UNIVERSITY OF PISA

Ph.D. Dissertation in Mathematics

Claudio Bonanno

Applications of information measures to chaotic dynamical systems

Advisors
Prof. Vieri Benci
Prof. Stefano Isola

Director
Prof. Fabrizio Broglia

July 2003

To the memory
of my mother

Acknowledgements

I want to express my gratitude to my two advisors, Vieri Benci and Stefano Isola, for their encouragement from the beginning to the end of this work. In particular I thank Vieri Benci for introducing me to the method of application of information measures to dynamical systems, and Stefano Isola for being a constant guide in the "world" of dynamical systems.

I also want to thank the people with whom in these years I have been spending hours talking about the problems I had to face. In particular I express my gratitude to Stefano Galatolo, for his tutorial job, and to Giulia Menconi, for her collaboration. Moreover some of my research, related to this work but not included, has been done in collaboration with Jacopo Bellazzini, Paolo Grigolini, Vincenzo Manca and Mirko S. Mega. Certainly the discussions I had with them have been very important for my understanding of complexity in dynamical systems.

Finally, I want to thank my friends for their support and for having tolerated my terrible mood. In particular I feel in debt to Irene, Giuditta, Pietro, Chiara, Rossella and Domenico, they had the bad luck to be the closest in the worst moments. A thanks also goes to the people that have shared with me the lunches, the dinners and the (many) breaks at the Department of Mathematics in Pisa, for having helped me in forgetting this work for a while. I don't want to forget anyone, so I prefer not to mention them explicitly, but I am sure that they know who has been the most important for me.

Last but not least, I cannot forget the support I have received from my family, even if they live far away. The words of my father and my sister, the happiness of Andrea and the joy of the young Gabriele, have been an important help in these years.

Contents

Introduction

For long time, in the theory of dynamical systems, the main problem was that of finding analytically the orbits that the bodies of the system follow. In particular it was the aim of celestial mechanics. But at the end of the nineteenth century, the work of Poincaré started to change this point of view. This change was due to the discovery, made by Poincaré, of a great degree of uncertainty in the orbits of the bodies of the three body problem, and in particular to the proof of the lackness of analytic first integrals for the system. Hence it was "impossible" to find the solutions analytically, and the only possibility was that of giving some properties of these solutions. However, even to draw approximately the solutions was too difficult, because (using the actual terminology) of the chaoticity of the solutions near an unstable fixed point with a homoclinic point. These problems moved the attention towards a *qualitative* approach to dynamical systems.

Another important discovery in the theory of dynamical systems was made by Lorenz in 1963. Studying numerically a simple model of the terrestrial atmosphere, he found that "small" differences in the initial conditions caused "great" differences in the solutions. This was the beginning of the theory of *chaos*. In the last forty years there has been a big development of the theory of chaos, and many have been the different approaches introduced to study and characterize chaotic dynamical systems.

The difficulties in the explicit determination of the orbits of a system led to the beginning of a *statistical* approach to dynamical systems. Hence, instead of looking at a single orbit, the system was studied as a whole ensemble of orbits, in the hypotheses that a single orbit covered "almost" all the phase space. This approach has began an independent theory for dynamical systems, the *ergodic theory*.

In Chapter 1 we give an overview of ergodic theory and of the theory of chaos, with a particular point of view. To formally study the statistical properties of a system, it is necessary to have a probability measure on the phase space that is *invariant* with respect to the action of a map (the dynamics). Given this measure, it is possible to study the properties of ergodicity of the system, and to try to characterize the chaoticity of the system according to the *Kolmogorov-Sinai entropy*.

This approach fails in two cases. In one case, a dynamical system can have null Kolmogorov-Sinai entropy with respect to all its invariant probability measures that are *physically relevant*. With physically relevant measure we mean a measure that can be "seen" in the numerical iterations of the system (for the formal definition see Section 1.3.2). However there are such systems with a high degree of unpredictability for their orbits. These systems are called in this work *weakly chaotic*. The other situation in which the classical theory fails is that of systems with invariant measures that are not finite. In this case the ergodic theory gives results that cannot characterize the systems, and the classical theory of chaos is not applicable.

As shown in Chapter 1, in recent years the interest towards these two cases is growing, and our attention in this work is dedicated to the study of dynamical systems in these categories.

To study chaotic dynamical systems we use a method from information theory. Indeed already from the definition of Kolmogorov-Sinai entropy, it is clear that the chaoticity of a system is related to the quantity of information that has to be given to an observer to describe the evolution of the orbits of the system. This quantity of information can be measured by *information functions*, that are functions defined on the set of finite words on a given alphabet. Hence the receipt of our method is the following: first of all, translate the orbits of a dynamical system into words on an alphabet (this is done by the method of *symbolic dynamics*); then "measure" the quantity of information that each word carries, and characterize the system according to the asymptotic behaviour of this quantity of information as the length of the word increases (the length of the word is related to the time for the evolution of the system).

In Chapter 2 we introduce the concept of information functions and describe three different examples, specifying the differences among them. In this work we use as information function the notion of *Algorithmic Information Content (AIC)*, because it does not depend on the choice of the invariant measure of the system, but still it is related to the classical notions of chaotic dynamical systems, hence give information to characterize the systems.

In Chapter 3 we start our method, showing how to generate an information source from a dynamical system, and applying the information functions to this source. In the case of a chaotic dynamical system with a probability invariant measure, if the Kolmogorov-Sinai entropy relative to the measure is positive, we show that the *complexity* of a symbolic orbit of the system (that is the rate of increase with length/time of the *AIC* of a word) is equal to the Kolmogorov-Sinai entropy for ergodic systems. This result is due to Brudno ([Br]), and can be easily generalized to non-ergodic systems.

For weakly chaotic systems and for systems with infinite invariant measures, the complexity of the orbits is null, hence it does not give apparently any information on the system. However some information can be deduced looking at the asymptotic behaviour of the *AIC* of the orbits, as the length/time increases. Indeed the *AIC* converges to infinity, and the fact that the complexity is null means that the *AIC* grows less than linearly. In Chapter 3 we introduce the notions of *global* and *local chaos indexes*, that measure the grow of the *AIC*. Then in Section 3.4, we study the relations between our chaos indexes and other classical indexes introduced to study chaotic systems: the sensitivity to initial conditions, the Poincaré return time and the decay of correlations.

The usefulness of our method in the characterization of dynamical systems is shown in Chapter 4. Here we study some well known chaotic dynamical systems, such as the rotations on the unit interval and the interval exchange transformations, the logistic map and the Manneville maps. In particular the first examples are weakly chaotic systems, whereas the Manneville maps are systems with infinite invariant measures. These maps are *intermittent* systems, in the sense that their behaviour alternates between regular phases and chaotic bursts, typical behaviour of turbulent fluids. The rate with which these two phases alternate is related to the rate of grow of the *AIC* of the symbolic orbits of the system. The maps depend on a parameter $z > 1$, and for $z > 2$ the *AIC* grows as n^α, where $\alpha = \frac{1}{z-1}$.

The logistic map is instead an example of a family of *unimodal* maps on the unit interval. This map is important because it was one of the first studied chaotic systems, thanks to the simplicity of its definition and to the rich amount of different behaviours it presents, as the parameter λ, on which it depends, varies between 1 and 4. In this work we consider two different values of the parameter. The first is the so-called *edge of chaos*. As λ grows from 1, the map has different stable periodic orbits, of period growing as a power of two, and each time the new periodic orbit is created in a *period doubling bifurcation*. The values of λ at which these bifurcations occur accumulate to a value λ_∞ (the edge of chaos). For $\lambda > \lambda_\infty$ the map becomes chaotic, even if there is also a countable number of windows of periodicity. The second value of λ we are interested in, is one of the values $\lambda > \lambda_\infty$ at which a periodic window is created. In particular we consider the period 3 window. The window is created in a *saddle-node bifurcation*, and it is interesting to our aims because, at the bifurcation, the map presents the phenomenon of intermittency, similar to that of the Manneville maps.

In the second part of this work we face the problem of the application of our method to the study of real time series. If one wants to obtain information about a system, simply from the time series obtained by a real experiment, one has to face the problem of the presence of noise in the data. This is a very deep problem, and the difficulty is that the effects of the noise can hide the properties of chaoticity of the system. Namely if one measures a given Kolmogorov-Sinai entropy, the question is how much it is due to the noise and how much to the dynamics. In Chapter 5 we show how the application of our method to real time series can help in the answer to this question. Our approach is based on the size of the partition of the phase space used to generate the symbolic words from the dynamical system. In experiments this corresponds to the scale of the observations of the system. The result is that the complexity of the symbolic orbits changes as the size of the partition decreases, and it has a change of behaviour when the size of the partition is comparable to the variance of the noise present in the data.

However in the application of our method, there is another important problem. In this work we use the notion of AIC to measure the information quantity contained in a word, but the AIC is proved not to be a computable function. That is it does not exist an algorithm able to compute the AIC for all the words on a given alphabet. Then what can be done is to approximate the AIC by a computable information function. This can be done using a *compression algorithm*. In Chapter 6 we recall some basic facts from the theory of compression algorithms, and in particular study the behaviour of two algorithms on dynamical systems: $LZ77$ ([ZL1]) and *CASToRe* ([Ar],[Be]). We also show some numerical experiments on the previously studied dynamical systems. We find that in these cases the theoretical results obtained using the AIC, are found also numerically using $LZ77$ and *CASToRe*. Hence these two algorithms seem to approximate well the behaviour of the AIC for our aims.

Part I

Theory and results

Chapter 1

An overview on dynamical systems

Since the introduction of the concept of *chaos*, it has appeared clear that the statistical approach to the study of chaotic systems was the right approach to obtain qualitative and quantitative results. In this framework were developed the ideas of *unpredictability*, *ergodicity* and *entropy*.

In this chapter we give a review of results in the theory of dynamical systems, with particular emphasis on extensions of the classical ergodic theory and on decay of correlations. For the classical results see for example [CFS], [KH], [Ma], [Pet], and for infinite ergodic theory see [Aa].

1.1 Basic definitions and notations

Let X be a compact metric space, and denote the metric on X by d. We denote by $\mathcal{B}$ the Borel σ-algebra on X relative to d. Let $T : X \to X$ be a measurable map with respect to $\mathcal{B}$, not necessarily invertible. If T is not invertible, given a set $B \in \mathcal{B}$, by $T^{-1}(B)$ we denote the counterimage of B. In the following when we talk of a *dynamical system*, we consider the triplet $(X, \mathcal{B}, T)$.

Let μ be a non-negative measure on $(X, \mathcal{B})$. When X is a subset of the real space $\mathbb{R}^n$, we denote by m_n the Lebesgue measure on X, and use $m = m_1$.

Definition 1.1. A measure μ on $(X, \mathcal{B})$ is *invariant* with respect to T if for any $B \in \mathcal{B}$ it holds $\mu(T^{-1}(B)) = \mu(B)$.

When we consider an invariant measure μ on $(X, \mathcal{B}, T)$, we talk of the *invariant dynamical system* (X, T, μ). In general we will always assume to deal with *non-singular* dynamical systems, that is, even when μ is not invariant, we assume that for all $B \in \mathcal{B}$ it holds that $\mu(T^{-1}(B)) = 0$ if and only if $\mu(B) = 0$.

An important distinction in the theory of dynamical systems is made between systems whose measure μ is finite or not. In particular we will call *finite dynamical system* the dynamical system (X, T, μ) with $\mu(X) = 1$, and *infinite dynamical system* the system with $\mu(X) = +\infty$. For the infinite systems we will always assume that the measure μ is σ-finite.

A particular example of dynamical system is the well-known *symbolic dynamical system*. Given a finite alphabet $\mathcal{A}$, for example $\mathcal{A} = \{0, 1, \dots, N-1\}$, we construct the set $\mathcal{A}^n$ of all the words s of length n made by symbols of the alphabet $\mathcal{A}$. So, if by s_i we denote the i-th letter of the word s, we have that $s \in \mathcal{A}^n$ if $s = (s_0, s_1, \dots, s_{n-1})$ with $s_i \in \mathcal{A}$ for all $i = 0, \dots, n-1$. In this way it is possible to construct the set $\mathcal{A}^*$ of all the finite words in the alphabet $\mathcal{A}$, that is

$$\mathcal{A}^* = \bigcup_{n \in \mathbb{N}} \mathcal{A}^n \tag{1.1}$$

where $\mathcal{A}^0 = \{\lambda\}$, being λ the empty word.

Analogously we define $\Omega = \mathcal{A}^{\mathbb{N}}$ to be the set of all the infinite words in the alphabet $\mathcal{A}$. That is

$$\Omega = \Big\{\omega = (\omega_i)_{i \in \mathbb{N}} \;/\; \omega_i \in \mathcal{A} \;\; \forall\, i \in \mathbb{N}\Big\} \tag{1.2}$$

From now on we call *sequences* the elements of Ω. If we consider the alphabet $\mathcal{A} \subset \mathbb{N}$ and consider the distance $\bar{d}$ on $\mathcal{A}$ given by $\bar{d}(i,j) = |i - j|$, then the space Ω can be equipped by the product topology, that induces the metric d between two sequences ω, ω' given by

$$d(\omega, \omega') = \sum_{i \in \mathbb{N}} \frac{|\omega_i - \omega'_i|}{2^i} \tag{1.3}$$

The space (Ω, d) is a compact metric space, and the neighbourhoods of a sequence ω are defined by

$$C(\omega, k) = \Big\{\omega' \in \Omega \;/\; \omega'_i = \omega_i \;\; \forall\, i = 0, \dots, k-1\Big\} \tag{1.4}$$

These sets are called *cylinders*, and the σ-algebra they induce, denoted by $\mathcal{C}$, is the *cylinders σ-algebra*.

On the space Ω, we consider the action of the so-called *shift map* τ, which acts on Ω continuously (with respect to d), shifting on the left the sequences. Given a sequence ω, the image $\tau(\omega) \in \Omega$ is given by

$$(\tau(\omega))_i = \omega_{i+1} \quad \forall\, i \in \mathbb{N} \tag{1.5}$$

The triplet $(\Omega, \mathcal{C}, \tau)$ is the *symbolic dynamical system*. One possible invariant probability measure μ on (Ω, τ) is defined in the following way: let

P be a probability measure on the discrete space $(\mathcal{A}, \bar{d})$, then we define the measure μ on the cylinders by

$$\mu(C(\omega, k)) = \prod_{i=0}^{k-1} P(\omega_i) \tag{1.6}$$

The measure μ on $(\Omega, \mathcal{C})$ is then defined by extending it from the cylinders to the cylinders σ-algebra.

The symbolic dynamics is an important tool to study dynamical systems. Indeed given a system (X, T, μ), we consider a finite partition $Z = \{I_0, \dots, I_{N-1}\}$ of the phase space X in measurable sets I_j. To the partition Z it is associated the alphabet $\mathcal{A} = \{0, \dots, N-1\}$. Then we construct the *symbolic representation* of the system (X, T, μ) by means of a function $\varphi_Z : X \to \Omega := \mathcal{A}^{\mathbb{N}}$ defined by

$$\varphi_Z(x) = \omega \iff T^i(x) \in I_{\omega_i} \quad \forall\, i \in \mathbb{N} \tag{1.7}$$

The image $\Omega' = \varphi_Z(X)$ is a closed subset of Ω which is invariant under the shift map τ.

For the function φ_Z it holds $\varphi_Z \circ T = \tau \circ \varphi_Z$, hence if it is also a homeomorphism of X onto Ω', then φ_Z is a *conjugacy* between the dynamical systems (X, T) and (Ω', τ). Further if μ is an invariant measure on (X, T) then φ_Z induces an invariant measure ν on (Ω', τ). We will use extensively the symbolic representation of a system.

1.2 Ergodic theory

The first statistical approach to dynamical systems dates to Poincaré and his *Recurrence Theorem.* The Poincaré Recurrence Theorem studies the probability for the orbits starting from a measurable set $B \in \mathcal{B}$ to return to B.

Theorem 1.2 (Poincaré Recurrence Theorem). *Let (X, T, μ) be a finite invariant dynamical system, then for all $B \in \mathcal{B}$ we have that*

$$\mu\left(\left\{x \in B \;/\; T^n(x) \in B \;\; \textit{for infinite } n \in \mathbb{N}\right\}\right) = \mu(B)$$

For the same recurrence property to hold for infinite dynamical systems, it is necessary to add the notion of *conservativity* of a dynamical system (X, T, μ).

Definition 1.3. Given a dynamical system (X, T, μ), a set $W \subset X$ is called *wandering* if $\{T^{-n}W\}_{n \in \mathbb{N}}$ are disjoint. The *dissipative part* of the dynamical system (X, T, μ) is a set $\mathcal{D}(T)$ given by a countable union of wandering sets, such that any wandering set is contained in $\mathcal{D}(T)$. The dynamical system (X, T, μ) is *conservative* if $\mu(\mathcal{D}(T)) = 0$.

In particular any finite invariant dynamical system is conservative. The following is a generalization of the Poincaré Recurrence Theorem, valid also in the case of an infinite dynamical system.

Theorem 1.4 (Halmos Recurrence Theorem). *Let (X, T, μ) be a conservative dynamical system, then for all $B \in \mathcal{B}$ we have that*

$$\sum_{n=1}^{\infty} \chi_B(T^n(x)) = \infty$$

for μ-almost all $x \in B$, where $\chi_B(\cdot)$ is the characteristic function of the set B.

In the following we will see that also the rate of returns to the measurable sets has an important role. Given a conservative dynamical system (X, T, μ), let B be a measurable set of positive measure, $\mu(B) > 0$. Then Theorem 1.4 assures that for μ-almost any $x \in B$ there exists[1] an integer $n(x) \geq 1$ such that $T^{n(x)}(x) \in B$. Then it is possible to define for μ-almost all $x \in B$ a function $n_B : B \to \mathbb{N}$ given by

$$n_B(x) = \inf\{n \geq 1 \ / \ T^n(x) \in B\} \tag{1.8}$$

We call the function n_B the *first return time in B*. This function is then used to define the *induced map* on B, the map $T_B : B \to B$ that associates to a point x its first return point in B. Namely

$$T_B(x) = T^{n_B(x)}(x) \tag{1.9}$$

Another important quantity is the first return time $\tau(B)$ for a set B to itself. The Halmos Recurrence Theorem implies that $\tau(B)$ is finite.

Definition 1.5. Let (X, T, μ) be a conservative dynamical system. Given a measurable set $B \subset X$ with $\mu(B) > 0$, we define the *Poincaré recurrence time* $\tau(B)$ as

$$\tau(B) = \min\left\{n \geq 1 \ / \ T^n B \cap B \neq 0\right\} = \inf\left\{n_B(x) \ / \ x \in B\right\}$$

One way to study the statistical properties of an invariant dynamical system (X, T, μ) is to consider the values of an observable $f : X \to \mathbb{R}$ along the orbits of the system, and to consider the mean of these values. The hope is that this mean gives information on the mean value of the observable f for the system. The notion of *ergodicity* has been created to fulfill this hope.

Given a function $f : X \to \mathbb{R}$ in $L^1(X, \mu)$ define

$$S_n f(x) = \sum_{i=0}^{n-1} f(T^i(x)) \tag{1.10}$$

then

[1] Actually there exist infinitely many such integers.

Theorem 1.6 (Birkhoff Ergodic Theorem). *For a finite invariant dynamical system (X, T, μ), for all $f \in L^1(X, \mu)$ the following properties hold:*

a) for μ-almost all $x \in X$ the limit $\lim_{n\to\infty} \frac{1}{n} S_n f(x)$ exists and let $\tilde{f}(x)$ be its value;

b) $\tilde{f}$ is an invariant function on X, that is $\tilde{f}(T(x)) = \tilde{f}(x)$ where it is defined;

c) $\tilde{f} \in L^1(X, \mu)$ and $\|\tilde{f}\|_{L^1} \leq \|f\|_{L^1}$;

d) for all invariant $B \in \mathcal{B}$, that is $T^{-1}(B) = B$, it holds $\int_B f(x)\, d\mu(x) = \int_B \tilde{f}(x)\, d\mu(x)$;

e) $\frac{1}{n} S_n f \to \tilde{f}$ as $n \to \infty$ in $L^1(B)$ for all invariant $B \in \mathcal{B}$.

As for the recurrence property, the extension of the ergodic theorem to infinite invariant dynamical systems needs the property of conservativity. The following is the generalization of the Birkhoff Ergodic Theorem, valid also for infinite dynamical systems.

Theorem 1.7 (Hopf-Stepanov Ergodic Theorem). *For a conservative invariant dynamical system (X, T, μ), for all $f, p \in L^1(X, \mu)$ and $p > 0$, it holds*

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} f(T^i(x))}{\sum_{i=0}^{n-1} p(T^i(x))} = \mathbb{E}_{\mu_p}\left(\frac{f}{p}\Big|\mathcal{I}\right)(x) \tag{1.11}$$

for μ-almost all $x \in X$, where $\mathbb{E}_{\mu_p}(\cdot|\mathcal{I})$ denotes the conditional expectation with respect to the σ-algebra $\mathcal{I}$ of the measurable T-invariant sets of X, and with respect to the measure μ_p given by $d\mu_p = p d\mu$.

In the Hopf-Stepanov Ergodic Theorem, notice that the conditional expectation $\mathbb{E}_{\mu_p}$ in the right hand side is a T-invariant function.

When $p \equiv 1$ and the dynamical system is finite, the Hopf-Stepanov Ergodic Theorem becomes the Birkhoff Ergodic Theorem, hence the right hand side of equation (1.11) is equal to the $\tilde{f}(x)$.

To obtain information on the system (X, T, μ) using the observable f, it would be important to have a value for $\tilde{f}$ and for the conditional expectation $\mathbb{E}_{\mu_p}$ independent on the initial condition $x \in X$. This is obtained for an *ergodic* dynamical system.

Definition 1.8. A measure μ on the dynamical system (X, T) is *ergodic* if for all T-invariant sets $B \in \mathcal{B}$, that is $T^{-1}(B) = B$, it holds $\mu(B) = 0$ or $\mu(B^c) = 0$, where $B^c = X \setminus B$.

The following are classical results on ergodic dynamical systems.

Proposition 1.9. *For a conservative invariant dynamical system (X, T, μ) the following conditions are equivalent:*

a) μ is ergodic;

b) all measurable observables $f : X \to \mathbb{R}$ that are T-invariant, that is $f(T(x)) = f(x)$ for μ-almost all $x \in X$, are constant μ-almost anywhere.

As a corollary to the previous proposition and to the ergodic theorems, we have

Corollary 1.10. *If (X, T, μ) is an ergodic conservative invariant dynamical system then for μ-almost all $x \in X$ and for all $f, p \in L^1(X, \mu)$ and $p > 0$, it holds*

$$\mathbb{E}_{\mu_p}\left(\frac{f}{p}\Big|\mathcal{I}\right)(x) = \frac{\int_X f(x)\, d\mu(x)}{\int_X p(x)\, d\mu(x)}$$

In particular if (X, T, μ) is also a finite system and $p \equiv 1$ then $\tilde{f}(x) = \int_X f(x)\, d\mu(x)$ for all $f \in L^1(X, \mu)$ and for μ-almost all $x \in X$. If instead (X, T, μ) is an infinite system, for μ-almost all $x \in X$ and for all $f \in L^1(X, \mu)$ it holds

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = 0$$

Corollary 1.11. *For an invariant dynamical system (X, T, μ), the probability measure μ is ergodic if and only if for all measurable $A, B \subset X$ it holds*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) = \mu(A)\mu(B) \tag{1.12}$$

If μ is infinite and ergodic and the system is conservative then for all measurable $A, B \subset X$ of finite measure

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) = 0 \tag{1.13}$$

It appears clear that when the invariant measure μ on the dynamical system (X, T) is a probability measure, the results obtained above give more important properties for the dynamical system.

By the Halmos Recurrence Theorem 1.4, it was possible to define the *induced map* T_B on a subset $B \in \mathcal{B}$ of positive measure (see equation (1.9)). To study ergodicity for the induced map of a transformation we have the following

Theorem 1.12 ([Aa]). *Let (X, T, μ) be a conservative dynamical system, then for all $B \in \mathcal{B}$ such that $0 < \mu(B) < \infty$ it holds:*

i) if (X, T, μ) is ergodic then the dynamical system (B, T_B, μ_B) is ergodic;

ii) if (B, T_B, μ_B) is ergodic and $\cup_{n \geq 1} T^{-n}(B) = X (mod\ \mu)$, then (X, T, μ) is ergodic.

where $T_B : B \to B$ is the induced map, and μ_B is the conditional measure.

Theorem 1.13 (Kac). *If (X, T, μ) is an ergodic conservative invariant dynamical system, then for all $B \in \mathcal{B}$ such that $0 < \mu(B) < \infty$ it holds*

$$\int_B n_B(x)\ d\mu(x) = \mu(X) \tag{1.14}$$

where $n_B : B \to \mathbb{N}$ is the first return time defined in equation (1.8).

In the case of a finite system, the result of Kac's Theorem can be restated by saying that the mean first return time to B for points $x \in B$ is $\frac{1}{\mu(B)}$.

The Halmos Recurrence Theorem (Theorem 1.4) assures that μ-almost any point x in a measurable set B of positive measure $\mu(B)$ will return to B, and the theorem of Kac studies the mean of the first return time $n_B(x)$ for points $x \in B$.

In particular, using the result of Kac's theorem for a finite ergodic invariant dynamical system, if we write equation (1.12) with $A = B$, we obtain

$$\lim_{n \to \infty} \frac{\frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(B) \cap B)}{\mu(B)} = \mu(B) = \frac{1}{\text{mean first return time}}$$

Giving the same interpretation to the result that we obtain in the case of an infinite ergodic invariant dynamical system, that is using equation (1.13), we have that the mean first return time to a set of finite measure for points in this set is infinite.

Another consequence of Kac's theorem is an estimate for the Poincaré recurrence time of a set to itself (see Definition 1.5). Indeed, for an invariant dynamical system with probability measure μ, if Poincaré Recurrence Theorem implies that $\tau(B)$ is finite, from the theorem of Kac we find that $\tau(B) \leq \frac{1}{\mu(B)}$.

Some possible extensions of ergodic theory come from an analysis of the results of Theorem 1.6, Proposition 1.9 and of Corollary 1.10. First of all it would be interesting to study the speed of convergence of $\frac{1}{n} S_n f$ to $\tilde{f}$. Moreover one could study some extensions of the ergodic theorem for two classes of problems: (i) in the case of an ergodic dynamical system with a probability invariant measure, the convergence for an observable with $\int_X f d\mu = \infty$; (ii) in the case of infinite invariant measure μ, the convergence of $\frac{1}{a_n} S_n f$ for an integrable observable f and for some sequence a_n such that

$\lim a_n = \infty$ and $\lim \frac{a_n}{n} = 0$ as $n \to \infty$; the hope is to obtain as limit something different from $\tilde{f} = 0$ for μ-almost all $x \in X$. We now briefly state the main results in the two directions. For more detailed discussions see for example [Aa] and [Pet].

Regarding the speed of convergence in the ergodic theorem, it is not possible to find a fixed speed of convergence for all observables f. Indeed we have

Theorem 1.14 (Krengel [Kr2]). *For any finite ergodic invariant dynamical system (X, T, μ), such that μ is non-atomic, and for any sequence $a_n \to 0$ as $n \to \infty$, there is a bounded measurable function $f : X \to \mathbb{R}$ such that*

$$\limsup_{n\to\infty} \frac{|\frac{1}{n}\ S_n f(x) - \int_X f(x)\ d\mu(x)|}{a_n} = \infty$$

for μ-almost any $x \in X$.

Some results can be obtained studying the order of the convergence for a fixed function f with respect to other statistical properties of the observable, such as the decay of the correlations ([Ka]).

Negative results are obtained also in the second direction. Namely the following two theorems tell us that in general it is not possible to get better results than those given by the Birkhoff Ergodic Theorem.

In the case of observables with infinite integral on (X, μ), in the case $\mu(X) = 1$, the ergodic theorem would tell us that $\frac{1}{n} S_n f \to \infty$. The idea is then to find a sequence a_n growing faster than n such that $\frac{1}{a_n} S_n f$ converges to some constant μ-almost anywhere. This is not the case.

Theorem 1.15 (Aaronson [Aa]). *Let (X, T, μ) be a finite ergodic invariant dynamical system. Let $f : X \to \mathbb{R}$ be a non-negative measurable function such that $\int_X f(x)\ d\mu(x) = \infty$. Then does not exists a sequence a_n such that the limit*

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{i=0}^{n-1} f(T^i(x))$$

exists and is equal to a real non-zero number for μ-almost all $x \in X$.

Instead in the case of an infinite ergodic measure μ, Corollary 1.10 tells us that $\frac{1}{n} S_n f(x) \to 0$ for μ-almost all $x \in X$. Hence the idea would be to find a sequence growing slower than n such that $\frac{1}{a_n} S_n f$ converges to some non-zero constant μ-almost anywhere. Again the next theorem shows that it is impossible.

Theorem 1.16 (Aaronson [Aa]). *Let (X, T, μ) be an infinite ergodic conservative invariant dynamical system. Let $f : X \to \mathbb{R}$ be a non-negative* *function in $L^1(X,\mu)$. Then does not exist a sequence a_n such that the limit*

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{i=0}^{n-1} f(T^i(x))$$

exists and is equal to a real non-zero number for μ-almost all $x \in X$.

1.2.1 Decay of correlations

We have seen that Birkhoff and Hopf-Stepanov Ergodic Theorems and their corollaries study the convergence for the so-called "temporal mean" $\frac{1}{n}S_n f$ of observables $f \in L^1$ along the trajectories of a dynamical system. In particular Corollary 1.11 is obtained studying the mean for the characteristic functions of measurable sets. The result in equation (1.12) can be seen as a result about the probabilistic independence of the two sets $T^{-n}(A)$ and B, just observing that $\mu(T^{-n}(A)) = \mu(A)$ for all $n \geq 0$. We can say that the correlation between the two events "to be in $T^{-n}(A)$" and "to be in B" decreases in mean as n grows to infinity.

