from this to LSA, we have seen that one needs to be careful at the s-FPC, as they are at the intersection of different *regimes*, where the Jacobian is not well-defined. Previous section has proposed how to handle this discontinuity with a stronger requirement for perturbations at s-FPC to be stable: that the eigenspectrum

of the limit of J taken in any direction must contain only negative values.

Figure 8 shows the procedure followed to find the stability of a s-FPC, where the maximum eigenvalue of J in the limit of each type of dynamics (LTD-LTD, LTD-LTP, etc.) is plotted as a function of system parameters u and ϕ . In particular, the figure includes results for the selective s-FPC solution $\mathbf{y} = (2, 0)$. It can be visually appreciated how obtaining the maximum across all dynamics is actually identical to considering LTD-LTP dynamics only. This was confirmed numerically, as the λ_{Max} of J for that *regime* (LTD-LTP) was always higher than for any other potentiation/depression *regime*. This was also the case for the other selective solution ($\mathbf{y} = (0, 2)$), although there it was LTD-LTP that governed stability.

As one would expect, in LTP-LTP dynamics, $\lambda_{\text{Max}} < 0$ for all values of u and ϕ . As we have seen, in this *regime* standard BCM is recovered, and it is no surprise that the selective s-FPC solutions are thus stable everywhere, as they are the two stable solutions in sBCM. In turn, this allows us to know beforehand that the non-selective sBCM fixed points cannot be stable (if one follows the requirements here described), as they are unstable in sBCM and will thus present some non-negative eigenvalue for all values of u and ϕ within LTP-LTP dynamics. From here on, when referring to s-FPC stability we thus refer only to the selective-FPC.

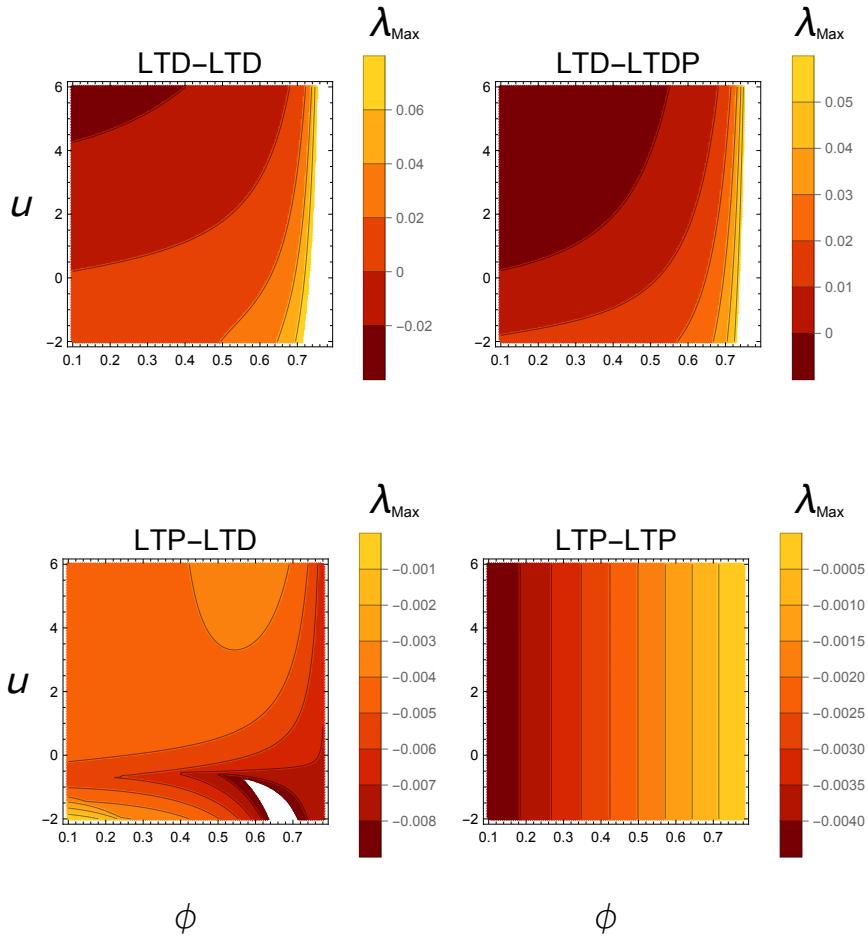


Figure 8: Heatmaps of maximum eigenvalue λ_{Max} of J at standard BCM fixed point $\mathbf{y} = (2, 0)$, taking the limit from different regions, as a function of u and ϕ . Because s-FPC are at the intersection of different *regimes*, the direction of the limit determines the analytical expression of J , that is different depending on \mathbf{d} . Plots show how the most restrictive dynamics, in terms of stability, are the ones corresponding to LTD-LTP (for this particular fixed point), as it has the largest eigenvalues at every region. As expected, the FP is always stable for LTP-LTP, as this FP is known to be stable in sBCM dynamics. LTP-LTP is also the only *regime* that is independent of u .

Figure 9 shows, for every type of FP Candidate, their *consistency* and *stability* across (u, ϕ) . Results show how, for every type of FPC, there is a region where the FPC is *consistent* and is *stable*, and another in which the FPC does not *consistent* (its stability cannot be evaluated because it is not an actual Fixed Point). The s-FPCs have an additional region where the FP is *consistent* (in this case is *accessible*) but it is unstable. It is observed that the stability regions are complementary (there is, for every pair (u, ϕ) , only one type of FPC that is stable). The *consistency* regions are almost perfectly complementary: at every point of the plane there is only one FPC that is *consistent*, with the exception of the orange region in s-FPC plot, where both w-FPC and s-FPC fixed points are *consistent*.

