Universidade Estadual de Campinas
Instituto de Física "Gleb Wataghin"

Elohim Fonseca dos Reis

Criticality in Neural Networks

Criticalidade em Redes Neurais

CAMPINAS
2015

Universidade Estadual de Campinas
Instituto de Física "Gleb Wataghin"

Elohim Fonseca dos Reis

Criticality in Neural Networks

Criticalidade em Redes Neurais

Dissertation presented to the Institute of Physics "Gleb Wataghin" of the University of Campinas in partial fulfillment of the requerements for the degree of Master in Physics.

Dissertação apresentada ao Instituto de Física "Gleb Wataghin" da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Mestre em Física.

Supervisor/Orientador: Prof. Dr. José Antônio Brum
Co-supervisor/Coorientador: Prof. Dr. Marcus Aloizio Martinez de Aguiar

Este exemplar corresponde à versão final da Dissertação defendida por Elohim Fonseca dos Reis e orientada pelo Prof. Dr. José Antônio Brum.

CAMPINAS
2015

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 2013/25361-6

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Física Gleb Wataghin
Valkíria Succi Vicente - CRB 8/5398

R275c

Reis, Elohim Fonseca dos, 1984-
Criticality in neural networks / Elohim Fonseca dos Reis. – Campinas, SP : [s.n.], 2015.

Orientador: José Antônio Brum.
Coorientador: Marcus Aloizio Martinez de Aguiar.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin.

1. Sistemas complexos. 2. Redes complexas. 3. Redes neurais (Neurobiologia). 4. Física estatística. I. Brum, José Antônio,1959-. II. Aguiar, Marcus Aloizio Martinez de,1960-. III. Universidade Estadual de Campinas. Instituto de Física Gleb Wataghin. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Criticalidade em redes neurais
Palavras-chave em inglês:
Complex systems
Complex networks
Neural networks (Neurobiology)
Statistical physics
Área de concentração: Física
Titulação: Mestre em Física
Banca examinadora:
José Antônio Brum [Orientador]
Osame Kinouchi Filho
Maurice de Koning
Data de defesa: 09-12-2015
Programa de Pós-Graduação: Física

MEMBROS DA COMISSÃO JULGADORA DA DISSERTAÇÃO DE MESTRADO DE **ELOHIM FONSECA DOS REIS – RA: 070715** APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA "GLEB WATAGHIN", DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 09/12/2015.

COMISSÃO JULGADORA:

- Prof. Dr. José Antônio Brum – (Orientador) – DFMC/IFGW/UNICAMP
- Prof. Dr. Osame Kinouchi Filho – DF/FFCLRP-USP
- Prof. Dr. Maurice de Koning – DFMC/IFGW/UNICAMP

OBS.: Informo que as assinaturas dos respectivos professores membros da banca constam na ata de defesa já juntada no processo de vida acadêmica do aluno.

CAMPINAS
2015

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Prof. José Brum for the continuous support, patience and motivation during this period not only in the scientific issues but also in my formation as a scientist.

Besides my advisor, I would like to thank my co-advisor Dr. Prof. Marcus Aguiar for the support and helpful discussions. I also would like to thank Dr. Prof. Gabriela Castellano, her students Raphael Casseb and Luis Herrera, and Dr. Prof. Rickson Mesquita for providing the experimental data and for their collaboration in the work; and Dr. Thierry Mora for supervising my work and for the support during the research internship at the École Normale Supérieure in Paris.

I thank CNPq and Fapesp for the funding support.

Last but not least, I would like to thank my family, specially my mother Ivone Fonseca and also my long time girlfriend Camilla Rehem, for supporting me throughout all this period and my life in all aspects.

Resumo

Este trabalho é dividido em duas partes. Na primeira parte, uma rede de correlação é construída baseada em um modelo de Ising em diferentes temperaturas, crítica, subcrítica e supercrítica, usando um algorítimo de Metropolis Monte-Carlo com dinâmica de *single-spin-flip*. Este modelo teórico é comparado com uma rede do cérebro construída a partir de correlações das séries temporais do sinal BOLD de fMRI de regiões do cérebro. Medidas de rede, como coeficiente de aglomeração, mínimo caminho médio e distribuição de grau são analisadas. As mesmas medidas de rede são calculadas para a rede obtida pelas correlações das séries temporais dos spins no modelo de Ising. Os resultados da rede cerebral são melhor explicados pelo modelo teórico na temperatura crítica, sugerindo aspectos de criticalidade na dinâmica cerebral.

Na segunda parte, é estudada a dinâmica temporal da atividade de um população neural, ou seja, a atividade de células ganglionares da retina gravadas em uma matriz de multi-eletrodos. Vários estudos têm focado em descrever a atividade de redes neurais usando modelos de Ising com desordem, não dando atenção à estrutura dinâmica. Tratando o tempo como uma dimensão extra do sistema, a dinâmica temporal da atividade da população neural é modelada. O princípio de máxima entropia é usado para construir um modelo de Ising com interação entre pares das atividades de diferentes neurônios em tempos diferentes. O ajuste do modelo é feito com uma combinação de amostragem de Monte-Carlo e método do gradiente descendente. O sistema é caracterizado pelos parâmetros aprendidos, questões como balanço detalhado e reversibilidade temporal são analisadas e variáveis termodinâmicas, como o calor específico, podem ser calculadas para estudar aspectos de criticalidade.

Abstract

This work is divided in two parts. In the first part, a correlation network is build based on an Ising model at different temperatures, critical, subcritical and supercritical, using a Metropolis Monte-Carlo algorithm with single-spin-flip dynamics. This theoretical model is compared with a brain network built from the correlations of BOLD fMRI temporal series of brain regions activity. Network measures, such as clustering coefficient, average shortest path length and degree distributions are analysed. The same network measures are calculated for the network obtained from the time series correlations of the spins in the Ising model. The results from the brain network are better explained by the theoretical model at the critical temperature, suggesting critical aspects in the brain dynamics.

In the second part, the temporal dynamics of the activity of a neuron population, that is, the activity of retinal ganglion cells recorded in a multi-electrode array was studied. Many studies have focused on describing the activity of neural networks using disordered Ising models, with no regard to the dynamic nature. Treating time as an extra dimension of the system, the temporal dynamics of the activity of the neuron population is modeled. The maximum entropy principle approach is used to build an Ising model with pairwise interactions between the activities of different neurons at different times. Model fitting is performed by a combination of Metropolis Monte Carlo sampling with gradient descent methods. The system is characterized by the learned parameters, questions like detailed balance and time reversibility are analysed and thermodynamic variables, such as specific heat, can be calculated to study critical aspects.

Contents

Chapter 1

Introduction

In the last few years, with the increasing availability of large data sets and the advances in computer power, the study of large networked systems has received the attention of many scientific communities [1, 4, 16, 41, 42]. With its intrinsic interdisciplinary character, network science is a fast-growing field flourishing across many areas of knowledge. Large complex networks arise in a vast number of natural and artificial systems. Technological networks, such as the Internet, are among the most studied networks. The brain is a complex network where neurons are interconnected; in ecological networks, species interact with each other and with elements of the environment. Different forms of relationship are represented by links in social networks. Large infrastructure networks, such as power grids or transport networks, are a major concern to modern society. To understand the behavior and uncover the underlying laws governing these systems, researchers have systematically studied connection patterns and the involved dynamical processes. Complex networked systems have in common the fact that they cannot be well described by the sum of their parts: non-linear features emerge from the collective behavior.

This kind of behavior associated with the large size and dynamic nature of these systems has attracted the attention of the statistical physics community. Statistical mechanics provides a language for connecting the interactions among many microscopic degrees of freedom to the collective macroscopic behavior. In this way, the study of large networked systems has favored the use of techniques related to the analysis of non-linear, equilibrium, and non-equilibrium physical systems, which have generated results of conceptual and practical relevance.

Within this context, the collective behavior of neural networks is a widely studied field with concepts of network science [54, 55]. The brain is genuinely a complex network, with approximately 10^{10} neurons and an even greater number of connections between them. Like other complex systems, its macroscopic features (such as memory, cognition or vision) emerge from simple microscopic interactions (neurons).

A first question we might ask is what kind of dynamical regime can describe some aspects of the collective behavior of such enormous network with astronomical dimensions and with such a complex functionality. Empirical evidences from real neural network measures together with analytical approaches based on statistical physics models and network science suggest that the emergence of collective behavior in such systems might be explained by a critical dynamics [6, 12, 13, 17, 19, 20, 22, 37, 38, 51, 57].

1.1 Criticality

Before continuing with the main discussion, it is worth at this point to define what is criticality. Critical behavior is associated with phase transition phenomena: the macroscopic state behavior, or phase, of the system under study is changed as an external parameter (such as temperature or pressure in physical systems) is varied, and between two phases there is a critical point. The critical point is marked by scaling and universality phenomena, the system properties drastically alters [11]. Hallmarks of criticality are observed such as the divergence of the correlation length, and other power-law divergence, ergodicity breaking and fractal behavior.

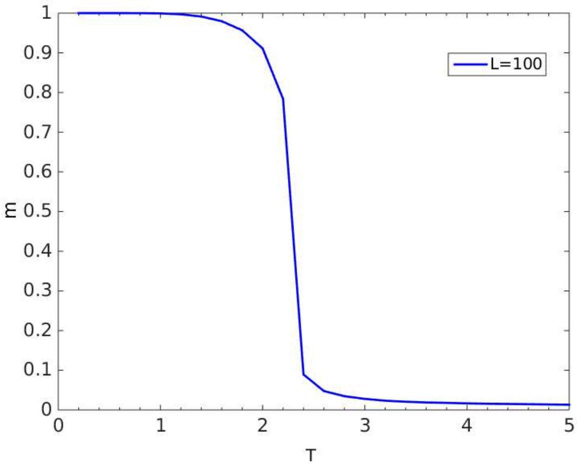
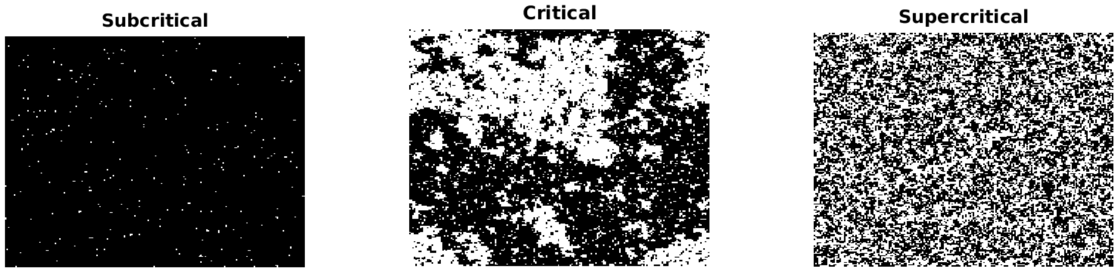


Figure 1.1: Magnetization per spin in function of the temperature from a numerical simulation of an Ising 2-D model in a lattice with size $L = 100$.

Let us consider the ferromagnetic phase transition, which is a second order, or continuous phase transition. Magnetic materials, from a microscopic point of view, are characterized by the presence of spins, which can be regarded as magnetic dipoles. In paramagnetic materials, the magnetization is zero in the absence of external field $h = 0$, while the spins align in the direction of any non-zero magnetic field $h \neq 0$. On the other hand, for ferromagnetic materials, even without the presence of an external field, neighbor spins tend to align in the same direction. This happens when temperature approaches the critical temperature T_c. Below this temperature, the system has a non-zero magnetization. Thus, the magnetization is the *order parameter* of the magnetic system. This is shown in Fig. 1.1, where a numeric simulation of an Ising model was performed for a square lattice of size $L = 100$. The Ising model is a paradigmatic model used to describe not only ferromagnetic systems but many complex systems, such as neural networks. The model will be mathematically defined in Chap. 3 and discussed all along this work. The spins in an Ising model can assume two values (+1 or -1). Disregarding the effects of finite size in Fig. 1.1, we see that the order parameter is zero above the critical temperature ($T_c \approx 2.3$), a paramagnetic phase, and non zero below the critical temperature even with zero field $h = 0$, a ferromagnetic phase. Between these phases there is the critical point.



(a) T $< T_c$ **(b)** $T \approx T_c$ **(c)** T $> T_c$

Figure 1.2: Three snapshots of a system for three different temperatures (subcritial, critial and supercritical) from a numerical simulation of a 2-D Ising model with size $L = 200$.

In this example the order parameter is a function of temperature and the intensity of the interactions. The situations are shown in Fig. 1.2, where three snapshots of a numerical simulation of a two dimensional Ising model, a square lattice with size $L = 200$ (the total number of spins is $N = L \times L$), are presented for three temperatures: subcritial, critical and supercritical. Above the critical temperature, the spins are flipping because of the available energy. These thermal fluctuations counteract the alignment between spins and the average sum of spin directions is zero, which can be regarded as a chaotic phase, the supercritial phase, as in Fig. 1.2c. In contrast, below the critical temperature the spins are aligned and there is a ordered phase, the subcritial phase, as in Fig. 1.2a. Between these two regimes the system lies at the critical point, as in Fig. 1.2b. The spins with the same direction coalesce into large domains. This is the point where the largest fluctuations are observed in the magnetization. The length of spatial correlation between spins diverge and a power-law of this long-range correlation is observed. The divergence of the susceptibility is also observed in a power-law. Even with a small but finite probability, the perturbation of a single spin can propagate through the whole system as an avalanche, reshaping the entire state of the system. It is important to observe that rigorously speaking, the divergence only in the thermodynamic limit ($N \to \infty$). These are amongst the dynamical properties that characterize the critical regime. Translating these terms to neural context, these are suggestive properties to describe the cooperative behavior in neural networks.

1.2 Critical Aspects in Brain Dynamics

The idea discussed in the last section can be transposed to the neural network context by observing how a neural system works. Making a connection with the Ising model, where spins have two possible states (+1 or -1), neurons can be regarded as spins such that they might be spiking (+1) or in silence (-1). This can be extended to brain regions in coarse-grained approaches.

In this way, to have proper functionality, we cannot have a situation when almost all neurons are firing, which would be a chaotic supercritical phase, neither when almost all neurons are in a quiescent state, which would be a subcritical phase. There must be a balance between these two situations. In terms of brain activity, when we are performing some task, different brain areas are being continuously activated and deactivated, which is analogous to the situation seen in Fig. 1.2b, where there is a large fluctuation of domains formation. This is where the critical hypothesis comes in: to perform the complex computations the brain must operate in a critical regime. Critical dynamics provides a tenuous trade-off between a highly active and chaotic phase and a quiescent phase, which would be impractical for living. It has also been argued that criticality implies optimal computational capabilities [31], large network stability [7], maximal variety of memory repertoires [5] and maximal sensitivity to stimuli [29].

1.3 Time Series Comparison

In the first part of this work, a theoretical model for simulating time series from brain activity is developed based on an Ising model [13, 19, 22]. It is then compared with experimental data.

The time series from the brain are measured via functional magnetic resonance (fMRI) in patients following a resting state protocol [18]. The brain is divided into many regions and it is possible to compute the correlation between the activity of these different regions. The experimental data is a work performed by Dr. Prof. Gabriela Castellano and her student Raphael Fernandes Casseb from the Neurophysics group at University of Campinas.

