Emergence of universal computations through neural manifold dynamics

Joan Gort Vicente[1]

[1]Universitat Autònoma de Barcelona (UAB), Facultat de Psicologia, Carrer de la Fortuna, s/n, 08193 Bellaterra, Barcelona

Abstract

There is growing evidence that many forms of neural computation may be implemented by low-dimensional dynamics unfolding at the population scale. However, neither the connectivity structure nor the general capabilities of these embedded dynamical processes are currently understood. In this work, the two most common formalisms of firing-rate models are evaluated using tools from analysis, topology and nonlinear dynamics in order to provide plausible explanations for these problems. It is shown that low-rank structured connectivity predicts the formation of invariant and globally attracting manifolds in both formalisms, which generalizes existing theories to different neural models. Regarding the dynamics arising in these manifolds, it is proved they are topologically equivalent across the considered formalisms.

It is also stated that under the low-rank hypothesis, dynamics emerging in neural models are universal. These include input-driven systems, which broadens previous findings. It is then explored how low-dimensional orbits can bear the production of continuous sets of muscular trajectories, the implementation of central pattern generators and the storage of memory states. It is also proved these dynamics can robustly simulate any Turing machine over arbitrary bounded memory strings, virtually endowing rate models with the power of universal computation. In addition, it is shown how the low-rank hypothesis predicts the parsimonious correlation structure observed in cortical activity. Finally, it is discussed how this theory could provide a useful tool from which to study neuropsychological phenomena using mathematical methods.

Keywords: Neural manifolds, firing-rate models, universal approximation theorems, mathematical neuroscience

1. Introduction

During the last decade, the proliferation of state-of-the-art technologies in the neuroscience community, such as microelectrode arrays, brain-computer interfaces or optogenetics, has given researchers the ability to record, control, and manipulate neural data in unprecedented ways (Cunningham & Yu, 2014; Jazayeri & Afraz, 2017; Vyas et al., 2020). With the previous techniques, it has been possible to obtain huge datasets of neural activity.

A common way to extract useful information from them is to apply dimensionality reduction methods (Altan et al., 2021; Cunningham & Yu, 2014) which in turn allow to recognize that, despite the large dimensionality of the recorded data, neural activation patterns seem to be embedded in a much lower dimensional region, often referred to as *neural manifold* (Gallego et al., 2017, 2018; Sadtler et al., 2014). Indeed, it has been the proceeding of many empirical investigations to obtain neuronal *in vivo* recordings from behaving animals and to subsequentially project the data into this lower dimensional region, in order to infer the most prominent features of the underlying neural dynamics (Chaisangmongkon et al., 2017; Chaudhuri et al., 2019; Churchland et al., 2012; Gallego et al., 2018; Mante et al., 2013).

To understand how neural ensembles perform task-specific computations is the inquiry of much contemporary research, and the role neural manifolds could play in this field is of special int(Cunningham & Yu, 2014; Gallego et al., 2017; Jazayeri & Afraz, 2017). In order to tackle this

challenge, many researchers have used recurrent neural networks (RNN) as a model for neural dynamics,

searching for a plausible yet tractable modelling capable to shed light on the neural machinery by which cortical computations are biologically implemented (Sussillo, n.d.; van Gerven & Bohte, 2017). These various types of networks describe the behavior of a neural circuit in terms of continuous population variables, primarily the passive somatic potential (Hopfield, 1984), synaptic conductance (B. Ermentrout, 2008; Gort Vicente, 2021), or an average of the firing rate of action potentials (Ermentrout & Terman, 2010a; Wilson & Cowan, 1973), which is why they are commonly referred to as "firing rate models" in theoretical neuroscience (G. B. Ermentrout & Terman, 2010a). Although they are crude approximations of neuronal dynamics that lack many nuances of the mechanisms of spike generation (Gort, 2021), their behavior exhibits many aspects of biological neuronal systems, such as recurrence, feedback, nonlinearity, and principal component activity (Sussillo et al., 2015), and thus they have appealed many neuroscientists who have found in them a way to simulate, interpret, and make sense of empirical data, showing that RNN setups are capable of replicating many experimental observations (Beiran et al., 2023; Chaisangmongkon et al., 2017; Hennequin et al., 2014; Hoellinger et al., 2013; Mante et al., 2013; Meijer et al., 2015; Sussillo et al., 2015; Sussillo & Barak, 2013; Wärnberg & Kumar, 2019))

As far as neural manifolds are concerned, RNN modelling could be a useful framework from which to understand the intrinsic mechanisms that drive

neuronal trajectories to settle in a lower dimensional embedding (Jazayeri & Ostojic, 2021) .

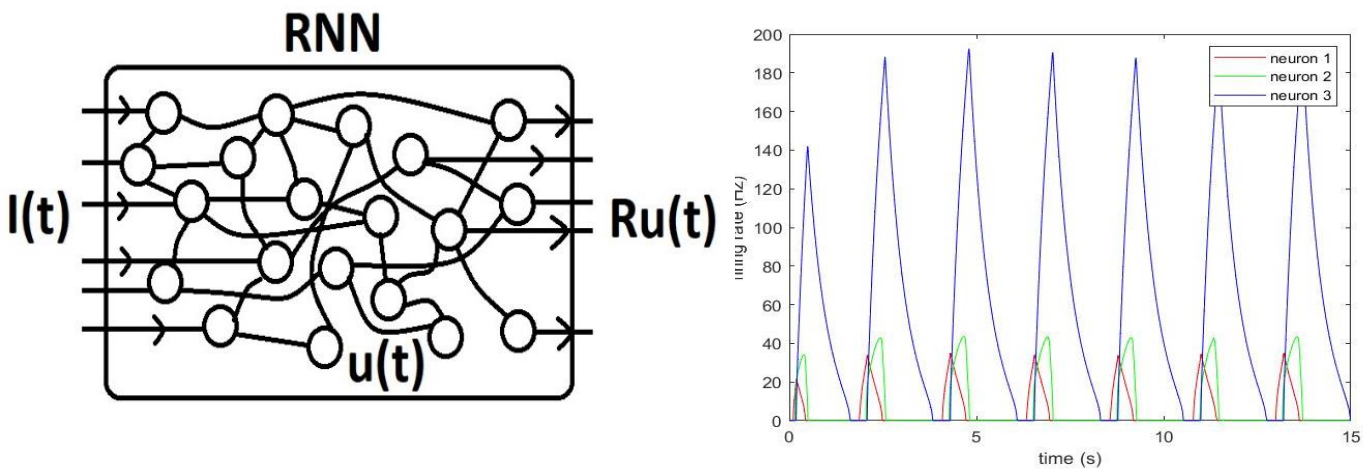

Figure 1. Left: a schematic representation of a recurrent neural network, being I(t) the input vector; u(t) the internal state array, formed by the measures of the activations of the recurrently connected nodes; Ru(t) the readout of the circuit, which extracts the output of the system through the action of a linear operator, R. Right: Plot showing the firing rate of a three-neuron network wired to simulate the periodic bursting of the Tritonia Diomedea central pattern generator. The model shows an intrinsic oscillator such as the one in the original circuit, which equips the organism with an escape swimming response. See (Katz, 2009) *for information about the biological circuit, and [16] for details on the computational modelling.*

In support of this proposal, many recent studies have focused on investigating the emergence of neural manifold phenomena in different types of neural networks, some of which having fully structured low-rank connectivity matrices (Beiran et al., 2021, 2023; Dubreuil et al., 2022; Herbert & Ostojic, 2022), while others possessing both a structured low-rank component and an unstructured random weight matrix (Darshan & Rivkind, 2022; Mastrogiuseppe & Ostojic, 2018; Schuessler et al., 2020).

Although interesting progress has been made in this direction, the approaches used so far rely mainly on computational tools

(Maheswaranathan et al., 2019) and statistical methods, such as mean-field theory (Beiran et al., 2021; Mastrogiuseppe & Ostojic, 2018; Schuessler et al., 2020). In some respects, there exist limitations in establishing sufficient conditions for the existence of neural manifolds and measuring their dimensionality (Altan et al., 2021). For example, the mean-field theory approach works only in the limiting case when the number of neurons tends to infinity (Sompolinsky et al., 1988), and thus cannot fully determine the evolution of a given finite-size neural system. Similarly, computational perspectives cannot make general statements concerning the mechanisms that unfold at the level of a large class of neural models, as they can only provide concrete numerical insights into a given system. Therefore, in these paradigms it is difficult to obtain general results, and one has to be content with interpreting the implementation of specific computational tasks in terms of concrete model configurations (Sussillo & Barak, 2013), without being able to make general statements about the universal computational capabilities with which neural networks are endowed.

Moreover, due to these pitfalls, most of the aforementioned works had to limit the size of the studied neural manifolds to as few as 2 variables (Herbert & Ostojic, 2022; Mastrogiuseppe & Ostojic, 2018; Schuessler et al., 2020), while empirical investigations show that these subspaces can consist of up to 10 independent components (Sadtler et al., 2014).

In the present study, we will consider the two best known formalisms of firing-rate models to study in detail their behavior from the point of view of dynamical systems theory, topology and analysis. We will prove that, for fully structured low-rank connectivity matrices, both models exhibit

invariant and globally attracting manifolds whose intrinsic dimensionality corresponds to the rank of the connectivity matrix.

We will then study the computational performance that results from low-rank models and show that they are universal approximators of input-driven control dynamical systems, extending results that have so far focused mainly on the autonomous case.

This will lead to announce some general results on the mechanisms underlying some important internal processes, such as the storage, control and production of muscle activity; the generation of attractive memory states and central patterns; and the universality to simulate effective procedures using serial symbolic manipulation with finite memory capacity. Finally, it is going to be shown how the low-rank connectivity hypothesis leads to empirically contrasted predictions, and the significance of these results is going to be discussed from a cognitive perspective.

2. Results

In this section, we will present statements concerning the dynamics of the following mathematical models:

$$u'_i(t) + u_i(t) = F_i(\sum_{j=1}^{N} w_{ij} u_j(t) + \sum_{j=1}^{m} \omega_{ij} I_j(t)\ ) \tag{1.1}$$

$$v'_i(t) + v_i(t) = \sum_{j=1}^{N} w_{ij} F_i(v_j(t)) + \sum_{j=1}^{m} \omega_{ij} I_j(t) \tag{1.2}$$

The first one describes the time evolution of a filtered measure of the firing rate (Miller & Fumarola, 2012), while the second does so for the

voltage of the cell when action potentials are removed (Ermentrout & Terman, 2010a). Each F_i models the relationship between the voltage and the spike frequency of each neuron, and they are going to be called transfer functions. The parameters, w_{ij}, ω_{ij} stand for the strengths of the synaptic projections coming from circuit and input neurons, respectively. We will often refer to the first formalism as the u-model or of the Wilson-Cowan type, and the second as the v-model or of the Hopfield type.

For the sake of simplicity, the above equations assume that all the nodes of the net have equal time constants. We will further suppose that for every F_i, there exists a scalar, θ_i, which stands for each neuron's action potential threshold, such that $F_i(x) = \sigma(x - \theta_i)$, being $\sigma: \mathbb{R} \to (0,1)$ a sigmoid function. The following results, however, hold for more general transfer functions as well.

The above formalisms are not the only ones that attempt to model collective neural behavior using physiologically relevant, continuous neural variables. However, they are analytically tractable and can be found among the most widely used models in systems neuroscience. Similar but more biophysically realistic models, also based on continuous variables and neural population measurements, can be found in (Pasquale, Sussillo, Abbott, 2023; Ekeberg, 1993; Thalmeier et al., 2016).

2.1 Conditions sufficing the formation of neural manifolds

In many cases, both computationally and experimentally, it is common to visualize data through dimensionality reduction methods (Hennequin et al., 2014; Sussillo & Barak, 2013). Although their estimates are good enough for many practical reasons, they can sometimes be deceptive

about the geometry or dimensionality of the studied lower-dimensional structures, especially when linear methods are involved (Altan et al., 2021).

For instance, even in the case of a network exhibiting chaotic high-dimensional fluctuations, which is known to be the dominant scenario for large recurrent networks with unstructured random connectivity (Mastrogiuseppe & Ostojic, 2018; Sompolinsky et al., 1988; Stern et al., 2014; Williamson et al., 2016), the solutions could settle on a strange attractor forming a relatively correlated data set that would subsequently yield some eigenvalues close to zero when performing a principal component analysis (PCA). However, this would not entail the dynamics to remain in a low-dimensional subspace, since the attractor itself could be full-dimensional, as numerical evidence suggests in figure 2.

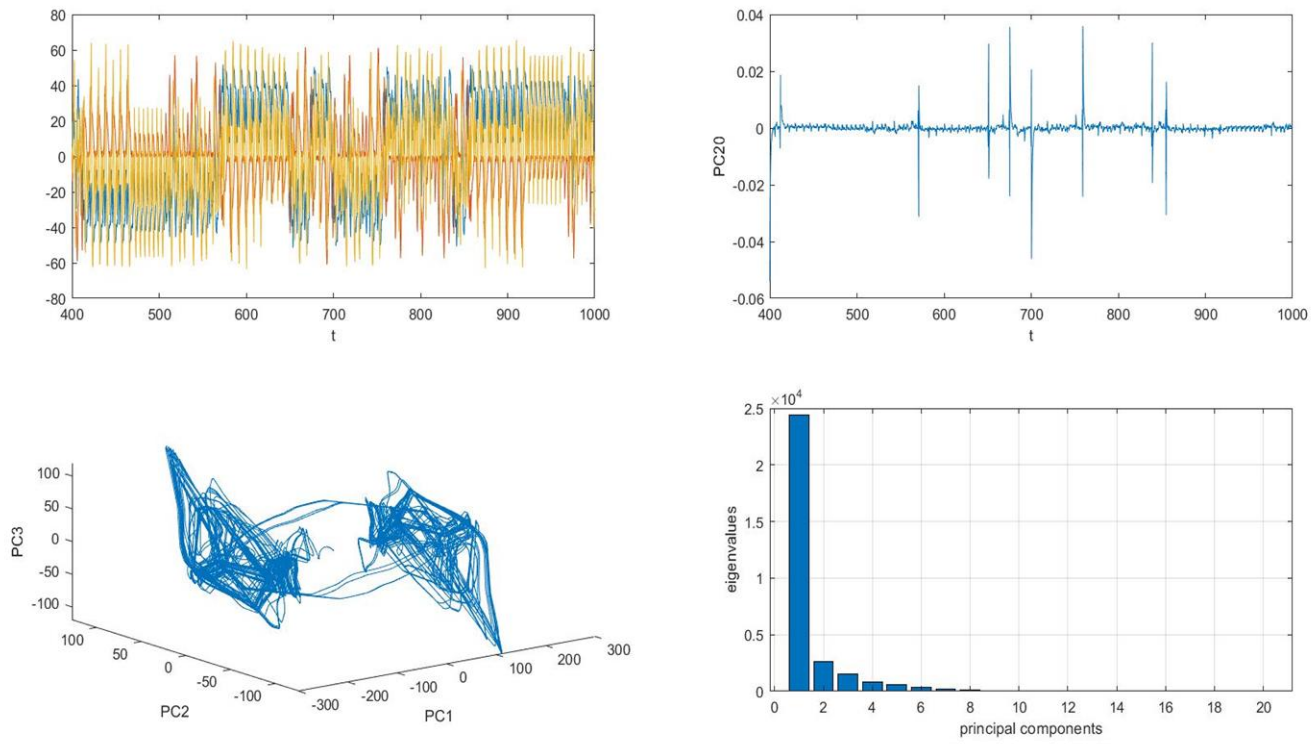

Figure 2 : Simulation of a 20 neuron Hopfield network whose connectivity was randomly wired for a given initial condition. In the first column, it is shown the activation of 3 units displaying chaotic behavior, and below the underlying double-scroll attractor is projected onto the 3 first principal components. In the second column, the evolution of the last principal component is plotted together with a histogram showing the eigenvalues of the covariance matrix. Although the last 10 eigenvectors could be neglected in a lower-dimensional model of the dynamics which could explain almost the whole variance of the system, the last component is nonconstant, and thus the network exhibits full-dimensional chaotic fluctuations.