The notion of *mixing* has been introduced for a dynamical system to have a stronger result in this direction.

Definition 1.17. An invariant dynamical system (X,T,μ) is *mixing* if when μ is finite then for all measurable sets $A, B \subset X$ it holds

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B) \; ; \tag{1.15}$$

whereas when μ is infinite and the system is conservative the limit in equation (1.15) is zero for all measurable sets $A, B \subset X$ of finite measure.

In the statistical approach to dynamical systems it is studied the action of the map T on some quantities defined on the system. Besides the observables defined by measurable functions $f : X \to \mathbb{R}$, it can also be studied the action of T on the family of probability measures $\mathcal{M}(X)$ defined on $(X, \mathcal{B})$, that are not necessarily invariant. A first result in this direction is the following.

Proposition 1.18. *If the invariant dynamical system (X,T,μ) is mixing then given a probability measure ν on $(X,\mathcal{B})$ absolutely continuous with respect to μ ($\nu << \mu$), if μ is a probability measure for all $B \in \mathcal{B}$ it holds*

$$\lim_{n\to\infty} \nu(T^{-n}(B)) = \mu(B) \; ;$$

whereas when μ is infinite and the system is conservative the previous limit is equal to zero for all measurable sets B.

In general it is possible to define an operator acting on the set of density functions for the probability space $(X, \mathcal{B}, \mu)$. Namely, let $D(X,\mu)$ be the space of $L^1(X,\mu)$ non-negative functions such that $\|\varphi\|_{L^1} = 1$. The space $D(X,\mu)$ is the space of *density functions.* In particular the function $\varphi(x) \equiv 1$ is in $D(X,\mu)$. Given the map T on $(X,\mathcal{B})$ such that (X,T,μ) is an invariant dynamical system and μ is a probability space, the *Perron-Frobenius operator* $\mathcal{P}$ studies the action of T on $D(X,\mu)$. Hence $\mathcal{P} : D(X,\mu) \to D(X,\mu)$ and for a density function $\varphi \in D(X,\mu)$ the image $\mathcal{P}\varphi$ is defined by the following relation

$$\int_B (\mathcal{P}\varphi)(x)\ d\mu(x) = \int_{T^{-1}(B)} \varphi(x)\ d\mu(x) \tag{1.16}$$

for all measurable $B \subset X$. In particular we notice that fixed points of the Perron-Frobenius operator correspond to invariant measures of the dynamical system. So the function $\varphi(x) \equiv 1$ is a fixed point for $\mathcal{P}$ since it corresponds to the measure μ.

The Perron-Frobenius operator $\mathcal{P}$ can be generalized to $L^1(X,\mu)$ functions using the same relation (1.16), and in general it satisfies the following important relation: given any $\varphi \in L^1(X,\mu)$ and any $\psi \in L^\infty(X,\mu)$ it holds

$$\int_X (\mathcal{P}\varphi)(x)\ \psi(x)\ d\mu(x) = \int_X \varphi(x)\ \psi(T(x))\ d\mu(x) \tag{1.17}$$

Going back to the mixing property, we recall that it can be interpreted as a property of "loss of correlation" for the characteristic functions of the phase space. Equation (1.17) gives a relation between this loss of correlation and the Perron-Frobenius operator. Indeed it is enough to notice that

$$\mu(T^{-n}(A) \cap B) = \int_X \chi_A(T^n(x))\ \chi_B(x)\ d\mu(x) = \int_X \chi_A(x)\ (\mathcal{P}^n \chi_B)(x)\ d\mu(x)$$

where χ_A and χ_B are the characteristic functions of the measurable sets A and B. Hence equation (1.15) becomes

$$\lim_{n\to\infty} \int_X \chi_A(T^n(x))\ \chi_B(x)\ d\mu(x) = \left(\int_X \chi_A(x)\ d\mu(x)\right)\left(\int_X \chi_B(x)\ d\mu(x)\right)$$

The idea is then to generalize this property to functions φ on some Banach space.

Definition 1.19. Given a finite invariant dynamical system (X,T,μ), the *correlation functions* $C_n(\varphi,\psi)$ for functions φ and ψ in some Banach space $\mathcal{F}$ are defined by

$$C_n(\varphi,\psi) = \int_X \varphi(T^n(x))\psi(x)\ d\mu(x) - \left(\int_X \varphi(x)\ d\mu(x)\right)\left(\int_X \psi(x)\ d\mu(x)\right) \tag{1.18}$$

Definition 1.20. An invariant dynamical system (X, T, μ) such that μ is a probability measure is called *exponentially mixing* on the Banach space of functions $\mathcal{F}$ if there exist a constant $C > 0$ and a real number $\tau < 1$ such that for all $\varphi, \psi \in \mathcal{F}$ it holds

$$|C_n(\varphi, \psi)| \leq C\ \tau^n \quad \forall\ n \in \mathbb{N}$$

In some cases it is possible that the Banach space $\mathcal{F}$ needed to obtain the property of exponential mixing does not contain the characteristic functions. Indeed for example let $T : (X, d) \to (X, d)$ be an *expanding* function, that is there exists $\sigma > 1$ such that $d(T(x), T(y)) \geq \sigma d(x, y)$ for all $x, y \in X$, then it holds

Theorem 1.21 ([Vi]). *If (X, T, μ) is an invariant dynamical system with a probability measure μ, and if T is expanding and of class $C^{1+\nu_0}$, then for all $\nu \in (0, \nu_0]$ the dynamical system (X, T, μ) is exponentially mixing on the Banach space $C^{0,\nu}(X, \mathbb{R})$ of the ν-Hölder real continuous functions on X.*

The speed with which the correlation functions decay has been widely studied in the last years, showing in particular the existence of systems with speed of decay different from the exponential one ([Is1],[LSV],[Is2]). Moreover the speed of decay of correlations has been related to other properties of the dynamical system, as for example the spectral properties of the Perron-Frobenius operator on the density functions ([Ba]) and the behaviour of the first return time (see equation (1.8)) for measurable subsets of X ([Yo]).

1.3 Chaotic systems

The notion of *chaos* is nowadays very common and there are many different definitions for a chaotic dynamical system. Anyway it is clear that it is associated to a lack of periodicity for the orbits of the system, and in particular to a high degree of unpredictability for the evolution of the system given its initial conditions. Many indicators have been introduced to study chaotic dynamical systems. In this section we present some classical definitions with special emphasis on *entropy*.

We now give the definition of chaos for a dynamical system $(X, \mathcal{B}, T)$ following [De].

Definition 1.22. The dynamical system $(X, \mathcal{B}, T)$ is *topologically transitive* if there exists a point $x \in X$ such that its orbit $\{T^i(x)\}_{i \in \mathbb{N}}$ is dense.

Definition 1.23. A dynamical system $(X, \mathcal{B}, T)$ has *sensitivity to initial conditions* if there exists $\delta > 0$ such that, for all $x \in X$ and for all neighbourhoods U of x, there exist $y \in U$ and $n \in \mathbb{N}$ such that $d(T^n(x), T^n(y)) > \delta$.

Definition 1.24. A dynamical system $(X, \mathcal{B}, T)$ is *chaotic* if it is topologically transitive, has sensitivity to initial conditions and the periodic points are dense in X.

The most interesting property of a chaotic dynamical system is for our aims the sensitivity to initial conditions. It shows how two orbits that have nearby initial conditions become sooner or later far, at least of a fixed distance δ. It is important also to remark that this behaviour must occur in any neighbourhood of any point $x \in X$. So there cannot be subsets of the phase space with some kind of attraction at any small scale.

In Chapter 4 we will study the *logistic map* at the chaos threshold, showing that it is not a chaotic dynamical system, but nevertheless its orbits are unpredictable. In that case we will see how this unpredictability is related to the separation of orbits with nearby initial conditions, even if the system has not *sensitivity to initial conditions.* In that case it is better to study the behaviour of the *separation sets* $B(x, \epsilon, n)$, introduced in [BK].

Definition 1.25. Given a dynamical system $(X, \mathcal{B}, T)$, for any $x \in X$ and for $\epsilon > 0$ and $n \in \mathbb{N}$, the (ϵ, n)*-separation set* of x is the set

$$B(x, \epsilon, n) = \Big\{ y \in X \ / \ d(T^i(x), T^i(y)) \le \epsilon \ \ \forall \ i = 0, \dots, n \Big\}$$

The system $(X, \mathcal{B}, T)$ has sensitivity to initial conditions if and only if there exists a δ such that for all $x \in X$ the (δ, n)-separation set of x shrinks to x as $n \to \infty$.

Anyway, when we talk of chaotic systems, we assume that the separation sets shrinks to a point as $n \to \infty$.

1.3.1 Entropy

One of the classical indicators that have been introduced to characterize chaotic dynamical systems is the notion of *Kolmogorov-Sinai (KS) entropy.* In this section we briefly review its basic properties.

Before introducing KS entropy, we talk of the *topological entropy*, which is a weaker version of KS entropy. Given a dynamical system $(X, \mathcal{B}, T)$, a positive integer n and a real $\epsilon > 0$, a set $A \subset X$ is called (n, ϵ)*-separated* for T if, for every pair of distinct points $x, y \in A$, there is at least one k with $0 < k < n$ such that $d(T^k(x), T^k(y)) > \epsilon$. We can now define the *number of different orbits of length n at scale ϵ* as

$$r(n, \epsilon, T) = \max\{\#A \ / \ A \subset X \text{ is a } (n, \epsilon)\text{-separated set for } T\} \qquad (1.19)$$

that is the number of points that, with resolution ϵ, cover the set X with their n-long orbits.

The *topological entropy* $h_{top}(T)$ *of a dynamical system* (X, T) is then defined as the limit of the growth rate of $r(n, \epsilon, T)$ as $\epsilon \to 0$. We have

$$h_{top}(T) = \lim_{\epsilon \to 0^+} \limsup_{n \to \infty} \frac{\log(r(n, \epsilon, T))}{n} \tag{1.20}$$

Hence topological entropy measures the exponential divergence of nearby orbits, giving the ratio of loss of information as the system evolves.

We remark that in the definition of the topological entropy, we considered the number of different orbits at a fixed scale and simply counted them. This is an information on the unpredictability of the system, since the faster the number of different orbits grows the more are the possible evolutions for our system given the initial condition. Hence in systems with high topological entropy we have a high degree of unpredictability.

However, topological entropy is only a first way to look at a dynamical system. Indeed to have more information on the possible evolutions of a system, it would be important to know how likely a single evolution occurs. This corresponds to not only count the number of different orbits, but weigh them according to their probability of occurrence. To this aim we need an invariant probability measure μ on the dynamical system $(X, \mathcal{B}, T)$. The measure needs to be invariant since we want the dynamics to evolve following the same probability law at any time. This is the idea of the *KS entropy.*

Definition 1.26. Given a finite invariant dynamical system (X, T, μ) and a finite measurable partition $Z = \{I_0, \dots I_{N-1}\}$ of the space X, the *entropy* $H_\mu(Z)$ of the partition Z is defined as

$$H_\mu(Z) = -\sum_{i=0}^{N-1} \mu(I_i)\ \log(\mu(I_i)) \tag{1.21}$$

with the assumption that $0 \log(0) = 0$.

Let now P and Q be two different finite measurable partitions of the space X. It is possible to define the *conditional entropy* $H_\mu(P|Q)$ of P with respect to Q. It is defined by

$$H_\mu(P|Q) = -\sum_{i,j} \mu(P_i \cap Q_j)\ \log\left(\frac{\mu(P_i \cap Q_j)}{\mu(Q_j)}\right) \tag{1.22}$$

On the space of the finite measurable partitions, this allows the definition of a metric, the so-called *Rokhlin metric* $d_{(R,\mu)}$, given by

$$d_{(R,\mu)}(P, Q) := H_\mu(P|Q) + H_\mu(Q|P) \tag{1.23}$$

As we have seen before, the idea of KS entropy is related to the increase of information needed to specify a single orbit as the system evolves with time. Hence it is natural to study the evolution of the entropy of the iterated partitions Z_n generated by a finite measurable partition Z. Let $T^{-1}Z$ be the partition given by the counter images $T^{-1}I_i$. Then let

$$Z_n = Z \vee T^{-1}Z \vee \cdots \vee T^{-n+1}Z \tag{1.24}$$

be the partition given by the sets of the form

$$I_{i_0} \cap T^{-1}I_{i_1} \cap \cdots \cap T^{-n+1}I_{i_{n-1}},$$

varying I_{i_j} among all the sets of Z. To know to which set of the partition Z_n a point x belongs to, corresponds to know what are the sets of the partition Z that the orbit of x visits up to time $n-1$. The *Kolmogorov-Sinai (KS) entropy* $h_\mu(T,Z)$ *relative to the partition* Z is defined as the limit

$$h_\mu(T,Z) = \lim_{n\to\infty} \frac{1}{n} H_\mu(Z_n) \tag{1.25}$$

As a remark we notice that the limit performed in equation (1.25) is an infimum, since the sequence on the right hand side is decreasing. There are other ways to compute the KS entropy relative to a partition, using the conditional entropy of a partition with respect to its future or to its past. Indeed we have

$$h_\mu(T,Z) = \lim_{n\to\infty} H_\mu(Z| \vee_{i=1}^{n-1} T^{-i}Z) = \lim_{n\to\infty} H_\mu(T^{-n}Z|Z_n) \tag{1.26}$$

We are now ready to define the *KS entropy.*

Definition 1.27. The *Kolmogorov-Sinai (KS) entropy* $h_\mu(T)$ of a finite invariant dynamical system (X,T,μ) is defined as

$$h_\mu(T) = \sup \left\{ h_\mu(T,Z) \ / \ Z \text{ finite measurable partition} \right\} \tag{1.27}$$

A fundamental property of the KS entropy which is evident from its definition is its invariance with respect to conjugacy. So two conjugate systems have the same KS entropy when calculated using the invariant measures related by the conjugacy.

For the explicit computation of the KS entropy, it is not necessary in most cases to perform the supremum over all partitions. Indeed there are some special partitions, called *generating*, for which the supremum is realized. Hence it is enough to compute the KS entropy relative to a generating partition to obtain the KS entropy of the system.

Definition 1.28. A finite measurable partition Z is *generating* for the dynamical system (X,T,μ) if the family of the iterated partitions Z_n is dense in the set of all partitions with respect to the Rokhlin metric $d_{(R,\mu)}$.

Theorem 1.29 (Kolmogorov-Sinai). *If Z is a generating partition for the finite invariant dynamical system (X, T, μ), then $h_\mu(T, Z) = h_\mu(T)$.*

The existence of a generating partition for dynamical systems is given by the following theorem:

Theorem 1.30 (Krieger Generator Theorem [Krie]). *For a finite ergodic invariant dynamical system (X, T, μ) on a Lebesgue space X, such that $h_\mu(T) < \infty$, there is a finite generating partition Z.*

From equations (1.21) and (1.25), it is clear that for a dynamical system (X, T, μ) the KS entropy relative to a partition Z is related to the evolution with time of the measure of the sets of the iterated partitions Z_n. This relation is clarified by the *Shannon-McMillan-Breiman Theorem* and by its corollaries. For each $x \in X$ and all n we denote by $Z_n(x)$ the set of the partition Z_n such that $x \in Z_n(x)$.

Theorem 1.31 (Shannon-McMillan-Breiman). *Let (X, T, μ) be a finite invariant dynamical system and let Z be a finite partition of X. Then the sequence $h_n(x) := -\frac{1}{n} \log(\mu(Z_n(x)))$ is convergent as $n \to \infty$ for μ-almost any $x \in X$ and as a sequence in $L^1(X, \mu)$. If $h_\mu(T, Z, x)$ denotes the limit function of $h_n(x)$ it holds*

$$\int_X h_\mu(T, Z, x) \; d\mu(x) = h_\mu(T, Z)$$

Corollary 1.32. *If (X, T, μ) is also ergodic then $h_\mu(T, Z, x)$ is a constant function μ-almost anywhere and for μ-almost all $x \in X$ it holds*

$$h_\mu(T, Z, x) = h_\mu(T, Z)$$

Corollary 1.33 (Equipartition Property). *For a finite ergodic invariant dynamical system (X, T, μ) and for any finite partition Z it holds: for all $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ the sets of the partition Z_n can be divided into two classes, the typical $\mathcal{T}$ and the negligible $\mathcal{N}$, such that (i) $\mu(\cup_{N \in \mathcal{N}} N) < \epsilon$; (ii) for all $T \in \mathcal{T}$ it holds*

$$2^{-(n+1)(h_\mu(T,Z)+\epsilon)} < \mu(T) < 2^{-(n+1)(h_\mu(T,Z)-\epsilon)} \tag{1.28}$$

The result in equation (1.28) tells us that if $h_\mu(T, Z) > 0$, the measure of the sets in the partitions Z_n converges to zero uniformly, with an exponential speed with constant given by the KS entropy relative to the partition Z. This results loses its importance in the cases $h_\mu(T, Z) = 0$, that are exactly the cases in which we are interested in this work.

The definition of KS entropy has been generalized to the class of infinite ergodic conservative invariant systems by Krengel ([Kr1]), and an equivalent of the Shannon-McMillan Breiman has been obtained for a restricted class of infinite systems ([Zw]).

The Shannon-McMillan-Breiman Theorem and its corollaries move the attention from a global view of the dynamical system to a local scale, studying how the iterated partition evolves in the neighbourhoods of the points of X. As we saw before, also the sensitivity to initial conditions is a local property, and in particular we have a result for the separation sets $B(x, \epsilon, n)$ of points x as defined in Definition 1.25.

Theorem 1.34 (Brin-Katok [BK]). *For a finite invariant dynamical system (X, T, μ), the function*

$$h(T, x) := \lim_{\epsilon \to 0} \lim_{n \to \infty} -\frac{1}{n} \log \Big(\mu(B(x, \epsilon, n)) \Big) \tag{1.29}$$

is defined for μ-almost any $x \in X$ and it holds $\int_X h(T, x)\, d\mu(x) = h_\mu(T)$.

1.3.2 Weak chaos

The KS entropy has been used for characterization of many dynamical systems, for example Ornstein classified Bernoulli shifts ([Or]). However there are systems for which all the classical indicators of chaos, KS entropy included, are null, but still the system has some unpredictability properties. In [FSS] it was introduced the term *weak chaos* to denote these systems.

In this work when talking of *weak chaos* we consider a chaotic dynamical system $(X, \mathcal{B}, T)$ for which the KS entropy is null when calculated with respect to any invariant probability measure that is *physically relevant*.

One way to define an invariant probability measure for a dynamical system is the following: given a point x, let $\delta_{T^i(x)}$ be the Dirac measures concentrated at the point $T^i(x)$ of the orbit of x; then let

$$\mu = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i(x)} \tag{1.30}$$

considering a subsequence when needed[2]. Then we give the following definition (see [ER]).

Definition 1.35. An invariant probability measure μ on $(X, \mathcal{B}, T)$ is a *physical measure* if equation (1.30) holds for all x in a subset $S \subset X$ such that $m_n(S) > 0$, where we recall m_n is the Lebesgue measure on the real space $\mathbb{R}^n$ in which X is embedded.

[2] The existence of this weak* limit is given by the weak* compactness of the space $\mathcal{M}_T(X)$ of invariant probability measures on $(X, \mathcal{B}, T)$.

Chapter 2

Measures of information

We now introduce the method of measuring the *information content* of sequences generated by *information sources*. In Chapter 3 we will apply this method to dynamical systems, using the method of symbolic representation (see equation (1.7)) to interpret a dynamical system as an information source. This chapter is based on [Be].

Let $\mathcal{A} = \{0, \dots, N-1\}$ be a finite alphabet of integers, and recall that we denote by $\mathcal{A}^*$ the set of all finite words s made by letters from $\mathcal{A}$, and by $\Omega = \mathcal{A}^{\mathbb{N}}$ the set of infinite sequences $\omega = (\omega_i)_{i \in \mathbb{N}}$, such that $\omega_i \in \mathcal{A}$ for all $i \in \mathbb{N}$.

Definition 2.1. Given an alphabet $\mathcal{A}$, an *information source* on $\mathcal{A}$ is a stationary stochastic process $\mathbb{Y} = \{Y_n\}_{n \in \mathbb{N}}$ with values in $(\Omega, \mathcal{C})$, where $\mathcal{C}$ is the cylinders σ-algebra generated by the cylinders defined in equation (1.4).

The sequences $\omega \in \Omega$ are then the single realizations of the process $\mathbb{Y}$. Let μ be the measure on $(\Omega, \mathcal{C})$ induced by the process $\mathbb{Y}$. The property of stationarity means that for all $k, h \in \mathbb{N}$ and all $\omega \in \Omega$, given the following definition

$$C(\omega, h, k) = \left\{ \omega' \in \Omega \;/\; \omega'_i = \omega_i \;\forall\; i = h, \dots, h+k-1 \right\}, \tag{2.1}$$

with the assumption that $C(\omega, k) := C(\omega, 0, k)$, it holds

$$\mu(C(\omega, k)) = \mu(C(\omega, h, k))$$

From the point of view of dynamical systems, the space (Ω, μ) is an invariant dynamical system when the map on Ω is the shift map τ defined in equation (1.5). Hence an information source on $\mathcal{A}$ is *ergodic* if such is the invariant dynamical system (Ω, τ, μ).

Given a word $s \in \mathcal{A}^*$, let $I(s)$ denote the *information content* of s. The intuitive idea is that

$I(s)$ is the length of the smallest message from which it is possible to reconstruct s

thus in general the *information content* is a function $I : \mathcal{A}^* \to \mathbb{N}$. Of course it is possible to define many different information functions according to the rule used to reconstruct the words.

In the following sections we analyze three different examples. These examples are fundamentally different for the properties they have. In particular we remark that the *Shannon entropy* is defined only when the word s is considered inside a particular context. That is for example a word of a particular language. Instead the other two notions are dependent only on the word itself. We will see that this difference is very important to our aims. Indeed if dynamical systems with an invariant measure can be considered as information sources with a context (given by the presence of the stationary measure, that represents the law with which words are produced), it is also important to have indicators of information content independent of the context if we want to draw information on an unknown dynamical system, simply analyzing its single orbits. The problem of effectively make computations on a dynamical system is studied in Chapter 6.

Before going into details for the different information functions, we remark that to analyze dynamical systems using measures of information, we need notions of information also for sequences $\omega \in \Omega$. This is because for dynamical systems we need to study the asymptotic behaviour of the orbits as time passes by.

Given a sequence ω we denote by ω^n the finite word made of the first n symbols of ω, namely $\omega^n = (\omega_0, \ldots, \omega_{n-1})$. We can then study the asymptotic behaviour of the information content of the words ω^n as n increases, compared to the length of the words.

Definition 2.2. Given a sequence $\omega \in \Omega$, we call *complexity* of ω, the quantity $K(\omega)$ given by

$$K(\omega) = \limsup_{n\to\infty} \frac{I(\omega^n)}{n} \tag{2.2}$$

Different complexities can be evaluated for a sequence using different information functions.

In the following we will generally study the behaviour for the complexity of sequences produced by ergodic information sources, for which the complexity is a constant μ-almost anywhere. When the information source is not ergodic the quantity to look at will be the mean of $K(\omega)$ over (Ω, μ), namely $\int_\Omega K(\omega) d\mu(\omega)$.

2.1 Shannon entropy

The *Shannon entropy* is the perfect equivalent of Kolmogorov-Sinai entropy (Section 1.3.1) for information sources. Indeed as we have seen before, an information source with finite measure μ can be considered as the finite invariant dynamical system (Ω, τ, μ), and hence it has an associated Kolmogorov-Sinai entropy.

In this section we briefly show how to construct the *Shannon entropy*, with emphasis on its interpretation as an information measure. This interpretation applies to dynamical systems with the obvious analogies. In Chapter 3 we clarify this equivalence.

In 1948 Shannon initiated the mathematical theory of information transmission ([Sh]). His theory was based on some basic assumptions: (i) the amount of information transmitted can be measured by the amount of uncertainty removed on the upcoming symbol; (ii) the uncertainty of the occurrence of a word is minus the logarithm in base 2 of the probability of the word received. So, for example, given the alphabet $\mathcal{A} = \{0, \ldots, N-1\}$ and the probabilities $(p_i)_{i=0,\ldots,N-1}$ of occurrence of a single letter of $\mathcal{A}$, the uncertainty of the occurrence of the i-th letter is given by $-\log_2 p_i$. This assumption comes from the intuitive idea that if we have to transmit a word, and we know that it is chosen in a set of n equiprobable words, then we can just order the words of the set and, if this ordering is known also to the receiver[1], transmit the position of our word in the set. So we have to transmit a number between 1 and n. Since to transmit a number it is usually needed an amount of information equal to the logarithm in base 2 of this number[2], the amount of information we need to transmit the word is of the order of $\log_2 n = -\log_2 p$, where p is the probability of occurrence of our word, and it is a constant since we assumed to have a set of equiprobable words.

Suppose first that each symbol is transmitted independently of what has come before. This is equivalent to say that the information source is made of independent random variables Y_n. In this case the Shannon entropy[3] h_μ of the information source is the average of the information transmitted with each symbol. So, if $(p_i)_{i=0,\ldots,N-1}$ are the probabilities of occurrence of the single letters of $\mathcal{A}$, we have that

$$h_\mu = -\sum_{i=0}^{N-1} p_i \log p_i \tag{2.3}$$

where in this case the probability measure μ induced by the process $\mathbb{Y}$ on $(\Omega, \mathcal{C})$ is the product measure $\bigotimes_{n\in\mathbb{N}} P$ of $\mathcal{A}^{\mathbb{N}}$, being P the probability measure on $\mathcal{A}$ defined by $P(j) = p_j$ for each $j \in \mathcal{A}$. In this case the measure μ is the τ-invariant probability measure defined in equation (1.6).

In the general case, the information source is not made of independent random variables, that is the probability of transmission of a symbol depends

[1] This can be done by simply transmit the rule before the message. This adds only a constant to the amount of information.

[2] This is equal to the number of digits in the binary expansion of an integer.

[3] We use the same notation as for the Kolmogorov-Sinai entropy, since when there is possibility of confusion they actually coincide.

on what has been received before. For example, if we know that we are receiving an English text (so we are in a specific context), the probability of receiving a "u" will rise if the previous symbol is a "q". So, to actually receive a "u" in this case will add very few information about the message. To concern with this general case, one studies the probabilities of occurrences for blocks of increasing length. So let $\mathcal{L}_k$ be the set of all words of length k that can be produced by the information source. This is equivalent to consider all the cylinders $C(\omega, k)$, varying $\omega \in \Omega$. Each word $s \in \mathcal{L}_k$ has a probability of occurrence $p(s)$, that is equal to $\mu(C(\omega, k))$ if $\omega^k = s$.

In this framework

Definition 2.3. The *Shannon entropy* h_μ of an information source $\mathbb{Y}$, with probability measure μ induced on $(\Omega, \mathcal{C})$, is equal to

$$h_\mu = \lim_{k\to\infty} \frac{1}{k} \sum_{s\in\mathcal{L}_k} p(s) \log\left(\frac{1}{p(s)}\right) = \lim_{k\to\infty} \frac{1}{k} \int_\Omega \log\left(\frac{1}{\mu(C(\omega,k))}\right) d\mu(\omega) \tag{2.4}$$

where $p(s)$ is the probability of occurrence of a word in $\mathcal{L}_k$.

The Shannon entropy h_μ is then the average amount of information which the source conveys with every symbol it produces, and equivalently it is the average degree of uncertainty the receiver of the message has about what the next symbol will be, whatever is the previous message. In particular, we remark that the Shannon entropy measures a characteristic of a source, and not of a single sequence of the source. Hence it is a measure of information that depends on the context.

Finally we remark that for an information source on an alphabet $\mathcal{A}$ with N letters, the highest possible Shannon entropy of the source is given by $\log_2 N$, that corresponds to the Shannon entropy of an independent information source, where each letter of $\mathcal{A}$ is equiprobable.

2.2 Empirical entropy

We now introduce the notion of *empirical entropy*. It is deeply related to the notion of Shannon entropy, in fact the complexity defined according to equation (2.2) using the empirical entropy is equal to the Shannon entropy for μ-almost any sequence of a finite ergodic information source (see Theorem 2.6 below).

The main importance of the empirical entropy is that, contrary to the Shannon entropy, it does not depend on the context. So it can be defined for a single given word.

Let s be a finite word of length n on the alphabet $\mathcal{A}$, namely $s \in \mathcal{A}^n$. For any given word $w \in \mathcal{A}^l$, with $1 \leq l \leq n$, we define the number of occurrences of w as a subword of s. Namely if s_i^j denotes the $(j-i)$-long subword of s made of the symbols $(s_i, s_{i+1}, \ldots, s_{j-1})$, let

$$\delta(s_i^{i+l}, w) = \begin{cases} 1 & if\ s_i^{i+l} = w \\ 0 & otherwise \end{cases} \qquad 0 \le i \le n-l \ .$$

Hence the number of occurrences of w as a subword of s is given by

$$N(s,w) = \sum_{i=0}^{n-l} \delta(s_i^{i+l}, w)$$

and we define $P(s,w)$ to be the relative frequency of w in s, that is the number $N(s,w)$ of occurrences of w as a subword of s divided by the total number of subwords of s of length l. Hence

$$P(s,w) = \frac{1}{n-l+1} \; N(s,w) \ .$$

Definition 2.4. Given a finite word $s \in \mathcal{A}^n$, the *l-th empirical entropy* $\hat{H}_l(s)$ of s is defined by

$$\hat{H}_l(s) = -\sum_{w \in \mathcal{A}^l} P(s,w) \log P(s,w) \tag{2.5}$$

For a sequence $\omega \in \Omega$ the *l-th empirical entropy* $\hat{H}_l(\omega)$ of ω is defined by

$$\hat{H}_l(\omega) = \limsup_{n \to \infty} \hat{H}_l(\omega^n) \tag{2.6}$$

We now use the notion of l-th empirical entropy to define the *empirical complexity* for sequences of an information source, following Definition 2.2.

