UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA CIVILE, EDILE E AMBIENTALE

Corso di Laurea Magistrale in Mathematical Engineering

Tesi di Laurea

Modelling Brain Dynamics with Recurrent Neural Networks

Relatore:
Prof./Dr. Amos Maritan
Correlatore:
Prof./Dr. Samir Simon Suweis

Laureando:
Luca Taffarello
2016823

Anno Accademico 2023/2024

Contents

Abstract

Recurrent Neural Networks (RNNs) are a powerful tool to shed light on how brain may work. RNNs can recapitulate different dynamical phases observed in cortical circuits, such as silent or chaotic state, and provide a simple explanation of asynchronous rate activity in neural systems and information processing capabilities. In this thesis, we will exploit advanced approaches in statistical physics to investigate emergent phases of RNNs with random coupling. Integrating tools from Dynamical Mean Field Theory and Random Matrix Theory with numerical simulations, we analyse the properties of the network with particular focus on its stability and the effect of quenched external inputs on the dynamical phases of the network. Finally, we will also study the equivalence of two frequently used forms of recurrent rate models with random interactions.

Introduction

The brain is one of the most complex systems in the universe. It contains approximately 10^{11} neurons that receive inputs from thousands of synapses, resulting in an overwhelmingly high total number of connections. In order to unravel the general principles behind brain functions and explain the emergent behaviors of cortical cirtuits, researchers are constantly challenged by the complexities arising from this intricate network of interactions.

Within the vast and interdisciplinary field of neuroscience, involving contributions from biology, psychology, mathematics and physics, theoretical neuroscience in becoming increasingly influential. To describe fundamental brain processes such as learning, memory and perception, this branch employs analytical and computational tools to understand how neural systems operate, taking advantage of statistical physics to model and analyse their behavior. The large number of degrees of freedom within neural networks and the non-linearity of the interactions make this branch of physics an ideal tool for modelling those systems.

By exploiting advanced approaches in statistical physics, the objective of this thesis is to understand how the underlying structure of the neural network determines its dynamical behavior and to characterize the different configurations (phases) that influence how the system responds to external inputs. On general grounds, our aim is to understand how the brain can flexibly adapt to process information from a complex environment and whether there is a state in which it is particularly sensi-

tive to external perturbations.

By treating the brain as a complex system with emergent behaviors, we can justify the inherent simplifications in modeling. As we focus on grasping the essential principles resulting from the interconnections between neurons, it becomes unnecessary to incorporate all features of isolated neurons or the intricate biological structure of the network within our models. Instead, we can neglect certain details while still capturing significant collective aspects.

The choice to model neural network dynamics using Recurrent Neural Networks (RNNs) composed of firing rate units with randomly distributed synaptic connections is based on this idea. Importantly, we consider asymmetric couplings because in this scenario the system cannot be described by an energy function, which is only possible when they are chosen to be symmetric. Consequently, the model does not converge to an attractor but instead displays a rich and complex dynamical behavior. In Chapter 1, we explore how to derive this model from a more biologically realistic description of spiking neurons. Our aim is to clarify the essential features while recognizing the simplifications made. Furthermore, we highlight the distinction between two frequently used forms of firing rate models that derive from this description. This involves capturing the different roles played by the neuronal variables involved in each model, a clarification that, to our knowledge, remains ambiguous in existing literature.

We use Dynamical Mean Field Theory (DMFT), an analytical tool from statistical physics, to understand the collective behavior of RNNs. DMFT simplifies the dynamics of a large ensemble of neurons into an equation for a single neuron embedded in a fluctuating field with self-consistent statistics, that accurately describes the system's behavior in the limit of the number of neurons approaching infinity. This equation can be obtained using the Martin-Sigga-Rose-De Dominicis-Janssen (MSRDJ) path integral formalism or the more intuitive dynamical cavity approach. In Chapter 2, we derive the DMFT equation and explain the underlying principles

and methodologies of both approaches.

In Chapter 3, we rigorously analyse the stationary state of the model by employing the stationary limit of the DMFT equation. In particular, we focus on the evaluation of the Largest Lyapunov Exponent (LLE). Interestingly, when the parameter regulating the synaptic connection strength increase over a critical threshold, the network undergoes a phase transition from a fixed point solution to a chaotic state. This transition bears significant functional implications from an information processing perspective.

In driven networks, the introduction of an external input to the neurons has been observed to stabilize their dynamics. In Chapter 4, we investigate this phenomenon by introducing a quenched Gaussian-distributed external input to the network. Analyzing the resulting phase diagram, we uncover bistable phases in which the network can settle into two distinct attractors based on the initial conditions.

In Chapter 5, we explore two main numerical methods used to investigate the dynamical behavior of RNNs: the Orbit Separation Method (OS) for computing the LLE through direct simulations, and a method for numerically solving the self-consistent DMFT equation. Other numerical approaches, such as those used for calculating the correlation function of neuronal activity and the Lyapunov exponent through DMFT, are comprehensively discussed in the corresponding chapters.

In the concluding chapter (Conclusions and Outlook), we provide a perspective on how to advance this research line, based on the critical hypothesis of the brain. Indeed, in our pursuit of uncovering fundamental principles, we often neglect a crucial aspect of the brain: its adaptability, essential for operating in a complex environment. On general grounds, adaptive models of complex living systems tune themselves close to a critical state, at the borderline between order and disorder, providing an excellent trade-off between flexibility and accuracy in information processing capabilities. Motivated by this insight, we propose a model in which synaptic couplings are dynamic variables, undergoing modifications according to a simple

Hebbian rule, meaning that when two neurons fires synchronously their connection will get stronger over time. Biologically, this process of enhancing or weakening the synaptic connection is known as synaptic plasticity and is a well-established phenomenon. We speculate that this simple mechanism could bring the system to self-tune close to a critical state. Finally, we emphasize the preparatory nature of this work, as it provides the crucial analytical and numerical tools to further exploration.

Chapter 1

Recurrent Neural Networks

In this chapter, we present the fundamental aspects of the model used in the thesis: a fully-connected recurrent neural network of firing rate units with random synaptic couplings. This model provides a simplified alternative to more complicated spiking models, allowing for a reasonably deep analytical analysis. We explain the reason for this choice, clarifying the simplifications and underlying assumptions.

However, before proceeding further, it is necessary to provide some insight into the fundamental biological features of the neuron's structure ad function. Neurons are specialized cells that communicate with each other by sending electrical impulses known as action potentials or, more simply, spikes. From a structural point of view, neurons are composed of distinct components such as dendrites, soma, and axons, each of them contributing to their unique functions. Dendrites, branching structures extending from the soma, receive incoming signals from neighboring neurons through the exchange of neurotransmitters, such as glutamate, dopamine and serotonin. The functional role of neurotransmitters is to regulate the opening and closing of ion channels, that allow the passage of ions across the cell membrane. The accumulation of ions within the soma modifies the membrane potential, which represents the voltage difference between the inside and outside of the neuron and, when a critical threshold is reached, a feedback process triggers the generation of an action potential,

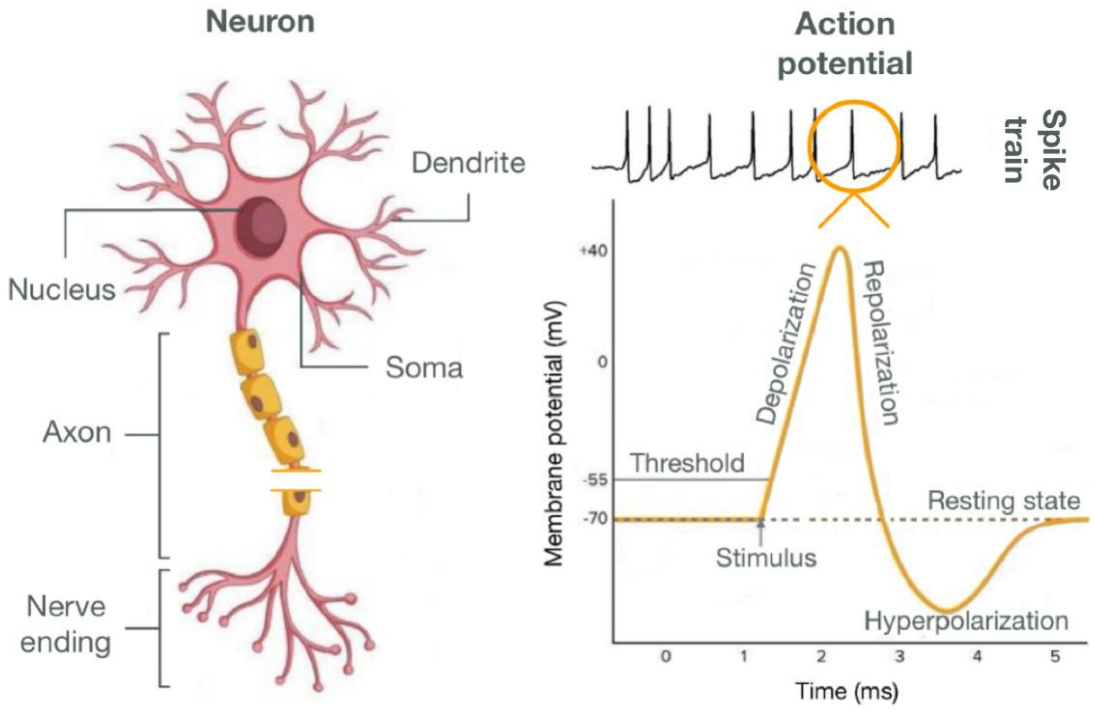


Figure 1.1: (Left) Schematic representation of the structure of a neuron. Soma (the cell body which contains the nucleus), dendrites and axon can be clearly distinguished. (Right) Example of a neuronal action potential, a short voltage pulse of the cell membrane potential of approximately 100 mV and with a typical duration of 1-2 ms. A chain of action potential emitted by a single neuron is called spike train.

i.e. a rapid fluctuation of the membrane potential (see figure 1.1). This spike, propagates via the axon, a slender fiber that connects the soma to neighboring neurons, allowing for intercellular communication. The juction that connect the terminal part of the axon of the transmitting (pre-synaptic) neuron to a dentrite of a receiving (post-synaptic) neuron is known as synapse.

1.1 From spikes to firing rate

Large synaptic connectivity represents a defining feature of neural circuits. As an example, a neuron in our cortex receives thousands of synaptic inputs. With network models, we can investigate the computational capabilities offered by such connectivity, employing both analysis and simulations. The most direct approach to simulate

neural networks involves the synaptical connection of spiking neurons. However, this method presents important analytical, computational, and interpretative challenges. For instance, the Hodgkin-Huxley models, which are useful to characterize the dynamics of the ion channels and identifying potential mechanisms to understand the patterns of firing neurons, are described by a complicated system of coupled differential equations [18]. Moreover, the dynamics of spiking models span a wide range of timescales, from channel opening that can happen in less than a millisecond, to collective network processes that may be many orders of magnitude slower.

Instead, we employ a simpler approach by constructing networks of neuron-like units with outputs represented by firing rates rather than action potentials. Firing rate models allow to neglect the modelling of short timescale dynamics required for simulating spike sequences, thereby rendering them considerably easier to simulate. Meanwhile, firing rate models provide an good alternative to investigate the dynamics of neural networks, as they enable us to study the input-output properties of neurons employing simpler analytical tools, and they can be easily extended to include interactions among different neurons.

Biological spiking neurons produce predictable spike patterns in response to injected current or synaptic input. However, deterministic models can produce reliable sequences of spikes only in the scenario where all inputs are known, which is unreasonable in the case of neurons inside a complex network. Consequently, while spiking models may appear to offer greater precision, this may not be attained in practice. Furthermore, while cortical neurons receive numerous inputs, the probability of finding a synaptic connection between randomly selected neurons is relatively low. To address this problem, network models often use averaging units to represent the collective response of multiple neurons with similar functional role. These units allow for denser interconnections and fewer units needed to build the model. If neural responses are characterized by firing rates, averaging is straightforward, but if responses are spikes, it's unclear how to average them.

Therefore, firing rate models allow us to construct a simplified networks, still capturing significant properties of neural assemblies.

Action potentials vary significantly in length, amplitude, and shape, but because of their short average duration (approximately 1 ms), we may describe them as point processes that transfer information through their timing. Mathematically, the series of spikes produced by a neuron is completely determined by the neural response function $\rho(t)$, which involves delta function spikes occurring at times t_i $(i = 1, \ldots, n)$ when a neuron discharged an action potentials:

$$\rho(t) = \sum_{i=1}^{n} \delta(t - t_i) \tag{1.1}$$

Clearly, the neural response function exhibits variability across different trials with the same stimulus presentation. Because of this it requires a statistical or probabilistic approach. As a consequence, in firing rate models, the accurate representation of a spike sequence given by $\rho(t)$ is substituted with an approximate description provided by the firing rate $r(t)$. Moving from spikes to firing rates requires counting the number of action potential within short time intervals and dividing by the interval duration (obtaining the so-called spike-count). However, for small time intervals that allows for high temporal resolution, the spike-count on any given trial is likely to be either 0 or 1, resulting in only two potential firing rate values. To avoid this issue while maintaining the required temporal accuracy, averaging across different trials is employed. Therefore, the continuous time firing rate can be defined as:

$$r(t) = \frac{1}{\Delta t} \int_{t}^{t+\Delta t} d\tau \, \langle \rho(\tau) \rangle \tag{1.2}$$

and it represents the probability density that a spike occurs during the time interval Δt. This approach enables us to switch from a point process to a continuous rate model. The validity of the latter depends on how well the average firing rate of the units represents the impact of actual spike sequences on the dynamical behavior

of the network. Substituting the neural response function with the firing rate can be justified by large amount of synaptic inputs received by each neuron, and it is acceptable if the network dynamics is not greatly affected by the variability in spike sequences from trial to trial. While individual synaptic inputs may vary significantly between trials, summing inputs from multiple synapses activated by independent pre-synaptic spike trains leads to a mean total input that increase proportionally with the number of synapses, while its variability increases only as the square root of it. Consequently, for uncorrelated pre-synaptic spike trains, the replacement of action potentials with firing rates is expected not to change the network dynamics in a relevant way. On the other hand, when a large portion of the inputs to a neuron are correlated, as when pre-synaptic neurons fires synchronously, a rate model may fail to accurately describe the system behavior. Moreover, the synaptic input resulting from a pre-synaptic neuron undergoes a filtering effect due to the propagation of the current from the synapse to the soma of the of the post-synaptic neuron. Slow synaptic or membrane dynamics contribute to temporal averaging, which diminishes the impact of variability in spike trains and supports the use of firing rates. In other words, when the propagation of the input signal is slow compared to the typical duration of the interspike intervals, firing rate models provide a better accuracy.

1.2 Synaptic input current

In order to develop a firing rate model, we follow the approach described by Dayan and Abbot [8]. We first need to understand how the the firing rate of the pre-synaptic neurons affect the overall synaptic input current received by the post-synaptic neuron. As a typical procedure to determine firing rate response curves (i.e. the neural response to a given stimulus) is obtained by experimentally injecting current into the soma, the easiest way to define the total synaptic input is by the total current reaching the soma due to pre-synaptic action potentials.

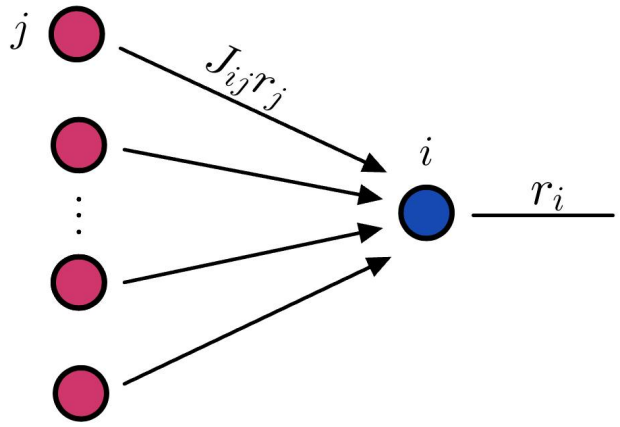


Figure 1.2: The firing rates r_j $(j = 1, \ldots, N)$ of the pre-synaptic neurons are weighted by the synaptic coupling J_{ij} and summed to compute the input current I_i to the i-th neuron (see eq. (1.4)) which then determine its output rate r_i.

Therefore, consider the i-th neuron of the model connected to N pre-synaptic neurons with firing rates r_j $(j = 1, \ldots, N)$ (fig. 1.2). Denote with I_i the total input current to the i-th neuron and with r_i its post-synaptic firing rate. Firstly, it is necessary to comprehend how the synaptic input I_i is influenced by the pre-synaptic spikes. When an action potential is generated by the neuron j at time $t = 0$, we express the synaptic current produced in the soma of the post-synaptic neuron at time t as $J_{ij}K_j(t)$. Here, J_{ij} represents the synaptic weight (or coupling efficacy) that determines the magnitude and sign of the synaptic current induced by the j-th neuron. $K_j(t)$ is the (normalized) synaptic kernel, which characterize the temporal profile of the synaptic current and it depends on the dynamics of the synaptic conductances activated by pre-synaptic spikes [8]. Then, if we consider that the effects of spikes at a single synapse combine linearly, the cumulative synaptic current at time t originating from a sequence of pre-synaptic spikes occurring at input j at times t_k can be described as:

$$i_j(t) = J_{ij} \sum_{t_{k,j}<t} K_j(t - t_{k,j}) = J_{ij} \int_{-\infty}^{t} d\tau \; K_j(t - \tau)\rho_j(\tau) \tag{1.3}$$

where $\rho_j(\tau)$ represent the neural response function of the j-th neuron (eq. 1.1) and the second equality is derived by integrating over the sum of delta functions in its definition.

Finally, assuming a linear interaction among various synaptic currents, the total synaptic current originating from all pre-synaptic inputs is straightforwardly obtained by summing:

$$\begin{aligned} I_i(t) &= \sum_{j=1}^{N} J_{ij} \int_{-\infty}^{t} d\tau \; K_j(t-\tau)\rho_j(\tau) \\ &\approx \sum_{j=1}^{N} J_{ij} \int_{-\infty}^{t} d\tau \; K_j(t-\tau)u_j(\tau) \end{aligned} \tag{1.4}$$

Note that in the second line we have substituted the neural response function $\rho_j(\tau)$ with the firing rate of the j-th neuron $u_j(\tau)$. This is the crucial step in developing a firing rate model and, as we have previously underlined, it is a good approximation when there are many input synapses, so that the trial-to-trial variability of the neural response function can be neglected (see eq. (1.2) and the consequent discussion).

Let's denote for the sake of simplicity:

$$F_j(t) = \int_{-\infty}^{t} d\tau \; K_j(t-\tau)u_j(\tau) \tag{1.5}$$

Now, we use an exponential kernel equal for all the synapses so that the index j can be dropped in its definition. This kernel assumes the form $K(t) = e^{-t/\tau_I}/\tau_I$, commonly used in firing rate models. By taking the derivative of equation (1.4) with respect to t, we obtain:

$$\begin{aligned} \tau_I \frac{dI_i}{dt} &= \sum_{j=1}^{N} J_{ij} \left(u_j(t) + \int_{-\infty}^{t} d\tau \; (-\frac{1}{\tau_I}) \exp(-\frac{t-\tau}{\tau_I}) u_j(t) \right) \\ &= \sum_{j=1}^{N} J_{ij} \left(u_j(t) - F_j(t) \right) = -I_i(t) + \sum_{j=1}^{N} J_{ij} u_j(t) \end{aligned} \tag{1.6}$$

where the time constant τ_I is the characteristic time of changes in synaptic input current and it is related to the decay of the synaptic conductance. Equation (1.6) is the differential equation describing the evolution of the total synaptic current entering the soma of the i-th neuron, depending on the firing rates of pre-synaptic neuron. We can interpret it from the perspective of Kirchhoff's laws considering neurons as an electrical element: the left-hand side represents the average change in potential; the first term on the right-hand side accounts for the average current leakage through the membranes of the neronal cells and the last term on the right-hand side (interaction term) describes the average current resulting from the activity of all other neurons.

1.3 Firing rate

In order to complete the model we need to determine the post-synaptic firing rate $r_i(t)$ from the knowledge of $I_i(t)$. On general grounds, firing rates do not follow changes in the total synaptic input current instantaneously. Indeed, as previously observed, action potentials are induced by the synaptic current affecting the membrane potential of the neuron. Because of the membrane's capacitance and resistance, its potential essentially acts as a low-pass filtered version of $I_i(t)$. As a result, the time-varying firing rate is often described as a low-pass filtered version of the steady-state firing rate:

$$\tau_R \frac{dv_i}{dt} = -v_i(t) + \phi(I_i(t)) \tag{1.7}$$

where the time constant τ_R determines how rapidly the firing rate converges to its steady-state value and how accurately can track rapid fluctuation for a time-dependent $I_i(t)$. Equivalently, it describes how relevant are the low-pass filtering properties of the membrane potential for the dynamics of the firing rate.

Note that in equation (1.7) we have introduced the activation function ϕ that defines the input-output relationship for each neuron i.e. it transforms currents to

firing rates. In this thesis, we will consider an activation function having a sigmoidal shape i.e. $\phi(\cdot) = \tanh(\cdot)$. The introduction of non-linearity is driven by the biological functioning of neuronal connections. As previously mentioned, neurons communicate via action potentials. However, neural signals are not directly shared but depend on the quantity and type of neurotransmitters released from the axon terminals of pre-synaptic neurons and absorbed by the dendrites of post-synaptic neurons. In a linear neural network (i.e. $\phi(x) = x$) the current in the model neuron is influenced by the pre-synaptic currents of other neurons weighted by the coupling strength. However, the neuron's output is not simply the sum of all inputs as it undergoes a nonlinear operation. For instance, neurons integrate inputs until a certain threshold is reached before an action potential occurs. Moreover, experimental findings indicate the presence of a saturating non-linearity for large inputs. Consequently, it is reasonable to conclude that the output (i.e. firing rate) of a neuron represents a nonlinear computation of the inputs from all connected neurons.

In light of our discussion so far, we can characterize the activity of each neuron in the network using a set of coupled differential equations. Therefore, in general, the ordinary differential equations (ODEs) describing the total synaptic input current to the i-th neuron $I_i(t)$ and its firing rate $r_i(t)$, which we rewrite here for convenience, are:

$$\begin{cases} \tau_R \dfrac{dr_i}{dt} &= -r_i(t) + \phi(I_i(t)) \\ \tau_I \dfrac{dI_i}{dt} &= -I_i(t) + \sum\limits_j J_{ij} r_j(t) \end{cases} \tag{1.8}$$

This model offers quite an accurate description of the dynamical activity exhibited by neurons within the network. However, its complexity renders it less commonly employed in the literature. Furthermore, as our focus is to understand the fundamental features driving network activity across various dynamic phases, we turn to simpler models still able to capture the essential dynamics.

1.4 Quasi-stationary approximations

When of the equations in (1.8) reaches equilibrium significantly faster than the other, the pair can be reduced to a single equation. This reduction simplifies the analysis and allow us to deep understanding of the system's behavior.

The explicit form of the coupled ODEs depends on the relation between the timescale τ_I of the synaptic input current dynamics and the firing rate time constant τ_R [8]. Assuming that $\tau_R \ll \tau_I$, we can employ quasi-stationary approximation of the equation for the firing rate and setting $r_i(t) = \phi(I_i(t))$. This assumption may be valid when the neurons in the network fire at a high frequency so that the low-pass filtering properties of the membrane potential are meaningless. Therefore, the dynamics derives entirely from the input current equation, and the system of equations describing the neural activity are provided by:

$$\begin{cases} r_i(t) & = \phi(I_i(t)) \\ \dfrac{dI_i}{dt} & = -I_i(t) + \sum\limits_j J_{ij} r_j(t) \end{cases} \tag{1.9}$$

where, for the sake of simplicity, we have set τ_I equal to one.

Instead, by assuming that $\tau_R \gg \tau_I$ we can make the quasi-stationary approximation of the equation for the current setting $I_i(t) = \sum_j J_{ij} r_j(t)$. This assumption leads to accurate results when the membrane potential stays below the threshold for long enough periods and its dynamics become relevant for the firing rate of the neuron. Hence, we can describe the activity of the neuron as:

$$\begin{cases} \dfrac{dr_i}{dt} & = -r_i(t) + \phi(I_i(t)) \\ I_i(t) & = \sum\limits_j J_{ij} r_j(t) \end{cases} \tag{1.10}$$

where, for the sake of simplicity, we have set τ_R equal to one.