It is thus that we are interested in seeking conditions accurately predicting the formation of neural manifolds. As previously announced, this will be achieved by restricting weight matrices to be non-invertible. Thereby, we present the following statements:

Theorem 1.1: Let W be the connection matrix of a given u-model with constant input. Then, it possesses an invariant, globally attractive manifold, $\mathcal{M}$, which can be parametrized by a homeomorphism $f: \mathbb{R}^{rank(W)} \to \mathcal{M}$.

The proof of this theorem relies on constructing a Cauchy sequence of functions converging to the map f.

Theorem 1.2: Let W be the connection matrix of a given v-model with constant input. Then, it possesses an invariant, globally attractive manifold whose intrinsic dimension is equal to *rank(W).* Furthermore, this manifold is Euclidean.

For the proofs, see methods, 5.1. Something that can be inferred from the previous theorems is that, although the manifolds emerging from each model could be geometrically different, they are always homoeomorphic. This does not imply that all observed neural manifolds must be topologically Euclidian, as on the contrary, for example, in the head direction circuits of mammals and insects ring attractors have been

observed (Chaudhuri et al., 2019; Kim et al., 2017). Nevertheless, these topologically more complex structures could well be considered as sub-manifolds embedded in the manifolds described in the previous theorems. Section 2.2 is going to study the dynamics evolving in these embeddings, showing that the submanifolds they enclose can be very rich indeed.

From now on, we will reserve the term “neural manifold” for the topological Euclidean manifolds described in the previous theorems, and we will refer to the dynamical systems evolving on these invariant regions as neural manifold dynamics.

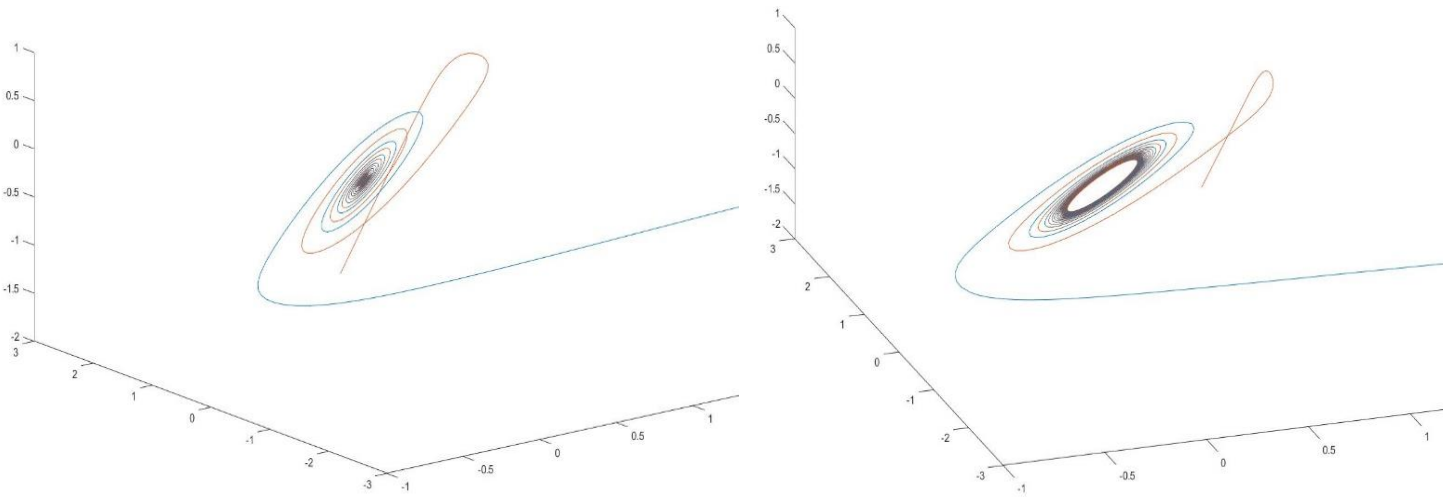

Figure 3: Two trajectories of a 3 neuron Wilson-Cowan system undergoing a Hopf bifurcation on a 2-dimensional manifold. The bifurcation parameter is a single weight whose modification leaves rank(W)=2 unchanged, so that the emergence of a globally attractive surface can be observed in both cases, according with the conclusions of theorem 1.1.

So far, we have studied u-models and v-models separately. Now that explicit conditions predicting the existence of dynamical embedding manifolds have been found, it could be asked what relation exists, if any, between neural manifold dynamics in both kinds of models. The following proposition addresses this problem:

Proposition 1.1: Suppose we have a u-model and a v-model both with equal transfer functions, connectivity arrays and constant inputs. Then, their neural manifold dynamics are topologically conjugate.

We say that the set of solutions of two systems are topologically conjugate if there is an invertible bicontinuous transformation that yield a trajectory of one system whenever applied to a solution of the other one. More technically, it means that there exists a homeomorphism which commutes with the flow map. See methods, section 5.1 for the proper definition, or (Hirsch et al., 2013a).

Intuitively, two dynamical systems being conjugate indicate that they are topologically equivalent, meaning their qualitative behavior remains unchanged, so that things like the number of fixed points, their stability, the presence of cycles or connection orbits as well as the existence of chaotic regions in state space are preserved by moving between the systems. This result provides a useful analytical tool for investigating the behavior of one type of network knowing the behavior of the other.

For example, the well-established fact that randomly wired networks of Hopfield neurons tend to behave chaotically as their dimensionality increases (Sompolinsky et al., 1988) can now be easily proved for Wilson-Cowan networks - we only have to use the fact that the most commonly

used definitions of chaos persist under conjugacy (Lu et al., 2013) and apply proposition 1.1.

Ultimately, this theorem assures that u-models and v-models are mutually consistent, being both equivalent expressions of the same process, from a topological perspective. In Miller & Fumarola, 2012, it was proven that both models were related via an affine map, but it could not be provided a way to find a complete trajectory of the u-model given a solution on the neural manifold of the v-model. In the previous proposition, it was proven that such an invertible connection between the networks exists in all cases, broadening the previous work whenever constant inputs are considered. See methods, section 5.1, for the proof and information about the conjugacy.

2.2 Exploring the universality of low-dimensional computations

The fact that low-rank structured systems have invariant manifolds implies that a dynamical system evolves within these hypersurfaces. In this section, we will explore the nature of these emergent, populationally distributed processes and the extent to which they can endow cortical systems with useful computational mechanisms.

From a dynamical systems perspective, it would be interesting to characterize both geometrically and topologically the various phenomena that take place in these manifolds. With respect to this problem, it is found, surprisingly, that neural manifold dynamics are, in some sense, dense in the space of dynamical systems, which means that for each system of differential equations there is a neural ensemble whose projected dynamics approximates the system up to a given, yet arbitrarily

small, degree of precision. This is stated more accurately in the following theorems, one for each of the models studied:

Theorem 2.1: Let $\Omega \subset \mathbb{R}^n$ be compact, $G \in C^1(\mathbb{R}^n \times \mathbb{R}^m, \mathbb{R}^n)$ and $x(t)$ the solution to the initial value problem

$$x'(t) = G(x(t), I(t)) | x(0) = x_0 \in \Omega$$

being $I: \mathbb{R} \to \mathbb{R}^m$ drawn from an uniformly bounded set of continuous functions. Then, there exists $N \in \mathbb{N}$, a matrix $R \in \mathbb{R}^{n \times N}$ and a u-model with input $I(t)$ and connectivity W such that, for any $\varepsilon > 0$ and any $T > 0$, being $[0, T]$ included in the maximal interval of existence, it is fulfilled that

$$\|x(t) - Ru(t)\| < \varepsilon \; \forall t \in [0, T]$$

given appropriate initial conditions, $u(0)$. Furthermore, $rank(W) = n$.

Theorem 2.2: Let $\Omega \subset \mathbb{R}^n$ be compact, $G \in C^1(\mathbb{R}^n \times \mathbb{R}^m, \mathbb{R}^n)$ and $x(t)$ the solution to the initial value problem

$$x'(t) = G(x(t), J(t)) | x(0) = x_0 \in \Omega$$

being $J: \mathbb{R} \to \mathbb{R}^m$ drawn from an uniformly bounded set of differentiable functions. Then, there exists $N \in \mathbb{N}$, a matrix $R \in \mathbb{R}^{n \times N}$ and a v-model with input $I(t) = J(t) + J'(t)$ and connectivity W such that, for any $\varepsilon > 0$ and any $T > 0$, being $[0, T]$ included in the maximal interval of existence, it is fulfilled that

$$\|x(t) - Rv(t)\| < \varepsilon \; \forall t \in [0, T]$$

given appropriate initial conditions, $v(0)$. Furthermore, $rank(W) = n$.

The proofs of these theorems rely on the universal approximation theorem from artificial neural network theory.

In the foregoing, the matrix R represents the readout which defines the output of the network once it is applied to the vector of neural states. The fact that rank(W)=n predicts, using the results of section 2.1, the emergence of stable invariant manifolds in the systems studied, at least for the case when the emergent dynamics are autonomous. This establishes a link between the theory of neural manifolds and the universal approximation properties of firing rate models, which will be further explored in the following subsections.

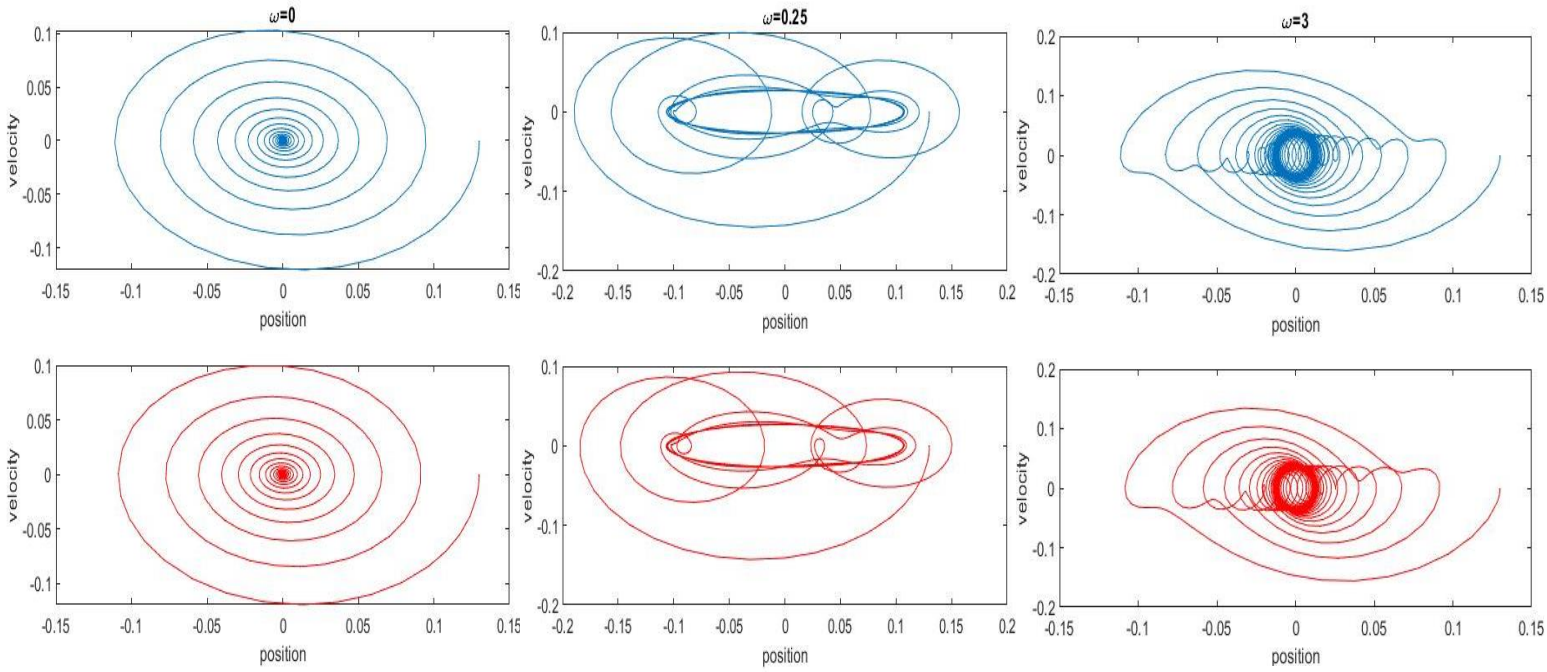

Figure 4: projection of a phase plane trajectory of a periodically driven damped harmonic oscillator. In the first row, in blue, the solutions of the original system; in the second row, in red, an approximation performed with a firing-rate network of the Wilson-Cowan type. Each column represents a different frequency for the driving force.

The results we have just outlined can be summarized saying that, for any driven nonautonomous dynamical system, there exists a neural network of each type whose outputs can approximate its solutions with arbitrary accuracy and during arbitrarily long time periods. This is a property that provides our neural models with a great power of simulation and control.

Biological neural networks must be able to perform a variety of different tasks, as they alone are responsible for a wide range of internal processes, ranging from the generation and control of the muscular patterns that constitute observable behavior to the computations that mediate between the stimulus and responses. For this reason, we will now turn to the behavioral and cognitive consequences that follow from the results obtained so far.

Rigorous propositions with relevant biological interpretations will be presented, concerning the coordination and execution of muscle patterns, the storage of mnemonic and behaviorally relevant attracting states and the universal implementation of symbolic procedures.

2.2.1 Storing compact sets of muscular trajectories

The generation of muscular activity patterns is one of the most studied topics within the dynamic systems approach to large-scale neuroscience (Vyas et al., 2020). This perspective has provided experimental and computational evidence suggesting that motor areas store large sets of muscle patterns whose production is triggered by switching between initial conditions during preparatory activity (Churchland et al., 2012; Hennequin et al., 2014; Michaels et al., 2016; Sussillo et al., 2015). It is known that there is a strong functional correlation between muscle

activity and the primary motor area of the cortex (Gallego et al., 2018; Miri et al., 2017) so that the dynamic properties of the latter can be readily mapped onto the temporal development of the former.

Here we will use the aforementioned theorems to make a statement about the large-scale storage of muscle patterns. Each of these responses is going to be modelled as a trajectory represented by a continuous mapping of the form $r: [0, T] \to \mathbb{R}^n, T > 0$. The set of all these trajectories are going to be equipped with the uniform norm:

$$\|r\|_\infty = \max_{0 \le t \le T} \|r(t)\|.$$

Proposition 2.1: Given any compact set of functions under the uniform norm, $\mathcal{A}$, whose elements are trajectories of the form $r: [0, T] \to \mathbb{R}^n$, $T > 0$, and given any $\varepsilon > 0$, there exists a u-network (respectively a v-network) such that, for every $r \in \mathcal{A}$, its output, given appropriate initial conditions, fulfils that

$$\max_{0 \le t \le T} \|r(t) - Ru(t)\| < \varepsilon$$

(respectively $\max_{0 \le t \le T} \|r(t) - Rv(t)\| < \varepsilon$).

This proposition claims that it is possible to develop a neural system capable of implementing arbitrary compact sets of muscle patterns, the production of which is preceded by pinpointing an appropriate initial condition, as claimed in previously mentioned studies (Churchland et al., 2012; Hennequin et al., 2014; Michaels et al., 2016; Sussillo et al., 2015).

