A bridge from neuroscientific models to recurrent neural networks

Derivation of continuous-time connectionist models from neuroscience computational principles

Joan Gort Vicente

Facultat de Psicologia, Universitat Autònoma de Barcelona
joan.gort@e-campus.uab.cat

Dr. Josep Maria Losilla Vidal

May 1st, 2021

Table of contents

Abstract

In the last years, recurrent neural networks with continuous dynamics have been applied to model many neurobiological phenomena. However, the literature on the physiological foundations of these connectionist networks is practically non-existent, as they are closer to artificial neural networks than neuroscientific computational models. In this article, we explicitly derive the equations of these recurrent connectionist systems from neuroscientific models, such as leaky integrate-and-fire (LIF) neurons and synaptic chemical kinetics. We specify under what conditions this modelling is supposed to hold, and we run simulations of networks wired like some simple neural circuits, such as those that possess species like *Tritonia Diomedea, Aplysia Californica* or lampreys, in order to show their similar behaviour. Finally, in the first annex we introduce some of the emerging properties of these networks, such as being universal approximators of dynamical systems, and we remark that this approach is congruent with the spontaneous synchronic activity that is known to take place in the cortex

Keywords: *Connectionism, computational neuroscience, integrate-and-fire models, synaptic chemical kinetics, realistic modelling, approximation theorems, emerging properties, neural manifolds, neural synchronies.*

1. Introduction

The connectionist perspective is a computational approach to the brain-mind binomial which states that the mental activity is an emerging property of neural systems, and therefore the study of that must refer to these. This perspective understands that the brain is not composed of symbolic processing units, enabled to manipulate explicit representations following a set of formal rules (Smolensky, 1988) but that instead these representations are found distributed in parallel, in a sub-symbolic level, spread over a whole set of interconnected elements, where each unit is not related to a single concept, as representations are understood to be a collective, emerging property of the whole net (Rumelhart et al., 1986). This is called Parallel Distributed Processing, or PDP. In practice, this is usually synonymous to neural network models that have some common features, such as being nets of units (neurons) that have an activation given by a nonlinear transformation of a propagation rule (how is each element connected to the neighbouring neurons) that is established by a weight matrix (the strength of the connexions) as well as some biases, determined by certain learning rules in interaction with an environment, where we assume certain stable distributions of probabilities in the presence of stimuli (Rumelhart et al., 1986; Thomas & McClelland, 2012).

Although one of the main objectives of the PDP perspective is to knit a biologically plausible modelling, enabled to provide an understanding about the nature of these emerging properties, the fact is that until now these models have followed a top-down approach, this is, the designed neural networks have been planned specifically to adjust to an expected functioning, by learning different abstract representations of certain given data (Lecun et al., 2015), that can go from pattern classification to speech recognition (Graves & Jaitly, 2014) or even the imitation of social skills (Weizenbaum, 1983). This opens the question of whether these devices are able to mimic some cognitive processes because they share a similar neural basis with real brains, or just because they have been configured to do the same tasks, in which case the PDP perspective would be interesting in machine learning and artificial intelligence, but not for the understanding of real biological information processing systems (van Gerven & Bohte, 2017).

In this article we focus on continuous time connectionist models, expressed in the language of non-linear differential equations, which have common features with discrete neural networks, both in their formalism as well as in their emerging computational properties (Hopfield, 1984; Hopfield & Tank, 1986). In the last years, their dynamics (Beer, 1995) and the stability of their behaviour (Hirsch, 1989) have been studied, and they have been proved to hold very interesting dynamical properties, as to be able to predict both autonomous and non-autonomous dynamical systems with arbitrary precision (Chow & Li, 2000; Funahashi & Nakamura, 1993), being therefore capable of approximating any trajectory on the activation space of the output units of the network (Trischler & D'Eleuterio, 2016).

In addition to these mathematical properties (which we will discuss later, given the relevant interpretations we think they can open in psychology) time continuous networks have been shown to model a very wide range of neurobiological phenomena, such as to mimic human ballistic movements (Cheron et al., 2007), reproducing locomotive patterns

(Hoellinger et al., 2013), interpreting the functioning of cells in the layer IVB of the primary visual cortex (Mineiro & Zipser, 1998), replicating the visual recognition of biological movements (Giese & Poggio, 2003), the EMG recordings of the reaching movements of monkeys (Susillo et al., 2015) or the locomotive patterns of salamanders (Ijspeert, 2001), among others.

However, despite the wide research and implementation, these models seem to merge out of the clear blue sky, as if they were, again, a product of the intentions of the developers, instead of an innocent reflex of the reality that they pretend to represent. We think, therefore, that there's a need to undo the path and seek for the neurobiological origins, to raise the walls on which the roof is already built.

In the following sections, we will derive different continuous time connectionist systems in a bottom-up fashion, reaching them explicitly from previous neuroscientific models, such as "integrate-and-fire" neurons or synaptic models (Gerstner et al., 2014). We will show that continuous time connectionist systems are valid approximations to the dynamics of given populations of neurons, whenever they follow certain conditions, and that they are something more than a useful invention, in the sense that, apart from being endowed with practical mathematical properties, they are also grounded on neuroscientific computational research.

That's why we will defend their relevance both on neuroscience and cognitive psychology. In regard to the first, we will perform different simulations using the derived models, which show that they can replicate the behaviour of some biological neural networks; with respect to the second, we will briefly discuss their emerging computational properties, showing that this approach is congruent with the synchronous spontaneous activity that it is known to occur in the cortex (Okun et al., 2015).

1.1 The ideal neuron

Models give schematic and simplified versions of reality, as on the physical side require idealizations and on the mathematical, approximations. Without idealizing the object of scientific interest, the study of nature would be impossible, given its great complexity. These idealizations are usually given by selecting the relevant elements of the study, simplifying them and approximating the basic laws that govern their dynamics (Fernández Rañada, 2005).

Some of the premises of our modelling are borrowed from "leaky integrate-and-fire" (LIF) neurons, since our units will have the same properties, as taking a single membrane potential value, a unique threshold over which action potentials are triggered, and considering the membrane as a passive capacitor which fires every time the threshold is reached. We will also neglect time delays, presuming they are much shorter than neurons time constants, as well as electrical phenomena associated with the shape of dendrites and axons (Hopfield & Tank, 1986).

With respect to synapses, we will presume them to have some grade of similarity, so that we can group them into different groups with shared properties, like volume, diffusivity, receptor kinetic parameters and density. We will consider different cases, where the

simplest one (this is, the case of a neuron with only one synaptic structure) will give birth to the well-known time continuous recurrent neural networks (Hopfield & Tank, 1986).

In addition, we will consider them capable to optimize the neurotransmitter release, meaning that the maximum output is capable to saturate postsynaptic receptors almost completely.

Finally, the considerations on synaptic dynamics will raise from a biochemical approach, studying their behaviour in terms of chemical equilibrium, enzymatic kinetics, diffusivity, affinity… from here, a set of differential equations will be derived, and their justification in biological and chemical terms will be addressed in the section on synaptic dynamics and on the second annex, where we will discuss the synaptic modelling on which the whole derivation is grounded. With regarding to approximations, relevant to mathematical aspects, they will be mentioned and justified whenever they occur.

1. Derivation of connectionist models

2.1"Integrate-and-fire models" and synaptic current: the basics of the model

"Integrate-and-fire" models are simplifications of the dynamics of a single neuron. Their precision level depends both on the nonlinear terms and the other equations that are added to the simplest one of those models, often called "leaky integrate-and-fire" (LIF)(Gerstner et al., 2014). The so called "adaptative exponential integrate-and-fire", for example, stands out for being able to predict about 95% of the action potentials, if the governing parameters are chosen correctly (Brette & Gerstner, 2005). From now on, however, we will be based on the simplest model, from which the rest are just improvements. This one consists of a passive membrane, whose potential can be obtained from a lineal differential equation, analogous to the one of RC electronic systems. If the potential reaches a certain threshold, a spike is released and the potential is reset (or, in other terms, you "integrate" the equation until a certain potential is reached, "and" then the neuron "fires").

From this analogy, the neuron membrane can be considered as an electric capacitor, able to store a certain amount of charge on both sides of the membrane given a voltage difference (membrane capacitance, C) with a certain tendence of charge filtering (membrane resistance, R). This lineal circuit is certainly capable of approximating the solutions of the potential when it comes to values close to the resting potential, but what about the highly nonlinear situation that takes place when an action potential is triggered? That's why, as we said, a threshold is added on, from which the neuron is assumed to generate a spike, after which the potential is set to a starting value again and again, form which the lineal sub-threshold dynamics are going to take on afterwards. The mathematical conditions are given by:

$$\tau \frac{dV}{dt} = -(V - V_{rest}) + RI(t) \tag{2.1.1}$$

if $V \geq \vartheta$ (the threshold), an action potential is generated and V takes automatically the Vr (reset) value (about -75 mV, given the hyperpolarization phenomena), where τ is the temporal constant of the membrane (which equals the product of the resistance for the

capacitance, RC); V is the membrane potential; V_{rest} is the resting neural potential (usually about -70mV); and *I(t)* is a function that governs the intensity of the injected current on the cell. (Gerstner et al., 2014)

So far, we have just talked about a spiking system, that little or nothing have to do with a connectionist modelling. Our goal will be to derive rate models from spiking ones, and so we have to move from action potentials to frequency.

Figure 2.1.1
F-I curve

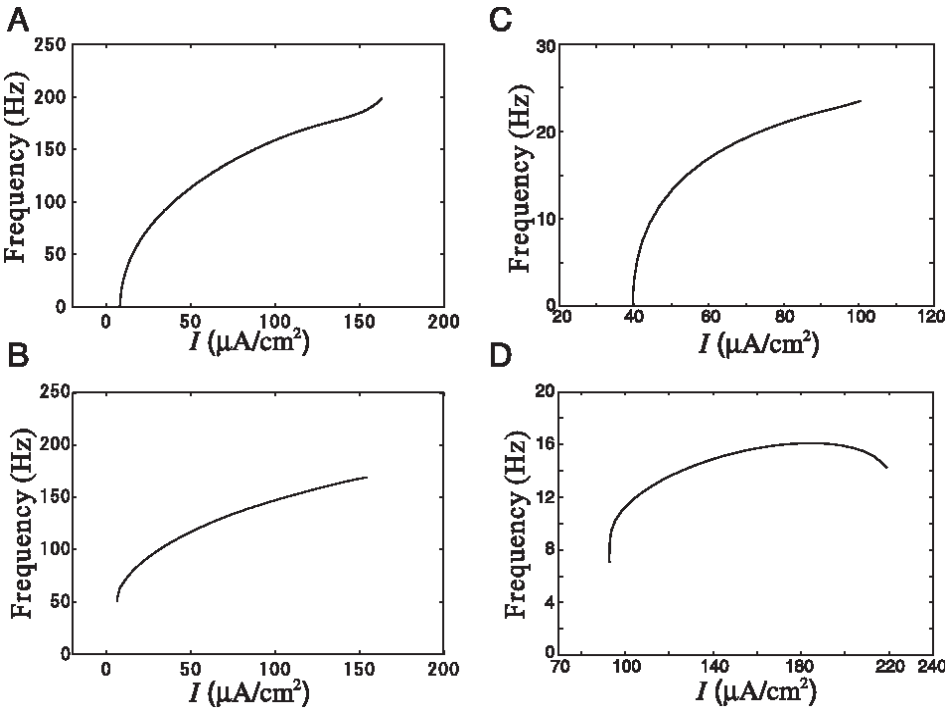


Image from (Tateno et al., 2004)

The image shows what are known as f-I curves, which relate the intensity of the injected current with the emitted action potential frequency (Tateno et al., 2004). The injected current, this is, the "input" of the studied neuron, depends on the output values of the neighbor neurons. Thus, the total membrane current can be thought as a propagation variable, which defines how the surrounding activations define the state of the unit in classical connectionist models (Rumelhart et al., 1986) and so, if we choose the firing frequency in Hertz as the activation units of the network elements, the f-I curve can be viewed as a kind of activation function.

The deal now is how to relate this kind of connectionist propagation with the activation (i.e., the firing rate) of surrounding neurons. To solve this problem from the ground up, we will have to resort to synaptic biochemical dynamics. This, however, will be addressed in the following section. Before, nevertheless, we will rearrange equation (2.1.1), fitting in the known expression for chemical synaptic current, and therefore expressing *I(t)* in terms of membrane potential, membrane permeability and the reversal potential of the ions involved, just as follows:

$$I(t) = \sum_{i,j} I_{ij}(t), \qquad I_{ij}(t) = -g_{ij}(t)\big(V - E_j\big) \tag{2.1.2}$$

Where $g_{ij}(t)$ is the membrane permeability of the *i*th synapse to the "j" ion, I_{ij} is the current associated to the same synapse for the same ion, and E_j is the reversal potential for the mentioned ion (Gerstner et al., 2014). Usually, $g_{ij}(t)$ is assumed to be a time dependent function that decays exponentially after each presynaptic spike. It can also be defined as:

$$g_{ij}(t) = \sum_k \overline{g}_{ijk} r_{ik}(t) \tag{2.1.3}$$

(Destexhe et al., 1994; Gerstner et al., 2014)

Where $\overline{g}_{ijk}$ is the màximum permeability that the k-th kind of ionotropic receptor have for the *j*th ion in the *i*th synapse and r_{ik} is the proportion of these channels that are open. Although it may seem that we're kicking the ball forward without solving the problem out, this expression will result of great value in the following sections.

To ease things and suppress adders and index, from now on we will develop the model for the hypothetic case of a neural network with just one ionotropic receptor permeable to a single ion. This restriction, nevertheless, is not necessary, and so we will soon introduce a general model for diverse receptors and ions. This shift is just to simplify the next steps and make them more accessible, from which we will later be able to generalize the resulting equations to a more realistic situation.

In this case, if we play the Russian dolls with the previous expressions, substituting (2.1.2) in (2.1.1), we find:

$$\tau \frac{dV}{dt} = -(V - V_{rest}) - z(t)\,(V - E) \tag{2.1.4}$$

Where *z* is just *Rg(t)* (the product of the membrane leak resistance for the membrane permeability to input current, which yields the dimensionless variable *z*). We have introduced this new function because, as we will soon see, it will have the honor to be the state variable of the system we are deriving.

To obtain our activation function, we will remember that we measure the activation in frequency units, and that frequency is the inverse of the period. Thus, we will integrate the previous equation as follows (Stein, 1967):

$$T = \int_0^T dt = \int_{Vr}^{\vartheta} \frac{dt}{dV} dV = -\tau \int_{Vr}^{\vartheta} \frac{dV}{V - V_{rest} + gR(V - E)}$$

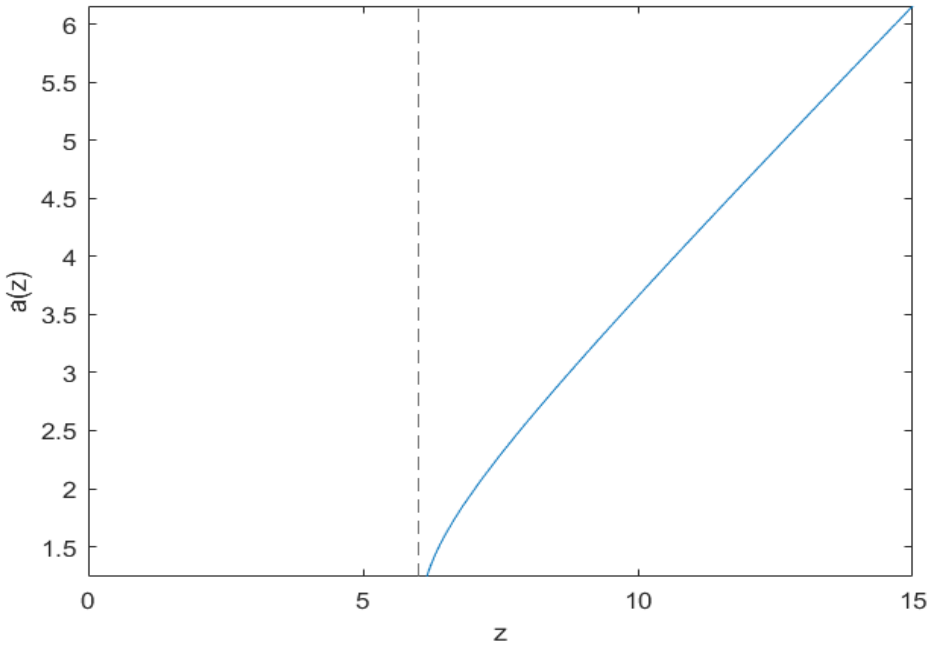
(2.1.5)

This yield:

$$a(z) = \frac{z+1}{\tau \ln \frac{(z+1)Vr - (V_{rest} + zE)}{(z+1)\vartheta - (V_{rest} + zE)}} \tag{2.1.6}$$

The following picture shows the graphics of the activation function (which is as we will label equation (2.1.6)) for the following values of the parameters: *Vr*=-80mV; V_{rest}=-70mV; E=0mV (we have selected sodium to be our reference ion, which has a reversal potential about 0mV); ϑ=10mV, $\tau = 1$.

Figure 2.1.2
Activation Function



As a summary, so far, we have based on "integrate-and-fire" models, that are approximations to the dynamics of a single neuron, which have permitted us to define a state variable, z, that as we will soon see enables us to relate the outcome of the surrounding neurons, just like in the rest of connectionist models happens with the propagation rule. At the same time, we have shown that the values of this function can be transformed into the frequency of the neuron via the activation function, equation (2.1.6). Finally, we have reduced the jungle of receptors and transmitters of the system just to make the next steps more accessible. At the end, however, we will return to more realistic situations.