Results indicate that the system can be partitioned into several regions where the different types of FPC transition in *consistency* and *stability*. This already seemed to be the case observed in simulations in previous sections, where u_i^c and u_s^c were defined. There, although ϕ was kept constant, these points of transition could already be appreciated. The critical values of inhibition presented previously (u_i^c and u_s^c) are now extended as curves that depend on ϕ and divide the (u, ϕ) plane as follows:

- $u(\phi) < u_i^c(\phi)$
The only *consistent* FPC are i-FPC. They are *stable*.
- $u_i^c(\phi) < u(\phi) < u_{s1}^c(\phi)$
The only *consistent* FPC are w-FPC. They are *stable*.
- $u_{s1}^c(\phi) < u(\phi) < u_{s2}^c(\phi)$
Both w-FPC and s-FPC are *consistent*. Only w-FPC are *stable*.
- $u_{s2}^c(\phi) < u(\phi)$
The only *consistent* FPC are s-FPC. They are stable.

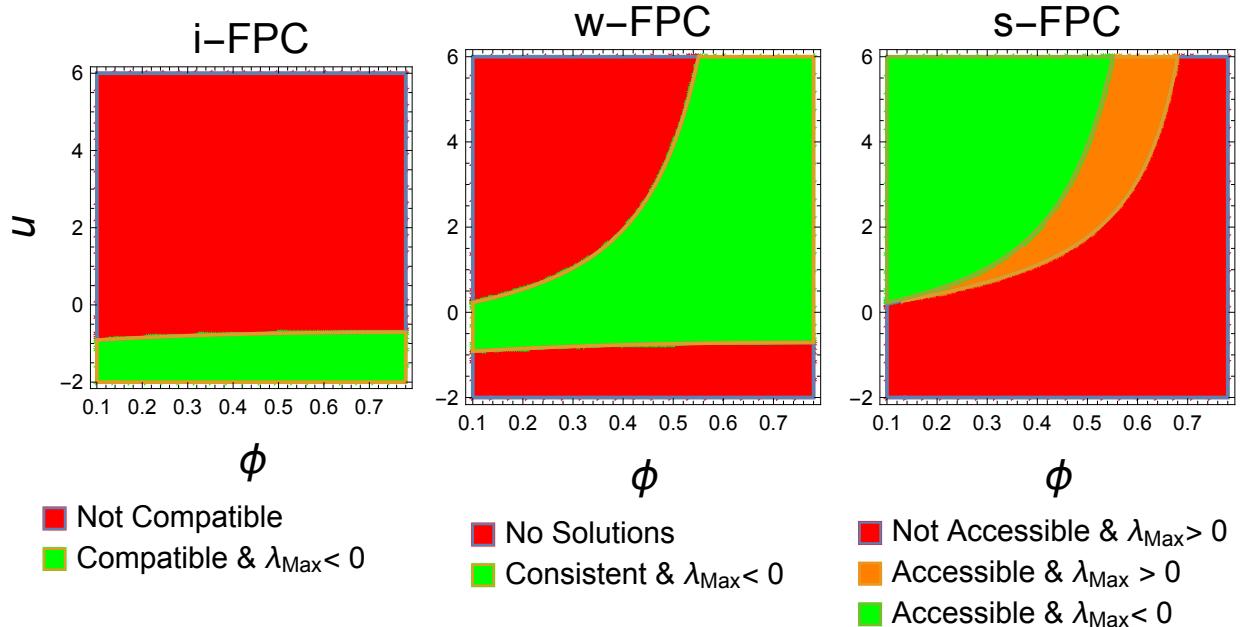


Figure 9: *Consistency* and *stability* regions in (u, ϕ) for every type of FPC. For s-FPC, λ_{Max} is taken across all possible eigenvalues of J in the limits of all possible dynamics. For i-FPC and w-FPC, λ_{Max} is taken across the eigenvalues of J assuming *compatible* dynamics (The value of J that corresponds to the *regime* that gives rise to the fixed point candidate). i-FPC: *consistency* depends on the *compatibility* of the FP. In green region, the fixed point is *compatible* (and thus *consistent*) and is *stable*. w-FPC: formal equations only admit solutions in green region, where the FPC exists, is *compatible* and is *accessible* (it is *consistent*) and is *stable*. s-FPC: The s-FPC are *accessible* in both orange and green regions, but are only *stable* in the green region.

To finalize the results section, we show the analytical expressions of the critical $(u(\phi), \phi)$ curves that have been found to divide the system in terms of its *consistent* and *stable* fixed points. These are obtained by imposing the conditions that we numerically observed to hold at each of the curves, and its derivation is included in the appendices (section B: B.3, B.4, B.5):

$$u_i^c(\phi) = -\frac{1}{\cos \phi + \sin \phi} \quad (39)$$

$$u_{s1}^c(\phi) = \frac{2 \sin \phi}{(\cos \phi + \sin \phi)(\cos \phi - \sin \phi)} \quad (40)$$

$$u_{s2}^c(\phi) = \frac{\sin 2\phi}{(\cos \phi - \sin \phi)^2 (\cos \phi + \sin \phi)} \quad (41)$$

The plot of each of the critical curves, together with the FPC properties of the regions that are delimited by them, is included in Figure 10. While they have been checked to actually be the same as the regions obtained numerically, one can already see by visual inspection that they do match the different plots shown in Figure 9.