The theoretical model simulates this situation. We start with a two dimensional Ising model an let the system evolve at a given temperature according to a Metropolis Monte-Carlo dynamics, discussed in Chap. 3. After a certain period, there is a time series for each spin. It is possible then to compute the correlations between the time series of each spin. At the end, we have a correlation matrix of the brain and of the network model, and analysis is performed. Quantities related to network theory are measured, like clustering coefficient, degree distribution, average shortest path length, allowing the characterization of the network. In general terms, what is observed in brain data is better explained by the theoretical model at the critical temperature, showing evidences of criticality in brain dynamics.

1.4 Maximum Entropy Models Inference

In the second part of this work, the temporal dynamics of the collective activity of a neuron population recorded directly from physiological data is studied [33, 37]. Real neural networks are complex networks that operate in a non-linear way. Neurons activities are correlated with each other, meaning that the collective behavior is not well formulated by the sum of its parts. In this way, physicists have been using statistical mechanics to study these systems. The information of the neural activity is encoded in a sequence of spikes and silences in time windows, forming a neural code-word. This suggests mapping the state of a neuron to a classical Ising spin, with up and down spin states being spikes and silences, and a code-word as being a state of the system. Studies have shown successful results in explaining how functions emerge collectively from a large neuron population described by a disordered Ising model with pairwise couplings [20, 37, 51].

Many studies that have used the analogy of Ising models and neural networks have focused on the distribution of code-words in a given time window, but with no regards to the time dynamics of the neural activity. To address this issue, in the second part of this work, time is treated as an additional dimension of the system. Thus, a model with pairwise interactions between different neurons in different time windows is built.

The principle of maximum entropy is used to model the activity of retinal ganglion cells recorded in a multi-electrode array [51]. To fit the model, a combination of Monte Carlo sampling with gradient descent method is used. It is possible then to study thermodynamic properties of the system. Questions like time-reversibility of neural activity, criticality and detailed balance violation can be studied. This work was done during a six months research internship at the École Normale Supérieure of Paris under the supervision of Dr. Thierry Mora.

Part I

Correlation Network

Chapter 2

Definitions

2.1 Adjacency Matrix

In general terms, a network is a system that can be represented in mathematical abstract way by a set of nodes, or vertices, and a set of links, or edges, that are pairs of connected nodes. The nodes are the elements of the network, such as neurons in neural networks, people in social networks, or species in ecological networks. The links represent a relation or interaction between the elements. Networks have been widely studied in graph theory and there are many ways to represent them mathematically. In this work, the network will be represented as an adjacency matrix. A network with N nodes will have a adjacency matrix A with $N \times N$ dimension. The elements of the adjacency matrix are defined as

$$A_{ij} = \begin{cases} 1 & \text{if there is a link from j ot i} \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

In the following applications, we make two considerations. First, there are no auto-links, that is, a node does not connect with itself, consequently the main diagonal is zero, $A_{ii} = 0$. Second, the links are not directed, that is, if there is a link from node i to node j, the opposite is true. In terms of the adjacency matrix, $A_{ij} = A_{ji}$, i.e., the adjacency matrix is symmetric, which defines an undirected network.

2.2 Mean Degree

The *degree* k_i of node i is defined as the number of neighbors, or the number of links, this node has k_i. This can be calculated as the sum of the node's respective row in the adjacency matrix (or column in the case were the adjacency matrix is symmetric).

$$k_i = \sum_j A_{ij} \tag{2.2}$$

The *mean degree* of the network, in its turn, is the average degree of the network nodes,

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i \tag{2.3}$$

.

The mean degree is a measure of how connected the network is. In the case were all nodes are connected to each other the mean degree is constant, and we have

$$\langle k \rangle = k = \binom{N}{2} = \frac{N(N-1)}{2}. \tag{2.4}$$

Another important concept regarding the degree definition is the *degree distribution* $P(k)$ of the network, which is a spectrum of how the vertices are related, or interact. In the case of undirected networks, the degree distribution is the probability of choosing a node at random that has degree k.

2.3 Clustering Coefficient

In complex networks as neural networks, we observe a clustering phenomenon which is a tendency of nodes to form a clique around a giving node. In other words, if a node i is connected to node j, and node j is also connected to another node l, there is a high probability that nodes i and l are also connected. This can be quantified in undirected networks by means of the *clustering coefficient* which measures the local group cohesiveness [59]. The clustering coefficient C_i of a given node i, with degree k_i, is defined as the ratio of the number of links e_i between the neighbors of i and the total possible number of such links $\binom{k_i}{2}$,

$$C_i = \frac{e_i}{k_i(k_i-1)/2}. \tag{2.5}$$

In terms of the adjacency matrix e_i is expressed as,

$$e_i = \frac{1}{2}\sum_{jl} A_{ij}A_{jl}A_{li} \tag{2.6}$$

so that,

$$C_i = \frac{\sum_{jl} A_{ij}A_{jl}A_{li}}{k_i(k_i-1)}. \tag{2.7}$$

The clustering coefficient of the whole network is just the average clustering coefficient of the nodes,

$$C = \frac{1}{N}\sum_i C_i. \tag{2.8}$$

2.4 Average Shortest Path Length

A *path* in a network is any sequence of nodes such that there is a link between a pair of consecutive nodes in the sequence. The *length* of a path in a network is the number of links traversed along the path. In terms of the adjacency matrix A, the total number $\mathcal{N}_{ij}^{(r)}$ of paths of length r from j to i is

$$\mathcal{N}_{ij}^{(r)} = (A^r)_{ij} \tag{2.9}$$

where $(\ldots)_{ij}$ denotes the ijth element of the adjacency matrix and A^r is the product of the adjacency matrix by itself r times.

Given two nodes i and j, the *geodesic distance* d_{ij} between them will be defined as the *shortest path distance*, that is, between all paths between i and j in the network the one with the shortest length. In mathematical terms, the geodesic distance between i and j is the smallest value of r such that $(A^r)_{ij} > 0$. Note that there might be more than one path with the shortest length r.

The average distance L_i between node i and all other nodes in the network is

$$L_i = \frac{1}{N-1}\sum_{j,j\neq i} d_{ij}. \tag{2.10}$$

Consequently, the *average shortest path length*, or the characteristic path length of the network is given by

$$L = \frac{1}{N}\sum_i L_i. \tag{2.11}$$

In other words, the average shortest path length is the average of d_{ij} over all possible pair of nodes in the network

$$L = \frac{1}{N(N-1)}\sum_{ij} d_{ij}. \tag{2.12}$$

Like the clustering coefficient, the average shortest path is a measure of connectivity of the network. In fact, if all nodes of the network are connected $L = 1$. It is also worth remarking that when a network has a small average shortest path and relative high clustering coefficient we observe the so-called *small-world effect* [59]. This means that, to go from one node of the network to any other node we must traverse a short path, or the distance between any given pair of nodes is small.

2.5 Diameter

Using the definition of the shortest path distance, the *diameter* of the network is

$$D = \max_{i,j}\{d_{ij}\}. \tag{2.13}$$

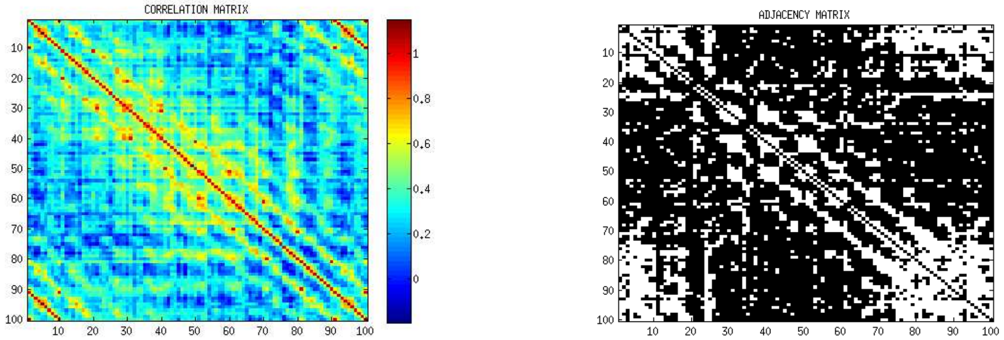
2.6 Correlation Network

For certain systems, nodes as well as links are not self-evident and easily identifiable. This is the case for functional networks of the brain, where time series of the activity of brain regions are measured. One way to approach this situation is considering the correlations between the time series, as done in [17, 19].

The network will be defined by means of the correlation coefficient r between nodes i and j, defined as

$$r(i,j) = \frac{\langle x_i(t)x_j(t)\rangle - \langle x_i(t)\rangle\langle x_j(t)\rangle}{sd[x_i(t)]sd[x_j(t)]} \tag{2.14}$$

which is the Pearson correlation, where $\langle\ldots\rangle$ is the time average, sd is the standard deviation $sd^2[x_i(t)] = \langle x_i^2(t)\rangle - \langle x_i(t)\rangle^2$ and $x_i(t)$ is the temporal series of the activity of node i, which can be the activity of a brain region from data or the time series of a node in the theoretical model (see Chap. 3).



(a) **(b)**

Figure 2.1: Examples of a correlation matrix (a) and a corresponding adjacency matrix (b).

In this framework, for a given correlation threshold ρ, a *link* between nodes i and j is established whenever the correlation between these nodes is greater or equal than the threshold: $r(i, j) \geq \rho \Rightarrow A_{ij} = 1$. By the same way, a *node* i of the network is a site with non-zero number of links: $k_i = \sum_j A_{ij} \neq 0$. This completes the definitions of the network nodes and links. In fig. 2.1 is presented an example of a correlation matrix (Fig. 2.1a) and a corresponding adjacency matrix (Fig. 2.1b) of a system with N nodes.

Chapter 3

Correlation Network Framework

3.1 Brain Data

All the brain data used to compare with the theoretical model in this and the following chapter (Chap. 4) is a work performed by Prof. Gabriela Castellano and her student Raphel Fernandes Casseb from the Neurophysics group at University of Campinas. In the following, for the sake of completeness, we will describe how the data is obtained and processed.

The brain data is measured via functional magnetic resonance (fMRI). Subjects were scanned following a typical resting state protocol, in which the individuals, while lying in the scanner, are asked not to engage in any specific task (mind wondering condition), to remain with their eyes closed, and to avoid falling asleep. Data were acquired for 27 subjects, in a Philips Achieva 3T machine; we obtained a structural (T1-weighted images; isotropic voxels of 1mm^3 and a voxel matrix of 240x240x180; TR/TE = 7/3.2 ms; flip angle = 8°) and a functional MR image (180 volumes; T2*-weighted image; isotropic voxels of 3x3x3 mm^3; gap of 0.6 mm; TR/TE = 2/0.03 s; and a voxel matrix of 80x80x40).

The brain activity is captured with the fMRI exam through the spatio-temporally resolved blood-oxygen-level-dependent (BOLD) signal [32]. The central idea is that neurons do not have an energetic reserve in the form of sugar and oxygen, both needed to perform the neural activity. In this way, the energy (glucose) must be brought quickly by the blood, and the oxygen is transported by the hemoglobin molecules. Through a process called hemodynamic response, blood flow is increased in the areas where neurons are more active. This process causes changes in the local rates of oxyhemoglobin and deoxyhemoglobin, what is captured by the fMRI scan because the the oxy- and deoxyhemoglobin have different magnetic properties.

The brain is divided in regions of interest (ROI), commonly called voxels, that are the tridimensional pixels that conform the image from fMRI exam. The BOLD signal of each voxel is recorded by the fMRI. In the end, we have a temporal series for each voxel, which is the brain activity of each brain ROI. As we are interested in brain activity, our analysis is restricted to the cerebral gray matter, since it is the tissue where the cell body of the neurons, responsible for information processing, are located. Some preprocessing steps (realignment of functional volumes; corregistration between the structural and the functional image; segmentation; normalization; and smoothing) are necessary, because the fMRI signal is weak and noisy. Afterwards, we apply a mask to exclude other tissues than gray matter; regress out movement parameters and spurious signal (from white matter and cerebrospinal fluid); remove linear trends of the signal; and use a bandpass filter to obtain data exclusively in the range of 0.008 and 0.01 Hz [8].

The time series obtained after this processing are then used to compute the correlation between each temporal series of the BOLD signal of each voxel by Eq. (2.14), where $x_i(t)$ in this case is the BOLD signal of voxel i. Having the correlation matrix, the adjacency matrix A is computed in function of a correlation threshold ρ with the method explained in Sec. 2.6, in order to obtain the correlation network.

3.2 Theoretical Model

The theoretical model is a correlation network built from a two dimensional Ising model [25]. The Ising model is a cornerstone in statistical physics approaches in complex systems. Its application goes way beyond physics, being the basis of modeling in several areas, such as social sciences, epidemics or neural networks. The Ising model is a structural lattice where in each site a spin is placed. The lattice can have many structures with different topologies, and may represent different situations where the links between sites can be real (chemical) bonds, some relation or statistical interactions. In this work the lattice is a two dimensional regular square lattice with periodic boundary conditions where each site has four neighbors. Each node is a spin s_i that can assume two values, $+1$ or -1. This can be regarded as an approximation to describe the spin direction, that is, $+1$ “up”, -1 “down”.

In the case of neural networks, it can be interpreted as +1 (active) and -1 (silence). The lattice has dimension L and the total number of spins is $N = L \times L$.

A system configuration is described by the states of all spins in that moment,

$$s = \{s_1, s_2, \ldots, s_N\}. \tag{3.1}$$

The interaction between spins is short-ranged, in order that the model includes only terms proportional to $s_i s_j$ in the Hamiltonian, coupled by pairwise interactions J_{ij}. For each configuration the system has an associated energy described by the function, or Hamiltonian

$$E = -\sum_{\langle ij \rangle} J_{ij} s_i s_j \tag{3.2}$$

where $\langle ij \rangle$ indicates that i and j are nearest neighbors. The couplings J_{ij} represent the energy reduction if spins i and j are aligned, that is $J_{ij} > 0$. Thus, the above Hamiltonian tells us that the minimum is at $T = 0$, when all spins are aligned, all spins up ($s_i = 1$) or all spins down ($s_i = -1$). In the model used in this first part we consider that all couplings are constant and equal to 1, $J_{ij} = J = 1$, so that

$$E = -\sum_{\langle ij \rangle} s_i s_j. \tag{3.3}$$

The model described by the Hamiltonian of Eq. (3.3) has analytical solution in two dimensions [23, 45], and the critical temperature for $J = k_B = 1$ is $T_c \approx 2.3$, where k_B is the Boltzmann constant.

The final goal of the model is to build an adjacency matrix, as defined in sec. 2.1. Given a certain temperature, we let the system defined above by the Ising model evolve according to the Metropolis Monte-Carlo algorithm [35] and store different configurations of the system. After a certain period of time there will be a time series for each spin $\{s_1(t), s_2(t), \ldots, s_N(t)\}$, and it is possible to calculate the correlation $r(i, j)$, as defined in Eq. (2.14), between each pair of spins in order to have a correlation matrix. Once we have the correlation matrix of the spins activity, we can compute the adjacency matrix given a correlation threshold, which is a correlation network, as defined in sec. 2.6.

3.3 Monte-Carlo Sampling

In the theoretical model, many samples, or system configurations, were needed in order to have temporal series of the spins. Those samples were obtained with a Monte-Carlo method. The algorithm used to implement the Monte-Carlo sampling was the Metropolis algorithm, which was introduced by Nicolas Metropolis and his co-workers in a 1953 paper on simulations of hard-sphere gases [35].