Our model of muscle activity is essentially autonomous and is controlled by the assignment of initial conditions. Therefore, theorems 1.1 and 1.2 predict that the manifold in which the relevant muscle pathways are embedded should be preserved and have the same geometry regardless of the evoked pattern or the elapsed time. This has indeed been reported by experimental studies (Gallego et al., 2018).

Moreover, the fact that autonomous and continuous dynamics evolving along neural manifolds follow the conditions of Picard's existence and uniqueness theorem implies that two different trajectories should never cross each other. This imposes conditions on the intrinsic dimensions of the manifold, which should be provided with additional degrees of freedom in order to disentangle different trajectories, as shown in methods, section 5.3. The existence of such control directions has also been reported earlier in empirical investigations (Hall et al., 2014; Russo et al., 2018).

Taken together, we believe that these results provide some initial evidence supporting our hypothesis.

2.2.2 Attracting sets of neural activity

One limitation of both theorems 2.1 and 2.2 is that, although the approximation of flows is universal and arbitrarily accurate, it can only be so, in general, for a time interval of limited duration. Nevertheless, this can be overcomed by providing stronger hypothesis. This will be achieved by assuring that all the approximated solutions eventually converge to stable limiting trajectories, as the following result suggests:

Proposition 2.2: Suppose we have an autonomous system of differential equations induced by a function $G \in C^1(\mathbb{R}^n, \mathbb{R}^n)$. Let $U \subseteq \mathbb{R}^n$ be open and $\Omega \subset U$ be a compact domain, and suppose further that there exist a family of $l \in \mathbb{N}$ periodic orbits, $\{s_i(t)\}_{i=1}^{l} \subseteq \Omega$, such that, if $x(t)$ is a solution given an initial condition x_0, then

$$\forall x_0 \in U \; \exists t_{x_0} \in \mathbb{R}, 0 \leq i \leq l : \lim_{t \to \infty} \left\| s_i\left(t - t_{x_0}\right) - x(t) \right\| = 0.$$

In this case, given $\varepsilon > 0$ and any trajectory, $x(t)$ with $x(0) \in \Omega$, there exists a u-network (respectively a v-network) whose output, given appropriate initial conditions, fulfils that:

$$\|x(t) - Ru(t)\| < \varepsilon \; \forall t \in \mathbb{R}^+$$

(respectively $\|x(t) - Rv(t)\| < \varepsilon \; \forall t \in \mathbb{R}^+$).

This allows to extend the previous results for all forward time, allowing our neural models to approximate phase trajectories *ad infinitum* whenever the target system is endowed with stable periodic orbits, being these either fixed points or limit cycles.

Attracting fixed points are traditionally used in the field of artificial neural networks as a tool for implementing auto-associative memory devices (Hopfield, 1982) which curiously often resort to low-rank weight matrices, just as in our approach (Amit et al., 1985; G. B. Ermentrout & Terman, 2010b), although there the matrices are further restricted to be symmetric, unlike in our case. This splitting of the neural state space into different basins of attraction has been observed in the study of context-dependent episodic memory in the hippocampus (Wills et al., 2009) .

Attracting fixed states of persistent neural responses also underpin much contemporary research on working memory, where stable bumps of neural activity have been found in the prefrontal cortex of animals prior to retrieval of task-related stored information (Wimmer et al., 2014).

Beyond fixed points, attractive limit cycles are also important objects in neurodynamics, as they are responsible for many robust oscillatory phenomena, such as in locomotion or respiratory patterns. An important class of biological neural circuits that have been shown to sustain intrinsic limit cycle oscillations are the so called central pattern generators (CPGs), which spontaneously support endogenous and repetitive fluctuations without the need of external periodic driving (Yuste et al., 2005).

These different types of neural phenomena all have in common that they are based on locally attracting periodic phase trajectories, as the ones assumed in the hypothesis of proposition 2.2.

Therefore, this result encompasses all these phenomena in the context of the universal approximative capabilities of firing rate networks.

The statement of the previous proposition does not only intend to show the ability of connectionist models to implement attracting memory states or CPGs, which has already been achieved (Hoellinger et al., 2013; Hopfield, 1984; Hopfield & Tank, 1986; Ijspeert, 2001), but also to demonstrate their universality in accomplishing these tasks, to prove that they can behave in this way persistently, and to show that an effective way to implement these functions is to use low-rank connectivity patterns, from which neural systems are expected to uphold neural manifold dynamics, as it has already been observed experimentally.

2.2.3. Universal simulation of finite memory Turing machines

In the study of computable functions, i.e., mappings for which there is an effective procedure to achieve the corresponding outputs, classical computational theory has relied on the Church-Turing thesis, which basically states that any algorithm can be implemented by a Turing machine, and thus relies on this theoretical construct in order to tackle many theoretical problems in computational science.

Intuitively, a Turing machine is composed of an infinite tape consisting of many ordered symbols borrowed from a finite set, the alphabet, where only one of them, the empty symbol, is allowed to occur infinitely often; a set of states, one of which is the initial and some others the final acceptors; and a hypothetical device that can move along the tape so that once it has read a symbol, it is able to rewrite the current cell of the tape and move in either direction to a new cell once it has updated its state, thus implementing a transition function. For a deeper understanding of these concepts we refer the reader to any textbook about computation theory, such as (Lewis & Papadimitriou, 1998).

The tape of the Turing machine store the problem's information given some formal language (Lewis & Papadimitriou, 1998), which the moving lecture device has the job to overwrite so as to terminate the computation. This sequentially manipulated tape, once the alphabet elements are indexed, can be thought as a natural number. We will say that a Turing machine have finite memory if it only performs interpretable computations on some compact set of $\mathbb{N}$, this is, if it is only defined over some set of tapes whose non-blank symbols are pushed to some common initial interval of finite length.

It is the quest of this subsection to explore firing rate networks power for serial symbolic computation. Much of this effort will rely on (Branicky, 1995), where it was proven that continuous time dynamical systems on $\mathbb{R}^3$ have the power of universal computation, meaning they can simulate any arbitrary Turing machine. The way such simulation can be set up is by encoding each Turing machine's configuration (this is, the present state, the whole infinite tape and the current cell location) as an open domain in phase space, and then to define the flow of the system such that it emulates the same transition function as that of the original Turing machine. See (Branicky, 1995) or section 5.3 for more details. Mixing this idea with our previous results give rise to the following statement.

Proposition 2.3: Given any compact set $S \subset \mathbb{N}$ and any computable function $f: S \to \mathbb{N}$, there exists a u-model (respectively a v-model) capable to implement an effective procedure for f. This simulation lies on a 3-dimensional, invariant and globally attractive manifold.

Using the Church-Turing thesis, we can think of any effective procedure as the operations performed by a Turing machine. Following the previous notions of simulation, we can think of the settings of this Turing machine as open regions in the predicted 3-D neural manifold, being those sets of states sewed together by the phase trajectories implementing the transition function. These trajectories would be in charge to transit between different domains of the state space with distinct symbolic interpretations. Interestingly, tunnelling trajectories joining different state space regions have been showed to implement robust computations in recordings of the monkey cortex during decision-making tasks (Chaisangmongkon et al., 2017).

Also backing what is been established experimentally, this proposition shows that it is not necessary to have great dimensionality in order to implement powerful neural computations (Gallego et al., 2017; Mante et al., 2013; Sadtler et al., 2014).

The fact that we have to settle for finite-memory computation is due to the fact that Turing universality is a theoretical notion, since no physical system such as neural circuits or modern computers can store infinite memory. Indeed, theoretical papers proving Turing universality of traditional discrete time neural networks either assume that these networks have unbounded precision (Sontag & Siegelmann, 1991), which is physically implausible, or that they can recruit more neurons, indefinitely, whenever they need more memory (Chung & Siegelmann, 2021), which make the size of these networks effectively infinite.

In the proof, section 5.3, it can be seen that our construction is robust under small perturbations of the neural trajectories, so that it could be, hypothetically, implemented by real noisy neural systems with intrinsic bounded precision.

2.3. Implications of low-rank connectivity

The fully structured low-rank hypothesis that it was imposed on the connectivity as to show both the existence of neural manifolds and the emergence of computationally useful dynamics constitutes a sufficient condition for these purposes, but not a necessary one. Indeed, since invertible matrices form an open and dense subset in $\mathbb{R}^{N\times N}$, one could arbitrarily approximate any non-invertible connectivity array by a full-rank matrix, therefore obtaining some full-rank connectivity matrix with the

same approximation power of the non-invertible one, thus discarding converse results. Nonetheless, the restriction of low-rank wirings could well have positive benefits for neural ensembles over other kinds of setups (Beiran et al., 2023).

For instance, take the number of parameters, say p, needed in order to fully define the connectivity matrix when implementing an n dimensional dynamical system. This measure will obviously depend on the size of the network, this is, the total number of neurons, N. If we consider that connectivity could lie anywhere in the $N \times N$ matrices space, the number of independent parameters needed to define the connections would be given by $p = N^2$, regardless of the dynamical system's dimensionality. However, if we just take the family of n-rank matrices into consideration, as theorems 2.1 and 2.2 suggest, since we just need to define n columns and the rest of them turn out to be linear combinations of the former, the amount of independent parameters needed in order to fully define the connections, this time, works to be given by $p = n(2N - n)$.

Therefore, the number of degrees of freedom of the matrix depend linearly on N, instead of doing so quadratically, thus regularizing the connectivity in a way it can implement difficult tasks relying on a structured parsimonious connectivity of reduced complexity. Indeed, it's been shown that low-rank structured networks generalize their behavior

to novel stimulus better than their full dimensional counterpart (Beiran et al., 2023).

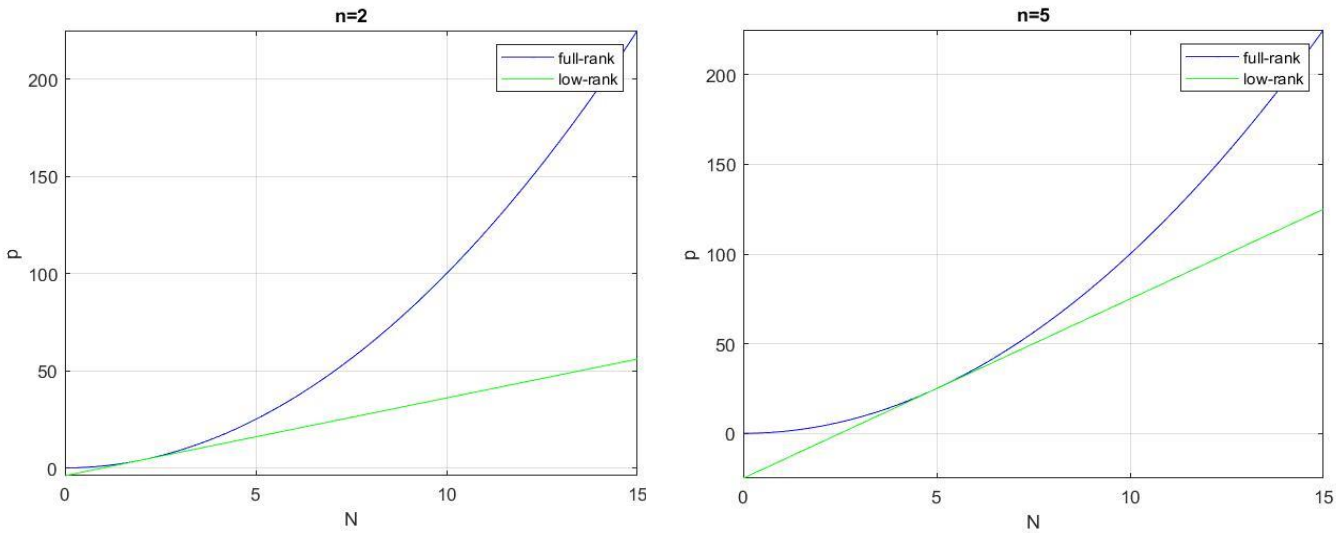

Figure 5: plots showing the number of parameters, p, needed to fully specify a connection matrix, versus the number of neurons, N, of the network. Different dimensionalities for the simulated dynamical systems, n, are supposed. It can be seen how the number of parameters needed in the low-rank setup is of order N, whereas in the full-rank scenario p equals the number of matrix components, N^2. It is further shown in the figure, as it

An appealing characteristic of RNN modelling is that it allows both to explain specific aspects of experimental data and to generate empirically testable predictions (Jazayeri & Ostojic, 2021; Vyas et al., 2020; Wärnberg & Kumar, 2019). These kinds of explanations have been emphasized in this study, where hypothetical mechanisms for the implementation of muscular control, attracting memory states, central pattern generation and serial symbolic processing were developed.

On this matter, interesting predictions can be done based on the low-rank regularization hypothesis. For instance, suppose we would like to know the value of some descriptive statistics of the model, like the correlation between pairs of individual units. This could be done individually for each

solution, subsequently averaging over some set of initial conditions in neural state space (see methods, 5.4). Under the low rank hypothesis, and given the existence and uniqueness theorem, any of these solutions would be completely specified once determined the initial conditions plus the $p = n(2N - n)$ number of parameters needed to fully determine the underlying connectivity. Even if we included individual firing thresholds and membrane time constants to the model, once we average over the relevant set of initial conditions the number of independent parameters needed in order to fully determine the correlation matrix of the model will be of order N, instead of the order N^2 values one would expect in the full-rank structure and in the random connectivity cases.

It turns out this somewhat surprising result have been empirically confirmed in the upper sensory cortex, where it was shown that the N correlations from individual neurons to the summated activity of the whole network suffice to determine a wide fraction of the entire matrix of pairwise correlations between network cells (Okun et al., 2015), showing it is only needed order N independent parameters so as to grasp the whole order N^2 dimensional correlation structure.

Again, here the low-rank hypothesis is a sufficient condition, but not a necessary one. Indeed, an alternative stronger hypothesis would be to suppose neural data remained on a completely flat *n*-dimensional surface, as then a straightforward PCA would show the covariance matrix, and thus also that of the correlations, would have rank equal to *n*, and therefore the number of independent defining parameters would also be of order N, using preceding arguments (methods 5.4).

Nonetheless, restricting the neural manifold to be flat is a rather strong and unrealistic assumption, as even in the v-model the observed manifold

is not Euclidian, since to obtain the firing rates we still must apply a non-linear transfer function, *F* (Ermentrout & Terman, 2010a). Thus, the low-rank hypothesis turns out to be a plausible frame from which to make sense of the seemingly parsimonious correlations observed between individual neurons in the cortex. Indeed, in (Okun et al., 2015) the connectivity matrix was already assumed to be low-rank, since in order to explain how enhanced nonspecific connectivity increases population coupling it was assumed that inhibitory strength relied exclusively on post-synaptic neurons, thus creating repeated columns in the connectivity matrix (see Okun et al., 2015, Supplementary Materials).

The prediction of this empirically confirmed cortical correlation structure supposes a further sign of the predictive power of the theory presented so far.

3. Discussion

Up to now, there have been presented theoretical results, based on rigorous mathematical tools, which have achieved the following goals: to give sufficient conditions for the formation of invariant, stable and low-dimensional manifolds in firing rate models, consisting on restricting the connectivity arrays to low-rank structured matrices; to prove these models can both implement, with arbitrary precision, input-driven control dynamical systems; to go through some of the theoretical consequences of the previous results, presenting propositions concerning the production of muscular activity patterns, the infinite time approximation near attracting states and the virtual power of universal computation, all of this within the predicted neural manifolds; and finally, to show how low-rank

connectivity could provide circuits of a useful regularization strategy, and to present how this hypothesis truly predicts empirically validated results concerning cortical correlations.

When it comes to the network setup, a strong link between universal approximation capabilities and neural manifolds have been revealed, showing that a condition guaranteeing the universal implementation of flows is also a signature of low-dimensional dynamics, namely, the low-rank connectivity structure.