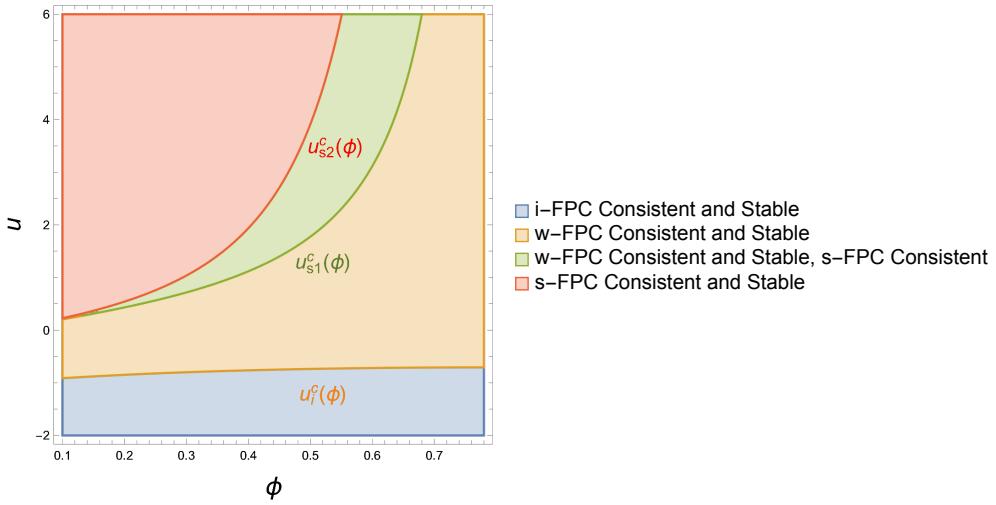


Figure 10: Critical FPC bifurcation curves. The three curves $u_i^c(\phi)$, $u_{s1}^c(\phi)$ and $u_{s2}^c(\phi)$ define 4 regions in the (u, ϕ) plane, where only one type of FPC is stable simultaneously. In three of the four regions (blue, orange and red) the stable FPC is also the only *consistent* type of FPC; in green region only w-FPC is stable but s-FPC is *consistent*, too.

4 DISCUSSION

We have studied weight-dependent BCM (wBCM) (Froc et al., 2021) in 2 dimensions as a dynamical system. This model is initially motivated by the experimental observation that relative depression seems to be independent of synaptic strength, together with the proposal of considering that only excitatory synapses are affected by BCM-like plasticity. Here, we have focused on finding the fixed points of this modified rule and their stability. The piece-wise nature of the dynamics has constrained us to define the so-called *Fixed Point Candidates* (FPC), that are the fixed point solutions that one obtains when assuming the different types of potentiation/depression *regimes* the system can be in. There are 3 types of fixed points that can, in principle, emerge from wBCM: i-FPC, w-FPC and s-FPC. For these solutions to be real fixed points (which we have called *consistent FPC*), we have seen that they must meet certain requirements that depend on system parameters. Both simulations and *Linear Stability Analysis* have shown the role of inhibition in the *consistency, stability* and *selectivity* of the different FP Candidates.

These results have implications in understanding how background and environmental conditions might affect plasticity during development and can make experimental predictions to further test the validity of wBCM. We have seen that when inhibition parameter changes sign, and can actually be considered background excitation, the stable fixed points are characterised by *quiescence* (the excitatory weights converge to 0). This is not surprising, as a lot of background excitation would lead to a very high threshold that would make most stimuli depressing. At the same time, increasing values of inhibition has been shown to lead to higher selectivity. Experiments investigating the relation between the activity of inhibitory and excitatory cells in primary visual cortex and the selectivity of the resulting receptive fields could shed some light on whether results obtained here can be extrapolated to biology.

The work here presented is connected to several research being done in the fields of neuroscience, computational models of learning or even the study of general dynamical systems. It builds upon results from (Castellani et al., 1999) and, specially, (Udeigwe et al., 2017) to formulate the study of the fixed points and their stability. Several choices that might seem arbitrary, as the input parametrization, the mean-field approximation or working with the activities instead of the weights have been inspired by this previous work. Notably, the validity of the mean-field approximation has been recently formalized in (Castellani et al., 2021) by studying the thermodynamical properties of different threshold dynamics as stochastic chemical processes. Some of the limitations present in the methods used in these models remain still to be addressed. For example, we have also used a linear model without rectification to facilitate analysis, which in (Udeigwe et al., 2017) is justified by understanding activities in terms of voltages instead of firing rates (Clopath & Gerstner, 2010). A field where the BCM model of plasticity is currently being put to use (see for example (Wang et al., 2020; Guo et al., 2020; Huang et al., 2021)), is that of *Neuromorphic Computing*. This area of research tries to generate chips that naturally mimic the key dynamics of neural computation in the brain, and thus an important effort is being made in benchmarking neuromorphic chips by their ability to reproduce models of weight dynamics in biological neural networks. This work (together with (Froc et al., 2021) where wBCM was first presented) can also be understood, besides its particular implementation in BCM, as another model of weight-dependent plasticity. It thus extends the work presented, for example, in Van Rossum et al. (2000), Rubin et al. (2001) or Morrison et al. (2007), where weight dependence in depression is used for stabilization of very large synapses that would otherwise need of additional homeostatic mechanism to avoid divergence.