In what follows, we will refer to the models (1.9) and (1.10) as the current model

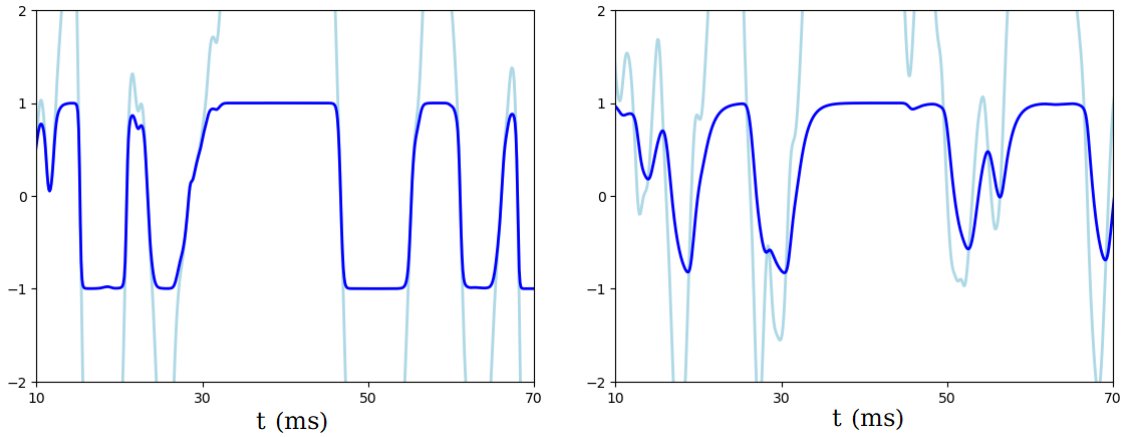


Figure 1.3: Comparison of the activity of a neuron in the network obtained by a direct simulation of the current model (1.9) (left) and the firing model (1.10) (right). In contrast to the firing model, the firing rate of the current model immediately follows changes of the input current, with no dampening or delay. Results obtained by simulating a network of $N = 1000$ neurons, with transfer function $\phi(\cdot) = \tanh(\cdot)$ and synaptic couplings J_{ij} i.i.d. random variables $\sim \mathcal{N}(0, g^2/N)$. The diagonal element of the connectivity matrix $J_{ii} = s = 1.6\ \forall i = 1, \ldots, N$ (self-couplings) and the gain parameter $g = 0.7$ are chosen such that the network exhibit a bimodal activity (see Section 3.4). Firing rate (dark blue) and input current (light blue) are plotted as a function of time.

and firing model, respectively. It is worth noting that, as is often the case with rate models, the biological interpretation of the dynamical variables and the time constant of their dynamics is not always clear in the literature. Our analysis aims to clarify whether these variables or parameters should be interpreted as characterizing neurons or synapses.

1.5 Synaptic connectivity

To complete the characterization of the models, we must introduce the network architecture, namely the synaptic connectivity matrix J. The matrix entries J_{ij} describe the properties of synaptic coupling between pre-synaptic neuron j and post-synaptic neuron i, thereby defining the network's topology.

In this thesis, we consider bidirectional connections between neurons and, in specific cases detailed later, even connections from a neuron to itself (self-couplings). Such networks are commonly referred to as recurrent neural networks (RNNs).

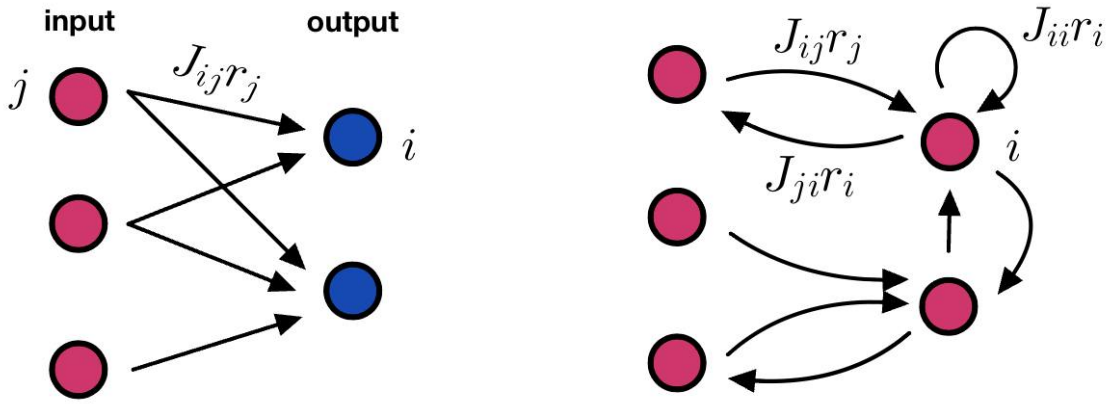


Figure 1.4: Schematic representation of the structures of the feed-forward neural network (left) and the (fully-connected) recurrent neural network (right). Not all the connections of the RNN have been drawn. The feed-forward neural network is organized hierarchically in distinct layers with data flowing from the input layer (through one or more hidden layers) to the output layer. In contrast, the RNN has not such a layered architecture.

Additionally, we assume a fully-connected network, meaning each of the N neurons can receive input from any other neuron and can transmit output to any other neuron. While full connectivity is not entirely biologically realistic, it is not completely unreasonable. Neurons typically receive approximately 10^4 synaptic inputs from other neurons, particularly in the cortex, where there are assemblies of densely interconnected neurons representing basic cortical processing modules. However, on a larger scale, connectivity becomes much sparser and frequently organized into layered structures. As previously discussed, we can address this issue by employing neuron-like units that aggregate neurons with similar properties. These units can be more densely connected and represent the collective activity of a population of neurons by averaging their firing rates.

The dynamical behavior of the network depends on the the connectivity matrix J. Nevertheless, we can distinguish two cases. If the matrix is symmetric (i.e. $J_{ij} = J_{ji}$) we can define an energy function that always decrease (or remains constant) as the system evolves according to its dynamical rule. The attractors are the local minima of the energy and the dynamics may be understood as the motion of a particle moving downhill on the energy surface under the effect of gravity, which pushes it down,

and friction, which prevents it from overshooting and allows it to come to rest at a minimum. Consider, for the sake of clarity, the differential equation in the current model (1.9). The appropiate energy (Lyapunov) function can be defined as [17]:

$$H = -\frac{1}{2}\sum_{ij} J_{ij} r_i r_j + \sum_i \int_0^{r_i} dr\ \phi^{-1}(r) \tag{1.11}$$

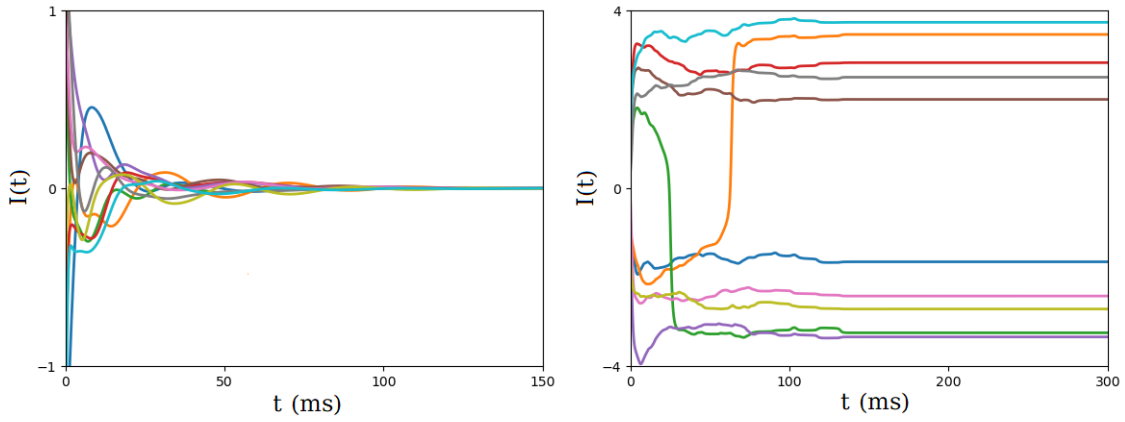
where we remark that $r_i = \phi(I_i)$ and the activation function $\phi(\cdot) = \tanh(\cdot)$. Differentiating H with respect to time (which enters implicitly through r_i), we get:

$$\begin{aligned}
\frac{dH}{dt} &= -\frac{1}{2}\sum_{ij} J_{ij}\frac{dr_i}{dt} r_j - \sum_{ij}\frac{1}{2} J_{ij}\frac{dr_j}{dt} r_i + \sum_i \phi^{-1}(r_i)\frac{dr_i}{dt} \\
&= -\sum_{ij} J_{ij}\frac{dr_i}{dt} r_j + \sum_i \phi^{-1}(r_i)\frac{dr_i}{dt} \\
&= -\sum_i \frac{dr_i}{dt}(\sum_j J_{ij} r_j - I_i) \\
&= -\sum_i \phi'(I_i)(\frac{dI_i}{dt})^2 \leq 0
\end{aligned} \tag{1.12}$$

Here we used the symmetry of J in obtaining the second line and we exploited equation (1.9) for the third line. The result derives from the monotonicity of the activation function and it is zero only if the system is at an equilibrium point, where $dI_i/dt = 0$ for every i. This demonstrate that the fixed points are attractors of the system and there is no other possibility for the long term behavior of the system rather that reaching one of these equilibrium points. Limit cycles, for example, are not allowed since the energy cannot decrease continuously around a closed curve.

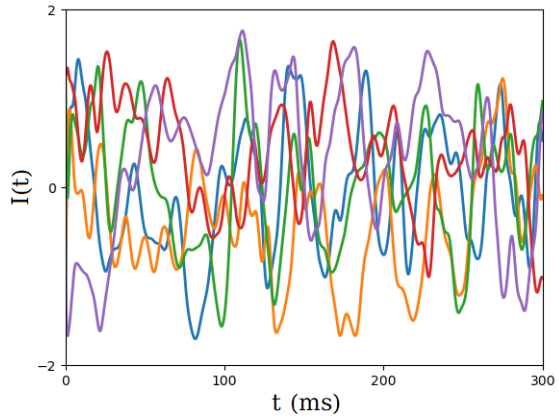
Assuming symmetry in the connectivity matrix is biologically unreasonable. However, by considering an asymmetric matrix ($J_{ij} \neq J_{ji}$), we cannot define an energy function and a richer set of steady-state behaviors emerges. In addition to fixed points, the system can also exhibits limit cycles and chaotic dynamics (fig. 1.5). For these reasons, this study will focus on asymmetric couplings.

Another important aspect to be taken into account when modelling neural connec-

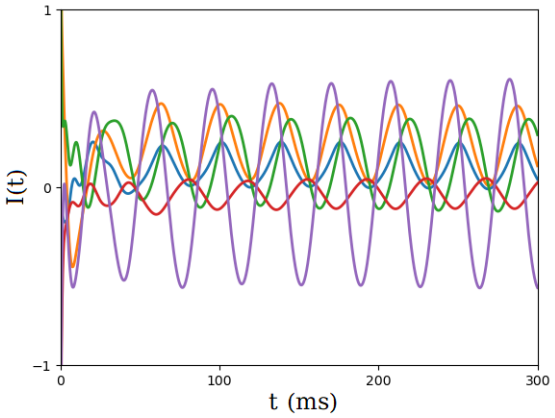


(a) Trivial fixed point

(b) Fixed points



(c) Chaotic



(d) Oscillatory

Figure 1.5: Network dynamics in a non-linear recurrent neural networks obtained from a direct simulation of the current-model (1.9) with $N = 1000$ neurons, transfer function $\phi(\cdot) = \tanh(\cdot)$ and synaptic couplings J_{ij} i.i.d. random variables $\sim \mathcal{N}(0, g^2/N)$ where g is the gain parameter. The diagonal element of the connectivity matrix $J_{ii} = s \ \forall i = 1, \ldots, N$ (self-couplings). With this choice, the connectivity matrix J is thus asymmetric and a rich set of dynamical behavior is observed. The input current $I_i(t)$ of a group of neurons is plotted as a function of time. Depending on the parameters of the model the activity of the network can be stable, chaotic or oscillatory. Network parameters: (a) $g = 0.95$, $s = 0$. (b) $g = 0.6$, $s = 2.5$. (c) $g = 1.5$, $s = 0$. (d) $g = 1.05$, $s = 0$.

tions is that accurately mapping them is an overwhelming, if not impossible, task. The intricacy and the complexity of the synaptic couplings, as well as factors like individual variability and sensitivity to environmental factors, ruled out the possibility of their precise measurement, even with the advancements in high-resolution brain imaging techniques. Moreover, as we are interested in uncovering general principles underlying brain dynamics and functions, we would like that our model would be independent on the particular realization of the connectivity matrix J. A general approach to realize this is the use of random connectivity matrices in neural network models. By employing randomly generated connectivity patterns, researchers can effectively capture the statistical properties of neural connections without the need for precise knowledge of individual ones within a biological network. Random neural networks can exhibit interesting emergent properties, as the presence of stable activity states or phase transitions in network dynamics. Investigating them allows us to shed light on the computational capabilities of biological neural networks and provides insights into the operational mechanisms underlying their functions. In other words, instead of focusing on specific values of the parameter, we study classes of them. From a statistical mechanics point of view, this corresponds to searching for universality classes of networks.

Effectively, we shall consider each element J_{ij} of the connectivity matrix as independent and identically distributed random variable drawn from a Gaussian distribution:

$$J_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \frac{g^2}{N}) \tag{1.13}$$

where the gain parameter g regulates the weight variability of the network or, equivalently, as we consider a zero mean value, the recurrent coupling strength. We remark that the scaling of the variance with N guarantees that the interaction term on the right hand side of the differential equations in (1.9) and (1.10) is of order $\mathcal{O}(1)$ as we take the thermodynamic limit ($N \to \infty$).

The assumption of zero average suggests an absence of a preferred synaptic connection type. However, this can be adjusted by introducing a finite average J_0/N, allowing for the tuning of preferred inhibitory ($J_{ij} < 0$) or excitatory ($J_{ij} > 0$) synaptic connections. Notably, the random neural network we introduced violates Dale's law, which establishes that a neuron can have only excitatory or inhibitory types of connections. However, this issues can be easily overcame by modelling a network in which the connection are independently drawn for two different sub-population of neurons in which are grouped together the exicitatory and the inhibitory ones.

Chapter 2

Dynamical Mean Field Theory

Analyzing the dynamical behavior of recurrent neural networks (RNN) is necessary to understand the neural information processing capabilities. However, the numerical simulation of the dynamics of RNNs, that consist of a large number of interconnected neurons, is computationally challenging. Moreover, our aim is to describe the complex dynamics of the brain into a more manageable framework, using minimal parameters to grasp the essential features underlying brain function. To address this, we turn to Dynamical Mean Field Theory (DMFT), a powerful analytical tool borrowed from statistical physics, to gain insights into the collective behavior of RNNs.

The concept of mean field theory for neural networks draws inspiration from the study of disordered systems, in particular spin glasses [29]. With the influential work of Sompolinsky et al. [38], this theoretical framework has gained particular relevance in the field of neuroscience. Their study, which focused on random network of non-linear rate units, has been further developed by future researches, based on their model to explore different aspects of network dynamics. Many qualitative features of the emerging properties of neuronal networks, such as correlated activity, stability, and response to inputs can be understood within this framework. In particular, large random networks of neuron-like units can display chaotic dynam-

ics, with significant functional consequences. For instance, these networks exhibit optimal information processing capabilities when operating at the edge of chaos, between order and disorder [22].

The DMFT equation reduces the dynamics of a large ensamble of N neurons into an equation of a single neuron embedded into a fluctuating field, the statistics of which are described by a time-lag dependent autocorrelation function that is determined self-consistently. A crucial difference with the standard mean field approximation is that this theory contains fluctuations, that results from a saddle point approximation of an auxiliary field that enters in the partition function of the system after averaging over the quenched disorder introduced by the randomly chosen couplings J_{ij}, as we will see later. Remarkably, this theory is exact in the limit $N \to \infty$ (thermodynamic limit).

The DMFT equation can be formally derived by using a field-theoretical approach called Martin-Sigga-Rose-De Dominicis-Janssen (MSRDJ) formalism which introduces the concepts of action and response field into a path-integral which express the dynamical equations of the system. The MSRDJ formalism bears concise mathematics, resulting in the mean field description of the original high dimensional dynamics. Unfortunately, it is not easy to access the essence of this tool and the underlying physics. However, the same equation can be derived using a more intuitive method, namely the dynamical cavity approach. Essentially, a neuron is added into the system, and its effects on other neurons are self-consistently determined using the linear response approximation. The dynamics of this newly added neuron exhibit similar characteristics to those of the other neurons, thus acting as a representative of the original high-dimensional dynamics.

In this chapter, we will first explore the derivation of the DMFT equations using the dynamical cavity method and then we shift our focus to explain the MSRDJ formalism.

2.1 Generic random neural network

We consider a RNN of N fully-connected neurons with random couplings. The activity of each neuron is described by the synaptic input current, which we will denote for convenience as $x_i(t)$ $(i = 1, \ldots, N)$. As previously discussed, this current obeys the nonlinear dynamical equation in (1.9), i.e.:

$$\frac{dx_i}{dt} = -x_i(t) + \sum_j J_{ij}\phi(x_j(t)) + \xi_i(t) \tag{2.1}$$

A Gaussian white-noise, denoted as $\xi_i(t)$, with zero mean and variance $\langle \xi_i(t)\xi_j(s)\rangle = \sigma^2\delta_{ij}\delta(t-s)$ is introduced in the model as it may represent an external input to each neuron in the network or the inherently stochastic nature of neural connections, such as those observed in cortical circuits. The parameter σ represents the strength of the noise. As mentioned before, the elements J_{ij} of the connection matrix are drawn from a Gaussian distribution with zero mean and variance g^2/N. For generality of explanation, we now also account for pairwise correlation between J_{ij} and J_{ji}. Consequently, the couplings are independent identically distributed Gaussian random variables, except for a potential asymmetric correlation characterized as:

$$\mathbb{E}[J_{ij}J_{ji}] = g^2\frac{\gamma}{N} \tag{2.2}$$

where $\gamma \in [-1, 1]$ describes the degree of asymmetry. More specifically, the wirings are fully symmetric when $\gamma = 1$, allowing for the definition of a Lyapunov function, while they are fully antisymmetric when $\gamma = -1$ and fully asymmetric when $\gamma = 0$. Furthermore, we set $J_{ii} = 0$ so that all the self-interaction are removed.

The system of equation (2.1) may be interpreted as a N-dimensional stochastic process with quenched disorder introduced by the couplings J_{ij} (given that they are time independent) and thermal disorder introduced by the external inputs $\xi_i(t)$.

Note that we denote as $\mathbb{E}[\cdot]$ (or simply $[\cdot]$ in the following) and $\langle\cdot\rangle$ the averages with

respect to the coupling probability distribution $P(J) \overset{\gamma=0}{=} \prod_{ij} P(J_{ij})$ (or quenched-disorder average) and the thermal average (over different noise trajectories), respectively.

2.2 DMFT for uncorrelated couplings

The DMFT equation of this generic model can be readily obtained when the correlation between the couplings is absent (i.e. $\gamma = 0$). The derivation is acquired by relying on the so called local chaos hypothesis initiated by Amari in the field of neural networks [2]. This hypothesis suggests that within a complex system, such as a neural network, there may be localized regions or components, like the neurons, exhibiting chaotic behaviors. Consequently, small variations in initial conditions or parameters can lead to vastly different outcomes over time. This translates into the assumption that the components of the system may be considered uncorrelated from each other and therefore, in a macroscopic (coarse-grained) description of the system, we can rely on simpler models. We remark the strong analogy with the molecular chaos hypothesis in the kinetic theory of gases that leads to the famous Boltzmann equation. Mathematically, it assumes that when N is large the system behaves as if the random variables $\phi(x_i(t))$ (i.e., the firing rates of the neurons) were independent of each other and of the random variables J_{ij} $(\forall\, i, j = 1, \dots, N)$. In this scenario, as we approach the thermodynamic limit $N \to \infty$, the input current received by each neuron through the coupling $\sum_j J_{ij}\phi(x_j(t))$ converges to a Gaussian field according to the central limit theorem. Hence, the basic idea of the mean field description is to replace the network interaction term in (2.1) with a Gaussian random variable η, as:

$$\dot{x_i}(t) = -x_i(t) + \eta_i(t) + \xi_i(t) \tag{2.3}$$

Then, we need to require that the moment of the η match the average over the coupling distribution of the synaptic input. Easily, exploiting the fact that $[J_{ij}] = 0$ and the correlation between the couplings and the firing rate are assumed to be zero thanks to the local chaos hypothesis, the mean of η vanishes:

$$\langle \eta_i(t) \rangle = \sum_j \big[J_{ij} \phi(x_j(t)) \big] = 0 \tag{2.4}$$

In a similar way (recalling that we assume $\gamma = 0$ here), for the second moment of η we have:

$$\begin{aligned} \langle \eta_i(t) \eta_j(s) \rangle &= \big[\sum_l J_{il} \phi(x_l(t)) \sum_k J_{jk} \phi(x_k(s) \big] \\ &= \sum_{l,k} \big[J_{il} J_{jk} \phi(x_l(t)) \phi(x_k(s) \big] \\ &= \delta_{ij} \frac{g^2}{N} \sum_k \big[\phi(x_k(t)) \phi(x_k(s)) \big] \\ &= \delta_{ij} g^2 C(t,s) \end{aligned} \tag{2.5}$$

where we have used the independence of J_{ij} and $[J_{ij}^2] = g^2/N$ in deriving the third line. In the last line we have introduced the (averaged) autocorrelation function of the firing rates $C(t,s)$ which measures the similarity between the state of the system at time t and the state at time s, given by:

$$C(t,s) = \frac{1}{N} \sum_j \big[\phi(x_j(t)) \phi(x_j(s)) \big] \tag{2.6}$$

Given that each neuron becomes statistically identical due to averaging over the coupling probability distribution, the index i becomes irrelevant in the mean field approach. Consequently, we can omit it and express the equations using a generic variable x. Therefore, the complex dynamics of the network can be reduced to an

effective single-neuron dynamics:

$$\dot{x}(t) = -x(t) + \eta(t) \tag{2.7}$$

where we have incorporated also external input $\xi(t)$ in the effective Gaussian noise $\eta(t)$. The temporally correlated variance of the latter is therefore determined by:

$$\langle \eta(t)\eta(s) \rangle = g^2 C(t,s) + \sigma^2 \delta(t-s) \tag{2.8}$$

Finally, we additionally require that the autocorrelation $C(t,s)$ converges to the its path average (with respect to the noise trajectories) in the large N limit, i.e.:

$$C(t,s) = \langle \phi(x(t))\phi(x(s)) \rangle \tag{2.9}$$

thereby closing the self-consistent DMFT equation.

However, the local chaos hypothesis has been found to be inaccurate in models with symmetric couplings [4]. This discrepancy is due to the fact that the mean field equation for the i-th neuron, derived under this assumption, fails to consider the so-called reaction term, which accounts for the influence of the i-th neuron on the others. Essentially, when considering the sum of the afferent currents, a significant correlation between J_{ij} and $\phi(x_j(t))$ arises through J_{ji} when $\gamma \neq 0$, leading to the breakdown of the central limit theorem. The symmetry of the coupling distribution gives rise to feedback effects that introduce a non-Gaussian contribution to the local field distribution. Therefore, from a general point of view, the main limitation of the local chaos hypothesis lies in its neglect of these feedback effects. Nevertheless, this assumption yields accurate results when the couplings are asymmetric, as the feedback effect vanishes in the thermodynamic limit.

In the general scenario, when the correlation between the couplings is present (i.e. $\gamma \neq 0$) to tackle this challenge one can rely on the dynamical cavity approach to

efficiently obtain the mean field description of the system.