Definition 2.5. For a sequence $\omega \in \Omega$, the *empirical complexity* $\hat{H}(\omega)$ is defined by

$$\hat{H}(\omega) = \lim_{l \to \infty} \frac{1}{l} \; \hat{H}_l(\omega) \ . \tag{2.7}$$

The existence of the limit in equation (2.7) is proved in [ZL2], and it intuitively comes from the following argument.

Recalling the interpretation of a finite information source as the finite invariant dynamical system (Ω, τ, μ), the relative frequency of a word $w \in \mathcal{A}^l$ in $s \in \mathcal{A}^n$ can be seen as the frequency of visit of the orbit of the system (Ω, τ, μ) with first n symbols given by s in the cylinder $C(w,l)$. Hence applying the Birkhoff Ergodic Theorem 1.6 to the characteristic function of $C(w,l)$, we find that

$$\lim_{n \to \infty} P(\omega^n, w) = \mu\Big(C(w,l)\Big)$$

for μ-almost any sequence $\omega \in \Omega$. Hence the existence of the limit for the empirical complexity is equivalent to the existence of the limit for the Shannon entropy or for the Kolmogorov-Sinai entropy.

Applying the Birkhoff Ergodic Theorem 1.6 in the ergodic case, it follows the next result, first proved in [ZL2].

Theorem 2.6 (Ziv-Lempel [ZL2]). *Given a finite ergodic information source with realizations in the space (Ω, μ), it holds $\hat{H}(\omega) = h_\mu$ for μ-almost any $\omega \in \Omega$.*

When the information source is not ergodic, applying the *Ergodic Decomposition Theorem* (see for example [KH]), it is proved

Theorem 2.7 ([Be]). *For a finite information source with realizations in the space (Ω, μ), it holds*

$$\int_\Omega \hat{H}(\omega)\ d\mu(\omega) = h_\mu$$

In the following we will study also information sources with an infinite invariant measure μ on $(\Omega, \mathcal{C})$ such that the shift map τ is conservative. We will talk of *infinite conservative information sources*. In this case, the notion of Shannon entropy is not defined.

Theorem 2.8. *For an infinite ergodic conservative information source with realizations in the space (Ω, μ), for μ-almost all $\omega \in \Omega$ we have $\hat{H}(\omega) = 0$.*

Proof. It easily follows from Corollary 1.10, by the previous interpretation of the empirical entropy with the use of the Birkhoff Ergodic Theorem. □

2.3 Algorithmic Information Content

One of the most important information function is the *Algorithmic Information Content (AIC)*. It is independent on the context as the empirical entropy, but its definition is much closer to the intuitive notion of measure of information. This implies that the theorems below that show its equivalence with the Shannon entropy are much deeper results.

In order to define the AIC, it is necessary to define the notion of partial recursive function. We limit ourselves to give an intuitive idea which is very close to the formal definition. We can consider a partial recursive function as a computer C which takes a program p (namely a binary word) as an input, performs some computations and gives a word $s = C(p)$, written in a given alphabet $\mathcal{A}$, as an output.

Definition 2.9. The *Algorithmic Information Content (AIC)* of a word $s \in \mathcal{A}^*$ is the length of the shortest binary program $p \in \{0,1\}^*$ which gives s as its output using a partial recursive function C. Namely

$$AIC(s,C) = \min\left\{|p| \ / \ C(p) = s\right\} \tag{2.8}$$

where $|p|$ denotes the length of p.

To make the definition more efficient, the computer C used is required to be a *universal computing machine*. Roughly speaking, a computing machine is called *universal* if it can simulate any other machine. In particular every real computer is a universal computing machine, provided that we assume that it has virtually infinite memory. For precise definitions of the notions we introduced see for example [LV] or [Cha].

The following theorem specifies the dependence of the AIC on the computer C.

Theorem 2.10 (Kolmogorov [Kol]). *If C and C' are universal computing machines then*

$$\left|AIC(s,C) - AIC(s,C')\right| \leq K(C,C')$$

where $K(C,C')$ is a constant which depends only on C and C' but not on the word s.

This theorem implies that the information content AIC of s with respect to C depends only on s up to a fixed constant. This implies that its asymptotic behaviour does not depend on the choice of C. For this reason from now on we will write $AIC(s)$ instead of $AIC(s,C)$.

As we said before, the notion of Algorithmic Information Content is the formal statement that the information content of a word is the length of the shortest program from which we can reconstruct s. This shortest program is then a sort of encoding of the word s, since all the information necessary to reconstruct the word is contained in the program.

Unfortunately this coding procedure cannot be performed by any algorithm ([Cha]). This is a very deep statement and, in some sense, it is equivalent to the Turing halting problem. The intuitive idea of this relation is that given a word $s \in \mathcal{A}^*$, to find $AIC(s)$ we could try all programs p of binary length less than or equal to the binary length of s on a universal machine C, and then choose the shortest that gives s as output. Unfortunately this cannot be done, since the Turing halting problem tells us that we cannot know, using an algorithm, if an universal machine would stop or not on a given program. Then the Algorithmic Information Content is not computable by any algorithm.

This fact has very deep consequences for our discussion as we will see later. For the moment we can say that the AIC cannot be used as a reasonable physical quantity since it cannot be measured, however it is very useful in proving general theorems.

Again, using the information function defined using the AIC, it is possible to define the notion of complexity associated to AIC.

Definition 2.11. The *AIC-complexity* $K(\omega)$ of a sequence $\omega \in \Omega$ is given by

$$K(\omega) = \limsup_{n\to\infty} \frac{AIC(\omega^n)}{n} \tag{2.9}$$

where we recall ω^n denotes the finite subword of ω, made of its first n symbols.

As we have seen for the empirical complexity, the AIC-complexity is related to the Shannon entropy of a finite information source by the following theorems.

Theorem 2.12 (Brudno [Br]). *Given a finite ergodic information source with realizations in the space (Ω, μ), it holds $K(\omega) = h_\mu$ for μ-almost any $\omega \in \Omega$.*

Actually when dealing with ergodic sources the AIC-complexity of a sequence can be obtained by simply performing the limit in equation (2.9). Indeed it holds

Theorem 2.13 (White [W1],[W2]). *For a finite ergodic information source with realizations in the space (Ω, μ), for μ-almost any $\omega \in \Omega$ it holds*

$$\liminf_{n\to\infty} \frac{AIC(\omega^n)}{n} = h_\mu \tag{2.10}$$

Again it is possible to obtain a generalization for the non-ergodic case, using the *Ergodic Decomposition Theorem.*

Theorem 2.14 ([Be]). *For a finite information source with realizations in the space (Ω, μ), it holds*

$$\int_\Omega K(\omega)\ d\mu(\omega) = h_\mu$$

It remains open the case of information sources with infinite invariant measure μ on $(\Omega, \mathcal{C})$. In this case the Shannon entropy of the source is not defined, but we can prove the same result as for the empirical complexity.

Theorem 2.15. *For an infinite ergodic conservative information source with realizations in the space (Ω, μ), for μ-almost all $\omega \in \Omega$ it holds $K(\omega) =$ 0.*

Proof. The theorem follows from Lemmas 1.7 and 1.9 in [Br], and from Corollary 1.10. □

Finally, we remark that the highest possible AIC-complexity for a sequence of an information source on an alphabet $\mathcal{A}$ of N letters is given by $\log_2 N$. This can be obtained as a consequence of Brudno's Theorem 2.12 for ergodic information sources and from the remark at the end of Section 2.1. However, in the next subsection we show some examples of the behaviour of the AIC on some different kinds of words, without using Shannon entropy, but simply by arguments on the definition of the AIC.

2.3.1 Examples

We now give some examples of the behaviour of the AIC on finite words. We will study words produced by three different kinds of information sources: periodic, intermittent, chaotic. These three examples represent the classes of dynamical systems that we will study in the following chapters.

Periodic source. Let's assume that the stationary stochastic process $\mathbb{Y}$ is periodic of first period k. By this we mean that for all $n \in \mathbb{N}$ it holds $Y_n = Y_{n+k}$, and that $Y_{n+i} \neq Y_n$ for all $i = 1, \dots, k-1$. Then the induced measure on $(\Omega, \mathcal{C})$ is finite and the Shannon entropy h_μ of the source is null.

Let then $\omega \in \Omega$ be a sequence produced by the source $\mathbb{Y}$. We study now the behaviour of $AIC(\omega^n)$. By Definition 2.9 we have to find a binary program p which produces ω^n on a universal machine. First of all we notice that in order to produce ω^n the program p has to contain the information about the number n. Hence its binary length cannot be smaller than $\log_2 n$. This comes from the fact that for almost any number n (in a sense clarified below in the chaotic case) the shortest binary program to produce n is of length $\log_2 n$. We have then that

$$AIC(\omega^n) \geq \log_2 n \tag{2.11}$$

We now show that the asymptotic behaviour of $AIC(\omega^n)$ as n increases is $\log_2 n$, by describing a program $\bar{p}$ of length $\log_2 n + c$, for a positive constant c, which produces ω^n.

The program $\bar{p}$ has the following steps: (i) transmit the length n of the word to produce; (ii) transmit the periodic pattern $(\omega_0, \dots, \omega_{k-1})$ to be repeated; (iii) repeat the periodic pattern for n symbols.

This program obviously produces ω^n for all $n \in \mathbb{N}$, and its binary length is given by: (i) $\log_2 n$ bits for the first step as explained above; (ii) a constant (with respect to the length n of the word to produce) c number of bits to make the other two steps. Hence $|\bar{p}| = \log_2 n + c$. By the definition of the AIC as the length of the shortest program p that produces ω^n, we have that

$$AIC(\omega^n) \leq |\bar{p}| = \log_2 n + c \tag{2.12}$$

Hence we have proved that all the sequences ω of a periodic source have AIC-complexity $K(\omega) = 0$ (hence it is equal to the Shannon entropy of the source), and the asymptotic behaviour of $AIC(\omega^n)$ is $\log_2 n$.

Chaotic source. Let's assume now that the information source $\mathbb{Y}$ is chaotic, in the sense that the Shannon entropy h_μ is defined and positive. In this case Theorems 2.12 and 2.14 give results for the AIC-complexity of sequences of the source in the ergodic and non-ergodic cases, respectively. We now show by means of an example what is the behaviour of $AIC(\omega^n)$ for a chaotic information source using Definition 2.9.

Let $\mathbb{Y}$ be a *Bernoulli trial of parameter* $q \in [0,1]$. By this we mean a stochastic process $\mathbb{Y}$ on the alphabet $\mathcal{A} = \{0,1\}$, such that the random variables Y_n are independent and are equally distributed with the Bernoulli distribution: $Prob\{Y_n = 1\} = q$, and $Prob\{Y_n = 0\} = 1 - q$. The measure μ induced by such a source on $(\Omega, \mathcal{C})$ is the shift invariant probability measure described in equation (1.6).

For simplicity let's study the case $q = 1 - q = \frac{1}{2}$. The following argument applies without loss of generality to the general cases.

By the remark at the end of Section 2.1, we see that in this case for the Shannon entropy it holds $h_\mu = \log_2 2 = 1$. We now show by an heuristic argument that $K(\omega) = 1$ for μ-almost any $\omega \in \Omega$.

In this case the measure of the cylinders $C(\omega, k)$ depends only on k. Indeed $\mu(C(\omega,k)) = \frac{1}{2^k}$ for all ω and $k \in \mathbb{N}$. For any given $c < 1$ let's study the measure of the set of sequences with complexity c. Fixed any $n \geq 1$, there are exactly $(2^{n+1} - 1)$ words in $\mathcal{A}^*$ of length less than or equal to n. In the same way there are[4] $(2^{cn+1} - 1)$ binary programs of length less than or equal to cn. Hence the measure of the set of sequences with complexity c is asymptotic to $\frac{2^{cn}}{2^n}$, which converges to 0 as $n \to \infty$, for all $c < 1$.

Hence, as it was expected, for μ-almost any $\omega \in \Omega$ it holds $K(\omega) = 1$, and the asymptotic behaviour of the Algorithmic Information Content is $AIC(\omega^n) \sim n\ h_\mu = n$, being $h_\mu = 1$ the Shannon entropy.

We now apply this result to the problem of transmitting a number n. One possible binary program p that produces n is its binary expansion, and it is well known that the binary length of this expansion grows as $\log_2 n$. Since the binary expansion of a number n can be considered as a word produced by a Bernoulli trial, we have that for almost any number n, the length of the shortest binary program that produces n on a universal machine grows as $\log_2 n$.

[4] We neglect for simplicity the integer part of cn.

Intermittent source. The phenomenon of *intermittency* has attracted more and more attention in the last years both from the mathematical and the physical community. In the first case, it has given the background to produce examples of behaviours, for some statistical indicators, different from the behaviours observed in many classical systems. Moreover, it has been observed experimentally in the behaviour of turbulent fluids. We will study *intermittency* better in Section 4.3. For the moment we just describe an information source that we will use to analyze intermittent dynamical systems.

In this case we assume that the information source $\mathbb{Y}$ is on an infinite alphabet $\mathcal{A} = \mathbb{N}$. This is not a problem for the application of our techniques, since when we study the behaviour of $AIC(\omega^n)$ it is possible to consider ω^n as a word on a finite alphabet.

The particular information source we consider has the following properties: (i) Y_0 is drawn randomly with a distribution such that $Prob\{Y_0 = k\} \sim k^{-(1+\alpha)}$, where α is real number in the interval $(0, 1)$; (ii) for each $n \geq 1$, we have that if $Y_{n-1} \neq 0$ then $Y_n = Y_{n-1} - 1$, if instead $Y_{n-1} = 0$ then Y_n is drawn randomly with the same distribution law of Y_0. This stochastic process is stationary with respect to a measure μ on $(\Omega, \mathcal{C})$ that is not a probability measure. Hence applying Theorem 2.15 we have that for μ-almost all $\omega \in \Omega$ the AIC-complexity $K(\omega)$ is null. As we will see in Section 3.3, in this case we are interested in the asymptotic behaviour of $AIC(\omega^n)$ as n increases.

Let ω be a sequence produced by this source. We now study the behaviour of $AIC(\omega^n)$. Let $\bar{p}$ be the following program: (i) transmit ω_0; (ii) for all $k \geq 1$, transmit ω_k if and only if $\omega_{k-1} = 0$. This can be proved to be the shortest binary program that produces ω^n, since it is the program that transmits the word $(\omega_0, \omega_{k_1}, \dots, \omega_{k_r})$ (where ω_{k_j} are such that $\omega_{k_j - 1}$ are all and the only occurrences of the letter "0" in ω^n), and this word can be considered as produced by an independent information source on a finite alphabet. Hence for almost all these words the AIC grows linearly with time, with a slope equal to the Shannon entropy. This implies in particular that for μ-almost any $\omega \in \Omega$, it holds $AIC(\omega^n) \sim |\bar{p}(\omega)|$, where $\bar{p}(\omega)$ denotes the program we described before, applied to the sequence ω. We remark that the length of the programs $\bar{p}$ increases with time for two reasons. First because the word $(\omega_0, \dots, \omega_{k_r})$ to be transmitted grows. But also because the alphabet by which this word is made has more and more letters. In Section 4.3 we will study the behaviour of the binary length of $\bar{p}$.

Chapter 3

Information content and chaotic dynamical systems

In this chapter we introduce the applications of the method of the information content (described in Chapter 2) to the study of chaotic dynamical systems. In particular we present the asymptotic behaviour of the information contained in an orbit of a dynamical system as an indicator of chaoticity. The usefulness of this indicator is clear in the case of weakly chaotic systems (see Sections 1.3.2 and 3.3). In the next chapter we show the applications to some examples of weakly chaotic systems, that represent some of the most studied systems.

3.1 Dynamical systems as information sources

The intermediate step, necessary to apply the method of information content to dynamical systems, is the representation of a system as an information source. This is done by using the *symbolic representation* of a system (see equation (1.7)).

Let $(X, \mathcal{B}, T)$ be a dynamical system, and let μ be a T-invariant measure on $(X, \mathcal{B})$. Let $Z = \{I_0, \dots, I_{N-1}\}$ be a finite measurable partition of X, and let $\mathcal{A} = \{0, \dots, N-1\}$ be the associated finite alphabet. Then the symbolic representation of the system (X, T, μ) is given by the function $\varphi_Z : X \to \Omega = \mathcal{A}^{\mathbb{N}}$ defined by

$$\varphi_Z(x) = \omega \iff T^i(x) \in I_{\omega_i} \quad \forall\, i \in \mathbb{N}$$

The image $\Omega' = \varphi_Z(X)$ is a closed subset of Ω which is invariant under the shift map τ. The use of the symbolic representation is meaningful in particular if we assume it to be injective on a subset of X of full Lebesgue measure. Hence, from now on, we assume that the function φ_Z is a homeomorphism, hence a conjugacy between the dynamical systems (X, T) and

(Ω', τ). Further if μ is an invariant measure on (X, T) we denote by ν the induced invariant measure on (Ω', τ).

At this point it is possible to apply the different measures of information introduced in the previous chapter to the orbits of the system. Given a finite measurable partition Z of X and the associated alphabet $\mathcal{A}$, and given an information function $I : \mathcal{A}^* \to \mathbb{N}$ we have

Definition 3.1. The *Z-information content* $I(x, n, Z)$ of the word $\varphi_Z(x)^n$ in $\mathcal{A}^n$ associated to a point $x \in X$, according to the information function I, is $I(x, n, Z) = I((\varphi_Z(x))^n)$. Analogously, the *$Z$-complexity* $K(x, Z)$ of a point $x \in X$, using I as information function, is given by

$$K(x, Z) = \limsup_{n\to\infty} \frac{I(x, n, Z)}{n} \tag{3.1}$$

Applying the empirical entropy or the algorithmic information content as information functions, we can define the Z-empirical complexity and the Z-AIC-complexity of a point $x \in X$, denoted respectively by $\hat{H}(x, Z)$ and $K(x, Z)$. For simplicity of notation, from now on when we talk of Z-complexity $K(x, Z)$ we mean the Z-AIC-complexity, when there is no possible confusion.

We can now apply the theorems about the equivalence between the various notions of complexity and the Shannon entropy of a source (see Sections 2.2 and 2.3), using a dynamical system as source, and using the Kolmogorov-Sinai entropy of the dynamical system relatively to a partition.

Theorem 3.2 ([Br], [Be]). *Let (X, T, μ) be a finite invariant dynamical system. Given a finite measurable partition Z of X, it holds*

$$\int_X K(x, Z)\ d\mu(x) = \int_X \hat{H}(x, Z)\ d\mu(x) = h_\mu(T, Z)$$

where $h_\mu(T, Z)$ denotes the Kolmogorov-Sinai entropy relative to Z of the dynamical system (see equation (1.25)). If the dynamical system (X, T, μ) is ergodic then for μ-almost all $x \in X$ we have

$$K(x, Z) = \hat{H}(x, Z) = h_\mu(T, Z)$$

Theorem 3.3. *Let (X, T, μ) be an infinite ergodic conservative invariant dynamical system. Given a finite measurable partition Z of X, for μ-almost all $x \in X$ it holds $K(x, Z) = \hat{H}(x, Z) = 0$.*

Proof. The theorem follows from the symbolic representation of a system, and from Theorems 2.8 and 2.15. □

Theorem 3.2 shows a perfect equivalence between the different indicators of chaos we introduced. As already remarked, the important property of the indicators of information content we introduced, as the empirical entropy and the AIC, is their independence on the context of the word. This implies that we can use these indicators to study single sequences. Hence it would be important to have a definition of complexity for a point not dependent on the partition Z. This problem is easily solved if the partition Z considered is *generating* (see Definition 1.28), since in this case the Kolmogorov-Sinai Theorem 1.29 tells us that the KS entropy $h_\mu(T,Z)$ relative to Z is equal to the KS entropy $h_\mu(T)$ of the system. Recall Theorem 1.30 for the existence of a finite generating partition. Unfortunately in real cases we are not always able to find a generating partition.

In general, analogously with the definition of the KS entropy from the KS entropy relative to a partition, one could think of computing the supremum of the Z-complexity $K(x,Z)$ of a point over all the finite measurable partitions Z. Unfortunately in [Br] it is proved that this approach fails, since for every non periodic point this supremum is infinite.

In [Br] the problem is solved by changing the definition of complexity of a point. Instead of using partitions of the phase space X, one could use finite open coverings of X. This approach has the drawback that some points $x \in X$ have more than one symbolic orbit in $\Omega = \mathcal{A}^{\mathbb{N}}$ (being $\mathcal{A}$ the finite alphabet associated to an open covering). Let $\mathcal{U}$ be a finite open covering of X, and denote by $\varphi_{\mathcal{U}}$ the symbolic function that to each point $x \in X$ associates the set of its symbolic orbits. Then the *$\mathcal{U}$-complexity* $K(x,\mathcal{U})$ of x is defined by

$$K(x,\mathcal{U}) = \limsup_{n\to\infty} \frac{1}{n} \min\left\{ AIC(\omega^n) \ / \ \omega \in \varphi_{\mathcal{U}}(x) \right\} \tag{3.2}$$

A different approach to solve the problem is introduced in [Ga1], [Ga2]. Instead of making the supremum of the Z-complexity of a point over all partitions, one can consider only a particular class of partitions, called *computable structures*. For example in $\mathbb{R}^n$ one can consider only the partitions generated by intersections of half spaces with rational coordinates.

We are not entering into the details of this problem for two reasons. The first is that in almost all the applications of this method we consider in this work, we are actually interested only in considering the KS entropy of a system relative to a finite measurable partition, and also the numerical experiments (see Chapter 6) are done with respect to a partition. The second reason is that in general we are interested in estimating the KS entropy of a system using the complexity, and not in just evaluating the complexity of a sequence by itself. Hence for our aims if we restrict ourselves to a countable family of finite measurable partitions Z^j with the property that $diam(Z^j) \to 0$ as $j \to \infty$, it holds

Proposition 3.4 ([Be]). *Let (X,T,μ) be a finite invariant dynamical system and $\{Z^j\}_{j\in\mathbb{N}}$ a family of finite measurable partitions Z^j with the property* *that* $diam(Z^j) \to 0$ *as* $j \to \infty$*, then*

$$\limsup_{j\to\infty} K(x, Z^j) = h_\mu(T)$$

for μ*-almost any* $x \in X$*.*

3.2 Positive entropy case

From Definition 3.1, it is clear that the Z-complexity of a point gives information on the asymptotic behaviour of the information content of an orbit of a dynamical system. Theorem 3.2 states that the Z-complexity is equal to the KS entropy relative to Z. Hence we have that from the KS entropy relative to Z we get information on $K(x, Z)$ and viceversa. This equivalence is strong in the case of positive KS entropy. In fact, as already remarked for the applications of the Equipartition Property of sets of the iterated partitions Z_n (Corollary 1.33), the case of positive entropy is the only one in which the results of Theorem 3.2 are significant.

Let (X, T, μ) be a finite ergodic invariant dynamical systems, such that the KS entropy $h_\mu(T, Z)$ relative to a finite measurable partition Z is positive. We will call such systems *fully chaotic*.

In this case, as a corollary to Theorem 3.2, we have that for μ-almost any $x \in X$ it holds

$$AIC(x, n, Z) \sim h_\mu(T, Z)\ n \tag{3.3}$$

As repeatedly remarked, the importance of the AIC of a word is that it does not depend on the context in which the word is. That is we can use this method without knowing the invariant measure μ. The importance of the result of equation (3.3) lays in the fact that estimating, for a randomly chosen initial condition x, the value of $AIC(x, n, Z)$, if we find a straight line we can say that the KS entropy of the dynamical system relative to Z is given by the angular coefficient of the line.

There are two problems when we really want to apply this method to a dynamical system. The first is what means to choose an initial condition randomly if we don't know the invariant probability measure μ. When making numerical experiments there are many ways one can choose a probability law to make the random choice of the initial condition. But the most natural is to use the Lebesgue distribution. Hence we are sure that this method actually works either when the measure μ is absolutely continuous with respect to the Lebesgue measure, or when we can prove (independently on μ) that equation (3.3) holds for almost any x in a subset of X with respect to the Lebesgue measure. The latter situation is equivalent to μ being a physical measure (see Definition 1.35).

The second problem is that the *AIC* is not a computable function (see Section 2.3), that is it is impossible to have an algorithm able to compute the *AIC* of all words. This problem is solved by looking for *compression algorithms* that can approximate the behaviour of the empirical entropy of a sequence. In Section 6.1 we will study this problem.

Much more difficult is the numerical approach to systems that are not ergodic. Indeed in this case, Theorem 3.2 implies that

$$\int_X AIC(x,n,Z)\ d\mu(x) \sim h_\mu(T,Z)\ n \tag{3.4}$$

Hence to know the KS entropy $h_\mu(T,Z)$ using the *AIC* of the orbits, it is necessary to have information on the invariant probability measure μ.

3.3 Weakly chaotic and infinite dynamical systems

So far we have seen the application of the information methods to finite invariant dynamical systems with positive KS entropy. In these cases, studying the asymptotic behaviour of the *AIC* of the symbolic orbits generated by an ergodic system, we expect to find for almost any initial condition straight lines with angular coefficient given by the KS entropy (see equation (3.3)).

When studying the symbolic orbits generated by a weakly chaotic system (a system with all its physical measures with null KS entropy) or by an infinite invariant system, the situation is much less clear than in the chaotic case. Indeed in these cases, by Theorems 3.2 and 3.3, for an ergodic system it holds $K(x,Z)=0$ for almost all x and for all finite partitions Z. We restrict our attention to infinite systems with invariant measure μ *equivalent* to the Lebesgue measure $\mu \sim m_n$, in the sense that $\mu << m_n$ and $m_n << \mu$, and to finite systems with a physical measure. We remark that for an ergodic conservative invariant dynamical system (X,T,μ), there is not an invariant measure μ' equivalent to μ. So, in particular, ergodic infinite systems with invariant measure $\mu \sim m_n$, do not have a physical measure equivalent to m_n. Viceversa, ergodic finite systems with physical measure $\mu \sim m_n$, do not have an infinite invariant measure equivalent to m_n.

In the case of a finite ergodic invariant system (X,T,μ) with null KS entropy, it is suggested in [Gr1] to look at the rate of decrease of the sequence $\frac{1}{n}H_\mu(Z_n)$ (see Section 1.3.1), for a finite partition Z, as a measure of complexity of the system. The same approach has been studied in [Bl]. We use the same approach for the AIC-complexity. Given an ergodic conservative invariant dynamical system (X,T,μ) with null AIC-complexity for μ-almost all $x \in X$, we look at the rate of increase of the sequence $AIC(x,n,Z)$ (see Definition 3.1), for a finite partition Z of X.

Recalling the results of Theorem 1.7 and Corollary 1.10, this approach can be compared to the search for a speed of convergence in the ergodic theorems. As stated by Theorem 1.14 for finite ergodic systems and by Theorem 1.16 in the infinite ergodic case, this search is hopeless for the speed of convergence of ergodic means of measurable functions. This suggests a negative result also in our case, that is we expect that there is not a sequence $a(n)$ of positive numbers, satisfying $a(n) \to \infty$ and $\frac{a(n)}{n} \to 0$ as n increases, such that

$$\lim_{n\to\infty} \frac{AIC(x,n,Z)}{a(n)}$$

exists finite and positive, and it is a constant for μ-almost all x of an ergodic conservative dynamical system. In Section 4.3, we will study an example of an infinite ergodic conservative dynamical system with this property.

To overcome this difficulty, we introduce the notions of *chaos index* (see Benci et al.[Be]) and *local chaos index*.

We first start giving the definitions for information sources. As explained in Section 3.1, this is equivalent to fix a finite partition for a dynamical system. Let $\mathbb{Y}$ be a conservative information source with realizations in the space (Ω, μ). We recall that it is equivalent to consider the conservative invariant symbolic dynamical system (Ω, τ, μ), where τ is the shift map defined in equation (1.5). If μ is an infinite measure, we consider a probability measure ν equivalent to μ. The measure ν is called a *reference measure*. If μ is a probability measure, it will be $\nu = \mu$.

Definition 3.5. The *upper chaos index* $\overline{q}(\Omega, \nu)$ is given by

$$\overline{q}(\Omega,\nu) = \inf\left\{ q \in (0,1) \ / \ \limsup_{n\to\infty} \int_\Omega \frac{AIC(\omega^n)}{n^q}\, d\nu(\omega) = 0 \right\} =$$

$$= \sup\left\{ q \in (0,1) \ / \ \limsup_{n\to\infty} \int_\Omega \frac{AIC(\omega^n)}{n^q}\, d\nu(\omega) = \infty \right\}$$

In the same way the *lower chaos index* $\underline{q}(\Omega, \nu)$ is defined using the inferior limit instead of the superior limit.

Analogously it is possible to define the *upper* and *lower local chaos indexes.*

Definition 3.6. The *upper local chaos index* $\overline{q}(\omega)$ is defined as

$$\overline{q}(\omega) = \inf\left\{ q \in (0,1) \ / \ \limsup_{n\to\infty} \frac{AIC(\omega^n)}{n^q} = 0 \right\} =$$

$$= \sup\left\{ q \in (0,1) \ / \ \limsup_{n\to\infty} \frac{AIC(\omega^n)}{n^q} = \infty \right\}$$

In the same way the *lower local chaos index* $\underline{q}(\omega)$ is defined using the inferior limit instead of the superior limit.