While the introduction of weight-dependence could extend virtually any past research in the BCM model, the present work leads to some specific questions that remain to be answered. An example of this is the barrier of analytical treatment for w-FPC. These type of fixed points are likely to be, due to the intermediate conditions that require, the ones of highest biological relevance. However, an analytical solution for them has not been found, and stability analysis has necessarily relied on numerical approximations. Nevertheless, the fact that their solutions and stability are characterized by the same critical curves as i-FPC and s-FPC, suggests that there could be some analytical method to also relate its dynamical properties to system parameters. Another question that one might make is what characterizes the fixed points in N dimensions, as neurons typically accept input from approximately 1000 other neurons. While these should likely be related to the different FPC found here, new mixed types (for example combinations of w-FPC with s-FPC or i-FPC) could also emerge by the additional degrees of freedom that would be added to the system. Simulations in (Froc et al., 2021), though, seem to evidence that the stable fixed points in this context (large N) are fundamentally of the w-FPC type, leading to receptive fields that are less selective than those obtained with sBCM.

References

- Abbott, L. F., & Nelson, S. B. (2000). Synaptic plasticity: taming the beast. *Nature neuroscience*, 3(11), 1178–1183.
- Bi, G.-q., & Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of neuroscience*, 18(24), 10464–10472.
- Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2(1), 32–48.
- Castellani, G., Cooper, L. N., De Oliveira, L. R., & Blais, B. S. (2021). Energy consumption and entropy production in a stochastic formulation of bcm learning. *Journal of Computational Biology*, 28(3), 257–268.
- Castellani, G., Intrator, N., Shouval, H., & Cooper, L. (1999). Solutions of the bcm learning rule in a network of lateral interacting nonlinear neurons. *Network: Computation in Neural Systems*, 10(2), 111.
- Clopath, C., & Gerstner, W. (2010). Voltage and spike timing interact in stdp—a unified model. *Frontiers in synaptic neuroscience*, 2, 25.
- Cooper, L. N., & Bear, M. F. (2012). The bcm theory of synapse modification at 30: interaction of theory with experiment. *Nature Reviews Neuroscience*, 13(11), 798–810.
- Cooper, L. N., Liberman, F., & Oja, E. (1979). A theory for the acquisition and loss of neuron specificity in visual cortex. *Biological cybernetics*, 33(1), 9–28.
- Debanne, D., Gähwiler, B. H., & Thompson, S. M. (1996). Cooperative interactions in the induction of long-term potentiation and depression of synaptic excitation between hippocampal ca3-ca1 cell pairs in vitro. *Proceedings of the National Academy of Sciences*, 93(20), 11225–11230.
- Debanne, D., Gähwiler, B. H., & Thompson, S. M. (1999). Heterogeneity of synaptic plasticity at unitary ca3–ca1 and ca3–ca3 connections in rat hippocampal slice cultures. *Journal of Neuroscience*, 19(24), 10664–10671.
- Frenkel, M. Y., & Bear, M. F. (2004). How monocular deprivation shifts ocular dominance in visual cortex of young mice. *Neuron*, 44(6), 917–923.
- Froc, M., Williamson, O., & van Rossum, M. (2021). Realistic receptive fields from weight dependent bcm plasticity. *preprint*.
- Gjorgjieva, J., Clopath, C., Audet, J., & Pfister, J.-P. (2011). A triplet spike-timing-dependent plasticity model generalizes the bienenstock–cooper–munro rule to higher-order spatiotemporal correlations. *Proceedings of the National Academy of Sciences*, 108(48), 19383–19388.
- Graupner, M., & Brunel, N. (2012). Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proceedings of the National Academy of Sciences*, 109(10), 3991–3996.
- Guo, J., Liu, Y., Li, Y., Li, F., & Huang, F. (2020). Bienenstock–cooper–munro learning rule realized in polysaccharide-gated synaptic transistors with tunable threshold. *ACS Applied Materials & Interfaces*, 12(44), 50061–50067.
- Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Psychology Press.
- Huang, Y., Liu, J., Harkin, J., McDaid, L., & Luo, Y. (2021). An memristor-based synapse implementation using bcm learning rule. *Neurocomputing*, 423, 336–342.
- Intrator, N., & Cooper, L. N. (1992). Objective function formulation of the bcm theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5(1), 3–17.
- Izhikevich, E. M., & Desai, N. S. (2003). Relating stdp to bcm. *Neural computation*, 15(7), 1511–1523.
- Kristan Jr, W. B. (2016). Early evolution of neurons. *Current Biology*, 26(20), R949–R954.
- Law, C. C., & Cooper, L. N. (1994). Formation of receptive fields in realistic visual environments according to the bienenstock, cooper, and munro (bcm) theory. *Proceedings of the National Academy of Sciences*, 91(16), 7797–7801.
- Loebel, A., Le Bé, J.-V., Richardson, M. J., Markram, H., & Herz, A. V. (2013). Matched pre-and post-synaptic changes underlie synaptic plasticity over long time scales. *Journal of Neuroscience*, 33(15), 6257–6266.
- Mioche, L., & Singer, W. (1989). Chronic recordings from single sites of kitten striate cortex during experience-dependent modifications of receptive-field properties. *Journal of neurophysiology*, 62(1), 185–197.