2.3 Dynamical cavity approach

The main idea of the dynamical cavity approach, initially proposed by Onsanger for studying spin systems, is to remove a spin from the model (hence the name cavity) and examining how this perturbation affects the dynamical properties of the system [29]. Following a similar principle, our method involves the inclusion of an additional neuron into the model and analyzing how it responds to such a perturbation.
Firstly, we introduce an external field $j_i(t)$ to the original dynamical equation (2.1):

$$\dot{x_i}(t) = -x_i(t) + \sum_j J_{ij}\phi(x_j(t)) + \xi_i(t) + j_i(t) \tag{2.10}$$

that will be necessary to determine the response of the system to a perturbation. Hereafter, we define the abbreviated notation $\phi(x_i(t)) = \phi_i(t)$.
Now, we add a neuron into the original system (denoted by the zero index) which clearly provides a new synaptic current $x_0(t)$ along with the associated synaptic connections J_{0i} and J_{i0} for $i = 1, \ldots, N$. As a consequence, this new neuron influences all the neurons in the original system. Considering this impact as a small perturbation in the limit of a large network ($N \to \infty$), we can employ linear response theory for the firing rates $\tilde{\phi}_i(t)$ in the presence of the added neuron as follows:

$$\begin{aligned}\tilde{\phi}_i(t) &= \phi_i(t) + \sum_k \int_0^t ds\, \frac{\delta\phi_i(t)}{\delta j_k(t)}\bigg|_{j=0} j_k(s) \\ &= \phi_i(t) + \sum_k \int_0^t ds\, R_{ik}(t,s)\,(J_{k0}\,\phi_0(s))\end{aligned} \tag{2.11}$$

where we have defined the linear response function $R_{ik}(t,s) = \frac{\delta\phi_i(t)}{\delta j_k(t)}\big|_{j=0}$ (the partial derivatives should be interpreted in a functional sense) and the small perturbation as $j_k(s) = J_{ko}\,\phi_0(t)$. Subsequently, we can substitute the perturbed firing rates in

the dynamical equation for the input current $x_0(t)$:

$$\begin{aligned}\dot{x_0}(t) &= -x_0(t) + \sum_j J_{0j}\tilde{\phi_j}(t) + \xi_0(t) + j_0(t) \\ &= -x_0(t) + \sum_j J_{0j}\Big(\phi_j(t) + \sum_k \int_0^t ds\ R_{jk}(t,s)\, J_{k0}\, \phi_0(s)\Big) + \xi_0(t) + j_0(t) \\ &= -x_0(t) + \sum_j J_{0j}\phi_j(t) + \int_0^t ds\ \sum_{j,k} J_{0j}R_{jk}(t,s)J_{k0}\, \phi_0(s) + \xi_0(t) + j_0(t)\end{aligned} \tag{2.12}$$

where the third term accounts for how the correlation between the couplings influences the state of the added neuron through the response function.

By taking the thermodynamic limit ($N \to \infty$), we investigate the statistical properties of all terms. The underlying concept is that, by construction, the input currents $x_i(t)$ are independent of J_{i0} and J_{0i} for $i = 1, \ldots, N$, allowing us to employ central-limit-like arguments. Firstly, we isolate the field without the influence of synaptic correlation as follows:

$$\eta_0(t) = \sum_j J_{0j}\phi_j(t) + \xi_0(t) \tag{2.13}$$

which, similarly to what we have seen in the previous section, converges to a centered Gaussian field with variance given by:

$$\langle \eta_0(t)\eta_0(s)\rangle = g^2 C(t,s) + \sigma^2\delta(t-s) \tag{2.14}$$

where $C(t,s)$ is the averaged autocorrelation function of the firing rates.

Now we focus on the third term in equation (2.12) and in particular we want to give a statistical estimate of $\sum_{ij} J_{0i}R_{ij}(t,s)J_{j0}$, which requires some additional attention. We assume that the response functions $R_{ij}(t,s)$ are random variables dependent on the synaptic connections J_{ij} for $i,j = 1,\ldots,N$ but are otherwise independent from J_{0i} and J_{j0}. We first consider the diagonal part $\sum_i J_{0i}R_{ii}(t,s)J_{i0}$ that, according to the central limit theorem, converges to its mean because of negligible variance (of

the order $\mathcal{O}(\frac{1}{N})$):

$$\sum_i J_{0i} R_{ii}(t,s) J_{i0} \xrightarrow{N\to\infty} N\big[J_{0i} R_{ii} J_{i0}\big] = N\big[J_{0i} J_{i0}\big]\big[R_{ii}\big] = g^2\gamma\big[R_{ii}\big] \tag{2.15}$$

Next, we examine the off-diagonal component $\sum_{i\neq j} J_{0i} R_{ij}(t,s) J_{j0}$. Clearly, its mean is zero given that $[J_{0i} J_{j0}] = 0$ for $i \neq j$ and thus we should consider the fluctuation. Let's introduce the random variable Z distributed as to a standard Gaussian distribution (i.e. $Z \sim \mathcal{N}(0,1)$). According to the central limit theorem, we get:

$$\begin{aligned}
\sum_{i\neq j} J_{0i} R_{ij}(t,s) J_{j0} &\xrightarrow{N\to\infty} N(N-1)\big[J_{0i} R_{ij} J_{j0}\big]_{i\neq j} + \sqrt{N(N-1)}\sqrt{\big[J_{0i}^2 R_{ij}^2 J_{j0}^2\big]_{i\neq j}} Z \\
&= N(N-1)\big[J_{0i} J_{j0}\big]_{i\neq j}\big[R_{ij}\big]_{i\neq j} + \sqrt{N(N-1)}\sqrt{\big[J_{0i}^2 J_{j0}^2\big]_{i\neq j}}\sqrt{\big[R_{ij}^2\big]_{i\neq j}} Z \\
&\approx 0 + N\,\frac{g^2}{N}\frac{1}{\sqrt{N}} Z
\end{aligned} \tag{2.16}$$

In deriving the last line we have assumed that $R_{ij}(t,s)$ is of the order $\mathcal{O}(\frac{1}{\sqrt{N}})$ for $i \neq j$. Indeed, in the equilibrium limit, the response function has exactly the same order of magnitude of the correlation function ($\mathcal{O}(\frac{1}{\sqrt{N}})$ in fully-connected mean field models). A proof employing perturbation theory in the strength of the interaction is provided in Roy et al. [34]. Therefore, the contribution of the off-diagonal part can be neglected in the thermodynamic limit.

Now, considering all the contributions, the dynamical equation for $x_0(t)$ can be simplified as follow:

$$\dot{x}_0(t) = -x_0(t) + \eta_0(t) + g^2\gamma \int_0^t ds\, [R_{ii}]\, \phi_0(s) + j_0(t) \tag{2.17}$$

Since the added neuron is not special with respect to the others, its dynamics can be regarded as representative of the typical behavior of the components of the system. Thus, we can drop the subscript 0 and express the mean field dynamical equation

as:

$$\dot{x}(t) = -x(t) + \eta(t) + g^2\gamma \int_0^t R(t,s)\phi(s)ds + j(t) \tag{2.18}$$

where $\eta(t)$ is the effective Gaussian noise with zero mean and correlation $g^2C(t,s) + \sigma^2\delta(t-s)$ (see eq. 2.14), which includes also the external input $\xi(t)$. To close the DMFT equation, we assume that in the large N limit, the correlation $C(t,s)$ and the response function $R(t,s)$ are self-consistently determined by:

$$C(t,s) = \langle \phi(t)\phi(s) \rangle \tag{2.19}$$

$$R(t,s) = \langle \frac{\delta\phi(t)}{\delta j(s)}\Big|_{j=0} \rangle \tag{2.20}$$

and hence they converges to theit path average (with respect to the noise trajectories η and the initials conditions). In summary, we began with an N-dimensional system of differential equations and arrived at a single stochastic differential equation with statistics that are self-consistently determined. It can be demonstrated that in the thermodynamic limit these two descriptions are equivalent, meaning there is a convergence in law between their statistic. It is worth noting that the influence of synaptic correlations among neurons in the N-dimensional system are transformed into an accumulated interaction term of the dynamical history within the one-neuron (mean field) description.

2.4 Path-integral approach

As we have seen, the dynamic of large size random neural network can be analysed relying is heuristic arguments based on Gaussian assumptions regarding the fluctuation of the system in the thermodynamic limit. This approach, however, is unable to give a systematic justification of the underlying assumptions, and extending it to more complex network architectures or dynamics results in a challenging task. Moreover, and most importantly, is not amenable in deriving in general important

properties like, for instance, the stability of the mean field equation and finite-size corrections [21]. In this section, we introduce the generating functional formalism, a systematic method that overcame this limitations. The main idea is to recast an high-dimensional dynamical equation into a path integral framework, which is helpful to reduce the dynamics to a low-dimensional mean field description. Again, the motivation that push us to recast into this general setting is the form of the interaction term in equation (2.1), namely:

$$\sum_j J_{ij}\phi(x_j(t)) := [J\phi(\mathbf{x}(t)]_i \tag{2.21}$$

which involves the sum of many weakly correlated contributions and, as we know, it converges toward a Gaussian random variable the better the larger the N. This phenomenon, known as concentration of measure, describes how the probability distribution becomes sharply focused around its mean value. By considering now the randomness of the couplings J_{ij}, which are referred to frozen or quenched disorder (due to their time independence), we are interested in determine how much the parameters of this Gaussian distribution vary across different realizations of the matrix J. If this variability diminishes as N grows, it suggest that also the variability between different neurons under the same J is small. This property, called self-averaging, allows us to obtain a simplified, low-dimensional description of the statistical behavior of a typical unit. It is implicitly assumed in neural network with random connectivity enabling us to expect that the observables (like firing rate and autocorrelation function) are independent of the particular realization of the connectivity matrix J in the large N limit. Therefore, their expected value can be computed by averaging over different network realizations.

For simplicity, we detailed the procedure to derive the DMFT equation for uncorrelated synaptic connections ($\gamma = 0$), although we remark that the result of the previous section can be equivalently obtained in this general framework.

To start, we recognize that the system of equations in our model, as written in (2.1), lacks rigorous meaning. It takes the form of what is known as a quasi-equation (Van Kampen) being the trajectories of a Wiener process (Brownian motion) never differentiable. Thus we should rewrite it in the form of N coupled stochastic differential equations (Langevin equation) as:

$$d\mathbf{x}(t) = \big(-\mathbf{x}(t) + J\phi(\mathbf{x}(t))\big)dt + \sigma d\mathbf{B}(t) \tag{2.22}$$

where $\mathbf{B}(t)$ is a N-dimensional Brownian motion with independent increments $\mathbf{B}(t) - \mathbf{B}(s) \sim \mathcal{N}(0, \mathbb{I}(t-s))$ where $\mathbb{I}$ denotes the identity matrix. Note that we have introduced an additive noise σ to be consistent with the previously defined system of equations (2.1).

Our objective now is to write the probability associated with the trajectory followed by the stochastic variable $\mathbf{x}(t)$ (input currents) that will be used to construct a moment-generating functional (a generalization of the generating function for random variable), represented as a functional path integral. Through this functional approach, we can readily compute moments of the paths, including correlation and response functions.

Firstly, we consider the dynamics on a discrete time lattice by partitioning the time interval $[0, T]$ into M segments of length Δt (the discrete index $a = 0, 1, \ldots, M$ will be used to indicate the time). For the discretization we follow the Ito convention, which in our case results simply in evaluating the deterministic drift in (2.22) at the beginning of the time interval. We remark that, in the continuum limit ($\Delta t \to 0$), the deterministic contribution is independent on the choice of the discretization procedure. However, attention should be paid when considering multiplicative noise instead of an additive one, as different behavior of the stochastic term are possible. The discretized Langevin equation for the i-th neuron of the network can be written

as:

$$\Delta x_i^a := x_i(t_a) - x_i(t_{a-1}) = f_i^a \delta t + \sigma \Delta B_i^a \tag{2.23}$$

with:

$$f_i^a = -x_i(t_a) + \big[J\phi(\mathbf{x}(t_a))\big]_i \qquad\qquad \Delta B_i^a = B_i(t_a) - B_i(t_{a-1})$$

Now, the joint probability distribution of a Brownian motions (denoted by the i-th index) to visit positions in the intervals (B_i^1, dB_i^1) at time t_1, (B_i^2, dB_i^2) at time t_2, ..., (B_i^M, dB_i^M) at time t_M can be easily obtained by exploiting the Markovian property of the process:

$$\begin{aligned} d\mathbb{P}(B_i^1, \ldots, B_i^M \big| B_i^0) &= W(B_i^M, t_M \big| B_i^{M-1}, t_{M-1}) W(B_i^{M-1}, t_{M-1} \big| B_i^{M-2}, t_{M-2}) \\ &\quad \ldots W(B_i^1, t_1 \big| B_i^0, t_0)\; dB_i^1 dB_i^2 \;\ldots dB_i^M \\ &= \prod_a \frac{dB_i^a}{\sqrt{2\pi\Delta t}} \exp\Big(- \sum_a \frac{(B_i^a - B_i^{a-1})^2}{2\Delta t} \Big) \end{aligned} \tag{2.24}$$

where $W(B_i, t \big| B_i^0, t_0) = \frac{1}{\sqrt{2\pi\Delta t}} \exp\big(- \frac{(B_i - B_i^0)^2}{2\Delta t} \big)$ is the propagator of the Brownian motion and it satisfies the Chapman-Kolmogorov relation that we have used to derive (2.24). The generalization to the multidimensional case is straightforward due to the independence of the N Brownian motions B_i with $i = 0, \ldots, N$, we get:

$$d\mathbb{P}(\mathbf{B}^1, \ldots, \mathbf{B}^M \big| \mathbf{B}^0) = \prod_a \prod_i \frac{dB_i^a}{\sqrt{2\pi\Delta t}} \exp\Big(- \sum_a \frac{(B_i^a - B_i^{a-1})^2}{2\Delta t} \Big) \tag{2.25}$$

As our interest is to have the corresponding probability measure in the $\mathbf{x}$-space, we need to transform the probability given in the last equation. By looking at equation (2.23) we can simply evaluate the Jacobian of the transformation as:

$$\mathcal{J} = \Big| \frac{\partial(x_1^1, \ldots, x_N^M)}{\partial(B_1^1, \ldots, B_N^M)} \Big| = \prod_a \prod_i \sigma \tag{2.26}$$

and therefore we obtain:

$$d\mathbb{P}(\mathbf{x}^1,\dots,\mathbf{x}^M\big|\mathbf{x}^0) = \mathcal{J}^{-1}d\mathbb{P}(\mathbf{B}^1,\dots,\mathbf{B}^M\big|\mathbf{B}^0)$$
$$= \prod_a\prod_i \frac{dx_i^a}{\sqrt{2\pi\sigma^2\Delta t}}\exp\Big(-\sum_a\sum_i \frac{(\Delta x_i^a - f_{i-1}^a\Delta t)^2}{2\sigma^2\Delta t}\Big) \quad (2.27)$$

Now, we move back to the continuous setting by taking the limit $M\to\infty$ and $\Delta t\to 0$ with $\Delta t\, M = T$ fixed. By recognizing that:

$$\sum_a \Delta t \to \int_0^T dt \qquad\qquad \frac{\Delta_i^a}{\Delta_t}\to \partial_t x_i \quad (2.28)$$

and defining:

$$\mathcal{D}\mathbf{x} := \lim_{M\to\infty}\prod_a\prod_i \frac{dx_i^a}{\sqrt{2\pi\sigma^2\Delta t}} \quad (2.29)$$

we can express the probability measure in a formal way as:

$$d\mathbb{P}\big(\{\mathbf{x}(t)\}\big|\mathbf{x}(0)\big) = \mathcal{D}\mathbf{x}\exp\big(-\mathcal{S}_{OM}[\mathbf{x}]\big) \quad (2.30)$$

where:

$$\mathcal{S}_{OM}[\mathbf{x}] := -\frac{1}{2\sigma^2}\int dt\sum_i \big(\partial_t x_i(t) - f_i(\mathbf{x}(t)\big)^2 \quad (2.31)$$

is the so called Onsanger-Machlup action (or stochastic Lagrangian).
Adding the term $\sum_i\int dt\; j_i(t)x_i(t)$ (source field) to the action and integrating over all paths we obtain the moment generating functional $\mathcal{Z}[\mathbf{j}]$:

$$\mathcal{Z}[\mathbf{j}] := \int \mathcal{D}\mathbf{x}\exp\Big(\mathcal{S}_{OM}[\mathbf{x}(t)] + \sum_i\int dt\; j_i(t)x_i(t)\Big) \quad (2.32)$$

Moments of the paths can be formally expressed as functional derivatives with respect to $\mathbf{j}(t)$ of the generating functional. For instance, we have:

$$\big\langle x_i(t)\cdots x_i(s)\big\rangle = \int \mathcal{D}\mathbf{x}\exp\big(\mathcal{S}_{OM}[\mathbf{x}(t)]\big)\, x_i(t)\cdots x_i(s)$$

$$= \frac{\delta}{\delta j_i(t)} \cdots \frac{\delta}{\delta j_i(s)} \mathcal{Z}[\mathbf{j}]\Big|_{\mathbf{j}=0} \tag{2.33}$$

It's crucial to understand that the integral notation is meant symbolically. For concrete computations of the path integral, it's necessary to use the discrete version with finite sums and the finite dimensional probability distribution (2.27). Integration should be carried out using this discrete approach, and the continuous limit has to be taken only afterwards.

Practical computations within the Onsager-Machlup formalism are complicated due to the square in the action (2.31) that introduces analytical challenges, especially when computing averages over the quenched disorder. We can remove this quadratic dependence by performing a Hubbard-Stratonovich transform (Gaussian integral), i.e. by using the identity:

$$\exp\big(-\frac{a}{2}y^2\big) = \frac{1}{i\sqrt{2\pi a}} \int_{-i\infty}^{i\infty} d\tilde{x}\ \exp\big(\frac{\tilde{x}^2}{2a} + \tilde{x}y\big) \tag{2.34}$$

with $\tilde{x}$ an auxiliary field, the so called response field. Employing this identity with $y_i = \frac{1}{\sigma}(\partial_t x_i - f_{i-1})$ leads to the MSRDJ action, that reads:

$$\mathcal{S}[\mathbf{x}, \tilde{\mathbf{x}}] = \tilde{\mathbf{x}}^T(\partial_t + 1)\mathbf{x} - \tilde{\mathbf{x}}^T J\phi(\mathbf{x}) + \frac{\sigma^2}{2}\tilde{\mathbf{x}}^T\tilde{\mathbf{x}} \tag{2.35}$$

where we have used the explicit form of the deterministic drift (f) and we have introduced the notation $\mathbf{x}^T\mathbf{y} = \sum_i \int dt\ x_i(t)y_i(t)$ for denoting the scalar product in time and in neuron space.

We remark that an alternative approach to derive the MSRDJ action involves expressing the joint probability distribution of the path as a product of δ functions enforcing the fulfillment of the stochastic differential equation, averaged over the noise. Representing these δ functions as Fourier integrals with the response fields as integration variables one arrives at the same expression obtained with the Hubbard-Stratonovich transform [16].

Finally, we can write the moment generatic functional in the MSRDJ formalism as:

$$\mathcal{Z}[\mathbf{j}](J) = \int \mathcal{D}\mathbf{x} \int \mathcal{D}\tilde{\mathbf{x}} \, \exp\left(\mathcal{S}_0[\mathbf{x}, \tilde{\mathbf{x}}] - \tilde{\mathbf{x}}^T J \phi(\mathbf{x}) + \mathbf{j}^T \mathbf{x}\right) \tag{2.36}$$

where we have isolated from the action the term that couples the neurons ($\tilde{\mathbf{x}}^T J \phi(\mathbf{x})$) and explicitly expressed the dependence on the interaction matrix J. The measures are defined as:

$$\int \mathcal{D}\mathbf{x} := \lim_{M\to\infty} \prod_a \prod_i \int_{-\infty}^{\infty} dx_i^a \qquad \int \mathcal{D}\tilde{\mathbf{x}} := \lim_{M\to\infty} \prod_a \prod_i \int_{-i\infty}^{i\infty} \frac{d\tilde{x}_i^a}{2\pi i} \tag{2.37}$$

Note the change in the definition of $\mathcal{D}\mathbf{x}$ as a result of the Hubbard-Stratonovich transform. With such a measures, the moment generating functional is properly normalized independently of the realization of J. This property allow us to introduce the disorder-averaged generating functional. Indeed, by assuming that the system shows self-averaging behavior, we expect that the moment generating functional shows concentration of measure as well. Therefore, in the limit of large network, observables that can be computed from $\mathcal{Z}[\mathbf{j}](J)$ can also be (approximately) obtained from its average over the coupling distribution $\langle \mathcal{Z}[\mathbf{j}](J) \rangle_J$. Following the derivation described by Helias [16], we define:

$$\bar{\mathcal{Z}}[\mathbf{j}] := \langle \mathcal{Z}[\mathbf{j}](J) \rangle_J = \left(\frac{N}{2\pi g^2}\right)^{N/2} \int \prod_{i,j} dJ_{ij} \exp\left(-\frac{N}{2g^2} J_{ij}^2\right) \mathcal{Z}[\mathbf{j}](J) \tag{2.38}$$

where we have used that $J_{ij} \sim \mathcal{N}(0, \frac{g^2}{N})$ are independently distributed, and thus the joint distribution over the couplings factorizes into the product of their individual distributions. Now, notice that the only term in the generating functional (2.36) that depends on J_{ij} is the coupling term ($\tilde{\mathbf{x}}^T J \phi(\mathbf{x})$). It appears linearly in the action, allowing for factorization as well ($\exp\left(-\int dt \, \sum_i \tilde{x}_i \sum_j J_{ij} \phi(x_j)\right) = \prod_{i,j} \exp\left(-J_{ij} \int dt \, \tilde{x}_i \phi(x_j)\right)$). Therefore, we can separately integrate over the dif-

ferent J_{ij} identical contributions of the from:

$$\sqrt{\frac{N}{2\pi g^2}}\int dJ_{ij}\ \exp\Big(-\frac{N}{2g^2}J_{ij}^2 - J_{ij}\int dt\ \tilde{x}_i\phi(x_j)\Big) = \exp\Big(\frac{g^2}{2N}\big(\int dt\ \tilde{x}_i\phi(x_i)\big)^2\Big) \tag{2.39}$$

where we have simply completed the square. Regrouping together all the contribution (i.e. product of exponentials as (2.39)) we are left with evaluating:

$$\begin{aligned}\frac{g^2}{2N}\sum_{i,j}\big(\int dt\ \tilde{x}_i(t)\phi(x_j(t))\big)^2 &= \\ &= \frac{g^2}{2N}\sum_{i,j}\int dt\,ds\ \tilde{x}_i(t)\phi(x_j(t))\tilde{x}_i(s)\phi(x_j(s)) \\ &= \frac{1}{2}\int dt\,ds\ \big(\sum_i \tilde{x}_i(t)\tilde{x}_i(s)\big)\big(\frac{g^2}{N}\sum_j \phi(x_j(t))\phi(x_j(s))\big)\end{aligned} \tag{2.40}$$

Here, we have used $\big(\int dt\ f(t)\big)^2 = \int\int dt\,ds\ f(t)f(s)$ in the second line and $\sum_{ij} x_i y_j = \sum_i x_i \sum_j y_j$ in the third. Collecting these observations, we can write the disorder-averaged moment generating functional as:

$$\begin{aligned}\bar{\mathcal{Z}}[\mathbf{j}] = \int \mathcal{D}\mathbf{x}\int \mathcal{D}\tilde{\mathbf{x}}\ \exp\big(\mathcal{S}_0[\mathbf{x},\tilde{\mathbf{x}}] + \mathbf{j}^T\mathbf{x}\big) \\ \times \exp\Big(\frac{1}{2}\int dt\,ds\big(\sum_i \tilde{x}_i(t)\tilde{x}_i(s)\big)\big(\frac{g^2}{N}\sum_j \phi(x_i(t))\phi(x_i(s))\big)\Big)\end{aligned} \tag{2.41}$$

In view of the fact that in the coupling term the sums run over all indices, the system represents a set of N neurons coupled with the others in an identical manner. Thus, as a result of the average over all possible realization of the connection matrix J, all neurons are treated identically. Now, our aim is to decouple the interaction term, which depends on four fields, into the product of pairs of fields. To this end, let's define:

$$Q(t,s) = \frac{g^2}{N}\sum_j \phi(x_j(t))\phi(x_j(s)) \tag{2.42}$$

This field represents an empirical average over N contributions that, in the large N limit and in the case of weak correlation, will converge to a Gaussian random variable by the central limit theorem. This heuristic argument can be made more rigorous: performing a saddle point approximation allows us to replace Q by its (self-consistent) expectation value. To this aim, we firstly enforce the constraint represented by the definition (2.42) by multiplying the disorder averaged moment generating functional by the following identity:

$$\begin{aligned} 1 &= \int \mathcal{D}\mathcal{Q}\, \delta\Big(Q(t,s) - \frac{g^2}{N}\sum_j \phi(x_j(t))\phi(x_j(s))\Big) \\ &= \int \mathcal{D}\mathcal{Q} \int \mathcal{D}\tilde{Q}\ \exp\Big(\int dt\, ds\ \tilde{Q}(t,s)\big(-\frac{N}{g^2}Q(t,s) + \sum_j \phi(x_j(t))\phi(x_j(s))\big)\Big) \end{aligned} \tag{2.43}$$

Here, the conjugate field $\tilde{Q}$ is purely imaginary and the normalization factor $(\frac{N}{2\pi g^2})$ is implicitly absorbed in the integral measure $\mathcal{D}\tilde{Q}$. Now, we consider the problem as a field theory for the auxiliary fields Q and $\tilde{Q}$. Thus, by introducing the notation $Q\tilde{Q}^T = \int dt\, ds\ Q(t,s)\tilde{Q}(t,s)$ and $\tilde{x}^T Q \tilde{x} = \int dt\, ds\ \tilde{x}(t)Q(t,s)\tilde{x}(s)$ we can rewrite the moment generating functional (2.41) as:

$$\bar{\mathcal{Z}}[k] = \int \mathcal{D}Q \int \mathcal{D}\tilde{Q}\ \exp\Big(-\frac{N}{g^2}Q^T\tilde{Q} + N\ \ln\Omega[Q,\tilde{Q}] + k^T Q\Big) \tag{2.44}$$

$$\Omega[Q,\tilde{Q}] := \int \mathcal{D}x \int \mathcal{D}\tilde{x}\ \exp\Big(\mathcal{S}_0[x,\tilde{x}] + \frac{1}{2}\tilde{x}^T Q\, \tilde{x} + \phi(x)^T \tilde{Q}\, \phi(x)\Big)$$

where we have dropped the original source term $\mathbf{j}^T\mathbf{x}$ and introduced the source term k for the auxiliary field Q. Moreover, by looking at the equations (2.41) and (2.43) it is readily apparent that the auxiliary fields depend only on sums of fields ($\sum_i x_i(t)x_i(s)$ and $\sum_j \phi(x_j(t))\phi(x_j(s))$ for Q and $\tilde{Q}$, respectively). So, by writing $N \ln \Omega[Q,\tilde{Q}]$ we have exploited the factorization of the generating functional for the fields $\mathbf{x}$ and $\tilde{\mathbf{x}}$ into the product of N identical factors $\Omega[Q,\tilde{Q}]$. Therefore, we have decoupled the interaction term in (2.41) and reduced the problem of N interacting

units to that of a single unit subject to external fields Q and $\tilde{Q}$.