As to the external control of the modelled behavioral and cognitive phenomena, we have used many times the hypothesis that this could often be supplied by the driven modification of the initial conditions along the manifold (Beiran et al., 2023; Remington et al., 2018). Concerning external continuous input driving, theorems 2.1 and 2.2 offer vast possibilities for the simulation of externally controlled on-line computing systems, similarly to the unstructured models of liquid state machines (Maass et al., 2002; Maass & Markram, 2004).

Some of the previous achievements have similar precedents in the literature. For instance, regarding the relation between low-rank connectivity, neural manifold dynamics and universal approximation properties, akin results were recently announced in (Beiran et al., 2021; Dubreuil et al., 2022) as well as in (Gort, 2021), annex 1. In all cases, however, these relations where only explored for the v-model, as we have no evidence concerning the existence of the same results for the u-model in the literature.

Regarding the universal approximation of dynamical systems by neural networks, similar theorems were also announced previously. For instance,

in (Funahashi & Nakamura, 1993), theorem 2.2 was proven for the less general class of autonomous systems; in (Trischler & D'Eleuterio, 2016), driving trajectories where included, although these where internally generated as the model continued to be essentially autonomous; in (Chow & Li, 2000; Li et al., 2005), results allowing external driving where proven for a third class of models with less relevance in the neuroscientific research; in (Doya, 1993), a result similar to our theorem 2.1 was stated, albeit a lesser level of detail.

Comparable evidence also appeared recently in (Beiran et al., 2021), although the proof there did not integrate biases in transfer functions and had thus to rely on restricted tonic inputs. We think the incorporation of variable firing thresholds in the model is not only computationally useful, but also supported by evidence (Fontaine et al., 2014).

A question arises from the universal realization of control dynamical systems by firing rate models: in top-down studies where RNN were optimized to perform animal behavioral tasks (Chaisangmongkon et al., 2017; Hennequin et al., 2014; Mante et al., 2013; Sussillo et al., 2015), was the similarity between the simulations and the data due to a robust physiologic correspondence between computational models and the targeted neural systems, or was it rather that both biological and artificial systems had adopted similar computational strategies in order to perform the same task? After all, universal simulation of dynamical systems applies under a large class of non-realistic transfer functions (in fact, it is possible for any non-polynomial mapping (Leshno & Schocken, 1993)), and thus the striking similarities between biologic and simulated neural systems could well be a matter of shared universal computational capabilities, without

needing a supposedly underlying common functional structure between them.

Concerning this dilemma, bottom-up research shows that firing rate models, whenever endowed with appropriate transfer functions and biophysically based parameters, are able to replicate the frequential behavior of more realistic spiking models under different kinds of input currents (Heiberg et al., 2018; Nordlie et al., 2010; Ostojic & Brunel, 2011), supporting the hypothesis that physiological features of individual neurons uphold the emergent computational phenomena unfolding at the population level.

Considering the limitations of our approach, the low-rank hypothesis is a rather strong assumption, since the provability of a random matrix to be singular is zero. Although we have seen that this hypothesis endows neural ensembles with vast computational capabilities, it would be expected from real circuits to have full rank connection matrices, which could nevertheless lay close enough to the mentioned low-rank arrays to possess the same properties.

In this wider scenario, however, the robustness of the derived neural manifolds to small connectivity perturbations should be considered. It should also be explored the mechanisms by which neural populations could, hypothetically, implement these nearly low-rank connectomes.

It could also be noted the relative simplicity of the models studied, which did not incorporate many biophysically relevant parameters, such as synaptic time constants, in order to make the equations analytically tractable. Regarding synaptic homogeneity, however, it could be assumed that, since the approach adopted so far has effectively ignored learning

mechanisms, the absence of different time constants could be a consequence of excluding synaptic potentiation from the modelled phenomena.

We think the results presented so far constitute a mathematical theory with a great explanatory power, which is consistent with many recent discoveries in systems neuroscience (Gallego et al., 2017; Sussillo, 2014; Vyas et al., 2020).

From this perspective, many internal processes controlling animal behavior, like the coordination of movements, the robust generation of central patterns or the computations mediating between stimulus and responses, are all understood as dynamical systems, externally controlled by either switching between initial conditions or by directly driving their vector fields. These continuous dynamics can be simulated by large scale neural ensembles, recurrently wired through low-rank connectivity patterns. Those mutually interacting units are capable to implement arbitrary flows through nonlinear feedback, being these dynamics parallelly and massively distributed across neural internal states, unfolding at the population level. Thus, neural correlates could not be understandable at the scale of the single neuron, whose individual activity reflects a raw mixture of the implemented dynamics, but only through the network-wide spread patterns of collective activity. These emergent motifs would, in turn, be embedded in a lower-dimensional manifold, a byproduct of the low-rank wiring, which with good reason could be considered the geometric scaffold of the implemented computations.

In particular, this paradigm possesses a great power to generate cognitive and psychologically relevant interpretations from modelled neural data. Within it, the study of the physiological basis of memory and cognition can

be elegantly interpreted from a geometrical lens. For instance, concerning the study of procedural memory, the short-term storage of task-relevant information, called working memory, could be understood as temporary persistent states of the manifold dynamics, while the long-term retention of the rules governing the adaptative performance of a given task, also known as reference memory, could be inferred from the dynamics defining the input-driven transitions in neural state space (Domjan, 2010).

In the same direction, and considering proposition 2.3, the dynamics governing the patterns evolving within the neural manifold could be understood as a syntactically structured set of rules, defining computations in terms of serial manipulation of symbols, being its implementation grounded in the network wiring. This emergentist proposal is backed by the connectionist approach to cognitive science, where neurons are understood to be sub-symbolic processors from which parallelly distributed processes can be inferred. This way, neural manifolds could be considered as a point of access to these spread, somewhat blurred computations, a place from where to elucidate the mechanisms by which conceptually intelligible phenomena emerge from raw sub-symbolic neural correlates (Berkeley, 2019; Smolensky, 1988; Thomas & McClelland, 2012).

In summary, the present research has tackled the study of various neural phenomena using analytical methods, proving, on the one hand, under which conditions attracting and invariant manifolds can be found in neural systems and, on the other hand, how these conditions are linked to the emergence of universal computational capabilities. In doing so, it has been possible to generalize previous results to a wider range of firing rate models, as well as to extend such developments to more general

frameworks of driven non-autonomous dynamics. Consequently, statements concerning the topological equivalence of various firing rate models have been achieved and, in addition, falsifiable predictions have been generated and subsequently contrasted, satisfactorily, with existing empirical data. All the above represents a contribution to strengthen our analytic understanding of the equations ruling firing rate networks, as well as to build a theoretical framework from which geometric and computational aspects of neural dynamics are inseparably understood.

4. Bibliography

Altan, E., Solla, S. A., Miller, L. E., & Perreault, E. J. (2021). Estimating the dimensionality of the manifold underlying multi-electrode neural recordings. *PLoS Computational Biology*, *17*(11), 1–23. https://doi.org/10.1371/journal.pcbi.1008591

Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1985). Storing Infinite Number of Patterns in a Spin-Glass Model of Neural Networks. *Physical Review Letters*, *55*(14), 1530–1533.

Azagra, D., Le Gruyer, E., & Mudarra, C. (2021). Kirszbraun's Theorem via an Explicit Formula. *Canadian Mathematical Bulletin*, *64*(1), 142–153. https://doi.org/10.4153/s0008439520000314

Barak, O., Sussillo, D., Romo, R., Tsodyks, M., & Abbott, L. F. (2013). From Fixed Points to Chaos: Three Models of Delayed Discrimination. *Prog Neurobiol.*, *103*, 214–222. https://doi.org/10.1016/j.pneurobio.2013.02.002. From

Beiran, M., Dubreuil, A., Valente, A., Mastrogiuseppe, F., & Ostojic, S. (2021). Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Computation*, *33*(6), 1572–1615. https://doi.org/10.1162/neco_a_01381

Beiran, M., Meirhaeghe, N., Sohn, H., Jazayeri, M., & Ostojic, S. (2023). Parametric Control of Flexible Timing Through Low-Dimensional Neural Manifolds. *Neuron*, 1–47.

https://doi.org/https://doi.org/10.1016/j.neuron.2022.12.016

Berkeley, I. S. N. (2019). The Curious Case of Connectionism. *Open Philosophy*, *2*(1), 190–205. https://doi.org/10.1515/opphil-2019-0018

Branicky, M. S. (1995). Universal computation and other capabilities of continuous and hybrid systems. *Theoretical Computer Science*, *138*(1), 67–100.

Brian DePasquale, David Sussillo, L.F. Abbott, M. M. C. (2023). The centrality of population-level factors to network computation is demonstrated by a versatile approach for training spiking networks. *Neuron*,https://doi.org/https://doi.org/10.1016/j.neuron.2022.12.007.

Chaisangmongkon, W., Swaminathan, S. K., Freedman, D. J., & Wang, X. J. (2017). Computing by Robust Transience: How the Fronto-Parietal Network Performs Sequential, Category-Based Decisions. *Neuron*, *93*(6), 1504-1517.e4. https://doi.org/10.1016/j.neuron.2017.03.002

Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A., & Fiete, I. (2019). The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience*, *22*(9), 1512–1520. https://doi.org/10.1038/s41593-019-0460-x

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

Chung, S., & Siegelmann, H. (2021). Turing Completeness of Bounded-Precision Recurrent Neural Networks. *Advances in Neural Information Processing Systems*, *34*(NeurIPS), 28431–28441.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S., & Shenoy, K. (2012). *Neural population dynamics during reaching*. *487*(7405), 51–56. https://doi.org/10.1038/nature11129.Neural

Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, *17*(11), 1500–1509. https://doi.org/doi:10.1038/nn.3776.

Cybenko, G. V. (1989). Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals and Systems*, *2*, 303–314. https://doi.org/https://doi.org/10.1007/BF02551274

Darshan, R., & Rivkind, A. (2022). Learning to represent continuous variables in heterogeneous neural networks. *Cell Reports*, *39*(1). https://doi.org/10.1016/j.celrep.2022.110612

Domjan, M. (2010). Compared cognition I: memory mechanisms. In *The principles of learning and behavior* (pp. 375–417). Wadsworth, Cengage Learning.

Doya, K. (1993). Universality of Fully-Connected Recurrent Neural Networks. *IEEE Transactions on Neural Networks*, *1*, 1–6.

Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F., & Ostojic, S. (2022). The role of population structure in computations through neural dynamics. *Nature Neuroscience*, *25*(6), 783–794. https://doi.org/10.1038/s41593-022-01088-4

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

Ermentrout, B. (2008). *Reduction of Conductance Based Models with Slow Synapses to Neural Nets*. 1–22.

Ermentrout, G. B., & Terman, D. H. (2010a). Firing Rate Models. In *Mathematical Foundations of Neuroscience* (pp. 331–367). Springer.

Ermentrout, G. B., & Terman, D. H. (2010b). Spatially Distributed Networks. In S. S. Antman, J. E. Marsden, L. Sirovich, & S. Wiggins (Eds.), *Mathematical Foundations of Neuroscience* (pp. 369–405). Springer.

Fontaine, B., Peña, J. L., & Brette, R. (2014). Spike-Threshold Adaptation Predicted by Membrane Potential Dynamics In Vivo. *PLoS Computational Biology*, *10*(4), 1–11. https://doi.org/10.1371/journal.pcbi.1003560

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

Gort Vicente, J. (2021). A bridge from neuroscientific models to recurrent neural networks Derivation of continuous-time connectionist models from neuroscience computational principles Joan Gort Vicente. *Dipòsit Digital de Documents de La UAB*. https://ddd.uab.cat/record/255161

Hall, T. M., deCarvalho, F., & Jackson, A. (2014). A Common Structure Underlies Low-Frequency Cortical Dynamics in Movement, Sleep, and Sedation. *Neuron*, *83*(5), 1185–1199. https://doi.org/10.1016/j.neuron.2014.07.022

Heiberg, T., Kriener, B., Tetzlaff, T., Einevoll, G. T., & Plesser, H. E. (2018). *Firing-rate models for neurons with a broad repertoire of spiking behaviors*. 103–132.

Hennequin, G., Vogels, T. P., & Gerstner, W. (2014). Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron*, *82*(6), 1394–1406. https://doi.org/10.1016/j.neuron.2014.04.045

Herbert, E., & Ostojic, S. (2022). The impact of sparsity in low-rank recurrent neural networks. *PLoS Computational Biology*, *18*(8), 1–22. https://doi.org/10.1371/journal.pcbi.1010426

Hirsch, M. W., Smale, S., & Devaney, R. L. (2013a). *Differential Equations, Dynamical Systems and an introduction to Chaos*. Academic Press.

Hirsch, M. W., Smale, S., & Devaney, R. L. (2013b). Existence and Uniqueness Revisited. In *Differential Equations, Dynamical Systems, and an Introduction to Chaos* (pp. 402–403). Academic Press.

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. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(8), 2554–2558. https://doi.org/10.1073/pnas.79.8.2554

Hopfield, J. J. (1984). Neurons with graded response have collective have collective computational properties like those of two-state neurons. *Biophysics*, *81*, 3088–3092.

Hopfield, J. 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. 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

Jazayeri, M., & Afraz, A. (2017). Navigating the Neural Space in Search of the Neural Code. *Neuron*, *93*(5), 1003–1014. https://doi.org/10.1016/j.neuron.2017.02.019

Jazayeri, M., & Ostojic, S. (2021). Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Current Opinion in N*, *70*, 113–120.

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

Kim, S. S., Rouault, H., Druckmann, S., & Jayaraman, V. (2017). Ring attractor dynamics in the Drosophila central brain. *Science*, *356*(6340), 849–853. https://doi.org/10.1126/science.aal4835

Leshno, M., & Schocken, S. (1993). Multilayer Feedforward Networks with Non-Polynomial Activation Functions Can Approximate Any Function. *Neural Networks*, *21*, 1–16. https://archive.nyu.edu/bitstream/2451/14384/1/IS-91-26.pdf

Lewis, H. R., & Papadimitriou, C. H. (1998). Elements of the Theory of Computation. In *Prentice-Hall, Inc.* https://doi.org/10.1145/300307.1040360

Li, X., Ho, J. K. L., & Chow, T. W. S. (2005). Approximation of Dynamical Time-Variant Systems by Continuous-Time Recurrent Neural Networks. *IEEE Transactions on Circuits and Systems*, *52*(10), 656–660.

Lu, T., Zhu, P., & Wu, X. (2013). The retentivity of chaos under topological conjugation. *Mathematical Problems in Engineering*, *2013*(2), 4–7. https://doi.org/10.1155/2013/817831

Maass, W., & Markram, H. (2004). On the computational power of circuits of spiking neurons. *Journal of Computer and System Sciences*, *69*(4), 593–616. https://doi.org/10.1016/j.jcss.2004.04.001

Maass, W., Natschläger, T., & Markram, H. (2002). *Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations*. *14*(11), 2531–2560. https://doi.org/10.1162/089976602760407955

Maheswaranathan, N., Williams, A. H., Golub, M. D., Ganguli, S., & Sussillo, D. (2019). Universality and individuality in neural dynamics across large populations of recurrent networks. *Advances in Neural Information Processing Systems*, *32*(NeurIPS).