2.2 Synaptic chemical dynamics

Our goal is to know, given a neural network with *N* elements, the evolution equations of the system, starting from the propagation of the units (z), which we have seen to depend only on the activations ($a(z)$) of the neighbor neurons.

Defined and interpreted these variables from a biological level, and having been categorized according to the connectionist scheme, now it's time to derive the rules that will determine the behavior of this system of dynamical interactions. To begin this task, we will start from the model:

$$\dot{r}_i = \alpha[T]_i(1 - r_i) - \beta r_i \tag{2.2.1}$$

Where r_i is the fraction of ligand bounded receptors in the ith synapse of the studied neuron; $\dot{r}_i$ is the variation rate of the mentioned variable; $[T]_i$ is the concentration of neurotransmitter in the i-th synaptic cleft; α and β are the kinetic coefficients of the reaction (Destexhe et al., 1994; Gerstner et al., 2014).

In the current case, in which we are just interested in neurons endowed with a single kind of receptor (so far), *T* could be glutamate and the receptor could be of the NMDA type, for example. We will study neural interactions only through ionotropic receptors, leaving aside metabotropic signalling.

This kinetic approximation enables us to acknowledge which is the fraction of "activated" neuroreceptors knowing the synaptic concentration of the transmitter, $[T]_i$, that in this model usually takes the shape of a function to choose, depending only on time. We, however, will add a companion to this equation, that will enable to describe $[T]_i$ through the activation state of the presynaptic neuron. This will be the first equation of own harvest, so it will regard a biological justification. Below we will set out the new parameters, and the deserved explanation of the biochemical sense of this new equation will be detailed in annex 2. The system takes the form:

$$\begin{cases} \dot{r}_i = \alpha[T]_i(1 - r_i) - \beta r_i \\ [\dot{T}]_i = -\alpha R_0[T]_i(1 - r_i) + \beta R_0 r_i - \frac{V_{max}[T]_i}{k_M + [T]_i} + \gamma\sigma(z_i) - \frac{DA}{V}[T]_i \end{cases} \tag{2.2.2}$$

Where $R_0 \equiv [R] + [TR]$ is the sum of the free receptor concentration plus that of the one bounded to the transmitter, and can be read as the receptor total concentration, which we will consider constant for our purposes; k_M is the transmitter reuptaker Michaelis-Menten constant; V_{max} is the maximum velocity at which glutamate can be reabsorbed; γ is a new constant of mol/(l*s) units, that we will discuss in the annex; *D* is the diffusion coefficient of the neurotransmitter from the synapse to the external environment, *A* and *V* are, respectively, the area that limits the synapse with the surrounding medium and the total volume of the synapse; and $\sigma(z_i)$ is the output value of the presynaptic neuron, which is defined as the composition of the output function with the activation function. The output function (σ) is defined, given an activation *a(z)*, as:

$$\sigma(z) = \frac{a(z)^n}{a_{\frac{1}{2}}^n + a(z)^n} \tag{2.2.3}$$

Where $a_{\frac{1}{2}}$ is the frequency at which half of the transmitter release is reached, and n is a constant to be determined empirically. For more details about this output function, the issue is discussed in the annex that verses on biochemical argumentation (annex 2).

A dilemma comes to mind when we pose (2.2.2). If what we want is an ODE system able to describe in a simplified and heuristic way the functioning of a biological neural network, how could we obtain it if we assume that each and every synapse, among the thousands with which every single neuron is endowed, already supposes two coupled differential equations? The complexity is huge, and so we will have to reduce the model using both techniques of phase plane analysis and the assumptions that have been pointed out previously.

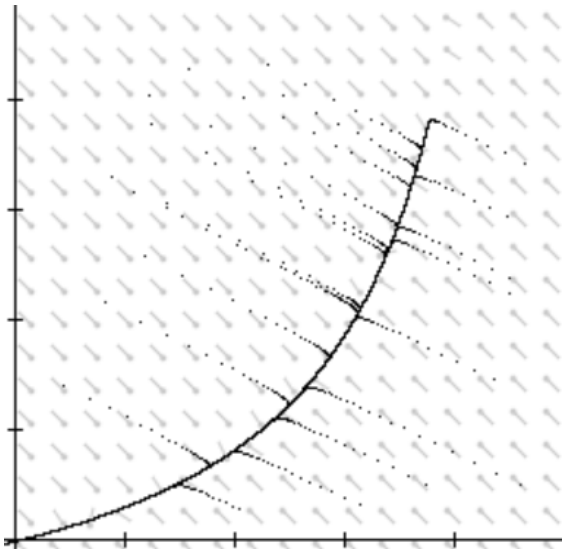
The first thing that comes to sight when analysing (2.2.2) parameters is the difference in its order of magnitude. On the one hand, the receptors are very specific to the transmitter. That's why their kinetic constants should be very large (especially when it comes to α, the parameter that governs the tendency of the neurotransmitter to join the receptor); On the other hand, the performance of the neurotransmitter reuptaker, that works against its gradient and thus belongs to the active transporter family (see annex 2) is going to be much slower compared to kinetic parameters. The same could be said about the diffusion coefficients.

For instance, if we take the NMDA receptors, as we have done before for exemplification purposes, the kinetic parameters take values about $\alpha = 1.7 * 10^{7} M^{-1} s^{-1}, \beta = 60 s^{-1}$ (Borschel et al., 2015), and when it comes to the enzymatic properties of the glutamate transporters, although they can vary largely on many biochemical aspects, for EAAT3 the parameters are known to take values that round, for the Michaellis-Menten constant, $k_M = 2'7 * 10^{-5} M$, (Sun et al., 2014) and for the V_{max} , according to the researchers, the values use to be of the same order of magnitude of k_M, depending on the transporter density.

Intuitively, this implies two different time scales on the behaviour of (2.2.2). On the one hand, the system evolves quickly to the chemical equilibrium of the transmitter-receptor dissociation reaction, which takes the form of a non-isolated fixed points curve on the phase plane, which analytic form is given by the chemical equilibrium equation (see annex 3). But on the other hand, this dissociation reaction doesn't happen on its own. As it is shown in annex 3, there is no conservation of mass inside the synapse, given that the transmitter can flow inward and outward. Thus, once the system has rapidly evolved to the mentioned curve of dissociative equilibrium, it is then slowly driven, thanks to the diffusion/transport/presynaptic signalling phenomena, towards the unique stable equilibrium point of the system.

That's why we could consider this curve of the dissociation reaction equilibrium as a center manifold (Carr, 2006). Afterwards, if we reduce the dynamics of the system to the dynamics on the central manifold (given this huge time-scale difference), the governing equations could be also reduced from two to one single ODE (Strogratz, 1994). The rigorous steps that involve this shift are, again, left to annex 2, where the simplified equation is derived using nondimensionalization and central manifold techniques.

Figure 2.2.1
Phase flow towards the slow manifold



In the image, some trajectories on the phase space of the problem. Separated dots indicate quick movements, while those that are together represent slow ones. It's easy to intuit the two time scales, as well as the central slow manifold.

As far as the synaptic current is concerned, the fraction of open ion channels, r_i, is the only variable to consider. That's why, using the mentioned techniques, we can reduce (2.2.2) to the following equation for r_i:

$$\dot{r_i} = \frac{1}{R_0}\left(\gamma\sigma(z_i) - \frac{V_{max} r_i}{\frac{k_M}{k_d} + \left(1 - \frac{k_M}{k_d}\right) r_i} - \frac{k_d DA}{V}\frac{r_i}{1 - r_i}\right) \tag{2.2.4}$$

Where k_d, the dissociation constant, is the equilibrium constant of the dissociation reaction between the receptor and the transmitter.

2.3 Approximation to a system of connectionist equations

It is time to close the circle, and to do so we will recover the connectionist vocabulary that we've lost in the last pages, when we seek to understand the chemical dynamics of the synapse. Now, we can use these results to conclude the development of the model.

We have claimed that our propagation rule takes the form $z = Rg(t) = R\sum_i g_i(t)$, assuming that just one ion and one receptor are on stage, and that $g_i(t)$, the ion permeability to the i-th synapse, was obtained:

$$g_i(t) = \overline{g}_i r_i(t) \tag{2.3.1}$$

(Destexhe et al., 1994)

To find the equation for the evolution of the state variable of our neuron, (z), which is the golden egg we're looking for, we could be tempted to plug (2.2.4) in the derivative of (2.3.1) and then sum over i. To make it possible, however, we still have to do another approximation before we move on.

Although we have done a great step after dropping one equation of the synaptic dynamics, (2.2.4) is still very complicated, and the fact that it is nonlinear frustrates the previous assignment. But we can account a couple of things.

First of all, if we consider (2.2.4) on the form $\dot{r_i}$=f(r_i, z_i), to simplify the notation, we can see that the domain of f(r_i, z) that concernes us is only the interval when $0<r_i<1$, which is trivial given the definition of r_i. Secondly, f(r_i, z_i) decreases monotonically on this interval as r_i increases (i.e. $\frac{\partial}{\partial r_i}$f($r_i, z_i$) < 0, which is easy to prove) . This means that the dynamics on this narrow segment are not quite complicate (the solutions can either go left or right asymptotically approaching the fixed point). Actually, this is similar to what happens in a linear ODE, so we shouldn't expect much difference between (2.4) and it's linearized version (Strogratz, 1994).

Therefore, it means that inside the (0,1) interval, f(r_i, z_i) doesn't differ that much from a negative slope affine function. Of course f(r_i, z_i) has a vertical assimptote at $r_i = 1$, but if we don't consider the extreme cases when r_i goes very close to that margin, the linear approximation will be acceptable (in fact, from (2.2.4) one can see that 1-r_i should be of $O(k_d DA/V)$ to be taken into consideration, and these parameters take really slow values (annex 3), so the effects of the asymptote on our approximation are negligible). We want to find for what slope our new affine function better fits the dynamics of the activated transmitter rate, r_i. A classical way to linearize would be to expand f(r_i, z_i) into Taylor series on the fixed point, assuming z_i to be a constant parameter, and then ignore the quadratic and higher order terms on r_i. The pity is that z_i is not a constant parameter, but instead a non-autonomous term. This means that our fixed point can move along the interval, and so our linearized dynamics slope would have to depend on z_i, what would make our following steps cumbersome.

With the aim of maintaining our heuristic spirit alive, we think that a good linear approximation to f(r_i, z_i) could be obtained using least squares method, that consists in minimizing the sum (or, in continuous cases like that of our interest, the integral) of the squared differences between the affine function and f(r_i, z_i). That is, we want to find the dynamical system of the form

$$\dot{r_i} = \frac{1}{R_0}(-kr_i + \gamma\sigma(z_i)) \tag{2.3.2}$$

That better mimics the evolution of (2.2.4) in our interval of interests. This is going to be achieved by minimizing the quadratic error like:

$$\frac{\partial E^2}{\partial k} = 0 \ , \qquad \frac{\partial}{\partial k}\int_0^1 (k\, r_i - (\frac{V_{max} r_i}{\frac{k_M}{k_d} + \left(1 - \frac{k_M}{k_d}\right) r_i} - \frac{k_d DA}{V}\frac{r_i}{1 - r_i}))^2 dr_i = 0 \tag{2.3.3}$$

Where k is the absolute value of the slope of the approximation and E^2 is the quadratic error, defined as the infinitessimal sum of the quadratic discrepance between the original function and the approximation. After showing that we can linearize (2.2.4) following this rule, the system will lose precision (the fixed point won't be situated exactly as before, and the rate of change of r_i will also have some discrepancies with respect to the original equation). However, we expect the new equation to reflect the overall evolution of the system as an exponentially fast asymptotic approach to a stable fixed point with a quite similar changing rate. This new approximation allows us to follow our path towards the evolution equations of the state variables of our modelling.

The quadratic error integrand depends just on biochemical parameters, so *k* is a constant specific of our synapse that governs the speed at which it tends to its equilibrium. Because we have assumed that synapses have some degree of similarity, we can set R_0 (this is, the concentration of total receptor) to be the same for all of them, and so we can introduce the next constant, aiming to reduce the number of parameters:

$$\tau \equiv \frac{R_0}{k} \Longrightarrow \ \tau \dot{r_i} = -r_i + \frac{\gamma}{k}\sigma(z_i) \tag{2.3.4}$$

We can simplify further the above expression by using the assumption that synapses optimize the quantity of neurotransmitter release. This means that at the maximum rate of neurotransmitter emission, its concentration in the synaptic cleft saturates the receptors of the postsynaptic cell. This can be expressed mathematically saying that when $\sigma(z_i)$=1 (the release is maximum), $r_i = 1$ (the receptors are saturated) and $\dot{r_i} = 0$ (the sistem lays in the equilibrium). Using this optimality argument, we can eliminate another parameter using (2.3.2):

$$-1 + \frac{\gamma}{k} = 0 \Longrightarrow \ \tau \dot{r_i} = -r_i + \sigma(z_i) \tag{2.3.5}$$

Where τ will be called the "temporal constant" of the system, as it is usually done in classic continuous time models. This simple linearized system allows us, finally, to get to the neural network equations. Since $z = R\sum_i \overline{g}_i r_i = R\overline{g}\sum_i r_i$, as synapses are all twins of each other (as we have assumed), all their maximum permeabilities take the same value, allowing us to remove the index. Then:

$$\tau \dot{z} = \tau R\overline{g}\sum_i \dot{r_i} = R\overline{g}\sum_i \left(-r_i + \sigma(z_i)\right) \tag{2.3.6}$$

On the one hand, when adding up each open channel fraction (r_i) weighted by the dimensionless maximum permeability of each synapse ($R\overline{g}$) we will find, simply, the neuron total permeability to the ion, z. On the other hand, first we remember that $\sigma(z_i)$ stands for the presynaptic emission of transmitter in the *i*th synapse. Of course, not all the synapses are wired to different neurons, so there are groups of synapses tied to the same node of the network. Therefore, if we define z_j not as the state of the specific presynaptic neuron joined to a given synapse, but as the state of the *j*th node of the neural network, and we let "n_j" to be the number of synapses in our studied neuron wired to the same *j*th presynaptic node, then:

$$R\overline{g}\sum_i \sigma(z_i) = R\overline{g}\sum_j n_j\sigma(z_j) \tag{2.3.7}$$

From here, (3.5) allows us to interpret the weights, the parameters that govern the strength of the connexions of the network, in biophysical terms. They can be read as the total maximum permeability to the *j*th presynaptic neuron of our studied node, scaled by the membrane leak resistance. In fact, this product just nondimensionalizes the first quantity, that is, a given node's total maximum permeability to a chosen presynaptic neuron. This is the same that happened to our propagation variable, z, interpreted as the permeability of the neuron to the ion scaled by *R*. Again, this product by the membrane's leak resistance nondimensionalizes the synaptic permeability, which yields a dimensionless parameter, but still with a strong biophysical interpretation. Just as then, this ground-up derivation allows us to enlighten the standard connectionist variables and parameters, allowing us to read them in a biophysical key. Here are the weights algebraic definition:

$$w_j = Rn_j\overline{g} \tag{2.3.8}$$

For all the above, the equation for the dynamics of our neuron of interest would be:

$$\tau\dot{z} = -z + \sum_j w_j\sigma(z_j) \tag{2.3.9}$$

And if instead of a single neuron we are talking of multiple interconnected nodes, the previous would let to the following system of ODEs, which is the final expression for the one ion one receptor model:

$$\tau\dot{z}_i = -z_i + \sum_j^N w_{ij}\sigma(z_j) \tag{2.3.8}$$

Where z_i is the *i*th neuron state function, w_{ij} is the weight between two given nodes, and N is the total number of elements of the net.

2.4 General case and formulation of a complex-valued model

For more realistic situations, where we can be faced with several ions, receptors and transmitters, (2.3.8) would be expressed as follows:

$$\tau_j\dot{z}_{ij} = -z_{ij} + \sum_k^N w_{ijk}\sigma(z_{k1,}\, z_{k2,} \ldots z_{kn}) \tag{2.4.1}$$

Where z_{ij} is the permeability of the *i*th neuron to a given ion through a concrete receptor denoted with the subscript *j*, τ_j is the temporal constant relative to each kind of receptor, *n* is the number of possible combinations between ions and receptors (the number of "z" functions that define the state of a neuron) and $\sigma(z_{k1,}\, z_{k2,} \ldots z_{kn})$ is the output function for the *k*th neuron, σ: $\mathbb{R}^n \to \mathbb{R}$.

The activation function for the general case can be obtained in the same way, and takes the following shape:

$$a\left(z_{k1,}\, z_{k2,}\, \dots z_{kn}\right) = \frac{1 + \sum_l^n z_{kl}}{\tau \ln\left(\frac{(1 + \sum_l^n z_{kl})Vr - V_{rest} - \sum_l^n z_{kl}\, E_l}{(1 + \sum_l^n z_{kl})\vartheta_k - V_{rest} - \sum_l^n z_{kl}\, E_l}\right)} \tag{2.4.2}$$

Where the term E_l refers to the different reversal potentials of the various ions that intervene in the synapses. The sub index *k* refers to the neuron numbered *k*, and *l* to the possible combination of the receptor and the ion in the *l*th position, out of *n* possible ones.

Supposing the relation between frequency (activation) and NT release do not vary between units (a condition which is not necessary, but we will assume for sake of simplicity), the output will be defined just as we did previously:

$$\sigma\left(z_{k1,}\, z_{k2,}\, \dots z_{kn}\right) = \frac{a\left(z_{k1,}\, z_{k2,}\, \dots z_{kn}\right)^n}{a_{\frac{1}{2}}{}^n + a\left(z_{k1,}\, z_{k2,}\, \dots z_{kn}\right)^n} \tag{2.4.3}$$

This model, while complete and detailed, involves throwing overboard all efforts that, until this moment, have been realized in order to elaborate a framework from which an understanding of neural behaviour can be built not only in a realistic manner, but also in a transparent and heuristic way. If we try, however, to interpret the sense of these equations making use of both our assumptions and empirical evidence, we can see that a model more suited to our interests can be revealed.