The N dependence of the action for the auxiliary fields Q and $\tilde{Q}$ in equations (2.44) implies that, in the thermodynamic limit, a saddle point approximation can be employed. This approximation entails disregarding fluctuations in the auxiliary fields and setting them to their expectation value, which corresponds to the point in the $(Q, \tilde{Q})$ space that yields the most significant contribution to the probability mass. In the saddle point approximation we search for the stationary point of the action by requiring that the two following conditions are attained:

$$0 = \frac{\delta \mathcal{S}[Q, \tilde{Q}]}{\delta\{Q, \tilde{Q}\}} = \frac{\delta}{\delta\{Q, \tilde{Q}\}} \Big(- \frac{N}{g^2} Q^T \tilde{Q} + N \ln \Omega[Q, \tilde{Q}] \Big) \tag{2.45}$$

from which we obtain a pair of equations:

$$\begin{aligned} 0 &= -\frac{N}{g^2} Q^*(t,s) + \frac{N}{\Omega} \frac{\delta \Omega[Q, \tilde{Q}]}{\delta \tilde{Q}(t,s)} \Big|_{(Q^*, \tilde{Q}^*)} \\ &\iff Q^*(t,s) = g^2 \langle \phi(x(t)\phi(x(s)) \rangle := g^2 C(t,s) \end{aligned} \tag{2.46}$$

$$\begin{aligned} 0 &= -\frac{N}{g^2} \tilde{Q}^*(t,s) + \frac{N}{\Omega} \frac{\delta \Omega[Q, \tilde{Q}]}{\delta Q(t,s)} \Big|_{(Q^*, \tilde{Q}^*)} \\ &\iff \tilde{Q}^*(t,s) = \frac{g^2}{2} \langle \tilde{x}(t) \tilde{x}(s) \rangle \equiv 0 \end{aligned} \tag{2.47}$$

where $C(t,s)$ is the autocorrelation function of the firing rates. We remark that the average $\langle \cdot \rangle$ has to be intended over all the paths of the dynamical process (different realization of x) evaluated at the saddle points $(Q^*, 0)$. It has to be computed self-consistently because the values of the saddle points, by equation (2.44), affect the statistics of the fields $\mathbf{x}$ and $\tilde{\mathbf{x}}$, which, in turn, determine the function $\mathcal{Q}^*$ by (2.46). Finally, inserting back the saddle point solutions in the generating functional (2.44), we can write:

$$\bar{\mathcal{Z}}^* \propto \int \mathcal{D}x \int \mathcal{D}\tilde{x} \exp \Big(\mathcal{S}_0[x, \tilde{x}] + \frac{g^2}{N} \tilde{x}^T C \tilde{x} \Big) \tag{2.48}$$

Hence, the replacement of Q in (2.41) with the saddle point Q^* (which depend only on the statistic of x, and not on its particular realization) allow us to decouple the system into non-interacting neurons driven by a common field with self consistently determined statistic. It is easy to prove that the second term in (2.48) represent a Gaussian field with correlation function $g^2C(t,s)$ that has to be interpreted as the input each neuron receive from the other units.

Thus, in the thermodynamic limit, the problem can be described by the effective dynamics of a single unit:

$$\dot{x}(t) = -x(t) + \eta(t) \tag{2.49}$$

where $\eta(t)$ is the effective Gaussian noise of the process (that incorporates also the external input) with autocorrelation $\langle \eta(t)\eta(s) \rangle = g^2C(t,s) + \sigma^2\delta(t-s)$. Note that this is the same result that we have derived in the previous section with heurist arguments.

Chapter 3

Chaotic activity in RNN

The brain processes information through the dynamics of large neural networks. Interestingly, cortical circuits exhibit intricate temporal patterns of spiking and display a high sensitivity to even weak disturbances in their ongoing activity, as the displacement of a single spike may have a significant impact on the timing of subsequent ones. These characteristics, coupled with analyses of single-cell recordings and electroencephalography data, suggest a chaotic dynamic within these circuits. Theoretical investigations have shown that chaos provides neural circuits with exceptional computational capabilities, influencing how incoming information is processed and received. Models of local cortical circuits that exhibit a chaotic dynamics can flexibly adapt to changes in external input and to explore a wide range of states encoding forthcoming stimuli, such as limit cycles and fixed points. Leveraging recent neurophysiological discoveries, the study of chaos in neural networks offers a great opportunity for understanding cognitive processes, motivating the appearance of spontaneous irregular patterns of activity in neural assemblies and the development of new technologies that take advantage of neural network parallel computing capacity and chaos control. From a biological point of view, intricate models have been used to simulate temporal phenomena in the brain, including oscillation synchronization for feature linking and the transitions between coherent states and chaos.

However, these models are complicated to analyse mathematically and have lacked a theoretical underground explaining network behavior in terms of a few crucial control parameters. One possibility to overcame these issues is to develop a model that, while potentially oversimplified in comparison to biological circuits, still captures essential features. The benefit of such a model is the analytical and computational control, providing accurate analysis of the time series of activity patterns.
In this section we will investigate the stationary (long time) behavior of a recurrent neural network with asymmetric coupling whose dynamics obey the input current model. Despite its simplicity, this model displays an astonishing variety of dynamical behaviors and a relatively deep mathematical analysis can be performed combining concepts and methods from dynamical systems theory, statistical physics and random matrix theory. As previously mentioned, the symmetry of synaptic connections ensures the existence of a Lyapunov function and, consequentely, a convergent dynamics. Instead, neural network with asymmetric synaptic weights exhibit a complex dynamics with a variety of dynamical regimes. In particular, in the limit of large network there is a sharp transition from a stationary to a chaotic state as a function of the gain parameter g.

3.1 Mean field and autocorrelation function

In the previous chapter, we have shown that in the thermodynamic limit a RNN with randomly coupled rate neurons behaves statistically identical to a model of an effective single neuron dynamics driven by a Gaussian noise with self consistently determined statistics. In what follows, we will conduct an analytical examination of the autocorrelation function of the network. To start, we consider an autonomous system (i.e. without the external input ξ that drives the network) and with no self-couplings s_i (the diagonal of the connection matrix J is set to zero). In the limit

$N \to \infty$, the effective mean field dynamics of this network is given by:

$$\dot{x}(t) = -x(t) + \eta(t) \tag{3.1}$$

where the autocorrelation of the Gaussian noise η is determined by $\langle \eta(t)\eta(s) \rangle = g^2 C(t,s)$. Following [38], our aim is to represent the autocorrelation function of the system as the motion of a particle in a self consistent potential. To see this, we firstly apply the Fourier transform on both sides of equation (3.1). Thus, we get:

$$(1 + iw)\hat{x}(w) = \hat{\eta}(w) \tag{3.2}$$

$$(1 - iw)\hat{x}(-w) = \hat{\eta}(-w) \tag{3.3}$$

where we have denoted as $\hat{f}(w)$ is the Fourier transform of a function $f(t)$ and $\hat{f}(-w)$ is the conjugated quantity of $\hat{f}(t)$. By multyplying both sides of equations (3.2) and (3.3), we obtain:

$$(1 + w^2)\hat{x}(w)\hat{x}(-w) = \hat{\eta}(w)\hat{\eta}(-w) \tag{3.4}$$

Now, perfroming an inverse Fourier transform of the right hand side of equation (3.4), we have:

$$\begin{aligned}
\frac{1}{2\pi}\int dw\; e^{iw\tau}\hat{\eta}(w)\hat{\eta}(-w) &= \frac{1}{2\pi}\int dw\; e^{iw\tau}\int dt\; \eta(t)e^{-iwt}\int ds\; \eta(s)e^{iws} \\
&= \frac{1}{2\pi}\int\int dt\, ds\; \eta(t)\eta(s)\int dw\; e^{iw(s+\tau-t)} \\
&= \int\int dt\, ds\; \eta(t)\eta(s)\delta(t-s+\tau) \\
&= \langle \eta(t)\eta(t+\tau) \rangle
\end{aligned} \tag{3.5}$$

A similar inverse Fourier transform to the left hand side reads:

$$\frac{1}{2\pi}\int dw\; (1 + w^2)\hat{x}(w)\hat{x}(-w)e^{iw\tau}$$

$$
\begin{aligned}
&= \frac{1}{2\pi} \int dw \, \big(1 - (iw)^2\big) \hat{x}(w) \hat{x}(-w) e^{iw\tau} \\
&= \big(1 - \frac{d^2}{d\tau^2}\big) \Delta(\tau)
\end{aligned} \tag{3.6}
$$

where we have defined the local field autocorrelation function $\Delta(\tau)$ which measures the similarity between the synaptic input current at time t and after a temporal separation of τ, i.e.:

$$
\Delta(\tau) = \langle x(t) x(t+\tau) \rangle \tag{3.7}
$$

It's worth noting that, given our focus on the stationary statistics of the system, the autocorrelation depends only on the time lag τ, i.e. $\Delta(t, s) = \Delta(\tau)$ where clearly $t - s = \tau$. In other words, in the long time limit, the dynamics is time-translation invariant. Now, collecting the results of equations (3.5) and (3.6) and exploiting our knowledge of the noise statistics, we obtain:

$$
\ddot{\Delta}(\tau) = \Delta(\tau) - g^2 C(\tau) \tag{3.8}
$$

where we have used the fact that the autocorrelation of the firing rates $C(t, s)$ is also time-traslation invariant and thus $C(t+\tau, t)$ is only a function of τ. We remark that we could have obtained this equation alternatively by multiplying equation (3.1) for time points t and s, taking the expectation with respect to the noise η on both sides, and then exploiting the time-translation invariance of the dynamics in the long time limit [16].

By solving the motion equation (3.8) governing the dynamics of $\Delta(\tau)$, we can determine the autocorrelation $C(\tau)$ of the firing rates of neurons separated by a time interval τ, depeding on the initial condition determined by the variance Δ_0. To solve equation (3.8), we need to express $C(\tau)$ as a function of $\Delta(\tau)$. We start by recognizing that both of these functions depends on $x(t)$. Furthermore, as demonstrated in the preceding chapter, the variable $x(t)$ can be effectively modeled as a Gaussian

random variable. This approximation stems from the fact that the expression in (2.48) corresponds to the moment generating functional of a Gaussian theory. Alternatively, one can heuristically justify this approximation by invoking the central limit theorem. The statistics of $x(t)$ are given by:

$$\langle x(t)\rangle = \langle x(t+\tau)\rangle = 0 \tag{3.9}$$

$$\langle x(t)x(t+\tau)\rangle = \Delta(\tau) \tag{3.10}$$

Now, we can exploit the following parametrization of the synaptic input current:

$$x(t) = \alpha y + \beta z \tag{3.11}$$

$$x(t+\tau) = \alpha y' + \beta z \tag{3.12}$$

where y, y' and z are independent Gaussian random variable with zero mean and unit variance and the coefficients α and β need to be determined in agreement to the relations (3.11) and (3.12). Indeed, we have:

$$0 = \langle x(t)\rangle = \langle \alpha y + \beta z\rangle \implies \langle y\rangle = \langle z\rangle = 0 \tag{3.13}$$

$$0 = \langle x(t+\tau)\rangle = \langle \alpha y' + \beta z\rangle \implies \langle y'\rangle = 0 \tag{3.14}$$

Therefore:

$$\Delta(0) := \Delta_0 = \langle(\alpha y + \beta z)^2\rangle = \alpha^2 + \beta^2 \implies \alpha = \sqrt{\Delta_0 - |\Delta(\tau)|} \tag{3.15}$$

$$\Delta(\tau) = \langle(\alpha y + \beta z)(\alpha y' + \beta z')\rangle \implies \beta = \sqrt{|\Delta(\tau)|} \tag{3.16}$$

In terms of the random Gaussian variables, we can express $C(\tau)$ as a function of $\Delta(0)$ and $\Delta(\tau)$ as:

$$C(\tau) = \int Dz\, Dy\, Dy'\, \phi(\alpha y + \beta z)\phi(\alpha y' + \beta z) = \int Dz \left(\int Dy\, \phi(\alpha y + \beta z)\right)^2$$

$$= \int Dz \Big(\int Dy \; \phi(\sqrt{\Delta_0 - |\Delta|}y + \sqrt{|\Delta|}z) \Big)^2 \tag{3.17}$$

where we have denoted as Dz the standard Gaussian measure, i.e.:

$$Dz = \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{3.18}$$

likewise for Dy and Dy'. Substituting the expression for $C(\tau)$ into equation (3.8) we get a closed system of equations for determining $\Delta(\tau)$ and $C(\tau)$. However, the latter expression for $C(\tau)$ is valid only in the case in which $\Delta(\tau) > 0$. Actually, a generalization of the equation (3.17) valid also for negative $\Delta(\tau)$ can be easily obtain by simple algebric manipulation, and reads:

$$C(\tau) = f_\phi(\Delta(\tau), \Delta_0) \tag{3.19}$$

where the function f_u is given by:

$$f_u(\Delta, \Delta_0) = \int Dz_1 \, Dz_2 \; u\Big(\sqrt{\Delta_0 - \frac{\Delta}{\Delta_0}} z_1 + \frac{\Delta}{\sqrt{\Delta_0}} z_2\Big) u\big(\sqrt{\Delta_0} z_2\big) \tag{3.20}$$

with z_1 and z_2 standard Gaussian random variables. This expression will be advantageous for the subsequent development and has been utilized in the numerical simulations.

3.2 Regimes of the network dynamics

Given the obtained result, it is worth to revisit equation (3.8). Conceptually viewing τ as a distinct time rather than merely a time lag, the mathematical structure of the equation resembles that governing the motion of a classical particle with unit mass subject to a force derived from the right-hand side. However, a slight complication arises. In classical Newtonian mechanics, Δ_0 would represent the particle's initial position and, in this context, the dependence of $C(\tau)$ on Δ_0 as indicated in equation

(3.19) suggests a subtle mechanical scenario. Here, the force acting on the particle not only depends on its instantaneous position ($\Delta(\tau)$) but also on its initial state (Δ_0). Nevertheless, it is useful to define a potential V that takes the form:

$$V(\Delta, \Delta_0) := -\frac{\Delta^2}{2} + g^2 f_\Phi(\Delta, \Delta_0) - g^2 f_\Phi(0, \Delta_0) \tag{3.21}$$

where $\Phi(\cdot) = \ln\cosh(x)$ is the integral of the transfer function $\phi(\cdot) = \tanh(\cdot)$ and the last term is an arbirtary constant that ensures that $V(0, \Delta_0) = 0$. Therefore, in terms of the potential V the equation of motion (3.8) takes the form:

$$\ddot{\Delta}(\tau) = -\frac{\partial V(\Delta, \Delta_0)}{\partial \Delta} \tag{3.22}$$

We emphasize that in the derivation of the potential, we employed the Price's theorem (i.e. the property $\frac{\partial}{\partial \Delta} f_\Phi(\Delta, \Delta_0) = f_{\Phi'}(\Delta, \Delta_0) = f_\phi(\Delta, \Delta_0)$) and that the dependence of $f_u(\Delta, \Delta_0)$ on τ arises only through $\Delta(\tau)$. The equation has to be solved self-consistently, as the initial position of the particle Δ_0 determines the shape of the effective potential $V(\cdot, \Delta_0)$ through equation (3.21).
Nevertheless, we can gain analytical insight. First of all, being a correlation function, $\Delta(\tau)$ must satisfy two physical constraint:

- $\Delta(\tau)$ is bounded i.e. $\Delta(0) \geq |\Delta(\tau)|$ and $\Delta(0) \geq 0$

- $\Delta(\tau)$ is a differentiable and even function i.e. $\Delta(\tau) = \Delta(-\tau)$ due to the time-traslation invariance in the long time limit. Hence, the initial kinetic energy must be zero i.e. $\dot{\Delta}(0) = 0$.

Now, if we are able to characterize the form of the potential $V(\Delta, \Delta_0)$ for any given value of g and initial condition Δ_0, we are can describe the evolution of $\Delta(\tau)$ over time (or lag), utilizing equation (3.22) along with the imposed physical constraints. Thus, to estimate its shape, we examine its derivatives, which can be

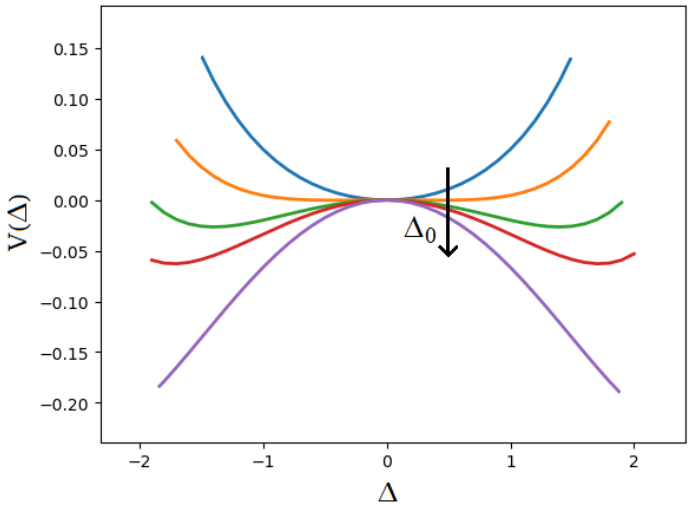


Figure 3.1: Self-consistent potential $V(\Delta, \Delta_0)$ as a function of the autocorrelation of the input current $\Delta(\tau)$ (see eq. (3.21)) for gain parameter $g = 2$ at increasing values of the variance $\Delta_0 = 1.6, 1.8, 1.924, 2, 2.2$ (following the black arrow). We observe that the shape of the potential changes from a single-well (blue) to a double-well with the height of the barrier that increases as Δ_0 is increased. The green potential is the one derived for the value of Δ_0 self-consistently determined by imposing $V(\Delta_0, \Delta_0) = 0$ (see main text). Results obtained from the DMFT of the current model (1.9) with transfer function $\phi(\cdot) = \tanh(\cdot)$ and i.i.d synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$, with $J_{ii} = 0 \; \forall i = 1, \dots, N$.

readily evaluated using Price's theorem, yielding:

$$\frac{\partial V(\Delta, \Delta_0)}{\partial \Delta} = -\Delta + g^2 f_\phi(\Delta, \Delta_0) \tag{3.23}$$

$$\frac{\partial^2 V(\Delta, \Delta_0)}{\partial \Delta^2} = -1 + g^2 f_{\phi'}(\Delta, \Delta_0) \tag{3.24}$$

Firstly, it's evident that $\frac{\partial V}{\partial \Delta}\big|\Delta = 0 = 0$ holds for an odd transfer function ($\phi(\cdot) = \tanh(\cdot)$). Additionally, considering that $0 < \phi' \leq 1$, the integral denoted by $f_{\phi'}(\Delta, \Delta_0)$ (see the expression (3.20)) has to be smaller than one, implying $\frac{\partial^2 V}{\partial \Delta} \leq g^2 - 1$. Consequently, if $g < 1$, $\frac{\partial^2 V}{\partial \Delta}$ will always be negative, and the potential $V(\Delta, \Delta_0)$ strictly concave. Conversely, for $g > 1$, two scenarios arise, as the second derivative depends on Δ_0. By analyzing $\frac{\partial^2 V}{\partial \Delta}\big|_{\Delta=0} = -1 + g^2\big(\int Dz \; \phi'(\sqrt{\Delta_0})\big)^2$, we observe that the potential exhibits either a positive or negative curvature for small and large values of Δ_0. In the intermediate regime, the potential assumes a double-well shape.

Therefore, the characteristic of the network dynamics can be classified depending on the shape of the potential into the following types:

- Concave potential (for $g < 1$): the only bounded solution (respecting the physical constraints) is the vanishing solution $\Delta(\tau) = \Delta(0) = 0$. This implies that the steady state of the network flows to the zero fixed point solution ($x_i = 0 \; \forall i$) for (almost) all the initial conditions.

- Convex potential (for $g > 1$ and small Δ_0): the trajectory of $\Delta(\tau)$ oscillates from Δ_0 to $-\Delta_0$ indicating a limit-cycle solution.

- Double well potential (for $g > 1$ and relatively large Δ_0): a lot of solutions exist in this case (static, oscillating and decaying solutions) depending on the initial value of Δ. However, the only stable solution is the one characterized by a decay to 0 of the autocorrelation for $\tau \to \infty$, that exhibit a positive Lyapunov exponent (chaotic solution) [19]. It's worth noting that, consistently with numerical simulations, finite-size networks exhibit oscillatory solutions near above the critical value of the gain parameter ($g = 1$). However, as the network size increases, these oscillations diminish, indicating a sharper transition to chaos with increasing N.