- Montgomery, J. M., Pavlidis, P., & Madison, D. V. (2001). Pair recordings reveal all-silent synaptic connections and the postsynaptic expression of long-term potentiation. *Neuron*, 29(3), 691–701.
- Morrison, A., Aertsen, A., & Diesmann, M. (2007). Spike-timing-dependent plasticity in balanced random networks. *Neural computation*, 19(6), 1437–1467.
- Nass, M. M., & Cooper, L. N. (1975). A theory for the development of feature detecting cells in visual cortex. *Biological cybernetics*, 19(1), 1–18.
- Owens, D. F., Boyce, L. H., Davis, M. B., & Kriegstein, A. R. (1996). Excitatory gaba responses in embryonic and neonatal cortical slices demonstrated by gramicidin perforated-patch recordings and calcium imaging. *Journal of Neuroscience*, 16(20), 6414–6423.
- Rittenhouse, C. D., Shouval, H. Z., Paradiso, M. A., & Bear, M. F. (1999). Monocular deprivation induces homosynaptic long-term depression in visual cortex. *Nature*, 397(6717), 347–350.
- Rubin, J., Lee, D. D., & Sompolinsky, H. (2001). Equilibrium properties of temporally asymmetric hebbian plasticity. *Physical review letters*, 86(2), 364.
- Turrigiano, G. G., & Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. *Nature reviews neuroscience*, 5(2), 97–107.
- Udeigwe, L. C., Munro, P. W., & Ermentrout, G. B. (2017). Emergent dynamical properties of the bcm learning rule. *The Journal of Mathematical Neuroscience*, 7(1), 1–32.
- Van Rossum, M. C., Bi, G. Q., & Turrigiano, G. G. (2000). Stable hebbian learning from spike timing-dependent plasticity. *Journal of neuroscience*, 20(23), 8812–8821.
- van Rossum, M. C., & Turrigiano, G. G. (2001). Correlation based learning from spike timing dependent plasticity. *Neurocomputing*, 38, 409–415.
- Wang, Z., Zeng, T., Ren, Y., Lin, Y., Xu, H., Zhao, X., ... Ielmini, D. (2020). Toward a generalized bienenstock-cooper-munro rule for spatiotemporal learning via triplet-stdp in memristive devices. *Nature communications*, 11(1), 1–10.
- Wiesel, T. N., & Hubel, D. H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of neurophysiology*, 26(6), 1003–1017.
- Zenke, F., Gerstner, W., & Ganguli, S. (2017). The temporal paradox of hebbian learning and homeostatic plasticity. *Current opinion in neurobiology*, 43, 166–176.

Appendices

A Weights → Activities Mapping

With the input pattern presented in the methods for the 2D case, one can express the activities (y_1 and y_2) due to each pattern as follows:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} \cos \phi & \sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \quad (42)$$

It can be useful to see how the \mathbf{w} and \mathbf{y} spaces map to each other. To do that, one can note how the directions $\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1, 1)$ and $\mathbf{u}_2 = \frac{1}{\sqrt{2}}(1, -1)$ are eigenvectors of the matrix that transforms \mathbf{w} into \mathbf{y}

$$\begin{pmatrix} \cos \phi & \sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \cos \phi + \sin \phi \\ \sin \phi + \cos \phi \end{pmatrix} = [\cos \phi + \sin \phi] \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (43)$$

$$\begin{pmatrix} \cos \phi & \sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} \cos \phi - \sin \phi \\ \sin \phi + \cos \phi \end{pmatrix} = [\cos \phi - \sin \phi] \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (44)$$

with eigenvalues $\lambda_1 = \cos \phi + \sin \phi$ and $\lambda_2 = \cos \phi - \sin \phi$. This is, given a vector \mathbf{w} , one can obtain its mapping to \mathbf{y} by multiplying its component along \mathbf{u}_1 with λ_1 and the component along \mathbf{u}_2 with λ_2 . Furthermore, because of the values λ_i take in the range $[0, \pi/2]$, it will always correspond to a stretching in the \mathbf{u}_1 direction and a compression in the \mathbf{u}_2 .