Our goal will be to put together different synaptic conductance functions (z_{ij}) under a same variable, enabling us to decrease the grade of complexity of (2.4.1). What have stopped us to do so until this moment? Two factors: the fact that different ions have different reversal potentials, which makes it impossible for us to group their conductances inside the activation function $a\left(z_{k1,}\, z_{k2,}\, \dots z_{kn}\right)$, and the fact that every kind of synapse has its own temporal constant, which a priori makes impossible the goal of joining together the different open channel fractions of different kind of synapses under the same permeability function.

As to the first question, it's true that every ion has its own reversal potential, but it is also true that there exists two kind of ionotropic receptors: excitatory and inhibitory. The first ones use to be permeable to both sodium and potassium, leaving a net reversal potential of about $\approx 0mV$. Second ones use to have potentials that round ≈ -70mV. (Purves et al., 2008). It is true that there exists some variability within receptors of the same kind (in excitatory glutamate receptors, NMDA kind receptors are also permeable to calcium, for example), but given the large difference on these potentials between excitatory and inhibitory receptors, we will reduce all the reversal potentials to just a couple of them: one for excitatory transmission and another one for that of inhibitory.

When it comes to the issue of temporal constants, we will follow the same idea that we used before. If we consider just two types of synapses (inhibitory and excitatory) and we presume that everyone of each kind has the same features, like the number or the sort of

receptors, volume, diffusivity (at least for each individual neuron, as we stated previously in our assumptions) then we could estimate the time constant for both excitatory and inhibitory synapses considering the proportion and the number of each type of receptor placed in them.

In fact, we can understand time constants as a quantity proportional to the time that every synapse requires in order to desaturize their postsynaptic receptors when treating them as a whole, this is, as if they had no diversity, once after averaging their properties. In practice, it will become an adjusting parameter in charge of fitting the rate at which a neuron is "heated" or "cooled", although we should not forget that it has units (of time) and a kinetical meaning as well.

If we follow the previous argumentation, considering just neurons equipped with "twin" synapses, we will be placed with the following system of two equations per neuron:

$$\begin{cases} \tau_i^+ \dot{z}_i^+ = -z_i^+ + \sum_j^N {w_{ij}}^+ \sigma(z_i^+, z_i^-) \\ \tau_i^- z_i^- = -z_i^- + \sum_j^N {w_{ij}}^- \sigma(z_i^+, z_i^-) \end{cases} \qquad (2.4.4)$$

Where + and – indices denote excitatory and inhibitory synapses, respectively. This is the set of connectionist equations that we would expect from a LIF network following the condition on synaptic similarity. A quite similar network has been used in the simulation of fish biological movements (Ekeberg, 1993; Ijspeert et al., 2004), so it looks like we're not going astray.

With (2.4.4) we could bring our seek to an end, since it is a model that fulfils all our assumptions and requirements. Nevertheless, this dynamical system can be expressed in a simpler and more compact way if we consider the equivalence between $\mathbb{R}^{2N}$ (the state space of (4.3)) and $\mathbb{C}^N$. Indeed, if we define a complex variable as $z_i = z_i^+ + iz_i^-$, and if we welcome input signals to the pitch (that we have ignored until now for being irrelevant to our purposes) we will be left with the following complex valued system, which is completely equivalent:

$$\alpha_i \dot{z}_i + \beta_i \dot{z}_i^* = -z_i + \sum_j^N w_{ij} \sigma(z_i) + I \qquad (2.4.5)$$

Where $z_i, w_{ij} \in \mathbb{C}$, ${w_{ij}} \equiv {w_{ij}}^+ + i{w_{ij}}^-$, $\alpha_i = \frac{\tau_i^+ + \tau_i^-}{2}$, $\beta_i = \frac{\tau_i^+ - \tau_i^-}{2}$ and $\dot{z}_i^*$ stands for the conjugate of the derivative.

Here, the activation function would be given by:

$$a(z_i) = \frac{1 + Re(z_i) + Im(z_i)}{\tau \ln\left(\frac{\big(1 + Re(z_i) + Im(z_i)\big)Vr - V_{rest} - Re(z_i)E^+ - Im(z_i)E^-}{\big(1 + Re(z_i) + Im(z_i)\big)\vartheta_i - V_{rest} - Re(z_i)E^+ - Im(z_i)E^-}\right)} \qquad (2.4.6)$$

Where $E^+$and E^- are, respectively, the excitatory and inhibitory reversal potentials. The output function remains the same.

2.5 A way to classical real-valued recurrent networks

Equation (2.4.5) resembles a classical time continuous recurrent neural network, if we leave out the fact that in (2.4.5) the state variables are complex-valued, while classical connectionist equations are real-valued. We now present two suppositions under which we can derive the exact classical model from (2.4.5):

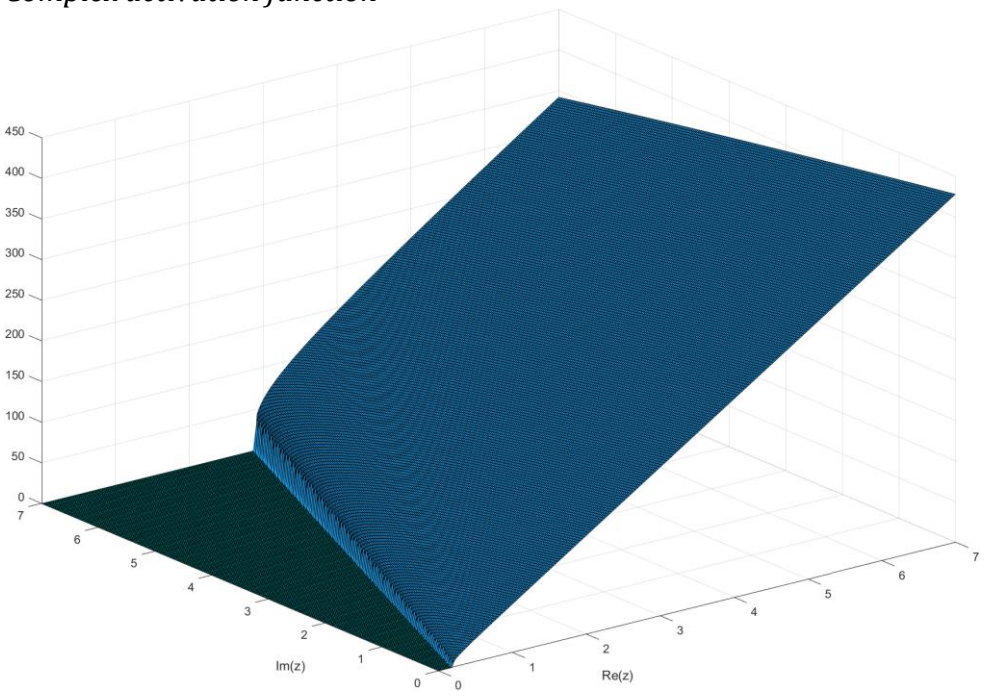
- There is just one time constant per neuron, even if there are both excitatory and inhibitory synapses
- The activation function has a continuous translational symmetry

First one is familiar, and we won't get into details since it follows the same explanation we gave in the previous section. For instance, if we put excitatory and inhibitory synapses in the same bag, and we average their synaptic properties just as we did before, pretending they have similar parameters of diffusivity, affinity, volume… we will end up with a single time constant per neuron, in contrast to the couple of them we found previously.

Again, time constants will play an adjusting role, ruling the speed at which a unit changes it state. Thus, if we consider a highly homogenic situation, where all the synapses have similar kinetic properties regardless of their associated transmitter, this approximation will fit the bill. In the case where we face connexions with different time constants, we will have to use some of the more general models that were announced previously.

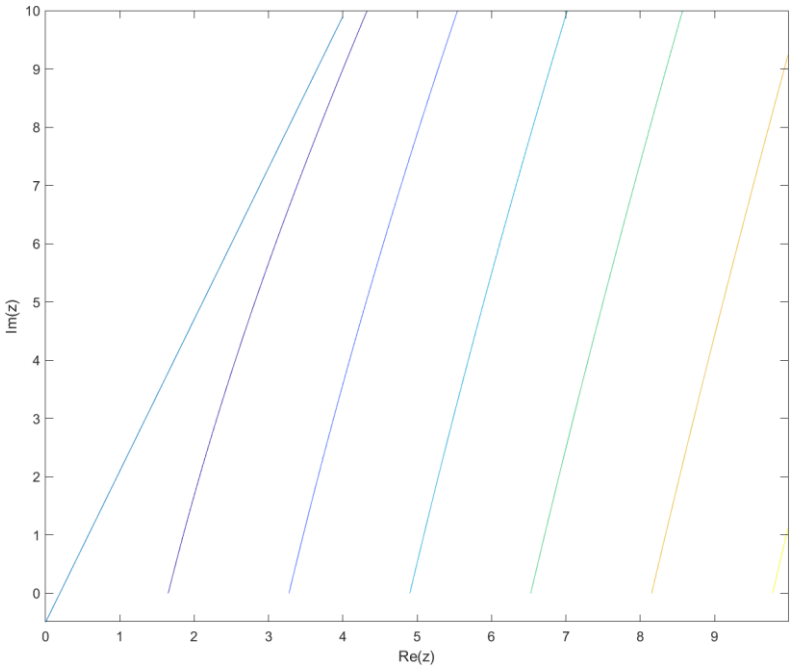
Second one is new to us. The steps will be like those of "canonical transformations" of Hamiltonian systems, in which a transformation of the phase space that leave the equations of motion unchanged is performed (Abramson, 2018). These transformations are useful because they allow to use the symmetries of the Hamiltonian function to reduce the system complexity by leaving some coordinates cyclic (this is, that the Hamiltonian don't depend explicitly on them). We understand a continuous symmetry as a continuous transformation that leaves the Hamiltonian (or in our case, the activation function) unchanged. Similarly, if we had some type of symmetry in the activation function, we could perform a change of variables that could leave one of them cyclic, and thus irrelevant to the evolutionary equations. Now the question is, do we have any sort of symmetry? Strictly speaking, the answer to this question is negative in the general case, but let's take a closer look to the activation function that we have discussed so far:

Figure 2.5.1
Complex activation function



This is the plot of the activation function over the complex plane for typical biophysical values. We can see that it tends asymptotically to a linear function as the real part increases, and that the domain of the mapping can be separated via a linear inequation between the null output region and the positive one. This separation line is not parallel to the contour lines of the plane at which the function tends, although it does not differ that much, as we can appreciate in the following contour plot:

Figure 2.5.2
Contour plot



Now, what would happen if all lines where parallel and straight? Let v be a complex number whose direction is parallel to the contour lines. Then, in this case, if we translate

the function along this direction, the output will remain unchanged for every point, this is,

$$\forall c \in \mathbb{R},\ a(z) = a(z + cv)$$

In this case, where a(z) is invariant under translation, we say that our activation function has a translational symmetry. This allows to express this function on the real domain. If we consider the equivalence $\mathbb{R}^2 \cong \mathbb{C}$, we can perform a change of basis on $\mathbb{R}^2$ by rotating the coordinate axis so that one of them is perpendicular and the other one is parallel to the contour lines. After this change of basis, the function won't depend on the parallel coordinate (we could call it cyclic, in analytical mechanics jargon), and so the activation function will depend on just one real input. We will leave the mathematical details in the annex 4, which yields:

$$(\vartheta - E^-)Im(z) = (E^+ - \vartheta)Re(z) - b \tag{2.5.1}$$

as the equation for the frontier line on the complex domain, where $b = (\vartheta - V_{rest})$ will be called the bias. The change of variables gives:

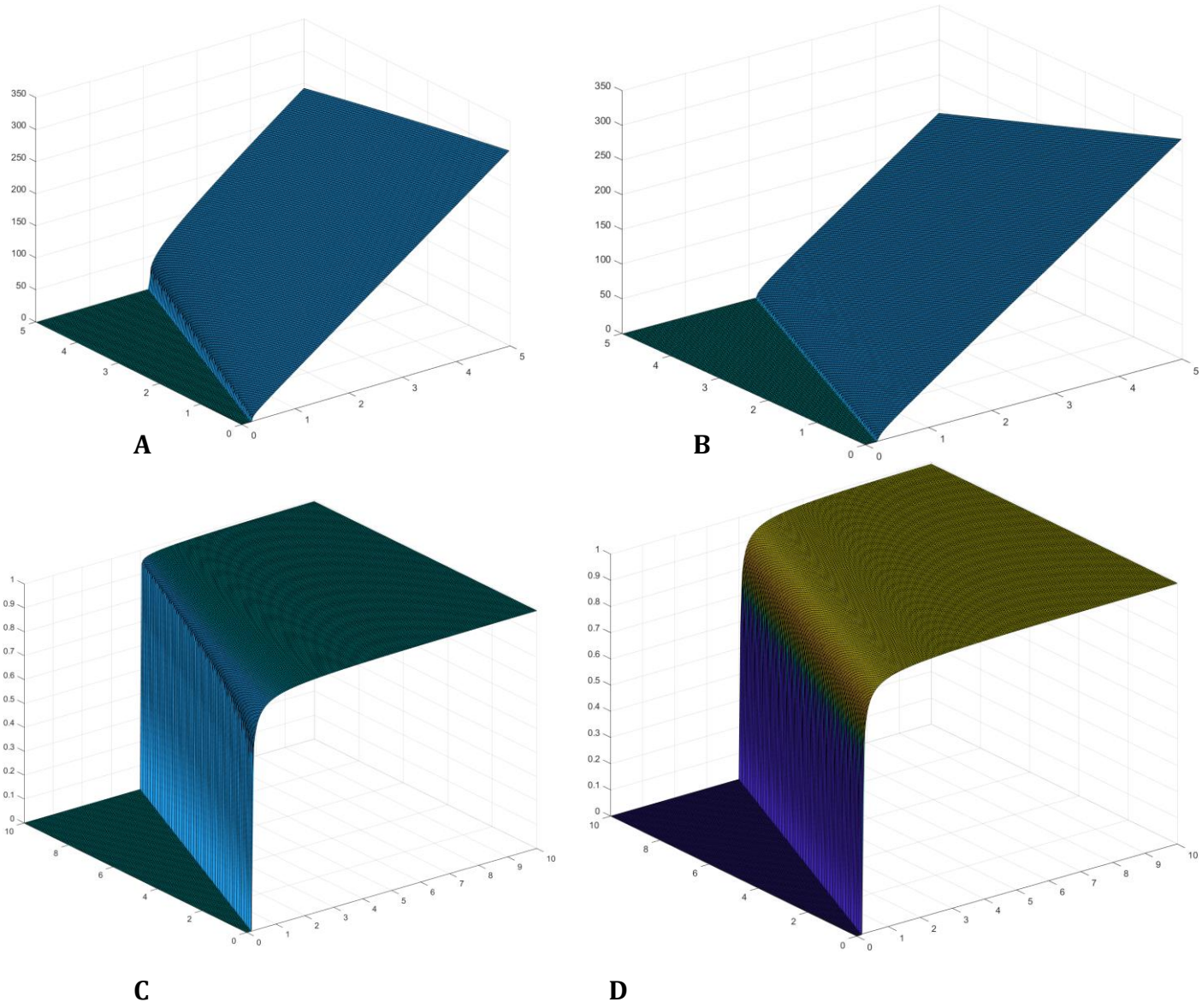
$$y = (E^+ - \vartheta)Re(z) - (\vartheta - E^-)Im(z) \tag{2.5.2}$$

as our new state variable. Now it's time to find an $\mathbb{R} \to \mathbb{R}$ map capable to approximate our $\mathbb{C} \to \mathbb{R}$ activation function, assuming it is almost symmetric. Our choice will be:

$$a(y) = \begin{cases} p\sqrt{y^2 - b^2} & if \ \ y \geq b \\ 0 & if \quad y < b \end{cases} \tag{2.5.3}$$

because it is much easier to compute than the previous one and it has a similar shape (although, probably, not the best fitting one). Here, *p* is just an adjusting parameter with no direct physical meaning. Below, the comparations between the original and the approximated activation function, as well as its composition with the output function:

Figure 2.5.3
Comparations



A) The previously used activation function on the complex domain. B) the same function modified so that it can possess a continuous symmetry. C) the output function defined previously. D) symmetric approximation of the same.

We can see that the approximation is better for low values of *Im(z)*. Thus, the greater the inhibitory conductances are in a neuron, the higher the discrepancies will be.

Once we've stated our approximation choice for the activation function, it's time to derive the new set of evolution equations. From (2.4.4), and taking our assumption about synaptic similarity into account:

$$\begin{aligned}\tau\dot{y}_i &= \tau\big((E^+ - \vartheta)\dot{z}_i^+ - (\vartheta - E^-)\dot{z}_i^-\big) \\ &= -(E^+ - \vartheta)z_i^+ + (\vartheta - E^-)z_i^- \\ &\quad + (E^+ - \vartheta)\sum_j^N {w_{ij}}^+\sigma\big(y_j, b\big) - (\vartheta - E^-)\sum_j^N {w_{ij}}^-\sigma\big(y_j, b\big)\end{aligned} \tag{2.5.4}$$

If we define $w_{ij} \equiv (E^+ - \vartheta){w_{ij}}^+ - (\vartheta - E^-){w_{ij}}^-$ and we add the inputs, we find the equations for the simplified model:

$$\tau\dot{y}_i = -y_i + \sum_j^N w_{ij}\sigma\left(y_j, b_j\right) + I_i \tag{2.5.5}$$

Which matches the classical time-continuous connectionist models that we have discussed in the introduction (Hopfield, 1984; Hopfield & Tank, 1986). We are now allowed to interpret the model variables and parameters in neurobiological terms, and to specify for which cases it will perform a good approximation to the studied phenomenon.