In order to determine the shape of the autocorrelation function $\Delta(\tau)$ we need to solve equation (3.22). The analogy with a mechanical system involving a particle in a potential well suggests a reciprocal conversion between kinetic and potential energy, ensuring the conservation of the particle's total energy. Let's denote with E this energy, we have:

$$E = \frac{\dot{\Delta}}{2} + V(\Delta, \Delta_0) = \frac{\dot{\Delta}}{2} - \frac{1}{2}\Delta^2 + g^2 f_\Phi(\Delta, \Delta_0) - g^2 f_\Phi(0, \Delta_0) \tag{3.25}$$

Now, at $\tau = 0$, we have seen that $\dot{\Delta} = 0$. Therefore, the initial energy is given by:

$$\begin{aligned} E = V(\Delta_0, \Delta_0) &= -\frac{\Delta_0^2}{2} + g^2 f_\Phi(\Delta_0, \Delta_0) - g^2 f_\Phi(0, \Delta_0) \\ &= -\frac{\Delta_0^2}{2} + g^2 \int Dz \left(\Phi(\sqrt{\Delta_0} z)\right)^2 - g^2 \Big(\int Dz \, \Phi(\sqrt{\Delta_0} z)\Big)^2 \end{aligned} \tag{3.26}$$

where in deriving the second line we have used the explicit expression of $f_\Phi(\cdot, \Delta_0)$ (eq. (3.20). By the argument of energy conservation, solutions to equation (3.22) must satisfy equation (3.25). Thus, substituting the value of E with the previous expression (3.26), it is evident that one solution is given by $\Delta(\tau) = \Delta(0) = 0 \; \forall \tau \geq 0$ (trivial fixed point solution of the network dynamics). To obtain the other solution, we impose the conditions $\Delta(\tau) \to 0$ and $\Delta(\tau) \to 0$ as $\tau \to \infty$, which are typical for an autocorrelation function measured for a chaotic state of the network. Thus, in the limit $\tau \to \infty$, we can write:

$$E \to g^2 f_\Phi(0, \Delta_0) - g^2 f_\Phi(0, \Delta_0) = 0 \tag{3.27}$$

and substituting again this value in equation (3.26), we get:

$$V(\Delta_0, \Delta_0) = 0 \tag{3.28}$$

The value of the initial position of the particle (Δ_0) as a function of the gain parameter g can be numerically obtain as the root of equation (3.28) (for example by using a bisectioning algorithm). Finally, the corresponding shape of the autocorrelation function $\Delta(\tau)$ can be readily derived by integrating equation (3.22). This task can be accomplished numerically by reformulating the second order differential equation as a coupled set of first order equations, given by:

$$\begin{cases} \dot{y}(\tau) & = \Delta - g^2 f_\phi(\Delta, \Delta_0) \\ \dot{\Delta}(\tau) & = y(\tau) \end{cases} \tag{3.29}$$

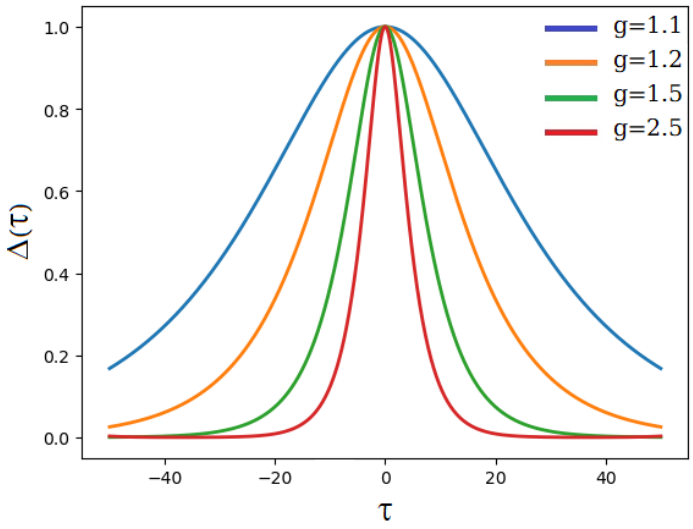


Figure 3.2: (Normalized) self-consistent solution of the autocorrelation function of the input current $\Delta(\tau)$ for different value of the gain parameter g as a function of the time lag τ in the chaotic regime of the network. The result is derived by solving (3.28) for the initial value Δ_0 and then integrating (3.29). We observe that the decay time of the autocorrelation function decreases as g is increased, indicating that the network activity becomes more and more chaotic. Results obtained from the DMFT of the current model (1.9) with transfer function $\phi(\cdot) = \tanh(\cdot)$ and i.i.d synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$, with $J_{ii} = 0 \; \forall i = 1, \ldots, N$.

with initial conditions:

$$\begin{cases} y(0) &= 0 \\ \Delta(0) &= \Delta_0 \end{cases} \tag{3.30}$$

The solution of this equation is shown in figure 3.2 for different values of the gain paramenter g. The decay of the autocorrelation function $\Delta(\tau)$ for $g > 1$ is indicative of a chaotic flow. Indeed, the activity of the neurons in the network (i.e. the synaptic input current) becomes uncorrelated as the temporal separation (τ) is increased and the network lose memory of its initial state. Notably, the decay is faster as the gain parameter is increased. It is worth noting that the autocorrelation of the firing rates $C(\tau)$, easily obtained by means of the relation (3.19), bares strong similarities with $\Delta(\tau)$. As illustrated in figure 3.3, $C(\tau)$ closely matches the results obtained from direct numerical simulations. The agreement improves with increasing size of the

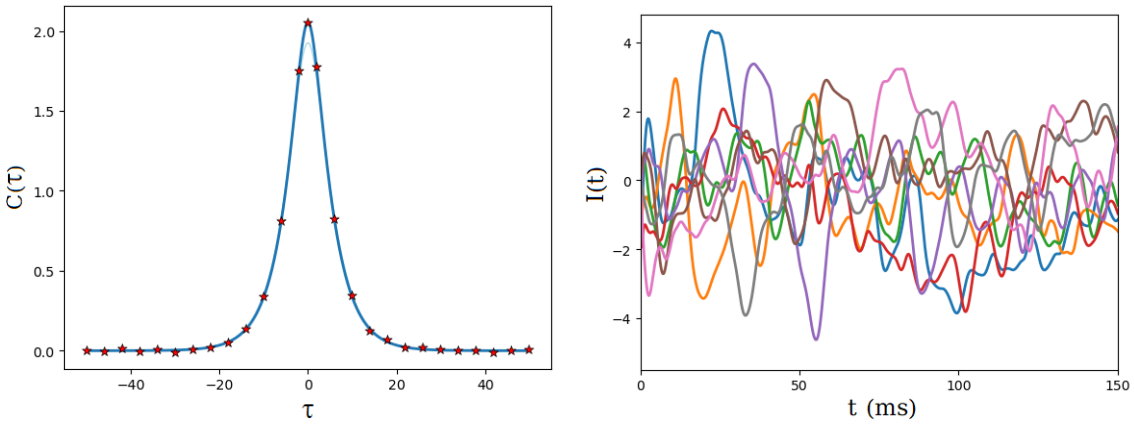


Figure 3.3: Left: autocorrelation function of the firing rate $C(\tau)$ (eq. 3.19) as a function of the time lag τ for gain parameter $g = 2$ (chaotic activity) multiplied by a scaling factor g^2 (blue) and the result obtained from a direct numerical simulation of a network with $N = 5000$ neurons (red stars). The theoretical result is derived by substituting in (3.19) the self consistent solution for $\Delta(\tau)$ (light-blue). Right: example of network activity for the same value of the gain parameter. Results obtained for the current model (1.9) with transfer function $\phi(\cdot) = \tanh(\cdot)$ and i.i.d synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$, with $J_{ii} = 0 \; \forall i = 1, \ldots, N$.

simulated network, confirming the validity of the mean field description as $N \to \infty$.

3.3 Largest Lyapunov exponent

To investigate the stability of the network, we examine how small perturbations evolve during the dynamics. Mathematically, we evaluate the Largest Lyapunov Exponent (LLE) of the system, which allows us determine the conditions for transitioning from a silent to a chaotic regime. The LLE measures the sensitivity of the dynamics to initial conditions: a positive LLE indicates that nearby trajectories (i.e. starting from nearly identical initial conditions) diverge exponentially over time. On the other hand, if the LLE is negative, trajectories with infinitesimally close initial conditions converge and the dynamics is stable.

The maximum Lyapunov exponent can be derived in the framework of Dynamical Mean Field Theory by considering two copies (replicas) of the network with identical coupling matrix J. In such a case, the LLE can be defined as the asymptotic

growth rate of the Euclidean distance between trajectories of the two replicas [16]. However, we consider here a slightly simpler approach. Our aim is to evaluate the time evolution of an infinitesimal perturbation $\delta x_i(t)$ from its original trajectory $x_i(t)$. To this end, we add the fluctuation $\delta x_i(t)$ to the input current model (eq. (1.9)), we get:

$$\frac{dx_i(t)}{dt} + \frac{d\delta x_i(t)}{dt} = -\big(x_i(t) + \delta x_i(t)\big) + \sum_j J_{ij}\big(\phi(x_j(t)) + \phi'(x_j(t))\delta x_j(t)\big) \quad (3.31)$$

where in the last term we have used the first order Taylor expansion of the transfer function $\phi(x_i(t) + \delta x_i(t))$. Subtracting the evolution equation of $x_i(t)$ we obtain the equation describing the dynamics of the perturbation, i.e:

$$\big(\partial_t + 1\big)\delta x_i(t) = \sum_j J_{ij}\phi'(x_j(t))\delta x_j(t) \quad (3.32)$$

In order to understand the evolution of the deviation over time, we analyse its autocorrelation function. To this aim, we make a time translation of τ to equation (3.32) i.e. considering the time point $s = t - \tau$, we can write:

$$\big(\partial_s + 1\big)\delta x_k(s) = \sum_l J_{kl}\phi'(x_l(s))\delta x_l(s) \quad (3.33)$$

Multiplying side by side equations (3.32) and (3.32), we get:

$$\big(\partial_t + 1\big)\big(\partial_s + 1\big)\delta x_i(t)\delta x_k(s) = \sum_{jl} J_{ij}J_{kl}\phi'(x_j(t))\phi'(x_l(s))\delta x_j(t)\delta x_l(s) \quad (3.34)$$

Now, averaging the above equation over the coupling probability distribution $P(J)$ and performing a similar reasoning to the one that allowed us to derive the DMFT equation in Section 2.2, we can write:

$$\big(\partial_t + 1\big)\big(\partial_s + 1\big)\Delta_\delta(t,s) = g^2 C_{\phi'}(t,s)\Delta_\delta(t,s) \quad (3.35)$$

where we have defined the autocorrelations:

$$C_{\phi'}(t,s) = \langle \phi'(x(t))\phi'(x(s)) \rangle \tag{3.36}$$

$$\Delta_{\delta}(t,s) = \langle \delta x(t)\delta x(s) \rangle \tag{3.37}$$

Note that we have dropped the neuronal indices (i, k) since averaging over J makes all the neuron statistically identical. Moreover, it is worth mentioning that in this scenario the autocorrelation function of the perturbations $\Delta_{\delta}(t,\tau)$ depends on time (and not only on the time lag τ), as it is relative to the time when the initial perturbation causing the deviation occurred (e.g. $t = 0$). Conversely, the autocorrelation $C_{\phi'}(t,s)$ is time translation invariant and in the stationary limit it can be exclusively represented as a function of the time lag τ. Similarly to what we have done with the autocorrelation $C(\tau)$ (see eq. (3.19)), we recognize that $C_{\phi'}(\tau)$ can be expressed as a function of $\Delta(\tau)$. Mathematically, this corresponds to writing $C_{\phi'}(\tau) = f_{\phi'}(\Delta(\tau), \Delta_0) = \frac{\partial C(\tau)}{\partial \Delta(\tau)}$ (the second equality being another instance of the Price theorem). Moreover, it is advantageous to represent $C_{\phi'}(\tau)$ in terms of the potential $V(\Delta(\tau), \Delta_0)$. By recalling equation (3.24), we can state:

$$g^2 C_{\phi'}(\tau) = \frac{\partial^2 V(\Delta, \Delta_0)}{\partial \Delta^2} + 1 \tag{3.38}$$

To proceed with the analytical developments, we can introduce the coordinate transformation $T = t + s$ and $\tau = t - s$ on the left-hand side of (3.35). Using the chain rule of differentiation i.e. $\partial_t f(T,\tau) = \partial_T f(T,\tau)\frac{\partial T}{\partial t} + \partial_\tau f(T,\tau)\frac{\partial \tau}{\partial t}$ (and similarly for the derivative with respect to s), we get:

$$\partial_t \to \partial_T + \partial_\tau \qquad\qquad \partial_s \to \partial_T - \partial_\tau \tag{3.39}$$

Therefore, the differential operator $\big(\partial_t + 1\big)\big(\partial_s + 1\big)$ on the left hand side of equation (3.35) can be reformulated in the new coordinates into the form $\big((1 + \partial_T)^2 - \partial_\tau^2\big)$.

Finally, collecting all the results, we can write the equation of motion for the autocorrelation of the deviation $\Delta_\delta(T,\tau)$ as:

$$\big((1+\partial_T)^2-\partial_\tau^2\big)\Delta_\delta(T,\tau)=\Big(\frac{\partial^2 V(\Delta,\Delta_0)}{\partial\Delta^2}+1\Big)\Delta_\delta(T,\tau) \tag{3.40}$$

Next, we are going to evaluate the LLE of the network analysing the growth of the perturbation dynamics. Firstly, we note that if $|\delta\mathbf{x}(t)|\sim|\delta\mathbf{x}(0)|e^{\boldsymbol{\lambda}t}$ and the largest exponent $\lambda=\max(\boldsymbol{\lambda})$ is positive, the divergence between the original trajectory and the trajectory influenced by the infinitesimal initial perturbation will be magnified, resulting in a chaotic state. Assuming that the perturbation at $t=0$ has unit norm (i.e. $\|\delta\mathbf{x}(0)\|=1$) the LLE can be computed as:

$$\begin{aligned}\lambda&=\lim_{t\to\infty}\frac{1}{t}\log\Big(\frac{\|\delta\mathbf{x}(t)\|}{\|\delta\mathbf{x}(0)\|}\Big)=\lim_{t\to\infty}\frac{1}{2t}\log\Big(\sum_i\big(\delta x_i(t)\big)^2\Big)\\&=\lim_{t\to\infty}\frac{1}{2t}\log\big(N\Delta_\delta(T,\tau=0)\big)=\lim_{t\to\infty}\frac{1}{2t}\log\big(\Delta_\delta(T,\tau=0)\big)\end{aligned} \tag{3.41}$$

where in deriving the second line we have assumed that the population-averaged square distance of the perturbation converges to the (equal-time) autocorrelation $\Delta_\delta(T,0)$ (i.e. $\frac{1}{N}\sum_i\big(\delta x_i(t)\big)^2\to\Delta_\delta(T,0)$). Considering the equation of motion for the autocorrelation of the perturbation (3.40) and the Lyapunov Exponents (3.41), we further assume a time-separation ansatz to express the former as:

$$\Delta_\delta(T,\tau)=e^{\frac{k}{2}T}\psi(\tau) \tag{3.42}$$

and therefore $\lambda=k/2$. Substituting this expression into equation (3.40), we obtain the eigenvalue equation:

$$\Big(-\partial_\tau^2-\frac{\partial^2 V(\Delta,\Delta_0)}{\partial\Delta^2}\Big)\psi(\tau)=\Big(1-\big(1+\frac{k}{2}\big)^2\Big)\psi(\tau) \tag{3.43}$$

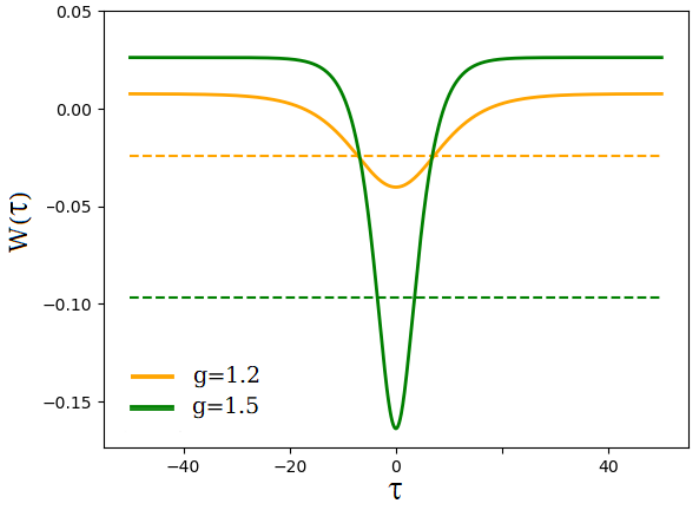


Figure 3.4: Quantum mechanical potential of the Schrödinger equation (3.43) as a function of the time lag τ for different values of the gain parameter g (solid lines) and the respective ground state energy E_0 (dashed lines) from which the LLE is computed as in (3.45). Results obtained from the DMFT of the current model (1.9) with transfer function $\phi(\cdot) = \tanh(\cdot)$ and i.i.d synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$, with $J_{ii} = 0 \; \forall i = 1, \ldots, N$.

which we recognize as the one-dimensional time-independent Schrödinger equation (with τ interpreted as the spatial coordinate). Drawing upon this analogy, $W(\tau) = -\frac{\partial^2 V(\Delta,\Delta_0)}{\partial \Delta^2}$ represents the quantum potential (plotted in figure 3.4) for different values of the gain parameter g), while $\left(1-\left(1+\frac{k}{2}\right)^2\right)$ is the energy E. The eigenvalues (or energies) E_n of the Schrödinger operator determines the exponential growth rate k_n of the perturbation ($\Delta_\delta(2t, 0) = e^{k_n t}\psi_n(0)$), and they are given by:

$$k_n^{\pm} = 2\big(-1 \pm \sqrt{1 - E_n}\big) \tag{3.44}$$

The fastest growing mode is clearly given by the ground state energy E_0 and choosing the plus sign in equation (3.44). Therefore, the LLE of the network can be estimated as:

$$\lambda = \frac{k_0^+}{2} = -1 + \sqrt{1 - E_0} \tag{3.45}$$

It is worth noting that in the case of silent activity of the network (zero fixed point solution), the quantum potential is constant ($W(\tau) = E_0 = 1 - g^2$) as can be readily

obtained by substituting the solution $\Delta(\tau) = \Delta(0) = 0$ in its definition. Thus, the trivial fixed point solution looses its stability as the critical value of the coupling strength ($g = 1$) is exceeded, and it is replaced by a chaotic state. Indeed, by looking at equation (3.45), when the lowest energy E_0 transitions to a negative value, the Lyapunov Exponent become positive, leading to the emergence of a chaotic solution, which is highly sensitive to small variations in the initial conditions.

The LLE obtained from the mean field theory described above as a function of the coupling strength g are depicted in figure 3.5, highlighting the remarkable agreement with a direct numerical simulation using an Orbit Separation (OS) algorithm (see Chapter 5). Moreover, we confirm the theoretically predicted linear growth for $g < 1$ (substituting $E_0 = 1 - g^2$ in (3.45) we get $\lambda = g - 1$). In the limits $g \to 1^+$ and $g \to \infty$ the leading behavior of $\lambda(g)$ becomes quadratic and logarithmic, respectively. This prediction can be theoretically proven by expanding the quantum potential in these limit [6].

We are now able explore the differences in the dynamical behavior between the two rate models (1.9) and (1.10): the input current model and the firing model, respectively. A comparison of the Lyapunov Exponents (LLE) as a function of the gain parameter g for the two models is depicted in figure 3.6. Surprisingly, both models exhibit identical behavior (or an equivalent degree of chaoticity), despite the distinct role of the non-linear transfer function in each case. To delve into the origin of this analogy, we numerically simulated the complete model (1.8), from which both the current and firing models emerge through different quasi-stationary approximations. Remarkably, we observed the same chaotic behavior for the LLE in the complete model, suggesting that the other two models derive their chaotic dynamics from the full model, making them equivalent in this sense.

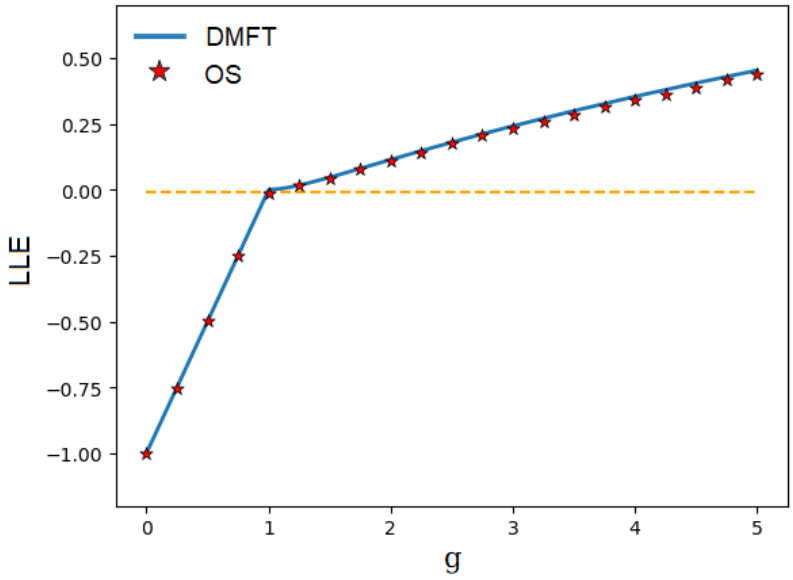


Figure 3.5: LLE of the network as a function of the gain parameter g. The theoretical (DMFT) prediction derived by computing the lowest eigenvalue (E_0) of the Schrödinger operator according to equation (3.45) (blue) is compared to a direct numerical simulation acquired through the Orbit Separation method (OS) of the network with $N = 5000$ neurons and 10000 time steps (red stars) (see Chapter 5). Results obtained for the current model (1.9) with transfer function $\phi(\cdot) = \tanh(\cdot)$ and i.i.d synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$, with $J_{ii} = 0 \; \forall i = 1, \ldots, N$.

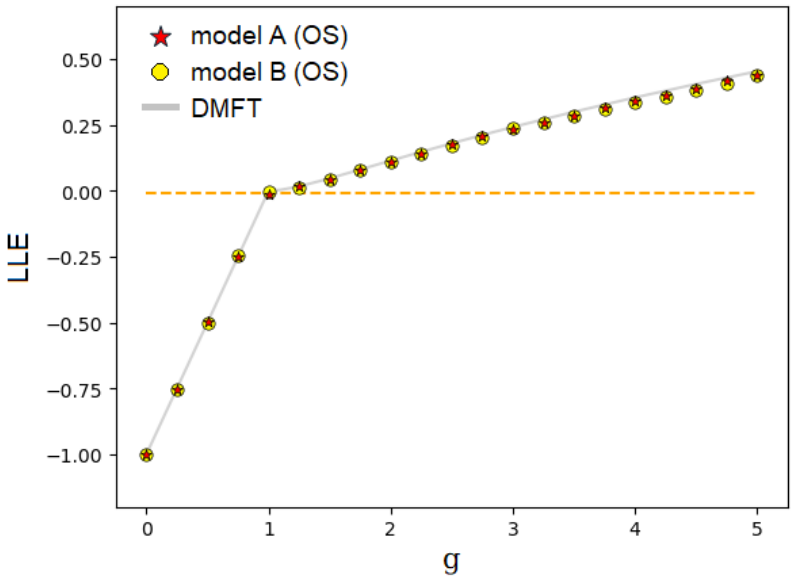


Figure 3.6: LLE as a function of the gain parameter g obtained from a direct numerical simulation with the Orbit Separation method (same parameters of figure 3.5) of the input current model (1.9) (red stars) and the firing rate model (1.10) (yellow circle). The theoretical DMFT prediction (eq. (3.45)) is presented for comparison (grey). Results derived for both models with transfer function $\phi(\cdot) = \tanh(\cdot)$ and i.i.d synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$, with $J_{ii} = 0 \; \forall i = 1, \ldots, N$.

3.4 Self-couplings

To further characterize the potential dynamical behaviors of RNNs, we introduce the possibility for the connectivity matrix J to contain non-zero diagonal elements (self-couplings s_i). As previously noted, firing rate models often group together neurons with similar properties into clusters, with the collective activity of these neurons described by the output of a single unit. From this perspective, interactions among neurons within a cluster are captured by self-couplings, i.e., feedback connections from a unit to itself, while interactions between clusters are represented by connections between units. RNNs with heterogeneous self-couplings offer a simple explanation for the hierarchy of timescales of the neural activity (typically quantified by the decay time of the autocorrelation function) as observed in cortical circuits [40]. In this section, we consider a RNN with homogeneous self-couplings (i.e. $s_i = s \ \ \forall\, i = 1, \ldots, N$), obeying the equation (1.9):

$$\frac{dx_i}{dt} = -x_i(t) + s\phi(x_i(t)) + \sum_j J_{ij}\phi(x_j(t)) \tag{3.46}$$

where we have explicitly written the diagonal element of the connectivity matrix J. Note that in this manner, the matrix J is the same as before, i.e. with $J_{ii} = 0$ and the off-diagonal elements are independent and identically distributed Gaussian random variables with zero mean and variance g^2/N.