Proposition 3.7. *The function $\overline{q}(\cdot)$ defined on $(\Omega, \mathcal{C})$, with values on the interval $[0,1]$ with the usual Borel σ-algebra, is a measurable function.*

Proof. We show that the sets $Q_\gamma = \{\omega \ / \ \overline{q}(\omega) < \gamma\}$ are measurable for all $\gamma \in (0,1)$. Given a fixed γ, we can write

$$Q_\gamma = \left\{\omega \ / \ \exists \ \bar{k}(\omega) \in \mathbb{N} \ : \ \forall \ k > \bar{k}(\omega) \ \ \limsup_{n\to\infty} \frac{AIC(\omega^n)}{n^{\gamma - \frac{1}{k}}} = 0\right\}$$

If, moreover we define

$$A_{\bar{k},\gamma} = \left\{\omega \ / \ \forall \ k > \bar{k} \ \ \limsup_{n\to\infty} \frac{AIC(\omega^n)}{n^{\gamma - \frac{1}{k}}} = 0\right\}$$

then $Q_\gamma = \cup_{\bar{k}\in\mathbb{N}} A_{\bar{k},\gamma}$. So we only need to show that the sets $A_{\bar{k},\gamma}$ are measurable. But we have that $A_{\bar{k},\gamma} = \cap_{k>\bar{k}} C_k$ where

$$C_k = \left\{\omega \ / \ \limsup_{n\to\infty} \frac{AIC(\omega^n)}{n^{\gamma - \frac{1}{k}}} = 0\right\}$$

The thesis is obtained by showing that

$$C_k = \bigcap_{h\in\mathbb{N}} \ \bigcup_{N\in\mathbb{N}} \ \bigcap_{n>N} \ \left\{\omega \ / \ AIC(\omega^n) < n^{\gamma - \frac{1}{k}} \ \frac{1}{h}\right\}$$

that shows the measurability of the sets C_k and hence of the sets $A_{\bar{k},\gamma}$. □

Analogously it is possible to prove that the function $\underline{q} : \Omega \to [0,1]$ is measurable. We now show that the local chaos indexes are invariant with respect to the shift map τ on Ω.

Proposition 3.8. *Let $(\Omega, \mathcal{C}, \tau)$ be a symbolic dynamical system, then for all $\omega \in \Omega$ it holds $\overline{q}(\omega) = \overline{q}(\tau\omega)$.*

Proof. From the definition of Algorithmic Information Content (see Definition 2.9) we have that

$$AIC(\omega^n) \leq AIC((\tau\omega)^n) + c(\mathcal{A}) + const. \tag{3.5}$$

since, given the word $(\tau\omega)^n$, we reconstruct the word ω^n by adding information on the symbol ω_0. This information is contained in the constant $c(\mathcal{A})$, which can be dependent only on the alphabet $\mathcal{A}$ from which the sequences $\omega \in \Omega$ are built.

It the same way, it is easy to prove that

$$AIC((\tau\omega)^n) \leq AIC(\omega^{n+1}) + const. \leq AIC(\omega^n) + c(\mathcal{A}) + const. \tag{3.6}$$

where the constants contain the information necessary to reconstruct the words.

Putting together equations (3.5) and (3.6), we have that for all $q \in (0,1)$ it holds

$$\frac{AIC(\omega^n)}{n^q} \leq \frac{AIC((\tau\omega)^n) + c(\mathcal{A}) + const.}{n^q} \leq \frac{AIC(\omega^n) + 2c(\mathcal{A}) + const.}{n^q} \tag{3.7}$$

and the proposition is proved. □

Theorem 3.9. *Let $\mathbb{Y}$ be an ergodic conservative information source with realizations in the space (Ω, μ), then for μ-almost any $\omega \in \Omega$ the functions $\overline{q}(\omega)$ and $\underline{q}(\omega)$ are constant.*

Proof. It is an easy consequence of Propositions 3.7, 3.8 and Proposition 1.9. □

Theorem 3.10. *For an ergodic conservative invariant symbolic dynamical system (Ω, τ, μ) with reference measure $\nu \sim \mu$, if for $q > \overline{q}$ the functions $\left(\frac{AIC(\omega^n)}{n^q}\right)_n$ satisfy the hypothesis of Fatou's Lemma from above, it holds*

$$\underline{q} \leq \underline{q}(\Omega, \nu) \leq \overline{q}(\Omega, \nu) \leq \overline{q}$$

where $\overline{q}$ and $\underline{q}$ are the constant local indexes. If $\underline{q} = \overline{q}$ then $\underline{q}(\Omega, \nu) = \overline{q}(\Omega, \nu) = \overline{q} = \underline{q}$.

Proof. The theorem easily follows from an application of Fatou's Lemma. Indeed we have

$$\int_X \liminf_{n\to\infty} \frac{AIC(\omega^n)}{n^q}\, d\nu(\omega) \leq \liminf_{n\to\infty} \int_X \frac{AIC(\omega^n)}{n^q}\, d\nu(\omega) \leq$$

$$\leq \limsup_{n\to\infty} \int_X \frac{AIC(\omega^n)}{n^q}\, d\nu(\omega) \leq \int_X \limsup_{n\to\infty} \frac{AIC(\omega^n)}{n^q}\, d\nu(\omega) \qquad \square$$

This theorem shows in particular that when $\underline{q} < \overline{q}$ then it becomes interesting to make a comparison between the different indexes. A particular interesting case is when $\underline{q} = 0$ and $\overline{q} = 1$. In this case the local indexes give no information on the system, and we have to compute the global chaos indexes.

Using the symbolic representation of a dynamical system (X, T, μ) for a given finite partition Z of X, with reference measure ν, we can define analogously the *global* and *local chaos indexes* $\overline{q}(X, \nu, Z), \underline{q}(X, \nu, Z), \overline{q}(x, Z)$ and $\underline{q}(x, Z)$, which are now dependent on the chosen partition Z. We recall our assumption that if μ is infinite then $\mu \sim m_n$. This implies that the reference measure in the infinite case can be chosen to be the Lebesgue measure m_n. All the previous results can now be easily extended to these indexes.

We remark that the choice of the reference measure m_n in the infinite case is motivated by the notion of physical measure. Indeed in this case, if μ is infinite, ergodic and equivalent to m_n, there is no physical measure equivalent to m_n. Hence to look for statistical properties of the physical measures, one can look for the asymptotic behaviour of statistical properties of m_n. For the same reason, we are interested in studying finite invariant systems such that the probability measure is a physical measure.

3.4 Relations with other indicators of chaos

We have so far built the structure of our analysis of dynamical systems. Our analysis is based on the asymptotic behaviour of the Algorithmic Information Content for the symbolic orbits generated by the systems. We have seen that in the classical cases of periodic systems and fully chaotic systems, our results coincide with the classical notion of KS entropy. Instead, in the weakly chaotic cases and for the infinite systems, we have from our analysis more information on the system. In Chapter 4, we will analyze some important systems in the second class, showing how they can be classified according to the chaos indexes introduced in the previous section.

In this section we study the relations between the chaos indexes and other statistical indicators of chaos, such as the sensitivity to initial conditions, the Poincaré recurrence time and the decay of correlations.

3.4.1 Chaos indexes and sensitivity to initial conditions

The property of sensitivity to initial conditions, introduced in Section 1.3, is related to the speed of separation of two orbits with nearby initial conditions. In [Ga1] and [Ga3], it has been formally studied the relation between the asymptotic behaviour of the AIC and the rate of decrease of the separation sets $B(x, \epsilon, n)$ (see Definition 1.25). This formal approach requires definitions and techniques from information theory (such as *constructivity* and *computable structures*), that are beyond the aims of this thesis. We now approach the argument with simplified hypotheses.

Let $(X, \mathcal{B}, T)$ be a dynamical system on a compact metric space X, such that the map $T : X \to X$ is piecewise continuous and it is "known" (in the sense that, given a point x, the value $T(x)$ can be calculated on a computer[1]). For simplicity we assume that $X = [0, 1]$. Let Z be a uniform finite measurable partition of $[0, 1]$ with diameter d, in the sense that Z is obtained by dividing the interval $[0, 1]$ in equal intervals of diameter d. We assume that a computer can decide to what set of Z a point belongs to. The aim is now to study the information content of the word $(\varphi_Z(x))^n$ for a

[1]This assumption recovers the notion of *constructivity*.

given initial condition x (recall the definition of symbolic representation of a dynamical system).

In this approach instead of studying the separation sets of Definition 1.25, we study the sets $B(x, n, Z)$, defined as the connected component of $Z_{n+1}(x)$ containing x, where we recall that $Z_{n+1}(x)$ denotes the set of the iterated partition Z_{n+1} containing x. Since we assumed that the symbolic interpretation φ_Z is injective on a set of m-full measure, we have that the sets $B(x, n, Z)$ shrinks to x as $n \to \infty$ for m-almost all $x \in [0, 1]$.

For a given initial condition $x \in [0, 1]$, let $\{z_j\}_{j \in \mathbb{N}}$ be a sequence of rational approximants of x. For example z_j can be the rational number obtained by considering the first j digits in the continued fractions expansion of x, or the first j digits in its binary expansion. Fix n and consider the word $(\varphi_Z(x))^n$. By definition of $B(x, n, Z)$, we have that for a given $z_{j(n)}$ the word $(\varphi_Z(z_{j(n)}))^n$ is equal to $(\varphi_Z(x))^n$ if and only[2] if $z_{j(n)} \in B(x, n, Z)$.

Let's consider the case of rational approximants z_j built by the binary expansion of x, and let ξ^j denote the word given by the binary digits of z_j. Hence $\xi^j \in \{0, 1\}^j$. For the fixed n, let $k(x, n, Z)$ be given by

$$k(x, n, Z) = \inf \left\{ k \in \mathbb{N} \ / \ B_x \left(\frac{1}{2^k} \right) \subset B(x, n, Z) \right\} \tag{3.8}$$

where $B_x(\delta)$ denotes a ball of radius δ around x.

Then since $z_{k(x,n,Z)} \in B(x, n, Z)$, the information content of $(\varphi_Z(x))^n$ is equal to that of $(\varphi_Z(z_{k(x,n,Z)}))^n$, and this word can be constructed by knowing $\xi^{k(x,n,Z)}$ and n. Indeed we just need to tell a computer what is the initial condition, and how many times it has to iterate the map T with respect to the partition Z. Then

$$AIC((\varphi_Z(x))^n) \leq \log n + AIC(\xi^{k(x,n,Z)}) + const \tag{3.9}$$

where $\log n$ is the number of bits necessary to transmit n (see Section 2.3.1), and the constant contains the information to perform the different operations.

Let's assume now to know the word $(\varphi_Z(x))^n$, we can then find the point x with an approximation increasing with n. In particular we look for the rational approximants of x given by its increasing binary expansions. If we define

$$h(x, n, Z) = \sup \left\{ h \in \mathbb{N} \ / \ B_x \left(\frac{1}{2^h} \right) \supset B(x, n, Z) \right\} \tag{3.10}$$

if we know the word $(\varphi_Z(x))^n$ then we know $\xi^{h(x,n,Z)}$. This implies that

$$AIC(\xi^{h(x,n,Z)}) \leq AIC((\varphi_Z(x))^n) \tag{3.11}$$

[2]We assume to look for rational approximants of x in the same connected component of $Z_{n+1}(x)$.

At this point we remark that, for m-almost all $x \in [0,1]$, we have that for all $j \in \mathbb{N}$ it holds $AIC(\xi^j) \sim j$, since the words ξ^j can be considered as produced by a Bernoulli trial of parameter $\frac{1}{2}$ (see Section 2.3.1). Hence, putting together equations (3.9) and (3.11), we find for all $q \in (0,1)$

$$\frac{h(x,n,Z)}{n^q} \leq \frac{AIC((\varphi_Z(x))^n)}{n^q} \leq \frac{\log n + k(x,n,Z) + const}{n^q} \tag{3.12}$$

Let now $\overline{q}(k(x,n,Z))$ be defined by

$$\overline{q}(k(x,n,Z)) = \inf \left\{ q \in (0,1) \ / \ \limsup_{n\to\infty} \frac{k(x,n,Z)}{n^q} = 0 \right\}$$

and define analogously $\underline{q}(k(x,n,Z))$, $\overline{q}(h(x,n,Z))$ and $\underline{q}(h(x,n,Z))$. From the definition of $h(x,n,Z)$ and $k(x,n,Z)$, we have that $\overline{q}(k(x,n,Z)) = \overline{q}(h(x,n,Z))$ and $\underline{q}(k(x,n,Z)) = \underline{q}(h(x,n,Z))$. Recalling the definition of local chaos index for a point x, we have proved

Theorem 3.11. *For a dynamical system $([0,1],\mathcal{B},T)$ with T piecewise continuous and "known", for a finite uniform partition Z of $[0,1]$, for m-almost all $x \in [0,1]$ it holds*

$$\overline{q}(x,Z) = \overline{q}(k(x,n,Z)) \qquad \underline{q}(x,Z) = \underline{q}(k(x,n,Z))$$

Hence the local chaos indexes of the point x relative to the partition Z depend on the rate with which the set $B(x,n,Z)$ shrinks.

Example 3.12. As an example, consider on the interval $[0,1]$ the map $T : x \to 2x(\text{mod } 1)$. It is the best known example of chaotic dynamical system, and it corresponds to the Bernoulli trial of parameter $\frac{1}{2}$. Applying the previous definitions we find that for all the finite uniform partitions Z and for m-almost all x it holds $\overline{q}(x,Z) = \underline{q}(x,Z) = 1$.

3.4.2 Chaos indexes and Poincaré recurrence time

In the local analysis of a chaotic system, it has been studied the behaviour of the Poincaré recurrence time (see Definition 1.5) for cylinders of a partition. This section is based on [BGI].

Given a dynamical system (X,T) and a finite measurable partition Z of X, let Z_n denote the iterated partition

$$Z_n = \bigvee_{i=0}^{n-1} T^{-i}Z$$

and $Z_n(x)$ the set of Z_n containing x. We now study the asymptotic behaviour of the Poincaré recurrence time $\tau(Z_n(x))$ as n increases. An elementary argument in [STV] shows that the set $Z_n(x)$ is completely determined by its first $\tau(Z_n(x))$ symbols. Indeed let $Z_n(x)$ be specified by

$(\omega_0, \ldots, \omega_{n-1})$. If j is the integer defined by $j\tau < n \leq (j+1)\tau$, where $\tau \equiv \tau(Z_n(x))$, then it holds

$$\omega_{i\tau}, \ldots, \omega_{(i+1)\tau-1} = \omega_0, \ldots, \omega_{\tau-1} \qquad \text{for } \; i = 0, \ldots, j-1$$

since, by definition of Poincaré recurrence time, there exists a point $y \in Z_n(x)$ such that $T^\tau(y) \in Z_n(x)$. This suggests that the information content of the word $(\omega_0, \ldots, \omega_{n-1})$ is contained in the word $(\omega_0, \ldots, \omega_{\tau-1})$. So it holds

$$AIC(x, n, Z) \leq AIC(x, \tau, Z) + \log n \tag{3.13}$$

since to transmit ω^n we have to transmit ω^τ, and the integer n needed to know how many times ω^τ has to be repeated.

As usual, to study the asymptotic behaviour of the Poincaré recurrence time, the first thing is to compare it with the straight lines.

Definition 3.13. Given a finite measurable partition Z of the space X, for each point $x \in X$ we define the *upper* and *lower recurrence rates* relative to Z respectively as

$$\overline{R}_Z(x) = \limsup_{n\to\infty} \frac{\tau(Z_n(x))}{n} \qquad \underline{R}_Z(x) = \liminf_{n\to\infty} \frac{\tau(Z_n(x))}{n}$$

Theorem 3.14 ([STV]). *Given a finite ergodic invariant dynamical system (X, T, μ), if for a finite measurable partition Z it holds $h_\mu(T, Z) > 0$, then $\underline{R}_Z(x) \geq 1$ for μ-almost all $x \in X$.*

We now study the relation between the local chaos indexes and the Poincaré recurrence time in a more general context, to obtain results for systems with null AIC-complexity.

Theorem 3.15 ([BGI]). *Let (X, T, μ) be an ergodic conservative invariant dynamical system and Z a finite partition. We assume that if μ is finite then the AIC-complexity (equal to the KS entropy relative to Z) is finite for μ-almost all $x \in X$, and if μ is an infinite measure then the AIC-complexity is null μ-almost anywhere. Then for μ-almost all $x \in X$ we have*

$$\liminf_{n\to\infty} \frac{AIC(x, n, Z)}{n^\beta} = 0$$

whenever $\beta > \alpha(x, Z)$, where

$$\alpha(x, Z) := \inf \left\{ \alpha \;/\; \liminf_{n\to\infty} \frac{\tau(Z_n(x))}{n^\alpha} = 0 \right\} .$$

Hence, $\underline{q}(x, Z) \leq \alpha(x, Z)$.

Proof. Using equation (3.13) and Theorems 3.2 and 3.3, we have, for $\beta > \alpha(x, Z)$,

$$0 \leq \liminf_{n\to\infty} \frac{AIC(x,n,Z)}{n^\beta} \leq \liminf_{n\to\infty} \left(\frac{AIC(x,\tau(Z_n(x)),Z)}{n^\beta} + \frac{\log n}{n^\beta} \right) =$$

$$= \lim_{n\to\infty} \frac{AIC(x,\tau(Z_n(x)),Z)}{\tau(Z_n(x))} \cdot \liminf_{n\to\infty} \frac{\tau(Z_n(x))}{n^\beta} = K(x,Z) \cdot 0 = 0$$

where in the case of a finite ergodic system we used Theorem 2.13. □

From the above we may deduce some consequences.

Corollary 3.16 ([BGI]). *Let (X, T, μ) be a finite ergodic invariant dynamical system, and suppose that for μ-almost all $x \in X$ we have $\alpha(x, Z) < 1$. Then $h_\mu(T, \mathcal{Z}) = 0$. If moreover $\alpha(x, Z) = 0$ then $\underline{q}(x, Z) = 0$.*

Another relation can be established for the upper recurrence rate.

Theorem 3.17 ([BGI]). *Let (X, T, μ) be an ergodic conservative invariant dynamical system and Z a finite partition. If $\overline{R}_Z(x) < 1$ then $\overline{q}(x, Z) = 0$. More specifically we have*

$$\limsup_{n\to\infty} \frac{AIC(x,n,Z)}{\log^2 n} \leq -\frac{1}{\log \overline{R}_Z(x)}$$

The proof of this theorem consists in a straightforward application of the following lemma with the identifications $d = \frac{1}{\overline{R}_Z(x)}$ and $a_n = AIC(x, n, Z)$.

Lemma 3.18 ([BGI]). *Let a_n be a real positive sequence such that there exist a real $d > 1$ and $\overline{n} \in \mathbb{N}$ such that: $\forall\, n > \overline{n}$ there is a real d_n satisfying $d_n > d$, $\frac{n}{d_n} \in \mathbb{N}$ and $a_n \leq a_{\frac{n}{d_n}} + \log n + C$. Then*

$$\limsup_{n\to\infty} \frac{a_n}{\log^2 n} \leq \frac{1}{\log d}$$

Proof. Set $A = \max\{a_n \;/\; n \leq \overline{n}\}$ and $f : \mathbb{N} \to \mathbb{N}$ be the map given by $f(n) = \frac{n}{d_n}$. Note that $f(n) < n$. For $n > \overline{n}$ we have

$$a_n \leq a_{f(n)} + \log n + C \leq a_{f^2(n)} + \log n + \log f(n) + 2C \leq$$

$$\leq \cdots \leq a_{f^k(n)} + k(C + \log n)$$

with $k = \min\{i \in \mathbb{N} \;/\; f^i(n) \leq \overline{n}\}$. One readily realizes that $k \leq \frac{\log(n/\overline{n})}{\log d}$. We then have $a_n \leq A + \frac{\log(n/\overline{n})}{\log d}\,(C + \log n)$. □

In Chapter 4, we will see an application of Theorem 3.17 to a specific example. This theorem will help in draw conclusions for the statistical behaviour of the system.

3.4.3 Chaos indexes and decay of correlations

In Section 1.2.1, we have seen that the statistical analysis of chaotic systems is based also on the study of the rate of decay of the correlation functions, as defined in Definition 1.19.

In Theorem 1.21 we gave an example of systems with exponential decay of correlations. However, in the last years, there have been many studies on systems with slower rates of decay. The main literature on this issue includes [Yo], [LSV], [Sa], [Go], [Is1], [ALP], [BLS]. In particular, an interesting argument is the extension of the study of correlations to infinite systems ([Is1]).

As already remarked in Section 1.2.1, the rate of decay of correlations is related to the loss of memory for the orbits of the system. This loss of memory is also reflected in the definition of KS entropy as in equation (1.26), that emphasizes the fact that the KS entropy is the mean asymptotic information needed to reconstruct each step of the orbits of a system. This information approach to a system is contained also in the definition of the complexity of orbits as in Chapter 2. In this section we give an argument leading to a possible relation between decay of correlations, convergence of KS entropy and chaos indexes.

Let (X,T,μ) be a finite mixing invariant system. In [Ka], one can find a first approach to relations between decay of correlations and the ergodic properties of the system, using the notion of *spectral measures* (see [CFS]). Let $f \in L^2(X,\mu)$ such that $\int_X f(x)\, d\mu(x) = 0$, and $C_n(f,f)$ the correlation functions

$$C_n(f,f) = \Big(f \circ T^n , f \Big)_{L^2} = \int_X f(T^n(x))\ f(x)\ d\mu(x)$$

The *spectral measure* σ_f associated to f is a probability Borel measure on the unit circle S^1, such that $C_n(f,f)$ are its Fourier coefficients. Since the system is mixing, $C_n(f,f) \to 0$ as $n \to \infty$. In [Ka] it is proved that there are strong relations between the rate of decay of $C_n(f,f)$, the behaviour of the measure σ_f near the origin and the variance $D(S_n f)$ of the ergodic sums $S_n f = \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i$. Another interesting result of [Ka] is a relation between the asymptotic behaviour, as $n \to \infty$, of $C_n(f,f)$ and of

$$\mu \left(\left\{ x \in X \ / \ \sup_{N \geq n} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) - \int_X f(x)\ d\mu(x) \right| \geq \epsilon \right\} \right)$$

for all $\epsilon > 0$. Hence a result about the speed of convergence in the Birkhoff Ergodic Theorem.

We now approach in a similar way the relation between the convergence of the KS entropy and the decay of correlations. Let Z be a finite measurable partition of X and let $h_\mu(T,Z)$ be the KS entropy relative to Z. By the Shannon-McMillan-Breiman Theorem 1.31 and by the ergodicity of the system (assured by the fact that the system is mixing), we have that for μ-almost any $x \in X$ it holds

$$\lim_{n\to\infty} -\frac{1}{n} \log(\mu(Z_n(x))) = h_\mu(T, Z) \tag{3.14}$$

where we recall that $Z_n(x)$ denotes the set of the iterated partition Z_n containing x.

Theorem 3.19. *Let (X, T, μ) be a finite ergodic invariant dynamical system. Let the correlation functions $C_n(f, g)$ decay with rate $\varphi(n)$ on a Banach space $\mathcal{F}$ of functions, containing the characteristic functions of measurable sets. Then for all $\epsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ it holds*

$$\mu\left(\mathcal{O}(\epsilon, n, Z)\right) \leq \frac{2}{n\epsilon^2} \sum_{k=0}^{n-1} \varphi(k)$$

where

$$\mathcal{O}(\epsilon, n, Z) = \left\{ x \in X \ / \ \left| -\frac{1}{n} \log(\mu(Z_{n+1}(x))) - h_\mu(T, Z) \right| > 3\epsilon \right\}$$

Proof. Let $P_n(x)$ denote the sequence

$$P_n(x) = -\frac{1}{n} \log(\mu(Z_{n+1}(x))) - h_\mu(T, Z)$$

For all integers n it holds

$$P_n(x) = -\frac{1}{n} \log(\mu(Z(T^n(x)))) - \frac{1}{n} \sum_{j=0}^{n-1} \log \frac{\mu(Z_{n+1-j}(T^j(x)))}{\mu(Z_{n-j}(T^{j+1}(x)))} - h_\mu(T, Z)$$

Let define

$$R_n(x) = -\frac{1}{n} \log(\mu(Z(T^n(x))))$$

$$S_n(x) = -\frac{1}{n} \sum_{j=0}^{n-1} \log \frac{\mu(Z_{n+1-j}(T^j(x)))}{\mu(Z_{n-j}(T^{j+1}(x)))} - h_\mu(T, Z)$$

that implies $P_n(x) = R_n(x) + S_n(x)$. Being Z a finite partition, for all x and for all n, we have

$$R_n(x) \leq -\frac{1}{n} \log d_\mu \tag{3.15}$$

where $d_\mu = \min\{\mu(I_i) \ / \ I_i \in Z\}$. Finally we let

$$f_n(x) = -\log \frac{\mu(Z_{n+1}(x))}{\mu(Z_n(T(x)))}$$

from which

$$S_n(x) = \frac{1}{n} \sum_{j=0}^{n-1} \left[f_{n-j}(T^j(x)) - h_\mu(T, Z) \right]$$

From the definition of the iterated partitions, it follows

$$Z_{n+1}(x) \subset T^{-1}(Z_n(T(x))) = (T^{-1} Z_n)(x)$$

hence $\mu(Z_{n+1}(x)) \leq \mu(Z_n(T(x)))$, and $f_n(x) \geq 0$ for all x. Moreover $f_n(x)$ are locally constant functions (if $Z_{n+1}(y) = Z_{n+1}(x)$ then $Z_n(T(y)) = Z_n(T(x))$, hence $f_n(y) = f_n(x)$), so are simple functions on the partition Z_n. In particular, for all n it holds $f_n \in L^\infty$ and $f_n \in \mathcal{F}$.

Lemma 3.20. *Let $(F_n)_n$ be a sequence of functions in $L^2(X, \mu)$, such that for all n it holds $\int_X F_n(x)\ d\mu = 0$. Then for all n we have*

$$\mu \left(\left\{ x \in X \ / \ \left| \sum_{j=0}^{n-1} F_{n-j}(T^j(x)) \right| > \epsilon \right\} \right) \leq \frac{1}{\epsilon^2} \sum_{k,h=0}^{n-1} C_{h-k}(n-h, n-k),$$

where

$$C_h(m, n) = \int_X F_m(T^k(x))\ F_n(x)\ d\mu,$$

if $h > 0$, and otherwise

$$C_h(m, n) = C_{-h}(n, m) = \int_X F_m((x))\ F_n(T^{-h}(x))\ d\mu.$$

To use Lemma 3.20, we write $S_n(x) = S_n^1(x) + S_n^2(x)$, where

$$S_n^1(x) = \frac{1}{n} \sum_{j=0}^{n-1} \left[f_{n-j}(T^j(x)) - h_\mu(T, Z) - \int_X \big(f_{n-j}(T^j(x)) - h_\mu(T, Z) \big)\, d\mu \right]$$

$$S_n^2(x) = \frac{1}{n} \sum_{j=0}^{n-1} \int_X \big(f_{n-j}(T^j(x)) - h_\mu(T, Z) \big)\, d\mu$$

From the properties of the functions $f_n(x)$, we have that we can apply Lemma 3.20 to the sequence $S_n^1(x)$. Moreover, since the correlation functions on the Banach space $\mathcal{F}$ decay as the function $\varphi(n)$, the functions $C_k(m, n)$ of the lemma are independent on m and n, and they only depend on the number k of iterations considered. From this one obtains

$$\mu \left(\left\{ x \in X \ / \ \left| S_n^1(x) \right| > \epsilon \right\} \right) \leq \frac{2}{n\epsilon^2} \sum_{k=0}^{n-1} \varphi(k) \tag{3.16}$$

We still have to study the behaviour of the sequence $S_n^2(x)$. Being $f_n(x)$ simple functions on Z_n, and being the probability measure T-invariant, we have

$$S_n^2(x) = \left[-\frac{1}{n}\sum_{j=0}^{n-1} \mu(Z_{n+1-j}(x)) \log \frac{\mu(Z_{n+1-j}(x))}{\mu((T^{-1}Z_{n-j})(x))} \right] - h_\mu(T,Z)$$

and using $Z_{n+1-j} = Z \vee T^{-1}Z_{n-j}$, we can moreover write

$$S_n^2(x) = \frac{1}{n}\sum_{j=0}^{n-1} H(Z|T^{-1}Z_{n-j}) - h_\mu(T,Z)$$

Hence we can conclude that $S_n^2(x)$ is a sequence independent on x and that

$$\lim_{n\to\infty} S_n^2 = \lim_{n\to\infty} H(Z|T^{-1}Z_n) - h_\mu(T,Z) = 0, \tag{3.17}$$

where the first equality is obtained looking at at the sequence S_n^2 as a Cesaro sequence, and the second equality comes from the definition of the KS entropy in equation (1.26).