This concludes our derivation.

2. Computational experiments

Regarding the implementation, it's been used MATLAB version R2021a to simulate the previous models. In order to perform numerical integration, we chose the program ode45, which allows to approximate solutions to systems of ordinary differential equations via Runge-Kutta method (Senan, 2017).

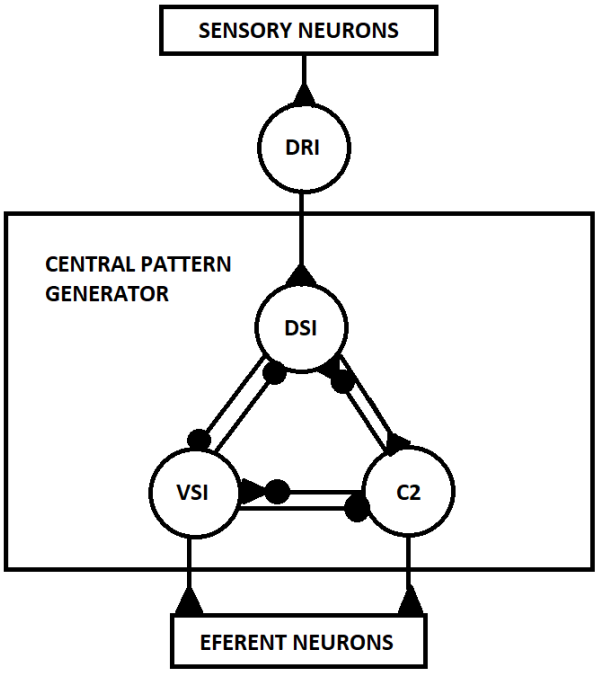
Below, some simulations of biological neural networks are performed, showing that the discussed modelling can replicate the behaviour of real neural circuits. You can find the code for each of the three simulations at the following repository:

https://github.com/joangort/ComputationalExperiments

3.1 Tritonia swim network

Tritonia Diomedea is a nudibranch mollusc that exhibits a characteristic escape-swim pattern executed via ventral and dorsal body flexions (Katz, 2009). The neural mechanisms underlying such cyclic behaviour have been isolated, registered and analysed (Getting, 1981; Popescu & Frost, 2002) and it has been revealed that such exhibited rhythmic patterns are generated by a central mechanism, where sensory inputs don't play any important role apart from triggering the action (Dorsett et al., 1973). Such central mechanism, enabled to create periodic signals by itself, is called a central pattern generator (CPG). Different neural circuits of this kind have been found and studied in different species (Katz, 2016).

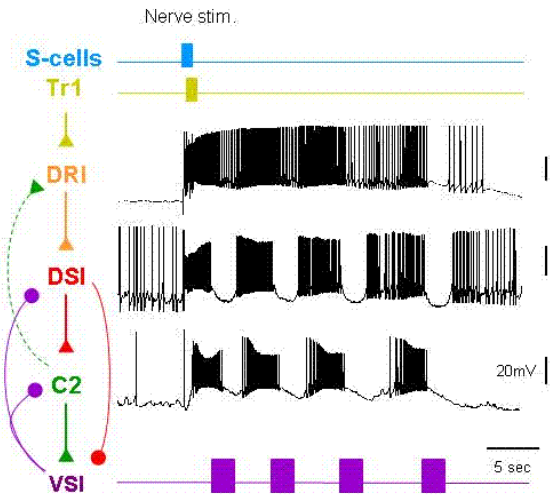
Figure 3.1.1
Tritonia CPG scheme



Schematic representation of Tritonia swim CPG that we've used in our implementation. Triangles represent excitatory synapses and circles inhibitory ones. For a much more detailed study of this network, see (Getting, 1981).

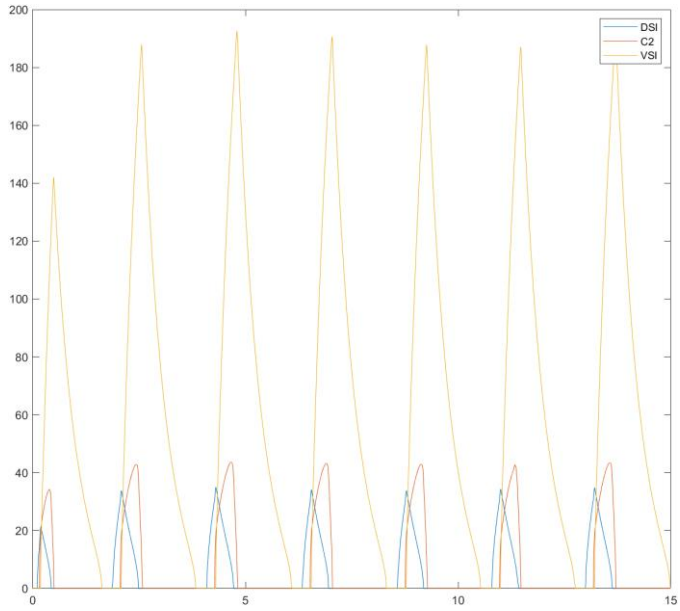
We will wire the derived dynamic recurrent neural network just as in the previous scheme, which boils down the *Tritonia* swim CPG to this simplified version (Katz & Frost, 1997). Below, we show that the resulting dynamics of the network exhibit sustained periodic oscillations in the presence of stimuli, just as in the *Tritonia* CPG.

Figure 3.1.2
Recordings and simulation



A

B



A) Recording of the Tritonia Diomedea CPG, exhibiting oscillatory behaviour (Katz, 2009). *B) Simulation of the connectionist network using the scheme of Fig.1, which also exhibits rhythmic oscillations.*

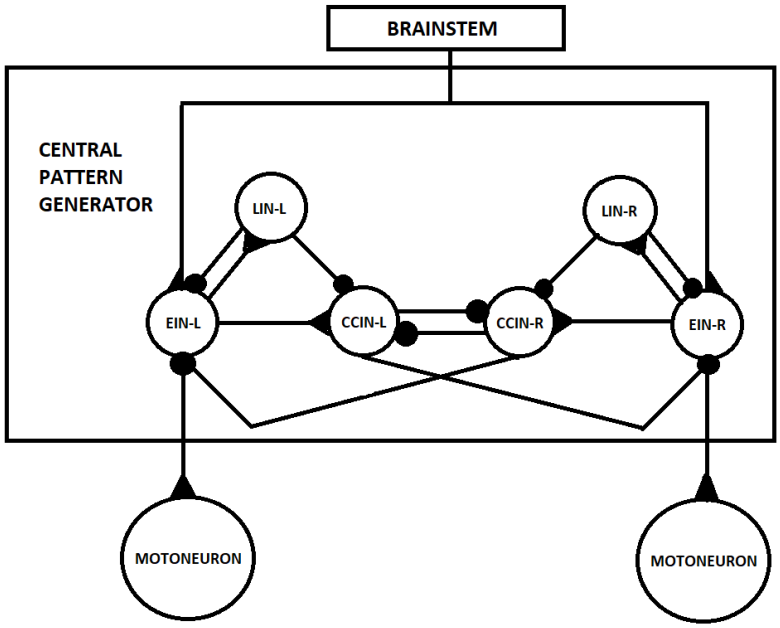
3.2 Half center oscillator

Oscillatory behaviour is widespread along the brain, and so the study of CPGs is of great importance in neuroscience. They could even play an important role in the understanding of the rich dynamics observed in the cortex, as it also behaves in an oscillatory fashion (Yuste et al., 2005). In lots of cases, however, we do not need only a periodic pattern of activity, but also a synchronization between different neurons.

Imagine, for example, the flight of a dragonfly. To raise over, the insect needs to swing their wings up and down. During this fluttering, two muscles need to be synchronized: one to extend the wing up, and another to bend it down. Therefore, when one muscle is tensed, the other needs to be relaxed. The kind of CPGs that allows this synchronized anti-phase pattern are called “half center oscillators”, and they play an important role in movement and locomotion, also in vertebrates and mammals (Katz, 2016; McCrea & Rybak, 2008). The simplest model of these kind of CPGs consists of two reciprocally inhibited neurons with adapting spike patterns (Skinner et al., 1994). Unfortunately, our modelled neurons are unable to reproduce this kind of modulations (as they are, in essence, LIF neurons).

Nonetheless, there have been proposed other kinds of more complex CPGs which can work with simpler neurons, just like ours (Guertin, 2009). We will try to reproduce the half center oscillations of a neural circuit of the Lamprey, a jawless fish endowed with a particularly well studied CPG in its spinal cord (Goulding, 2009). This circuit consists of two groups of interneurons that are wired symmetrically (Grillner et al., 1998; Grillner & Matsushima, 1991), which we schematically represent below.

Figure 3.2.1
Lamprey CPG scheme

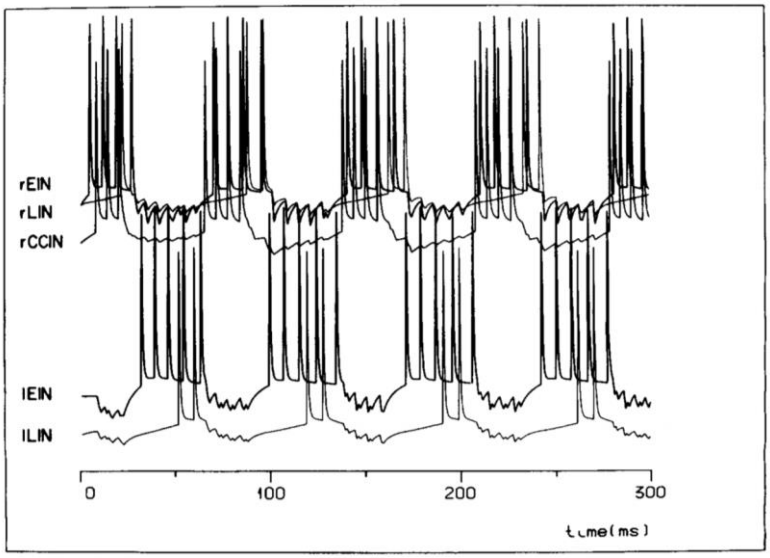


A schematized version of the Lamprey swimming CPG, that we have implemented. For more information about the network and it's functioning, see (Grillner et al., 1995) *or the previously cited articles.*

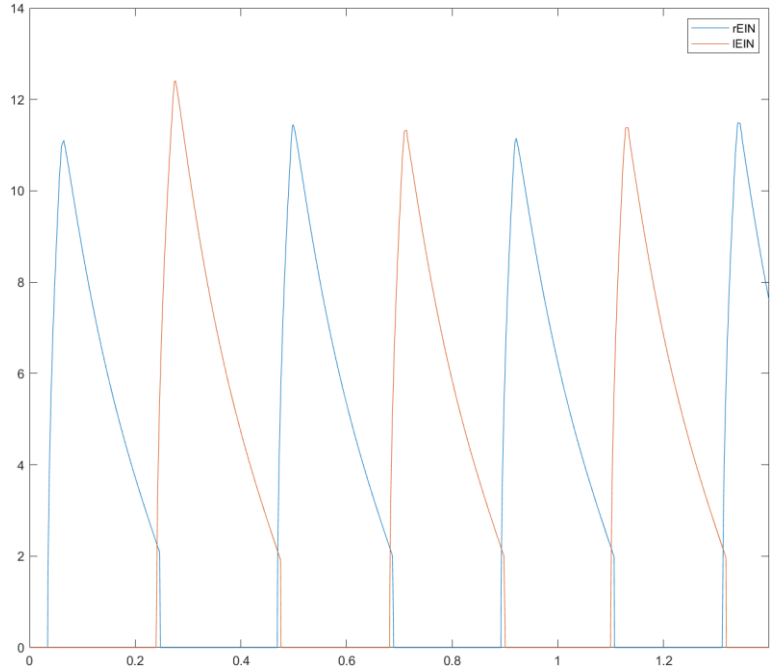
A continuous stimulus of the brainstem origins the half center oscillations of the circuit, which in turn evoque the synchronized swinging of the fish body while swimming (Ekeberg & Grillner, 1999;Grillner et al., 1991). Below, we compare the prototypical functioning of this network with our simulation using the derived continuous recurrent network.

Figure 3.2.2
Simulations

A



B



A) Network behaviour according to (Grillner et al., 1988). *B) Simulation using our model. In both cases, the half center oscillations are appreciable*

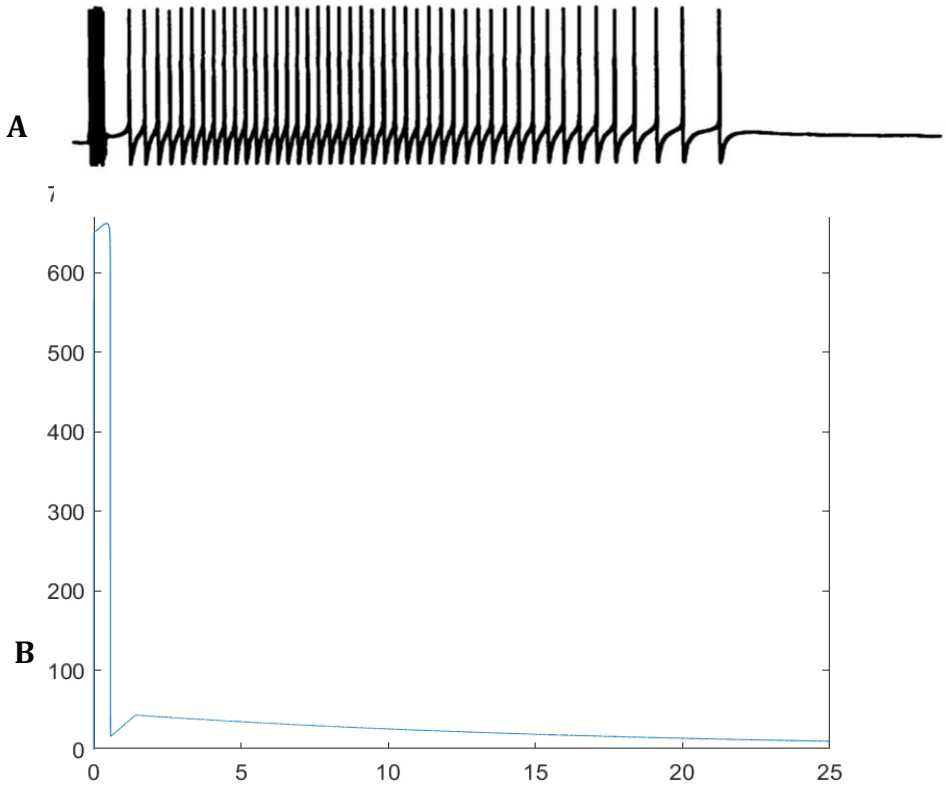
We would like to point out that the simulation of the lamprey swimming had already been done previously using continuous connectionist models. For instance, it's been realized with a system of three-state variables per neuron that is very similar to our complex variable model, as we have discussed previously (Ekeberg, 1993; A. Ijspeert et al., 2004) and with classical dynamic neural networks (A. J. Ijspeert & Kodjabachian, 1999). In both cases, it is shown that continuous time connectionist networks are able to model the lamprey locomotion.

3.3 Aplysia withdrawal reflex

Aplysia Californica is a type of sea slug which has been extensively studied for its withdrawal reflex, similar to that of snail horns, because it has enabled the research about neural mechanisms involved in sensitization and habituation processes (Castellucci et al., 1970; Rayport & Schacher, 1986) as well as classical conditioning studies (Carew et al., 1981; Hawkins et al., 1983).

Here, we will present a simulation of its neural circuit using our connectionist modelling. This neural network can be schematized as a set of sensory neurons that wire with some motoneurons as well as with some interneurons, which postsynaptic time constants are way higher than those of sensory neurons (White et al., 1993). This fails to comply the suppositions we made previously about synaptic similarity, so the real valued recurrent network won't hold here. Therefore, in this occasion we will have to use a form of the previous higher dimensional generalized models (see section 2.4). We set the parameters the way it's done in the mentioned paper (see "Table 1" and "Table 2" of the same). Below, we compare the results after a brief stimulation of the sensory neurons.

Figure 3.3.1
Simulations



A) Motoneuron response according to a simulation in White et al., 1993. B) motoneuron response using our connectionist modelling. In both cases, we can see a slow initial burst followed by a moment of decreased activity, after which the motoneuron spikes tonically and gradually turns off.

3. Conclusions

Throughout these pages, we have provided a method to derive continuous-time connectionist networks from neuroscientific computational models. Starting from spiking LIF neurons and the biochemical kinetics of the synapse, we have been able to obtain different neural networks that, depending on the assumptions we make, are meant to simulate the behaviour of neural structures with specific conditions.

This effort is aimed as an allegation of the biological plausibility that corresponds to these models, and by that we mean connectionist networks with continuous dynamics.

If it is true that the restrictions that made possible the series of approximations that led to this model narrow the diversity of neurons which they are capable to represent, ignoring neural populations endowed with more complex firing patterns (Gerstner et al., 2014) or connexions with a large difference on their kinetic parameters within the same neuron, the existence of some groups of neurons which follow these rules allows to apply some

well-known results about this kind of networks to the study of at least some kind of neural populations.

We have shown that such kind of circuits can be found, by reproducing the behaviour of some simple neural circuits using numerical solutions of the derived models, that gave similar behaviours to those observed.

However, although these simulations have given promising results, they are not exhaustive and are based on simplified versions of the real networks, and thus we consider that their significance is restricted. Although we think that showing a direct path towards their derivation from previous models is in itself an evidence of validation, we think it's also necessary to sum direct evidence, based on empirical data, to widen the validation on the performance of continuous connectionist models. This task has only been timidly hinted in the performed simulations, and we think a wider research in this direction is necessary.