By varying the parameters g and s, a rich set of dynamical behaviors of the network emerges: a zero fixed point state, a chaotic state, and a transient chaotic activity that converges to a non-trivial fixed point. The phase diagram in parameter space, along with examples of the network dynamics, is depicted in figure 3.7. Following [41], an intuitive understanding of how these different types of activity emerge can be provided by considering two special cases. We have already seen that when $s = 0$, the network exhibits chaotic activity above the critical coupling $g = 1$ and silent activity when it is below. Now, let's consider a specific case where the gain parameter

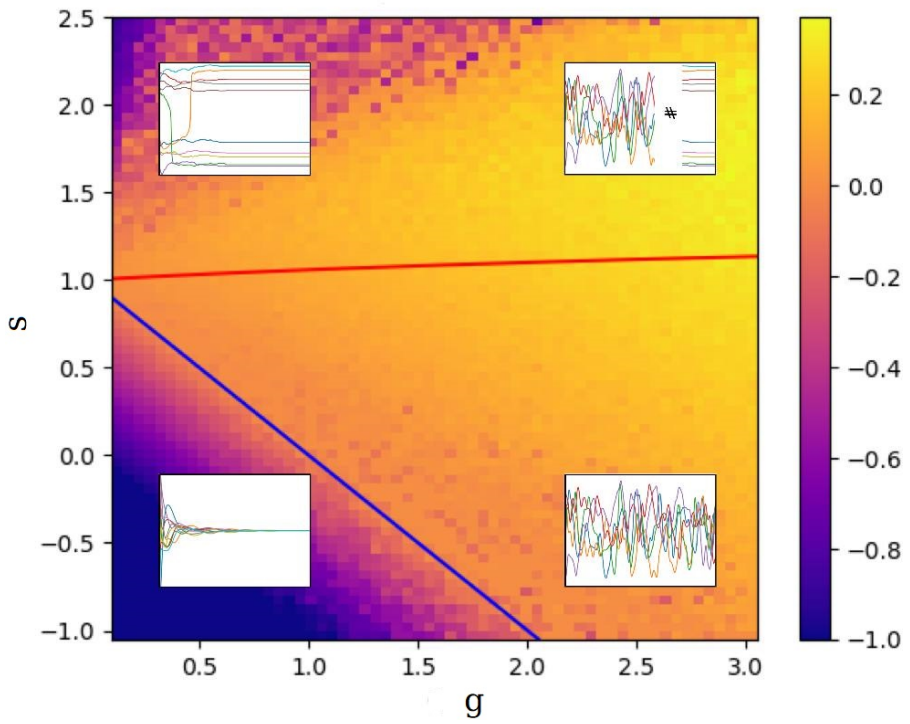


Figure 3.7: Network phase diagram obtained by a direct numerical simulation of the LLE (with Orbit Separation method) as a function of the gain parameter g and the self-coupling s. The phase space in divided into three regions: below the blue line the activity of the network decays to the zero fixed point solution. Above the red curve the network exhibit transient chaotic activity that eventually settles into a non zero stable fixed point solution (increasing the value of g results in a longer time required for convergence to a fixed point). Between the two curves, the activity is chaotic. Examples of network dynamics are shown in the insets. Results obtained from the DMFT of the current model (1.9) with transfer function $\phi(\cdot) = \tanh(\cdot)$ and i.i.d synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$, with $J_{ii} = s \; \forall i = 1, \ldots, N$. Figure reproduced from Sompolinsky et al. [41].

vanishes ($g = 0$). In this scenario, the units become uncoupled and, at stationarity, the solution to equation (3.46) is determined by $x_i = s\phi(x_i)$. When $s < 1$, the only stable solution is $x_i = 0$, leading to the decay of all unit activity to zero from any initial state. Conversely, when $s > 1$, two stable solutions exist (while the zero solution is unstable) with equal magnitude and opposite signs. This implies that the units exhibit bistability, and the network dynamics converge to a non-zero fixed point solution. When both s and g are non zero, an interesting interplay between chaos and bistability emerges.

In order to obtain the critical condition that separates a stable to a chaotic region in the phase space, we can analyse how the different fixed point solutions are destabilized by changing the parameters of the network. To describe the general approach, we focus on simple example of the network in the trivial fixed point state. We already know that $x_i = 0 \;\; \forall\, i = 1, \ldots, N$ is a stationary solution of the system of equation (3.46). To analyse its stability, we compute the Jacobian matrix D of the system and study its eigenvalues. Let's call F_i the right hand side of equation (3.46), we get:

$$D_{ij} = \frac{\partial F_i}{\partial x_j} = \big(-1 + s\phi'(x_i)\big)\delta_{ij} + J_{ij}\phi'(x_j) \tag{3.47}$$

where δ_{ij} is the Kronecker delta and $\phi'(\cdot) = \text{sech}^2(\cdot)$. Evaluating the Jacobian at the trivial fixed point solution we obtain:

$$D_{ij} = (-1 + s)\delta_{ij} + J_{ij} \tag{3.48}$$

since $\text{sech}^2(0) = 1$. Now, in order to determine the phase boundary that separates silent activity from chaotic regime, we can exploit the (generalized) circular law from Random Matrix Theory [37, 31] by employing the statistical properties of the connectivity matrix J. The eigenvalues of the stability matrix D are uniformly distributed as a disk in the complex plane with center in $(-1 + s)$ in the x-axis and radius g. The real part of the eigenvalues needs to be negative in order to ensure the

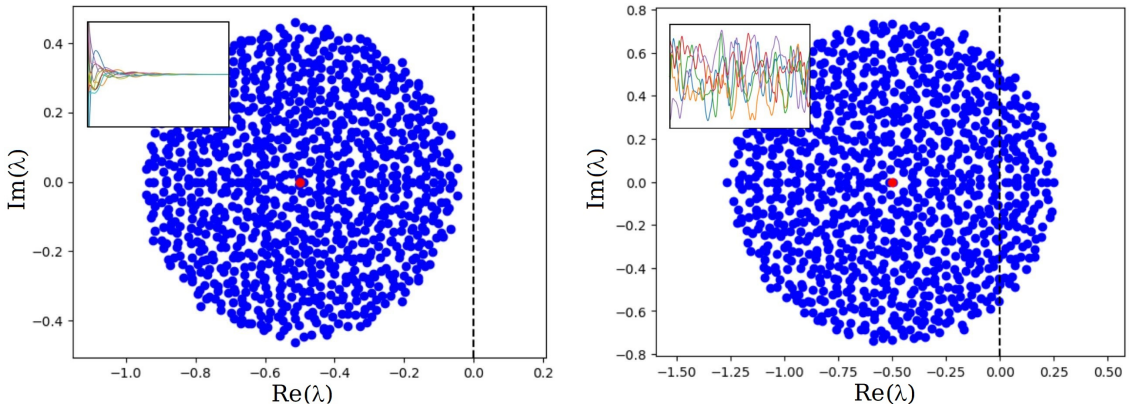


Figure 3.8: Real (x-axis) and imaginary (y-axis) parts of the eigenvalues of the stability matrix D for the trivial fixed point solution (eq. (3.48)) obtained from a direct numerical simulation of a network with N=1000 neurons and $s = 0.5$. The gain parameter is $g = 0.45$ (left) and $g = 0.75$ (right). The eigenvalues (blue dots) are distributed as a disk in the complex plane with center in $(s-1)$ in the x-axis (red dot) and radius g. As the real part of the eigenvalues grows above zero (dotted line) the fixed point solution loses stability and is replaced by a chaotic solution. Examples of the network activity are shown in the insets. Results obtained for the current model (1.9) with transfer function $\phi(\cdot) = \tanh(\cdot)$ and i.i.d synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$, with $J_{ii} = s \; \forall i = 1, \ldots, N$.

stability of the zero fixed-point solution, a condition that is satisfied by the simple inequality $(-1 + s) + g < 0$. Consequently, the phase boundary is given by the condition $s = 1 - g$, that is represented by the blue line in figure 3.7. For $s > 1 - g$, the eigenvalues exhibit a real positive part (see figure 3.8), and the zero fixed point solution loses its stability and it is replaced by a chaotic activity. It's worth noting that in the scenario of vanishing self-coupling ($s = 0$), we retrieve the previously determined critical gain parameter value of $g = 1$ that governs the transition.

When dealing with non-trivial fixed points, computing the eigenvalue distribution of the stability matrix (3.48) is not as straightforward as in the case of trivial fixed points. A detailed derivation strategy using Random Matrix Theory is outlined by Ahmadian et al. [1], while Stern et al. [41] have also investigated the results. The boundary separating chaotic and non-zero fixed points corresponds to the red line in figure 3.7.

To further understand the nature of chaotic activity, we simulated the network in regions where both transient and persistent chaotic behavior occur. In these

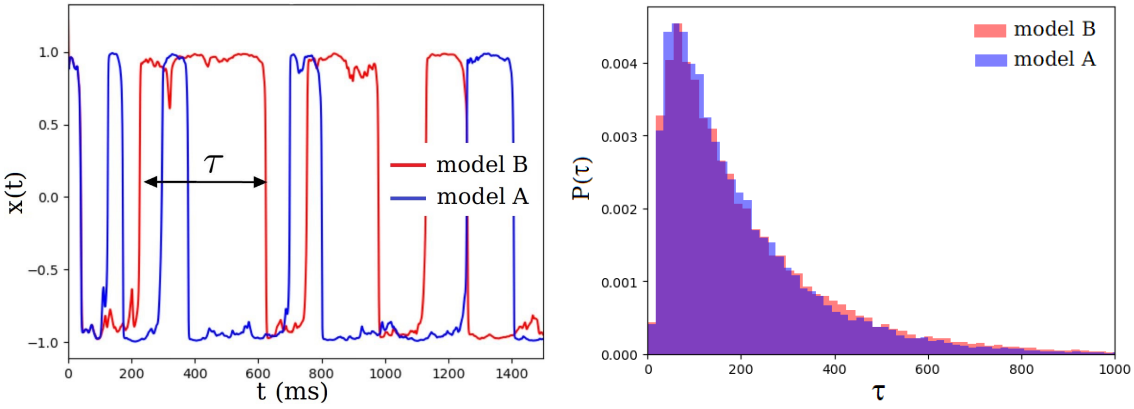


Figure 3.9: Bimodal activity neurons in the transient chaotic region (parameters of the simulation: $s = 1.6$, $g = 0.7$, $N = 1000$). Left: firing rate a neuron in the network for the input current model (1.9) (blue) and the firing rate model (1.10) (red). Right: the characteristic time in a state (τ) follows a log-normal distribution with average $\langle\tau\rangle = 198.6\,\text{ms}$ for the input current model (blue bars) and $\langle\tau\rangle = 194.5\,\text{ms}$ for the firing rate model (red bars). Results obtained from a direct simulation of both models with transfer function $\phi(\cdot) = \tanh(\cdot)$ and i.i.d synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$, with $J_{ii} = s\ \forall i = 1, \ldots, N$.

regions, the firing rate of neurons exhibits a bimodal pattern, beginning in the persistent chaotic region and becoming more pronounced where fixed points exist. Bimodality is characterized by intermittent switches between states that fluctuate near the values $\phi(x_i) = \pm 1$ (figure 3.9). The average time between these switches, τ, follows a log-normal distribution. Interestingly, the firing rate model (1.10) exhibits the same behavior, suggesting again an equivalence between the two models.

Chapter 4

External input to a RNN

In the previous chapter we have studied the convergence of the dynamics of RNNs in absence of a signal. Autonomous systems indeed can exhibit a diverse range of steady-state behaviors, including the interesting emergence of chaotic states. On general grounds, neural activity results from the interplay between spontaneous firing within neural circuits and responses to external stimuli. In this context, Freeman et al. conducted a notable experiment on the olfactory bulb of rabbits [12, 13]. They proposed that spontaneous activity in the cortical circuit might exhibit chaos. However, they observed that recognizing a previously learned smell led to a temporary reduction in chaotic activity. In other words, retrieving a previously stored pattern was interpreted as the reduction of the chaotic (strange) attractor to an attractor of lower dimension (fixed point solutions). In this scenario, during the alert waiting state, the network explores a wide range of its phase space through chaotic dynamics. Then, when a stimulus is introduced, the dynamics shift to a lower-dimensional attractor that was formed during the learning process. Building on these concepts, our objective is to explore how a neural network that inherently produces chaotic activity patterns can maintain sensitivity to external inputs and what are the implications for its dynamics and information processing capabilities.

4.1 Baseline control of RNN

To explore these ideas, let's once again turn to a simplified model of biologically plausible cortical circuits. As usual, we consider a fully-connected RNN with randomly distributed synaptic connections between neurons. Additionally, each neuron in the network receives an external input ξ_i, drawn from a Gaussian distribution with mean μ and variance σ^2. In this setting, the discrete-time version of the input current model (1.3) reads:

$$x_i(t+1) = \sum_j J_{ij}\phi(x_j(t)) + \xi_i \tag{4.1}$$

where we've employed a simple Euler discretization of the continuous-time equation for the input current. Note that the external input ξ_i does not explicitly depend on time, as we treat the input received by each neuron as quenched (time-independent). This can be understood as follows: if we consider the RNN as a model of the cortical circuit, the external input in this context can be interpreted as a synaptic current originating from areas like the thalamus and other subcortical regions (e.g. locus coeruleus), which send a widespread signal to cortical areas and modulate their dynamics, as shown by intracellular recordings [30, 27]. In this scenario, changes in the mean and variance of the input ξ_i reflects alterations in this baseline signal which represent changes in behavioral states or other contextual modulations [42, 23]. For instance, baseline modulation can represent arousal or movements that enhance the visual and gustatory capabilities of the sensory circuits responsible for processing information from stimuli [25, 20]. Because the timescale of behavioral modulation is typically much slower than the intrinsic timescale of cortical circuit dynamics, we can approximate the effect of baseline modulation as quenched inputs.

The advantage of working in a discrete-time setting lies in the ease of estimating the Largest Lyapunov Exponent (LLE) of the network. In this context, simple self-consistent equations for the order parameters of the model (mean m and variance

q of the single neuronal activity) can be rigorously derived using Dynamic Mean Field Theory, allowing for a straightforward calculation of the LLE. An heuristic argument (similar to the one used in Section 2.2) to obtain those equations could be outlined as follow. Averaging equation (4.1) with respect to the quenched disorder (joint probability distribution for the couplings and the baseline input), we get:

$$\langle x_i(t+1)\rangle = \sum_j \langle J_{ij}\phi(x_j(t))\rangle + \mu \tag{4.2}$$

Assuming that in the thermodynamic limit we can neglect the correlation between the random variables $x_i(t)$ from each other and from the random variables J_{ij} (local chaos hypothesis), we obtain $\langle x(t+1)\rangle = J_0\langle\phi(x(t))\rangle + \mu$. Note that, as previously observed in Chapter 2, we have dropped the neuronal indices since averaging over J makes all the neuron statistically identical. Moreover, we allow the synaptic connection to have a non vanishing mean J_0 (i.e. $J_{ij} \sim \mathcal{N}(\frac{J_0}{N}, \frac{g^2}{N})$). In the stationary regime, the distributions of $x_i(t+1)$ and $x_i(t)$ are identical, and according to the central limit theorem, they follow a Gaussian distribution with mean m and variance q. Therefore, we can write:

$$m = \mu + J_0 \int Dx\ \phi\big(\sqrt{q}x + m\big) \tag{4.3}$$

where we denoted with Dx the standard Gaussian measure (i.e. $Dx = \frac{dx}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$). Taking the second order moment of equation (4.1) and applying the same assumptions, we get that the variance of the synaptic input current distribution can be expressed as:

$$q = \sigma^2 + g^2 \int Dx\ \phi\big(\sqrt{q}x + m\big)^2 \tag{4.4}$$

Solutions to the self-consistent equations (4.3) and (4.4) can be easily obtained numerically by employing a bisectioning algorithm (root-finding).

With these expressions for the mean and variance, the Lyapunov exponent can

then be easily obtained by considering the evolution of the mean quadratic distance between two trajectories (replicas) of the system in the phase space. In this setting, the LLE is defined as:

$$\lambda = \lim_{t\to\infty} \lim_{|x^1(0)-x^2(0)|\to 0} \frac{1}{2t} \log \frac{\left(x^1(t) - x^2(t)\right)^2}{\left(x^1(0) - x^2(0)\right)^2} \tag{4.5}$$

where $x^1(t)$ and $x_2((t)$ are two initially arbitrary closed trajectories (orbit). In the thermodynamic limit, the LLE can be estimated as [9]:

$$\lambda = \frac{1}{2} \ln \left(g^2 \int Dx\ \phi'(\sqrt{q}x + m)^2 \right) \tag{4.6}$$

where ϕ' denotes the derivative of the transfer function and m and q are the self-consistent solutions to equations (4.3) and (4.4). Compared to the continuous setting (see Section 3.3), the above formula can be easily implemented in numerical simulations, providing us with a faster computational approach for investigating the chaotic behavior of the network and define its the phase diagram.

To understand how the external input ξ_i affects the phase boundary between the network's fixed point and chaotic solutions, we numerically computed the LLE by solving equation (4.6) for different values of the mean μ of the external input distribution and vanishing quenched variance (fig. 4.1). In contrast to the autonomous case (i.e. without external inputs), we observed an increase in the critical coupling strength g that leads to the phase transition. Specifically, in the autonomous system, g is critical at 1, while with an external input, the critical value of g increases. Moreover, as the mean μ of this external input increases, so does the critical value. In general, external input drives the network response, leading to a suppression of ongoing activity and ultimately eliminating chaos through a phase transition [32]. This behavior can be explained through a mean field approach, by analyzing the interaction between the quenched baseline and the recurrent synaptic inputs with the single neuron transfer function. Chaotic behavior often occurs when a substan-

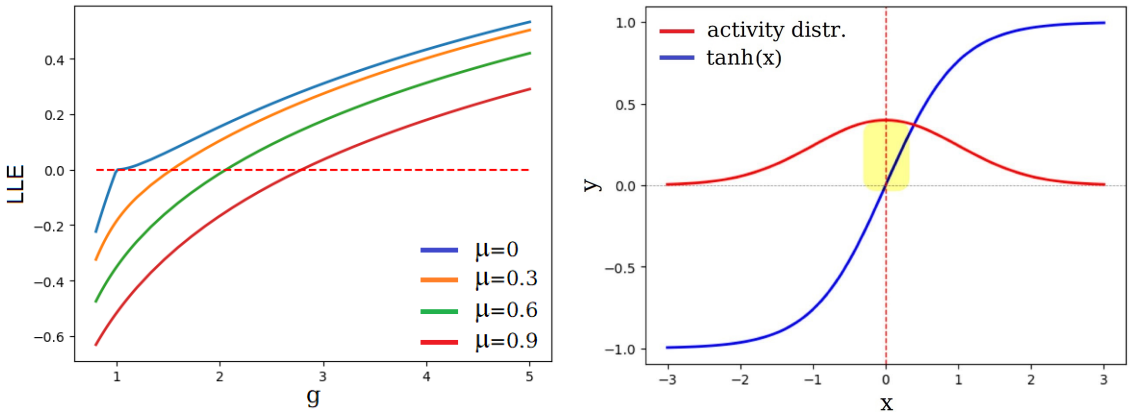


Figure 4.1: Left: LLE of the network obtained from DMFT (eq. (4.6)) with transfer function $\phi(\cdot) = \tanh(\cdot)$ as a function of the gain parameter g for different values of the quenched mean μ. The critical coupling g leading to the transition to chaos (LLE>0) increases for increasing μ. Right: transfer function $\tanh(\cdot)$ (blue) and Gaussian activity distribution of the network with mean m (eq. (4.3)) and variance q (eq. (4.4)) (red). The highlighted area represents the high gain region of the transfer function (see main text). Results obtained from the discrete-time current model (4.1) with i.i.d. synaptic couplings $J_{ij} \sim \mathcal{N}(J_0/N, g^2/N)$ and quenched external input $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$. Network parameters: (left) $J_0 = 0$, $\sigma^2 = 0$; (right) $g = 1$, $J_0 = 0$, $\mu = 0$.

tial portion of the synaptic input distribution is concentrated in the region where the gradient of ϕ is large. This region is commonly referred to as the high-gain region of the transfer function. Indeed, typical random RNN models use a activation function with a slope at the inflection point determined by the gain parameter (i.e., $\phi(x) = \tanh(gx)$), and the variance of the recurrent coupling is simply given by $1/N$. We remark that this representation of the model is equivalent to the one considered here. In the high-gain region of the transfer function $\phi'(\cdot)^2 \sim \mathcal{O}(1)$, resulting in a large Lyapunov exponent (see eq. (4.6)). The distribution of synaptic inputs (which matches the activity distribution) has a mean m and variance q that can be computed self-consistently by solving equations (4.3) and (4.4). In the absence of quenched input, a significant portion of synaptic inputs accesses the high-gain region, inducing chaotic activity in the network. By increasing the mean of the quenched input μ, the mean of the activity distribution shifts proportionally to μ, and the fraction of synaptic inputs in the high-gain region gradually decreases, suppressing chaos. Generally, decreasing the variance of the quenched input σ^2 (see eq.

(4.4)) narrows the activity distribution, ensuring a significant portion of synaptic inputs is in the high-gain region, thus restoring chaotic activity.

4.2 Bistability and ergodicity breaking

So far, we have only considered the input current model (both in discrete and continuous versions) with the hyperbolic tangent transfer function, $\phi(\cdot) = \tanh(\cdot)$. In the discrete case (similarly to the continuous one), the self-consistent solution of the equations (4.3) and (4.4) has several solutions corresponding to attractors in phase space (silent solution, fixed points, chaotic activity). In this scenario, we have observed that regardless of the value of the gain parameter, there is always one attractor. Now, following the work of Ogawa et al. [28], we introduce a more biologically plausible transfer function, namely:

$$\phi(x) = \frac{1}{2}\big(\tanh(x-\theta)+1\big) \tag{4.7}$$

which is positive definite and includes a soft rectification (non-linearity) as well as thresholding (neurons become active if the synaptic input signal exceeds a certain threshold θ). Surprisingly, this transfer function results in interesting patterns of neural activity. To explore this behavior, we performed numerical simulations of the LLE across various mean μ values (assuming variance $\sigma^2 = 0$) as a function of the gain parameter g (see fig. 4.2). Initially, we note that for low μ values, the LLE does not show a monotonic relationship with g, unlike what we observed with $\phi(\cdot) = \tanh(\cdot)$. Moreover, at low g values, the order parameter stabilizes at a single value, and neural activity converges to a single attractor. With increasing g, a bifurcation occurs for specific μ values. This suggests the existence of regions of the phase space where two different LLE values can emerge for a fixed g and the same values of the recurrent and external synaptic inputs, resulting in neural activity converging to two different attractors based on initial conditions in with

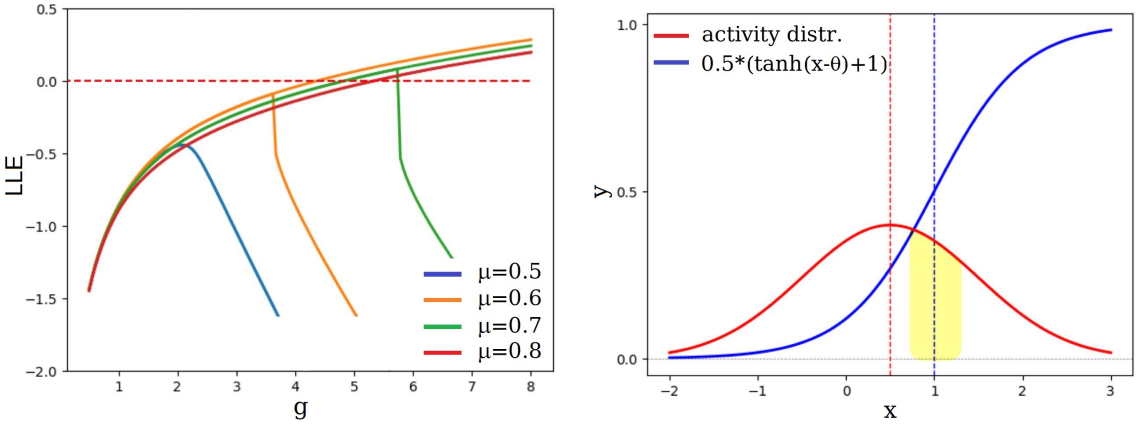


Figure 4.2: Left: LLE of the network obtained from DMFT (eq. (4.6)) with transfer function $\phi(x) = \frac{1}{2}\big(\tanh(x-\theta)+1\big)$ as a function of the gain parameter g for different values of the quenched mean μ. For certain values of μ, at a critical point of g, the LLE exhibits a saddle-node bifurcation (orange and green curves), indicating bistable network activity (see main text). Right: transfer function $\phi(x) = \frac{1}{2}\big(\tanh(x-\theta)+1\big)$ (blue) and Gaussian activity distribution of the network with mean m (eq. (4.3)) and variance q (eq. (4.4)) (red). The highlighted area represents the high gain region of the transfer function. Results obtained from the discrete-time current model (4.1) with i.i.d. synaptic couplings $J_{ij} \sim \mathcal{N}(J_0/N, g^2/N)$ and quenched external input $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$. Network parameters: (left) $\theta = 1$, $J_0 = 0.5$, $\sigma^2 = 0$; (right) $\theta = 1$, $g = 1$, $J_0 = 0$, $\mu = 0.5$.

the network is initiated. In other words, crossing this bifurcation point leads to a breakdown of system ergodicity, where different solutions coexist, and their basin of attraction divides the phase space. Clearly, the other order parameters of the network (the mean activity m and the variance q) shows a similar behavior.

To further investigate this phenomenon and get a clear representation of the system behavior, we analyse the phase diagram of the network in the (μ, σ) space (see fig. 4.3). By varying the parameters of the baseline external input distribution μ and σ^2 the system can access multiple phases. In addition to the monostable phases (fixed points and chaos) encountered previously, there are also bistable phases with coexistence of chaos and fixed points, two fixed points, or, for large values of the gain parameter g, two different chaotic phases with two different positive Largest Lyapunov Exponents (weak and strong chaos) (see fig. 4.4).