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742

Mastrogiuseppe, F., & Ostojic, S. (2018). Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks. *Neuron*, *99*(3), 609-623.e29. https://doi.org/10.1016/j.neuron.2018.07.003

Meijer, H. G. E., Eissa, T. L., Kiewiet, B., Neuman, J. F., Schevon, C. A., Emerson, R. G., Goodman, R. R., McKhann, G. M., Marcuccilli, C. J., Tryba, A. K., Cowan, J. D., van Gils, S. A., & van Drongelen, W. (2015). Modeling focal epileptic activity in the Wilson-Cowan model with depolarization block. *Journal of Mathematical Neuroscience*, *5*(1). https://doi.org/10.1186/s13408-015-0019-4

Michaels, J. A., Dann, B., & Scherberger, H. (2016). Neural Population Dynamics during Reaching Are Better Explained by a Dynamical System than Representational Tuning. *PLoS Computational Biology*, *12*(11), 1–22. https://doi.org/10.1371/journal.pcbi.1005175

Miller, K. D., & Fumarola, F. (2012). *Mathematical Equivalence of Two Common Forms of Firing-Rate Models of Neural Networks*. *24*(1), 25–31. https://doi.org/doi:10.1162/NECO_a_00221. Mathematical

Miri, A., Warriner, C. L., Seely, J. S., Elsayed, G. F., Cunningham, J. P., Churchland, M. M., & Jessell, T. M. (2017). Behaviorally Selective Engagement of Short-Latency Effector Pathways by Motor Cortex. *Neuron*, *95*(3), 683-696.e11.

https://doi.org/10.1016/j.neuron.2017.06.042

Munkres, J. R. (2002). *Topology, 2nd edition*. Pearson Education, S.A.

Nordlie, E., Tetzlaff, T., & Einevoll, G. T. (2010). Rate dynamics of leaky integrate-and-fire neurons with strong synapses. *Frontiers in Computational Neuroscience*, *4*(December), 1–13. https://doi.org/10.3389/fncom.2010.00149

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

Ortega Aramburu, J. M. (2002). *Introducció a l'Anàlisi Matemàtica*. Universitat Autònoma de Barcelona, Servei de Publicacions.

Ostojic, S., & Brunel, N. (2011). From Spiking Neuron Models to Linear-Nonlinear Models. *PLoS Computational Biology*, *7*(1). https://doi.org/10.1371/journal.pcbi.1001056

Remington, E. D., Narain, D., Hosseini, E. A., & Jazayeri, M. (2018). Flexible Sensorimotor Computations through Rapid Reconfiguration of Cortical Dynamics. *Neuron*, *98*(5), 1005-1019.e5. https://doi.org/10.1016/j.neuron.2018.05.020

Rudin, W. (1976). *Principles of Mathematical Analysis* (A. A. Arthur & S. Levine Langman (eds.)). McGraw-Hill, Inc.

Russo, A. A., Bittner, S. R., Perkins, S. M., Seely, J. S., London, B. M., Lara, A. H., Miri, A., Marshall, N. J., Kohn, A., Jessell, T. M., Abbott, L. F., Cunningham, J. P., & Churchland, M. M. (2018). Motor Cortex Embeds Muscle-like Commands in an Untangled Population Response. *Neuron*, *97*(4), 953-966.e8. https://doi.org/10.1016/j.neuron.2018.01.004

Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., Yu, B. M., & Batista, A. P. (2014). Neural constraints on learning. *Nature*, *512*(7515), 423–426. https://doi.org/10.1038/nature13665

Schuessler, F., Dubreuil, A., Mastrogiuseppe, F., Ostojic, S., & Barak, O. (2020). Dynamics of random recurrent networks with correlated low-rank structure. *Physical Review Research*, *2*(1), 13111. https://doi.org/10.1103/PhysRevResearch.2.013111

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

Sompolinsky, H., Crisanti, A., & Sommers, H. J. (1988). Chaos in random neural networks. *Physical Review Letters*, *61*(3), 259–262. https://doi.org/10.1103/PhysRevLett.61.259

Sontag, E. D., & Siegelmann, H. T. (1991). *Turing Computability with Neural Nets*. *4*(6), 77–80.

Stern, M., Sompolinsky, H., & Abbott, L. F. (2014). Dynamics of random neural networks with bistable units. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *90*(6), 1–19. https://doi.org/10.1103/PhysRevE.90.062710

Sussillo, D. (n.d.). ScienceDirect Neural circuits as computational dynamical systems. *Current Opinion in Neurobiology*, *25*, 156–163. https://doi.org/10.1016/j.conb.2014.01.008

Sussillo, D., & Barak, O. (2013). Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Computation*, *25*(3), 626–649. https://doi.org/10.1162/NECO_a_00409

Sussillo, D., Churchland, M. M., Kaufman, M. T., & Shenoy, K. V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nature Neuroscience*, *18*(7), 1025–1033. https://doi.org/doi:10.1038/nn.4042.

Thalmeier, D., Uhlmann, M., Kappen, H. J., & Memmesheimer, R. M. (2016). Learning Universal Computations with Spikes. *PLoS Computational Biology*, *12*(6), 1–29. https://doi.org/10.1371/journal.pcbi.1004895

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

Tinchcombe, M., Hornik, K., & White, H. (1989). *Multilayer Feedforward Networks are Universal Approximators*. *2*, 359–366.

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

van Gerven, M., & Bohte, S. (2017). Editorial: Artificial Neural Networks as Models of Neural Information Processing. *Frontiers in Computational Neuroscience*, *11*, 114. https://doi.org/10.3389/fncom.2017.00114

Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation through Neural Population Dynamics. *Annual Review of Neuroscience*, *43*, 249–275. https://doi.org/10.1146/annurev-neuro-092619-094115

Wärnberg, E., & Kumar, A. (2019). Perturbing low dimensional activity manifolds in spiking neuronal networks. *PLOS Computational Biology*, *15*(5), 1–23. https://doi.org/10.1371/journal.pcbi.1007074

Williamson, R. C., Cowley, B. R., Litwin-Kumar, A., Doiron, B., Kohn, A., Smith, M. A., & Yu, B. M. (2016). Scaling Properties of Dimensionality Reduction for Neural Populations and Network Models. *PLoS Computational Biology*, *12*(12), 1–27. https://doi.org/10.1371/journal.pcbi.1005141

Wills, T. J., Lever, C., Cacucci, F., Burgess, N., & Keefe, J. O. (2009). *Attractor Dynamics in the Hippocampal Representation of the Local Environment*. *308*(5723), 873–876. https://doi.org/10.1126/science.1108905.Attractor

Wilson, H. R., & Cowan, J. D. (1973). *A Mathematical Theory of the Functional Dynamics of Cortical and Thalamic Nervous Tissue*. *80*.

Wimmer, K., Nykamp, D. Q., Constantinidis, C., & Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, *17*(3), 431–439. https://doi.org/10.1038/nn.3645

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

5. Methods

Throughout this section, we will often consider equations (1.1) and (1.2) in vector notation, defining the vector transfer function $F: \mathbb{R}^N \to \mathbb{R}^N$ as $F(x) = (F_1(x_1), \dots, F_N(x_N))^T$. Then, if W, ω are the connectivity matrices, the equations for the u-model and the v-model are, respectively,

$$u'(t) + u(t) = F(Wu(t) + \omega I(t)) \tag{5.1}$$

$$v'(t) + v(t) = WF\big(v(t)\big) + \omega I(t) \tag{5.2}$$

In the following sections, all the results presented in the text are going to be proved using diverse mathematical techniques from dynamical systems theory, topology and real analysis. Before every proof the most crucial results concerning these mathematical methods are going to be presented. However, some basic knowledge about the previous subjects will be assumed, and so the reader will be redirected to introductory textbooks on these topics whenever it is necessary.

5.1. Proof of theorem 1.1, theorem 1.2 and proposition 1.1.

Throughout this paper, the notions of invariant and attracting sets have been used many times. We now formalize these concepts:

Definition 5.1.1: Let $x(t)$ be some solution of a continuous, finite dimensional dynamical system for a given initial condition, x_0. We say a set S is invariant if $x(t) \in S \; \forall x_0 \in S, t \in \mathbb{R}$; we say a set S is positively invariant if $x(t) \in S \; \forall x_0 \in S, t \in \mathbb{R}^+$.

Definition 5.1.2: Let $p \in \mathbb{R}^n, S \subset \mathbb{R}^n$; Let $d(p,S) = \inf_{q \in S} \|p - q\|$; Let $x(t)$ be some solution of a continuous, finite dimensional dynamical system for a given initial condition, x_0. We say S is globally attracting if $\lim_{t \to +\infty} d(x(t), S) = 0 \; \forall x_0 \in \mathbb{R}^n$.

The notion of manifold has also appeared ubiquitously in the text. We present here its rigorous definition:

Definition 5.1.3: A n-dimensional manifold is any Hausdorff, second-countable topological space, X, for which for every point $x \in X$ there exists a neighbourhood homeomorphic to some open set of the Euclidian space, $\mathbb{R}^n$.

Here it is not intended to make a comprehensive introduction to manifold topology, so we refer the reader to any elementary textbook on topology covering manifolds, like (Munkres, 2002), for a thorough exposition of these and other topics. Nonetheless, we present here the notion of homeomorphism, as it is going to be fundamental.

Definition 5.1.4: Let X, Y be topological spaces. A map $h: X \to Y$ is a *homeomorphism* whenever it is continuous, invertible and have a continuous inverse.

Homeomorphic sets can be considered as equivalent in a topological sense, since each neighbourhood in one space will be mapped to a unique neighborhood of the other, and vice versa. It is thus that homeomorphisms are regarded as topological isomorphisms.

To prove the mentioned results, we present the following lemmas:

Lemma 5.1.1: Suppose $W \in \mathbb{R}^{N\times N}$. Then, $rank(W) = n$ if and only if there exist some rank n matrices $B \in \mathbb{R}^{N\times n}$, $A \in \mathbb{R}^{n\times N}$ such that $W = BA$

Proof: If $W = BA$, the columns of W are all linear combinations of the n linearly independent columns of B. Since $rank(A) = n$, the columns of W span the same column space of B, proving the forward result; If $rank(W) = n$, suppose $\{b_{1,} \dots, b_n\}$ is a basis for the image of W. Then, there exist unique scalars a_{ij} such that $W = (\sum_{i=1}^{n} a_{i1}b_i, \dots, \sum_{i=1}^{n} a_{iN}b_i) = BA$, where the columns of B are given by b_i and the elements of A by a_{ij}. By the fundamental theorem of linear algebra, $rank(A) = rank(W) = n$. This concludes the proof ■

Lemma 5.1.2: Suppose $\{f_n\}$ is a sequence of continuous functions of the form $f_n: \mathbb{R}^n \to \mathbb{R}$, and suppose it is Cauchy under the uniform norm, meaning that

$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}$ such that $|f_n(x) - f_m(x)| < \varepsilon \; \forall n, m \geq n_0, \; x \in \mathbb{R}^n$

Then, there exists a continuous function f fulfilling that $\{f_n\} \to f$ uniformly, this is,

$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \varepsilon \; \forall n \geq n_0, x \in \mathbb{R}^n$.

For a proof of this important result, see any textbook on real analysis covering successions and series of functions, such as (Ortega Aramburu, 2002; Rudin, 1976). We now turn to prove the theorems of section 1.1.

Proof of theorem 1.1: If $rank(W) = N$, it is verified trivially that $\mathcal{M} = \mathbb{R}^N$. If, for the contrary, $rank(W) = n < N$, we know from lemma 5.1 that $W = BA$, where both the columns of B and the rows of A are linearly

independent vectors. We are going to perform the change of variables $\binom{x(t)}{z(t)} = \binom{A}{M}u(t)$, where the matrix M has the $N-n$ basis vectors of the kernel of W as its rows.

In this new basis, the system can be expressed as:

$$\begin{cases} x'(t) = -x(t) + AF(Bx(t) + \omega I)\ (*) \\ z'(t) = -z(t) + MF(Bx(t) + \omega I) \end{cases}$$

where I is, by hypothesis, a constant input. Let's see, for later, that if $\binom{x(t)}{z(t)}, \binom{x(t)}{\tilde{z}(t)}$ are both solutions of the previous system of ordinary differential equations (ODEs), then

$|z_i(t) - \tilde{z}_i(t)| = |z_i(0) - \tilde{z}_i(0)|e^{-t}\ \forall 1 \leq i \leq N-n$, since

$\frac{d}{dx}\big(z_i(t) - \tilde{z}_i(t)\big) = -\big(z_i(t) - \tilde{z}_i(t)\big)$

We will now focus on the study of the system of ODEs given by $(*)$ plus the i-th component of $z(t)$, which is passively driven by $x(t)$. We will label this $n+1$ dimensional autonomous system by $(**)$, and their solutions are going to be represented as $\binom{x(t)}{z_i(t)}$.

Now, we proceed to define some sets: let $S_{i_0}^{+} = \left\{\binom{x}{z_i} \in \mathbb{R}^{n+1} : z_i = \kappa\right\}$, where $\kappa := \sum_{i,j}|m_{ij}|$, being m_{ij} the components of M; if M_i is the i-th row of M, then let

$$S_{i_t}^{+} = \left\{\binom{x}{z_i} \in \mathbb{R}^{n+1} :\ z_i = e^{-t}(\kappa + \int_0^t e^s M_i F(Bx(s) + \omega I)ds), \forall x : x'(s) = -x(s) + AF(Bx(s) + \omega I)\right\}$$

so that the previous set is the image of $S_{i_0}^{+}$ under the flow of $(**)$ after a time t; we can define, analogously, $S_{i_0}^{-} = \left\{\binom{x}{z_i} \in \mathbb{R}^{n+1}: z_i = -\kappa\right\}$,and $S_{i_t}^{-}$, in the same way, making again the change $\kappa \mapsto -\kappa$; finally, we define the set $T_{i_0} = \left\{\binom{x}{z_i} \in \mathbb{R}^{n+1}: -\kappa \leq z_i \leq \kappa\right\}$, so that T_{i_0} is the closed set which includes the origin and whose frontier is $\partial T_{i_0} = S_{i_0}^{+} \cup S_{i_0}^{-}$.

It can be checked that T_{i_0} is positively invariant. Indeed,

$z_i\left(-z_i + M_i F(Bx + \omega I)\right) < 0$ for every $\binom{x}{z_i} \in \mathbb{R}^{n+1} \backslash T_{i_0}$, since, by definition, F is bounded by 1. Thus, no trajectory can escape $T_{i_0,}$ as, by the mean value theorem, the contrary would imply the existence of a time $c: z_i(c)\, z_i{}'(c) > 0$. A similar argument can be used to show that T_{i_0} is globally attracting.

We now prove that, for all $0 \leq t$, there exists a continuous function $f_{i_t}^{+}: \mathbb{R}^n \to \mathbb{R}$ fulfilling that

$$S_{i_t}^{+} = \left\{\binom{x}{z_i} \in \mathbb{R}^{n+1}: z_i = f_{i_t}^{+}(x)\right\}$$

Indeed, given $x_0 \in \mathbb{R}^n$, let $x(t)$ be the solution of $(*)$ with $x(0) = x_0$, which exists and is unique for all $\mathbb{R}$, since the system $(*)$ is globally Lipschitz. Then:

$f_{i_t}^{+}(x_0) = e^{-t}(\kappa + \int_0^t e^s M_i F(Bx(s-t) + \omega I) ds)$. The fact that $f_{i_t}^{+}$ is continuous comes from the continuity of the flow. Analogously, we can also define continuous functions $f_{i_t}^{-}$ such that

$$S_{i_t}^{-} = \left\{\binom{x}{z_i} \in \mathbb{R}^{n+1}: z_i = f_{i_t}^{-}(x)\right\}$$

Let's now see that $f_{i_t}^{+} \leq f_{i_s}^{+}\ \forall s < t$. If the contrary held,

$\exists x_0 \in \mathbb{R}^n, s < t:\ f_{i_s}^+(x_0) < f_{i_t}^+(x_0)$. In this case, let $x(t)$ be the solution of $(*)$ with $x(0) = x_0$, and $z_i(t), \tilde{z}_i(t)$ be such that $\binom{x(t)}{z_i(t)}, \binom{x(t)}{\tilde{z}_i(t)}$ are both solutions of the $n+1$ dimensional system, $(**)$. Suppose further that $z_i(0) = f_{i_s}^+(x_0)$, $\tilde{z}_i(0) = f_{i_t}^+(x_0)$. Then, by the definition of $S_{i_t}^+$, $z_i(-s) = \tilde{z}_i(-t) = \kappa$. But since we showed T_{i_0} is positively invariant, $\tilde{z}_i(-s) \leq \kappa$. Thus, using Boltzano's theorem, $\exists c \in [-s, 0)$ such that $z_i(c) = \tilde{z}_i(c)$. This, however, contradicts Picard's existence and uniqueness theorem, thus proving our claim that the functions $f_{i_t}^+$ are monotonically decreasing. Using the same procedure, it can be shown that $f_{i_s}^- \leq f_{i_t}^-\ \forall s < t$, being this complementary set of functions monotonically increasing.