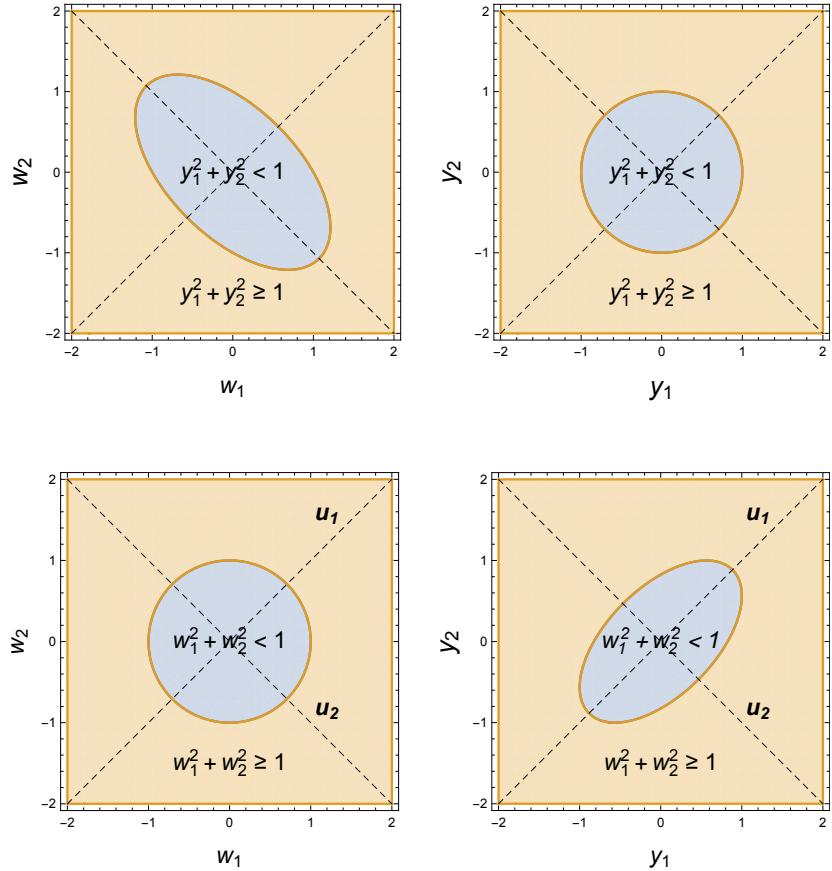


Figure 11: $\mathbf{y} \rightarrow \mathbf{w}$ (Top) and $\mathbf{w} \rightarrow \mathbf{y}$ (Bottom) transformation of a circle. The transformation $\mathbf{y} \rightarrow \mathbf{w}$ produces a stretching along the $\frac{1}{\sqrt{2}}(1, -1)$ direction, while the transformation $\mathbf{w} \rightarrow \mathbf{y}$ produces a stretching along the $\frac{1}{\sqrt{2}}(1, 1)$ direction.

B Derivations

Proposition B.1. If, for a given fixed point, in 2D,

$$\dot{w}_i^{(1)} = -\dot{w}_i^{(2)} \neq 0 \quad \forall i \quad (45)$$

then one can express one weight as a linear function of the other (and inhibition).

Proof. We will follow the derivation for the case where stimulus \mathbf{x}_1 induces LTD and \mathbf{x}_2 induces LTD. Then one has:

$$(w_1 + u)x_{11}y_1(y_1 - \theta) = -x_{21}y_2(y_2 - \theta) \quad (46)$$

$$(w_2 + u)x_{12}y_1(y_1 - \theta) = -x_{22}y_2(y_2 - \theta) \quad (47)$$

(48)

Then, because neither $y_1(y_1 - \theta)$ or $y_2(y_2 - \theta)$ are zero, one can divide the equations and

$$\frac{(w_1 + u)x_{11}}{(w_2 + u)x_{12}} = \frac{x_{21}}{x_{22}} \implies w_1 = \frac{x_{21}x_{12}}{x_{22}x_{11}}w_2 + \left(\frac{x_{21}x_{12}}{x_{22}x_{11}} - 1\right)u \equiv kw_2 + (k-1)u \quad (49)$$

The derivation for the case where each stimulus leads to opposite sign of weight change is identical. \square

Proposition B.2. The Jacobian of the system has always (for all ϕ and u) a discontinuity at the standard BCM fixed points. If the limit of its elements J_{ij} , taken in two different directions, is different, the absolute value of the difference is

$$|(w_j + u)J_{ij}^0 - J_{ij}^0| \quad (50)$$

where J^0 is the value of J evaluated with standard BCM dynamics.

Proof. The Jacobian J is defined as

$$J = \begin{pmatrix} \partial_{y_1} F_1 & \partial_{y_2} F_1 \\ \partial_{y_1} F_2 & \partial_{y_2} F_2 \end{pmatrix}, \quad F_i = \frac{dw_i}{dt} \quad (51)$$

In the mean-field approximation,

$$F_i = [w_i + u]^{d_1} z_{1i} + [w_i + u]^{d_2} z_{2i}, \quad z_{ji} \equiv \frac{x_{ji}y_j(y_j - \theta)}{2\tau_w} \quad (52)$$

In the limit $\mathbf{y} \rightarrow (0, \theta)$, $\mathbf{y} \rightarrow (\theta, 0)$, $\mathbf{y} \rightarrow (\theta, \theta)$ and $\mathbf{y} \rightarrow (0, 0)$ (the standard BCM fixed points) d_1 and d_2 take different values depending on the direction in which the limit is taken (as the system is at the intersection of the different LTP/LTD regimes).

$$J_{ij}(d_1, d_2) = z_{1i}\partial_{y_j}[w_i + u]^{d_1} + z_{2i}\partial_{y_j}[w_i + u]^{d_2} + [w_i + u]^{d_1}\partial_{y_j}z_{1i} + [w_i + u]^{d_2}\partial_{y_j}z_{2i} \quad (53)$$

Here one can use that $z_{ji} \rightarrow 0$ (independently of the direction of the limit) and that $\partial_{y_j}z_{ki} = 0$ for $k \neq j$ and the expression can be simplified:

$$J_{ij}(d_1, d_2) = [w_j + u]^{d_j}\partial_{y_j}z_{ji} \quad (54)$$

In matrix form, depending on the region containing the limit trajectory, one has:

LTD-LTD

$$J = J(1, 1) = \begin{pmatrix} (w_1 + u)\partial_{y_1}z_{11} & (w_1 + u)\partial_{y_2}z_{21} \\ (w_2 + u)\partial_{y_1}z_{12} & (w_2 + u)\partial_{y_2}z_{22} \end{pmatrix} \quad (55)$$