As observed in the original work by Mazzuccato et al. [28], these bistable phases can be utilized for various binary decision-making tasks without modifying the re-

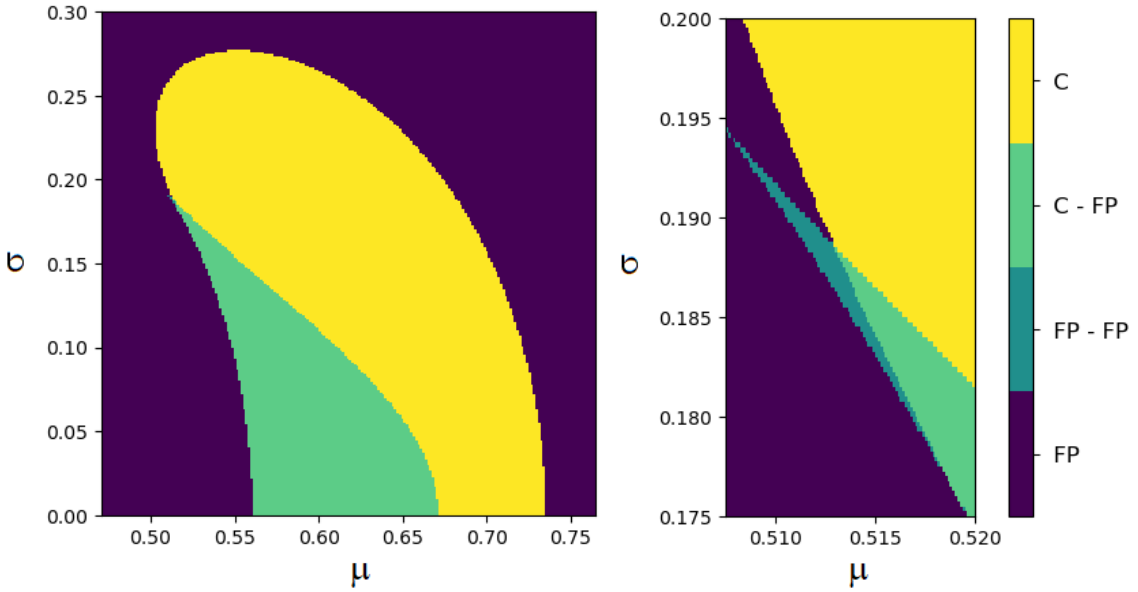


Figure 4.3: Phase diagram in the (μ, σ) space obtained by computing the LLE according to equation (4.6) with transfer function $\phi(x) = \frac{1}{2}\big(\tanh(x-\theta)+1\big)$. The network have access to four different phases reported in the legend (colorbar): fixed point (FP), chaos (C), bistable phases with coexistence of either chaos and fixed points (C-FP) or two fixed point solutions (FP-FP) highlighted by a zoom (right) on the critical point where the three previous phases merge together. The root-finding algorithm for the self-consistent solutions of the mean activity m (see eq. (4.3)) was initialized with large positive and negative values of m to obtain different solutions in the bistable phases. Results obtained from the discrete-time current model (4.1) with i.i.d. synaptic couplings $J_{ij} \sim \mathcal{N}(J_0/N, g^2/N)$ and quenched external input $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$. Network parameters: $\theta = 1$, $g = 5$, $J_0 = 0.5$. Figure reproduced from Mazzuccato et al. [28].

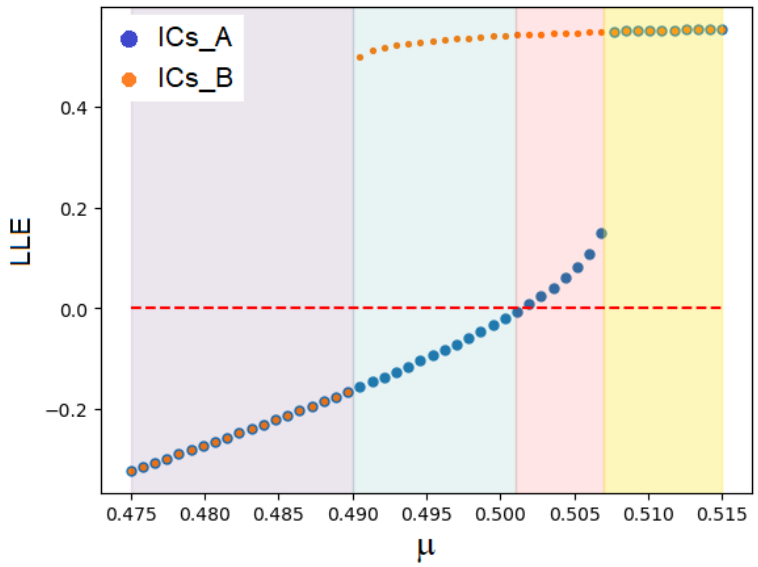


Figure 4.4: LLE of the network as a function of the baseline mean μ. The two branches are obtained by solving the DMFT equation (4.6) for different values of the initial conditions. The highlighted regions represent the different phases of the network: fixed points (purple), chaos (yellow), coexistence of fixed points and chaos (blue) and coexistence of weak and strong chaos (red). Results obtained from the discrete-time current model (4.1) with transfer function $\phi(x) = \frac{1}{2}\big(\tanh(x-\theta)+1\big)$, i.i.d. synaptic couplings $J_{ij} \sim \mathcal{N}(J_0/N, g^2/N)$ and quenched baseline input $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$. Network parameters: $\theta = 1$, $g = 15$, $J_0 = 0.5$, $\sigma = 0.2$.

current couplings, and thus without actual training. The potential outcomes of this binary decision are represented by the two different branches of a bistable phase, in which the network's activity settles after stimuli have been presented. By varying the statistics of the baseline, the network can access different bistable phases that can perform different tasks. Additionally, as we will see in the next section, near a phase boundary between chaotic and fixed point phases, the network exhibits excellent memory capacity, demonstrating the benefit of baseline control in terms of information processing capabilities.

It is worth remarking that there is an excellent agreement between the LLE results obtained with the mean field theory and those obtained from a direct numerical simulation of the network. Additionally, both the input current model (1.9) and the firing model (1.10) produce equivalent results, as observed earlier. However, in this scenario, there is a slight complication. Indeed, when the system is in a phase space

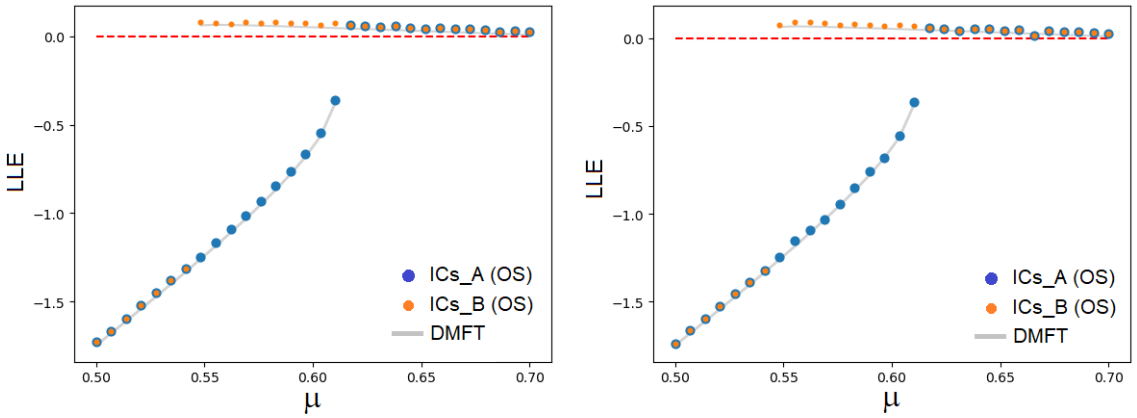


Figure 4.5: LLE of the network as a function of the quenched mean μ obtained by a direct numerical simulation of the input current model (1.9) (left) and firing rate model (1.10) (right) in discrete time with the Orbit Separation method ($N = 1000$, 10000 time steps) for different values of the initial conditions (orange and blue dots). The theoretical DMFT prediction (eq. (4.6)) is presented for comparison (grey). Results obtained for both models with transfer function $\phi(x) = \frac{1}{2}\big(\tanh(x-\theta)+1\big)$, i.i.d. synaptic couplings $J_{ij} \sim \mathcal{N}(J_0/N, g^2/N)$ and quenched external input $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$. Network parameters: $\theta = 1$, $g = 5$, $J_0 = 0.5$ $\sigma = 0.1$.

region where two different solutions exist (bistability), it is crucial to judiciously select the initial conditions to obtain the same behavior, given their importance in determining which of the two solutions the system will reach at steady state. To understand how to choose the initial conditions for the firing model, we first highlight the following observation. We start from the firing model equation, that we rewrite here for simplicity:

$$\dot{r}_i(t) = -r_i(t) + \phi\Big(\sum_j J_{ij} r_j(t) + b_i\Big) \tag{4.8}$$

Now, by defining the variable $I_i(t) = \sum_j J_{ij} r_j(t) + b_i$ (i.e. the input current), we have that it satisfies the input current model (1.9). Indeed, by taking the time derivative of $I_i(t)$ and using equation (4.8), we get:

$$\begin{aligned} \dot{I}_i(t) &= \sum_j J_{ij} \dot{r}_j(t) = \sum_j J_{ij} \Big(-r_j(t) + \phi\Big(\sum_k J_{jk} r_k(t) + b_j\Big)\Big) \\ &= -I_i(t) + \sum_j J_{ij} \phi(I_j(t)) + b_i \end{aligned} \tag{4.9}$$

Therefore, to ensure equivalent steady-state solutions for the system in both descriptions, the initial conditions for the firing model must be chosen based on the definition of $I_i(t)$, as follows:

$$\mathbf{r}(0) = J^{-1}\big(\mathbf{I}(0) - \mathbf{b}\big) \tag{4.10}$$

where J^{-1} represents the inverse of the matrix J. We note that, however, from a computational point of view, this is impractical as the inversion of the connectivity matrix J results in a full matrix, and both the numerical costs of inversion and storing are overwhelmingly high (especially for very large systems). Moreover, and most importantly, in real biological networks the inverse operation can seldom be performed due to the often singular nature of J.

4.3 Optimal sequence memory

The dynamical state of a neural network influence its ability to process information. As we have seen, a neural network of randomly and recurrently interconnected neurons produces complex dynamics which is suitable for various computational tasks. It has been demonstrated that the performance of the network peaks near the boundary between stationary and chaotic states [36]. This boundary depends on the parameters and characteristics of the inputs, and keeping the network near this boundary requires careful adjustment of the recurrent connections, which is biologically costly and slow as it usually requires synaptic plasticity. As shown by Mazzuccato et al. [28], it is possible to set the network near a transition boundary simply by modifying the statistics of the baseline external input, without the need of fine tuning the recurrent connections. In the following, we examine these ideas by computing the memory capacity of the system.

The sequential memory of a neural network refers to its ability to retrieve information from a previous input at a later time. Essentially, it indicates the capacity of the

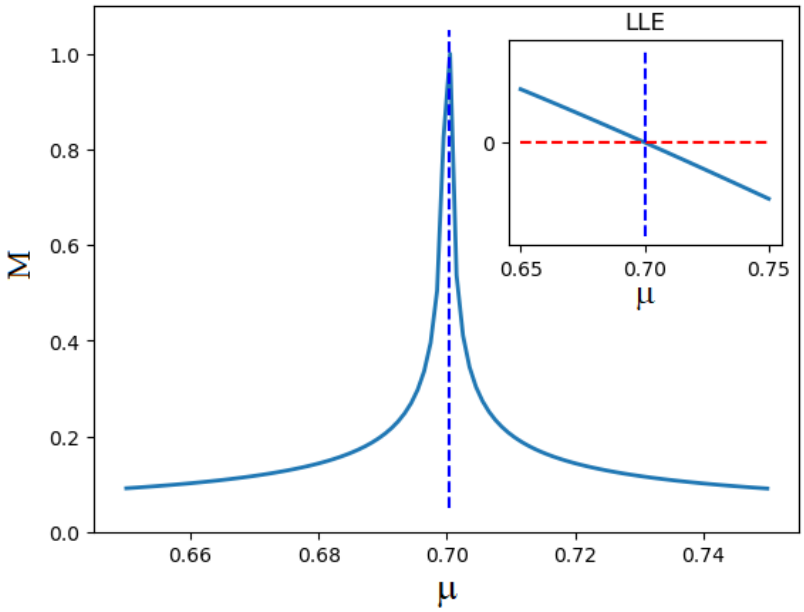


Figure 4.6: (Normalized) memory capacity M of the network obtained accordingly to equation (4.12) as a function of the baseline mean μ. M is optimal at the edge of chaos (see fig. 4.3) where the LLE crosses zero (inset). Results obtained from the discrete-time current model (4.1) with transfer function $\phi(x) = \frac{1}{2}\big(\tanh(x-\theta)+1\big)$, i.i.d. synaptic couplings $J_{ij} \sim \mathcal{N}(J_0/N, g^2/N)$ and quenched external input $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$. Network parameters: $\theta = 1$, $g = 5$, $J_0 = 0.5$ $\sigma = 0.16$.

network to remember sequences of inputs and use this information for subsequent tasks. Mathematically, it is defined as the capability of the system to reconstruct an input signal $z(t)$ at a later time $t+\tau$ based on the network state, using a sparse linear readout $\sum_{j=1}^{K} w_j x_j(t+\tau)$, where K is the number of readout neurons and $\mathcal{O}(K) \ll \mathcal{O}(\sqrt{N})$. To assess how accurately the input signal is reconstructed by the readout neurons, one can calculate the memory curve $m(\tau)$. For optimal readout weights w_i minimizing the error between the input and the readout, $m(\tau)$ can be evaluated as in reference [7]:

$$m(\tau) = \frac{\langle \mathbf{x}(t+\tau)z(t)\rangle^T \langle \mathbf{x}(t)\mathbf{x}(t)^T\rangle^{-1} \langle \mathbf{x}(t+\tau)z(t)\rangle}{\langle z(t)^2\rangle} \tag{4.11}$$

where $\langle\cdot\rangle$ denotes the (discrete) time average. In the thermodynamic limit $N \to \infty$ one can perform a field-theoretical calculation using an approach similar to the one utilized in Section 2.4. Thus, the memory capacity of the system, defined as the as

the integral of the memory curve $m(\tau)$, can be estimated as [36]:

$$M = \int_0^\infty d\tau \; m(\tau) \sim \frac{\sigma^2}{\sqrt{1 - g^2 \langle \phi'(x)^2 \rangle}} \tag{4.12}$$

As shown in figure 4.6, the memory capacity of the system peaks at the phase boundary between fixed point solutions and chaotic phase, where the Largest Lyapunov Exponent crosses zero. At this critical point, the network achieves optimal balance between stability, which is important to mantain information over time, and sensitivity to external inputs. Therefore, the control of the baseline statistics in order to settle the system at the edge of chaos allows for the maximization of information processing capabilities without the need to alter the recurrent couplings.

Chapter 5

Numerical methods

In this chapter, we describe the numerical methods used to investigate the dynamical behavior of random recurrent neural networks. Numerical methods are fundamental to explore the complexities of systems with large number of interacting units. Moreover, they provide insights that are often difficult to derive analytically and allow for comparisons with theoretical results, when available. Our aim is to introduce two distinct approaches: one for computing the Largest Lyapunov Exponent through direct simulations, and another for solving the self-consistent Dynamic Mean Field Theory equations.

The first method we discuss is the Orbit Separation Method (OS), a simple and standard numerical technique used to compute the LLE of dynamical systems. Throughout this thesis, we have widely employed it to compare theoretical predictions with numerical simulations.

The second method we focus on is the solution of the self-consistent equations of the DMFT. Although this approach has not been commonly used in the thesis since we primarily focus on steady state analysis, we describe it because it is essential to understand the theoretical background underlying our research and helps to illustrate the nature of the DMFT equations.

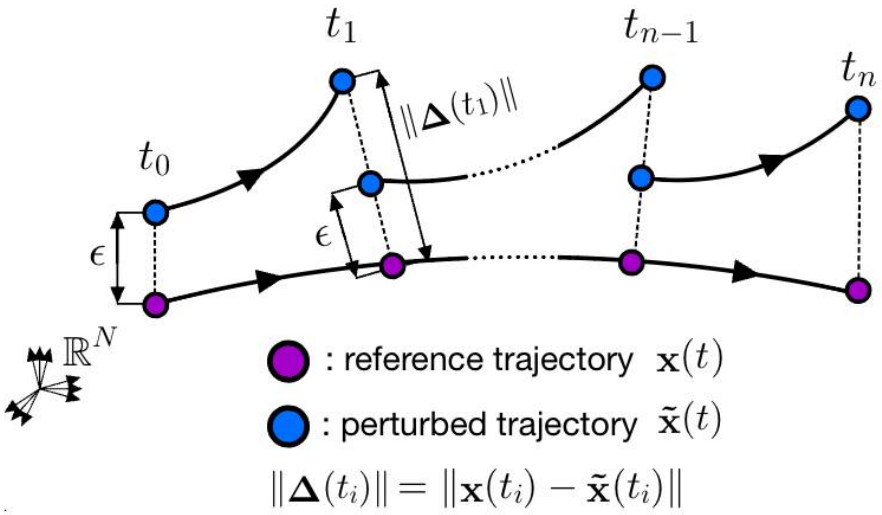


Figure 5.1: Schematic representation of the OS method for computing the LLE.

5.1 Orbit Separation Method

The Lyapunov exponents of a dynamic system are frequently computed using numerical methods. The OS method is a flexible numerical technique for calculating the LLE that can be applied to any system in any dimension and it does not require the explicit evaluation of the Jacobian (monodromy) matrix of the system. We remark that a Jacobian-based method for computing the full spectrum of Lyapunov exponent that allows us to study interesting physical properties, such as dynamical entropy rate and attractor dimensionality, is presented in [11].

Essentially, the OS method relies on numerically integrating and repeatedly comparing two orbits (trajectories) of the system with very close initial conditions, and evaluating the average logarithm of the distance between these two orbits in phase space. According to [39], the numerical procedure, illustrated in figure 5.1, can be described as follows:

- Initialize two orbits of the system $\mathbf{x}(t)$ and $\tilde{\mathbf{x}}(t)$ with the same initial conditions (i.e. $\mathbf{x}(0) = \tilde{\mathbf{x}}(0)$). Next, perturb the second orbit by a small constant ϵ in any

(random) direction, as:

$$\tilde{\mathbf{x}}(0) = \mathbf{x}(0) + \epsilon \frac{\mathbf{\Delta}(0)}{\|\mathbf{\Delta}(0)\|} \tag{5.1}$$

where $\mathbf{\Delta}(0)$ is a random vector of dimension N. The perturbation magnitude ϵ should be much smaller than the time scale on which the flow changes, but several orders of magnitude larger than the numerical precision. We use $\epsilon = 10^{-10}$ in double precision.

- Apply a numerical method (e.g. Runge Kutta 4^{th} order) to integrate once the equation describing the evolution of the system for each initial condition and determine $\mathbf{x}(t_1)$ and $\tilde{\mathbf{x}}(t_1)$. We may consider the first orbit as the reference one and the second as the perturbed one.

- Determine the distance vector $\mathbf{\Delta}(t_1) = \mathbf{x}(t_1) - \tilde{\mathbf{x}}(t_1)$ and evaluate (and store) the natural logarithm of the relative separation λ_1:

$$\lambda_1 = \ln\left(\frac{\|\mathbf{\Delta}(t_1)\|}{\epsilon}\right) \tag{5.2}$$

- Apply a normalization procedure to ensure that at every time-step the orbits are ϵ apart while maintaining the direction oriented to the one of maximum expansion. In other words, readjust the second orbit as follows:

$$\tilde{\mathbf{x}}(t_1) = \mathbf{x}(t_1) + \epsilon \frac{\mathbf{\Delta}(t_1)}{\|\mathbf{\Delta}(t_1)\|} \tag{5.3}$$

- Iterate the procedure outlined above by considering at each time-step the initial conditions $\mathbf{x}(t_k)$ and $\tilde{\mathbf{x}}(t_k)$ given by (5.3) for the reference and perturbed orbit, respectively.

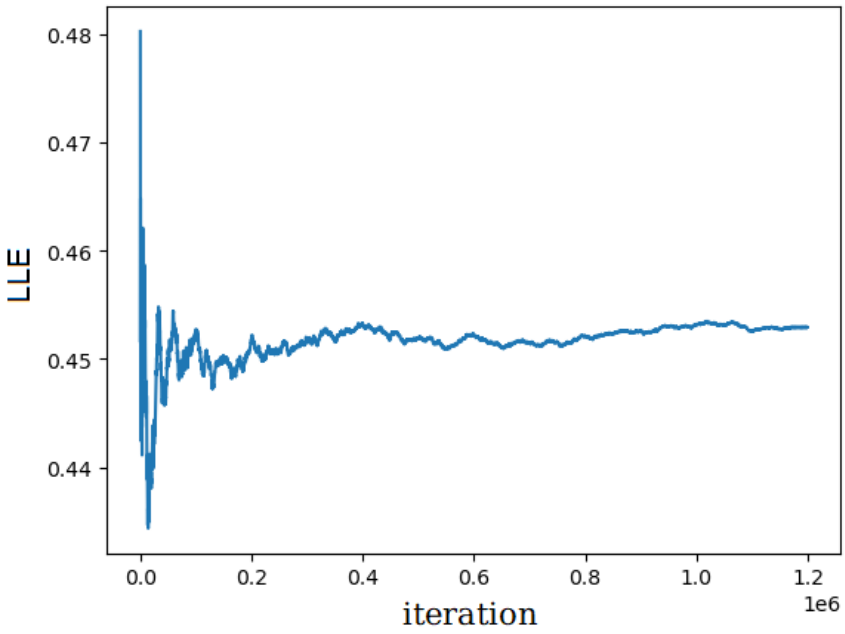


Figure 5.2: Convergence off the LLE of the network gain parameter $g = 5$ obtained by applying the Orbit Separation method with $N = 1000$ and averaging over 10 different initial conditions and realization of the connectivity matrix J. Total simulation time $T = 15000$ with $\Delta t = 0.01$ (1/5 of the total iteration were discarded to allow the system reach the attractor). The LLE value, $\lambda = 0.4529$, is in excellent agreement with the DMFT prediction of $\lambda = 0.4534$. Result obtained for the input current model (1.9) with transfer function $\phi(\cdot) = \tanh(\cdot)$ and i.i.d. synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$.

- Compute the LLE exponent of the system as:

$$\lambda = \frac{1}{\Delta t}\Big(\frac{1}{n}\sum_{i=1}^{n}\lambda_i\Big) \tag{5.4}$$

  where n represent the total number of iterations, and the division by the time-step size Δt ensures that the units are correct (as the system is a flow rather than a map).

It is worth remarking that for a more accurate value of the Lyapunov exponent, this numerical procedure is performed for several initial conditions of the system and realizations of the connectivity matrix J. The LLE is then calculated by averaging those obtained with these different realizations. Additionally, in order to improve the convergence of the method, it is recommended to exclude from the calculation

of the mean in equation (5.4) the initial iterations where the orbit is not sufficiently close to the (strange) attractor.

The convergence of the LLE with respect to the number of iterations of the OS method is illustrated in figure 5.2, demonstrating excellent agreement with the LLE obtained analytically through DMFT.

5.2 Numerical solution of DMFT equation

As we have demonstrated in Chapter 2, the DMFT equation for the the generic random neural network (2.1), which we rewrite here for convenience, are given by:

$$\dot{x}(t) = -x(t) + \eta(t) + g^2\gamma \int_0^t R(t,s)\phi(s)ds \tag{5.5}$$

where $\eta(t)$ is the effective Gaussian noise with zero mean and variance $\langle \eta(t)\eta(s) \rangle = g^2 C(t,s) + \sigma^2\delta(t-s)$ and $R(t,s)$ is the response function arising from correlated couplings ($\gamma \neq 0$). Clearly, the DMFT equation lacks an analytical solution in closed form and needs to be solved numerically. However, solving the self-consistent DMFT equation is a challenging task compared to solving systems in equilibrium or steady state. This difficulty primarily arises from the presence of time-dependent functions in the self-consistent iterations of the numerical scheme, such as the two-point autocorrelation of the firing rate $C(t,s)$ and the response function $R(t,s)$.