We now define a sequence of functions, $\{f_{i_n}\}$, given by $f_{i_n} = \frac{1}{2}(f_{i_n}^- + f_{i_n}^+)$. We claim this sequence is Cauchy under the uniform norm. To see this, let $\varepsilon > 0$ be given; choose a natural number $n_0 > ln(\frac{2\kappa}{\varepsilon})$ so that

$\left|f_{i_{n_0}}^-(x) - f_{i_{n_0}}^+(x)\right| = 2\kappa e^{-n_0} < \varepsilon\ \ \forall x \in \mathbb{R}^n$, using what we proved earlier. Because of monotony, $f_{i_{n_0}}^- \leq f_{i_n}^- \leq f_{i_n} \leq f_{i_n}^+ \leq f_{i_{n_0}}^+\ \forall n_0 \leq n$, and therefore

$$\left|f_{i_n}(x) - f_{i_m}(x)\right| < \left|f_{i_{n_0}}^-(x) - f_{i_{n_0}}^+(x)\right| < \varepsilon\ \forall n_0 \leq n, m.$$

Since $\{f_{i_n}\}$ is a Cauchy sequence, lemma 5.2 stablishes the existence of a continuous function, f_i, such that $\{f_{i_n}\} \to f_i$ uniformly.

Define $T_{i_n} = \left\{\binom{x}{z_i} \in \mathbb{R}^{n+1}: f_{i_n}^-(x) \leq z_i \leq f_{i_n}^+(x)\right\}$, which forms a sequence of nested closed sets, $T_{i_m} \subset T_{i_n} \forall n \leq m$. Notice that every T_{i_n} is positively invariant and globally attracting given the flow of $(**)$, since

each of them is an image of T_{i_0}under the flow map. Moreover, $\bigcap_{n\in\mathbb{N}} T_{i_n} = \left\{\binom{x}{z_i} \in \mathbb{R}^{n+1}: z_i = f_i(x)\right\}$, since $\{f_{i_n}^+\}, \{f_{i_n}^-\} \to f_i$ uniformly. With all this, the set $\bigcap_{n\in\mathbb{N}} T_{i_n}$ is found to be positively invariant and globally attractive in the $n+1$ dimensional dynamical system given by $(**)$. The fact that this set is also invariant for all backward time comes from the fact that $\mathbb{R}^{n+1} \setminus \bigcap_{n\in\mathbb{N}} T_{i_n}$ is disconnected, being it the union of two connected invariant sets.

To finish the proof, we repeat the previous reasoning for every $1 \leq i \leq N-n$, finding that the set $\mathcal{M} = \left\{\binom{x}{z} \in \mathbb{R}^N: z = f(x)\right\}$ is invariant and globally attracting in the original N dimensional system, where $f(x) = (x_1, \dots, x_n, f_1(x), \dots, f_{N-n}(x))$. Since $f: \mathbb{R}^n \to \mathcal{M}$ is a homeomorphism, $\mathcal{M}$ is also a manifold. This concludes the proof ■

Proof of theorem 1.2: Again, if $rank(W) = N$ it is verified trivially that our Euclidean manifold is given by $\mathbb{R}^N$. If $rank(W) = n < N$, we decompose $W = BA$. It is going to be proven that $\mathcal{M} = \{v \in \mathbb{R}^N: v = Bx + \omega I, x \in \mathbb{R}^n\}$ is globally attractive and invariant. Let M be the matrix whose rows span the normal space of $\mathcal{M}$, so that MB gives the null $(N-n) \times n$ matrix. Let $v(t)$ be any solution of the v-model. Then: $\frac{d}{dt} Mv(t) = -Mv(t) + M\omega I$. Solving this linear differential equation, we find that $Mv(t) = (Mv(0) - M\omega I)e^{-t} + M\omega I$. Therefore, for any $t \in \mathbb{R}$:

$$v(t) \in \mathcal{M} \Leftrightarrow M(v(t) - \omega I) = 0 \Leftrightarrow M(v(0) - \omega I) = 0 \Leftrightarrow v(0) \in \mathcal{M}$$

With this we proved $\mathcal{M}$ is invariant. To show it is also globally attractive it is enough to see that, for any initial condition, $\lim_{t\to+\infty} M(v(t) - \omega I) = 0$ ■

For proposition 1.1, we define the notion of topological conjugacy:

Definition 5.1.5: Suppose we have two C^1 finite-dimensional dynamical systems, and suppose there exists a homeomorphism, h, such that, for any solution of the first system, $x(t)$, we have that $h(x(t))$ is also a solution of the second one. Then, we say that these dynamical systems are *topologically conjugate*. The homeomorphism, h, is called the conjugacy.

A more general definition, which the previous one can be shown to fulfil, would be to say that two dynamical systems are conjugate whenever there exists a conjugacy which commutes with the flow map. However, the previous definition is enough for the scope of this work. With this, we are ready to prove the last result of section 1.1, which will rely heavily on the previous proofs.

Proof of proposition 1.1: Suppose $u(t), v(t)$ are solutions of the u-model and the v-model, respectively. In the case where $rank(W) = N$, choose $h: \mathbb{R}^N \to \mathbb{R}^N: h(x) = Wx + \omega I$ to be the conjugacy, and verify that, for every solution $u(t)$ of the u-model, we have that

$$\frac{d}{dt}(Wu(t) + \omega I) = -Wu(t) + WF(Wu(t) + \omega I) + \omega I - \omega I =$$
$$-h\big(u(t)\big) + WF\left(h\big(u(t)\big)\right) + \omega I$$

Thus, $v(t) = h(u(t)$ is a solution of the v-model, and therefore h is a conjugacy.

In the case of $rank(W) = n < N$, decompose $W = BA$. Without loss of generality, we can assume the columns of B form an orthonormal basis of the image of W. Let's call $\mathcal{N}$ to the n-dimensional Euclidean manifold obtained in theorem 1.2, and x to the n-dimensional vector whose elements define the coordinates of $\mathcal{N}$ in the base given by B. Then, $v(t) \in$

$\mathcal{N} \Leftrightarrow v(t) = Bx(t) + \omega I$. Since $x(t) = B^T(v(t) - \omega I)$, we have that any $x(t)$ is the solution of

$$x'(t) = -x(t) + AF(Bx(t) + \omega I) \qquad (*)$$

This defines the neural manifold dynamics in $\mathcal{N}$.

Now consider any solution, $u(t)$, on the u-model's neural manifold, $\mathcal{M}$, and decompose $W = BA$ in the same way. Suppose that $x(t)$ is some solution of $(*)$, and perform a change of basis such that $\binom{x(t)}{z(t)} = \binom{A}{M}u(t)$, where the matrix M has a basis of W's kernel as it's rows, as done in the proof of theorem 1.1. There, it was proved that the solutions of the u-model in their neural manifold, $\mathcal{M}$, are given by

$$\binom{x(t)}{z(t)} = f\big(x(t)\big).$$

To finish the proof, define the homeomorphism:

$$h: \mathcal{N} \to \mathcal{M}: h(v) = \begin{pmatrix} A \\ M \end{pmatrix}^{-1} f\big(B^T(v - \omega I)\big) \ \forall v \in \mathcal{N}.$$

It is now straightforward to see that, for every solution $v(t) \in \mathcal{N}$ of the v-model, we have a solution $h(v(t)) \in \mathcal{M}$ of the u-model, as we wanted.

■

5.2. Proof of theorems 2.1 and 2.2

In this subsection, theorems expressing the universal approximation capabilities of firing rate models are going to be proved. For this reason, we present the following preliminary work:

Theorem 5.2.1 (approximation by sigmoidal superposition): Let K be a compact set of $\mathbb{R}^n$, and $f: K \to \mathbb{R}^m$ a continuous function. Then, given an

arbitrarily small $\varepsilon > 0$, there are $N \in \mathbb{N}$ and matrices A ($m\times N$), B ($N \times n$), θ ($N \times 1$) such that:

$$\max_{x\in K}\|f(x) - A\sigma(Bx + \theta)\| < \varepsilon$$

where $\sigma: \mathbb{R} \to (0, 1)$ is a *sigmoid map*, this is, a monotonically increasing continuous function such that $\lim_{x\to-\infty} \sigma(x) = 0$, $\lim_{x\to+\infty} \sigma(x) = 1$. The theorem also holds for more general functions, as long as they are non-polynomic (Leshno & Schocken, 1993).

For proofs of this important result in neural network theory, see (Cybenko, 1989; K. I. Funahashi, 1989; Tinchcombe et al., 1989).

In the following proofs, we will need a stronger definition of continuity:

Definition 5.2.1: Let $U \subset \mathbb{R}^m$. We say a function $f: U \to \mathbb{R}^m$ is Lipschitz continuous if there exists a constant $C > 0$ such that $\|f(x) - f(y)\| \leq C\|x - y\| \forall x, y \in U$; we say a function defined as before is locally Lipschitz if for every $p \in U$ there exists a neighborhood of p, U_p, where $f|U_p$ is Lipschitz continuous.

Lemma 5.2.1: If a function is differentiable, then it's locally Lipschitz; if a function is locally Lipschitz, it is Lipschitz in every compact set $K \subset U$.

For a definition of compactness, see the following subsection.

Lemma 5.2.2: Let $U \subset \mathbb{R}^n \times \mathbb{R}$ be an open set containing $(x_0, 0)$ and suppose that the maps $F, \tilde{F}: U \to \mathbb{R}^n$ are locally Lipschitz. Suppose also that $\left\|F(x,t) - \tilde{F}(x,t)\right\| < \varepsilon \quad \forall (x,t) \in U$ for some $\varepsilon > 0$. Let C be a Lipschitz constant in x for F . If $x(t), y(t): I \subset \mathbb{R} \to \mathbb{R}^n$ are solutions, respectively, of the systems of equations given by $x'(t) = F(x(t), t)$, $y'(t) = \tilde{F}(y(t), t)$, on some interval I and these solutions fulfil that $x(0) = y(0) = x_0$, then:

$$\|x(t) - y(t)\| < \frac{\varepsilon}{C}\left(e^{C|t|} - 1\right) \ \forall t \in I$$

For the proofs of the previous lemmas, we redirect the reader to (Hirsch et al., 2013b).

Lemma 5.2.3: let $v(t) \in \mathbb{R}^n$ solution of the system

$$v'(t) = -v(t) + f\big(v(t)\big),$$

where $f: \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz, and $u(t) \in \mathbb{R}^n$ be solution of

$u'(t) = -u(t) + f\big(v(t)\big)$. Then, $u(0) = v(0) \Rightarrow u(t) = v(t) \ \forall t \in \mathbb{R}$

Proof: Knowing that solutions are unique and defined for all time, it's only left to show that, for all the components of the solution vectors:

$\frac{d}{dt}\big(v_i(t) - u_i(t)\big) = -\big(v_i(t) - u_i(t)\big) \Leftrightarrow v_i(t) - u_i(t) = \big(v_i(0) - u_i(0)\big)e^{-t}$ ■

Proof of theorem 2.1: Let $\varepsilon, T > 0$ be given. Let $x(t)$ be the solution of the initial value problem of the system to be approximated for some $x(0) = x_0 \in \Omega$, and suppose $[0, T]$ is included in its maximal interval. Let $\widetilde{\Omega} = \{x(t) \in \mathbb{R}^n : x(0) \in \Omega, \ t \in [0, T]\}$ be the image of Ω under the flow map for the time interval $[0, T]$.

Start supposing that $W = BA$, being B, A, matrices as the ones described in lemma 5.1.1, whose precise value is going to be provided later. Define $y(t) := Au(t)$, so that $y'(t) = -y(t) + AF(By(t) + \omega I(t))$. Let C_G be a Lipschitz constant for G in Ω, whose existence is guaranteed since G is assumed to be differentiable, and thus locally Lipschitz (lemma 5.2.1). Then, theorem 5.2.1 guarantees the existence of $N \in \mathbb{N}$ and matrices $\omega(n \times N)$, $A\ (n \times N)$, $B\ (N \times n)$, such that:

$\|AF(By + \omega I) - y - G(y, I)\| < \frac{C_G \varepsilon}{e^{C_G T} - 1}$ for every $(y, I)^T \in \widetilde{\Omega} \times K_I$,

whenever F is endowed with appropriate thresholds. Here,

$$K_I := \{x \in \mathbb{R}^n : \|x\| \leq b\}$$

where b is taken to be big enough so that $I(t) \in K_I \; \forall t \in \mathbb{R}^+$ for all input trajectories drawn from the uniformly bounded set of functions considered in the statement of the theorem.

Choose $u(0)$ such that $Au(0) = x_0$. This system is solvable since, as stated earlier, we can assume that $rank(A) = n$, as these matrices form an open and dense set in $\mathbb{R}^{n \times N}$. Then, by lemma 5.2.2:

$$\|x(t) - Au(t)\| < \varepsilon \; \forall t \in [0, T].$$

Letting $R = A$, we found what we wanted to prove. Since $W = BA$, lemma 5.1.1 guarantees that $rank(W) = n$. ■

Proof of theorem 2.2: Let $\varepsilon, T > 0$ be given. Let $x(t)$ be the solution of the initial value problem of the system to be approximated for some $x(0) = x_0 \in \Omega$, and suppose $[0, T]$ is included in its maximal interval. Let $\widetilde{\Omega} = \{x(t) \in \mathbb{R}^n : x(0) \in \Omega, \; t \in [0, T]\}$ be the image of Ω under the flow map for the time interval $[0, T]$.

Define $y(t) \in \mathbb{R}^n$ as the solution of the initial value problem given by

$$y'(t) = -y(t) + AF\big(By(t) + DJ(t)\big) \,|y(0) = x_0$$

Where A is some $n \times (N - n)$matrix, B is $(N - n) \times n$ and D is of order $(N - n) \times m$. Let C_G be a Lipschitz constant for G in Ω, whose existence is guaranteed since G is assumed to be differentiable, and thus locally Lipschitz. It is clear that $AF(Bv + DJ) = AF((B \; D)\binom{v}{J})$. With this, we can make use of theorem 5.2.1, which assures that there exist matrices $A, (B \; D)$ as the ones we just defined such that, for *N* sufficiently big, ratify that:

$$\|AF(Bv+\omega J)-v-G(v,J)\| < \varepsilon \frac{C_G}{e^{C_G T}-1} \quad \forall (v,J)^T \in \widetilde{\Omega} \times K_J$$

whenever F is endowed with appropriate thresholds. Here,

$$K_J := \{x \in \mathbb{R}^n : \|x\| \leq b\},$$

where b is taken to be big enough so that $J(t) \in K_I \; \forall t \in \mathbb{R}^+$ for all the possible driving trajectories, J , drawn from the uniformly bounded set of functions considered in the statement of the theorem.

With all the previous, lemma 5.2.2 establishes that

$$\max_{t\in[0,T]} \|x(t)-y(t)\| < \varepsilon.$$

Let's define the connectivity matrices of the net, $W\ (N \times N), \omega(N \times m)$, as $W = \begin{pmatrix} 0 & A \\ 0 & BA \end{pmatrix}$, $\omega = \begin{pmatrix} 0 \\ D \end{pmatrix}$; we will split the solutions of the v-model, making $v(t) = (v^1(t), v^2(t))^T$, where $v^1(t) \in \mathbb{R}^n, v^2(t) \in \mathbb{R}^{N-n}$.