To finish the proof of the theorem, we finally notice that

$$\mu(\{x \in X \ / \ |P_n(x)| > 3\epsilon\}) \leq \mu(\{x \in X \ / \ |R_n(x)| + |S_n^1(x)| + |S_n^2(x)| > 3\epsilon\})$$

and, using equations (3.15) and (3.17), we have that for all $\epsilon > 0$ there exists an $n_0 > 0$ such that for all $n \geq n_0$ it holds $-\frac{1}{n}\log d_\mu < \epsilon$ and $|S_n^2| < \epsilon$, hence

$$\mu(\{x \in X \ / \ |P_n(x)| > 3\epsilon\}) \leq \mu(\{x \in X \ / \ |S_n^1(x)| > \epsilon\}) \leq \frac{2}{n\epsilon^2}\sum_{k=0}^{n-1}\varphi(k) \quad \square$$

Proof of Lemma 3.20. Using Chebyshev inequality, we can write

$$\mu\left(\left\{x \in X \ / \ \left|\sum_{h=0}^{n-1} F_{n-h}(T^h(x))\right| > \epsilon\right\}\right) \leq \frac{1}{\epsilon^2}\int_X \left|\sum_{h=0}^{n-1} F_{n-h}(T^h(x))\right|^2 d\mu$$

The lemma now follows from the following equalities:

$$\begin{aligned}\int_X \left|\sum_{h=0}^{n-1} F_{n-h}(T^h(x))\right|^2 d\mu &= \left(\sum_{h=0}^{n-1} F_{n-h}(T^h(x)), \sum_{h=0}^{n-1} F_{n-h}(T^h(x))\right)_{L^2} = \\ &= \sum_{h=0}^{n-1}\int_X \left[F_{n-h}(T^h(x))\sum_{k=0}^{n-1} F_{n-k}(T^k(x))\right] d\mu = \\ &= \sum_{h=0}^{n-1}\int_X \left[\sum_{k=0}^{h} F_{n-h}(T^{h-k}(x))F_{n-k}(x) + \right. \\ &\left. + \sum_{k=h+1}^{n-1} F_{n-h}(x)F_{n-k}(T^{k-h}(x))\right] d\mu = \sum_{h,k=0}^{n-1} C_{h-k}(n-h, n-k) \quad \square\end{aligned}$$

We have thus obtained a relation between the speed of convergence in measure for the Shannon-McMillan-Breiman theorem and the decay of correlations. To use the methods of information theory introduced in Chapter 2, we can define an information function I on the dynamical system (X, T, μ) with partition Z, given by

$$I(x, n, Z) = -\log(\mu(Z_n(x))) \tag{3.18}$$

and the complexity $K_I(x, Z)$, as in Definition 3.1, by

$$K_I(x, Z) = \limsup_{n\to\infty} \frac{I(x, n, Z)}{n}$$

The Shannon-McMillan-Breiman theorem tells us that for a finite ergodic invariant dynamical system (X, T, μ), it holds $K_I(x, Z) = h_\mu(T, Z)$ for μ-almost all $x \in X$. Analogously for an infinite ergodic conservative system it holds $K_I(x, Z) = 0$ for μ-almost all x.

Let (X, T, μ) be a finite ergodic invariant dynamical system and let Z be a finite measurable partition of X. From the Shannon-McMillan-Breiman theorem one can deduce that for μ-almost all $x \in X$ it holds

$$|I(x, n, Z) - h_\mu(T, Z)\ n| = G(x, n, Z) \tag{3.19}$$

where the functions $G(x, n, Z)$ satisfy

$$\lim_{n\to\infty} \frac{G(x, n, Z)}{n} = 0$$

From equation (3.19), we see that the first order behaviour of the functions $I(x, n, Z)$ is independent on x, and it is given by the KS entropy relative to Z. We now turn our attention to the second order, that is the functions $G(x, n, Z)$. Let's remark that in the case $h_\mu(T, Z) = 0$, the functions $G(x, n, Z)$ become the first order approximation of $I(x, n, Z)$.

We now study the asymptotic behaviour for the functions $G(x, n, Z)$. To this aim we define the analogous of the upper chaos index. Let

$$\overline{q}_G(X, \mu, Z) = \inf \left\{ q \in (0, 1)\ /\ \limsup_{n\to\infty} \int_X \frac{G(x, n, Z)}{n^q}\ d\mu(x) = 0 \right\} \tag{3.20}$$

Let's assume that the system has correlation functions decaying as a function $\varphi(n)$, for all functions in a Banach space $\mathcal{F}$ containing the characteristic functions of the measurable sets. We can then use Theorem 3.19. We now start an argument that leads us to a conjecture.

For two sequences $\{a_n\}$ and $\{b_n\}$, if there exist two positive constants c_1 and c_2, and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ it holds

$$c_2 b_n \leq a_n \leq c_1 b_n$$

we write that $a_n \asymp b_n$.

With this notation, let's assume for a moment a stronger result for Theorem 3.19, that is

$$\mu(\mathcal{O}(\epsilon, n, Z)) \asymp \frac{2}{n\epsilon^2} \sum_{k=0}^{n-1} \varphi(k)$$

Then, writing for a $q > \overline{q}_G(X, \mu, Z)$

$$\int_X \frac{G(x,n,Z)}{n^q}\, d\mu(x) \geq \int_{\mathcal{O}(\epsilon,n,Z)} 3\epsilon n^{1-q}\, d\mu(x) \geq \frac{6c_2}{\epsilon n^q} \sum_{k=0}^{n-1} \varphi(k)$$

for a positive constant c_2, we can draw the following conclusions. If $\varphi(n) \asymp \vartheta^n$, for a constant $\vartheta \in (0,1)$, so the system is exponentially mixing on $\mathcal{F}$, this inequality gives no problem for all $\overline{q}_G(X,\mu,Z)$. If instead $\varphi(n) \asymp n^{-\alpha}$ for a real $\alpha \in (0,1)$, then the previous inequality implies that $\overline{q}_G(X,\mu,Z) \geq 1-\alpha$. We now state our conjecture.

CONJECTURE. For a finite ergodic invariant system (X, T, μ) with finite partition Z, with correlations decaying as a function $\varphi(n)$ on a Banach space $\mathcal{F}$ containing the characteristic functions of measurable subsets of X, it holds:

- $\varphi(n) \asymp \vartheta^n$, for a constant $\vartheta \in (0,1)$, implies $\overline{q}_G(X,\mu,Z) = 0$;
- $\varphi(n) \asymp n^{-\alpha}$ for a real $\alpha \in (0,1)$ implies $\overline{q}_G(X,\mu,Z) = 1-\alpha$.

In Section 4.3.1, we will study a dynamical system satisfying this conjecture, if we restrict ourselves to a particular class of characteristic functions. For that system, we can go further, studying also the situation in the case of an infinite invariant measure. In this case, one studies the so-called *scaling rate* $\frac{\mu(B\cap T^{-n}(B))}{(\mu(B))^2}$ (see [Is1]), for subsets B in the same particular class of characteristic functions, and the result is that

$$\frac{\mu(B \cap T^{-n}(B))}{(\mu(B))^2} \sim \frac{I(x,n,Z)}{n}$$

where $I(x,n,Z) = -\log(\mu(Z_n(x))) = G(x,n,Z)$, being null the KS entropy relative to Z.

We finally discuss the problem of actually compute the rate of decay of correlations of a dynamical system. For results on this direction see [Li]. We notice that our conjecture could help in this direction. Indeed, in Chapter 6 we will study the problem of actually computing the behaviour of the Algorithmic Information Content of a system. Moreover, the *AIC* is, in a sense explained in Section 2.3, the best information function. So we could hope that our conjecture could be extended to the upper chaos index $\overline{q}(X,\mu)$ defined using the AIC instead of the information function I of equation (3.18). So, in conclusion, being able to compute the behaviour of the AIC, we would be able to compute the rate of decay of the correlations. All this is of course only work in progress.

Chapter 4

Examples of weak chaos

In this chapter we study some examples of chaotic dynamical systems, showing the applications of the methods introduced in the previous chapters. As explained before, the importance of this approach release in the classification of weakly chaotic and infinite dynamical systems. Hence we restrict to study examples of this kind, and show that we find many possible asymptotic behaviours for the Algorithmic Information Content of the symbolic orbits.

4.1 Rotations and Interval Exchange Transformations

Let's consider rigid transformations of the unit circle that preserve the Lebesgue measure m. The simplest example of such a transformation is given by the rigid rotation of an angle $\alpha \in (0,1)$.

Let $X = S^1$, considered as the interval $[0,1]$ with end points identified. Consider on X the Borel σ-algebra of the euclidean topology, and the transformation $T : X \to X$ given by

$$T(x) = x + \alpha \ (\mathrm{mod}\ 1) \tag{4.1}$$

for a real $\alpha \in (0,1)$. The Lebesgue measure m on $[0,1]$ is invariant. It is well known that the properties of the rotation T depend on α. If $\alpha \in \mathbb{Q}$, then the map T is periodic on all $x \in X$, and hence the dynamical system (X,T,m) is regular. If instead $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, then all orbits are dense in X (T is *minimal*) and m is the unique invariant measure for (X,T). Hence m is the physical measure and it is also ergodic. We restrict ourselves to study the latter case.

Let us consider the partition $Z = \{I_0, I_1\}$ of X into the half-open arcs $I_0 = [0,1/2)$ and $I_1 = [1/2,1)$. Since T is minimal, Z is generating, and moreover the iterated partition Z_n is made out of $2n$ arcs. This can be easily realized by induction: when passing from Z_{n-1} to Z_n, one has to add to the endpoints of the arcs belonging to Z_{n-1}, the two new points $T^{-n}(0)$ and $T^{-n}(1/2)$. Thus,

$$0 \leq h_m(T, Z) \leq \lim_{n\to\infty} \frac{\log(2n)}{n} = 0$$

and the dynamical system (X, T) is then weakly chaotic (according to the definition given in Section 1.3.2).

In our approach to rotations, we will make use of the expansion in *continued fractions* for numbers in the unit interval $[0, 1]$. So let us briefly recall the main definitions and properties. For more details see [RS]. For a real $\alpha \in (0, 1)$ its continued fractions expansions is given by a sequence of integers $(a_i)_{i\geq 1}$ such that

$$\alpha = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} \equiv [a_1, a_2, a_3, \dots] \tag{4.2}$$

We recall that rational numbers have a finite expansion and denote by $\frac{p_k}{q_k} = [a_1, a_2, \dots, a_k]$ the sequence of the rational approximants of a real number α with expansion $[a_1, a_2, a_3, \dots]$. In the following we need the following estimate for the approximation of α by the rationals $\frac{p_k}{q_k}$

$$\frac{1}{q_k^2 \cdot (a_{k+1} + 2)} < \left| \alpha - \frac{p_k}{q_k} \right| < \frac{1}{q_k^2 \cdot a_{k+1}} \tag{4.3}$$

We now study the asymptotic behaviour of the AIC for the symbolic sequences of the ergodic rotation (X, T, m), with $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

Let σ denote the infinite sequence in $\mathbb{N}^{\mathbb{N}}$ associated to the angle of rotation α, and in the same way let $[\xi_1, \xi_2, \dots, \xi_n, \dots]$ be the continued fraction expansion for a number $x \in (0, 1)$ (the initial condition), and denote by s the associated word. We also denote by $\frac{p_k}{q_k}$ and $\frac{r_h}{t_h}$, respectively, the rational approximants of α and x:

$$\frac{p_k}{q_k} = [a_1, a_2, \dots, a_k] \ , \quad \frac{r_h}{t_h} = [\xi_1, \xi_2, \dots, \xi_h] \ .$$

Theorem 4.1 ([BGI]). *For all $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and for m-almost all $x \in X$ we have $\overline{q}(x, Z) = 0$, where Z is the partition described above. Hence for the global chaos index it holds $\overline{q}(X, m, Z) = 0$.*

Proof. By the estimate (4.3) we have that the knowledge of the first $k + 1$ digits (we remark that n digits specify p_i and q_i for $i = 1, \dots, n$) in the continued fractions expansion of α allows us to say that $T^n(x)$ is contained in the interval

$$I_n^k := \left[(x + n\frac{p_k}{q_k} - \frac{n}{q_k^2 a_{k+1}}) (\operatorname{mod} 1) \, , (x + n\frac{p_k}{q_k} + \frac{n}{q_k^2 a_{k+1}}) (\operatorname{mod} 1) \right]$$

The knowledge of the endpoints of this interval yields unambiguously the symbol ω_n in the symbolic orbit of x provided I_n^k does not contain 0 or 1/2. Let T_k be the translation of the angle $\frac{p_k}{q_k}$ and ω^k be the symbolic orbit of the point x with the map T_k on Z: we have $\omega_n^k = \omega_n$ provided I_n^k does not contain 0 or 1/2. Set

$$A(k,n) := \left\{ x \in X \,:\, \omega_n \neq \omega_n^k \right\}$$

A short calculation yields

$$m(A(k,n)) \leq \frac{4n}{q_k^2 a_{k+1}}$$

for all k and n. The measure of the set of initial points x that have at least one symbol among $\omega_1 \dots \omega_n$ different from the corresponding symbols in the word ω^k is then estimated as

$$m\left(\bigcup_{i=1}^{n} A(k,i)\right) \leq \sum_{i=1}^{n} m(A(k,i)) \leq \sum_{i=1}^{n} \frac{4i}{q_k^2 a_{k+1}} = \frac{2n(n+1)}{q_k^2 a_{k+1}}$$

Let now $k(n)$ be a sequence of integers such that

$$\sum_{n=1}^{\infty} \frac{4n}{q_{k(n)}^2 a_{k(n)+1}} < \infty \tag{4.4}$$

By Borel-Cantelli Lemma we have that for m-almost all $x \in X$ there is $n_0 > 0$ such that $x \in X \setminus A(k,n)$ for all $n \geq n_0$ and therefore

$$AIC(x,n) = AIC(\omega_0\omega_1 \dots \omega_{n-1}) \leq AIC(\omega_o^{k(n)}\omega_1^{k(n)} \dots \omega_{n-1}^{k(n)}) + \text{const}$$

that is to specify the first n symbols of the sequence ω it is enough to specify the first n symbols of the symbolic word $\omega^{k(n)}$ obtained by the orbit of the point x drawn using the periodic map $T_{k(n)}$, defined by $T_{k(n)}(x) = x + \frac{p_{k(n)}}{q_{k(n)}} \pmod 1$.

We have now to estimate the information content of the first n symbols of the word ω^k. First of all we have that to reconstruct ω^k it is necessary to know $\frac{p_k}{q_k}$, hence we need to know the first k symbols of the sequence σ.

Now it remains the uncertainty about the initial condition we use to generate ω^k. We repeat the same argument as before. Let $[\xi_1, \xi_2, \dots, \xi_n, \dots]$ be the expansion in continued fractions of the initial condition x. If we know only the first h symbols in this expansion, we can approximate x by the rational number $\frac{r_h}{t_h}$ so that $x + n\frac{p_k}{q_k} \pmod 1$ is contained in the interval

$$J_n^h := \left[(\frac{r_h}{t_h} + n\frac{p_k}{q_k} - \frac{1}{t_k^2\,\xi_{k+1}}) \pmod 1,\ (\frac{r_h}{t_h} + \frac{p_k}{q_k} + \frac{1}{t_k^2\,\xi_{k+1}}) \pmod 1 \right]$$

Reasoning as above and choosing $h(n)$ so that

$$\sum_{n=1}^{\infty} \frac{4}{t_{h(n)}^2\, \xi_{h(n)+1}} < \infty \tag{4.5}$$

we conclude that for m-almost all $x \in X$ there exists $n_1 > 0$ such that for all $n \geq n_1$ it is possible to determine ω_n^k just knowing $[\xi_1, \xi_2, \dots, \xi_{h(n)}]$.

From this argument we can conclude that for almost any $x \in X$ we have that there exist integer functions $k(n)$ and $h(n)$ such that for n large enough

$$\begin{aligned} AIC(x,n) &\leq AIC(\omega_0^{k(n)} \omega_1^{k(n)} \dots \omega_{n-1}^{k(n)}) + \text{const} \leq \\ &\leq AIC(\sigma_{k(n)}) + AIC(s_{h(n)}) + \log_2 n + \text{const} \end{aligned} \tag{4.6}$$

where $AIC(\sigma_{k(n)})$ and $AIC(s_{h(n)})$ denote the Algorithmic Information Content of the first $k(n)$ and $h(n)$ symbols of the sequences σ and s, respectively, whereas the term $\log_2 n$ is the information needed to reconstruct the $h(n)$-long symbolic orbit of the rational point $\frac{r_{h(n)}}{t_{h(n)}}$ by the map $T_{k(n)}$.

We now study the general problem of estimating $AIC(\sigma_k)$. A general estimate is the following

$$AIC(\sigma_k) \leq k\ \log_2(\max\{a_i\ :\ i = 1, \dots, k\}) \tag{4.7}$$

where we recall that a_i are the partial quotients in the expansion of α. This estimate comes from the general estimate of the Algorithmic Information Content of a word written on a given alphabet by the length of the word times the logarithm of the number of letters in the alphabet.

To conclude the proof we need the following lemma.

Lemma 4.2. *Let $f(k) = \max_{i \leq k} a_i$ and $k(n)$ a sequence such that*

$$\sum_{n=1}^{\infty} \frac{4n}{q_{k(n)}^2 a_{k(n)+1}} < \infty \tag{4.8}$$

Then it is possible to choose $k(n)$ such that $k(n) \log(f(k(n))) = o(n^\epsilon)$ for all $\epsilon > 0$.

Applying this lemma to $AIC(\sigma_{k(n)})$ and, with an entirely analogous argument, to $AIC(s_{h(n)})$, we obtain the thesis of the theorem by equation (4.6). □

Proof of Lemma 4.2. Given $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, the best bounds we can give on the denominators q_k are

$$\frac{1}{2 \prod_{i=0}^{k-1} G^i(\alpha)} < q_k < \frac{1}{\prod_{i=0}^{k-1} G^i(\alpha)} \tag{4.9}$$

where $G : [0,1] \to [0,1]$ is the Gauss transformation given by $G(x) = \{1/x\}$ for $x > 0$ and $G(0) = 0$. From this we obtain $q_k \geq \frac{1}{2} f(k)$ for all k. Moreover, it is well known that the golden mean $\bar{\alpha} = \frac{\sqrt{5}-1}{2}$ is the number for which the denominators $\bar{q}_k$ are the smallest, and they are equal to the Fibonacci numbers F_k. Hence in general, for all irrational α it holds $q_k \geq F_k$ for all k.

We have thus obtained two different estimates for the denominators q_k, and the two estimates have the same order if the function $f(k)$ grows exponentially.

When $f(k) = \mathcal{O}(exp(k))$, we can use the estimate $q_k \geq F_k$. Hence

$$\sum_{n=1}^{\infty} \frac{4n}{q^2_{k(n)} a_{k(n)+1}} \leq \sum_{n=1}^{\infty} \frac{4n}{F^2_{k(n)}} \sim \sum_{n=1}^{\infty} \frac{n}{\left(\frac{1}{\bar{\alpha}}\right)^{2k(n)}}$$

Hence one can choose $k(n) \sim \log n^{\frac{3}{2}}$ to obtain condition (4.8). Moreover $k \log(f(k)) = \mathcal{O}(k^2)$, from which we obtain the thesis of the lemma.

If instead $f(k)$ grows faster than an exponential, then the better estimate for the denominators q_k is $q_k \geq f(k)$. In this case

$$\sum_{n=1}^{\infty} \frac{4n}{q^2_{k(n)} a_{k(n)+1}} \leq \sum_{n=1}^{\infty} \frac{4n}{f^2(k(n))}$$

and condition (4.8) is obtained choosing $f(k(n)) \sim n^{\frac{3}{2}}$. From this is then easy to see that $k(n) = o(n^\epsilon)$ for all $\epsilon > 0$, hence $k(n)\log(f(k(n))) \sim k(n) \log n = o(n^\epsilon)$ for all $\epsilon > 0$. $\square$

In [BGI], ergodic rotations are studied also from the point of view of the Poincaré recurrence time.

From the proof of Theorem 4.1, it comes that the uncertainty on the orbits of points $x \in X$ depends on the uncertainty on the initial condition x, but also on the uncertainty on the angle α, that being irrational, needs an infinite amount of information to be known.

We now study the Interval Exchange Transformations. Let the phase space X to be again the unit circle S^1, obtained as the unit interval $[0,1]$ with identified end points. Let Z be a partition into a number $r \geq 2$ of intervals $I_j = [d^j, d^{j+1})$, such that $d^1 = 0$ and $d^{r+1} = 1$. Let moreover π be a permutation of the set of integers $\{1, \dots, r\}$.

Definition 4.3. The *Interval Exchange Transformation (IET)* associated to the partition Z of X and to the permutation π is the transformation $T(\pi, Z)$ of X such that: (i) on each interval I_j the transformation is of the form $T(x) = x + \alpha^j \pmod 1$, for real numbers $\alpha^1, \dots, \alpha^r$ in $(0,1)$; (ii) if $I'_j := TI_j$ then $Z' = \{I'_{\pi(1)}, \dots, I'_{\pi(r)}\}$ is a partition of X. These two conditions imply that the numbers α^j are uniquely defined given the points $d^1, \dots, d^r$ and the permutation π.

Let's notice that for $r = 2$ we obtain a rotation of the unit circle. An IET preserves the Lebesgue measure m, but m is not the only invariant measure. We moreover assume that the IET is aperiodic, in the sense that it has no periodic points. In this case the Lebesgue measure m is also ergodic. For more details and properties of IET we refer to [CFS].

To study the behaviour of the AIC for orbits of an IET, we use again the continued fraction expansions. Given the left end points $d^1 = 0, d^2, \ldots, d^r$ of the partition Z of X, we denote their continued fraction expansions by $[\delta_{(i,1)}, \ldots, \delta_{(i,n)}, \ldots]$, for $i = 2, \ldots, r$. Let $\frac{p_{(i,k)}}{q_{(i,k)}} = [\delta_{(i,1)}, \ldots, \delta_{(i,k)}]$ be the sequence of rational approximants of d^i, then it holds (see equation (4.3)) for all $i = 2, \ldots, r$

$$\left| d^i - \frac{p_{(i,k)}}{q_{(i,k)}} \right| \leq \frac{1}{q^2_{(i,k)}\ \delta_{(i,k+1)}}$$

We denote by i_0 the index i such that the approximation $\frac{1}{q^2_{(i,k)}\ \delta_{(i,k+1)}}$ is the worst.

From the definition of an IET, we have that the angles α^j of rotation for the different intervals can be easily obtained from the points d^i and the permutation π. In particular we find that, using the rational numbers $\frac{p_{(i,k)}}{q_{(i,k)}}$ instead of the points d^i, we have that the angles α^j are approximated by rational numbers β^j, such that for all $j = 1, \ldots, r$

$$\left| \alpha^j - \beta^j \right| \leq \frac{r}{q^2_{(i_0,k)}\ \delta_{(i_0,k+1)}} \tag{4.10}$$

Theorem 4.4. *For an aperiodic interval exchange transformation $T(\pi, Z)$, associated to the permutation π and to the partition Z, we have that for m-almost all $x \in X$ it holds $\overline{q}(x, Z) = 0$, hence $\overline{q}(X, m, Z) = 0$.*

Proof. The proof is analogous to that of Theorem 4.1. Let $x \in X$ be an initial condition for the IET $T(\pi, Z)$, and denote by $[\xi_1, \ldots, \xi_n, \ldots]$ its continued fractions expansion. Again let $\frac{r_h}{t_h} = [\xi_1, \ldots, \xi_h]$ be the rational approximants of x. From equation (4.3) we have

$$\left| x - \frac{r_h}{t_h} \right| \leq \frac{1}{t_h^2\ \xi_{h+1}}$$

It is evident that for m-almost all $x \in X$ there exist h and k such that the rational numbers $\frac{r_h}{t_h}$ and $\frac{p_{(i,k)}}{q_{(i,k)}}$, for $i = 2, \ldots, r$, are well ordinated on the real line. Hence we obtain the value of $\omega_0(x)$, that is the first symbol in the symbolic sequence $\omega(x) \in \Omega := \{1, \ldots, r\}^{\mathbb{N}}$ associated to the IET by the usual method of symbolic representation.

Given that we know $\omega_0(x)$, we have that $T(x) \in A_1^{(k,h)}$, where

$$A_1^{(k,h)} = \left[\frac{r_h}{t_h} + \beta^{\omega_0(x)} - \epsilon_1(h,k),\ \ \frac{r_h}{t_h} + \beta^{\omega_0(x)} + \epsilon_1(h,k) \right]$$

$$\epsilon_1(h,k) = \frac{1}{t_h^2\ \xi_{h+1}} + \frac{r}{q^2_{(i_0,k)}\ \delta_{(i_0,k+1)}}$$

Hence $\omega_1(x)$ is well defined if the set $A_1^{(k,h)}$ does not intersect the intervals

$$B(i,k) = \left[d^i - \frac{1}{q^2_{(i_0,k)}\ \delta_{(i_0,k+1)}},\quad d^i + \frac{1}{q^2_{(i_0,k)}\ \delta_{(i_0,k+1)}} \right]$$

for all $i = 1, \dots, r$. Then if we define

$$P_1^{(k,h)} = \left\{ x \in X\ /\ A_1^{(k,h)} \cap B(i,k) \neq \emptyset \ \text{ for some } i \right\}$$

we have easily that

$$m(P_1^{(k,h)}) \leq r \left(\frac{2}{t_h^2\ \xi_{h+1}} + \frac{2(r+1)}{q^2_{(i_0,k)}\ \delta_{(i_0,k+1)}} \right) \tag{4.11}$$

Let us assume now that we know $\omega_0(x)$ and $\omega_1(x)$. We can then say that $T^2(x) \in A_2^{(k,h)}$, where now

$$A_2^{(k,h)} = \left[\frac{r_h}{t_h} + \beta^{\omega_0(x)} + \beta^{\omega_1(x)} - \epsilon_2(h,k),\quad \frac{r_h}{t_h} + \beta^{\omega_0(x)} + \beta^{\omega_1(x)} + \epsilon_2(h,k) \right]$$

$$\epsilon_2(h,k) = \frac{1}{t_h^2\ \xi_{h+1}} + \frac{2r}{q^2_{(i_0,k)}\ \delta_{(i_0,k+1)}}$$

Again $\omega_2(x)$ is well defined if the set $A_2^{(k,h)}$ does not intersect the intervals $B(i,k)$. Then we define

$$P_2^{(k,h)} = \left\{ x \in X\ /\ A_2^{(k,h)} \cap B(i,k) \neq \emptyset \ \text{ for some } i \right\}$$

and obtain

$$m(P_2^{(k,h)}) \leq r \left(\frac{2}{t_h^2\ \xi_{h+1}} + \frac{2(2r+1)}{q^2_{(i_0,k)}\ \delta_{(i_0,k+1)}} \right) \tag{4.12}$$

Iterating this argument, we can conclude that for a given n, the measure of the set of initial conditions x for which the knowledge of $\frac{r_h}{t_h}$ and $\frac{p_{(i,k)}}{q_{(i,k)}}$, for $i = 2, \dots, r$, is not enough to construct $(\omega_0(x), \dots, \omega_{n-1}(x))$, is estimated from above by

$$m\left(\bigcup_{j=1}^{n} P_j^{(k,h)} \right) \leq \sum_{j=1}^{n} r \left(\frac{2}{t_h^2\ \xi_{h+1}} + \frac{2(jr+1)}{q^2_{(i_0,k)}\ \delta_{(i_0,k+1)}} \right) \tag{4.13}$$

At this point the conclusion of the theorem follows as for Theorem 4.1. From the Borel-Cantelli Lemma, choosing sequences $k(n)$ and $h(n)$ such that

$$\sum_{n=1}^{\infty} r\left(\frac{2}{t^2_{h(n)}\ \xi_{h(n)+1}} + \frac{2(nr+1)}{q^2_{(i_0,k(n))}\ \delta_{(i_0,k(n)+1)}}\right) < \infty \tag{4.14}$$

we have that for m-almost all initial conditions x

$$\begin{aligned} AIC(x,n,Z) \ \leq \ & \text{const } + \log_2 n + AIC(\xi_1,\dots,\xi_{h(n)}) + \\ & + \textstyle\sum_{i=2}^{r} \ AIC(\delta_1^i,\dots,\delta_{k(n)}^i) \end{aligned} \tag{4.15}$$

where the constant contains also the information necessary to transmit the permutation π to which the IET is associated, and the term $\log_2 n$ accounts for the information necessary to transmit the number n of iterations that have to be performed using the rational numbers $\frac{r_h}{t_h}$ and $\frac{p_{(i,k)}}{q_{(i,k)}}$, for $i = 2,\dots,r$.

At this point to conclude the theorem we simply apply equation (4.7) and Lemma 4.2. □

4.2 The logistic map at the chaos threshold

We now restrict our attention to study the behaviour of the information content of one of the best known examples of one dimensional dynamical systems, the *logistic map at the chaos threshold*. The results are an improvement of those presented in [BM].

To study the behaviour of the logistic map, we use the theory of *unimodal maps*. For references on these subjects see [CE], [DS] and [De].

The *logistic map* is the dynamical system defined on the space $X = [0,1]$ by

$$T_\lambda(x) = \lambda x(1-x)\ , \quad 1 \leq \lambda \leq 4, \tag{4.16}$$

where λ is a parameter. For some values of the parameter λ, this is a very simple example of a map with an extremely complicated behaviour. The dynamics of the map have been studied extensively, and there are many important results that have been generalized to one dimensional dynamical systems.

Before studying the behaviour of the logistic map we recall two basic definitions in the theory of one dimensional dynamical systems.

Definition 4.5. Given a map T on the unit interval $[0,1]$ such that $T \in C^3$, the *Schwarzian derivative* $ST(x)$ of T at a point x, where $T'(x) \neq 0$, is given by

$$ST(x) := \frac{T'''(x)}{T'(x)} - \frac{3}{2}\left(\frac{T''(x)}{T'(x)}\right)^2 \tag{4.17}$$

Definition 4.6. A map $T \in C^1$ on $[0,1]$ is *unimodal* if there exists exactly one critical point $c \in (0,1)$ and $T(0) = T(1) = 0$.

An important property for maps of the unit interval is to have $ST < 0$ at all points where the Schwarzian derivative is definite.