The Monte-Carlo method is used to generate samples of system states, that is, given a certain value of the external parameter, such as temperature, the algorithm generates a configuration of the system that corresponds to some thermodynamic state with a certain probability. In this way, if the system is in an initial state μ, the transition probability of generating a new state ν from the old state μ is $P(\mu \to \nu)$. To build a Monte-Carlo algorithm, the transition probability is broken down into two parts [43],

$$P(\mu \to \nu) = g(\mu \to \nu) R(\mu \to \nu), \tag{3.4}$$

where $g(\mu \to \nu)$ is the selection probability and $R(\mu \to \nu)$ is the acceptance probability. The selection probability is the probability that the algorithm will generate a new target state ν given an initial state μ. The acceptance probability, in its turn, says that, given that the system is at the initial state μ, the new configuration ν will be accepted with probability $R(\mu \to \nu)$, otherwise the system remains in the old configuration μ.

In the Metropolis algorithm, it is common to flip a single spin every iteration of the algorithm to generate a new target state. An algorithm that uses this procedure is said to have a single-spin-flip dynamics. The selection probabilities $g(\mu \to \nu)$ in the Metropolis algorithm are all chosen to be equal for each of the possible states μ, and zero for all the other states. In a system with N spins, the algorithm can choose between N of the spins to flip. Thus, the selection probability is given by

$$g(\mu \to \nu) = \frac{1}{N}. \tag{3.5}$$

The Metropolis algorithm is then characterized by the acceptance probability. To guarantee that the stationary distribution of the Markov chain is canonical, we must have the detailed balance condition, which is

$$p_\mu P(\mu \to \nu) = p_\nu P(\nu \to \mu). \tag{3.6}$$

where p_μ is the probability that the system is at a given state μ. Equation (3.6) states that the rate at which the system makes transition from a given state μ to a new state ν must be equal to the rate at which it makes transition from ν to μ. We wish that the equilibrium distribution to be the Boltzmann distribution, so that we choose the values of p_μ to be the Boltzmann probabilities. The equation of the detailed balance condition tells us that the transition probabilities should satisfy

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{p_\nu}{p_\mu} = e^{-\beta(E_\nu - E_\mu)}. \tag{3.7}$$

With the selection probabilities (3.5), the detailed balance condition, Eq. (3.7) takes the form

$$\frac{P(\mu \to \nu)}{P(\nu \to \mu)} = \frac{g(\mu \to \nu)R(\mu \to \nu)}{g(\nu \to \mu)R(\nu \to \mu)} = \frac{R(\mu \to \nu)}{R(\nu \to \mu)} = e^{-\beta(E_\nu - E_\mu)}. \tag{3.8}$$

Now, the the acceptance probabilities must be chosen to satisfy this equation. Given a constraint like (3.7), to produce the most efficient algorithm the acceptance ratios must be maximized. This is done by setting the largest of the two ratios the largest value – namely 1 – and adjusting the other one to satisfy the constraint. Given the variation of energy between states μ and ν, $\Delta E = E_\nu - E_\mu$, the optimal algorithm is defined by

$$R(\mu \to \nu) = \begin{cases} e^{-\beta\Delta E} & \text{if } \Delta E > 0 \\ 1 & \text{otherwise,} \end{cases} \tag{3.9}$$

where β is the inverse temperature, for $k_B = 1$ we have $\beta = 1/T$. What Eq. (3.9) says is that whenever the energy of the new selected state ν is lower than the old one, the transition to the new state will always be accepted. Otherwise, there still a chance to accept a new state with higher energy with probability given by the Boltzmann weights $e^{-\beta\Delta E}$, which is a function of the temperature. This defines the Metropolis algorithm used to sample the system with a Monte-Carlo method.

3.4 Algorithm

The Metropolis algorithm then consists in flipping a spin and testing the new configuration according to Eq. (3.9). Every time a spin is flipped, which is a Monte-Carlo step, the change in energy must be calculated. So, in a Monte-Carlo step a spin k is chosen at random and then the algorithm decide to flip it or not. If the system is at an initial state μ with energy E_μ before flipping the spin and the energy of the new state ν is E_ν, the change in energy ΔE due to the flip of a spin k is

$$\begin{aligned} \Delta E &= E_\nu - E_\mu \\ &= -\sum_{\langle ij \rangle} s_i^\nu s_j^\nu + \sum_{\langle ij \rangle} s_i^\mu s_j^\mu \\ &= -\sum_{i \ n.n. \ k} s_i^\mu (s_k^\nu - s_k^\mu). \end{aligned} \tag{3.10}$$

In the last line, 'i $n.n.$ k' means the sum is over the i nearest neighbors of spin k, and it was also considered the fact that only the spin k flips and all the others spins remain unchanged, that is, $s_i^\nu = s_i^\mu$ for $i \neq j$. However, if $s_k^\mu = -1$, then after flipping, $s_k^\nu = +1$ and $(s_k^\nu - s_k^\mu) = +2$. On the other hand, if $s_k^\mu = +1$, then after flipping, $s_k^\nu = -1$ and $(s_k^\nu - s_k^\mu) = -2$. Thus, we can write

$$(s_k^\nu - s_k^\mu) = -2 s_k^\mu \tag{3.11}$$

and the change in energy is

$$\Delta E = 2 s_k^\mu \sum_{i \ n.n. \ k} s_i^\mu. \tag{3.12}$$

The implementation of the Metropolis algorithm is the following

1. Choose a site (k) at random;

2. Calculate the change in energy ΔE due to the flip of spin s_k;

3. If $\Delta E \leq 0$, accept the move and flip the spin: $s_k \to -s_k$;

4. Else, if $\Delta E > 0$, flip the spin with probability $p = e^{-\beta \Delta E}$;

This is a Monte-Carlo step. For a system with N spins, when N Monte-Carlo steps are performed, it is said that the algorithm completed one sweep of the lattice, which is on average giving the chance to every spin to flip. In the sampling process, for every sweep, one configuration of the system is stored. If we want M configurations of the system then the algorithm will perform M sweeps in the sampling process. The whole process is given by

1. Sampling

   Initialize the system: $s_i = \pm 1$ for $i = 1, \ldots, N$;

   Thermalization: N^2 sweeps

   Monte-Carlo sampling → M sweeps → temporal series

$$\{s_i(t)\} \quad i = 1, \ldots, N \quad t = 1, \ldots, M \tag{3.13}$$

2. Correlation matrix

$$r(i,j) = \frac{\langle s_i(t) s_j(t) \rangle - \langle s_i(t) \rangle \langle s_j(t) \rangle}{\sigma[s_i(t)]\sigma[s_j(t)]}; \tag{3.14}$$

3. Adjacency matrix (ρ)

$$A_{ij} = \begin{cases} 1 & \text{if} \quad r(i,j) \geq \rho \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

Chapter 4

Results

4.1 Correlation Distribution

Three theoretical simulations were performed, one with the system at a subcritical temperature ($T = 2$), one with the system at a supercritical temperature ($T = 3$), and one at the critical temperature ($T = 2.3$). In all cases, the lattice had a size of $L = 100$, that is, a lattice with $N = 10000$ spins. The sampling process was performed in order to have 2000 configurations of the system.

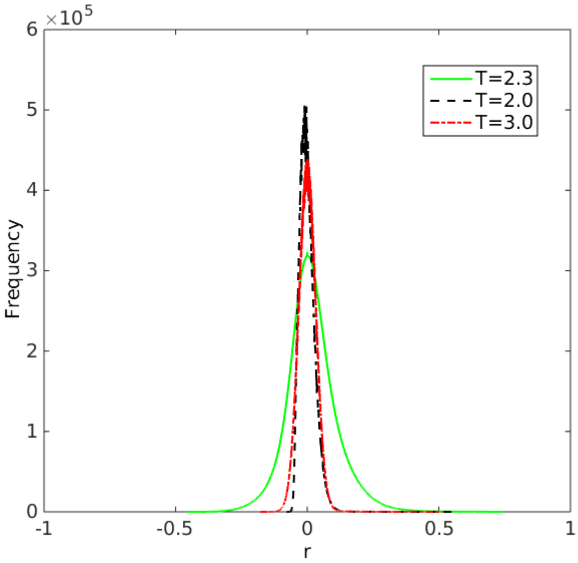


Figure 4.1: Correlation distribution for three temperatures: subcritical (black), critical (green) and supercritical (red).

The plot in Fig. 4.1 shows the distributions of the correlations between the spin temporal series for the three temperatures. It possible to observe that for the sub- and supercritical regimes the behavior of the correlations are similar. However, at the critical temperature there is a noticeable increase in the variance of correlations. This behavior is explained by the fact that it is only at the critical point that the spins with the same direction coalesce and large domains emerge, the system experience the largest fluctuation in this regime.

4.2 Mean Degree vs. Correlation Threshold

In Sec. 2.6 the correlation network was defined. The correlation network is represented by its corresponding adjacency matrix. For each value of the correlation threshold ρ, a different adjacency matrix is obtained and for each adjacency matrix the corresponding correlation network has a different mean degree $\langle k \rangle$.

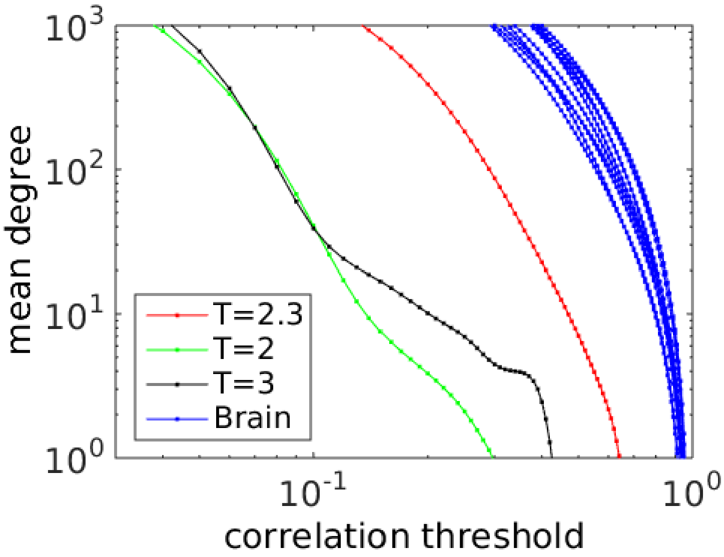
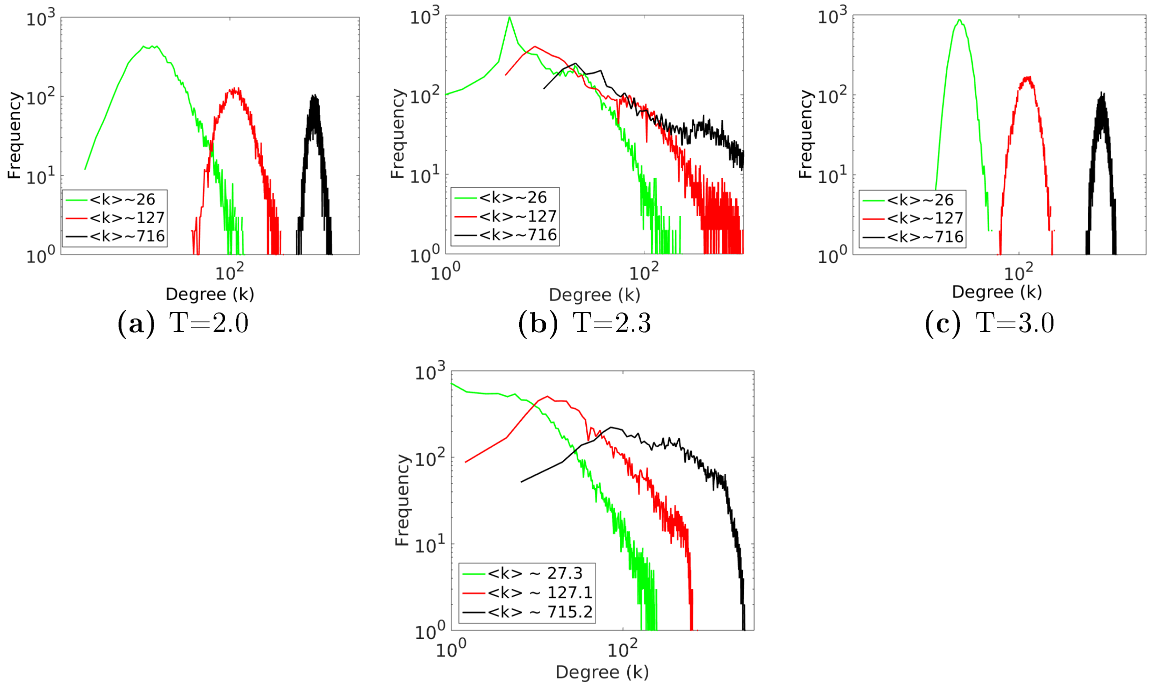


Figure 4.2: Mean degree $\langle k \rangle$ in function of the correlation threshold ρ for the theoretical model network in three temperatures (subcritical (green), critical (red) and supercritical (black)), and for the brain network of eleven subjects (blue curves).

The relation between the correlation threshold and the mean degree of the network is shown in Fig. 4.2. The blue curves are brain networks of eleven subjects and the other curves are the Ising correlation network at three temperatures, subcritical (green), critical (red) and supercritical (black). The curves are related by an inverse relation, that is, as ρ grows the mean degree gets smaller. In other words, higher correlation thresholds generate networks with fewer connections and thus with a smaller mean degree. This pattern is observed in the theoretical model and in the brain curves as well. It is also worth noticing that the red curve corresponding to the critical temperature is the one that has the closer behavior to the brain set of curves.

4.3 Degree Distribution

The degree distribution of the network is one of the most important characteristics of the network. It is the probability $P(k)$ of choosing a vertex of the network at random and that this vertex has degree k.



(a) T=2.0 **(b)** T=2.3 **(c)** T=3.0

(d) Brain

Figure 4.3: Degree distributions of the theoretical model at three temperatures: (a) subcritical, (b) critical and (c) supercritical, for networks with three mean degrees: $\langle k \rangle \approx 26$ (green), $\langle k \rangle \approx 127$ (red), and $\langle k \rangle \approx 716$ (black). At the bottom, (d) degree distribution of the brain network with $N = 12699$ and three mean degrees: $\langle k \rangle \approx 27$ (green), $\langle k \rangle \approx 127$ (red), and $\langle k \rangle \approx 715$ (black).

The degree distributions of the theoretical model network are shown in the top three plots of Fig. 4.3 and the brain network distributions in the bottom. The degree distribution of the theoretical model are presented for three mean degrees, $\langle k \rangle \approx 26$ (green), 127 (red), and 716 (black), by setting appropriate values of ρ. In the same way, the bottom graph shows the degree distribution of the brain network with $N = 12699$ for $\langle k \rangle \approx$ 27 (green), 127 (red), and 715 (black). As expected, a drastic change in the behavior of the theoretical Ising model is observed at the critical temperature. In Fig. 4.3b, at the criticality the system shows power-law tails with a cutoff, while at the subcritical temperature in Fig. 4.3a and in the supercritical temperature in Fig. 4.3c the system shows a Poisson distribution. Above all, the behavior observed in the theoretical model is also observed in the distribution of the brain network.

4.4 Network Metrics

In Chap. 2 some network metrics were defined. Here the Ising correlation network metrics are compared with the brain network. The metrics were calculated using the Brain Connectivity Toolbox [50]. The same metrics are also presented in the case of a random network, that is, a randomized network with its degree distribution preserved.