Now, we check that $v^2(t) = By(t) + DJ(t)$ is the value of the last components of the solution vector, $v(t)$, whenever $v^2(0) = Bx_0 + DJ(0)$, since:

$$\begin{aligned} v^{2'}(t) = By'(t) + DJ'(t) &= -By(t) + BAF\big(By(t) + DJ(t)\big) + DJ'(t) \\ &= -By(t) - DJ(t) + BAF\big(By(t) + DJ(t)\big) + DJ'(t) \\ &+ DJ(t) = -v^2(t) + BAF(v^2(t)) + DI(t) \end{aligned}$$

Using that, by definition, $I(t) = J'(t) + J(t)$. Now, since the equations of the first components of $v(t)$ are given by $v^{1\prime}(t) = -v^1(t) + AF\big(v^2(t)\big)$, choosing $v^1(0) = x_0$ we have, by lemma 5.2.3, that $v^1(t) = y(t)$. We define R as the matrix fulfilling that $Rv(t) = v^1(t)$. Therefore: $\max_{t\in[0,T]} \|Rv(t) - x(t)\| = \max_{t\in[0,T]} \|y(t) - x(t)\| < \varepsilon$

To finish the proof, we see that $\binom{A}{BA} = \binom{I_n}{B}A$, where I_n is the n-th order identity matrix, and thus, by lemma 5.1.1, $rank(W) = n$. With this, everything that was claimed in theorem 2.2 has been proved. ■

5.3. Proof of propositions 2.1 to 2.3

We start by presenting a handful of preliminary concepts that will be needed for the proof of proposition 2.1. The notion of compacity is of special interest in the fields of topology and analysis, and we have used it implicitly in the previous results. Because in the next proof we will need a more concise grasp of this concept, we present it here its formal definition.

Definition 5.3.1: Let X be a topological space and let $K \subseteq X$. Suppose $\{U_i\}_{i\in I}$is a family of open sets indexed by some set I. We say it is an open cover of K if $K \subset \bigcup_{i\in I} U_i$; We say $\left\{U_{i_j}\right\}_{j=1}^{n}$ is a finite subcover of K if $\left\{U_{i_j}\right\}_{j=1}^{n} \subseteq \{U_i\}_{i\in I}$ and $K \subset \bigcup_{j=1}^{n} U_{i_j}$; finally, we say that K is compact if for every open cover there exists a finite subcover for K.

In Euclidean spaces, the Heine-Borel theorem assures that a set is compact if and only if it is closed and bounded. Although the previous are necessary conditions in general topological spaces, they are not sufficient in more general topological spaces. This is the case in function spaces equipped with the uniform topology, as the one we deal in proposition 2.1, where distances are induced by the norm $\|r\|_\infty = \min_{0\le t\le T} \|r(t)\|$. The existence of compact sets in these Banach spaces is guaranteed by the Arzelà-Ascoli theorem (Munkres, 2002). We now present an extension theorem:

Theorem 5.3.1 (Kirszbraun): Suppose $K \subset \mathbb{R}^n$ is a compact set. Suppose also that $f: K \to \mathbb{R}^m$ is Lipschitz continuous, with a given constant $C > 0$. Then, there exists a function $\tilde{f}: \mathbb{R}^n \to \mathbb{R}^m$ with the same Lipschitz constant, C, such that $\tilde{f}\big|K = f$

For a recent constructive proof of this theorem, see (Azagra et al., 2021). Finally, we have the following lemma:

Lemma 5.3.1: Under the uniform norm, C^∞ functions are dense in the Banach spaces of continuous functions with compact domain.

The proof is a straightforward application of the Stone-Weierstrass theorem. See (Ortega Aramburu, 2002; Rudin, 1976) for a statement of this crucial result in mathematical analysis.

Proof of proposition 2.1: Considering the uniform norm, suppose $\mathcal{A}$ is a compact set of continuous functions of the form $r: [0, T] \to \mathbb{R}^n, T > 0$, as the statement of the proposition suggests. Let $\varepsilon > 0$ be given. Let $B_{\frac{\varepsilon}{3}}(r) = \left\{x \in C([0, T], \mathbb{R}^n): \|x - r\|_\infty < \frac{\varepsilon}{3}\right\}$ be the ball of radius $\frac{\varepsilon}{3}$ centred at r. Since $\left\{B_{\frac{\varepsilon}{3}}(r)\right\}_{r \in \mathcal{A}}$ is an open cover of the compact set $\mathcal{A}$, there exist a finite set of trajectories, $\{r_1, \dots, r_l\} \subseteq \mathcal{A}$ for which the inclusion $\mathcal{A} \subset \bigcup_{i=1}^{l} B_{\frac{\varepsilon}{3}}(r_i)$ holds.

We define $\hat{r}_i \in C^\infty([0, T], \mathbb{R}^n)$ such that $\|r_i - \hat{r}_i\|_\infty < \frac{\varepsilon}{3}, 1 \leq i \leq l$, using lemma 5.3.1.

Now consider the l trajectories in $\mathbb{R}^{n+l}$, $\{\tilde{r}_1, \dots, \tilde{r}_l\}$, whose components are given by

$\tilde{r}_{i_j}(t) = \hat{r}_{i_j}(t)$ if $j \leq n$; $\tilde{r}_{i_j}(t) = 0$ if $n < j, j \neq n + i$; ; $\tilde{r}_{i_j}(t) = 1 + t$ if $n < j,\ j = n + i$

It can be checked that $\tilde{r}_i(t) \neq \tilde{r}_j(t) \forall t \in [0,T], i \neq j$, and that $\tilde{r}_i(t) \neq \tilde{r}_i(s)$ $\forall 1 \leq i \leq l,\ s \neq t$, so these new trajectories will never cross each other. Now consider the compact set of $\mathbb{R}^{n+l}$ defined as follows:

$$K = \bigcup_{\substack{t \in [0,T] \\ 1 \leq i \leq l}} \tilde{r}_i(t)$$

Consider also the function $g: K \to \mathbb{R}^{n+l}$ which fulfils that $g(x) = \frac{d}{dt}\tilde{r}_i(t)$ $\forall x = \tilde{r}_i(t)$. Since every $\tilde{r}_i(t)$ is C^∞, and because these trajectories never cross, the function is well-defined. Moreover, since they all have continuous derivatives, by lemma 5.2.1 g is Lipschitz. Then, theorem 5.3.1 allows to extend g to a Lipschitz function $G: \mathbb{R}^{n+l} \to \mathbb{R}^{n+l}$.

Suppose we want to approximate an arbitrary trajectory $r \in \mathcal{A}$. To this end, from compactness we know that $\exists 1 \leq i \leq l$ such that $\|r_i - r\|_\infty < \frac{\varepsilon}{3}$. Now let $x(t)$ be the solution of the initial value problem $x'(t) = G\big(x(t)\big) | x(0) = \tilde{r}_i(0)$. From the existence and uniqueness theorem, it follows that $x(t) = \tilde{r}_i(t)$ $\forall t \in [0,T]$

Although we presented theorems 2.1 and 2.2 in terms of dynamical systems induced by C^1 functions to avoid technicalities, there is nothing restricting us to apply the same results to more general Lipschitz functions. Thus, by theorem 2.1, there exists a u-model whose output fulfils that $\left\|\tilde{r}_i(t) - \tilde{R}u(t)\right\| < \frac{\varepsilon}{3}$ $\forall t \in [0,T]$. Now, let $M \in \mathbb{R}^{n \times (n+l)}$ be given by $M = (I_n \quad 0)$. Let $R = M\tilde{R}$. Then, using the triangle inequality:

$$\begin{aligned}
\|r(t) - Ru(t)\| &\leq \|\hat{r}_i(t) - Ru(t)\| + \|r_i(t) - \hat{r}_i(t)\| + \|r_i(t) - r(t)\| \\
&\leq \left\|\tilde{r}_i(t) - \tilde{R}u(t)\right\| + \|r_i(t) - \hat{r}_i(t)\| + \|r_i(t) - r(t)\| \\
&< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon
\end{aligned}$$

For the v-model, use instead theorem 2.2 to obtain the same result. This concludes the proof.

■

Now we turn to the proof of proposition 2.2, which involves periodic solutions of differential equations, this is, trajectories s for which $\exists T > 0: s(t) = s(t+T)\ \forall t \in \mathbb{R}$. The orbits fitting in this definition consists of either fixed points or limit cycle oscillations.

The following deduction will require of new definitions and preliminary results. The first one involves the rigorous formulation of the notion of a flow, which we eluded so far in order to leave aside the more technical concepts from the main text.

Definition 5.3.2: Let $U \subseteq \mathbb{R}^n$ be open. A map $g: U \times \mathbb{R} \to U$ is called a *flow* if for every $s, t \in \mathbb{R}$ and $x \in U$, it fulfils the following axioms:

- $g_0(x) = x$
- $g_{s+t}(x) = g_s(g_t(x))$

In the case of continuous time dynamical systems induced by Lipschitz functions, the existence and uniqueness theorem allow to express the bundle of its solutions as a flow fulfilling the equation

$$\frac{\partial}{\partial t} g_t(x) = G(g_t(x)),$$

being x the values for the initial conditions. A remarkable property of this flow is the following:

Theorem 5.3.2: Let $G: U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ have Lipschitz constant C_G, and let the flow map of its induced dynamical system be given by $g: U \times \mathbb{R} \to U$. Then, it is fulfilled that

$$\|g_t(x) - g_t(y)\| \leq \|x - y\| e^{C_G|t|}\ \ \forall x, y \in U$$

In other words, this result assures that the flow mapping is continuous. For a proof, see (Hirsch et al., 2013b).

In the case of the flow on a neural manifold, $\mathcal{M}$, of a low-rank system, we will design it by $v_t(f(x))$, where $x \in \mathbb{R}^n$ stand for the initial position on the intrinsic coordinate system of the manifold. As it was seen in the proof of theorem 1.1 (and as it can be seen trivially in the proof of theorem 1.2), the coordinate chart of $\mathcal{M}$ can be expressed by a continuous bijection, given by $f: \mathbb{R}^n \to \mathcal{M}$, that can be chosen so that its inverse is defined by the linear surjection $R: \mathbb{R}^N \to \mathbb{R}^n$ restricted to $\mathcal{M}$, where R was the matrix set to define the readout of the network in theorems 2.1 and 2.2. This way, in order to study neural manifold dynamics, we assure the initial condition of the network, $f(x)$, lays on the manifold, and thus that the correspondent solution remains there forever.

Before we start proving proposition 2.2, we present another important result concerning the topology of compact spaces:

Theorem 5.3.3: Let X be a metric space. Then, it is compact if and only if for every sequence $\{x_n\} \subset X$ there exists a subsequence $\{x_{n_k}\} \subset \{x_n\}$ such that $\{x_{n_k}\} \to x$, for some $x \in X$.

An interesting idea used in the proof of this theorem, which is not going to be presented here, is the notion of a Lebesgue number. Given an open cover of a compact set, there exists a number, $\lambda > 0$, such that the ball with radius λ centred in any point on the compact region is included in some of the open sets forming the cover. For a proof, see (Ortega Aramburu, 2002). Now proposition 2.2 is proved.

Proof of proposition 2.2: Let $\varepsilon > 0$ be given. Let $S = \bigcup_{t \in \mathbb{R}} s(t)$ be the orbit set traced by some attracting solution. We will suppose there is just

one of them, as the process below can be repeated for each of the open sets composing the original set U, which is necessarily disconnected in the case there are many stable trajectories. Let V be a bounded open set such that $\Omega \subseteq \bar{V} \subset U$, being $\bar{V}$ its closure. It can be constructed as the union of a set of balls with sufficiently small radius.

Let $T_{\frac{\varepsilon}{3}}$ be big enough in order to satisfy that

$$\|g_t(x) - s(t - t_x)\| < \frac{\varepsilon}{3} \ \forall t \geq T_{\frac{\varepsilon}{3}}, x \in \bar{V}.$$

To prove the existence of such number, negate the claim to find a succession $\{x_n\} \subset \bar{V}$ for which, for each term, $\exists t \geq n$: $\left\|g_t(x_n) - s(t - t_{x_n})\right\| \geq \frac{\varepsilon}{3}$. As $\bar{V}$ is compact, by theorem 5.3.3 $\lim_k x_{n_k} = x \in \bar{V}$. But then, by theorem 5.3.2, $g_t(x)$ should not converge to $s(t - t_x)$, thus obtaining a contradiction. The open set $g_{T_{\frac{\varepsilon}{3}}}(V)$ has the property that

$$x \in g_{T_{\frac{\varepsilon}{3}}}(V) \Rightarrow \|g_t(x) - s(t - t_x)\| < \frac{\varepsilon}{3} \ \forall t \geq 0$$

Let $\lambda > 0$ be such that $\bigcup_{x \in S} B_\lambda(x) \subset g_{T_{\frac{\varepsilon}{3}}}(V)$. Given that $g_{T_{\frac{\varepsilon}{3}}}(V)$ is an open cover of S, λ can be taken to be the Lebesgue number of that cover. It's obvious that $\lambda < \frac{\varepsilon}{3}$.

Let $T_{\frac{\lambda}{2}}$ be defined analogously, so that $\|g_t(x) - s(t - t_x)\| < \frac{\lambda}{2} \ \forall t \geq T_{\frac{\lambda}{2}}, x \in \bar{V}$. Again, the existence of such time is proved as with $T_{\frac{\varepsilon}{3}}$. We define a function $z: U \to S : z(x) = \lim_n g_{nT}(x)$, so that it represents to which phase of the attracting orbit does the solution with initial value x converge. To conclude this list of definitions, let $W_x = \{y \in U: z(y) =$

$z(x)\}$, so that it is the set of initial values which converge to the same orbit with the same phase.

First, we prove $z(x)$ is continuous. To this end, let $\eta > 0$ be given, and let $n \in \mathbb{N}$ such that $T_{\frac{\eta}{3}} \leq nT$, where $T_{\frac{\eta}{3}}$ is defined as in the previous cases. Then, by theorem 5.3.2:

$$\exists \delta > 0: x, y \in V, \|x - y\| < \delta \Rightarrow \|g_{nT}(x) - g_{nT}(y)\| < \frac{\eta}{3}.$$

But then, by the definition of $T_{\frac{\eta}{3}}$: $\|z(x) - z(y)\| \leq \|z(x) - g_{nT}(x)\| + \|g_{nT}(x) - g_{nT}(y)\| + \|z(y) - g_{nT}(y)\| < \frac{\eta}{3} + \frac{\eta}{3} + \frac{\eta}{3} = \eta$

Now, by theorem 2.1 (resp. theorem 2.2) it can be built a u-model (resp. a v-model) whose flow fulfils that $\|g_t(x) - Rv_t(f(x))\| < \frac{\lambda}{2}$ $\forall x \in \bar{V}, 0 \leq t \leq T_{\frac{\lambda}{2}} + 2T$. From the triangle inequality, it follows that $\|s(t - t_x) - Rv_t(f(x))\| \leq \|s(t - t_x) - g_t(x)\| + \|g_t(x) - Rv_t(f(x))\| < \lambda$ for every $T_{\frac{\lambda}{2}} \leq t \leq T_{\frac{\lambda}{2}} + 2T$.