Theorem 4.7 ([DS]). *Let T be a C^3 unimodal map on $[0,1]$ such that $ST < 0$. If $T'(0) \geq 1$, that is 0 is a repelling fixed point, then T has at most one attracting periodic orbit. Moreover, each attractor is such that the map T restricted to it is topologically transitive (see Definition 1.22), and it has one of the following structures:*

1. *a periodic orbit;*
2. *a Cantor set given by the topological closure of the orbit of the critical point;*
3. *a finite union of N intervals, determined by the first $2N$ iterations of the critical point, such that the critical point belongs to the interior of this union and T^N restricted to any of these intervals is chaotic (in the sense of Definition 1.24) and topologically conjugate to a piecewise linear map.*

From the definitions, it is easy to see that the logistic map T_λ is a C^3 unimodal map with critical point $c = \frac{1}{2}$ and with $ST < 0$. So the conclusions of the theorem apply.

Consider first $\lambda \in (1,3)$. There are two fixed points, $x_0 = 0$ and $x_1 = (\lambda - 1)/\lambda$. The fixed point at the origin is unstable and the point x_1 is stable. Then every orbit ends up on the fixed point x_1. Hence in this range of values for λ, the only physical measure is the Dirac delta at the point x_1, and the system is regular.

When $\lambda = \lambda_1 = 3$, we have the first bifurcation. The derivative of T_3 at the point x_1 is equal to -1, and the fixed point is now neutrally stable. Moreover one periodic orbit of period 2 is generated and it becomes stable as soon as λ becomes greater than λ_1. This is a period doubling bifurcation.

This kind of bifurcation repeats over and over at different values λ_n of the parameter. That is, when $\lambda \in (\lambda_n, \lambda_{n+1})$, there is one stable periodic orbit of period 2^n, and unstable periodic orbits of period 2^j, $j = 0, \dots, n$. Hence the system is regular and the only physical measure is concentrated on the stable periodic orbit. When $\lambda = \lambda_{n+1}$, the stable periodic orbit loses its stability and a new periodic orbit of period 2^{n+1} is generated.

One of the main features of this bifurcation sequence is that the bifurcation parameters λ_n accumulate at $\lambda = \lambda_\infty$, where

$$\lambda_\infty = 3.569945671870944901842\dots \tag{4.18}$$

In the following we will use the fact, showed in [Fei], that the bifurcation sequence (λ_n) can be generated by the relation

$$\lambda_n = \lambda_\infty - \frac{b}{\delta^n}, \tag{4.19}$$

where δ is the so-called *Feigenbaum constant* and b is a suitable constant.

We call the value λ_∞ the *chaos threshold*, since for $\lambda < \lambda_\infty$ the system is periodic, hence we are in case 1 of Theorem 4.7, whereas there are values of $\lambda > \lambda_\infty$ for which the system (X, T_λ) is in case 3 of Theorem 4.7.

In this section we study the behaviour of the dynamical system (X, T_{λ_∞}). The following theorem gives that in this case we obtain case 2 of Theorem 4.7.

The dynamical system (X, T_{λ_∞}) have countably many unstable periodic orbits of periods 2^j, for all $j \in \mathbb{N}$, and an attractor Ω that is a Cantor set, the so-called *Feigenbaum attractor*. It holds the following theorem:

Theorem 4.8 ([CE],Theorem III.3.5). *The logistic map T_{λ_∞} at the chaos threshold has an invariant Cantor set Ω such that*

1. *there is a decreasing chain of closed subsets*
$$J^{(0)} \supset J^{(1)} \supset J^{(2)} \supset \dots,$$
each of which contains the critical point $c = 1/2$, and each of which is mapped onto itself by T_{λ_∞};

2. *each $J^{(i)}$ is a disjoint union of 2^i closed intervals. $J^{(i+1)}$ is constructed by deleting an open subinterval from the interior of each of the intervals making up $J^{(i)}$;*

3. *T_{λ_∞} maps each of the intervals making up $J^{(i)}$ onto another one; the induced action on the set of intervals is a cyclic permutation of order 2^i;*

4. *$\Omega = \cap_i J^{(i)}$. T_{λ_∞} maps Ω onto itself in a one-to-one fashion. Every orbit in Ω is dense in Ω;*

5. *for each $k \in \mathbb{N}$, T_{λ_∞} has exactly one periodic orbit of period 2^k. This periodic orbit is repelling and does not belong to $J^{(k+1)}$. Moreover this periodic orbit belongs to $J^{(k)} \setminus J^{(k+1)}$, and each point of the orbit belongs to one of the intervals of $J^{(k)}$;*

6. *every orbit of T_{λ_∞} either lands after a finite number of steps exactly on one of the periodic orbits enumerated in 5, or converges to the Cantor set Ω in the sense that, for each k, it is eventually contained in $J^{(k)}$. There are only countably many orbits of the type enumerated in 5.*

The dynamical system (X, T_{λ_∞}) has null topological entropy (see equation (1.20)), and hence, by the *Variational Principle*, the KS entropy of the system is null with respect to any invariant probability measure. The only physical measure μ is concentrated on the Cantor set Ω.

Moreover we now prove that the logistic map at the chaos threshold has no *sensitivity to initial conditions* (see Definition 1.23), hence it is not a chaotic system, according to Definition 1.24.

Theorem 4.9 ([BM]). *The logistic map T_{λ_∞} at the chaos threshold has no sensitivity to initial conditions. Indeed there exists a subset $Y \subset [0,1]$ with $m(Y) = 1$, for the Lebesgue measure m, such that for all $\delta > 0$ and for all $x \in Y$, there exists a neighbourhood $U(\delta)$ of x such that $\forall y \in U(\delta)$ and $\forall n \in \mathbb{N}$ we have $d(T^n_{\lambda_\infty}(x), T^n_{\lambda_\infty}(y)) < \delta$.*

Proof. By Theorem 4.8 (6), we have that eventually almost every orbit, with respect to the Lebesgue measure m, ends up on $J^{(k)}$ for all k. Moreover by Theorem 4.8 (1), we have that $T_{\lambda_\infty}(J^{(k)}) \subset J^{(k)}$.

Let now $\delta > 0$ be fixed. Being Ω a Cantor set, there exists k_δ such that $m(J_i^{(k_\delta)}) < \delta$ for all $i = 1, \ldots, 2^{k_\delta}$, where by $J_i^{(k_\delta)}$ we denote the intervals making up $J^{(k_\delta)}$. We have (Theorem 4.8 (3)) that the action of T_{λ_∞} on $J^{(k_\delta)}$ is a cyclic permutation of the 2^{k_δ} intervals. Then if $x \in J_i^{(k_\delta)}$, for some i, and $U(\delta)$ is a neighbourhood of x contained in $J_i^{(k_\delta)}$, then, for all n, $T^n_{\lambda_\infty}(U(\delta)) \subset J_j^{(k_\delta)}$, for some j. Hence $\forall y \in U(\delta)$ and $\forall n \in \mathbb{N}$, we have $d(T^n_{\lambda_\infty}(x), T^n_{\lambda_\infty}(y)) < \delta$.

Let now Y be the set of the points whose orbit is eventually contained in $J^{(k_\delta)}$. By Theorem 4.8 (6), we have $m(Y) = 1$. Let now be $x \in Y$, then there exists a neighbourhood $U(\delta)$ of x, such that, if $T^N_{\lambda_\infty}(x) \in J_i^{(k_\delta)}$ for some i, then $T^N_{\lambda_\infty}(U(\delta)) \subset J_i^{(k_\delta)}$. The neighbourhood $U(\delta)$ exists by continuity of the map T_{λ_∞}. Now, we can repeat the same argument as before, showing that $\forall y \in U(\delta)$ and $\forall n \in \mathbb{N}$, we have $d(T^n_{\lambda_\infty}(x), T^n_{\lambda_\infty}(y)) < \delta$ (if necessary it is possible to shrink the neighbourhood $U(\delta)$ to have the previous relation for all $n \in \mathbb{N}$). $\square$

However, for the map T_{λ_∞} numerical experiments show a power law divergence of nearby orbits. Namely if $\Delta x(0)$ denotes the initial separation between two orbits, it has been observed that it holds the relation

$$\lim_{\Delta x(0) \to 0} \frac{\Delta x(n)}{\Delta x(0)} \simeq n^k \tag{4.20}$$

where $\Delta x(n)$ denotes the distance between the two orbits at time n. The constant k has been found to be approximately equal to $1.3236 \ldots$ ([TPZ]).

The presence of this divergence can be explained looking at the behaviour of the *separation sets* $B(x, \epsilon, n)$ introduced in Definition 1.25.

Using the notion of separation sets, we apply Theorem 4.9 to obtain the following

Corollary 4.10. *For all $\epsilon > 0$ and for m-almost all $x \in [0,1]$, there exists a $N(x,\epsilon) \in \mathbb{N}$, such that $B(x,\epsilon,n) = B(x,\epsilon,m)$ for all $n, m \geq N(x,\epsilon)$.*

Hence, given m-almost any point $x \in X$ and ϵ, the set $B(x,\epsilon,n)$ shrinks as n increases, until n reaches the value $N(x,\epsilon)$. At this point the set $B(x,\epsilon,n)$ doesn't change any more. But, if we are interested in estimate the order of divergence of nearby orbits at the point x, we have to consider the limit

$$m(B(x,n)) = \lim_{\epsilon \to 0^+} m(B(x,\epsilon,n)) \tag{4.21}$$

that corresponds to the limit for $\Delta x(0) \to 0$ in equation (4.20). We find that

$$\lim_{\epsilon \to 0^+} N(x,\epsilon) = +\infty, \tag{4.22}$$

for m-almost all $x \in X$, hence the function $m(B(x,n))$ is a decreasing function of n, whose order gives information on the order of the local divergence.

To study the behaviour of the *AIC* for the symbolic sequences of the logistic map at the chaos threshold, we use *renormalization theory.* We now briefly present only the basic definitions that we need for our aims. For details we refer to [DS].

Definition 4.11. Let $T : [0,1] \to [0,1]$ be a unimodal map and let $c \in (0,1)$ be the critical point. We say that T is *renormalizable* if there exist an interval $J \subset [0,1]$ and an integer p such that: i) $c \in J$; ii) $T^p(J) \subseteq J$ and $T^i(J) \cap T^j(J) = \emptyset$ for all $i,j = 0,\dots,p-1$ such that $i \neq j$; iii) $T^p|_J$ is unimodal with critical point in c.

We use the fact the logistic map at the chaos threshold is *infinitely renormalizable.* It means that there exist a sequence of intervals $J^0 \supset J^1 \supset \dots \supset J^n \supset \dots$ and a sequence of integers $\{p_n\}$ such that

1. $c \in J^0$, T is renormalizable on the interval J^0 with integer p_0, and let $T_1 := T^{p_0}|_{J^0}$;
2. for all $n \geq 1$ it holds $c \in J^n$, T_n is renormalizable on the interval J^n with integer p_n, and define recursively $T_{n+1} := T_n^{p_n}|_{J^n}$.

For the logistic map T_{λ_∞} all the integers p_n are equal to 2. Moreover we remark that the closed subsets $J^{(n)}$ of Theorem 4.8 are given by the union of $(T_{\lambda_\infty})_n^i(J^n)$ for $i = 0,\dots,p_n - 1$.

Another important notion for one dimensional multimodal dynamical systems is the *kneading theory.* It is based on the symbolic representation of the orbits of the system, with respect to the partition made by the critical points. In particular for unimodal maps it is done with respect to the partition $Z = \{I_0, I_1\} = \{[0,c), [c,1]\}$. The main tool of the kneading theory is the *kneading invariant* $K(T)$ of a map T, given by the symbolic sequence of the point $c_1 = T(c)$. For more detail we refer to [DS].

For the logistic map we have that for $\lambda < \lambda_\infty$, being the map periodic, also the kneading invariant $K(T_\lambda)$ is periodic. Instead $K(T_{\lambda_\infty})$ can be built by the following algorithm: (i) begin with 1; (ii) double the sequence, changing the last symbol. Hence we have

$$K(T_{\lambda_\infty}) = 1\ 0\ 11\ 1010\ 10111011\ 1011101010111010\ldots$$

This property for the kneading invariant $K(T_{\lambda_\infty})$ comes from the fact that the map T_{λ_∞} is infinitely renormalizable, always with period 2.

The natural choice for the partition to use, to study the behaviour of the *AIC* for the orbits of T_{λ_∞}, would then be $Z = \{I_0, I_1\} = \{[0,c), [c,1]\}$. But first we study the problem with other partitions.

Let Z be a finite partition of $[0,1]$ into intervals, such that no end point belongs to the attractor Ω. In this case the conclusion is easy. Indeed, by Theorem 4.8 (6), it follows that m-almost any point $x \in [0,1]$ eventually enters any subset $J^{(n)}$. Moreover being the attractor Ω, given by $\Omega = \cap_i J^{(i)}$ (Theorem 4.8 (4)), a Cantor set of null m-measure, we have that there exists a n_0 such that for $n \geq n_0$ the sets $J^{(n)}$ do not contain any end point of the partition Z. These two facts, together with the property that the dynamics of T_{λ_∞} restricted to $J^{(n)}$ is a periodic permutation (Theorem 4.8 (3)), imply that eventually the symbolic orbit of m-almost any point $x \in [0,1]$ is periodic, with respect to such a partition Z. Hence for m-almost any point x, it holds $\overline{q}(x,Z) = 0$, and in particular $AIC(x,n,Z) = \log_2 n + const.$

More difficult is the situation for the partition $Z = \{[0,c), [c,1]\}$. We recall that, by Theorem 4.8 (6), m-almost any point $x \in [0,1]$ eventually enters any subset $J^{(n)}$, and that T_{λ_∞} is a cyclic permutation on the intervals $J_i^{(n)}$ making up $J^{(n)}$ (Theorem 4.8 (3)). Moreover if one orders the 2^n intervals of $J^{(n)}$ so that $c = \frac{1}{2} \in J_0^{(n)} = J^n$ for all n, then we can write $T_{\lambda_\infty}(J_i^{(n)}) = J_{i+1}^{(n)}$ for $i = 0, \ldots, 2^n - 2$, and $T_{\lambda_\infty}(J_{2^n-1}^{(n)}) = J_0^{(n+1)}$. This is implied by the fact that the end points of the intervals $J_i^{(n)}$ are determined by the forward orbit of the critical point $c = \frac{1}{2}$. Let now $N(k,x)$ be the first integer n such that $(T_{\lambda_\infty})^n(x) \in J^k$. We now prove the following

Theorem 4.12. *For the logistic map T_{λ_∞} at the chaos threshold, it holds $\overline{q}(x,Z) = 0$ for all $x \in [0,1]$, where the partition Z is given by $Z = \{[0,c), [c,1]\}$.*

Proof. The proof is based on the properties of renormalization of the system, and on the construction of the kneading invariant by the explained algorithm. The argument is iterative.

First of all by Theorem 4.8 (6), we have that a point $x \in [0,1]$ either lands after a finite number of steps on a periodic orbit or converges to the attractor Ω. Hence for the former points we have

$$AIC(x,n,Z) \leq \log_2 n + const$$

being eventually periodic points.

For the latter class of points, we first show

$$AIC(x,n,Z) \leq \log_2 n + \sum_{i=1}^{k} \frac{N(i+1,x) - N(i,x)}{2^i} + const \tag{4.23}$$

where k is such that $N(k,x) \leq n < N(k+1,x)$.

Let $N(1,x) \leq n < N(2,x)$. If ω denotes the symbolic sequence of a point x, we have that ω^n, the word given by the first n symbols of ω, can be reconstructed by the knowledge of $(\omega_0, \dots, \omega_{N(1,x)-1})$ and of $(\omega_{N(1,x)}, \dots, \omega_n)$ separately. But since $(T_{\lambda_\infty})^{N(1,x)}(x) \in J^1$, it holds $(T_{\lambda_\infty})^n(x) \in J^{(1)}$ for all $n \geq N(1,x)$. In particular $(T_{\lambda_\infty})^{N(1,x)+2h+1}(x) \in J^{(1)} \setminus J^1$ for all $h \in \mathbb{N}$, hence we have that $\omega_{N(1,x)+2h+1} = 1$ for all $h \in \mathbb{N}$. Indeed we have that the point x has to follow the iterates of the critical point c, when outside J^1. So its symbolic sequence is given by the kneading invariant. The only uncertainty is about the iterates of x that are in J^1, since we don't know what is its position relative to the critical point.

So for all n satisfying $N(1,x) \leq n < N(2,x)$, we have

$$\begin{aligned} AIC(x,n,Z) \leq \ & AIC(\omega, 0, N(1,x)-1) + \log_2 n + \\ & + AIC\left(\omega_1, N(1,x), \left[\tfrac{n-N(1,x)}{2}\right]\right) + const \end{aligned}$$

where by $AIC(\omega,p,q)$ we denote the information content of the word given by the symbols of the sequence ω from the p-th to the q-th, by ω_1 we denote the sequence

$$(\omega_{N(1,x)}, \omega_{N(1,x)+2}, \dots, \omega_{N(1,x)+2h}, \dots)$$

and let $[\cdot]$ be the integer part of a number. In this first case we obtain equation (4.23), since $k=1$, and

$$AIC\left(\omega_1, N(1,x), \left[\frac{n-N(1,x)}{2}\right]\right) \leq \frac{N(2,x)-N(1,x)}{2}$$

To obtain equation (4.23) in the general case, the iterative argument can be applied to the sequence ω_1, since it is nothing but the symbolic sequence of the point x by the map $(T_{\lambda_\infty})_1 := (T_{\lambda_\infty})^2|_{J^1}$.

Let n be such that $N(2,x) \leq n < N(3,x)$. Hence ω_1 can be written as $((\omega_1)_{N(1,x)}, \dots, (\omega_1)_{N(2,x)-2})$ plus $((\omega_1)_{N(2,x)}, \dots, (\omega_1)_n)$. Hence

$$AIC(\omega_1, N(1,x), n) \leq AIC(\omega_1, N(1,x), N(2,x)-2) + AIC(\omega_1, N(2,x), n)$$

But now

$$AIC(\omega_1, N(1,x), N(2,x) - 2) \leq \frac{N(2,x) - N(1,x)}{2}$$

and the iterative argument implies that

$$AIC(\omega_1, N(2,x), n) \leq \frac{N(3,x) - N(2,x)}{4}$$

being half of the symbols of ω_1 specified by the kneading invariant of the map.

At this point the iteration of this argument implies that for $N(k,x) \leq n < N(k+1,x)$ it holds equation (4.23).

To conclude, we notice that $N(i+1,x) - N(i,x) = 2^i$, since $T_{\lambda_\infty}(J^{(i)}_{2^i-1}) = J^{(i+1)}_0 = J^{i+1}$. Moreover if k is such that $N(k,x) \leq n < N(k+1,x)$ then

$$n - N(1,x) = \sum_{j=1}^{k-1} 2^j = 2^k - 3$$

hence $k = \log_2(n - N(1,x) + 3) \sim \log_2 n$. It follows that equation (4.23) can be rewritten as

$$AIC(x,n,Z) \leq \log_2 n + \sum_{i=1}^{k} 1 + const = \log_2 n + k + const \sim 2\log_2 n + const$$

hence it follows $\overline{q}(x,Z) = 0$ for all $x \in [0,1]$ that converge to the attractor Ω. $\square$

We have studied the behaviour of the information content of sequences of the logistic map T_{λ_∞} also numerically. In Chapter 6, Section 6.2.2, the results are shown. They confirm Theorem 4.12.

4.3 Intermittency

So far in this chapter we have studied the behaviour of the Algorithmic Information Content for dynamical systems with a finite invariant probability measure. In particular we have been able to give results for almost any point with respect to the Lebesgue measure, even if it was not the physical measure.

We now change our interest to infinite dynamical systems, for which some interesting results exist for the quantities introduced in Chapter 3. As specified in Section 3.3, we restrict our attention to systems with an infinite measure that is equivalent to the Lebesgue measure. Hence the Lebesgue measure will be the reference measure to compute the global chaos indexes.

The systems we study in this section are called *intermittent*, since their orbits present an alternation between long regular phases, called *laminar*, and short irregular phases, called *turbulent*.

We first study the family of *Manneville maps*, that are the basic examples of intermittent dynamical systems. The following section is based on [Bo1].

4.3.1 The family of Manneville maps

In [Mann], Manneville introduced a simple map of the interval $I = [0,1]$ into itself as an example of a discrete dissipative dynamical system with *intermittency*. This behaviour had been observed in fluid dynamics experiments and in chemical reactions, and the map introduced by Manneville was a simple model displaying this complicated behaviour. We consider the family of *Manneville maps* from the interval $I = [0,1]$ into itself, defined by

$$T_z(x) = x + x^z (\text{mod } 1) \tag{4.24}$$

with z being a real parameter varying on the half line $(1, +\infty)$. In Figure 4.1 it is plotted the Manneville map for $z = 2$. After their introduction, thanks to the simplicity of their form, the Manneville maps gave a strong impulse to the search for this intermittent behaviour in dynamical systems. In particular in the physics literature we can find many approaches to different systems using intermittency. Just to mention some examples, there are applications of the Manneville maps in dynamical approaches to DNA sequences ([Al1],[Al2]) and ion channels ([TL]), and in non-extensive thermodynamical problems ([BGP]).

The family of Manneville maps also attracted much attention from the mathematics community, being an example of a dynamical system with peculiar properties. In particular the properties that have been studied are the existence of invariant measures ([Th], [CF], [CI2]) and the decay of correlations ([LSV], [Yo], [Sa], [Is1], [Is2]).

The study of the Manneville maps using the Algorithmic Information Content has been suggested by Gaspard and Wang ([GW1]).

The behaviour of the Manneville maps drastically changes as the parameter z crosses the value $z = 2$. Indeed for $z < 2$, the dynamical system $([0,1], T_z)$ has a physical measure μ, which is ergodic and equivalent to the Lebesgue measure. Moreover the KS entropy $h_\mu(T_z)$ is positive, hence we have that $AIC(x, n, Z) \sim h_\mu(T_z, Z)n$ for μ-almost any point $x \in [0,1]$, for a partition Z with positive entropy. Instead for $z \geq 2$, the only physical measure of the system $([0,1], T_z)$ is the Dirac delta at the origin (an *indifferent fixed point* for the map), and all the probability invariant measures have null KS entropy. Hence the systems $([0,1], T_z)$ for $z \geq 2$ are weakly chaotic. Moreover they have an infinite invariant measure μ, which is equivalent to the Lebesgue measure and make the system conservative and ergodic. Hence we will now restrict our attention to the case $z \geq 2$.

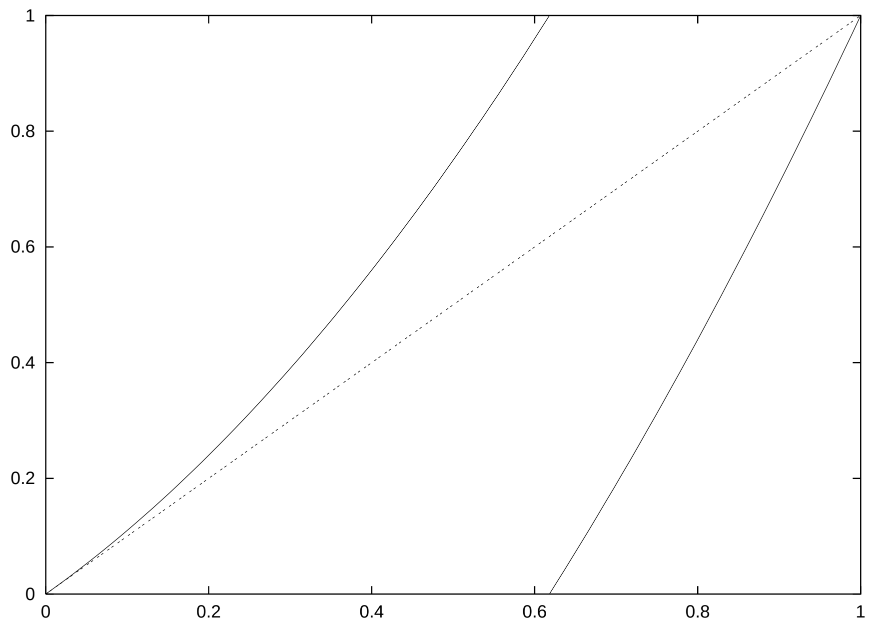

Figure 4.1: *The Manneville map* T_2

Gaspard and Wang ([GW1]) showed that the Manneville maps with $z \geq 2$ exhibit a behaviour that they called *sporadicity*. Namely, the Algorithmic Information Content of a word n symbols long, behaves in mean like $n^{\alpha}(\log(n))^{\beta}$, with either $0 < \alpha < 1$ or $\alpha = 1$ and $\beta < 0$.

We now give a formal proof of the results obtained by Gaspard and Wang, giving more precise estimates for the AIC of a word generated by the map. But the most important generalization is that we find our results for the Manneville maps as a particular case of a general theorem concerning a large family of maps L, defined in equation (4.26). This family of maps exhibits an extremely wide range of behaviours (for the AIC of the generated sequences), and sporadicity is only one possible case. Then we find a family of maps that can be classified with respect to the *global chaos index* (see Definition 3.5) of the maps.

In the following, we study a family of piecewise linear maps L on the interval $X = [0, 1]$, which are topologically equivalent to the Manneville maps T_z. Using the maps L, all the theorems have an easier interpretation and computations can be done exactly. Moreover all our results are extendible through metric isomorphism, hence we shall define on the interval X two different measures, that make the topological equivalence between L and T_z a metric isomorphism. Then we can extend all the results that we find for the maps L to the Manneville maps T_z.

Let's start defining the piecewise linear maps L that we consider. We use here the same approach as in [Ga3]. A natural way to get a partition of the interval $X = [0,1]$ from the Manneville maps T_z is the following: let's call x_0 the point of X such that $T_z(x_0) = 1$ with $x_0 \neq 0,1$, and x_1 the preimage of x_0 in the interval $[0, x_0]$; then we define recursively $x_n = \{T_z^{-1}(x_{n-1})\} \cap [0, x_{n-1}]$. Then the sub-intervals $B_k = (x_k, x_{k-1}]$, for $k \geq 1$, and $B_0 = (x_0, 1]$ are a partition of X.

Define $\{\epsilon_k\}_{k\in\mathbb{N}}$ a sequence of positive real numbers, that is strictly monotonically decreasing and converging towards zero, with the property that

$$\frac{\epsilon_{k-1} - \epsilon_k}{\epsilon_{k-2} - \epsilon_{k-1}} < 1 \quad \forall\ k \in \mathbb{N} \tag{4.25}$$

The piecewise linear maps that we consider are defined by

$$L(x) = \begin{cases} \frac{\epsilon_{k-2}-\epsilon_{k-1}}{\epsilon_{k-1}-\epsilon_k}(x - \epsilon_k) + \epsilon_{k-1} & \epsilon_k < x \leq \epsilon_{k-1}, \quad k \geq 1 \\ \frac{x-\epsilon_0}{1-\epsilon_0} & \epsilon_0 < x \leq 1 \\ 0 & x = 0 \end{cases} \tag{4.26}$$

where we define $\epsilon_{-1} = 1$. These piecewise linear maps L clearly depend on the definition of the sequence $\{\epsilon_k\}_{k\in\mathbb{N}}$, but a particular choice for this sequence is important only later. For the moment we consider any possible sequence, with the properties specified above. Let's define sub-intervals $A_i = (\epsilon_i, \epsilon_{i-1}]$, and $A_0 = (\epsilon_0, 1]$. These intervals form a partition of the interval $X = [0,1]$. We prove the following

Theorem 4.13 ([Bo1]). *Any piecewise linear map L defined as in equation (4.26) is topologically equivalent to the Manneville maps T_z.*

Proof. We have to find a homeomorphism $h : X \to X$ such that $h(T_z(x)) = L(h(x))$ for each $x \in X$. To find such a homeomorphism we use the partitions (B_j) and (A_j). We define $h(x_n) = \epsilon_n$ for each $n \in \mathbb{N}$ and $h(0) = 0$, $h(1) = 1$, and such that $h(B_j) = A_j$, for all $j \geq 0$. To define the homeomorphism h we use a dense set of X, define h on this set and, then, simply extends the definition of h to the whole interval X by continuity.

Let's consider a sub-interval B_k. By definition of the Manneville maps T_z, we have that $T_z(B_k) = B_{k-1}$, for $k \geq 1$, and $T_z(B_0) = I$. Then it follows that $T_z^k(B_k) = B_0$ and $T_z^{k+1}(B_k) = I$. So, within each B_k we can find sub-intervals B_{kj}, with $k, j \in \mathbb{N}$, defined by $T_z^{k+1}(B_{kj}) = B_j$. These sub-intervals form a partition of each B_k. We can continue this partition of the intervals B_k, defining, by the same rule, sub-intervals B_{kji} that form a partition of B_{kj}. We write then any sub-interval of the form $B_{k_1k_2\ldots k_n}$, with $k_i \in \mathbb{N}$ for each $i = 1, \ldots, n$, as $B_{k_1k_2\ldots k_n} = (x_{k_1k_2\ldots k_n}, x_{k_1k_2\ldots(k_n-1)}]$. The set $\{x_{k_1k_2\ldots k_n},\ n \in \mathbb{N},\ k_i \in \mathbb{N}\}$ is a countable dense set of the interval I. Analogously, we can define a set of points $\{\epsilon_{k_1k_2\ldots k_n},\ n \in \mathbb{N},\ k_i \in \mathbb{N}\}$, for the map L, with the same property as $x_{k_1k_2\ldots k_n}$. We define then $h(x_{k_1k_2\ldots k_n}) = \epsilon_{k_1k_2\ldots k_n}$ for each $n \in \mathbb{N}$, and extend the function h continuously to the whole interval X. We have thus obtained a continuous function h such that $h(B_{k_1k_2\ldots k_n}) = A_{k_1k_2\ldots k_n}$, for each $n \in \mathbb{N}$, where the sub-intervals $A_{k_1k_2\ldots k_n}$ are defined as $A_{k_1k_2\ldots k_n} = (\epsilon_{k_1k_2\ldots k_n}, \epsilon_{k_1k_2\ldots(k_n-1)}]$.