Nevertheless, we think that the previous work can help to consolidate the theoretical bridge that links connectionist foundations to neurobiological modelling. Besides, on the one hand, our line of reasoning could be applied starting from more complex models than LIF, like Generalized LIF or Adaptative Exponential Integrate-and-Fire models, which would allow to increase the range of described firing patterns and frequency responses (Gerstner & Brette, 2009; Teeter et al., 2018) and thus widening the class of represented neurons, at the expense of increasing the number of state variables of the system. Thus, we think that the kind of rate models like those we derived could be of interest in neuroscientific research.

On the other hand, as far as psychology is concerned, the dynamical properties of these kind of networks could gave birth to a new set of scientifical interpretations within cognitive perspectives, as their mathematical features, like the ones we discuss in annex 1, could be labelled in terms of learning, behaviour and information processing, allowing to infer the computational capabilities that emerge collectively from neural activity.

Therefore, we think that research on these systems should be spurred beyond their technological applications, as they could constitute a point of access to the link between neural and mental phenomena.

4. References

Abramson, G. (2018). *Mecánica clásica*. http://ricabib.cab.cnea.gov.ar/696/1/mecanica_2.pdf

Beer, R. D. (1995). On the Dynamics of Small Continuous-Time Recurrent Neural Networks. *Adaptive Behavior*, *3*(4), 469–509. https://doi.org/10.1177/105971239500300405

Bernard, C. (2012). Alterations in synaptic function in epilepsy. In *Jasper's Basic Mechanisms of the Epilepsies*. https://doi.org/10.1111/j.1528-1167.2010.02828.x

Borschel, W. F., Cummings, K. A., Tindell, L. K., & Popescu, G. K. (2015). Kinetic contributions to gating by interactions unique to N-methyl-D-aspartate (NMDA) receptors. *Journal of Biological Chemistry*, *290*(44), 26846–26855. https://doi.org/10.1074/jbc.M115.678656

Brette, R., & Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, *94*(5), 3637–3642. https://doi.org/10.1152/jn.00686.2005

Carew, T. J., Walters, E. T., & Kandel, E. R. (1981). Classical conditioning in a simple withdrawal reflex in Aplysia californica. *Journal of Neuroscience*, *1*(12), 1426–1437. https://doi.org/10.1523/jneurosci.01-12-01426.1981

Carr, J. (2006). Center Manifold. In *Scholarpedia* (p. 1(12):1826). http://www.scholarpedia.org/article/Center_manifold

Castellucci, V., Pinsker, H., Kupfermann, I., & Kandel, E. R. (1970). Neuronal Mechanisms of Habituation and Dishabituation of the Gill-Withdrawal Reflex in Aplysia. *Science*, *167*(3926), 1745–1748. https://doi.org/10.1126/science.167.3926.1745

Cheron, G., Cebolla, A. M., Bengoetxea, A., Leurs, F., & Dan, B. (2007). Recognition of the physiological actions of the triphasic EMG pattern by a dynamic recurrent neural network. *Neuroscience Letters*, *414*(2), 192–196. https://doi.org/10.1016/j.neulet.2006.12.019

Chow, T. W. S., & Li, X. D. (2000). Modeling of continuous time dynamical systems with input by recurrent neural networks. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *47*(4), 575–578. https://doi.org/10.1109/81.841860

Destexhe, A., Mainen, Z. F., & Sejnowski, T. J. (1994). *An Efficient Method for Computing Synaptic Conductances Based on a Kinetic Model of Receptor Binding*. 3–6.

Dorsett, D. A., Willows, A. O. D., & Hoyle, G. (1973). Neuronal basis of behavior in Tritonia. IV. The central origin of a fixed action pattern demonstrated in the isolated brain. *Journal of Neurobiology*, *4*(3), 287–300.

Ekeberg, Ö. (1993). A combined neuronal and mechanical model of fish swimming. *Biological Cybernetics*, *69*(5–6), 363–374. https://doi.org/10.1007/bf00199436

Ekeberg, Ö., & Grillner, S. (1999). Simulations of neuromuscular control in lamprey swimming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *354*(1385), 895–902. https://doi.org/10.1098/rstb.1999.0441

Fernández Rañada, A. (2005). 2.6 Modelos, sus aspectos físicos y matemáticos. In A.

Retif (Ed.), *Dinámica Clásica* (p. 657). Fondo de Cultura Económica.
Funahashi, K. I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks*, *2*(3), 183–192. https://doi.org/10.1016/0893-6080(89)90003-8
Funahashi, K. ichi, & Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, *6*(6), 801–806. https://doi.org/10.1016/S0893-6080(05)80125-X
Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural Manifolds for the Control of Movement. *Neuron*, *94*(5), 978–984. https://doi.org/10.1016/j.neuron.2017.05.025
Gallego, J. A., Perich, M. G., Naufel, S. N., Ethier, C., Solla, S. A., & Miller, L. E. (2018). Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature Communications*, *9*(1), 1–13. https://doi.org/10.1038/s41467-018-06560-z
Gerstner, W., & Brette, R. (2009). Adaptive exponential integrate-and-fire model. In *Scholarpedia* (p. 4(6):8427). https://doi.org/doi:10.4249/scholarpedia.8427
Gerstner, W., Kistler, W., Naud, R., & Paninski, L. (2014). *Neuronal Dynamics: From single neurons to networks and models of cognition*. Cambridge University Press. https://neuronaldynamics.epfl.ch/
Getting, P. A. (1981). Mechanisms of pattern generation underlying swimming in Tritonia. I. Neuronal Network Formed by Monosinaptic Connexions. *Journal of Neurophysiology*, *46*(1), 65–79. https://doi.org/10.1152/jn.1983.49.4.1017
Giese, M. A., & Poggio, T. (2003). Cognitive neuroscience: Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*(3), 179–192. https://doi.org/10.1038/nrn1057
Goulding, M. (2009). Circuits controlling vertebrate locomotion: moving in a new direction. *Nature Reviews Neuroscience*, *10*(7), 507–518. https://doi.org/10.1038/nrn2608
Graves, A., & Jaitly, N. (2014). Towards end-to-end speech recognition with recurrent neural networks. *31st International Conference on Machine Learning, ICML 2014*, *5*, 3771–3779.
Grillner, S., Buchanan, J. T., & Lansner, A. (1988). Simulation of the segmental burst generating network for locomotion in lamprey. *Neuroscience Letters*, *89*(1), 31–35. https://doi.org/10.1016/0304-3940(88)90476-4
Grillner, S., Deliagina, T., El Manira, A., Hill, R. H., Orlovsky, G. N., Wallén, P., Ekeberg, Ö., & Lansner, A. (1995). Neural networks that co-ordinate locomotion and body orientation in lamprey. *Trends in Neurosciences*, *18*(6), 270–279. https://doi.org/10.1016/0166-2236(95)80008-P
Grillner, Sten, Ekeberg, Ö., El Manira, A., Lansner, A., Parker, D., Tegnér, J., & Wallén, P. (1998). Intrinsic function of a neuronal network — a vertebrate central pattern generator. *Brain Research Reviews*, *26*(2), 184–197. http://www.sciencedirect.com/science/article/pii/S0165017398000022
Grillner, Sten, & Matsushima, T. (1991). The neural network underlying locomotion in lamprey-synaptic and cellular mechanisms. *Neuron*, *7*(1), 1–15. https://doi.org/10.1016/0896-6273(91)90069-C
Grillner, Sten, Wallén, P., Brodin, L., & Lansner, A. (1991). Neuronal network generating locomotor behavior in lamprey: Circuitry, transmitters, membrane properties, and simulation. *Annual Review of Neuroscience*, *14*, 169–199. https://doi.org/10.1146/annurev.ne.14.030191.001125
Guertin, P. A. (2009). The mammalian central pattern generator for locomotion. *Brain*

Research Reviews, *62*(1), 45–56. https://doi.org/10.1016/j.brainresrev.2009.08.002

Hawkins, R. D., Kandel, E. R., Prince, E., Physiol, J. C., Mpitsos, G. J., & Cohan, C. S. (1983). *A Cellular Mechanism of Classical Conditioning in Aplysia: Activity-Dependent Amplification of Presynaptic Facilitation*. *219*(4583), 400–405. https://doi.org/10.1126/science.6294833

Hirsch, M. W. (1989). Convergent activation dynamics in continuous time networks. *Neural Networks*, *2*(5), 331–349. https://doi.org/10.1016/0893-6080(89)90018-X

Hoellinger, T., Petieau, M., Duvinage, M., Castermans, T., Seetharaman, K., Cebolla, A. M., Bengoetxea, A., Ivanenko, Y., Dan, B., & Cheron, G. (2013). Biological oscillations for learning walking coordination: Dynamic recurrent neural network functionally models physiological central pattern generator. *Frontiers in Computational Neuroscience*, *7*(MAY), 1–15. https://doi.org/10.3389/fncom.2013.00070

Hopfield, J J. (1984). Neurons with graded response have collective. *Biophysics*, *81*(May), 3088–3092.

Hopfield, John J., & Tank, D. W. (1986). Computing with neural circuits: A model. *Science*, *233*(4764), 625–633. https://doi.org/10.1126/science.3755256

Ijspeert, A., Hallam, J., & Wyse, D. (2004). *Evolving Swimming Controllers for a Simulated Lamprey with Inspiration from Neurobiology*. *June*.

Ijspeert, A. J. (2001). A connectionist central pattern generator for the aquatic and terrestrial gaits of a simulated salamander. *Biological Cybernetics*, *84*(5), 331–348. https://doi.org/10.1007/s004220000211

Ijspeert, A. J., & Kodjabachian, J. (1999). Evolution and development of a central pattern generator for the swimming of a lamprey. *Artificial Life*, *5*(3), 247–269. https://doi.org/10.1162/106454699568773

Jolivet, R., Lewis, T. J., & Gerstner, W. (2004). Generalized integrate-and-fire models of neuronal activity approximate spike trains of a detailed model to a high degree of accuracy. *Journal of Neurophysiology*, *92*(2), 959–976. https://doi.org/10.1152/jn.00190.2004

Katz, B., & Miledi, R. (1970). Further study of the role of calcium in synaptic transmission. *The Journal of Physiology*, *207*(3), 789–801. https://doi.org/10.1113/jphysiol.1970.sp009095

Katz, P. S. (2009). Tritonia swim network. In *Scholarpedia* (p. 4(5):3638). https://doi.org/10.4249/scholarpedia.3638

Katz, P. S. (2016). Evolution of central pattern generators and rhythmic behaviours. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1685). https://doi.org/10.1098/rstb.2015.0057

Katz, P. S., & Frost, W. N. (1997). Removal of spike frequency adaptation via neuromodulation intrinsic to the Tritonia escape swim central pattern generator. *Journal of Neuroscience*, *17*(20), 7703–7713. https://doi.org/10.1523/jneurosci.17-20-07703.1997

Kvalheim, M. D. (2018). *Aspects of Invariant Manifold Theory and Applications*. *https://deepblue.lib.umich.edu/bitstream/handle/2027.42/147642/kvalheim_1.pdf?sequence=1*

Lay, D. (2007). Álgebra Lineal Y Sus Aplicaciones. In *Pearson Educación* (Vol. 53, Issue 9). http://publications.lib.chalmers.se/records/fulltext/245180/245180.pdf%0Ahttps://hdl.handle.net/20.500.12380/245180%0Ahttp://dx.doi.org/10.1016/j.jsames.2011.03.003%0Ahttps://doi.org/10.1016/j.gr.2017.08.001%0Ahttp://dx.doi.org/10.1016/j.precamres.2014.12

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Llinás, R. R. (1982). Calcium in Synaptic Transmission. *Scientific American*, *247*(4), 56–65.

Luczak, A., Barthó, P., & Harris, K. D. (2009). Spontaneous Events Outline the Realm of Possible Sensory Responses in Neocortical Populations. *Neuron*, *62*(3), 413–425. https://doi.org/10.1016/j.neuron.2009.03.014

Magi, S., Piccirillo, S., Amoroso, S., & Lariccia, V. (2019). Excitatory amino acid transporters (Eaats): Glutamate transport and beyond. *International Journal of Molecular Sciences*, *20*(22). https://doi.org/10.3390/ijms20225674

McCrea, D. A., & Rybak, L. A. (2008). Organization of mammalian locomotor rhythm and pattern generation. *Brain Research Reviews*, *57*(1), 134–146. https://doi.org/10.1016/j.brainresrev.2007.08.006

Mineiro, P., & Zipser, D. (1998). Analysis of Direction Selectivity Arising from Recurrent Cortical Interactions. *Neural Computation*, *10*(2), 353–371. https://doi.org/10.1162/089976698300017791

Okun, M., Steinmetz, N. A., Cossell, L., Iacaruso, M. F., Ko, H., Barthó, P., Moore, T., Hofer, S. B., Mrsic-Flogel, T. D., Carandini, M., & Harris, K. D. (2015). Diverse coupling of neurons to populations in sensory cortex. *Nature*, *521*(7553), 511–515. https://doi.org/10.1038/nature14273

Popescu, I. R., & Frost, W. N. (2002). Highly dissimilar behaviors mediated by a multifunctional network in the marine mollusk Tritonia diomedea. *Journal of Neuroscience*, *22*(5), 1985–1993. https://doi.org/10.1523/jneurosci.22-05-01985.2002

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (2008). *Neuroscience*. Sinauer Associates, Inc.

Rabinovich, M. I., Varona, P., Selverston, A. I., & Abarbanel, H. D. I. (2006). Dynamical principles in neuroscience. *Reviews of Modern Physics*, *78*(4), 1213–1265. https://doi.org/10.1103/RevModPhys.78.1213

Rayport, S. G., & Schacher, S. (1986). Synaptic plasticity in vitro: Cell culture of identified Aplysia neurons mediating short-term habituation and sensitization. *Journal of Neuroscience*, *6*(3), 759–763. https://doi.org/10.1523/jneurosci.06-03-00759.1986

Rojas, R. (1996). The Backpropagation Algorithm. *Neural Networks*, 149–182. https://doi.org/10.1007/978-3-642-61068-4_7

Rumelhart, D. E., Hinton, G. E., & Mcclelland, J. L. (1986). A General framework for Parallel Distributed Processing. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, January 1986*, 45–76.

Senan, N. A. F. (2017). *A brief introduction to using ode45 in MATLAB*. *1*(5), 1–6.

Skinner, F. K., Kopell, N., & Marder, E. (1994). Mechanisms for oscillation and frequency control in reciprocally inhibitory model neural networks. *Journal of Computational Neuroscience*, *1*(1–2), 69–87. https://doi.org/10.1007/BF00962719

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*(1), 1–23. https://doi.org/10.1017/S0140525X00052432

Stein, R. B. (1967). The frequency of nerve action potentials generated by applied currents. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *167*(6), 64–86. https://doi.org/10.1098/rspb.1967.0013

Strogratz, S. H. (1994). *Nonlinear Dynamics and Chaos*. Westview Press.

Sun, W., Shchepakin, D., Kalachev, L. V., & Kavanaugh, M. P. (2014). Glutamate transporter control of ambient glutamate levels. *Neurochemistry International*,

73(1), 146–151. https://doi.org/10.1016/j.neuint.2014.04.007

Susillo, D., Churchland, M. M., Kaufman, M., & Shenoy, K. V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nature Neuroscience*, *18(7)*, 1025–1033.

Tateno, T., Harsch, A., & Robinson, H. P. C. (2004). Threshold firing frequency-current relationships of neurons in rat somatosensory cortex: Type 1 and type 2 dynamics. *Journal of Neurophysiology*, *92*(4), 2283–2294. https://doi.org/10.1152/jn.00109.2004

Teeter, C., Iyer, R., Menon, V., Gouwens, N., Feng, D., Berg, J., Szafer, A., Cain, N., Zeng, H., Hawrylycz, M., Koch, C., & Mihalas, S. (2018). Generalized leaky integrate-and-fire models classify multiple neuron types. *Nature Communications*, *9*(1), 1–15. https://doi.org/10.1038/s41467-017-02717-4

Thomas, M. S. C., & McClelland, J. L. (2012). Connectionist Models of Cognition. *The Cambridge Handbook of Computational Psychology*, *44*(0), 23–58. https://doi.org/10.1017/cbo9780511816772.005

Trischler, A. P., & D'Eleuterio, G. M. T. (2016). Synthesis of recurrent neural networks for dynamical system simulation. *Neural Networks*, *80*, 67–78. https://doi.org/10.1016/j.neunet.2016.04.001

Tsodyks, M., Kenet, T., Grinvald, A., & Arieli, A. (1999). *Linking Spontaneous Activity of Single Cortical Neurons and the Underlying Functional Architecture*. *286*(5446), 1943–1946. https://doi.org/10.1126/science.286.5446.1943

van Gerven, M., & Bohte, S. (2017). Artificial Neural Networks as Models of Neural Information Processing. *Frontiers in Computational Neuroscience*.

Weizenbaum, J. (1983). ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine. *Communications of the ACM*, *26*(1), 23–28. https://doi.org/10.1145/357980.357991

White, J. A., Ziv, I., Cleary, L. J., Baxter, D. A., & Byrne, J. H. (1993). The role of interneurons in controlling the tail-withdrawal reflex in Aplysia: A network model. *Journal of Neurophysiology*, *70*(5), 1777–1786. https://doi.org/10.1152/jn.1993.70.5.1777

Yuste, R., MacLean, J. N., Smith, J., & Lansner, A. (2005). The cortex as a central pattern generator. *Nature Reviews Neuroscience*, *6*(June), 477–483. https://doi.org/10.1016/B978-075066268-0/50005-6

Annex 1

Emerging properties and their link to neural synchronization

So far, it has been shown that continuous time connectionist models can be understood as a simplification of the dynamics of biological neural networks for certain kinds of neurons, and that they are therefore capable to simulate their behaviour, as we have just shown. However, there exist many other neuroscientific models that are able to perform more accurate simulations of this kind, being able to predict even individual spikes with different grades of precision (Brette & Gerstner, 2005; Jolivet et al., 2004; Teeter et al., 2018). So why are these connectionist models supposed to be useful? And in what terms can they contribute to psychology?