LTD-LTP

$$J = J(1, 0) = \begin{pmatrix} (w_1 + u)\partial_{y_1}z_{11} & \partial_{y_2}z_{21} \\ (w_2 + u)\partial_{y_1}z_{12} & \partial_{y_2}z_{22} \end{pmatrix} \quad (56)$$

LTD-LTD

$$J = J(0, 1) = \begin{pmatrix} \partial_{y_1}z_{11} & (w_1 + u)\partial_{y_2}z_{21} \\ \partial_{y_1}z_{12} & (w_2 + u)\partial_{y_2}z_{22} \end{pmatrix} \quad (57)$$

LTP-LTP

$$J = J(0, 0) = \begin{pmatrix} \partial_{y_1} z_{11} & \partial_{y_2} z_{21} \\ \partial_{y_1} z_{12} & \partial_{y_2} z_{22} \end{pmatrix} \equiv J^0 \quad (58)$$

From these expressions it is clear that, if one takes two directional limits of J_{ij} , its values can either be the same or be one J_{ij}^0 and the other $(w_j + u)J_{ij}^0$, hence the absolute value of the difference will be

$$|(w_i + u)J_{ij}^0 - J_{ij}^0| \quad (59)$$

□

For completeness, in Figure 12, the evolution of the elements of the Jacobian J_{ij} along one of these discontinuities (parametrized by m , $\mathbf{y} = m(1, 0)$) is plotted (solid lines). For J_{11} and J_{21} (the only discontinuous elements), the value of J_{ij}^0 (dotted lines) and $(w_j + u)J_{ij}^0$ (dashed lines) is also included. The plot is in agreement with the theoretical quantification of the discontinuity. In both J_{11} and J_{21} the value of J for $m < 2$, follows the dotted line (J_{ij}^0). Exactly at $m = 2$, which this corresponds to the sBCM fixed point $\mathbf{y} = (2, 0)$, there is a discontinuity and the value of J_{ij} jumps, for $m > 2$, to $(w_j + u)J_{ij}^0$ (dashed line), which results in a total increment of $(w_j + u)J_{ij}^0 - J_{ij}^0$.

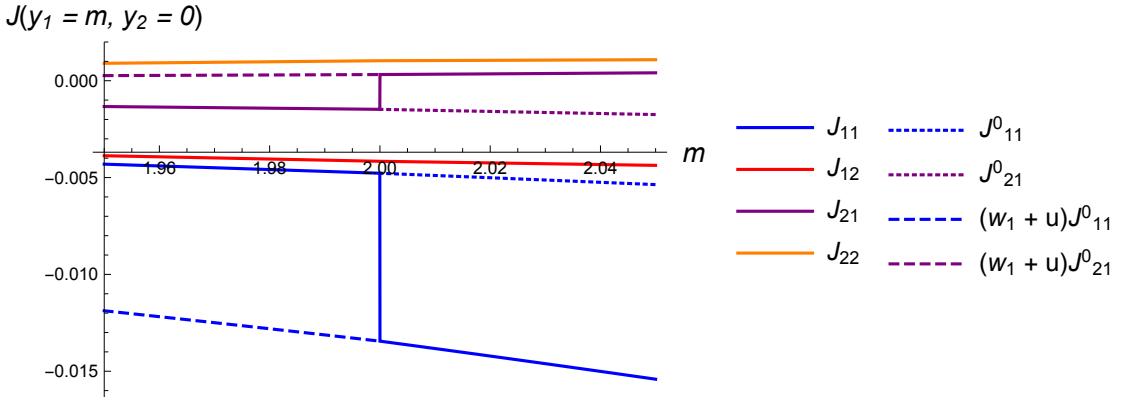


Figure 12: Discontinuities of the Jacobian at the selective BCM fixed point $\mathbf{y} = (0, 2)$ across $\mathbf{y} = m(1, 0)$. Continuous lines represent numerically evaluated values of J_{ij} , dotted lines the values of J_{ij}^0 and dashed lines $(w_j + i)J_{ij}^0$.

Proposition B.3. *Given ϕ , the maximum u for which the i -FPC is compatible (and thus consistent) is the same as the maximum u for which the i -FPC is stable, and has value*

$$u_i^c(\phi) = -\frac{1}{\cos \phi + \sin \phi} \quad (60)$$

Proof. The requirement for *compatibility* is, in the case of i-FPC, that it leads to activities y_1, y_2 such that the system is under LTD-LTD ($\mathbf{d} = (1, 1)$). The i-FCP has solutions $\mathbf{w} = (-u, -u)$, so substituting in equations (13) and (14)

$$y_1 = y_2 = u(\cos \phi + \sin \phi) \quad (61)$$

Because both inputs \mathbf{x}_1 and \mathbf{x}_2 lead to the same activity ($y_1 = y_2$), the system can only be in LTD-LTD or LTP-LTP. Following equation (9), we are interested in finding the value of u for which there is a sign change in $y(y - \theta)$. Substituting y with y_1 (or y_2):

$$y(y - \theta) = u^2(\sin 2\phi + 1)(u(\sin \phi + \cos \phi) + 1) \quad (62)$$

The only sign change must come from the last term (both u^2 and $\sin(2\phi) + 1$ will always be ≥ 0). Making the last term 0 one obtains the critical value of u for which the system transitions from LTD-LTD to LTP-LTP:

$$u(\sin \phi + \cos \phi) + 1 = 0 \implies u = -\frac{1}{\sin \phi + \cos \phi} \quad (63)$$

That it is *the maximum value* (and not *the minimum*) for which the system is in LTD-LTD can be easily checked substituting the expression for some value of u different than the critical.