The injectivity of h follows by contradiction. Let's suppose to have two points $x < y \in X$ such that $h(x) = h(y) = z$. By the density of the set $\{x_{k_1k_2\ldots k_n},\ n \in \mathbb{N},\ k_i \in \mathbb{N}\}$, we can find a point $\bar{x}_{k_1k_2\ldots k_{\bar{n}}} \in (x, y)$. This implies that there exists a $\bar{n}$ such that $x \in B_{k_1k_2\ldots(k_{\bar{n}}-1)}$ and $y \in B_{k_1k_2\ldots k_{\bar{n}}}$. Then we have $h(x) \neq h(y)$. The subjectivity of h follows immediately by the definition.

Then the inverse function h^{-1} exists and h is a homeomorphism because it is a continuous invertible function from a compact to a Hausdorff space. $\square$

The topological equivalence between L and T_z can be used in particular to obtain a metric isomorphism. If we have a measure ν on the system $(X, \mathcal{B}, L)$, where $\mathcal{B}$ is the Borel σ-algebra, then the homeomorphism h carries ν into another measure $\mu = h^*\nu$ on $(X, \mathcal{B}, T_z)$, and with respect to these measures h is a metric isomorphism.

To study the behaviour of the system $(X, \mathcal{B}, L)$, we use the symbolic representation of this system with respect to the partition $Z = \{A_i\}_{i\in\mathbb{N}}$ used to define the map L (see equation (4.26)). Hence we obtain an information source on an infinite alphabet (see Chapter 2). In particular for the system $(X, \mathcal{B}, L)$ we have that the function $\varphi_Z : [0,1] \to \Omega = \mathbb{N}^{\mathbb{N}}$ (see equation (1.7)) is a conjugacy between the dynamical system $(X, \mathcal{B}, L)$ and $(\Omega_M, \mathcal{C}, \tau_M)$, where Ω_M is the closed subset of Ω given by the sequences ω that respect the restrictions given by the so-called *transition matrix* M. The matrix M is a $\mathbb{N} \times \mathbb{N}$ matrix with entries $m_{ij} \in \{0, 1\}$, and $\omega \in \Omega_M$ if and only if for all $n \in \mathbb{N}$ and for all $i, j \in \mathbb{N}$ it holds

$$\omega_n\ \omega_{n+1} = i\ j \iff m_{ij} = 1$$

The map τ_M is the restriction of the shift map to the invariant subset Ω_M. For the system $(X, \mathcal{B}, L)$ the transition matrix M is given by

$$M = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 & 1 & \cdots \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}. \tag{4.27}$$

The study of the *AIC* for the symbolic orbits generated by the system $(X, \mathcal{B}, L)$ needs some results from the theory of stochastic processes, and in particular from the theory of *Markov chains*. We now recall some basic definitions and results introduced by Parry in [Pa].

Definition 4.14. A *non-atomic stochastic process* is $(\Sigma_M, \mathcal{C}, \tau_M, \mu)$ where $\Sigma_M = \{\omega = (\omega_0, \omega_1, \dots) \ / \ \omega_n \in \mathbb{N} \ \forall n \in \mathbb{N}, \quad \omega_n\omega_{n+1} = ij \ \Leftrightarrow \ m_{ij} = 1 \ \forall n \in \mathbb{N} \ \forall \ i, j \in \mathbb{N}\}$, $\mathcal{C}$ is the σ-algebra generated by the cylinders, μ is a non-atomic probability measure on $\mathcal{C}$, and τ_M is the forward shift on X. In the theory of stochastic processes the transition matrix M on Σ_M is called a *structure matrix.*

Definition 4.15. A stochastic process is called *transitive of order* k if for all $(\omega_1, \dots, \omega_k)$ and $(\omega'_1, \dots, \omega'_k)$ with $\mu(\omega_1, \dots, \omega_k) > 0$ and $\mu(\omega'_1, \dots, \omega'_k) > 0$, there exists a finite $(\psi_1, \dots, \psi_n)$ such that

$$\mu(\omega_1, \dots, \omega_k; \psi_1, \dots, \psi_n; \omega'_1, \dots, \omega'_k) > 0.$$

When $k = 1$, a stochastic process is simply called *transitive.*

Definition 4.16. A stochastic process is said to be *intrinsically Markovian of order* k if $\mu(\omega_1, \dots, \omega_n) > 0$ and $\mu(\omega_{n-k+1}, \dots, \omega_{n+1}) > 0$ imply

$$\mu(\omega_1, \dots, \omega_n, \omega_{n+1}) > 0.$$

When $k = 1$ it is simply called *intrinsically Markovian.*

Given a probability measure ν on the system $(X, \mathcal{B}, L)$, where we remark that we haven't supposed ν to be L-invariant, there is a metric isomorphism with the system $(\Omega_M, \mathcal{C}, \tau_M)$ with probability measure P induced by the conjugacy φ_Z. Then we can easily prove

Proposition 4.17. *The space* $(\Omega_M, \mathcal{C}, \tau_M, P)$ *is a non-atomic stochastic process, which is transitive and intrinsically Markovian.*

Definition 4.18. Given a stochastic process $(\Sigma_M, \mathcal{C}, \tau_M, \mu)$, a measure p makes the process $(\Sigma_M, \mathcal{C}, \tau_M, p)$ *compatible* with the original when $p(C^n_r) > 0$ if and only if $\mu(C^n_r) > 0$, for any cylinder C^n_r.

At this point we use the notion of non-atomic stochastic processes to obtain a compatibility between the maps L and a Markov chain, through the symbolic dynamical system on countable symbols. Before giving the theorems in this direction, we briefly recall the theory of Markov chains (see [Chu]).

Definition 4.19. Given a probability space $(\Lambda, \mathcal{F}, \bar{P})$ and a countable space W with the discrete σ-algebra, a *Markov chain* is a sequence $(Z_n)_{n\in\mathbb{N}}$ of random variables $Z_n : \Lambda \to W$ such that

i) If, given $y_0, \dots, y_{n+1} \in W$, we have $\bar{P}[Z_n = y_n, Z_{n-1} = y_{n-1}, \dots, Z_0 = y_0] > 0$, then

$$\bar{P}[Z_{n+1} = y_{n+1} \mid Z_n = y_n, \dots, Z_0 = y_0] = \bar{P}[Z_{n+1} = y_{n+1} \mid Z_n = y_n],$$

ii) If $x, y \in W$ and $m, n \in \mathbb{N}$ are such that $\bar{P}[Z_m = x] > 0$ and $\bar{P}[Z_n = x] > 0$, then

$$\bar{P}[Z_{m+1} = y \mid Z_m = x] = \bar{P}[Z_{n+1} = y \mid Z_n = x].$$

In particular the numbers $p(x, y) = \bar{P}[Z_{n+1} = y \mid Z_n = x]$ form a matrix $\Pi = (p(x, y))_{x,y \in W}$, called the *transition matrix.* Moreover the probability measure on W defined by $\nu(y) = \bar{P}[Z_0 = y]$ is called the *initial distribution.*

The transition matrix Π is a *stochastic matrix*, that is $p(x, y) \geq 0$ and $\sum_{y \in W} p(x, y) = 1$ for all $x \in W$.

Theorem 4.20 ([Chu]). *Given any countable space W, a transition matrix Π and an initial distribution ν, it is possible to construct a probability space $(\Lambda, \mathcal{F}, \bar{P})$ and a sequence of random variables $Z_n : \Lambda \to W$, such that the constructed Markov chain has Π as transition matrix and ν as initial distribution.*

Proof. For the proof of the theorem see Chung [Chu]. We simply say what is the constructed probability space $(\Lambda, \mathcal{F}, \bar{P})$. The space Λ is $W^{\mathbb{N}}$ and is called the *realizations space*, the σ-algebra $\mathcal{F}$ is given by the cylinders, and the probability $\bar{P}$ is defined on the cylinders by

$$\bar{P}(C_r^n) = \nu(r_0)p(r_0, r_1)p(r_1, r_2) \ldots p(r_{n-1}r_n).$$

The random variables are defined as the projections of Λ on W. □

Theorem 4.21. *Given any intrinsically Markovian, transitive stochastic process $(\Sigma_M, \mathcal{C}, \tau_M, \mu)$ and a stochastic matrix Π such that $p(i, j) > 0$ if and only if $m_{ij} = 1$, for the structure matrix M of the process, there is a probability p on Σ_M such that $(\Sigma_M, \mathcal{C}, \tau_M, p)$ is compatible with $(\Sigma_M, \mathcal{C}, \tau_M, \mu)$, and it is a Markov chain with Π as transition matrix.*

Proof. We have just to apply Theorem 4.20 to the matrix Π and to an initial distribution ν, being Σ_M already in the form of the realizations space. The probability p is then the probability $\bar{P}$ defined as above. □

Corollary 4.22. *Our space $(\Omega_M, \mathcal{C}, \tau_M, P)$ is compatible with a Markov chain.*

Proof. The corollary is proved thanks to Theorem 4.21 and Proposition 4.17. We have just to choose a stochastic matrix that satisfies the hypothesis, that is $p(i, j) > 0$ if and only if $m_{ij} = 1$ for the structure matrix defined as in equation (4.27). The measure P can be used as initial distribution. □

We have thus completed our equivalence, in the sense of Corollary 4.22, between the maps L defined on the space $(X, \mathcal{B}, \nu)$, for any probability measure ν, and a Markov chain, that is denoted simply by a stochastic matrix Π and an initial distribution that we can choose to be ν itself.

Now, we want to relate the dynamics of the Markov chain with our dynamical system. A natural choice for the stochastic matrix Π is

$$\Pi = \begin{pmatrix} p_0 & p_1 & p_2 & \cdots & p_n & p_{n+1} & p_{n+2} & \cdots \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{4.28}$$

where p_i are the probabilities of transition from the sub-interval A_0 to the sub-interval A_i, and in terms of the measure ν are given by

$$p_i = \frac{\nu(A_0 \cap L^{-1}(A_i))}{\nu(A_0)} \tag{4.29}$$

It is evident, in the definition of Π, the dependence on the particular sequence (ϵ_k) that we consider to define the map L, and on the probability measure ν.

It has been shown by Gaspard and Wang ([GW1]), that a way to estimate the *AIC* of a word obtained from our dynamical system is the theory of *recurrent events* for Markov chains ([Fel2],[Fel1]). From the theory of Markov chains, we have that our stochastic matrix Π is *irreducible*, and that the state A_0 is *persistent*, in the sense that

$$p[Z_m = 0 \text{ for some } m > n | Z_n = 0] = 1 \quad \forall\ n$$

where p is the probability measure on $(\Omega_M, \mathcal{C}, \tau_M)$ that makes it a Markov chain, with (Z_n) as random variables from Ω_M to $\mathbb{N}$ (see Theorems 4.20, 4.21).

If we consider as recurrent event $\mathcal{E}$ the passage from the sub-interval A_0, we have that $\mathcal{E}$ is *certain* and that the *mean recurrence time* m_0 is given by

$$m_0 = \sum_{k=1}^{+\infty} k\ p_{k-1} \tag{4.30}$$

If $m_0 = +\infty$ the state A_0 is called *null*, otherwise it is *ergodic*. Thanks to the irreducibility of the stochastic matrix Π, we have that all the states A_i are of the same kind of A_0, and then either $m_i = +\infty$ for all $i \in \mathbb{N}$ or $m_i < +\infty$ for all $i \in \mathbb{N}$.

For the recurrent event $\mathcal{E}$ two random variables can be introduced: $X_k : \Omega_M \to \mathbb{N}$ given by 1 plus the number of trials between the $(k-1)$-th and k-th occurrence of $\mathcal{E}$; $N_k : \Omega_M \to \mathbb{N}$ given by the number of realizations of $\mathcal{E}$ in k trials. The random variables X_k have all the same probability distribution given by

$$p[X_k = r] = p_{r-1},$$

and their mean $E_p[X_k] = m_0$, where the subscript p specifies the measure we use to find the mean. For our problem it will be very important also the form of the distribution function $F(x) = \sum_{r=0}^{[x]} p_r$ of the X_k. Finally, we consider also the probabilities u_n that $\mathcal{E}$ occurs at the n-th trial. It holds that

$$\lim_{n\to+\infty} u_n = \frac{1}{m_0} \tag{4.31}$$

Let's now explain the link between the AIC of a word generated by the map L and the theory of recurrent events ([GW1]). Given an initial point $x \in [0,1]$ we obtain a sequence $\omega \in \Omega_M$ such that $L^k(x) \in A_{\omega_k}$ for all $k \in \mathbb{N}$. The sequence ω is, for example, of the form

$$\omega = (76543210543210021032 10\dots)$$

One possible way to give an estimate for the AIC of the sequence is to consider a compression of the word, and study the binary length of the compressed word. That is we assume to have a compression algorithm $\mathcal{Z}$, and we consider the information function $I_{\mathcal{Z}} : \mathcal{A}^* \to \mathbb{N}$ given by $I_{\mathcal{Z}}(s) = |\mathcal{Z}(s)|$, where $\mathcal{Z}(s)$ denotes the output of the compression algorithm on the word s and $|\cdot|$ denotes the binary length of a word. The output $\mathcal{Z}(s)$ is a word in $\{0,1\}^*$ from which it is possible to reconstruct the word $s \in \mathcal{A}^*$. We study compression algorithms and the approximation of the AIC by compression in Chapter 6.

One possible compression of the sequence ω is given by

$$S = (75023\dots)$$

that is the sequence of recurrence times for $\mathcal{E}$. So to the word ω^n, obtained by the first n symbols of ω, we associate the word

$$S^{N_n} = (\omega_{q_1}\omega_{q_2}\dots\omega_{q_{N_n}}) \tag{4.32}$$

with $\omega_{q_i - 1} = 0$ for all i, where N_n is the number of realizations of $\mathcal{E}$. Then to have an idea of the behaviour of the AIC of a sequence, we consider the information content of the words S^{N_n}, hence we have to estimate the behaviour of the random variables N_n.

In [Fel1], some possible behaviours for $E_p[N_n]$ have been studied, for particular forms of the distribution function $F(x)$. In particular

Theorem 4.23 (Feller). *If the recurrence time of $\mathcal{E}$ has finite mean m_0 and variance V, then*

$$E_p[N_n] \sim \frac{n}{m_0} + \frac{V - m_0 + m_0^2}{2m_0^2}.$$

If, instead $V = +\infty$, and the distribution function $F(x)$ satisfies

$$F(x) \sim 1 - Ax^{-\alpha}$$

with a constant A and $0 < \alpha < 2$, then:
i) If $1 < \alpha < 2$,

$$E_p[N_n] \sim \frac{n}{m_0} + \frac{A}{(\alpha-1)(2-\alpha)m_0^2}\, n^{(2-\alpha)};$$

ii) If $0 < \alpha < 1$,

$$E_p[N_n] \sim \frac{\sin \alpha\pi}{A\alpha\pi} n^{\alpha}.$$

If the variance V of the recurrence time is infinite but the distribution function has a form different from that studied in Theorem 4.23, then we can show that

Theorem 4.24 ([Bo1]). *If the state A_0 is ergodic then $E_p[N_n] \sim kn$, if instead A_0 is a null state then $E_p[N_n]$ is an infinite of order less than n.*

Proof. The proof is based on a characterization of the mean $E_p[N_n]$. In [Fel1], it is shown that $E_p[N_n] = U_n - 1$, where

$$U_n = \sum_{i=0}^{n} u_i$$

Then it follows that

$$\lim_{n\to+\infty} \frac{E_p[N_n]}{n} = \lim_{n\to+\infty} \frac{U_n}{n} = \lim_{n\to+\infty} u_n = \frac{1}{m_0}$$

Then, if $m_0 < +\infty$, that is A_0 is ergodic, then $E_p[N_n]$ is linear on n. Whereas if $m_0 = +\infty$, that is A_0 is a null state, then $E_p[N_n]$ is an infinite of order less than n. □

The information content I, calculated with the compression we have chosen, can be linked to the random variables N_n by the following theorem

Theorem 4.25 ([Bo1]). *For any sequence $\omega \in \Omega_M$ it holds*

$$(N_n - 1) + \log_2(n - N_n + 2) \le I(\omega^n) \le N_n \log_2 \left(\frac{n + N_n}{N_n}\right) \tag{4.33}$$

Proof. Given $n \in \mathbb{N}$, we have that $\Omega_M = \cup C_r^n$, where the union is made on all the possible cylinders C_r^n, with $r \in \mathbb{N}^n$. Moreover we write

$$I(\omega^n) = \sum_{i=1}^{N_n} \log_2(\omega_{q_i} + 2)$$

for the information content, where $\omega_{q_i} + 2$ is used instead of ω_{q_i}, to have $\log_2(\omega_{q_i} + 2) \geq 1$ for all i.

First of all, we consider only the cylinders C_r^n with $r_n = 0$. This is done because we want to study the sequences whose compression changes when we increase our given n. Indeed, if the compression wouldn't change, we wouldn't have any hint on the behaviour of the information content with respect to the length of the sequence.

We start with some special cases. Let's consider first the case $N_n = n$. The only possible cylinder is then C_r^n with $r = (0, \ldots, 0)$. Then our compression doesn't change any word in this cylinder, and $I(\omega^n) = n$ for any word.

In the case $N_n = (n-1)$, the possible cylinders are given by $r \in \mathbb{N}^n$ with only one symbol 1 and all the others 0. For all the words in these cylinders, the compression is $(n-1)$ symbols long, and $I(\omega^n) = (n-2) + \log_2 3$.

In the case $N_n = 1$, the only possible cylinder is given by $r_i = (n - i)$, and for the words in this cylinder $I(\omega^n) = \log_2(n+1)$.

Let now be in general $N_n = n - h$, for some $h < n$. The compression of such words is then N_n-symbols long. Moreover the compression is such that $\sum_{i=1}^{N_n} \omega_{q_i} = h$. We now want to find the maximum and the minimum of the function

$$\sum_{i=1}^{N_n} \log_2(\omega_{q_i} + 2)$$

with the condition $\sum_{i=1}^{N_n} \omega_{q_i} = k$. The maximum is attained for equal $\omega_{q_i} \neq 0$, and the minimum for all the $\omega_{q_i} = 0$ but one which is equal to h. Then the maximum is given by $\omega_{q_i} = \frac{h}{n-h}$ for all i, and the information content for the words in such a cylinder is given by $\sum_{i=1}^{n-h} \log_2 \left(\frac{h}{n-h} + 2 \right) = N_n \log_2 \left(\frac{n+N_n}{N_n} \right)$, and the minimum is given by $(n - h - 1) + \log_2(h+2) = (N_n - 1) + \log_2(n - N_n + 2)$.

Looking back at the special cases we studied before, we see that for $N_n = n$ and $N_n = 1$, the maximum and the minimum are the same, and give exactly the value of the information content we found building the sequences. For the case $N_n = (n-1)$, the only possible value for the information content is equal to the minimum we found. This shows that the maximum is not attained always, but there are cases in which it is attained. For example let

$n = 8$, $N_n = 4$, and consider the word $\omega = (10101010)$. Its compression is then given by $S = (1111)$, and $I(\omega^n) = 4\log_2 3 = N_n \log_2\left(\frac{n+N_n}{N_n}\right)$.

Finally we remark that we have tacitly assumed that our words do not begin with the symbol 0. If it happened, the estimates wouldn't change significantly.

We have thus proved that for a subset of Ω_M of full measure with respect to all the cylinders with 0 as last symbol, the information content of a word can be estimated using the value of the random variables N_n. □

Our plan is the following: given a probability measure ν on $(X, \mathcal{B}, L)$, we find the stochastic matrix Π, equation (4.28), the distribution function $F(x)$, and, the mean $E_p[N_n]$. Then we can link the mean $E_\nu[I(\omega^n)]$, where ν is the probability measure on $(X, \mathcal{B})$, with the mean $E_p[N_n]$, by Theorem 4.25.

Another important aspect is the existence of an invariant measure for the Markov chain associated to our dynamical system. Given a stochastic matrix Π of the form of equation (4.28) we have

Theorem 4.26 ([Chu]). *There is a measure $\bar{p}$ on the space $(\Omega_M, \mathcal{C}, \tau_M)$, invariant for the stochastic matrix Π, defined by*

$$\bar{p}(k) = \sum_{n=0}^{+\infty} p_{k+n}.$$

This measure is a probability measure if and only if the mean recurrence time m_0 is finite.

This theorem allows us to induce on $(X, \mathcal{B}, L)$ an L-invariant measure $\bar{\mu}$, which is a probability measure if and only if A_0 is ergodic.

So far, we have obtained results for the behaviour of the information content of the symbolic sequences generated by the intermittent dynamical system $(X, \mathcal{B}, L)$. But the results depend on the initial distribution ν of the Markov chain associated to the system, and on the sequence $\{\epsilon_k\}$ we used to define the map L (see equation (4.26)). Moreover we have used the information content I defined by the compression given in equation (4.32). We now look for results for the AIC.

When we apply the notion of AIC to a sequence obtained from a dynamical system, the choice of this sequence depends on the choice of the initial point x that we use to generate the orbit of the dynamical system. This choice can be made randomly, and the most natural way to introduce a probability distribution on the choice of the initial point is by using the Lebesgue measure m on the space. Hence we apply the previous results to the system $(X, \mathcal{B}, L, m)$. Using the Lebesgue measure m, thanks to the piecewise linearity of the map L, the probabilities of transition p_i given in equation (4.29) assume a particular simple form. Indeed we find that $p_i = m(A_i)$ for all $i \in \mathbb{N}$.

With respect to the Lebesgue measure m, it is also possible to prove that the compression we have introduced for words given by any map L is the best possible. Hence the asymptotic behaviour of the AIC of a symbolic sequence is the same as that of the information content I.

Theorem 4.27 ([Bo1]). *Given any sequence $\{\epsilon_k\}$ of the form (4.25), the best compression for the symbolic words*[1] *generated by the dynamical system $(X, \mathcal{B}, L, m)$, where m is the Lebesgue measure, is the compression given in equation (4.32), for m-almost any initial condition. Namely there exists a constant $c > 0$ such that $AIC(x, n, Z) = cI(x, n, Z)$, for m-almost all $x \in X$.*

Proof. The compression we introduced gives a bijective relation between our space $(\Omega_M, \mathcal{C}, \tau_M, p)$ and the space of all possible infinite sequences built on countable symbols, without any restriction given by a transition matrix. We denote this space as Ω. We introduce on Ω the σ-algebra of the cylinders $C_r'^n$, and a probability measure p' inherited in some way from p. We define p' by

$$p'(C_r'^n) = p(C_R^N) \tag{4.34}$$

where the cylinder C_R^N is built in such a way that compression of words that belong to it gives words belonging to the cylinder $C_r'^n$. At this point we ask if it is possible to compress any more words belonging to the space $(\Omega, \mathcal{C}', p')$. But, if on this space we consider the usual shift map, we find a p'-invariant map with positive KS entropy $h_{KS}(\tau, Z)$. This is given by a direct computation, using the piecewise linearity of the map L that gives a simple form for the measure p'. This gives the thesis with $c = h_{KS}(\tau, Z)$. $\square$

The proof of the previous theorem is based on the properties of the *induced map* $g : [0, 1] \to [0, 1]$ that can be defined from the map L by

$$g(x) = L^k(x) \quad \text{if } x \in A^{k-1}$$

The properties of this map have been extensively studied (see [CI1], [CI2], [CI3], [CF], [CGS], [Is1], [Is2], [LSV], [Th], [Yo]).

The second point is the choice of the sequence $\{\epsilon_k\}$. We know that for any sequence, we can find an isomorphism of the map L with the Manneville maps T_z of equation (4.24). At this point we present some cases for the choice of $\{\epsilon_k\}$, and then study a particular choice.

[1] We remark that we always consider the partition $Z = \{A_i\}_{i \in \mathbb{N}}$, where the intervals A_i are defined by $A_i = (\epsilon_i, \epsilon_{i-1}]$ (see equation (4.26)).

Example 4.28. Let the sequence be given by $\epsilon_k = \frac{1}{k^\alpha}$ with $\alpha > 0$. This sequence has all the properties we need for the definition of the map L. Then

$$p_i = m(A_i) = \frac{1}{(i-1)^\alpha} - \frac{1}{i^\alpha} \sim \frac{1}{i^{\alpha+1}}.$$

The mean recurrence time m_0 is given by

$$m_0 = \sum_{r=1}^{+\infty} r\ p_{r-1}$$

and it is finite if and only if $\alpha > 1$. Then we can find an invariant measure $\bar{\mu}$ for the system $(X, \mathcal{B}, L)$, such that $\bar{\mu}(A_i) \sim \frac{1}{i^\alpha}$. This measure $\bar{\mu}$ is a probability measure if and only if $\alpha > 1$.

The variance V of the recurrence time is given by

$$V = \sum_{r=1}^{+\infty} r^2\ p_{r-1}$$

and then $V < +\infty$ if and only if $\alpha > 2$. If we compute the distribution function $F(x)$, we have that

$$F(x) \sim 1 - Ax^{-\alpha},$$

then we can apply Theorems 4.23 and 4.25, and obtain that:

i) if $\alpha > 1$ then $E_m[AIC(x,n,Z)] \sim E_p[N_n] \sim n$, hence for the global chaos indexes (see Definition 3.5) it holds $\overline{q}(X,m) = \underline{q}(X,m) = 1$;

ii) if $\alpha < 1$ then $E_p[N_n] \sim n^\alpha$ and $n^\alpha \leq E_m[AIC(x,n,Z)] \leq n^\alpha \log_2 n$ (see Theorem 4.25), and $\overline{q}(X,m) = \underline{q}(X,m) = \alpha$.

Example 4.29. Let now the sequence $\{\epsilon_k\}$ be given by $\epsilon_k = \frac{1}{a^k}$, where $a \in \mathbb{N}$ and $a > 1$. In this case it is easy to verify that $p_i \sim \frac{1}{a^i}$. Then the mean recurrence time m_0 and the variance V of the recurrence time are always finite, and we can deduce that $E_m[AIC(x,n,Z)] \sim E_p[N_n] \sim n$. This implies that $\overline{q}(X,m) = \underline{q}(X,m) = 1$. The invariant probability measure $\bar{\mu}$ exists and is given by $\bar{\mu}(A_i) \sim \frac{1}{a^i}$.

Example 4.30. Finally let's consider the case in which the sequence $\{\epsilon_k\}$ is given by $\epsilon_k = \frac{1}{k^\alpha (\log k)^\beta}$ with either $\alpha = 1$ and $\beta > 1$ or $\alpha > 1$. If we compute the variance V of the recurrence time, we obtain that it is finite if and only if either $\alpha > 3$ or $\alpha = 3$ and $\beta > 1$. In this case we obtain from Theorem 4.23 that $E_m[AIC(x,n,Z)] \sim n$, hence $\overline{q}(X,m) = \underline{q}(X,m) = 1$. But if $V = +\infty$, we cannot find an explicit form for the distribution function $F(x)$ similar to that of Theorem 4.23, but we can use Theorem 4.24. Indeed we have that the state A_0 is ergodic if and only if either $3 > \alpha > 2$ or $\alpha = 2$ and $\beta > 1$, in which case $E_m[AIC(x,n,Z)] \sim n$. For other values of α and β, we know that $E_p[N_n]$ is an infinite of order less than n, and for $E_m[AIC(x,n,Z)]$ we can apply Theorem 4.25 to obtain an estimate for its order of infinite.

Example 4.31. In a particular case, we can say something more about the order of infinite of $E_p[N_n]$ using the theory of recurrent events ([Fel1]) and of power series ([Ti]). Indeed, choosing the sequence $\epsilon_k \sim \frac{1}{\log k}$, we have that the distribution function $F(x) \sim \left(1 - \frac{1}{\log x}\right)$. Then from the characterization of $E_p[N_n]$ in terms of the generating function of the random variables X_k, we obtain that asymptotically $E_p[N_n] < n^\alpha$ for all $\alpha > 0$. Hence, from Theorem 4.25, $E_m[AIC(x,n,Z)]$ is an infinite of order smaller than any power law. This implies that $\overline{q}(X,m) = \underline{q}(X,m) = 0$.

We have thus found that changing the sequence $\{\epsilon_k\}$ it is possible to obtain all a set of behaviours for the global chaos index of the system, using as reference measure the Lebesgue measure. It is clear that the different behaviours are induced by the rate with which the derivative of the map L increases, rate that depends on the order with which $\{\epsilon_k\}$ converges to 0. It is then evident that this must be also the criterion to distinguish between different behaviours of the information function of the Manneville maps T_z, for different values of the parameter z in equation (4.24). This is the degree of *non-hyperbolicity* of the map T_z in $x = 0$. We have then to find a way to associate to a particular value of z a given sequence $\{\epsilon_k\}$. Since we have to maintain a given rate of increasing of the derivative, given a value for z, we look for the sequence $\{\epsilon_k\}$ such that $\epsilon_k \sim x_k$, where x_k is the sequence of preimages of the point x_0.

We have that $x_k \sim \frac{1}{k^\alpha}$ with $\alpha = \frac{1}{z-1}$. We are then in the first case we studied, and $\alpha > 0$ being $z > 1$. We can apply all the results we found, in particular:

- if $z < 2$ it holds $E_m[AIC(x,n,Z)] \sim n$, $\overline{q}(X,m) = \underline{q}(X,m) = 1$, and there is an L-invariant probability measure $\bar{\mu}$ such that $\bar{\mu}(A_i) \sim \frac{1}{i^\alpha}$;
- if $z > 2$ it holds $n^\alpha \leq E_m[AIC(x,n,Z)] \leq n^\alpha \log_2 n$, $\overline{q}(X,m) = \underline{q}(X,m) = \alpha = \frac{1}{z-1}$, and the invariant measure $\bar{\mu}$ is not a probability measure.