Table 4.1: Network metrics of the Ising correlation network.

T	N	$\langle k \rangle$	C	L	D	C_{rand}	L_{rand}	D_{rand}
				Ising Network				
2.0	10000	197	0.078	2.06	3	0.022	2.01	3
2.3	10000	198	0.544	4.2	16	0.148	2.25	4
3.0	10000	195	0.091	2.09	3	0.021	2.01	3
				Brain Network				
	12699	198	0.549	4.13	11	0.0658	2.26	4

The measures presented in Table 4.1 are the average degree $\langle k \rangle$, the clustering coefficient C, the average shortest path length L, and the diameter D. Four situations are considered in table 4.1, the Ising model in three regimes (critical, sub- and supercritical), and the brain network. In all cases, the correlation thresholds ρ were set in order to have networks with the same average degree. The most remarkable aspect is the fact that the theoretical network turns out to show a small-world behavior at criticality. In this regime, a substantial grow is observed on the clustering coefficient while the average shortest path length, in comparison, almost remains constant and the diameter also increases. This is a consequence of the dynamics of the network, the fact that the system is at the critical temperature. The same effect is not observed when the system time series are randomized. Looking at the brain network metrics, they are in reasonable agreement with what is observed in the theoretical model at criticality, a high clustering coefficient and a comparatively small average shortest path length.

The brain has a similar dynamics to the critical dynamics observed in the last sections [12, 13, 19]. In order to compute tasks at the same time, even with the person at rest (it is impossible to shutdown all our senses), some areas of the brain are being activated while other areas must deactivate. This is something needed, there must be such a balance. It is not viable to a healthy living animal to have all neurons in a quiescent state, this would mean that the brain has stopped, neither all neurons in a super active state when they are all firing. Although this might be intuitively trivial, the way the brain achieves this balance is something complex and still lacking a good explanation.

Part II

Maximum Entropy Models

Chapter 5

Maximum Entropy Models Framework

The principle of maximum entropy (MaxEnt) was first announced by E. T. Jaynes in two papers in 1957 [26, 27], in which he makes a connection between information theory and statistical mechanics. He was able to derive important probability distributions in statistical mechanics by maximizing the Shannon entropy [53] subject to appropriate constraints. The method consist in inferring probability distributions consistent with limited information and otherwise has the least bias as possible with the other degrees of freedom.

The main idea is that, in some situations, it is possible to accurately obtain measures from data, such as means, variances, etc., even without having the distribution that characterizes the system. In this context, the MaxEnt principle is a strategy to find a distribution consistent with these observed measures, but otherwise has as little structure as possible.

Another important concept is the Kullback-Leibler divergence [30] that measures a discrepancy, or the difference, between two distributions. The idea is to minimize the distance between the real distribution of the system and the model distribution that maximizes the entropy, by means of the empirical expected values. This concept is used to implement the algorithm to find the model distribution.

5.1 Entropy

Entropy is a basic concept of information theory. It is a measure of uncertainty or randomness of the probability distribution of the system. In terms of coding length, it is closely related to the code length needed to characterize a random variable. The entropy

S for a discrete random variable X with a distribution P is defined as

$$S[P] = -\sum_i P(X = a_i) \log P(X = a_i) \tag{5.1}$$

where a_i are the possible values of X.

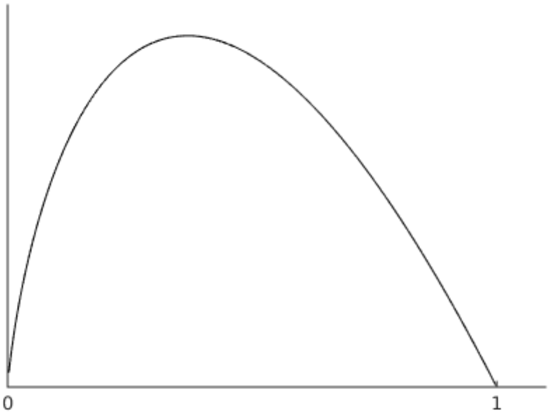


Figure 5.1: The function f plotted in the interval [0,1]

An important property of the entropy is its concavity. Consider the function f

$$f(p) = -p \log p, \qquad \text{for } 0 \leq p \leq 1 \tag{5.2}$$

This is a nonnegative function that is zero for $p = 0$ and $p = 1$, and it is positive for values in between, as plotted in Fig. 5.1. Moreover, we have $f'(p) = -(1 + \log p)$ and $f''(p) = -1/p$. Thus, $f(p)$ is a concave function of p. Since the sum of concave functions is a concave function, the entropy $S[P] = \sum_i f(P(X = a_i))$ is a concave function. This is a very useful property, because a local maximum will be the global maximum for a concave function subject to linear constraints.

5.2 Maximum Entropy

In situations in which data is limited and the system has many degrees of freedom, getting a good estimate of the probability distribution from data can be impractical. Even though, it might be possible to estimate with accuracy the average of chosen observables that describe important properties of the system. In this way, the goal of the maximum entropy method is to build a model for the distribution $P(\sigma)$ over the states of the system with

N elements, $\sigma \equiv \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$, that generate the correct values of the observables, but otherwise the distribution is as random as possible [9, 10, 28, 36, 39, 47]. As an ingredient there are M samples $\{\sigma^1, \sigma^2, \ldots, \sigma^M\}$ from which the averages of the observables are calculated. The observables will be represented by $f \equiv \{f_1, f_2, \ldots, f_k\}$ where $f_\mu = f_\mu(\sigma)$. This is an optimization problem where $P(\sigma)$ is the solution from maximizing the entropy subject to constraints that give the correct expectation values.

Lets define the real, unknown, distribution of the system as

$$P_r(\sigma) = \frac{1}{M} \sum_{n=1}^{M} \delta(\sigma, \sigma^n) \tag{5.3}$$

where $\delta(a, b)$ is the Kronecker delta that is equal to one when $a = b$ and zero otherwise.

The maximum entropy problem, then, is to find the model distribution $P_m(\sigma)$ that maximizes the entropy

$$S[P_m] = -\sum_{\sigma} P_m(\sigma) \log P_m(\sigma) \tag{5.4}$$

subject to the normalization condition

$$\sum_{\sigma} P_m(\sigma) = 1, \tag{5.5}$$

and the constraints

$$\langle f_\mu(\sigma) \rangle_m = \langle f_\mu(\sigma) \rangle_r. \tag{5.6}$$

where $\langle \ldots \rangle_m$ stands for the expectation value with respect with the model distribution and $\langle \ldots \rangle_r$ with respect with the real distribution.

Using the technique of the Lagrange multipliers to maximize the entropy $S[P_m]$, we introduce the quantity $\tilde{S}$, with a Lagrange multiplier g_k for each of the k constraints given by Eq. (5.6) and another Lagrange multiplier λ for the normalization condition Eq. (5.5).

$$S[\tilde{P_m}] = -\sum_{\sigma} P_m(\sigma) \log P_m(\sigma) - \lambda \Big[\sum_{\sigma} P_m(\sigma) - 1 \Big] - \sum_{\mu=1}^{k} g_\mu \Big[\sum_{\sigma} P_m(\sigma) f_\mu(\sigma) - \langle f_\mu \rangle_r \Big] \tag{5.7}$$

where we used the fact that $\langle f_\mu \rangle_m = \sum_\sigma P_m(\sigma) f_\mu(\sigma)$.

Maximizing $\tilde{S}$,

$$0 = \frac{\partial \tilde{S}}{\partial P_m(\sigma)} = -\log P_m(\sigma) - (1+\lambda) - \sum_{\mu=1}^{k} g_\mu f_\mu(\sigma) \tag{5.8}$$

Rearranging the last equation,

$$\log P_m(\sigma) = -(1+\lambda) - \sum_{\mu=1}^{k} g_\mu f_\mu(\sigma) \tag{5.9}$$

$$P_m(\sigma|g) = \frac{1}{Z(g)} \exp\left[-\sum_{\mu=1}^{k} g_\mu f_\mu(\sigma)\right] \tag{5.10}$$

where $\{g_1, g_2, \ldots, g_k\} \equiv g \in \mathbb{R}^k$ are the Lagrange multipliers. As usual, from the normalization condition (5.5), we have that $\exp(1+\lambda) = \sum_\sigma \exp\left[-\sum_\mu g_\mu f_\mu(\sigma)\right] = Z(g)$ is the partition function that normalizes the distribution to one. Note that the constraints conditions given by Eq. (5.6) are obtained by differentiating $\tilde{S}$ with respect to the Lagrange multipliers g_μ. Thus, Eq. (5.10) gives us the form of the maximum entropy distribution.

At this point, it is worth noting that we can make use of an analogy with statistical mechanics to introduce the concept of an energy function or the Hamiltonian of the system. If the only constraint imposed was the expected value of the energy, $f(\sigma) = E(\sigma)$, the Boltzmann distribution, $P_m(\sigma) = Z^{-1} \exp -\beta E(\sigma)$, is recovered. Where, in this case, $\beta = 1/kT$ is the inverse temperature. In a more general way, the sum inside the exponential in Eq. (5.10) can be interpreted as an energy function: $E(\sigma) = \sum_{\mu=1}^{k} g_\mu f_\mu(\sigma)$.

It is still missing to adjust the constants g_μ in order that the model distribution $P_m(\sigma)$ gives the correct expectation values measured from data. For this end, we need to match the expectation values given by model distribution, $\langle f_\mu(\sigma)\rangle_m = \sum_\sigma P_m(\sigma) f_\mu(\sigma)$, with the empirical expectation values $\langle f_\mu(\sigma)\rangle_r$.

There is a unique set of Lagrange multiplier that satisfy all the constraints. But, this is a computationally difficult inverse problem. Classically, we see direct problems in statistical mechanics, i.e., given a set of parameters $g \equiv \{g_1, g_2, \ldots, g_k\}$, we ask what is the probability of finding the system in a given state $\sigma \equiv \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$, $P(\sigma|g)$. The inverse problem is then: if we observe M independent samples of the system $\{\sigma^1, \sigma^2, \ldots, \sigma^M\}$, how do we estimate the parameters g that characterizes the Hamiltonian? The parameters cannot be determined exactly with a finite amount of data. We need a probabilistic approach to estimate them. This is done by maximizing the likelihood function $P(g|\sigma)$ or, in a more convenient way, the log-likelihood. In other words, we need to find the set of parameters g that make the obtained empirical values most likely.

In statistical mechanics, the logarithm of the partition function is the free energy of the system and the derivative of the the free energy is a expectation value.

$$\langle f_\mu(\sigma)\rangle_m = -\frac{\partial \log Z(\sigma)}{\partial g_\mu}. \tag{5.11}$$

Thus, matching with the empirical expectation values means solving the equation

$$\langle f_\mu(\sigma)\rangle_r = \frac{1}{M}\sum_{n=1}^{M} f_\mu(\sigma^n) = -\frac{\partial \log Z(\sigma)}{\partial g_\mu} \tag{5.12}$$

where the upper index in σ means the nth sample.

Eq. (5.12) can be written as

$$\begin{aligned}\frac{1}{M}\sum_{n=1}^{M} f_\mu(\sigma^n) &= \frac{1}{M}\sum_{n=1}^{M}\frac{\partial}{\partial g_\mu}\sum_{\nu=1}^{k} g_\nu\, f_\nu(\sigma^n) = -\frac{\partial \log Z(g)}{\partial g_\mu}\\ 0 &= \frac{\partial}{\partial g_\mu}\frac{1}{M}\sum_{n=1}^{M}\left[-\log Z(g) - \sum_{\nu=1}^{k} g_\nu\, f_\nu(\sigma^n)\right]\\ &= \frac{\partial}{\partial g_\mu}\frac{1}{M}\sum_{n=1}^{M}\log P_m(\sigma^n)\end{aligned}$$

$$\therefore \qquad \frac{\partial}{\partial g_\mu}\log\prod_{n=1}^{M} P_m(\sigma^n) = 0 \tag{5.13}$$

The quantity being maximized in Eq. (5.13) is the log-likelihood $\mathcal{L}$ that the experimental data was produced by the model, and it can be rearranged in a convenient way.

$$\begin{aligned}\mathcal{L} &= \log\prod_{n=1}^{M} P_m(\sigma^n)\\ &= \sum_{n=1}^{M}\log P_m(\sigma^n)\\ &= \sum_{n=1}^{M}\sum_{\sigma}\delta(\sigma,\sigma^n)\log P_m(\sigma)\\ &= M\sum_{\sigma} P_r(\sigma)\log P_m(\sigma)\end{aligned}$$

$$\therefore \qquad \mathcal{L} = M\{S[P_r(\sigma)] - D_{KL}(P_r|P_m)\} \tag{5.14}$$

where $D_{KL}(P_r|P_m)$ is a measure, from information theory, known as the Kullback-Leibler (KL) divergence (Appendix A). It is a measure of the difference, or divergence, between the distribution for which the system is built to work in a optimal way, the real distribution in this case $P_r(\sigma)$, from an approximate or model distribution $P_m(\sigma)$. The KL divergence is defined as

$$D_{KL}(P_r|P_m) = \sum_{\sigma} P_r(\sigma) \log \frac{P_r(\sigma)}{P_m(\sigma)}. \tag{5.15}$$

Technically, the KL divergence measures the difference between the entropy of the real distribution, $S[P_r]$, and the cross-entropy of the real distribution and the model distribution, $S[P_r, P_m]$, which can easily been seen from Eq. (5.15).

Therefore, maximizing the log-likelihood, Eq. (5.13), is equivalent to minimizing the KL divergence

$$\frac{\partial D_{KL}(P_r|P_m)}{\partial g_\mu} = \langle f_\mu(\sigma)\rangle_m - \langle f_\mu(\sigma)\rangle_r \tag{5.16}$$

which ensures that the constraints are satisfied at the minimum. Between the properties of the KL divergence, there is the convexity property (Appendix A) that establishes that the local minimum is the global minimum [28]. To implement an algorithm that learns the appropriate parameters, the gradient descent method is used [24].

5.3 Gradient Descent

The gradient descent approach consists in minimizing a function with respect to a parameter array or set. In our case, we want to minimize the KL divergence $D_{KL}(P_r|P_m)$ with respect to the Lagrange multipliers g as in Eq. (5.16). Starting from an initial point $g(0)$, the function $D_{KL}(P_r|P_m)$ is iteratively minimized by calculating its gradient at this point and then moving to the steepest descent direction by a suitable distance. At this new point, the procedure is repeated, and so on. For some $t > 0$, we have

$$g_\mu(t) = g_\mu(t-1) - \eta \left.\frac{\partial D_{KL}}{\partial g_\mu}\right|_{g_\mu = g_\mu(t-1)} \tag{5.17}$$

where the gradient is computed at the point $g_\mu(t-1)$. The parameter η is the 'learning rate' and gives the length of the step in the gradient direction. The iteration continues until some convergence criteria is reached. The convergence criteria will be defined in Sec. 7.2.

Using Eq. (5.16), it is more convenient to represent the update rule (5.17) in conformity with programming languages,

$$g_\mu \leftarrow g_\mu + \eta \, [\, \langle f_\mu(\sigma) \rangle_r - \langle f_\mu(\sigma) \rangle_m \,] \tag{5.18}$$

The inverse problem is in fact broken down into two tasks: (1) the direct problem of estimating the expected values $\langle f_\mu(\sigma) \rangle_m$ with respect to the model distribution $P_m(\sigma)$ for a given set of parameters $\{g_\mu\}$; and (2) the inverse problem of implementing an update rule such as (5.18) that converges to the parameters $\{g_\mu\}$ that give the correct empirical values $\langle f_\mu(\sigma) \rangle_r$, which is a machine learning process. In principle, the direct problem should be summed over all possible states of the system σ, which makes it computationally costly. Some approximate methods have been proposed to get around this problem. In this work, Monte Carlo algorithms are used, but approximate analytic methods, such as high temperature expansion or message-passing algorithms, were also developed and shown good results. This subject will be addressed in Chap. 7.