We describe the numerical procedure used to solve the DMFT equation for the generic neural network model, based on the method outlined in [34]. Essentially, this procedure involves the simulation of a large number of trajectories of the DMFT equation. The observables of the system (e.g. autocorrelation $C(t,s)$ and mean activity $m(t)$) are computed in each iteration by averaging over different paths, until convergence. Clearly, the numerical scheme operates in discrete time and requires defining a total simulation time T as well as the time-step Δt for numerical integration. In this setting, the autocorrelation $C(t,s)$ and the response function

$R(t,s)$ become squared matrices of dimension $T/\Delta t$, initialized as identity matrices. However, to avoid excessive notation, we will still refer to the entries of these matrices as (t,s), noting that these variables take values in the discrete time lattice (i.e. $t,\ s = 0, \Delta t, \ldots, T$).

In each iteration of the algorithm, the operations can be described as follows:

- Sample M effective noise trajectories $\{\eta_i(t = 0, \ldots, T)\}_{i=1}^{M}$, each of them from a multivariate Gaussian distribution $\mathcal{N}(0, g^2 C(t,s) + \sigma^2/\Delta t)$, where the discretization of the Dirac delta function leads to the appearance of Δt.

- Apply a numerical method to integrate the DMFT equation for each the noise paths, obtaining M current trajectories $\{x_i(t = 0, \ldots, T)\}_{i=1}^{M}$. By using a simple Euler scheme, we have:

$$x_i(t + \Delta t) = (1 - \Delta t)x_i(t) + \Delta t\, \eta_i(t) + g^2 \gamma \Delta t^2 \sum_{s=0}^{t} R(t,s)\phi(x_i(s)) \tag{5.6}$$

  The initial conditions $\{x_i(0)\}_{i=1}^{M}$ are drawn randomly (e.g. from an uniform distribution $\mathcal{U}(0,1)$).

- Compute the new autocorrelation matrix $C(t,s)$ and the response matrix $R(t,s)$ by averaging over the paths. In particular, the self consistent functions are evaluated by:

$$C(t,s) = \frac{1}{M} \sum_{i=1}^{M} \phi(x_i(t))\phi(x_i(s)) \tag{5.7}$$

$$R(t,s) = \frac{1}{M} \sum_{i=1}^{M} \chi_i(t,s)\phi'(x_i(t)) \tag{5.8}$$

  where $\chi_i(t,s)$ represents the response function of the input currents $x_i(t)$. This function can be determined by solving the following integro-differential equation

that can be derived in the MSRDJ formalism [43] and, in discrete time, reads:

$$\chi_i(t+\Delta t, s) = (1-\Delta t)\chi_i(t,s) + \delta_{t\,s} + g^2\gamma\Delta t^2 \sum_{t'=s}^{t} R^{\text{old}}(t,t')R_i(t',s) \tag{5.9}$$

Here, $R^{\text{old}}(t,s)$ represent the response function estimated in the last iteration of the algorithm, and $R_i(t,s) = \phi'(x_i(t))\chi_i(t,s)$.

- Update the autocorrelation function $C(t,s)$ and the response function $R(t,s)$ and start a new iteration of the algorithm. The updating rule is based on soft reinjection, as follows:

$$C^{\text{updated}}(t,s) = (1-\alpha)C^{\text{old}}(t,s) + \alpha C(t,s) \tag{5.10}$$

where α is the reinjection parameter (we use $\alpha = 0.3$). The same updating applies to $R(t,s)$. This method is essential to ensure the convergence of the algorithm, preventing it from jumping unpredictably from one function to another.

It is worth noting that in the DMFT setting, each noise trajectory $\eta_i(t)$ is independent from the others. A simple strategy to generate them all is to start from a random (noise) matrix A of dimension $(T/\Delta t, M)$, where the entries are independent identically distributed random variables drawn from a standard Gaussian distribution. Next, we diagonalize the covariance matrix $\Gamma(t,s) = g^2C(t,s) + \sigma^2/\Delta t$ and multiply it with the orthogonal basis matrix of $\Gamma(t,s)$ to obtain the matrix B. The trajectories are then given by the rows of the product matrix AB.

The comparison between the observables obtained from a direct numerical simulation of the network and the DMFT solution, obtained with the algorithm described previously (with $\gamma = 0$), is shown in figure 5.3 for two different values of the gain parameter g: one above the critical coupling ($g = 1$) and one below it. We observe an excellent agreement between the two sets of data. Additionally, in the chaotic

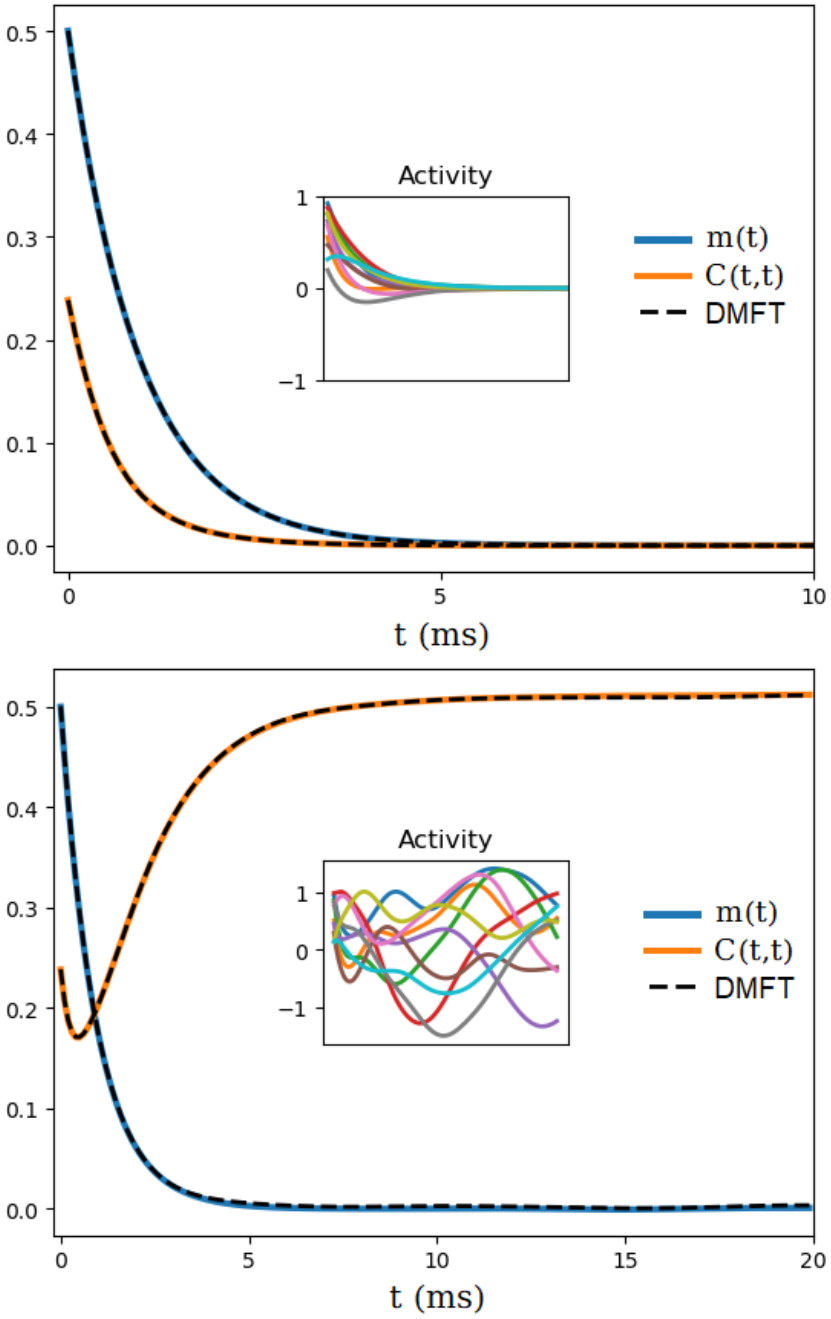


Figure 5.3: Mean activity $m(t)$ and same-time autocorrelation function of the firing rate $C(t,t)$ as a function of time for $g = 0.5$ (top) and $g = 2$ (bottom). Observables obtained from a direct numerical simulation of a network with $N = 10000$ averaged over 100 different realization of the connectivity matrix J and initial conditions (drawn from an uniform distribution $\mathcal{U}(0,1)$) are compared with the DMFT results (dotted line) obtained by averaging over 100000 trajectories. The network dynamics is shown in the insets. Results obtained for generic current model (2.1) with transfer function $\phi(\cdot) = \tanh(\cdot)$, synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$ with a potential asymmetric correlation $\langle J_{ij}J_{ji}\rangle = g^2\gamma/N$ and Gaussian white noise with zero mean and variance $\langle \xi_i(t)\xi_j(s)\rangle = \sigma^2\delta_{ij}\delta(t-s)$. Network parameters: $\gamma = 0$, $\sigma = 0$.

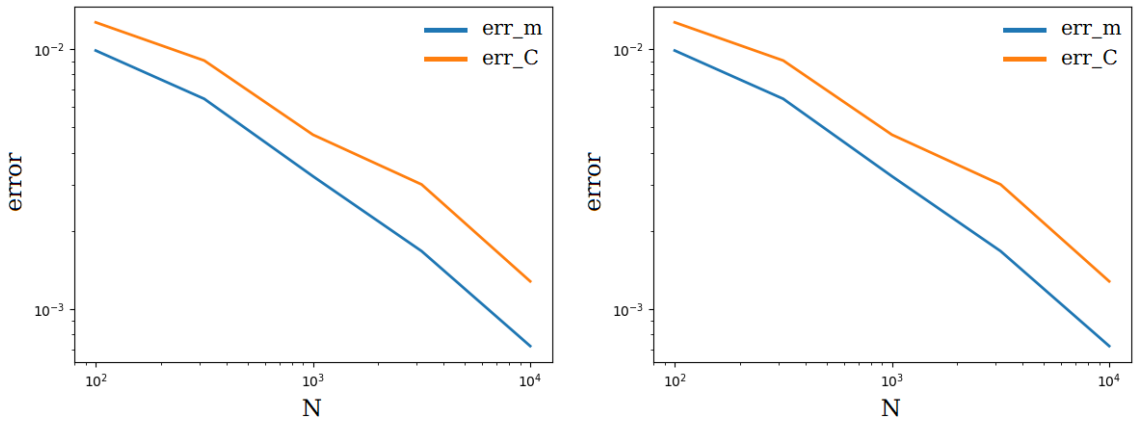


Figure 5.4: Relative error between direct numerical simulation and DMFT results as a function of N the for the mean activity $m(t)$ (blue) and same-time autocorrelation function $C(t,t)$ (orange) for $g = 0.5$ (left) and $g = 2$ (right) computed according to eq. (5.11). Simulations results are obtained by averaging over 100 different realization of the connectivity matrix J and initial conditions. Results obtained for generic current model (2.1) with transfer function $\phi(\cdot) = \tanh(\cdot)$, synaptic couplings $J_{ij} \sim \mathcal{N}(0, g^2/N)$ with a potential asymmetric correlation $\langle J_{ij}J_{ji}\rangle = g^2\gamma/N$ and Gaussian white noise with zero mean and variance $\langle \xi_i(t)\xi_j(s)\rangle = \sigma^2\delta_{ij}\delta(t-s)$. Network parameters: $\gamma = 0$, $\sigma = 0$.

phase of the network, there is an emergence of a non-zero same-time autocorrelation function of the firing rate $C(t,t)$ at stationarity, conistently with what we observed in Chapter 3. In order to investigate further on the differences, we evaluate the relative error between the simulation results (variables with tilde) and DMFT results, computed by:

$$\text{err}(\mathbf{m}) = \frac{\|\mathbf{m} - \tilde{\mathbf{m}}\|_2}{\|\mathbf{m}\|_2} \qquad \text{err}(C) = \frac{\|C - \tilde{C}\|_F}{\|C\|_F} \tag{5.11}$$

where the vector $\mathbf{m} = m(t = 0, \dots, T)$ is the mean firing rate activity and $\|\cdot\|_2$ and $\|\cdot\|_F$ are the $\mathcal{L}^2$ norm and Frobenious norm, respectively. As illustrated in figure 5.4, the relative error decreases as the number of neurons N in the direct numerical simulation of the network increases. This confirms the hypothesis that the DMFT equation describes the characteristic behavior of the system in the large N limit.

Conclusions and Outlook

In this thesis, we investigated the dynamical behavior of a fully-connected Recurrent Neural Network with randomly Gaussian-distributed couplings. Despite representing an interesting physical systems, their analysis is complicated due to the inherent stochastic activity, the non linearity of the input-output transfer function of a single neuron and the strong interaction effects deriving from densely recurrent connections. To grasp the collective behaviors emerging from these complexities, we reformulated the stochastic differential equations in a path integral formalism. Leveraging the large number of synaptic inputs and the self-averaging properties of the system (the value of an observable in a single realization of the quenched disorder converges to its average over the disorder), we applied Dynamic Mean Field Theory (DMFT) to reduce the interacting system to an effective equation of a single neuron embedded in a fluctuating field with self-consistently determined statistics. In the case of an autonomous system, the mean-field perspective enabled us to analyse the time-lag dependent autocorrelation function of a typical unit through the motion of a particle in a self-consistent potential. As the gain parameter g is increased beyond unity, an emergent non-zero autocorrelation function is observed with a decay time inversely proportional to the value of g, indicating that neuronal activity becomes uncorrelated at subsequent times and loses memory of its initial state. Correspondingly, a phase transition from a fixed-point solution to a chaotic state of the network is observed at the critical value $g = 1$. Accordingly, the Largest

Lyapunov Exponent (LLE) of the network, calculated by determining the ground state energy of a quantum mechanical problem (time-independent Schrödinger equation), shifts from negative to positive values. The same critical condition is predicted by Random Matrix Theory. In particular, according to the circular law, the transition to chaos arises when the spectral radius of the stability matrix exceeds unity. Interestingly, for finite size systems, numerical simulations reveal that the transition from a stationary to a chaotic state passes through intermediate stages of limit-cycles, oscillatory behavior with increasing complexity. In the limit $N \to \infty$, a sharp transition emerges as predicted by DMFT.

Significantly, the numerical results obtained through a direct simulation of the network show excellent agreement with the theoretical predictions. Moreover, surprisingly, both the input current model and the firing rate model, display the same behavior. In particular, they both draw from the complete model the same critical condition for the chaotic transition.

In the case of a stochastically driven network, chaotic activity emerges at higher coupling parameter g. This input driven suppression of chaos can be understood in a mean-field perspective by considering how the activity distribution interact with the single-neuron transfer function. We argued that chaos emerges when the activity distribution is concentrated around the saddle point of the transfer function. Turning on a external input alters this distribution, thus stabilizing the dynamics.

Surprisingly, by changing the mean μ and variance σ of the quenched external input a rich set of dynamical behavior is observed. We delineated the phase diagram of the network in the (μ, σ) space by computing the LLE through DMFT. Rather than fixed point and chaotic states, several bistable phases emerge as the network activity breaks ergodicity. Depending on the initial conditions, for the same values of μ and σ, the network can settle into a fixed point or a chaotic state, two different fixed points or two chaotic states (weak and strong chaos). This bistable phases can be harnessed to perform multiple tasks (binary decision-making) without any

weight optimization, simply by changing the statistics of the baseline quenched input [28]. Overall, the impact of this baseline modulation can be understood in terms of changes in the slope of an effective transfer function $\Phi_{\text{eff}}(r) = \int Dx\ \phi(\sqrt{q}x + m + r)$, where m and q are the self-consistent mean and variance of the activity distribution, and Dx the standard Gaussian measure. This observation reflects experimental findings which suggest that changes in behavioral states are influenced by gain modulation, which allow organisms to adapt their responses to changing environmental conditions and behavioral demands [42]. An intuitive understanding of this phenomenon is given by examining the effective potential $E(r) = \int dr\ \big(r - \Phi_{\text{eff}}(r)\big)$, derived from an effective mean-field theory [25]. This potential exhibits a double-well shape, and the height of the barrier separating the two energy minima, which represent two distinct attractors for the network, is directly proportional to the slope of the effective transfer function. Increasing the variance of the quenched input reduces this slope, thereby decreasing the barrier height and allowing for quicker transitions between different states and faster encoding of upcoming stimuli.

To investigate on the information processing capabilities, we evaluated the sequential memory of the network. Interestingly, it peaks in the edge of chaos, where the network achieves an optimal balance between stability and flexibility. Indeed, at the borderline between order and disorder, the network is very sensitive to external perturbation. At this critical state, it is able to explore a wide range of dynamic states and respond flexibly to external inputs. Despite this sensitivity, the system maintains enough stability to store and retrieve information consistently over time. While baseline modulation offers a straightforward mechanism to bring the network close to a critical point, the fundamental question of how living systems dynamically tune themselves in the proximity of a critical state remains open. Indeed, this requires a deep understanding of the intricate mechanisms of neural circuits responsible for the transmission and the modulation of these signals. As an alternative, we propose a model based on synaptic plasticity, where the synaptic connections

are dynamic variables subject to a Hebbian updating rule. A straightforward implementation, in discrete time, is given by:

$$J_{ij}(t+1) = J_{ij}(t) + \frac{\alpha}{N}\big(\phi_i(t+1) - \theta\big)\big(\phi_j(t) - \theta\big)H\big(\phi_j(t) - \theta\big)$$

Here, the parameter $\alpha > 0$ represents the intensity of the Hebbian plasticity, $H(\cdot)$ denotes the Heaviside step function, and we used abbreviated notation to indicate the neuronal firing rate ($\phi(x_i(t)) = \phi_i(t)$). Additionally, the parameter θ defines a threshold that allows us to determine if a neuron is active at time t ($\phi_i(t) > \theta$) or silent. From a biological perspective, the effect of the plasticity rule is to enhance the synaptic connection J_{ij} if the pre-synaptic neuron j is active at time t and the post-synaptic neuron i is active at time $t+1$. On the other hand, if the neuron j is active and the neuron i is inactive the synaptic connection is weakened.

This process is related to the ability of the brain to reinforce neural connections associated with correlated and relavant activity patterns, while weakening those associated with uncorrelated activity. On general grounds, we expect this dynamical modification of the synaptic couplings to drive the neural network towards the edge of chaos, where it achieves a delicate balance between order and disorder. This idea is supported by the following intuitive reasoning. When the network is initialized in the chaotic phase, increasing the connections between active neurons induces the emergence of correlated activity patters, that could bring the network closer to an ordered phase. On the other hand, if the initial condition is chosen to be in a stable state, the modification of the synaptic connections could introduce variability on the dynamics of the network, eventually pushing the system towards a more disordered state. We leave the investigation of this phenomenon for future research.

Bibliography

[1] J. Ahmadian, F. Fumarola, and D. M. Kenneth. Properties of networks with partially structured and partially random connectivity. *Physical Review E*, 91(1), 2013.

[2] S. I. Amari. Characteristics of random nets of analog neuron-like elements. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-2(5), 1972.

[3] D. J. Amit. *Modelling brain function: the world of attractor neural networks*. Cambridge University Press, 1989.

[4] B. Cessac. Increase in complexity in random neural networks. *Journal de Physique I*, 5(3), 1995.

[5] B. Cessac and M. Samuelides. From neuron to neural networks dynamics. *The European Physical Journal Special Topics*, 142(1), 2007.

[6] A. Crisanti and H. Sompolinsky. Path integral approach to random neural networks. arXiv:1089.06042v2, 2018.

[7] J. Dambre, D. Verstaeten, Schrauwen B., and S. Massar. Information processing capacity of dynamical systems. *Scientific Reports*, 2(1), 2012.

[8] P. Dayan and L. F. Abbott. *Theoretical neuroscience: computational and mathematical modelling of neural systems*. The MIT Press, 2005.

[9] B. Doyon, B. Cessac, M. Quoy, and M. Samuelides. Mean-field equations, bifurcation map and chaos in discrete time, continuous state, random neural networks. *Acta Biotheoretica*, 43(1-2), 1995.

[10] J. P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, 57(3), 1985.

[11] R. Engelken, F. Wolf, and L. F. Abbott. Lyapunov spectra of chaotic recurrent neural networks. *Physal Review Research*, 5(4), 2023.

[12] W. J. Freeman. Simulation of chaotic eeg patterns with a dynamic model of the olfactory system. *Biological Cybernetics*, 56, 1987.

[13] W.J. Freeman, Y. Yao, and B. Burke. Central pattern generating and recognizing in olfactory bulb: A correlation learning rule. *Neural Networks*, 1(4), 1988.

[14] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninsky. *Neuronal dynamics: from single neurons to networks and models of cognition*. Cambridge University Press, 2005.

[15] O. Harish and D. Hansel. Asynchronous rate chaos in spiking neuronal circuits. *PLoS Computational Biology*, 11(7), 2015.

[16] M. Helias and D. Dahmen. *Statistical field theory for neural networks*. Springer, 2020.

[17] J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the theory of neural computation*. CRC Press, 2018.

[18] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4), 1952.

[19] H. Huang. *Statistical mechanics of neural networks*. Springer, 2021.

[20] D. Hulsey, K. Zumwalt, L. Mazzuccato, D. A. McCormick, and S. Jaramillo. Decision-making dynamics are predicted by arousal and unistructed movements. *Cell Reports*, 43(2), 2024.

[21] J. Kadmon and H. Sompolinsky. Transition to chaos in random reuronal retworks. *Physical Review X*, 5, 2015.

[22] R. Legenstein and W. Maass. Edge of chaos and prediction of computational performance for neural circuit models. *Neural Networks*, 20(3), 2007.

[23] L. Mazzuccato. Neural mechanism underlying the temporal organizzation of naturalistic animal behavior. arXiv:2203.02151v1, 2022.

[24] L. Mazzuccato, A. Fontanini, and G. La Camera. Dynamics of multistable states during ongoing and evoked cortical activity. *The Journal of Neuroscience*, 35(21), 2015.

[25] L. Mazzuccato, G. La Camera, and A. Fontanini. Expectation-induced modulation of metastable activity underlies faster coding of sensory stimuli. *Nature Neuroscience*, 22, 2019.

[26] K. D. Miller and F. Fumarola. Mathematical equivalence of two common forms of firing-rate models of neural network. *Neural Computation*, 24(1), 2012.

[27] D. B. Nestvogel and D. A. McCormick. Visual thalamocortical mechanisms of waking state-dependent activity and alpha oscillations. *Neuron*, 110(1), 2022.

[28] S. Ogawa, F. Fumarola, and L. Mazzuccato. Multi-tasking via baseline control in recurrent neural networks. *Proceedings of the National Academy of Sciences*, 120(33), 2023.

[29] G. Parisi, M. Mézard, and M. A. Virasoro. *Spin Glass Theory and Beyond*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific, 1987.

[30] P. O. Polack, J. Friedman, and P. Golshani. Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nature Neuroscience*, 16(9), 2013.

[31] K. Rajan and L. F. Abbott. Eigenvalue spectra of random matrices for neural networks. *Physical Review Letters*, 97(18), 2006.

[32] K. Rajan, L. F. Abbott, and H. Sompolinsky. Stimulus-dependent suppression of chaos in recurrent neural networks. *Physical Review E*, 82(1), 2010.

[33] S. Recanatesi, U. Pereira-Obilinovic, M. Murakami, A. Mainen, and L. Mazzuccato. Metastable attractors explain the variable timing of stable behavioral action sequences. *Neuron*, 100, 2022.

[34] F. Roy, G. Biroli, G. Bunin, and C. Cammarota. Numerical implementation of dynamical mean field theory for disordered systems: application to the lotka-volterra model of ecosystems. *Journal of Physics A: Mathematical and Theoretical*, 52(48), 2019.

[35] J. Schücker, S. Goedeke, D. Dahmen, and M. Helias. Functional methods for disordered neural networks. arXiv:1605.06758v2, 2016.

[36] J. Schücker, S. Goedeke, and M. Helias. Optimal sequence memory in driven random networks. *Physical Review X*, 8(4), 2018.

[37] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein. Spectrum of large random asymmetric matrices. *Physical Review Letters*, 60(19), 1988.

[38] H. Sompolinsky, A. Crisanti, and H. J. Sommers. Chaos in random neural networks. *Physical Review Letters*, 61(3), 1988.

[39] J. C. Sproot. *Chaos and time series analysis*. OUP Oxford, 2003.

[40] M. Stern, N. Istrate, and L. Mazzuccato. A reservoir of timescales in random neural network. arXiv:2110.09165v2, 2022.

[41] M. Stern, H. Sompolinsky, and L. F. Abbott. Dynamics of random neural networks with bistable units. *Physical Review E*, 90(6), 2014.

[42] D. Wyrick and L. Mazzuccato. State-dependent regulation of cortical processing speed via gain modulation. *The Journal of Neuroscience*, 41(18), 2021.

[43] W. Zou and H. Huang. Introduction to dynamical mean-field theory of randomly connected neural networks with bidirectionally correlated couplings. *SciPost Physics Lecture Notes*, 79, 2024.