The answer we give is that connectionist models have been meticulously studied for their long list of applications in both science and engineering. By now, we have tried to show that this modelling does not only offer applications for prediction and classification, but also descriptions of real neural processes. Since the mathematical formalism is the same, we can use these results to infer the emerging computational properties of real neural circuits. In other words, we can use some results on "artificial neural networks" to understand biological ones, since we have shown that the behaviour of at least some kinds of neural circuits can be assimilated to that of continuous-time recurrent networks.

For instance, it opens the possibility to apply a well-known theorem, that states the following:

Theorem 1:(Funahashi & Nakamura, 1993)

Let D be an open subset of $\mathbb{R}^n$, $F: D \to \mathbb{R}^n$ a C^1mapping and $K \subset D$ a compact subset. Then, for any solution $\boldsymbol{x}(t) \in K$of an initial value problem of the form

$$\dot{\boldsymbol{x}}(t) = F\big(\boldsymbol{x}(t)\big)\,, \qquad \boldsymbol{x}(0) \in K\,,$$

defined for $t \in I$, where $I = [0, T], 0 < T < \infty$, and given an arbitrarily small $\varepsilon > 0$, there exists a recurrent neural network with *n* output units and *m* hidden units such that

$$\forall t \in I, \| \boldsymbol{x}(t) - \boldsymbol{u}(t) \| < \varepsilon$$

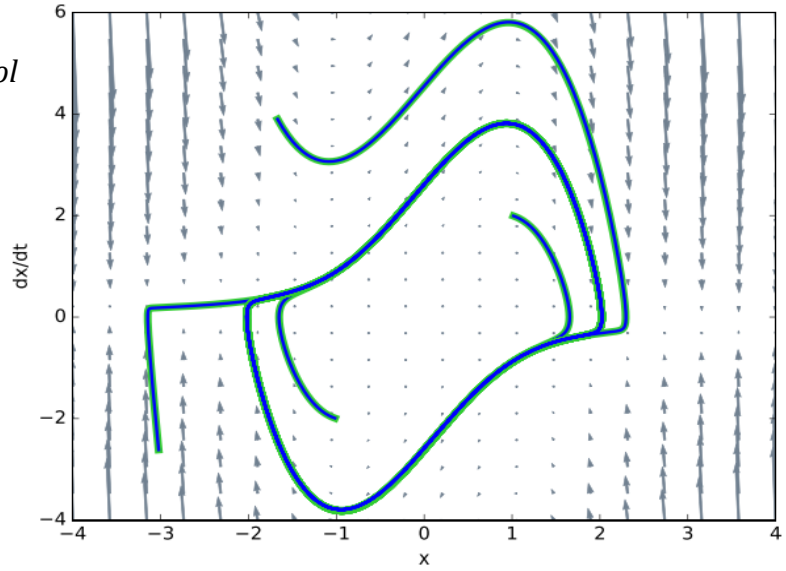
Where $\boldsymbol{u}(t)$ is the vector of the internal states (what we have called propagation or state variables throughout these pages) of the output units of the network.

This means that a neural network like those we derived previously, of the form:

$$\tau \dot{y}_i = -y_i + \sum_{j}^{n+m} w_{ij}\sigma\big(y_j, b_j\big) \qquad (a.1.1)$$

can approximate arbitrarily well any dynamical system that lays on a bounded region during some interval of time, if it has enough hidden units.

Figure A.1.1
Fitting the van der Pol oscillator



Plot showing some trajectories of the van der Pol system. In green, the original solutions; in blue, the ones approximated by a recurrent neural network, according to Trischler & D'Eleuterio, 2016. *The overlaping shows how accurate these aproximations can be.*

This represents an emerging property that, since some neural circuits have been shown to behave as recurrent networks of this kind, can also be attributed as a characteristic of some real neural populations, stating that these possess the ability of approximating any kind of output trajectory, this is, any kind of pattern of behaviour.

And there's a clear example of how this modelling can be useful to psychology, in order to infer the immanent dynamic ground on which cognitive and behaviour processes are hold in the brain, from where this or that conduct is raised after learning and conditioning.

By studying these emerging properties of neural structures, like that of being universal approximators of dynamical systems, we think there can be built a systematic way to shorten the shadows that wander between stimuli and behaviour, the long-awaited noon of cognitive science, using mathematical methods.

But how do these kind of networks approximate dynamical flows? First, we assume a system where the output units are uniquely determined by the hidden units, with no reciprocal connections between these layers. Then, the weight matrix takes the following form:

$$W = \begin{pmatrix} 0 & A \\ 0 & C \end{pmatrix}$$

and the vector of the state values of the system can be written as $\boldsymbol{y}(t) = (\boldsymbol{u}(t), \boldsymbol{h}(t))^T$, where $\boldsymbol{u}$ stands for the vector of the output units and $\boldsymbol{h}$ for the hidden ones. A and C are, respectively, $n \times m$ and $m \times m$ matrices. Therefore, the evolution of the output units of such a network, using the derived connectionist model, can be written in vector notation like:

$$\tau \dot{\boldsymbol{u}} = -\boldsymbol{u} + A\sigma(\boldsymbol{h}, \boldsymbol{b}) \qquad (a.1.2)$$

Where τ is the time constant, which we will suppose the same for all neurons of the whole network, in order to ease the following steps, just as it's done in the proof of theorem 1 (Funahashi & Nakamura, 1993). The components of the vector $\sigma(\boldsymbol{h}, \boldsymbol{b})$ are just $\sigma(h_i, b_i)$, where σ is a single-valued function that is continuous and bounded, called sigmoid function (in the previous pages, a function coherent with this definition has been called output function).

But the previous equation is set to approximate a dynamical system, like the one defined in theorem 1, inside a domain *K*, and therefore:

$$\forall \boldsymbol{u} \in K, \quad \dot{\boldsymbol{u}} = \tilde{F}(\boldsymbol{u}) \rightarrow A\sigma(\boldsymbol{h}, \boldsymbol{b}) = \tau\tilde{F}(\boldsymbol{u}) + \boldsymbol{u} \qquad (a.1.3)$$

Where $\tilde{F}$ is the resulting mapping that approximates the desired flow. In the above equation, the right-hand side depends only on $\boldsymbol{u}$, and the left-hand side does so for $\boldsymbol{h}$ (**b** is a constant vector, not a variable). How could the system set $A\sigma(\boldsymbol{h}, \boldsymbol{b})$ so that $\tilde{F}$ mimics the target flow as accurately as possible? The *universal approximation theorem* states that a feed-forward network with output $A\sigma(B\boldsymbol{u}, \boldsymbol{b})$ can approximate any continuous function on a compact domain with an arbitrary grade of precision, given enough hidden units, *m,* and an appropriate configuration of the parameters (Funahashi, 1989). Although the theorem is more general, here we will stick to $\mathbb{R}^n \rightarrow \mathbb{R}^n$ mappings, since $\tilde{F}$ belongs to this class. Therefore, B is a $m \times n$ matrix.

In order to find the best set of parameters *A*, *B* , $\boldsymbol{b}$ there exists some well-known methods, like the backpropagation algorithm (Rojas, 1996), which is massively used in machine learning. Although the biological plausibility of these algorithms is still a subject of research (Lecun et al., 2015), the universal approximation theorem assures that, regardless of the learning algorithm, such configuration exists.

A consequence about this setting towards the approximation of the flow is that

$$\boldsymbol{h} = B\boldsymbol{u} \qquad (a.1.4)$$

which is a consequence of the universal approximation theorem. These implies that hidden states depend linearly on output units, and thus the trajectories on the hidden space must rely on a *n*-dimensional linear subspace, in order to approximate the mentioned dynamical system. Not only that, but this linear manifold is also attracting.

Here's the reasoning: the dynamics on the hidden space are given by

$$\tau\dot{\boldsymbol{h}} = -\boldsymbol{h} + C\sigma(\boldsymbol{h}, \boldsymbol{b}) \qquad (a.1.5)$$

analogously to the system (a.1.2). From (a.1.4) we can see that

$$\tau\dot{\boldsymbol{h}} = \tau B\dot{\boldsymbol{u}} = -B\boldsymbol{u} + BA\sigma(\boldsymbol{h}, \boldsymbol{b}) \qquad (a.1.6)$$

for every $\boldsymbol{h}$ laying on the mentioned subspace where the flow can be approximated.

This implies, for equation (a.1.6) to equal (a.1.5), that *C=BA*, since *W* does not depend on $\boldsymbol{h}$. The matrix *B* results from the optimization procedure, and its columns are nothing but the set of the vectors that span the subspace where $\boldsymbol{h}$ should lie for $\boldsymbol{u}$ to approximate a given flow, since (a.1.4) is given in vector parametric form (Lay, 2007). We could call these vectors $\boldsymbol{v}_i$, :

$$B = (\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n) \qquad (a.1.7)$$

Now we will define a set of vectors $\{\boldsymbol{n}_1, \boldsymbol{n}_2, \cdots, \boldsymbol{n}_{m-n}\}$ which are both linearly independent and perpendicular to the manifold defined by (a.1.4). To see the evolution of the system projected onto each of these normal vectors, we can compute the scalar product of these vectors with both sides of (a.1.5):

$$\tau\langle \boldsymbol{n}_i, \dot{\boldsymbol{h}}\rangle = -\langle \boldsymbol{n}_i, \boldsymbol{h}\rangle + \langle \boldsymbol{n}_i, BA\sigma(\boldsymbol{h}, \boldsymbol{b})\rangle$$

But since $\langle \boldsymbol{n}_i,\ \boldsymbol{v}_j\rangle = 0$, as the vectors $\boldsymbol{n}_i$ are defined to be perpendicular to the subspace, the term $\langle \boldsymbol{n}_i, BA\sigma(\boldsymbol{h}, \boldsymbol{b})\rangle$ must vanish, since the product of *B* by any *n*-dimensional vector gives a linear combination of the set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$. As $\langle \boldsymbol{n}_i, \frac{d}{dt}\boldsymbol{h}\rangle = \frac{d}{dt}\langle \boldsymbol{n}_i, \boldsymbol{h}\rangle$, since these normal vectors are, of course, constant, we find the following linear differential equation:

$$\tau\frac{d}{dt}\langle \boldsymbol{n}_i, \boldsymbol{h}\rangle = -\langle \boldsymbol{n}_i, \boldsymbol{h}\rangle \qquad (a.1.8)$$

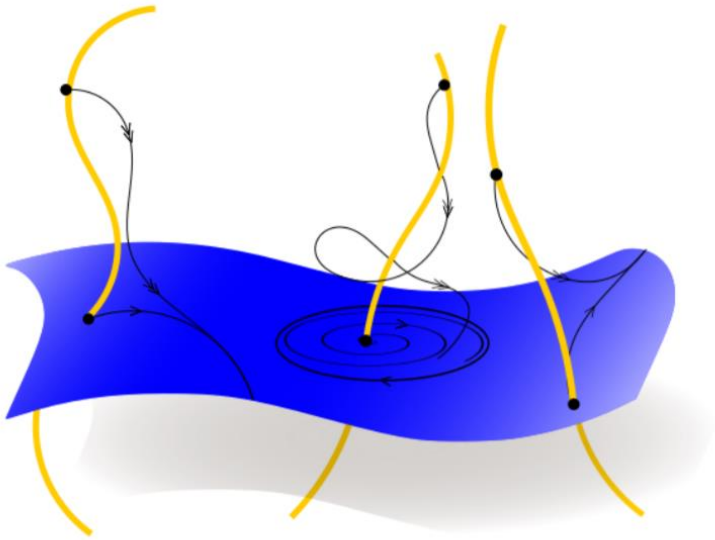
The solution of the previous equation yields:

$$\langle \boldsymbol{n}_i, \boldsymbol{h}\rangle = \langle \boldsymbol{n}_i, \boldsymbol{h_0}\rangle e^{\frac{-t}{\tau}} \qquad (a.1.9)$$

Which says that the scalar product of any normal vector $\boldsymbol{n}_i$ with any trajectory laying in the hidden unit's phase space tends to zero exponentially fast. Therefore, the manifold on which the approximation of a dynamical system is performed is an invariant, stable, lower dimensional and globally attracting set where all the phase trajectories tend as time goes on.

We have been able to proof that for any system of the form (a.1.1) with feed-forward connections between the hidden and the output layer, whose output's internal states are able to mimic a given flow for some compact domain, by making use the approximation universality of feedforward networks (Funahashi, 1989), there exists a unique globally and asymptotically stable manifold on the hidden units space where all the phase trajectories tend exponentially fast.

Figure a.1.2
Attracting invariant manifold



A stable manifold like the ones we have just discussed in neural spaces. We can see that this subspace is invariant, meaning all trajectories embedded in the manifold remain wandering there forever. Illustration from (Kvalheim, 2018).

This result is conclusive with neurophysiological data. It's been reported a phenomenon by which different neuron firings tend to couple, existing a correlation between their activities that cause their patterns to live in a lower dimensional manifold of the neural space (Gallego et al., 2017, 2018; Rabinovich et al., 2006).

This coupling restricts the possible dynamics of the population to a subspace embedded in the larger space of possible neural configurations, in a synchronous activity that have been registered in the cortex of different animals (Okun et al., 2015; Tsodyks et al., 1999) which is not merely a product of sensory inputs, but rather characteristic of the neural population itself (Luczak et al., 2009).

This can be explained if we understand the neural structures of the cortex as universal approximators of dynamical systems. Indeed, the previous analysis showed that all networks which adopt a series of optimization procedures, based on a feed-forward network structure optimization, include a linear attracting invariant manifold on the hidden states space. Of course, we have seen in section 2.5 that the state variables are nothing but a combination of the synaptic conductances, and hence are not directly measurable. But the firing rates could be obtained by applying the activation function. From here we can conclude that there must exist an attracting manifold also in the activation space of the hidden neurons, that will take the form:

$$\boldsymbol{a} = a(\boldsymbol{h}) = a(B\boldsymbol{u}) \qquad (a.1.10)$$

This manifold is not linear, but since the activation function is continuous and locally bounded, the resulting subspace will keep these properties.

As we see, this implies that the firing rate vector spontaneously tends to a manifold, and it remains there in the lack of perturbations. This implies that neural activities are not independent, but rather tied to each other, since they are restricted to a lower-dimensional subspace, where the activity of some units are determined by some others, which implies a correlation between them. This is exactly the kind of phenomenon that has been reported to take place in the cortex, so it could be a signal that some populations of neurons really behave as universal approximators of dynamical systems, which is a specific characteristic of the networks we have derived in the present work.

With this, we have shown that we can apply at least some results derived from the study of well-known neural networks to infer the emerging properties that real neural populations could hold in the nervous system, and that the understanding of neural circuits as universal approximators of flows is coherent with the synchronies that have been already observed in the cortex (Okun et al., 2015).

Annex 2

Biochemical justification of the adopted synaptic model

The first equation of the system (2.2.2) is just equation (2.2.1), already discussed, which describes the tax variation of r_i. The second equation, that we added on, describes the change of $[T]_i$, and to study it in detail we can rewrite it as:

$$[\dot{T}]_i = -R_0\dot{r}_i - \frac{V_{max}[T]_i}{k_M+[T]_i} + \gamma b(\phi_i) - \frac{DA}{V}[T]_i \qquad (a.2.1)$$

The term $-R_0\dot{r}_i$ is deduced from the chemical kinetic of the reaction between the transmitter and the receptor in absence both of transporters, reuptakers and presynaptic release. The stoichiometric equation of the mentioned reaction is:

$$T + R \leftrightharpoons TR \qquad (a.2.2)$$

Where *T* is the transmitter, *R* the free receptor and *TR* the ligand bonded receptor. As it can be seen, for every mole of *TR* that is produced a mole of each reactant is decreased, so ignoring the rest of the synaptic biochemical aspects, the reaction velocity of *TR* shall be of the same magnitude and opposite sign as that of T, relation that is represented by the first term of (4.3), since $-R_0\dot{r}_i$=-$[\dot{TR}]_i$.

The above can also be justified by the law of the conservation of mass. If the system is closed (in the dissociation reaction stated in (a.2.2), the system doesn't exchange mass with the outside) the amount of mass is therefore constant. This can be stated, through the conservation of the total concentration of the transmitter, in the following way:

$$[\text{T}]+[\text{TR}]=\text{constant} \Rightarrow$$
$$\frac{d}{dt}([T]+[TR]) = [\dot{T}] + [\dot{TR}] = [\dot{T}] + R_{0i}\dot{r} = 0 \Rightarrow$$
$$[\dot{T}] = -R_{0i}\dot{r}$$

However, the synapse is not a closed system. We have, on the one side, transporters, reuptakers or enzymes that diminishes the synaptic concentration of transmitter, and on the other side the presynaptic axon button, that tends to increase it. Besides, we should also considerate the diffusion of the transmitter outside of the synapse. That's why we should incorporate the terms that specify which way this exchange is given in the above expression.