In the limit cycle case, without loss of generality, we can assume that λ is small enough so that $S \nsubseteq z(B_\lambda(s(t)))\forall t \in \mathbb{R}$. Since $Rv_t(f(x)) \in B_\lambda(s(t - t_x))$ $\forall t \in \left[T_{\frac{\lambda}{2}}, T_{\frac{\lambda}{2}} + 2T\right]$, parametrizing S and using the intermediate value theorem, it can be shown that there exists some value of time $\tau_1 \in \left[T_{\frac{\lambda}{2}}, T_{\frac{\lambda}{2}} + 2T\right]$ where $z(Rv_{\tau_1}(f(x))) = z(x)$. The same result is evident in the fixed-point scenario. Thus, in both situations one has that $Rv_{\tau_1}(f(x)) \in W_x \cap g_{T_{\frac{\varepsilon}{3}}}(V)$.

Let $n \in \mathbb{N}$ and suppose $\exists \tau_n \in \mathbb{R}^+$ such that $Rv_{\tau_n}(f(x)) \in W_x \cap g_{T_{\frac{\varepsilon}{3}}}(V)$. Then:

$$\begin{aligned}
&\left\|Rv_t\big(f(x)\big) - s(t-t_x)\right\| \\
&\qquad \le \left\|Rv_{t-\tau_n}\left(f\left(Rv_{\tau_n}\big(f(x)\big)\right)\right) - g_{t-\tau_n}\left(Rv_{\tau_n}\big(f(x)\big)\right)\right\| \\
&\qquad + \left\|g_{t-\tau_n}\left(Rv_{\tau_n}\big(f(x)\big)\right) - s(t-t_x)\right\| < \frac{2\varepsilon}{3} \qquad \forall t \\
&\qquad \in \left[\tau_n, \tau_n + T_{\frac{\lambda}{2}} + 2T\right]
\end{aligned}$$

In addition, repeating the argument used for τ_1, there exists a $\tau_{n+1} \in \left[\tau_n + T_{\frac{\lambda}{2}}, \tau_n + T_{\frac{\lambda}{2}} + 2T\right]$ such that $Rv_{\tau_{n+1}}\big(f(x)\big) \in W_x \cap g_{T_{\frac{\varepsilon}{3}}}(V)$. Thus, by induction, we have, for all $t \ge T_{\frac{\lambda}{2}} + 2T$:

$$\begin{aligned}
\left\|g_t(x) - Rv_{\tau_n}\big(f(x)\big)\right\| &\le \|g_t(x) - s(t-t_x)\| \\
&\quad + \left\|Rv_t\big(f(x)\big) - s(t-t_x)\right\| < \varepsilon
\end{aligned}$$

■

To finish this subsection, the proof of proposition 2.3 is going to be presented, relating universal approximation capabilities of neural networks to the simulation of symbolic computations whenever finite (yet arbitrary) memory is required. From now on, we will follow the Church-Turing thesis so that for any decidable function, f, we will assume that an effective procedure capable of computing its outputs can be implemented by a Turing machine. In order to link these theoretical devices with dynamical system theory, we present the following theorem, which claims that discrete time dynamical systems possess the power of universal computation.

Theorem 5.3.4: Every Turing machine is equivalent to an iterated map of the form $F: \mathbb{N} \to \mathbb{N}$.

Proof: Describe the configuration of a given Turing machine as follows: suppose that $\{i_n\}$ is the sequence describing the tape, being each term the number codifying each symbol of the tape once a particular order has been assigned to the alphabet set, where $i = 0$ stands for the blank symbol, the only appearing infinitely many times; $\{p_n\} = \{2, 3, 5, 7, 11 \dots\}$ is the sequence of prime numbers; $s \in \mathbb{N}$ the state, being $s = 1$ the codification of the initial state; $N \in \mathbb{N}$ the position on the tape. Then, the natural number

$$x = 2^s 3^N \prod_{n=1}^{\infty} p_{n+2}^{i_n}$$

codifies uniquely the configuration of the Turing machine. Thus, it can be constructed a function $F: \mathbb{N} \to \mathbb{N}$ implementing the transition function with which the given Turing machine operates.

■

However, concerning neural computations, since the models studied in this paper take continuous values of both time and phase variables, it can be questioned how these kinds of systems could effectively implement an essentially discrete model of computation. It is thus that we should define in which way it could be understood that a continuous system simulates a discrete one. For that purpose, we present some theory developed by Branicky in (Branicky, 1995).

Definition 5.3.3 (S-simulation): Let X, Y be topological spaces. Let $g: X \times \mathbb{R} \to X$ be the flow of a continuous dynamical system, and $F: Y \to Y$ an iterated map describing a discrete one. We say the former system S-simulates the later whenever there exists a continuous surjective function $\psi: D \subseteq X \to Y$ and some $T_0 \in \mathbb{R}^+$ such that

$$\psi\left(g_{nT_0}(x)\right) = F^n\big(\psi(x)\big)\ \forall x \in D, n \in \mathbb{N}$$

where F^n denotes the map F n times composed with itself.

This definition is similar to definition 5.1.4, which presented the notion of topological conjugacy. In this case, however, the function ψ needs not to be injective, and thus topological conjugacy is a stronger hypothesis compared to S-simulation.

With this, we could set up a neural simulation of some Turing machine as follows: choose some intrinsic dimensionality for the neural manifold (it is going to be proved that 3 dimensions are sufficient) and place a collection of open sets in this immersed space, each one of which will stand for a particular configuration of the Turing machine's tape, position and current state. Then, use theorems 2.1 and 2.2 to define a dynamical system capable to implement the transition function of the algorithm. This could be performed by joining the mentioned open regions with concrete phase trajectories, indicating the system which steps should it follow to simulate the given computation. In order to formalize the previous intuition, we present the following theorem.

Theorem 5.3.5: Let $F: Y \subset \mathbb{Z}^n \to \mathbb{Z}^n$ define a discrete dynamical system and suppose Y is a compact subset. Then, it can be S-simulated by a continuous-time dynamical system in $\mathbb{R}^{2n+1}$ whose dynamics are induced by a Lipschitz function.

For a proof see (Branicky, 1995), theorem 5.7.

It is interesting to see that this dynamical system is robust, being able to carry out the simulation also in the cases where the flow is slightly

perturbed or when some small enough amount of noise is added to the model.

Indeed, in the original proof (Branicky, 1995) it was constructed a continuous nearest integer function like

$$[\cdot]_C: \mathbb{R} \to \mathbb{R}: [x]_C = \begin{cases} i & i - \frac{1}{3} < x \leq i + \frac{1}{3} \\ 3x - 2i - 1 & i + \frac{1}{3} < x \leq i + \frac{2}{3} \end{cases}$$

for all $i \in \mathbb{Z}$, subsequently defining the map inducing the S-simulation, $\psi: \mathbb{R}^{2n+1} \to \mathbb{R}^n$, like $\psi(x) = ([x_1]_C, \dots, [x_n]_C)^T \forall\, x \in \mathbb{R}^{2n+1}$.

Now suppose that $g_t(x)$ is the flow simulating the given iterated map. To say that this system S-simulates a given discrete time dynamical system is equivalent to say that for every initial condition $x \in \psi^{-1}(F(Y))$ it is fulfilled that $g_{nT_0}(x) \in \psi^{-1}\left(F^n(\psi(x))\right)$ for all $n \in \mathbb{N}$. It can be seen from the definition of ψ that $\psi^{-1}(Y)$ has nonempty interior, since for every $y \in Y$, if $x = (y, y, 0)^T$ then $B_{\frac{1}{3}}(x) \subset \psi^{-1}(y)$. Therefore, given any $M \in \mathbb{N}$, for every $x \in B_{\frac{1}{3}}(x)$one can find some $r > 0$ such that $B_r(g_{nT_0}(x)) \subset \psi^{-1}\left(F^n(\psi(x))\right)$ $\forall n \leq M$. This is saying that if some small enough perturbation is added to some trajectory starting in an open set, it will still be able to efficiently carry on the computing process during a sufficiently long period of time, thus ensuring that our setup is robust. In the proof of proposition 2.3, which is presented below, it is shown how robustness is preserved by neural manifold implementations.

Proof of proposition 2.3: By the Church-Turing thesis, take the effective procedure which sustain the computable function f to be performed by some Turing Machine. Let S be a set of natural numbers, each of which

codifying a given tape uniquely as it was done in the proof of theorem 5.3.4. Since we restrict our computations to require only a finite amount of memory, there exists $l \in \mathbb{N}$ such that the length of the nontrivial segment of the tape is less than l for every tape codified in S. Let $M \in \mathbb{N}$ be the maximum number of epochs necessary to terminate the program in any input of the given set, S.

By theorem 5.3.4, there is a map $F: Y \subset \mathbb{N} \to \mathbb{N}$ implementing the computation, where we define the set $Y = \bigcup_{i=0}^{M} F^i(S)$. It is compact since it is a finite set of natural numbers, each of which represents some configuration of the Turing machine. It is invariant since after a time M every possible computation has been completed, reaching thus a stable state of the mapping.

By theorem 5.3.5, there exists a Lipschitz flow $g: \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3$ S-simulating the dynamical system produced by F, where the function ψ defining the simulation can be the one we already presented.

Then, choose some $0 < \eta < \frac{1}{3}$ and let $\Omega = \{x \in \mathbb{R}^3 : \|x - y\| \geq \eta \ \forall y \in \mathbb{R}^3 \backslash \psi^{-1}(S)\}$. It can be seen that $\Omega \subset \psi^{-1}(S)^\circ$, where $\psi^{-1}(S)^\circ$ denotes the interior of $\psi^{-1}(S)$, and that Ω is compact.

Let $\widetilde{\Omega} = \bigcup_{i=0}^{M} g_{iT_0}(\Omega)$, which is still compact. Since g S-simulates the discrete dynamics, it follows that $\widetilde{\Omega} \subset \psi^{-1}(Y)^\circ$, where $\circ$ stands again for the interior. This is because the Lipschitz condition makes g_{T_0}a homeomorphism, and thus it maps interiors to interiors.

Since $\psi^{-1}(Y)^\circ$ is an open cover of $\widetilde{\Omega}$, let $\lambda > 0$ be a Lebesgue number. Using theorem 2.1 in the u-model case or theorem 2.2 for the v-model, we have can find a recurrent network such that

$\|g_t(x) - Rv_t(f(x))\| < \lambda \; \forall t \in [0, MT_0], x \in \Omega.$

With this, it is seen from the definition of $\widetilde{\Omega}$ that $Rv_{nT_0}(f(x)) \in \psi^{-1}\left(F^n(\psi(x))\right) \; \forall n \leq M, x \in \Omega$, which implies that

$\psi(Rv_{nT_0}(f(x)) = F^n(\psi(x)).$

It only remains to be shown that the neural manifold of this system is 3-dimensional. Indeed, theorems 5.3.4 and 5.3.5 together predict that g can be a 3-d flow. Thus, using theorems 2.1 and 2.2 $rank(W) = 3$ in both models. Therefore, theorems 1.1 and 1.2 confirm that the emerging manifolds will be 3-dimensional. This concludes the proof. ■

Using wat it is presented in the previous proof, since Ω has nonempty interior, if $x \in \Omega^{\circ}$, using that $\|g_t(x) - Rv_t(f(x))\| < \lambda$ and the same arguments we fenced to show g is robust, it also follows that the computational process held by the neural manifold trajectory $v_t(f(x))$ will tolerate small perturbations without affecting its performance, both if we slightly modify the trajectory or if we do so for its initial condition by a sufficiently small amount.

5.4. Descriptive statistics of low-rank models

In this subsection a precise formulation of the statement presented in section 2.3, where it was claimed that in low-rank wirings the correlation matrix's number of degrees of freedom was of order N, is going to be presented. We start by making the idea of order more precise:

Definition 5.4.1: We say that a real valued function f is of order g, being g a positive valued function, whenever $\lim_{N\to\infty} \frac{|f(N)|}{g(N)} \leq K$, for some $K \in \mathbb{R}^+$, and we write it $f \sim O(g)$.

The flow of the set of ODE's defining our neural systems is going to be given by ν: $\mathbb{R}^N \times \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}^N$ and it is going to be written like $\nu_t(x, \alpha)$, where $\alpha \in \mathbb{R}^p$ stand for the vector of bifurcating parameters defining the system, which include the weights, the firing thresholds or other scalars that could be added to improve the model's fitting to data, such as membrane time constants or parameters describing different input-spiking rate relations.

Let's now describe the statistical measures which are going to be required: $E[\nu_t(x, \alpha)_i]$ will stand for the mean of the i-th component of the trajectory starting in x with bifurcation vector α during some arbitrary lapse of time $[0, T], T > 0,$ and will be given by

$$E[\nu_t(x, \alpha)_i] \coloneqq \frac{1}{T} \int_0^T \nu_t(x, \alpha)_i \, dt$$

Taking the previous definition, $C_{ij}(x, \alpha) \coloneqq E\left[\nu_t(x, \alpha)_i \nu_t(x, \alpha)_j\right] - E[\nu_t(x, \alpha)_i] E\left[\nu_t(x, \alpha)_j\right]$ is going to stand for the covariance between the i-th and the j-th component of a trajectory starting in x and whose definitory parameter vector is α. Of course, the behavior of a single trajectory need not to be representative. Thus, if we are interested in studying the behavior of the model in some bounded open region of the phase space, we can define

$$C_{ij}(\alpha) \coloneqq \frac{1}{\mu(U)} \int_U C_{ij}(x, \alpha) dx$$

to be the covariance of the neural model specified by the parameter vector α averaged over the given domain, U. Here, it was used that $\mu(U) \coloneqq \int_U 1dx$ is the Lebesgue measure of the open region U.

Finally, in order to find the correlation matrix, P, one could normalize the covariance matrix using the formula

$$\mathrm{P}(\alpha) = S^{-1}(\alpha)C(\alpha)S^{-1}(\alpha)$$

Where $S(\alpha)$ is the diagonal matrix whose entries consist on the variances of each component of the solution vector, this is, $S_{ij}(\alpha) = C_{ij}(\alpha)\delta_{ij}$, where $\delta_{ij}$is the Kronecker delta. From here, it could be seen that the matrix $\mathrm{P}(\alpha)$ is defined by a continuous map of the form $\mathrm{P}\colon \mathbb{R}^p \to \mathbb{R}^{N\times N}$.

Thus, in order to fully characterize the matrix $\mathrm{P}(\alpha)$ we must specify the p independent parameters which completely define the flow of our neural models.

In a general frame, it could be assumed that three different classes of parameters exist. In the first place, we could find parameters which affect the whole network, like the activation of a neuron of a nearby region that projects outputs to the system. These parameters do not depend on the size of the network, and thus remain constant with N.

Then, there are the ones that depend on individual neurons, like spike thresholds, time constants or the afferent synaptic strengths coming from outside the network. The number of such scalars increases linearly with the network size.

Finally, we have the parameters that define pairwise relations between units, like synaptic strengths between neurons of the network itself.

In general, if no specific structure is imposed on these last types of arrays, or if their components are assumed to be mutually independent, then their number grows quadratically as the network size increases. In this scenario $p \sim O(N^2)$.

However, if we assume these parameters are mutually dependent, as in the case we impose connection matrices to be restricted to have a given lower dimensional rank, then their size also grows in a linear fashion, as it was shown in section 2.3, and thus one has that $p \sim O(N)$. Indeed, it has been already shown how the number of parameters needed to fully specify a $N \times N$ rank-n matrix is given by $p = n(2N - n)$. Thus, if we also take into consideration the other types of parameters, whose cardinality cannot increase faster than linear, one could divide by N to find that $p \sim O(N)$, as claimed.

In section 2.3 it was also assured that if the dynamics were confined to a lower dimensional Euclidean manifold the same would hold true. In this case, if it is assumed the manifold is n-dimensional, then a PCA would find that the covariance matrix would only have n non-null eigenvalues, which in turn would imply that $rank(\mathrm{P}) = rank(C) = n$. Then, since we already know that the complexity of a rank-n matrix is given by $p = n(2N - n)$, we would also have that $p \sim O(N)$, as before.