Chapter 6

Model Framework

In this chapter the theory from Chap. 5 is applied to the neural network. It is shown that the system can be represented by an Ising model as the least structured model within the experimental data framework.

Time is treated as an extra dimension in order to study the time dynamics of the network. For this end, couplings between different neurons at different time moments are defined.

As mentioned at the end of Chap. 5, the problem is broken down into two tasks, a direct and an inverse problem. The expected values of the observables and the respective Lagrange multipliers are defined with respect to the data from the neural network.

6.1 Experimental data

To understand how the collective behavior of neurons works, it is crucial having their dynamics recorded simultaneously. Thus, the experimental data consists on physiological data of retinal ganglion cells recorded from a multi-electrode array, which captures the activity of all neurons simultaneously [33]. The recordings were performed on the rat retina. The retina is isolated from the rat's eye and the ganglion cell side is put against the multi-electrode. This work is done by Olivier Marre at the Vision Institute in Paris.

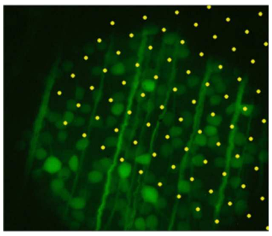


Figure 6.1: Example of a retinal patch over the multi-electrode.

The retina is an excellent system to record the neural population activity. It is a built-in system that has a layer that captures visual inputs, encodes this information in the mid layers and outputs the information in the form of a neural code, with spikes and silences. The activity of nearby firing neurons are captured by one or more electrode. Hence, a deconvolution step is needed. For this end, a spike sorting algorithm is used to sort the record.

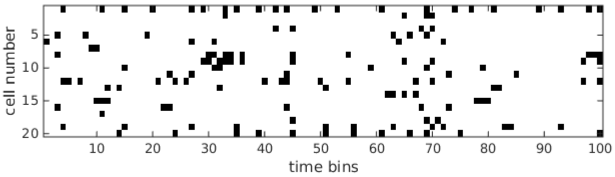


Figure 6.2: Raster with temporal sequences of spikes and silences. Columns are time bins and rows are neuron activities.

In the end, what we have is a raster, which is a temporal sequence of spikes and silences of the neural activity as in Fig. 6.2. Each column is a time bin and each row is the activity of a neuron.

6.2 Connection with statistical physics

The system is a network of N nodes. Each node i can been regarded as a spin that represents the activity of a neuron and it has a binary stochastic variable associated with its state σ_i. The values it can assume are $\sigma_i = 0$ for silence or $\sigma_i = 1$ for a spike. It can also be seen as a spin up or down. At a given moment, we have a code word or the state of the system $\sigma \equiv \{\sigma_i\}$ where $i = 1, 2, \ldots, N$.

Associated with the state σ we can define an energy function or the Hamiltonian of the system $E(\sigma)$. In the framework of maximum entropy models, we need to chose functions $f_\mu(\sigma)$ that describe interesting features of the system. Given that the state of the neurons are binary, the system has 2^N possible states. Our hope is to find a much smaller number of functions, $\mu = 1, 2, \ldots, k$ with $k << 2^N$, which will be sufficient to capture the collective behavior of the neural population.

The idea of the maximum entropy principle is to build models that are consistent with some measured observables from data, but otherwise have the little structures as possible. In this sense, the first term we can think about for the energy function is the independent firing of a neuron σ_i caused by an external input, or field, h_i. Thus, our first choice of function is equivalent to constraining the mean firing rate $\langle \sigma_i \rangle_r$ calculated from data.

Studies have shown that the activity of neurons are correlated [20, 51, 56, 57], this is also an intuitive idea for a complex network as the neural system. Thus, the second simplest function is a pairwise correlation between two neurons $\sigma_i \sigma_j$ coupled by a link J_{ij}. This is equivalent to constraining the correlation between these two neurons $\langle \sigma_i \sigma_j \rangle$. Then, the energy function is given by

$$E(\sigma) = -\sum_i h_i \, \sigma_i - \sum_{i<j} J_{ij} \, \sigma_i \sigma_j. \tag{6.1}$$

The Eq. (6.1) is the Hamiltonian of a pairwise Ising model. In statistical physics, the first term is the Hamiltonian for paramagnetism and the second accounts for ferromagnetism. Hence, we can make use of analogies from the Ising model to study our model of the neural network, although what we really have is mathematical correspondence only. In this theoretical framework, the statistical distribution of code words is studied in a given time window. However, the activity of a neuron has correlation in time, with itself and with other neurons. To study the time dynamics, time is regarded as an additional dimension of the system.

6.3 Time as an extra dimension

Treating time as an extra dimension means that the states of the system are the entire multi-neuron spike train $\sigma \equiv \{\sigma_{i,t}\}$ where $i = 1, \ldots, N$ and $t = 1, \ldots, L$. The network is regarded not as system of N spins, but as a network of size $N \times L$, where L is the number of time bins. This is done by establishing couplings between the activities of different neurons at different moments. For each neuron i at time t it is associated the activity $\sigma_{i,t}$ (=0 or 1). Then, we have a coupling $J_{ij,\tau}$ between a neuron i at time t and a neuron j at time $t + \tau$. The parameter τ can be seen as distance in the time dimension and its maximum absolute value is defined as u.

Following the same reasoning of last section, the energy function is defined as

$$E(\sigma) = -\frac{1}{2}\sum_{t=1}^{L}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\tau=-u}^{u} J_{ij,\tau}\,\sigma_{i,t}\,\sigma_{j,t+\tau} - \sum_{t=1}^{L}\sum_{i=1}^{N} h_i\,\sigma_{i,t} \tag{6.2}$$

Note that τ can assume positive and negative values, $\tau = 0, \pm 1, \pm 2, \ldots, \pm u$. By construction, the system is time-invariant under time translation and the coupling matrix is symmetric under time reversibility $J_{ij,\tau} = J_{ji,-\tau}$.

The constraints are the mean firing rates

$$\langle \sigma_{i,t} \rangle_r = \langle \sigma_{i,t} \rangle_m \tag{6.3}$$

and the correlation functions

$$\langle \sigma_{i,t}\,\sigma_{j,t+\tau} \rangle_r = \langle \sigma_{i,t}\,\sigma_{j,t+\tau} \rangle_m. \tag{6.4}$$

Using the technique of the Lagrange multipliers as shown in Chap. 6, we have that the model distribution is given by

$$P_m(\sigma) = \frac{1}{Z}\exp\left[-E(\sigma)\right] \tag{6.5}$$

where $E(\sigma)$ is given by Eq. (6.2).

6.4 Direct Problem

The direct problem consists in estimating some functions given a set of parameters. In this case, the set of parameters is $g = \{h_i, J_{ij,\tau}\}$ and the set of observables being estimated is $f = \{\sigma_{i,t}, \sigma_{i,t}\sigma_{j,t+\tau}\}$. In order to take the averages we must have samples of the theoretical model, which is done using a Metropolis Monte-Carlo method, which can be seen as samples drawn from the distribution $P_m(\sigma)$. In this way, the algorithm generates n samples. If the experiment has a total time T, then the time dimension is set to L to have the identity $nL = T$, so that an average take over nL summation is equivalent to the average taken from experimental dataset. In this way, the model mean firing rate is given by

$$m_i^{model} = \langle \sigma_{i,t} \rangle_m = \sum_\sigma P_m(\sigma) \left(\frac{1}{L} \sum_{t=1}^{L} \sigma_{i,t} \right) = \frac{1}{n} \sum_{\mu=1}^{n} \left(\frac{1}{L} \sum_{t=1}^{L} \sigma_{i,t}^{\mu} \right). \tag{6.6}$$

The real firing rate is given by,

$$m_i^{real} = \langle \sigma_{i,t} \rangle_r = \frac{1}{T} \sum_{t=1}^{T} \sigma_{i,t} \tag{6.7}$$

The model correlation is given by,

$$c_{ij,\tau}^{model} = \langle \sigma_{i,t} \sigma_{j,t+\tau} \rangle_m = \frac{1}{n} \sum_{\mu=1}^{n} \left(\frac{1}{L} \sum_{t=1}^{L} \sigma_{i,t}^{\mu} \sigma_{j,t+\tau}^{\mu} \right) \tag{6.8}$$

and the real correlation,

$$c_{ij,\tau}^{real} = \langle \sigma_{i,t} \sigma_{j,t+\tau} \rangle_r = \frac{1}{T} \sum_{t=1}^{T} \sigma_{i,t}^{\mu} \sigma_{j,t+\tau}^{\mu} \tag{6.9}$$

The last function that is interesting in capturing the collective behavior of neural network is the connected correlation

$$cc_{ij,\tau} = \langle \sigma_{i,t} \sigma_{j,t+\tau} \rangle - \langle \sigma_{i,t} \rangle \langle \sigma_{j,t} \rangle. \tag{6.10}$$

6.5 Inverse Problem

The inverse problem is a learning process where, as opposed to last section, the parameters $\{h_i, J_{ij,\tau}\}$ must be estimated from the observed states of the system $\{\sigma_{i,t}\}$ to give the correct empirical values of the mean firing rates $\langle \sigma_{i,t} \rangle$ and the correlations $\langle \sigma_{i,t} \sigma_{j,t+\tau} \rangle$. As discussed in Chap. 5, the learning process is implemented by the gradient descent algorithm. The update learning rule (5.18) for the couplings is given by

$$J_{ij,\tau} \quad \longleftarrow \quad J_{ij,\tau} + \alpha [\, c_{ij,\tau}^{real} - c_{ij,\tau}^{model} \,] \tag{6.11}$$

and for the fields

$$h_i \quad \longleftarrow \quad h_i + \gamma [\, m_i^{real} - m_i^{model} \,]. \tag{6.12}$$

Chapter 7

MaxEnt Algorithms

The method of making models with maximum entropy approach, in terms of computer algorithm, has two main branches, the direct and the inverse problem, as exposed in Chaps. 5 and 6. The direct problem is implemented with the Metropolis Monte Carlo (MC) algorithm to produce samples, that we can think of as samples drawn from the model distribution $P_m(\sigma)$, to estimate the functions $\{m_i,\, c_{ij,\tau},\, cc_{ij,\tau}\}$ given the parameters $\{h_i, J_{ij,\tau}\}$. The inverse problem uses with the gradient descent algorithm to implement a machine learning process with update learning rules.

7.1 Metropolis Monte-Carlo

The network has NL nodes, or spins. After performing NL Monte Carlo steps, we say we have completed one sweep of the network. It means that each spin $\sigma_{i,t}$ is given a chance to flip ($\sigma_{i,t} \to 1 - \sigma_{i,t}$) on average. Periodic boundary conditions were used in the time dimension to avoid boundary effects.

Sweep:

```
for $i = 1, \dots, NL$
    Choose a site (k,t) at random;
    Calculate the change in energy $\Delta E$ due to the flip of spin $\sigma_{k,t}$, using Eq. (6.2)
```

$$\Delta E = (2\sigma_{k,t} - 1)\Bigg(\sum_{j=1}^{N} \sum_{\tau=-u}^{u} J_{kj,\tau}\sigma_{j,t+\tau} + h_k \Bigg); \qquad (7.1)$$

If $\Delta E \leq 0$, accept the new configuration and flip the spin: $\sigma_{k,t} \to 1 - \sigma_{k,t}$;

Else, if $\Delta E > 0$, flip the spin with probability $\mathcal{P} = e^{-\Delta E}$, otherwise keep the old configuration;

The whole Metropolis algorithm consists in many sweeps. First, there is a thermalization period and afterwards n sweeps are performed to computed the expected values of the functions.

Inputs:

number of spins: N;
time dimension: L;
thermalization sweeps: T;
system sweeps: n;

Algorithm:

```
initialize system with random spins
for $i = 1, \ldots, T$
    system sweep
for $i = 1, \ldots, n$
    system sweep
    calculate correlations and firing rates
```

7.2 Machine learning

The machine learning process uses both the gradient descent algorithm, to update the parameters, and the Metropolis Monte Carlo algorithms, to run the sweeps of the system.

Input

empirical observations: m_i^{real}, $c_{ij,\tau}^{real}$;

parameters:

number of spins: N;

time dimension: L;

```
thermalization sweeps: $T$;
system sweeps: $n$;
maximum correlation distance on time: $u$;
coupling learning rate: $\alpha$;
field learning rate: $\gamma$;
precision: $\bar{p}$;
check point: $checkpt$;
```

Algorithm:

```
initialize parameters $\{h_i,\ J_{ij,\tau}\}$
initialize model expected values $\{m_i^{model},\ c_{ij,\tau}^{model}\}$
initialize system with random spins
for $i = 1, \ldots, T$
    system sweep
do
    update parameters $\{h_i, J_{ij,\tau}\} \leftarrow \{h_i + \gamma\Delta h_i,\ J_{ij,\tau} + \alpha\Delta J_{ij,\tau}\}$
    reset expected values $\{m_i^{model} = 0,\ c_{ij,\tau}^{model} = 0\}$
    for $j = 1, \ldots, n$
        system sweep
        compute functions $\{\sigma_{i,t},\ \sigma_{i,t}\sigma_{j,t+\tau}\}$
    calculate expected values $\{m_i^{model},\ c_{ij,\tau}^{model}\}$
while(check if precision $< \bar{p}$)
```

After every Monte Carlo run, the program checks if the precision of the program p is within the precision $\bar{p}$ wanted. The precision of the program is calculated as the relationship between two errors. One is the error between the ‘real’ and ‘model’ expected values (E_{real}). The other is the error associated with the inherent stochasticity of the Monte Carlo method (E_{MC}), which is the error of two consecutive Monte Carlo runs with the same parameters. The precision is given by $p = E_{real}/E_{MC}$. The algorithm is made in order that the program stops when $p = E_{real}/E_{MC} \leq \bar{p}$. The two errors E_{real} and E_{MC} are the mean absolute error, which is the mean absolute value of the difference between the two values being compared. As an example, the error between the 'real' and 'model' mean firing rate is given by

$$E_{real}(\{m_i\}) = \frac{1}{N}\sum_{i=1}^{N} |\langle m_i \rangle_r - \langle m_i \rangle_m| \tag{7.2}$$

The errors are calculated for the mean firing rates $\{m_i\}$ and the correlation functions $\{c_{ij,\tau}\}$.

Checking process algorithm:

```
if (checkpt)
    calculate the learning errors ($E_{real}$)
    run a MC algorithm with the same parameters for test
    calculate MC test errors ($E_{MC}$)
    compare errors with precision: $p = E/E_{MC} \leq \bar{p}$ ?
```

Chapter 8

Results

Before running the program with the real data, some tests were made with synthetic data and analytic approximations. First, the Metropolis Monte Carlo algorithm was loaded with artificial values of couplings and fields, and the estimated connected correlation was compared with the high temperature expansion analytical approximation [14, 52]. Then, the machine learning algorithm was tested against synthetic data produced from the Metropolis algorithm to see if it was giving the correct parameters to reproduce the measured values from the synthetic data. Once the program was tested and it has given reliable parameters, the real data was analyzed.