Now we are interested in doing a similar analysis in terms of its stability. To do that, we will obtain the value of the Jacobian J evaluated at $\mathbf{w} = (-u, -u)$ and see how the sign of roots of its characteristic polynomial is related to u and ϕ . The value of J can be obtained analytically because we know that this type of FPC is always in LTD-LTD.

For this solution, the value of J can be simplified to be, without taking into consideration τ_w ,

$$J = \begin{pmatrix} -\frac{u^2(\sin 2\phi+1)(u(\sin \phi+\cos \phi)+1)}{\tan \phi-1} & -\frac{u^2 \sin \phi(\sin 2\phi+1)(u \cot \phi+u+\csc \phi)}{\cot \phi-1} \\ -\frac{u^2 \sin \phi(\sin 2\phi+1)(u \cot \phi+u+\csc \phi)}{\cot \phi-1} & -\frac{u^2(\sin 2\phi+1)(u(\sin \phi+\cos \phi)+1)}{\tan \phi-1} \end{pmatrix} \quad (64)$$

To obtain the characteristic polynomial we do $\text{Det}(J - \lambda I)$, which has roots (eigenvalues of J):

$$\lambda_1 = \frac{u^2(\sin \phi + \cos \phi)^3(u(\sin \phi + \cos \phi) + 1)}{\cos \phi - \sin \phi} \quad (65)$$

$$\lambda_2 = u^2(\sin 2\phi + 1)(u(\sin \phi + \cos \phi) + 1) \quad (66)$$

Again, in both cases, the terms u^2 and $\sin 2\phi + 1$ have value ≥ 0 . Also, because ϕ is restricted to the interval $(0, \pi/4)$, we always have $\sin \phi + \cos \phi \geq 0$. Both eigenvalues J change sign, thus, due to the same term as *consistency* does: $u(\sin \phi + \cos \phi) + 1$. \square

Proposition B.4. *Given ϕ , the minimum u for which the selective s-FPC are accessible (and thus consistent), has value*

$$u_{s1}^c(\phi) = \frac{2 \sin \phi}{(\cos \phi - \sin \phi)(\cos \phi + \cos \phi)} \quad (67)$$

Proof. For a FPC to be *accessible*, it must be within hard-bounds $w_i \geq -u$. To find the critical value of u at which that happens, we can map the solutions in \mathbf{y} to \mathbf{w} and then obtain the minimum value of u for which the solutions are *accessible*. Because we are interested in the selective FP solutions, we have either $\mathbf{y} = (0, 2)$ or $\mathbf{y} = (2, 0)$. We will follow the derivation for the first case as, given the symmetry of the hard-bounds, the solution must be the same. If $\mathbf{y} = (0, 2)$,

$$w_1(y_1 = 0, y_2 = 2) = \frac{2 \cos \phi}{(\cos \phi - \sin \phi)(\cos \phi + \cos \phi)} \quad (68)$$

$$w_2(y_1 = 0, y_2 = 2) = -\frac{2 \sin \phi}{(\cos \phi - \sin \phi)(\cos \phi + \cos \phi)} \quad (69)$$

From these, the most negative one is w_2 . \mathbf{w} is *accessible* when $w_2 \geq -u$:

$$-\frac{2 \sin \phi}{(\cos \phi - \sin \phi)(\cos \phi + \cos \phi)} \geq -u \implies u \geq \frac{2 \sin \phi}{(\cos \phi - \sin \phi)(\cos \phi + \cos \phi)} \quad (70)$$

So the minimum value of u for which the selective s-FPC is *consistent* is the critical value of the inequality. \square

Proposition B.5. *Given ϕ , the minimum u for which the selective s-FPC are stable, has value*

$$u_{s2}^c(\phi) = \frac{\sin 2\phi}{(\cos \phi - \sin \phi)^2(\cos \phi + \cos \phi)} \quad (71)$$

Proof. We will again follow the derivation for $\mathbf{y} = (0, 2)$. We have seen that in the case of s-FPC the 4 *Jacobians* (the 4 analytical expressions of J that come from the 4 different *regimes* of depression/potentiation) need to have all its eigenvalues strictly negative. To obtain the critical curve, we will omit the evaluation for LTD-LTD, LTP-LTD and LTP-LTP, as we have seen in results that stability depends on the LTD-LTDP dynamics (to strictly follow the full derivation, though, we should also prove that the eigenvalues under these dynamics are also always the highest).

Under LTD-LTP, and not considering τ_w , the selective s-FPC $\mathbf{y} = (0, 2)$ has value of J

$$J = \begin{pmatrix} \tan 2\phi - u \cos \phi & -\sin \phi \\ -\sin \phi(u + 2 \cos \phi \sec 2\phi) & -\cos \phi \end{pmatrix} \quad (72)$$

The eigenvalues expression is omitted due to its length, but both $\lambda_1 = 0$ and $\lambda_2 = 0$ can be solved for u to find the critical value of sign change, and the solution is in both cases

$$u = \frac{\sin 2\phi}{(\cos \phi - \sin \phi)^2 (\cos \phi + \cos \phi)} \quad (73)$$

□