Figure a.2.1
Schematic representation of the glutamatergic synapse

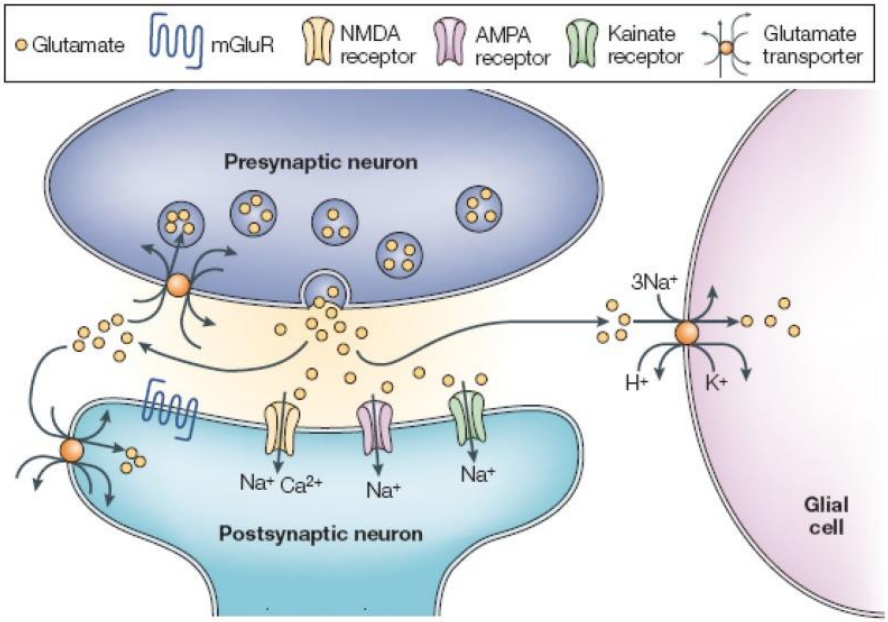


Image from (Bernard, 2012), in the book Jasper's Basic Mechanisms of the Epilepsies

For example, in the case of glutamatergic synapses, the managers of decreasing the glutamate concentrations are the glutamate transporters. Those form part of the EAATs (Excitatory Amino Acid Transporters), transmembrane proteins with a specific domain for this kind of transmitters that carry out the active transport of the same, using sodium electrochemical gradient as a source of energy to realize a change of conformation that enables the amino acid to cross to the intracellular space against the gradient (Magi et al., 2019).

While it is true that these transporters are not enzymes (as enzymes catalyse spontaneous reactions without varying the equilibrium, changing only the intermedium state and decreasing the activation energy, while this kind of transporters perform a non-spontaneous process, therefore altering the equilibrium of the diffusion), their kinetics can be modelled as that of an enzyme, as it is explained below. The derivation of the kinetic equation of an enzyme starts from the stoichiometric equation:

$$E + S \rightleftharpoons ES \rightarrow E + P \qquad (a.2.3)$$

According to this scheme, the enzyme binds to the substrate to form an intermedium product, which will give birth to the products of the reaction plus to the starting enzyme, following some equilibriums. In the case of the simport of glutamate and sodium, the substrates would be extracellular glutamate and sodium, and the products would consist in intracellular glutamate and sodium, which would enter the cytoplasm thanks to a change of conformation of the transporter facilitated due to the energy proportionated by the sodium electrochemical gradient.

Due to the considerable electrochemical gradient of sodium, together with the action of the sodium-potassium pump, that re-establishes it quickly, we could assume that the gradient of sodium stays constant (the ions of sodium that enters the cell are compensated by the ones that the pump expels, being able to consider thus the flux of sodium across the membrane practically null). Therefore, if the concentrations of sodium stay constant, we can neglect this variable in the study of glutamate transport, being the only substrate to consider the extracellular transmitter, and the only product the intracellular glutamate.

Thus, although EAATs are not enzymes, as they don't catalyse any spontaneous reaction, we can approximate their kinetics to that of Michaelis-Menten, as the reaction scheme agrees with the stated in the previous equation, and therefore it could be regarded to follow the steady state assumption from which the Michaelis-Menten equation is constructed. Therefore, given the various equilibrium constants and initial conditions of the system, the decrease of neurotransmitter can be approximated to the classic equation of enzymatic kinetics. Although we have focused on glutamate and the EAATs, the existence of transporters, reuptakers and degrading enzymes is common for all the other transmitters, so this argument could also be applied for other transporters and, of course, for all kinds of degrading enzymes that follow a Michaelis-Menten kinetics.

Let's now talk about the $\gamma\sigma(z_i)$ term. Which is the increment of NT concentration for unit of time as a function of the frequency, $a(z_i)$?

The relation between the release of transmitter and presynaptic spike frequency cannot be linear. This is because the release of the vesicles of the axon terminal button is mediated by calcium ions (Llinás, 1982), and since they have a reversal potential, their intracellular concentration is bounded, and so must be the transmitter release. This dependence on the calcium influx causes the transmitter release to be sigmoid shaped with respect to the presynaptic activation (Katz & Miledi, 1970). That's why we will model the output function in the following way:

$$\sigma(z) = \frac{a(z)^n}{a_{\frac{1}{2}}{}^n + a(z)^n}$$

Which is a sigmoid function with $a_{\frac{1}{2}}$ being the activation at which half of the output transmission is reached and *n* is a parameter to be empirically determined, which we will usually set to 2 or 1. γ stands for the maximum neurotransmitter release, in moles/l*s, that is reached when the presynaptic activation is maximum (in our case, where we haven't imposed any upper bound, when it tends to infinity).

Finally, let's deal with the term referring to diffusion. First Fick law says that the flux is proportional to the gradient of the concentration of the solute. This yields an equation in partial derivatives that doesn't fit the "ordinary" spirit of the rest of the derivatives of the system, so instead of understanding the extracellular space as a continuum (as it is set in the classic formulation of Fick laws), we will reduce it to a discrete space, formed by two regions that have a uniform concentration of solute: an intrasynaptic region and another for extrasynaptic space. This makes sense if the diffusion between these two spaces is much slower than the flux inside of them. In this case, the concentration distribution would be step-shaped, and because this is not a continuous function, and hence it's not

differentiable, we won't be able to compute the gradient at the boundary, substituting it for a difference:

$$J_{int} = -D([T]_i - C_{ext}) \qquad (a.2.5)$$

J_{int} is the flux of transmitter inside the synapse and $C_{ext}$is the concentration of that in the outside. Taking the definition of the flux as the passage of moles of solute per unit of time and area, we can see that $J_{int} = \frac{V}{A}\frac{d[T]_i}{dt}$. We can think that the concentration outside the synapse will be very low, thanks to the action of transporters, glia cells and enzymes, so if we set C_{ext} to be zero and we substitute for the previous expression of the flux, we find the term that describes the change of synaptic concentration caused by diffusion, completing the system we used in section 2.2:

$$\begin{cases} \dot{r}_i = \alpha[T]_i(1 - r_i) - \beta r_i \\ [\dot{T}]_i = -\alpha R_0[T]_i(1 - r_i) + \beta R_0 r_i - \frac{V_{max}[T]_i}{k_M + [T]_i} + \gamma b(\phi_i) - \frac{DA}{V}[T]_i \end{cases} \qquad (a.2.6)$$

Annex 3

Reducing the dimensionality of the model

In order to start analysing (2.2.6), we will nondimensionalize the system. First equation is partially nondimensionalized, as:

$$r_i = \frac{[TR]_i}{R_0} = \frac{moles/litres}{moles/litres} \tag{a.3.1}$$

As it can be seen, units vanish mutually leaving r_i as a dimensionless variable. We will do the same with all the terms in the system, defining the following dimensionless variables, that scale the original transmitter concentration and time to remove their units:

$$s_i = \frac{[T]_i}{R_0}, \tau = \frac{V_{max}}{R_0} t \tag{a.3.2}$$

We will start by plugging s_i in the system, resulting in:

$$\begin{cases} \dot{r}_i = \alpha R_0 s_i (1 - r_i) - \beta r_i \\ \dot{s}_i = -\alpha R_0 s_i (1 - r_i) + \beta r_i - \frac{\frac{V_{max}}{R_0} s_i}{\frac{k_M}{R_0} + s_i} + \frac{\gamma}{R_0}\sigma(z_i) - \frac{DA}{V} s_i \end{cases} \tag{a.3.3}$$

And to erase time units we will use the chain rule to set de derivatives respect dimensionless time, τ :

$$r'_i = \frac{dr_i}{d\tau} = \frac{dt}{d\tau}\dot{r}_i = \frac{R_0}{V_{max}}\dot{r}_i \tag{a.3.4}$$

using (a.3.2) to compute $\frac{dt}{d\tau}$. It's easy to see that we would obtain an equivalent expression for s'_i . Substituting the dimensionless derivatives:

$$\begin{cases} r'_i = \frac{\alpha {R_0}^2}{V_{max}} s_i (1 - r_i) - \frac{\beta}{V_{max}} R_0 r_i \\ s'_i = -\frac{\alpha {R_0}^2}{V_{max}} s_i (1 - r_i) + \frac{\beta R_0}{V_{max}} r_i - \frac{s_i}{\frac{k_M}{R_0} + s_i} + \frac{\gamma}{V_{max}}\sigma(z_i) - \frac{D*A*R_0}{V_{max}*V} s_i \end{cases} \tag{a.3.5}$$

Finally, we will define the dimensionless groups of the system,

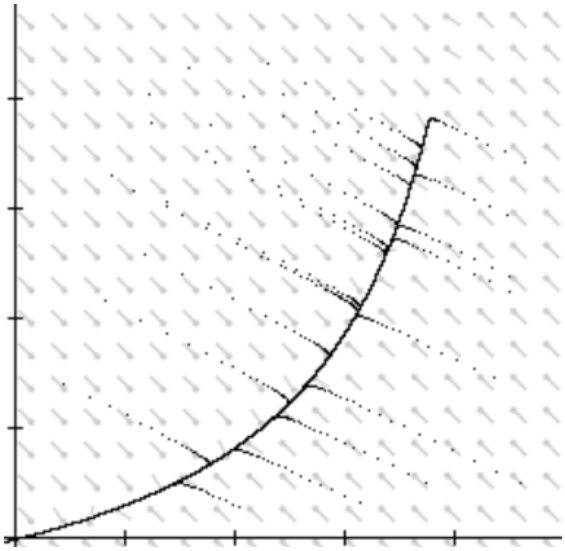
$$a \equiv \frac{\alpha {R_0}^2}{V_{max}}, b \equiv \frac{\beta R_0}{V_{max}}, c \equiv \frac{\gamma}{V_{max}}, {k'}_M \equiv \frac{k_M}{R_0}, \delta \equiv \frac{D*A*R_0}{V_{max}*V} \tag{a.3.6}$$

Thus, nondimensionalization yields:

$$\begin{cases} r'_i = a s_i (1 - r_i) - b r_i \\ s'_i = -a s_i (1 - r_i) + b r_i - \frac{s_i}{{k'}_M + s_i} + c\sigma(z_i) - \delta s_i \end{cases} \tag{a.3.7}$$

In order to reduce the dimensionality of this system we will ask, what would happen in the extreme case where *a* and *b* were much larger than the rest of the groups? If kinetic constants were much higher than V_{max}, how would this system evolve? This is the case we have discussed in section 2.3, showing empirical evidence (Borschel et al., 2015; Sun et al., 2014), and the following task will be to use this piece of information to consolidate the intuition that we have gained about this biochemical system, justifying the presence of two time scales and the existence of a slow manifold.

Figure 2.2.1
Phase flow towards the slow manifold



In the image, some trajectories on the phase space of the problem. Separated dots indicate quick movements, while those that are together represent slow ones. It's easy to intuit the two time scales, as well as the central slow manifold.

Which is the interpretation of the system's phase portrait? Substrate and receptor concentrations tend to equilibrium with great celerity. Once they meet on the manifold, instead of remaining motionless, concentrations travel towards an equilibrium that is given by the input signal: while terms refering to the neurotransmitter transporters as well as diffusion pull concentration down the curve, the term related with presynaptic output pushes it up, compensating both effects in a stable equilibrium, laying inside the slow manifold, that can be classified as a nonlinear node.

We could now ask, what is this curve over which transmitter and bound receptor quantities move? In the case where $a,b>>1$:

$$\frac{b}{a} = \frac{s_i(1-r_i)}{r_i} \qquad (a.3.8)$$

This is nothing but the equation that determines the chemical equilibrium of the dissociation reaction between receptor and transmiter, in dimensionless form. The original expression, recovering the original units, is given by:

$$k_d = \frac{[T][R]}{[TR]} \qquad (a.3.9)$$

Where k_d is the equilibrium constant, known as the dissociation constant, used in biochemistry to know the specificity, in this case, of the receptor. Thus, using the definitions of dimensionless groups, it can be seen that:

$$\frac{b}{a} = \frac{k_d}{R_0} \qquad (a.3.10)$$

The following step will be to replace the parameter *b* in the first equation of (a.3.7) by the previous expression:

$$r'_i = as_i(1 - r_i) - a\frac{k_d}{R_0}r_i = a\left(s_i(1 - r_i) - \frac{k_d}{R_0}r_i\right) = af(r_i, s_i) \qquad (a.3.11)$$

Where the function $f(r_i, s_i)$ is just to shorten the notation. As it can be seen from (a.3.8) and (a.3.10), in the central manifold, $f(r_i, s_i) = 0$. When the output of this function is of order one, *O(1)*, and remembering we are dealing with the case where *a>>1,* it's easy to see that the magnitude of r'_i will be large, $r'_i \approx O(a)$. This confirms what we saw in the phase portrait: trajectories starting outside of the slow manifold tend to it quickly. Now, what happens in the case where trajectories lay almost along the curve? If we set

$$f(r_i, s_i) \approx O\left(\frac{1}{a}\right) \implies r'_i \approx O(1)$$

This means that after a brief state of transition where r'_i has a large magnitude, trajectories aproach the manifold, where they evolve more slowly, but still with considerable speed. Thus, while it is true that when measuring a system like that of our interest we could find configurations placed outside of the slow manifold, we know that they would not last long, as they would tend quickly to the central curve, where they would continue evolving, more slowly once there, towards the equilibrium.

Therefore, we could reduce the dynamics of the whole system to that found on the center manifold (Carr, 2006), from a two dimensional system to a single ODE.

To do so, we will take the second equation of (a.3.7) and we will rewritte it in the following way:

$$s'_i = -r'_i - \frac{s_i}{k'_M + s_i} + c\sigma(z_i) - \delta s_i \qquad (a.3.12)$$

And we will restrict the flow on the center manifold, that we know for (a.3.8) that can be written as:

$$r_i = \frac{s_i}{\frac{k_d}{R_0} + s_i} \qquad (a.3.13)$$

Expressing the curve in this explicit form, we can apply the chain rule to compute the following derivative:

$$r'_i = \frac{dr_i}{ds_i} s'_i,\ \frac{dr_i}{ds_i} = \frac{k_d}{R_0(s_i+\frac{k_d}{R_0})^2} \Rightarrow r'_i = \frac{k_d}{R_0\left(s_i+\frac{k_d}{R_0}\right)^2}(-r'_i - \frac{s_i}{k'_M+s_i} + c\sigma(z_i) - \delta s_i)$$

After some algebra, expressing s_i as a function of r_i using (a.3.8), and isolating r'_i afterwards, we can finally find the ODE that wewere searching, which describes the behaviour of ligand bound receptors over the equilibrium curve where they quickly tend:

$$r'_i = \frac{(1-r_i)^2}{\frac{k_d}{R_0}+(1-r_i)^2}\left(c\sigma(z_i) - \frac{r_i}{\frac{k_M}{k_d}+\left(1-\frac{k_M}{k_d}\right)r_i} - \delta\frac{k_d}{R_0}\frac{r_i}{1-r_i}\right) \qquad (a.3.14)$$

Nevertheless, we could simplify the previous expression by taking into account the order of magnitude of the constants that show up, as we did before. Reembering that we have set $k_d = 3.53 * 10^{-6}M$ (Borschel et al., 2015), and assuming that R_0 is large, as it stands for the concentration of the total receptor in the synapse, we can therefore approximate:

$$\frac{(1-r_i)^2}{\frac{k_d}{R_0}+(1-r_i)^2} \approx 1 \qquad (a.3.15)$$

Which will be acceptable as long as the receptors are not extremely saturated, this is, unless r_i tends to one. This tranforms (a.3.14) into:

$$r'_i = c\sigma(z_i) - \frac{r_i}{\frac{k_M}{k_d}+\left(1-\frac{k_M}{k_d}\right)r_i} - \delta\frac{k_d}{R_0}\frac{r_i}{1-r_i} \qquad (a.3.16)$$

Finally, if we divide by the dimensionless time factor, $\frac{R_{0i}}{V_{max}}$, the above expression yields, using the previous parameters and original time units:

$$\dot{r_i} = \frac{1}{R_0}(\gamma\sigma(z_i) - \frac{V_{max}r_i}{\frac{k_M}{k_d}+\left(1-\frac{k_M}{k_d}\right)r_i} - \frac{k_dDA}{V}\frac{r_i}{1-r_i}) \qquad (a.3.17)$$

Which is the equation stated in section 2.2 and used in the steps of 2.3. With this, we conclude the dimensionality reduction.