8.1 Synthetic data

The synthetic data is produced by the Metropolis Monte Carlo algorithm with artificial values of couplings and fields. We are in a regime of weak interactions. The couplings were, therefore, drawn from a Gaussian distribution with zero mean and a small standard deviation *sd*. A separate process is done for the fields. The fields were made constant and their values were calculated with the approximation of weak interactions, i.e., considering $J_{ij,\tau} = 0$ in Eq. (6.2). They are calculated as follows.

The probability of finding the system in a specific state $\{\sigma_{i,t}\}$ factorizes over neurons and over time, and reads

$$P(\{\sigma_{i,t}\}) = \prod_{t=1}^{L} \prod_{i=1}^{N} p_i^{\sigma_{i,t}} q_i^{1-\sigma_{i,t}} \tag{8.1}$$

where p_i is the probability of neuron i with a firing rate r_i spiking in a time bin δt,

$p_i = r_i\,\delta t$, and $q_i = 1 - p_i$. But, from Eq. (5.10) and using the approximation $J_{ij,\tau} \sim 0$, the distribution is also equal to

$$P(\{\sigma_{i,t}\}) = \frac{1}{Z} \exp\left[- \sum_{t=1}^{L} \sum_{i=1}^{N} h_i \sigma_{i,t} \right]. \tag{8.2}$$

Taking the log and equating the probabilities,

$$\sum_{i,t} [\sigma_{i,t} \log p_i/q_i + \log q_i] = \sum_{i,t} h_i \sigma_{i,t} - \log Z. \tag{8.3}$$

Comparing both sides and considering constant firing rates $r_i = r$

$$h_i = h = \log \frac{r\,\delta t}{1 - r\,\delta t}. \tag{8.4}$$

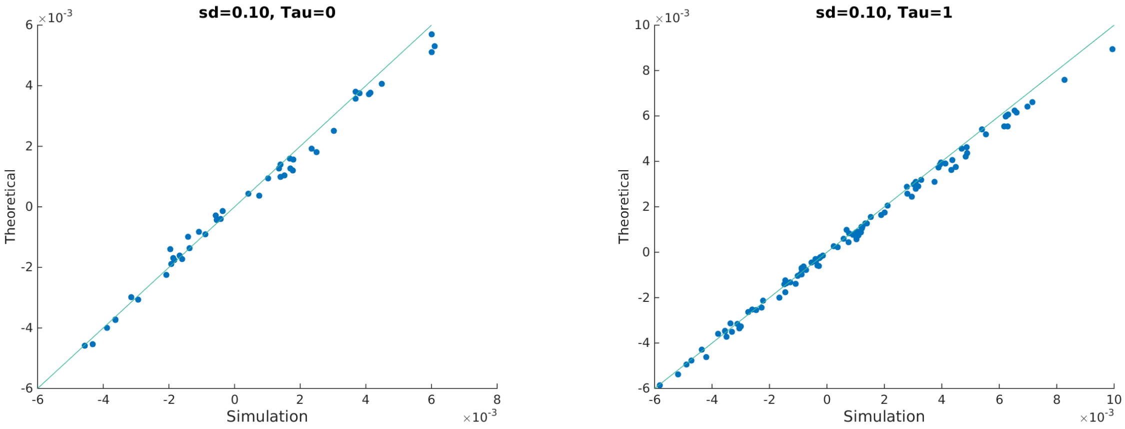
To perform the simulations it was considered a firing rate of $r = 10$ Hz and time bins of t =10 ms, which gives $h \sim -2.2$.

The high temperature expansion approximation was used to compare the connected correlation estimated from the Monte Carlo algorithm. It is given by

$$cc_{ij,\tau} = \tanh\left(\frac{J_{ij,\tau}}{4} \right) \left[1 - (2\langle \sigma_i \rangle - 1)^2\right] \left[1 - (2\langle \sigma_j \rangle - 1)^2\right]. \tag{8.5}$$

8.1.1 Metropolis Monte-Carlo: direct problem

The Metropolis Monte Carlo algorithm for the given theoretical framework was tested with many configurations of synthetic data. Scatter plots compare the connected correlation estimated from the Monte Carlo algorithm (Simulation) with the high temperature expansion analytical approximation (8.5) (Theoretical).



(a) $\tau = 0$

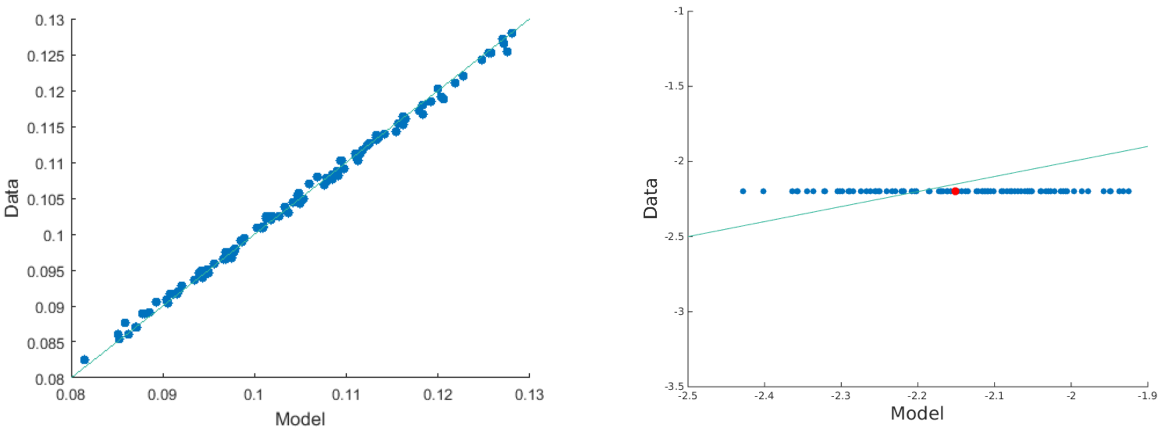
(b) $\tau = 1$

Figure 8.1: Comparison between the connected correlation estimated from the Monte Carlo algorithm (Simulation) and the high temperature expansion (Theoretical) with time correlation distances of (a) $\tau = 0$ and (b) $\tau = 1$.

As an example, it is presented in Fig. 8.1 a configuration with N=10 neurons, a time dimension of L=1000 time bins, a maximum distance of interaction in the time dimension of u=1, n=20000 sweeps, sd=0.1 and time correlation distance of $\tau = 0$ in Fig. 8.1a and $\tau = 1$ in Fig. 8.1b. The scatter plot shows that the algorithm is estimating correct expected values withing the theoretical approximation considered.

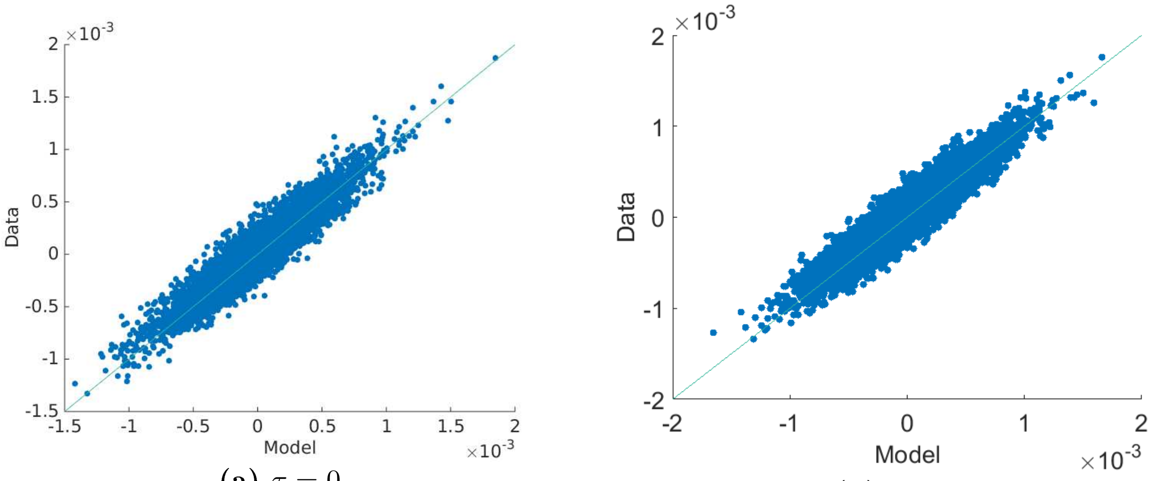
8.1.2 Machine learning: inverse problem

The machine learning algorithm was tested with many configurations as well. Scatter plots show a comparison between the synthetic data (Data) and the data inferred from the model (Model). As an example, it is shown a configuration of N=100, L=1000, u=1, n=800 every MC run, sd=0.04, and learning rates α=0.85 for the interactions and γ=0.025 for the fields. Finding learning rate values that give appropriate steps towards the minimum of the Kullback-Leibler divergence is a difficult task, it has to be done in an empirical way.



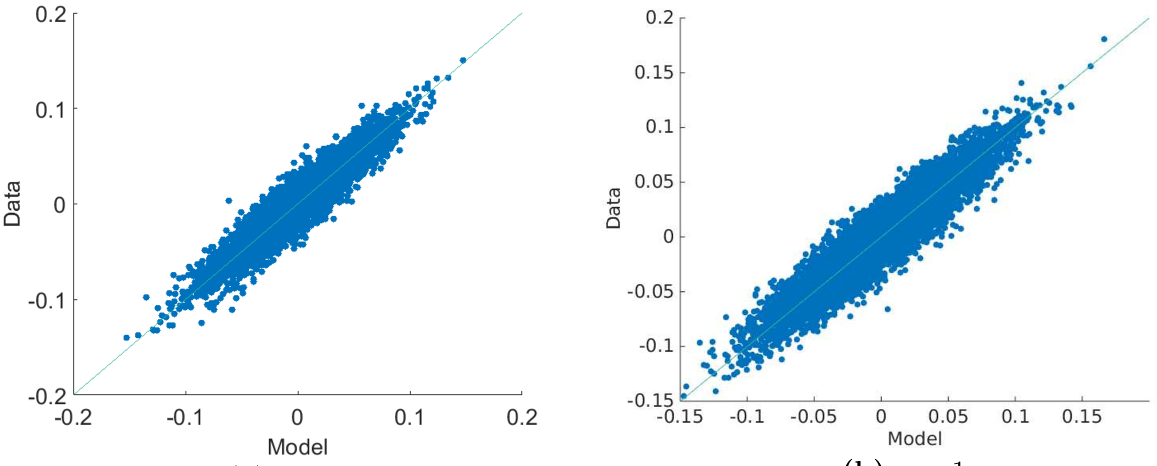
(a) **(b)**

Figure 8.2: Scatter plot of (a) mean firing rates $\{m_i\}$ and (b) fields $\{h_i\}$.



(a) $\tau = 0$ **(b)** $\tau = 1$

Figure 8.3: Scatter plot of connected correlations $\{cc_{ij,\tau}\}$ for (a) $\tau = 0$ and (b) $\tau = 1$.



(a) $\tau = 0$ **(b)** $\tau = 1$

Figure 8.4: Scatter plot of interactions $\{J_{ij,\tau}\}$ for (a) $\tau = 0$ and (b) $\tau = 1$.

The scatter plots of Figs. 8.2–8.4 show that the machine learning algorithm is giving reliable values of couplings and fields since it is reproducing the synthetic mean firing rates and connected correlations. The fields were made constant and it is not expected that the algorithm set them all to a constant value. The scatter plot of Fig. 8.2a shows that field values are very close to the expected value ($h \sim -2.2$) within an error.

8.2 Real data

The real data spike train $\{\sigma_{i,t}\}$ is obtained from the flickering checkerboard stimuli projected on the retina, with time bins $\delta t = 10$ ms. Three configurations were analyzed

$$N = 185, \quad L = 185000;$$

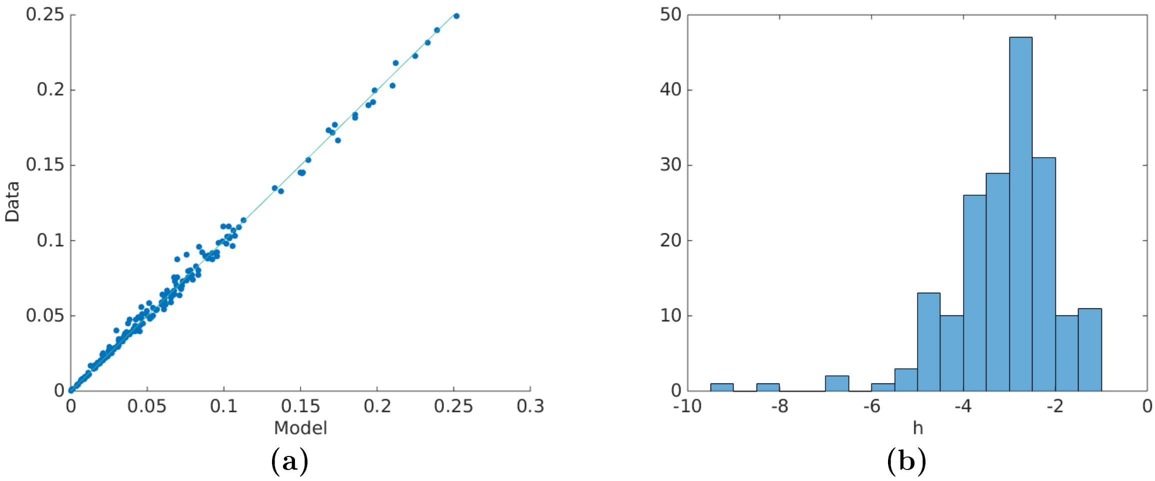
$$N = 97, \quad L = 185000;$$

$$N = 61, \quad L = 185000;$$

where N is number of neurons and L is the time dimension, i.e., the total time of the experiment is $T = L\delta t$.

8.2.1 Time range u=5

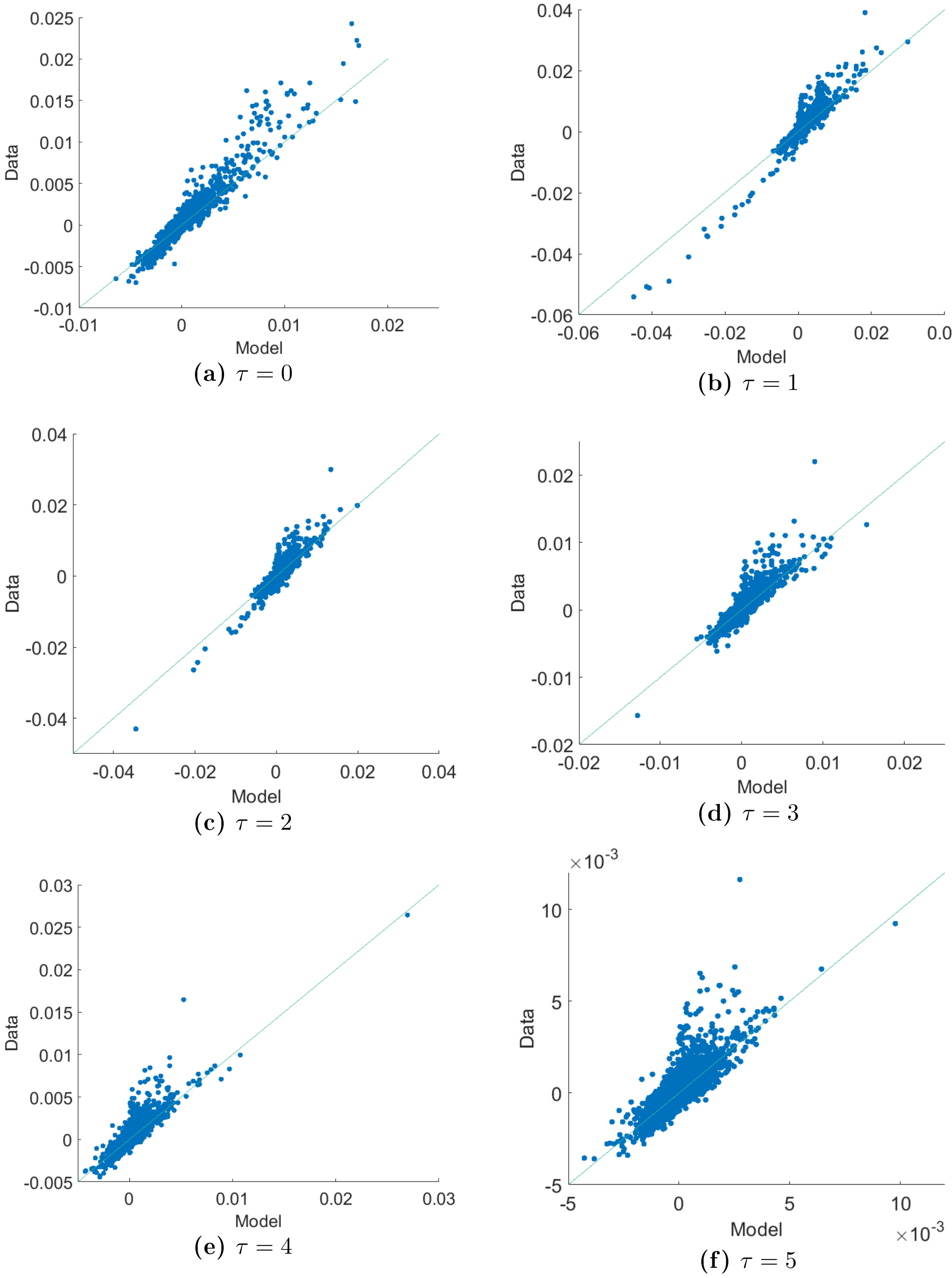
Time interaction distances from $\tau = 0$ to $\tau = 5$ were studied. Here is presented the largest range studied with $u = 5$ and with the configuration N=185 neurons, L=500 time dimension, n=360 sweeps every MC run, α=0.25 couplings learning rate, γ=0.025 field learning rate and precision $\bar{p}$=3.



(a) (b)

Figure 8.5: Scatter plot of (a) mean firing rates and (b) fields distribution.

It can be seen from Fig. 8.5 that the algorithm is estimating the firing rates with good accuracy, making the fields values to be reliable.



(a) $\tau = 0$

(b) $\tau = 1$

(c) $\tau = 2$

(d) $\tau = 3$

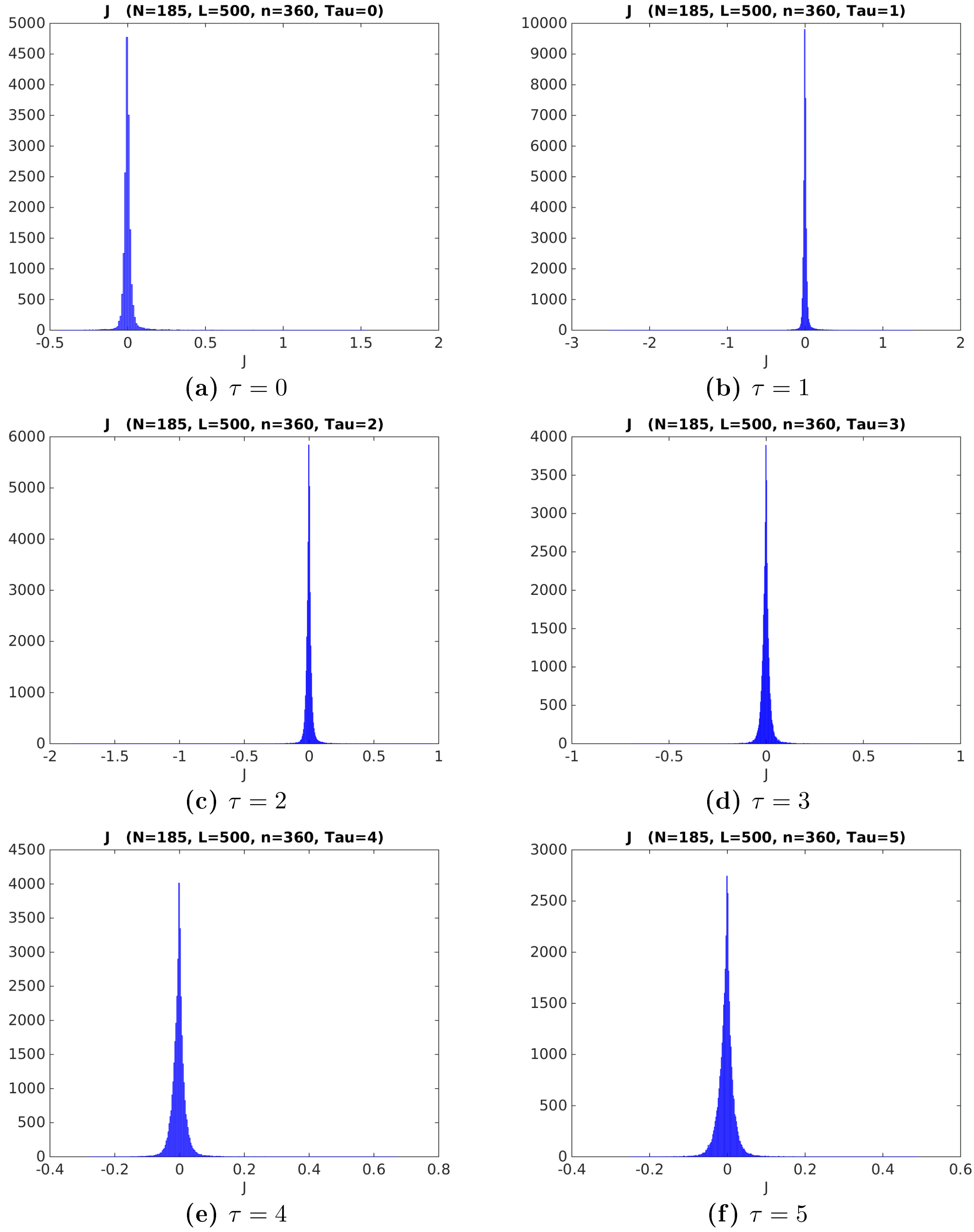
(e) $\tau = 4$

(f) $\tau = 5$

Figure 8.6: Scatter plots of connected correlation with τ ranging from 0 to 5.

Also the connected correlation are being estimated with good accuracy (see Figs. 8.6a–

8.6f). Nevertheless, it can be observed that there are many outliers for the larger absolute values. Those outliers are a result of the high self correlation, although it is not clear why are self correlations so high.



(a) $\tau = 0$

(b) $\tau = 1$

(c) $\tau = 2$

(d) $\tau = 3$

(e) $\tau = 4$

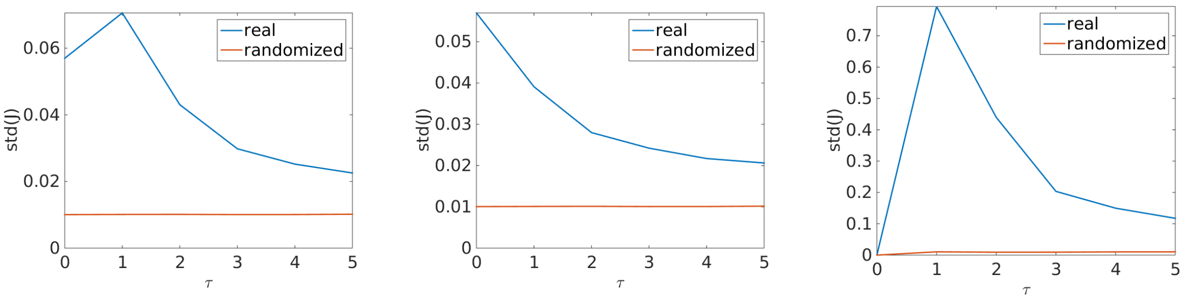
(f) $\tau = 5$

Figure 8.7: Interactions distributions with τ ranging from 0 to 5.

The interactions distributions, Figs. 8.7a–8.7f, show that they are picked around zero, in other words, the interactions are weak. In spite of that, there are many outliers for larger absolute values, what is consistent with the connected correlations in Figs. 8.6a–8.6f.

8.2.2 Couplings range

It is expected that there is a time-scale on which the time auto-correlation falls off. For time differences τ larger than the time auto-correlation, the activities of neurons are not correlated anymore. This is observed in real data in terms of the decay of the coupling range.



(a)

(b)

(c)

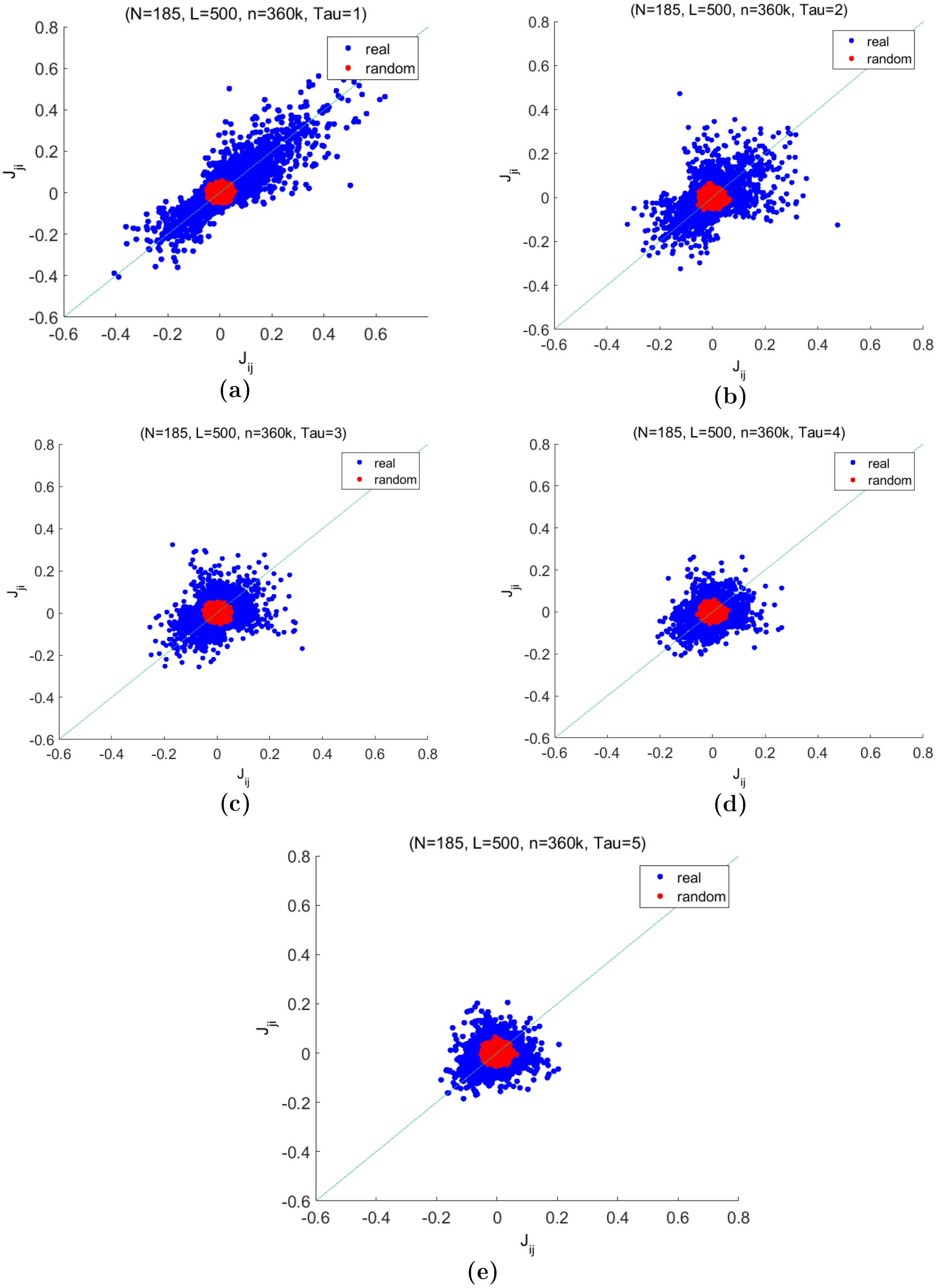
Figure 8.8: Standard deviation of the couplings distribution in function of τ of the real data (blue) and the randomized data (red). (N=185, L=500, u=5, α=0.25, γ=0.025): (a) complete, (b) without main diagonal, (c) only main diagonal.

As the time distance τ of interaction increases, the couplings range decreases. This can be observed in Fig. 8.8, where the standard deviation of the couplings of real data (blue line) decreases. In Fig. 8.8a all interactions are present, and there is an increase in the coupling range from $\tau = 0$ to $\tau = 1$, because in this case there is no self correlation at $\tau = 0$ ($J_{ii,0} = 0$), in terms of neuron activity, which are responsible for most of the outliers. In Fig. 8.8b the main diagonal was removed and the curve is monotonically decreasing. In Fig. 8.8c there is only the main diagonal and the values are much higher than the other two graphs. Moreover, in Fig. 8.8, the red line stands for the standard deviation of the couplings of the randomized data, that is, every neurons has its time series randomized in the original experimental dataset. It is observed that the blue line approaches the red line as τ increases. Therefore, it is expected that, for large τ, the correlations goes to zero.

8.2.3 Detailed balance

The system studied is out of equilibrium. In general, out of equilibrium systems do not obey a detailed balance equation. It can also be regarded as an irreversibility on time.

Thus, the detailed balance can be studied be means of the interactions $J_{ij,\tau}$, because of the time symmetry property $J_{ij,\tau} = J_{ji,-\tau}$.



(a)

(b)

(c)

(d)

(e)

Figure 8.9: Detailed balance: scatter plots of $J_{ji,\tau}$ versus $J_{ij,\tau}$.

The graphs in Fig. 8.9 are scatter plots of the interactions matrix $J_{ij,\tau}$ and the transposed matrix $J_{ji,\tau}$ for $\tau = 1, 2, 3, 4, 5$. Although we cannot make conclusion yet, the system does not seem to obey the detailed balance. This must be investigated with more careful and quantified methods.

Another remarkable point is that there is always the question if the measured data is not completely random, i.e., that neurons are spiking with no correlation. To investigate this, the real data was randomized in the time dimension. The code word of each neuron is maintained with the same number of zeros and ones, but the order is randomized, and the machine learning process performed again. If the real data is random, it should be indistinguishable from the randomized data. In all graphs of Fig. 8.9 the red dots stands for the randomized data, and they are all concentrated at the center of the scatter plots. This shows that the real data is not a product of a dynamics of random spiking neurons. It is also possible to observe that the real data gets closer to the randomized data as τ grows.