Annex 4

Changing variables

In order to perform a transformation such that the new expression for the activation function does not depend on one of the new variables, we have carried out the following steps:

If *x* refers to the real part (this is, the state variable referring to the activation permeability) and *y* to the imaginary part (or, what is the same, the state associated with the inhibitory transmission), we want to find the line defined by *a(x,y)=0*. This is the frontier from where both activation and output functions increase their values monotonically, and we will consider that, under the symmetrical approximation, this is the direction of symmetry, parallel to the rest of contour lines of both functions. From numerical evidence, we infer this is a straight line, so we will consider the explicit form *y=mx+n* and we will substitute it in the activation function, which yields:

$$\frac{(m+1)x+n+1}{\tau \ln\left(\frac{\big(Vr(m+1)-(E^{+}+mE^{-})\big)x-V_{rest}-nE^{-}+Vr(n+1)}{\big(\vartheta(m+1)-(E^{+}+mE^{-})\big)x-V_{rest}-nE^{-}+\vartheta(n+1)}\right)}=0 \qquad (a.4.1)$$

If we find *m,n* such that the denominator of the logarithm vanishes, the logarithm will tend to infinity and thus the function will reduce to zero. As it must cancel regardless of the value of *x*, both $\vartheta(m+1)-(E^{+}+mE^{-})$ and $-V_{rest}-nE^{-}+\vartheta(n+1)$ should vanish separately. This implies a system of equations that, after simplification, takes the following shape:

$$\begin{cases}(\vartheta-E^{-})m+\vartheta-E^{+}=0\\(\vartheta-E^{-})n+\vartheta-V_{rest}=0\end{cases} \qquad (a.4.2)$$

Below, the obtained expressions for the coefficients:

$$m=\frac{E^{+}-\vartheta}{\vartheta-E^{-}} \quad , \quad n=\frac{V_{rest}-\vartheta}{\vartheta-E^{-}} \qquad (a.4.3)$$

And so, the frontier line is given by $Im(z)=\frac{E^{+}-\vartheta}{\vartheta-E^{-}}Re(z)+\frac{V_{rest}-\vartheta}{\vartheta-E^{-}}$. It's director vector is $(\vartheta-E^{-},E^{+}-\vartheta)^{T}$. As we will choose our new basis as a set of orthogonal axis, the component along this line is given by the dot product with any point in the plane *(x,y)*, which recovering complex notation gives: $y'=(\vartheta-E^{-})Re(z)+(E^{+}-\vartheta)Im(z)$. This variable is going to remain cyclic, as the activation function won't depend explicitly on its value, so it will be discarded. With respect to the second axis, we find a vector perpendicular and pointing in the direction of growth, like $(E^{+}-\vartheta,-(\vartheta-E^{-}))^{T}$. Using the same method, we find the only state variable that matters in the translational symmetric case, which is:

$$y=(E^{+}-\vartheta)Re(z)-(\vartheta-E^{-})Im(z) \qquad (a.4.4)$$

Un pont des dels models neurocientífics a les xarxes neuronals recurrents

RESUM EXECUTIU

Joan Gort Vicente

Treball de Final de Grau

Tutor: Josep Maria Losilla Vidal

Curs 2020-2021

19 de Maig de 2021

1. Agents clau i context del treball

Aquest treball es proposa estudiar el vincle entre els modelatges computacionals provinents de la neurociència i aquells en que es basa la perspectiva connexionista, sovint relacionats amb les xarxes neuronals artificials. Per tal de dur a terme aquesta tasca, s'ha adoptat un enfocament interdisciplinari que utilitza i integra diferents perspectives i mètodes, com ara models neurocientífics que parteixen de fenòmens biofísics, fisiològics i de cinètica química; xarxes neuronals artificials, centrant-nos en el seus fonaments i la seva estructura matemàtica; l'estudi de les equacions diferencials, els sistemes dinàmics i l'ús de metodologies computacionals.

És per això que pensem que la present recerca pot interpel·lar a aquelles persones interessades en camps tant diversos com les neurociències, la psicologia cognitiva, l'aprenentatge automàtic, la intel·ligència artificial o altres camps de les matemàtiques aplicades, a les quals els preocupi la qüestió del modelatge realista dels processos neuronals i cognitius.

2. Introducció:

Durant les últimes dècades, l´ús de modelatges matemàtics inspirats en el funcionament de les neurones han suposat una revolució tecnològica, establint les bases de l'aprenentatge automàtic i la intel·ligència artificial, tot permetent la implementació d'algoritmes capaços de reconèixer, classificar i predir fenòmens i objectes tant diversos que, a hores d'ara, és difícil anomenar algun camp de la ciència o la tecnologia on no se'n hagi fet cap mena d'ús.

Aquest treball, però, no tracta de lloar les virtuts de les xarxes neuronals artificials en tant a eines, si no més aviat d'estudiar-les en tant a reflex, com una representació de l'activitat neuronal, és a dir, de mostrar el seu potencial per a explicar fenòmens neurobiològics reals, més enllà de la seva utilitat a l'hora d'implementar algoritmes.

El naixement de l'aprenentatge automàtic i la intel·ligència artificial mai ha estat desvinculat de la psicologia. El connexionisme, perspectiva que entén els processos psicològics com a fenòmens emergents de l'activitat col.lectiva i distribuïda en paral.lel de les neurones, s'ha basat des dels seus orígens en el que avui coneixem com a xarxes neuronals artificials, entenent-les com una via capaç de donar resposta a les preguntes que planteja la ciència cognitiva.

Aquesta perspectiva, però, està basada en uns fonaments que podríem considerar poc ferms, tremolosos, travessats per una qüestió que els posa en dubte i els amenaça de ruïna, la qual es presenta a continuació: si bé es cert que els modelatges connexionistes han permès la implementació de molts fenòmens que podríem considerar de caire psicològic, tals com l'aprenentatge, el reconeixement de patrons o el condicionament als estímuls, què ens permet mantenir-nos ferms a l'hora de defensar que aquests processos es donen al compartir la mateixa base neural que trobem al cervell? En altres paraules, són aquests algoritmes un reflex innocent d'una realitat compartida amb els cervells reals? O són, per contra, un producte interessat de la mà humana, un resultat deliberat que no comparteix les premisses, si no tant sols uns resultats que no es poden assimilar a l'acció d'un mateix mecanisme?

En aquest treball, investigarem la base neurobiològica d'un modelat connexionista concret: les xarxes neuronals recurrents contínues en el temps. Aquestes, si bé s'han emprat per simular molts processos associats als sistemes nerviosos reals, manquen d'una fonamentació teòrica que justifiqui el seu bon funcionament en termes neurobiològics. És així que pretendrem mostrar la relació d'aquestes xarxes tant amb els models neurocientífics computacionals com amb les observacions de fenòmens neurofisiològics.

Amb això, pretenem donar impuls a la seva validació biocomputacional, així com defensar la idea de que algunes xarxes neuronals artificials tenen una relació amb els circuits neuronals reals que va més enllà d'una mera inspiració, cosa que podria fer-les menys artificials del que havíem pensat.

3. Objectius, aportació i rellevància:

El nostre objectiu principal és el de derivar explícitament aquestes xarxes connexionistes partint de models neurocientífics ben estudiats, com els *integrate-and-fire* (que prediuen l'emissió dels potencials d'acció de les neurones partint dels seus corrents iònics) o de la cinètica química de la sinapsi, on s'estudien les reaccions de dissociació que es donen entre receptors i transmissors i com aquestes són influïdes per l'acció dels transportadors, els recaptadors i l'activitat presinàptica. Tal és així que, per una banda, pretenem determinar quines són les premisses i idealitzacions sobre les quals es pot erigir aquest procés de derivació, detallant per a quins grups de neurones i sota quines circumstàncies aquest modelat connexionista pot suposar una bona aproximació; per altra banda, també aspirem a expressar totes les variables i paràmetres d'aquestes xarxes en termes neurobiològics, remarcant la correspondència entre el model i els processos neuronals que aquest pretén representar.

A més, també realitzarem simulacions numèriques per tal de contrastar si el comportament d'aquestes xarxes és compatible amb els enregistraments realitzats en circuïts neuronals biològics, mostrant la rellevància que podrien tenir per tal de simular processos neuronals reals.

Per últim, indagarem en les propietats emergents d'aquestes xarxes, com ara que, de tenir suficients neurones, són capaces d'aproximar qualsevol tipus de comportament observable.

Pensem que aquestes tasques permetran refermar les bases en que es sustenta la perspectiva connexionista, aportant evidències de validació a nivell teòric i computacional, i mostrant que l'ús del seu formalisme pot ser una eina útil tant en el camp de la neurociència com en el de la psicologia.

4. Mètodes

Per assolir aquests objectius, utilitzarem recursos matemàtics provinents de l'estudi dels sistemes dinàmics i les equacions diferencials, com ara l'anàlisi de l'estabilitat d'òrbites, subespais i punts d'equilibri; la linealització d'equacions diferencials ordinàries; l'adimensionalització de les seves components; la reducció del nombre de variables d'estat en presència de varietats centrals; o la disminució de la dimensionalitat del model utilitzant transformacions que n'aprofitin les simetries.

Així doncs, l'ús de totes aquestes tècniques d'anàlisi, juntament amb l'ús d'aproximacions i premisses pertinentment enunciades i justificades, ens permetran realitzar una derivació explícita i exhaustiva del formalisme de les xarxes neuronals recurrents, partint sempre de principis neurobiològics i computacionals.

Pel que respecta als mètodes numèrics, les simulacions es realitzen emprant el mètode d'integració de Runge-Kutta, implementat amb la funció ode45 del programa MATLAB, versió R2021a. Aquestes simulacions s'han realitzat en base als paràmetres i l'estructura de circuits biològics estudiats en mol·luscs i peixos, per tal de comparar enregistraments dels mateixos amb el comportament del model.

5. Resultats

El procés de derivació ha conduït a diversos models de caire connexionista, amb diferents capacitats de ser generalitzats depenent de les consideracions adoptades. En el cas de neurones que tinguin patrons d'activació senzills, on no es considerin retards en el temps de transmissió i se suposi un alt grau de similitud pel que fa als paràmetres cinètics, morfològics i de difusivitat de les sinapsis, s'ha obtingut que la dinàmica d'una xarxa formada per aquestes unitats pot ser aproximada per un sistema d'equacions diferencials que es correspon amb el formalisme de les xarxes neuronals recurrents, mostrant que aquestes poden ser derivades explícitament de principis neurocientífics. Amb això, donem per assolit l'objectiu de derivació, enunciant les condicions de generalització que delimiten l'aplicabilitat del model i traduint el significat de paràmetres i variables en termes neurobiològics.

Per tal de mostrar que les neurones representades es poden aplicar a circuïts reals, s'han realitzat simulacions numèriques d'aquestes xarxes, cablejant-les tot imitant circuïts biològics ben estudiats. Els resultats han mostrat que aquests models poden reproduir-ne el seu comportament. Pensem, però, que cal acumular més evidència en aquesta direcció.

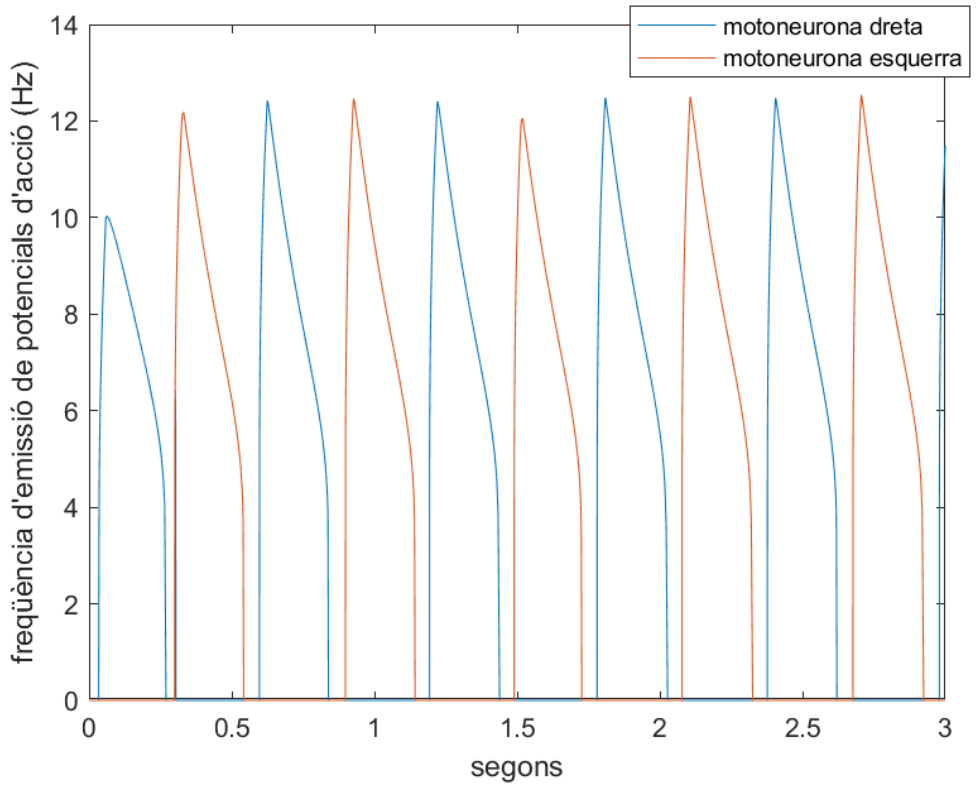


Figura 1: en molts moviments, sobretot locomotors, diverses neurones s'han de sincronitzar per tal de generar un comportament periòdic. Pensi's en l'aleteig d'una libèl·lula: quan el múscul encarregat d'extendre l'ala cap amunt es tensa, el que la flexiona ha de relaxar-se, i viceversa. Quelcom similar passa amb la natació de molts peixos, com els Hiperoartis, els quals posseeixen un circuit senzill que hem pogut simular emprant el model connexionista.

Per últim, també hem analitzat una de les propietats emergents d'aquestes xarxes, la de poder aproximar qualsevol trajectòria de sortida (llegeixi's en termes psicològics com a conducta), i que això implica la correlació de l'activitat de les unitats de la xarxa. Aquesta correlació ha sigut observada al còrtex, on moltes neurones es sincronitzen espontàniament. Això, pensem, reforça la idea de que els model estudiat pot ser útil en l'estudi dels fenòmens neurals.

6. Conclusions

Creiem que el marc connexionista té potencial per esdevenir una perspectiva capaç tant de modelar processos neuronals reals, com d'inferir-ne els processos emergents que d'aquests es deriven, doncs creiem haver mostrat amb aquest treball que alguns models connexionistes no tant sols són interpretables i heurístics, si no que també poden ser bastits partint de principis neurocientífics, de manera sistemàtica, i predir fenòmens de caire fisiològic, tals com el comportament de circuits específics o la sincronització observada en grups de neurones a gran escala.

Per això, creiem que és necessari encoratjar-ne la seva investigació, per tal d'ampliar-ne les evidències de validesa computacional; avançar cap a modelatges més complerts i generals, capaços de representar poblacions més àmplies de neurones; entendre les seves propietats matemàtiques, i construir, a partir d'elles, interpretacions rigoroses i falsejables sobre la fenomenologia dels processos psíquics, partint directament de la base neural que els origina.

Les xarxes neuronals artificials, més reals del que pensàvem?

El formalisme matemàtic de la intel·ligència artificial pot ser derivat directament del comportament real de les neurones, segons l'estudi

La intel·ligència artificial ha sigut, durant dècades, un tòpic de la ciència ficció: des de les novel·les d'Isaac Asimov fins al "*2001, una odissea a l'espai*" de Kubrick, la possibilitat de que les màquines esdevinguin éssers racionals ha inundat l'imaginari col·lectiu.

Paral·lelament a la fantasia i l'entreteniment, **la idea de les màquines intel·ligents ha conduït a moltes persones a plantejar-se la naturalesa mateixa del pensament i el raciocini**, una idea que posa en tela de dubte la vella concepció de l'ànima immaterial com a causa de la intel·ligència i el lliure albir com el seu atribut. Podríem ser tots i totes, al cap i a la fi, una espècie màquines, tant lligades i determinades pels seus propis mecanismes com els objectes que ens envolten ho estan a les lleis de la natura?

Durant les últimes dècades, la idea de que el cervell és una màquina de pensar ha conduït a la creació de **models matemàtics, inspirats en el funcionament de les neurones,capaços d'aprendre, reconèixer, predir i reaccionar als estímuls, coneguts com a xarxes neuronals artificials.** Amb elles, aquella utopia futurista s'ha començat a implementar, i actualment aquests programes són capaços des de de llegir o reconèixer tota mena de patrons fins a inclús simular comportaments empàtics i expressions afectives, entre d'altres.

És així com dintre la psicologia es va plantejar la possibilitat de que aquests models poguessin constituir un nucli d'accés realista als fenòmens cognitius, plantejant al cervell com un sistema de còmput equipat amb propietats similars a les d'aquestes xarxes neuronals artificials.

Aquesta idea, associada a un moviment conegut com a connexionisme, no s'ha pogut deslliurar d'una qüestió encara oberta, que penja sobre seu com una espasa i que amenaça la seva validesa en tant a teoria científica: si és cert que aquests modelatges (les xarxes neuronals) poden replicar alguns comportaments de caire psicològic, que ens fa pensar que això sigui degut a la seva similitud al cervell, i no pas a les finalitats pràctiques per les que van ser creades? És realment el seu funcionament un reflex innocent d'una base neuronal compartida amb els cervells reals?

En la present recerca, **hem aconseguit derivar explícitament el formalisme d'un tipus de modelat connexionista** (les xarxes neuronals recurrents dinàmiques**) des d'una vessant únicament biològica**, partint d'equacions de la neurociència computacional, camp que es dedica a estudiar les neurones reals partint de models matemàtics. Així, mostrem l'existència d'un fil conductor que uneix els models connexionistes als neurocientífics.

A més, també **hem sigut capaços de simular grups de neurones reals** utilitzant aquests models connexionistes, amb resultats satisfactoris. Per últim, hem evidenciat que en el comportament d'aquests models es donen fenòmens observats al còrtex, com la sincronització espontània de l'activitat de les neurones.

Amb això, pretenem impulsar la investigació d'aquests modelatges, mostrar-ne el potencial més enllà de les tecnologies, i alimentar la flama de la teoria connexionista, des d'on la intel·ligència de les màquines podria ser menys artificial del que pensàvem.

Joan Gort Vicente

Universitat Autònoma de Barcelona (UAB),
Facultat de Psicologia