8.2.4 Thermodynamic Variables

Once the system is characterized by the learned parameters $\{h_i; J_{ij,\tau}\}$, it is possible to calculate thermodynamic variables by considering an artificial temperature. The model distribution is than given by,

$$P(\sigma) = \frac{1}{Z} \exp -\beta E(\sigma), \tag{8.6}$$

where $\beta = 1/T$ is the inverse temperature and it has no direct physical significance, it is just an external parameter.

In the way that the model was built, the temperature of the system is $T = 1$. So that, thermodynamic variables can be compared to different temperatures and $T = 1$ is the system temperature, or the temperature of operation of the system. For each temperature variables can be calculated using a Monte-Carlo method to give n samples. The energy is given by,

$$E(T) = \frac{1}{n} \sum_{\nu=1}^{n} E_\nu(\sigma, T). \tag{8.7}$$

The magnetization can be calculated,

$$\langle m(T)\rangle = \frac{1}{n}\sum_{\nu=1}^{n} m_\nu(T) \tag{8.8}$$

where, for $s_{i,t}^{\nu} = \pm 1$,

$$m_\nu(T) = \frac{1}{NL}\sum_{i,t} s_{i,t}^{\nu} = \frac{1}{NL}\sum_{i,t}(1 - 2\sigma_{i,t}^{\nu}) = 1 - 2\langle .\sigma_{i,t}^{\nu}\rangle . \tag{8.9}$$

We can also define a specific heat,

$$c(T) = \frac{1}{NL}\frac{\langle \delta E^2(T)\rangle}{T^2} \tag{8.10}$$

where, $\delta E(T) = E(T) - \langle E(T)\rangle$.

These are thermodynamic variables that have a specific behavior at the critical point. The magnetization is the system order parameter and the specific heat diverges at the critical point. The analysis of these variables will allow us to characterize the criticality of the system. This work is the object of a future project.

Chapter 9

Conclusion

This work was divided in two parts. In this first part, a theoretical model was developed to simulate the brain network built from measured BOLD temporal series. A computer simulation of a spin lattice was performed and the system evolved according to a Metropolis algorithm with single-spin-flip dynamics. Using the Monte-Carlo sampling method, each spin had a time series in the end. Correlations between each pair of spin time series were obtained and a correlation network was build for the Ising model. The simulations were performed in three temperatures, critical, subcritial, and supercritical. Network measures were also calculated and analysed. A comparison between the brain network of the experimental results from the Neurophysics group at University of Campinas and the Ising correlation network showed that what was observed in brain data was better explained by the theoretical model when it was at the critical temperature. This shows evidences that the brain operates in a critical regime, or that critical aspects are present in the brain dynamics.

The next steps of this work is to enhance the model. A first approach would be to include a field term in the Hamiltonian. Another approach is to work with a variable with more than two dynamical states as the spin, given that the experimental data from brain has temporal series with a continuous signal which can assume arbitrary values in principle. Something that is also possible and somehow equivalent is to study different topologies in the original lattice, other than a square lattice. Finally, we intend to investigate theoretical models to simulate brain data from subjects with pathologies, like epilepsy.

In the second part, maximum entropy models approach was used to study the collective dynamics of a large neuron population. This work was performed during a six months internship at the École Normale Supérieure of Paris under the supervision of Dr. Thierry Mora, and the data was obtained from his collaborator Dr. Olivier Marre from the Vision Institute of Paris. Data was recorded directly from the activity of the retina of mice. A movie was projected on the retina and the activity of ganglion cells was recorded by a multi-electrode array. In the end, the dataset is a raster of time bins of the whole system where each neuron can be spiking or in silence in that time bin, which is a code-word. Mapping each neuron into a classic spin with two states, "up" (spike) or "down" (silence), it was possible to construct an Ising model with pairwise interactions.

Many works have studied this kind of dynamics, but no attention was given to the temporal dynamics, considering each code-word as time independent. In this work, to study the temporal dependence, an Ising model was built treating time as an extra dimension of the system, putting interactions between different spins at different moments of time. Using a Monte-Carlo sample method together with a gradient descent algorithm a machine learning process was used in model fitting, and the system was characterized by the parameters learned from the computer program. Aspects of the dynamics, such as the interactions range on time and detail balance are being studied.

The work is in progress and now we are looking to improve the learning process in order to have more reliable parameters. Thermodynamics variables will be calculated, such as the specific heat and magnetization, which show a specific diverging behavior at criticality, so that the critical hypothesis can be studied.

In a collaboration with the Neurophysics group at University of Campinas, we consider to apply this framework to brain data. Near-infrared spectroscopy (NIRS) together with fMRI are being used to capture brain activity, which can be studied with the maximum entropy inference approach.

Bibliography

[1] Albert, R., and Barabasi, A. L. Statistical mechanics of complex networks. *Rev Mod Phys 74*, 1 (2002), 47–97.

[2] Barabási, A. The network takeover. *Nature 8*, 1 (2012), 14–16.

[3] Barabási, A. L., and Albert, R. Emergence of Scaling in Random Networks. *Science 286*, October (1999), 509–512.

[4] Barrat, A., Barthelemy, M., and Vespignani, A. *Dynamical Processes on Complex Networks*. Cambridge University Press, Cambridge, UK, 2010.

[5] Beggs, J. M. Neuronal Avalanches Are Diverse and Precise Activity Patterns That Are Stable for Many Hours in Cortical Slice Cultures. *J Neurosci 24*, 22 (2004), 5216–5229.

[6] Beggs, J. M., and Plenz, D. Neuronal Avalanches in Neocortical Circuits. *J Neurosci 23*, 35 (2003), 11167–11177.

[7] Bertschinger, N., and Natschläger, T. Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comp 16*, 7 (jul 2004), 1413–36.

[8] Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Mag Res Med 34*, 5 (1995), 537–41.

[9] Broderick, T., Dudík, M., Tkacik, G., Schapire, R. E., , and Bialek, W. Faster solutions of the inverse pairwise ising problem, 2007. arXiv:0712.2437v2 [q-bio.QM].

[10] Cavagna, A., Giardina, I., Ginelli, F., Mora, T., Piovani, D., Tavarone, R., and Walczak, A. M. Dynamical maximum entropy approach to flocking. *Phys Rev E 89* (2014), 042707.

[11] Chaikin, P. M., and Lubensky, T. C. *Principles of Condensed Matter Physics.* Cambridge University Press, Cambridge, UK, 1995.

[12] Chialvo, D. R. Emergent complex neural dynamics. *Nat Phys 6*, 10 (2010), 744–750.

[13] Chialvo, D. R. Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis. *Front in Physiol 3*, February (2012), 1–12.

[14] Cocco, S., Leibler, S., and Monasson, R. Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *PNAS 106*, 33 (2009), 14058–14062.

[15] Dorogovtsev, S. N., Goltsev, A. V., and Mendes, J. F. F. Critical phenomena in complex networks. *Rev Mod Phys 80* (2008).

[16] Easley, D., and Kleinberg, J. *Networks, Crowds, and Markets – Reasoning about a Highly Connected World.* Cambridge University Press, Cambridge, UK, 2010.

[17] Eguíluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., and Apkarian, A. V. Scale-Free Brain Functional Networks. *Phys Rev Lett 018102*, January (2005), 1–4.

[18] Fox, M. D., and Raichle, M. E. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci 8* (2007), 700–711.

[19] Fraiman, D., Balenzuela, P., Foss, J., and Chialvo, D. R. Ising-like dynamics in large-scale functional brain networks. *Phys Rev E 79* (2009), 61922.

[20] Gašper, Marre, O., Mora, T., Amodei, D., Berry, M. J., and Bialek, W. The simplest maximum entropy model for collective behavior in a neural network. *J Stat Mech 2013*, 03 (2013), P03011.

[21] Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. Mapping the Structural Core of Human Cerebral Cortex. *PLoS Biol 6*, 7 (2008), e159.

[22] Haimovici, A., Tagliazucchi, E., Balenzuela, P., and Chialvo, D. R. Brain organization into Resting State Networks Emerges at Criticality on a Model of the Human Connectome. *Phys Rev Lett 110* (2013), 178101.

[23] Huang, K. *Statistical Physics.* Jon Wiley & Sons, New York, USA, 1987.

[24] Hyvärinen, A., Karhunen, J., and Oja, E. *Independent Component Analysis.* John Wiley & Sons, 2001.

[25] Ising, E. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik 31*, 1 (1925), 253–258.

[26] Jaynes, E. T. Information theory and statistical mechanics. *Phys Rev 106* (1957), 620–630.

[27] Jaynes, E. T. Information Theory and Statistical Mechanics II. *Phys Rev 108* (1957), 171–190.

[28] Kapur, J. N., and Kesavan, H. K. *Entropy Optimization Principles with Applications.* Academic Press, 1992.

[29] Kinouchi, O., and Copelli, M. Optimal dynamical range of excitable networks at criticality. *Nat Phys 2*, 5 (2006), 348–351.

[30] Kullback, S., and Leibler, R. A. On information and sufficiency. *Ann Math Statist 22*, 1 (1951), 79–86.

[31] Legenstein, R., and Maass, W. Edge of chaos and prediction of computational performance for neural circuit models. *Neural Net 20*, 3 (apr 2007), 323–34.

[32] Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature 412*, 6843 (2001), 150–7.

[33] Marre, O., Amodei, D., Deshmukh, N., Sadeghi, K., Soo, F., Holy, T. E., and Berry, M. J. Mapping a complete neural population in the retina. *J Neuroscience 32*, 43 (2012), 14859–14873.

[34] Mesquita, R. C., Franceschini, M. A., and Boas, D. A. Resting state functional connectivity of the whole head with near-infrared spectroscopy. *Biomed Opt Express 1*, 1 (2010), 676–682.

[35] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equation of state calculations by fast computing machines. *J Chem Phys 21*, 6 (1953), 1087–1092.

[36] Mora, T., and Bialek, W. Are Biological Systems Poised at Criticality? *J Stat Phys 144*, 2 (2011), 268–302.

[37] Mora, T., Deny, S., and Marre, O. Dynamical Criticality in the Collective Activity of a Population of Retinal Neurons. *Phys Rev Lett 114* (2015), 78105.

[38] Moretti, P., and Munoz, M. A. Griffiths phases and the stretching of criticality in brain networks. *Nat Commun 4* (jan 2013), 2521.

[39] Nasser, H., and Cessac, B. Parameter Estimation for Spatio-Temporal Maximum Entropy Distributions: Application to Neural Spike Trains. *Entropy 16*, 4 (2014), 2244–2277.

[40] Nasser, H., Marre, O., and Cessac, B. Spatio-temporal spike train analysis for large scale networks using the maximum entropy principle and Monte Carlo method. *J Stat Mech 2013*, 03 (2013), P03006.

[41] Newman, M. E. J. The structure and function of complex networks. *SIAM Rev 45*, 2 (2003), 167–256.

[42] Newman, M. E. J. *Networks: An Introduction.* Oxford University Press, Oxford, UK, 2010.

[43] Newman, M. E. J., and Barkema, G. T. *Monte Carlo Methods in Statistical Physics.* Oxford University Press, Oxford, UK, 2001.

[44] Nicosia, V., Valencia, M., Chavez, M., Díaz-Guilera, A., and Latora, V. Remote synchronization reveals network symmetries and functional modules. *Phys Rev Lett 110* (2013), 174102.

[45] Onsager, L. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Phys. Rev. 65* (1944), 117–149.

[46] Plenz, D. Neuronal avalanches and coherence potentials. *Eur Phys J Special Topics 301* (2012), 259–301.

[47] Pressé, S., Ghosh, K., Lee, J., and Dill, K. A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev Mod Phys 85*, 3 (2013), 1115–1141.

[48] Roudi, Y., and Hertz, J. Mean field theory for nonequilibrium network reconstruction. *Phys Rev Lett 106*, 4 (2011), 1–4.

[49] Roudi, Y., and Hertz, J. Mean field theory for nonequilibrium network reconstruction - Supplements. *Phys Rev Lett 106* (2011), 1–3.

[50] Rubinov, M., and Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage 52*, 3 (sep 2010), 1059–69.

[51] Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature 440* (2006), 1007–1012.

[52] Sessak, V., and Monasson, R. Small-correlation expansions for the inverse ising problem. *J Phys A 42*, 5 (2009), 055001.

[53] Shannon, C. E. A mathematical theory of communication. *Bell Syst Tech J 27*, July 1928 (1948), 379–423.

[54] Sporns, O. *Networks of the Brain.* MIT Press, Cambridge, UK, 2011.

[55] Sporns, O., Tononi, G., and Kötter, R. The human connectome: A structural description of the human brain. *PLoS Comput Biol 1*, 4 (2005), e42.

[56] Tkačik, G., Marre, O., Amodei, D., Schneidman, E., Bialek, W., and Berry, M. J. Searching for Collective Behavior in a Large Network of Sensory Neurons. *PLoS Comp Bio 10*, 1 (2014).

[57] Tkačik, G., Mora, T., Marre, O., Amodei, D., Palmer, S. E., Berry, M. J., and Bialek, W. Thermodynamics and signatures of criticality in a network of neurons. *PNAS 112*, 37 (2015), 11508–11513.

[58] Vasquez, J. C., Marre, O., Palacios, a. G., Berry, M. J., and Cessac, B. Gibbs distribution analysis of temporal correlations structure in retina ganglion cells. *J Physiol - Paris 106*, 3-4 (2012), 120–127.

[59] Watts, D. J., and Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature 393*, 6684 (1998), 440–2.

[60] Yacoub, E., and Harel, N. High-field fMRI unveils orientation columns in humans. *PNAS 105*, 30 (2008).

Appendix A

Kullback-Leibler divergence

Given two probability distributions, P_r and P_m, the Kullback-Leibler (KL) divergence is defined as

$$D_{KL}(P_r|P_m) = \sum_\sigma P_r(\sigma) \log \frac{P_r(\sigma)}{P_m(\sigma)}. \tag{A.1}$$

It is a non-symmetric measure of discrepancy or divergence between two distributions. In terms of code length, if a system is built to operate optimally with a given distribution, say P_r, the KL divergence measures the extra length of bits that must be coded by the system to operate with the other distribution, P_m in Eq. (A.1), and produce the same message. Technically speaking, the KL divergence measures the difference between the entropy of the optimal distribution, $S[P_r]$, and the cross-entropy of P_r and P_m, $S[P_r, P_m] = -\sum_\sigma P_r(\sigma) \log P_m(\sigma)$.

$$\begin{aligned} D_{KL}(P_r|P_m) &= -\sum_\sigma P_r(\sigma) \log P_m(\sigma) + \sum_\sigma P_r(\sigma) \log P_r(\sigma) && \text{(A.2)} \\ &= S[P_r, P_m] - S[P_r]. && \text{(A.3)} \end{aligned}$$

Between the properties of the KL divergence, some of them are worth mentioning here:

1. *non-negativity*: $D_{KL}(P_r|P_m) \geq 0$;

2. *identity*: $D_{KL}(P_r|P_m) = 0$ if, and only if, $P_r = P_m$;

3. *continuity*: $D_{KL}(P_r|P_m)$ is a continuous function of P_r and P_m;

4. *convexity*: $D_{KL}(P_r|P_m)$ is a convex function of both P_r and P_m.

Properties (1) and (2) are essential for any measure of discrepancy or divergence. Property (4) ensures that the local minimum is the global minimum, which is used as convergence property of the process of finding the appropriate Lagrange multiplier for the model distribution $P_m(\sigma)$.