We have then found as a particular case the same results of [GW1]. But, we would like to have a behaviour of the AIC valid for almost any orbit with respect to the Lebesgue measure m. We have

Theorem 4.32 ([Bo1]). *For almost any point $x \in X$ with respect to the Lebesgue measure m and for all $\delta > 0$, we have that $E_{\mu_\delta}[AIC(x,n,Z)]$ is* *asymptotically equivalent to* $E_m[AIC(x,n,Z)]$, *where with* μ_δ *we denote the measure given by the Lebesgue measure* m *concentrated on* $U_\delta = (x-\delta, x+\delta)$.

Proof. The proof is based on a simple application of the method we used before. Indeed, given $x \in X$, let's consider $U_\delta = (x-\delta, x+\delta)$ for some δ. If we now want to estimate the value of $E_\delta[N_n]$, that is the mean made with respect to the measure on Ω_M induced by μ_δ, we notice that, from the properties of the map L, it is clear that there exists a $R(\delta) \in \mathbb{N}$ such that $L^{R(\delta)}(U_\delta) = [0,1]$. We can deduce that $E_\delta[N_n]$ is the same as $E_p[N_n]$, where p is the measure on Ω_M induced by m, for $n > R(\delta)$. Then it follows that $E_{\mu_\delta}[AIC(x,n,Z)] \sim E_m[AIC(x,n,Z)]$ for $n > R(\delta)$. □

Corollary 4.33 ([Bo1]). *The AIC of the Manneville maps* T_z *is given by*

- $n^\alpha \leq E_{\mu_\delta}[AIC(x,n,Z)] \leq n^\alpha \log_2 n$ *with* $\alpha = \frac{1}{z-1}$, *for* $z > 2$;
- $E_{\mu_\delta}[AIC(x,n,Z)] \sim n$ *for* $z < 2$

for almost any point $x \in X$ *with respect to the Lebesgue measure* m.

The result of Corollary 4.33 is that we have information on the behaviour of the AIC on average, with respect to any probability measure μ_δ uniformly distributed on a ball of radius δ and center x (for m-almost any $x \in X$). At this point we would like to let δ shrink to 0, to obtain the pointwise behaviour of $AIC(x,n,Z)$. This would lead to prove $\underline{q}(x,Z) = \overline{q}(x,Z) = \underline{q}(X,m) = \overline{q}(X,m)$ for m-almost any $x \in X$. Unfortunately it is not the case. We recall that this approach corresponds to an ergodic theorem for infinite conservative systems. Hence, from Theorem 1.16, we expect that it is impossible to find a sequence of positive integers $\{a_n\}$, satisfying $a_n \to \infty$ and $\frac{a_n}{n} \to 0$ as $n \to \infty$, such that

$$\lim_{n\to\infty} \frac{AIC(x,n,Z)}{a_n} = const > 0$$

for almost any $x \in X$ with respect to the infinite invariant measure $\bar{\mu}$, induced on X by the measure $\bar{p}$ of Theorem 4.26 (we recall that the infinite invariant measure $\bar{\mu}$ is equivalent to the Lebesgue measure). In [CI2] it is shown that for the Manneville map T_2 ($z = 2$), it holds the following

Theorem 4.34 ([CI2]). *For any continuous real function* f *compactly supported on* $(0,1]$ *and for any* $\epsilon > 0$, *it holds*

$$\lim_{n\to\infty} m\left(\left\{x \in [0,1] \;/\; \left|\frac{1}{a_n}\sum_{k=0}^{n-1} f(T_2^k(x)) - \int_0^1 f(x)\, d\bar{\mu}(x)\right| \geq \epsilon\right\}\right) = 0$$

where $a_n = \frac{cn}{\log n}$ *for a positive constant* c.

In [CI2] it is also shown that this result cannot be sharpened to a pointwise result. Moreover it is supported the conjecture that $\overline{q}(x,Z) = 1$ for $\bar{\mu}$-almost any $x \in X$.

We remark that the problem in sharpening the result of Theorem 4.34 is given by the behaviour of the random variables N_n, for which there are weak pointwise results. It is evident since, in the same hypotheses of Theorem 4.34, it is shown in [CI2] that for $\bar{\mu}$ almost any $x \in X$ it holds

$$\lim_{n\to\infty} \frac{1}{N_n(\omega(x))} \sum_{k=0}^{n-1} f(T_2^k(x)) = \int_0^1 f(x)\ d\bar{\mu}(x)$$

where we recall that $\omega(x)$ denotes the symbolic sequence associated to a point $x \in X$ according to the partition Z.

Further results for the local chaos indexes and the pointwise behaviour of the sequence $AIC(x,n,Z)$ can be found in [BG], where some deeper results in the theory of recurrent events are used.

Another interesting property of the Manneville maps is the behaviour of the correlation functions. We restrict our attention to the case of characteristic functions. In particular we consider the measurable sets $B \subseteq [0,1]$ contained in

$$\mathcal{B}_+ := \bigcup_{\epsilon>0} \{B \in \mathcal{B}\ /\ m(B) > 0,\ B \subseteq [0,1] \setminus (0,\epsilon)\}$$

For the *mixing rate* of a set $B \in \mathcal{B}_+$, defined in the case $z < 2$ as

$$r_n(B) := \frac{\bar{\mu}(B \cap T_z^n(B)) - (\bar{\mu}(B))^2}{(\bar{\mu}(B))^2}$$

being $\bar{\mu}$ the probability invariant measure, it is shown in [Is1] that for all $B \in \mathcal{B}_+$, when $z > \frac{3}{2}$ it holds

$$r_n(B) \sim n^{1-\frac{1}{z-1}} \sim n^{-\alpha}$$

with $\alpha = \frac{1}{z-1} - 1$. In the case $z \geq 2$, one studies the *scaling rate* of a set $B \in \mathcal{B}_+$, defined as

$$\sigma_n(B) := \frac{\bar{\mu}(B \cap T_z^n(B))}{(\bar{\mu}(B))^2}$$

being $\bar{\mu}$ the infinite invariant measure. Then it holds ([Is1]) for all $B \in \mathcal{B}_+$

- $z = 2 \ \Rightarrow\ \sigma_n(B) \sim \frac{1}{\log n}$;
- $z > 2 \ \Rightarrow\ \sigma_n(B) \sim n^{\frac{1}{z-1}-1} \sim n^{-\alpha}$ with $\alpha = 1 - \frac{1}{z-1}$.

Let study now the behaviour of the index $\overline{q}_G(X,\mu,Z)$ introduced in Section 3.4.3. From the results of Theorems 4.23 and 4.25, we obtain that

- $\frac{3}{2} < z < 2 \;\Rightarrow\; \overline{q}_G(X, \bar{\mu}, Z) = 2 - \frac{1}{z-1}$;
- $z > 2 \;\Rightarrow\; \overline{q}_G(X, \bar{\mu}, Z) = \frac{1}{z-1}$.

Hence $\overline{q}_G(X, \bar{\mu}, Z) = 1 - \alpha$ for all $z \in (\frac{3}{2}, 2) \cup (2, +\infty)$. This result supports our conjecture of Section 3.4.3.

Finally we apply the results of Theorems 3.15 and 3.17 to the Manneville maps T_z with $z > 2$. We obtain some indications about the behaviour of the recurrence time to the sets of the iterated partitions Z_n, for the given partition Z. We recall from Section 3.4.2, that if $Z_n(x)$ denotes the set of the iterated partition Z_n containing a point $x \in X$, then $\tau(Z_n(x))$ is the Poincaré recurrence time of the set $Z_n(x)$ (see Definition 1.5). Moreover the *upper* and *lower recurrence rates* are defined respectively by

$$\overline{R}_Z(x) = \limsup_{n\to\infty} \frac{\tau(Z_n(x))}{n} \qquad \underline{R}_Z(x) = \liminf_{n\to\infty} \frac{\tau(Z_n(x))}{n}$$

From Theorem 3.17, and from the result $\overline{q}(X, m) = \underline{q}(X, m) = \frac{1}{z-1}$ for T_z with $z > 2$, we have that on a set of positive Lebesgue measure it holds $\overline{R}_Z(x) \geq 1$.

Further results on the Manneville maps and on their recurrence rates will appear in [BG] and [BGI].

4.3.2 Intermittency in unimodal maps

The phenomenon of intermittency has been found in other one-dimensional dynamical systems. In this section, following [HY], we just give an example of generation of intermittency in families of unimodal maps. In Chapter 6 we study, using a particular compression algorithm, this situation for the logistic map.

Let $\{T_\gamma\}_{\gamma\in\mathbb{R}}$ be a one-parameter family of unimodal maps (see Definition 4.6) on the interval $X = [0, 1]$. Let assume that for all $\gamma \in \mathbb{R}$, the following conditions hold:

- T_γ is of class C^3 with a unique maximum at the critical point c (the critical point is the same for all γ);
- $T''_\gamma(c) < 0$ and $T_\gamma(\partial X) \subset \partial X$;
- $ST_\gamma(x) < 0$ for all $x \in X$, where $ST(x)$ denotes the Schwarzian derivative (see Definition 4.5).

A possible scenario of generation of intermittency is a *saddle-node* bifurcation. The family $\{T_\gamma\}$ is said to unfold a saddle-node at $\gamma = 0$ if

(i) for an integer k, there is a k-periodic point p for the map T_0, such that $DT_0^k(p) = 1$;

(ii) $D^2T_0^k(p)\,\frac{\partial T_\gamma^k(p)}{\partial\gamma} > 0$ at $\gamma = 0$.

Condition (i) assures that there is an indifferent fixed point for the map T_0^k, that has the same role of the origin for the Manneville maps (see Figure 4.1). Condition (ii) on one hand describes the behaviour of the maps T_γ for values of γ close to 0. If we assume that both $D^2T_0^k(p)$ and $\frac{\partial T_\gamma^k(p)}{\partial\gamma}$ are positive, then for small negative values of γ there are, close to p, two fixed points for the map T_γ^k, one is stable and the other one is unstable; whereas for small positive γ there are no fixed point close to p for T_γ^k, and there is a small "channel" that the orbits cross in an amount of time that decreases as γ increases. On the other hand, Condition (ii) assures that $D^2T_0^k(p) \neq 0$, hence if the behaviour of the map T_0^k close to p is written in the form $T_0^k(x) = p + (x-p) + c(x-p)^z + hot$, it has to be $z = 2$. Hence we find the behaviour of the Manneville map at $z = 2$.

The family of *logistic maps* T_λ, defined in equation (4.16), satisfies the condition imposed on the family T_γ, and moreover there exists a sequence of parameter values $\{\lambda_m\}$, with $\lambda_m > \lambda_\infty$ and such that $\lambda_m \to \lambda_\infty$ as $m \to \infty$, so that the family T_λ unfolds a saddle-node at each λ_m.

In [HY], given a small neighbourhood U of the periodic orbit of p which does not contain any critical point of T_0^k, it has been studied the behaviour of the function $\Phi_U(x,\gamma)$, the relative frequency with which the orbit of a point x visits U. Namely, for any measurable set $B \subseteq X$,

$$\Phi_B(x,\gamma) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \chi_B(T_\gamma^i(x))$$

whenever the limit exists. The following theorem is proved in [HY].

Theorem 4.35 ([HY]). *In the previous hypotheses for the family $\{T_\gamma\}$, unfolding a saddle-node at $\gamma = 0$, for each $\gamma_0 > 0$ and each rational number $q \in (0,1)$, there exists $\gamma \in (0,\gamma_0)$ such that*

$$\Phi_U(x,\gamma) = q$$

for all x in a subset of X of full Lebesgue measure. Moreover there exist sets Γ of parameter values of positive measure and positive density at 0, that is

$$\lim_{\gamma\to 0^+} \frac{m(\Gamma\cap[0,\gamma))}{\gamma} > 0$$

so that, restricting to $\gamma \in \Gamma$, the function $\Phi_U(x,\gamma)$ is a constant $\Phi_U(\gamma)$ for m-almost all $x \in X$, and it depends continuously on γ at $\gamma = 0$. For each such set Γ, there exists a constant $C > 0$ such that

$$\lim_{\gamma\in\Gamma,\ \gamma\to 0^+} \frac{1-\Phi_U(\gamma)}{\sqrt{\gamma}} = C \tag{4.35}$$

The fact that $\Phi_U(0) = 1$ is a consequence of condition (ii). Indeed being the map T_0^k equivalent, close to p, to the Manneville map at $z = 2$, the only physical measures of this map are the Dirac deltas concentrated on the points of the periodic orbit of p (that is the indifferent fixed points for T_0^k). Hence the limit in the definition of Φ_U has to converge, for $\gamma = 0$, to $\chi_U(p) = 1$. It is interesting to our aims the result of equation (4.35), that gives the order of convergence. In Chapter 6 we will study this behaviour for the family of logistic maps T_λ using the AIC, and in particular a compression algorithm to approximate it.

Part II

Applications to real time series

Chapter 5

Randomly perturbed dynamical systems

In the first part of this work we have introduced the tools from information theory that we use to study the behaviour of chaotic dynamical system. Moreover we have studied the relations between our approach and the classical indexes in the theory of dynamical systems, such as the sensitivity to initial conditions, the return time to a set and the decay of correlations. Then we have applied our approach to some well known dynamical systems, that have been a good test for our method, showing its importance in the classification of weakly chaotic dynamical systems.

In this second part, we want to test the applicability of our method to real time series, that is outputs of numerical experiments. This problem has two aspects. The first is that when studying numerical experiments one has to face the unavoidable presence of noise in the outputs. Hence one needs to know how to distinguish the chaoticity due to the noise and that due to the intrinsic properties of the system. The second aspect has to do with the already remarked difficulty in actually evaluating the *AIC* of a symbolic word.

In this chapter we study the former problem, showing how, studying the behaviour of the *AIC*, one can have some information about the relation between the noise and the intrinsic properties of a dynamical system. In particular we use the equivalence between the complexity of a symbolic sequence, and the Kolmogorov-Sinai (KS) entropy of the dynamical system that has produced the sequence. This chapter is based on [Bo2].

As it has been proved in [Ki] for a typical situation, the KS entropy of a dynamical system perturbed by a sequence of random variables is infinite for all $\sigma > 0$; here σ is a real parameter measuring the size of the perturbation. This result shows that the KS entropy is not the right quantity to look at in randomly perturbed dynamical systems. In [Ki] it has been suggested to study the KS entropy relative to a partition.

We study the perturbation of a dynamical system (X, T, μ) by some noise. By noise we mean a discrete stochastic processes of independent and identically distributed (i.i.d.) random variables $\{w_n\}_{n\in\mathbb{N}}$ defined on a probability space $(W, \mathcal{P}, P)$ with values on the interval $[-\sigma, \sigma]$, for a positive real number σ. The number σ is the parameter that measures the size of the perturbation on the original system.

We also assume that for all σ there is an ergodic invariant probability measure μ_σ on the perturbed system (X, T_σ) such that the probability measure μ is the weak limit of the sequence μ_σ as $\sigma \to 0$ (for the formal definition of μ_σ see Section 5.1).

In [Ki] it has been proved that, under these hypothesis, it holds

$$\limsup_{\sigma\to 0^+} h_{\mu_\sigma}(T_\sigma, Z) \leq h_\mu(T, Z) \leq h_\mu(T) \ , \tag{5.1}$$

being $h_\mu(T, Z)$ and $h_{\mu_\sigma}(T_\sigma, Z)$ the KS entropies relative to the partition Z of the original and the perturbed system respectively, and by $h_\mu(T)$ we denote the KS entropy of the unperturbed dynamical system. The equality holds for certain hyperbolic systems ([Ki]).

In equation (5.1) it is used the KS entropy relative to a fixed partition, avoiding the problem of having infinite KS entropy, and it is performed the limit for the size of the perturbation going to zero. This result shows that the KS entropy relative to a partition for an unperturbed dynamical system is well approximated by the same quantity for the perturbed system, when the perturbation is small enough.

Unfortunately when we want to study time series obtained by measurements of real systems, it is impossible to know how small is the perturbation to the original system. Hence it is impossible to perform the limit in equation (5.1).

Real systems are observed using a symbolic representation of the measurements. That is the phase space is divided into a finite number of sets and we study the time series given by the symbols of the sets of the partition visited by the system as time grows.

The main idea to study real systems is to look at the system using different scales of observation, which corresponds to use different sizes for the partition of the phase space. When the partition is coarse the effects of the noise are negligible, whereas, when the partition is very fine, then the effects of the noise hide the dynamics of the original system. Then the right approach to randomly perturbed systems is to study the behaviour of the KS entropy while the partition is refined and the noise is fixed.

In this chapter we prove that, when the diameter of the partition is big with respect to the size of the noise, the KS entropy of the perturbed system is a good approximation for the unperturbed KS entropy. Our main result is contained in equations (5.8) and (5.9), which in particular imply equation (5.1). Moreover we obtain precise estimates for the effects of the noise, in order to distinguish between the perturbation and the dynamics of the system, a very important problem in the analysis of time series ([Ce]). We remark that the main extension of the result contained in equation (5.1) is its applicability to time series obtained by observations of real systems.

To apply our method we have to face the problem of the measurement of the KS entropy in real systems (that is the KS entropy of observed time series). This is an important problem (for a review see for example [Ab]), and many papers have been dedicated to the approximation of the KS entropy by different indicators. We recall the approach of Grassberger and Procaccia [GP1] based on the generalized Renyi entropy K_2, and that of Cohen and Procaccia [CP] based on the correlation integral for a time series ([GP2]).

In Sections 5.1 and 5.2 we apply the method of the AIC to randomly perturbed systems.

As stated above, our approach deals with the KS entropy of a system relative to a measurable partition Z of the phase space X. We recall that, by the Kolmogorov-Sinai Theorem 1.29, if the chosen partition Z is generating, then the KS entropy $h_\mu(T,Z)$ relative to the partition is equal to the KS entropy $h_\mu(T)$ of the system (for an invariant probability measure μ). Moreover the existence of a finite generating partitions is assured by the Krieger Generator Theorem 1.30.

We now introduce slightly different notations that we use in this chapter. For a finite invariant dynamical system (X,T,μ), let ϵ be the diameter of a uniform partition Z of the space X, and simply denote by $h_\mu(\epsilon)$ the KS entropy $h_\mu(T,Z)$ relative to the partition Z. This notation has the aim to enhance the role of the diameter of the partition considered. The quantity $h_\mu(\epsilon)$ is also called the *ϵ-entropy* of the dynamical system (see [GW2] for a review). The Krieger Generator Theorem 1.30 implies that, if $h_\mu(T)$ is finite, there exists ϵ_0 such that $h_\mu(\epsilon) = h_\mu(T)$ for all $\epsilon \leq \epsilon_0$.

It is also possible to establish the behaviour of the ϵ-entropy for discrete stochastic processes of i.i.d. random variables.

Theorem 5.1 (Gaspard-Wang [GW2]). *Let $\{w_n\}_{n\in\mathbb{N}}$ be a discrete stochastic process of i.i.d. random variables with values on the interval $[0,1]$. Then as $\epsilon \to 0$ it holds*

$$h(\epsilon) \sim -\log \epsilon \ ,$$

where the entropy is computed with respect to the invariant probability measure of the system (the induced measure) absolutely continuous with respect to the Lebesgue measure.

When dealing with randomly perturbed systems, it has been shown using numerical experiments ([GW2], [SD]) that the expected behaviour is

$$h_{\mu_\sigma}(\epsilon) := h_{\mu_\sigma}(T_\sigma, Z) \sim \begin{cases} h_\mu(\epsilon) & \text{for } \epsilon >> \sigma \\ -\log \epsilon & \text{for } \epsilon << \sigma \end{cases} \tag{5.2}$$

where σ denotes the standard deviation of the noise (that is the square root of the variance of the random variables w_n). The same result is expected for the K_2 entropy ([SK]), that is the generalized Renyi entropy K_q for $q = 2$.

We recall that for randomly perturbed systems the KS entropy $h_{\mu_\sigma}(T_\sigma)$ is infinite, hence the Krieger Generator Theorem is not applicable to the system (X, T_σ). But we assume that the unperturbed system (X, T) has finite KS entropy $h_\mu(T)$ with generating partition of diameter ϵ_0. Hence $h_\mu(\epsilon) = h_\mu(T)$ for all $\epsilon \leq \epsilon_0$. This implies that if the size σ of the noise is small with respect to ϵ_0, approximating $h_\mu(\epsilon)$ by $h_{\mu_\sigma}(\epsilon)$, we find a good approximation to the KS entropy of the unperturbed system.

In the following we prove (5.2) using the behaviour of the *AIC* of the symbolic words generated by the perturbed system. We remark that the behaviour of equation (5.2) is of great importance, since it would give us a method to estimate the size of the random perturbation on the deterministic dynamics of the system we are observing.

5.1 Estimates of the relative KS entropy

We now introduce formal definitions for random perturbations of a dynamical system (following [Ki]). We also give some estimates for the KS entropy relative to a partition for the perturbed system using the tool of symbolic dynamics.

Let $\{w_n\}$ be a stochastic process of i.i.d. random variables, with each w_n defined on the probability space $(W, \mathcal{P}, P)$ with values on the interval $[-\sigma, \sigma]$, and q^σ as the induced distribution.

Definition 5.2. A *random perturbation* of the finite invariant dynamical system (X, T, μ) is a family of Markov chains f^n_σ on the probability space $(W, \mathcal{P}, P)$ with values on X and with transition probabilities given by

$$P^\sigma(x, B) := P\{T^{n+1}_\sigma \in B \mid T^n_\sigma = x\} = q^\sigma(B - T(x)) ,$$

for all Borel sets $B \subset X$.

The meaning of this definition is that the point x is moved to the point $T(x)$ under the unperturbed dynamics and then it disperses randomly with distribution q^σ.

We will consider two different possible actions of the random perturbation (see [Arg]). We talk of *output noise* when the dynamics is driven only by the unperturbed map T, and the dispersion is caused by a non exact observation of the point $T(x)$. That is given a point $x \in X$, its orbit is given by $x_n = T(x_{n-1})$ and our data are measured observing the orbit $y_n = T_\sigma(x_n) = x_n + w_n$. Instead we talk of *dynamical noise* when the dispersion is intrinsic in the dynamics and it is not caused by the observation, in this case the dynamics is driven by T_σ. So, for a point $x \in X$ we have $x_n = T^n_\sigma(x) = T(x_{n-1}) + w_n$.

Definition 5.3. A probability measure μ_σ on X is an *invariant measure* of the randomly perturbed dynamical system (X, T_σ) if

$$\mu_\sigma(B) = \int_X P^\sigma(x, B)\ d\mu_\sigma(x)$$

for all Borel sets $B \subset X$.

We now start from the following assumptions:

(i) there is a finite invariant ergodic dynamical system (X, T, μ) with finite KS entropy, and denote by $h_\mu(\epsilon)$ its KS entropy relative to a finite measurable partition Z of diameter ϵ;

(ii) we can analyze data produced by a random perturbation (X, T_σ) of the system (output or dynamical noise) with fixed σ;

(iii) there is an invariant and ergodic probability measure μ_σ on (X, T_σ) and the KS entropy relative to the partition Z for the randomly perturbed system is denoted by $h_{\mu_\sigma}(\epsilon)$.

To analyze our system by the *AIC* we have to use the method of symbolic dynamics. Given the finite measurable partition $Z = \{I_0, \dots, I_{N-1}\}$ of diameter $\epsilon = \frac{1}{N}$, we can associate a symbolic orbit in $\Omega = \{1, \dots, N\}^{\mathbb{N}}$ to any orbit in the space X using the map φ_Z defined in equation (1.7).

In the case of the unperturbed system (X, T, μ) we denote by $\Omega' \subseteq \Omega$ the image of the map φ_Z. The set Ω' is closed with respect to the action of the usual shift map τ defined on Ω. It is also possible to define the probability measure ν on Ω' induced by φ_Z. So we can work on the finite invariant ergodic dynamical system (Ω', ν, τ) with KS entropy $h_\nu = h_\mu(\epsilon)$.

Analogously we define the map $\varphi_{Z,\sigma}$ from the product space $X \times W$ to Ω in the following way: for each $x \in X$ the map $\varphi_{Z,\sigma}(x, \cdot)$ is the symbolic map for the Markov chain T_σ^n. So for every point $x \in X$ there is a set of possible symbolic orbits in $\Omega'_\sigma := \varphi_{Z,\sigma}(X \times W)$ that correspond to different realizations of the stochastic process $\{w_n\}$. Let ν_σ be the probability measure induced on Ω'_σ by the measure μ_σ. Then we study the finite invariant ergodic dynamical system $(\Omega'_\sigma, \nu_\sigma, \tau)$ with KS entropy $h_{\nu_\sigma} = h_{\mu_\sigma}(\epsilon)$.

Since now we will denote by ψ and by ω the symbolic orbits in Ω' and Ω'_σ respectively.

From the previous definitions easily follow the following propositions

Proposition 5.4 ([Bo2]). $\Omega' \subseteq \Omega'_\sigma$

Proposition 5.5 ([Bo2]). $h_\mu(\epsilon) \leq -\log \epsilon\ , \quad h_{\mu_\sigma}(\epsilon) \leq -\log \epsilon\ .$

Proposition 5.6 ([Bo2]). *If $\Omega' = \Omega'_\sigma$ and ν is absolutely continuous with respect to ν_σ ($\nu << \nu_\sigma$) then $h_\mu(\epsilon) = h_{\mu_\sigma}(\epsilon)$.*

Proof. For ν_σ-almost any $\omega \in \Omega'_\sigma$ it holds

$$\lim_{n\to\infty} \frac{AIC(\omega^n)}{n} = h_{\mu_\sigma}(\epsilon) \ , \tag{5.3}$$

thanks to Theorems 2.12 and 2.13. Hence, since for ν-almost any $\psi \in \Omega'$ it holds

$$\lim_{n\to\infty} \frac{AIC(\psi^n)}{n} = h_{\mu}(\epsilon) \ , \tag{5.4}$$

and $\nu << \nu_\sigma$ then $h_\mu(\epsilon) = h_{\mu_\sigma}(\epsilon)$. □

We conclude this section with a lower bound for the ϵ-entropy of the random perturbed system.

Proposition 5.7 (Kifer [Ki] Theorem 2.4). *Suppose that all transition probabilities $P^\sigma(x,\cdot)$ have bounded densities $p^\sigma(x,y) \leq K$, then $h_{\mu_\sigma}(\epsilon) \geq -\log\epsilon - \log K$.*

Hence the KS entropy of a randomly perturbed system is always infinite, since it can be computed as the limit of $h_{\mu_\sigma}(\epsilon)$ for ϵ going to 0.

5.2 The *AIC* for noisy systems

5.2.1 The case of output noise

We now study the behaviour of the *AIC* relative to a partition in randomly perturbed dynamical systems. We start with the case of output noise (see previous section) which is easier to be studied and hence it is useful to introduce the arguments we will use for the more important case of dynamical noise.

Let Ω' and Ω'_σ be the symbolic spaces of the unperturbed and the perturbed systems respectively. We will study the ϵ-entropy of the perturbed system using the AIC of the symbolic orbits, ψ and ω, of the original system and of its perturbed version.

At any step of our dynamical system the perturbation induced by the noise changes the position of the point, and so it could change the set of the partition in which the point is. So the n-th symbol of the sequence ω could be different from ψ_n, the corresponding symbol of the sequence ψ. We recall that since we are in the case of output noise, at the step $n+1$ there is no memory of the noise at the previous step. Moreover since the random variables w_n are independent, the probability q of ω_n being different from ψ_n does not depend on the step n, neither on the set of the partition we consider.

Let $H(q)$ be the KS entropy of a *Bernoulli trial of parameter* $q \in [0,1]$, namely the stochastic process of independent Bernoulli variables ϑ_n with parameter q. It is well known that

$$H(q) = -q \ \log(q) - (1-q) \ \log(1-q) \ .$$

Hence applying Theorems 2.12 and 2.13 it holds

$$\lim_{n\to\infty} \frac{H_n(q)}{n} = H(q) \ , \tag{5.5}$$

where $H_n(q)$ denotes the average of the AIC over all the sequences produced by the q-Bernoulli trial.

In this framework, we have

Proposition 5.8 ([Bo2]). *If the random variables w_n have values in the interval $[-\sigma, \sigma]$, then*

$$\mathbb{E}_{\mu_\sigma}[AIC(\omega^n)] \leq \mathbb{E}_{\mu}[AIC(\psi^n)] + nq \log\left(2\left\lceil\frac{\sigma}{\epsilon}\right\rceil\right) + H_n(q) \ ,$$

where $\lceil\cdot\rceil$ denotes the upper integer part of a real number and $\mathbb{E}[\cdot]$ denotes the mean value.

Proof. To prove the proposition we show an algorithm to describe the word ω^n and compute the information it needs. Let assume that we know the word ψ^n. We have then just to specify the symbols of ω^n that are different from those of ψ^n. We call i_n this information. We show that for all words ω^n and ψ^n it holds

$$i_n \leq r_n = (\#\{i \ / \ \psi_i \neq \omega_i\}) \ \log\left(2\left\lceil\frac{\sigma}{\epsilon}\right\rceil\right) + H_n(q) \ . \tag{5.6}$$

The algorithm to describe ω^n works as follows. First it describes a binary word $s = (s_0, \ldots, s_{n-1})$, where $s_i = 0$ implies that $\omega_i = \psi_i$ and $s_i = 1$ otherwise. To this aim we need on average $H_n(q)$ bits of information, since the word s is obtained as a word of a q-Bernoulli trial.

Moreover the algorithm needs also to specify how far on the left or on the right the noise has moved the point. To this aim we need $\log\left(2\left\lceil\frac{\sigma}{\epsilon}\right\rceil\right)$ bits of information for each symbol of ω^n different from the corresponding symbol of ψ^n. The factor $\left\lceil\frac{\sigma}{\epsilon}\right\rceil$ counts exactly how many sets of the partition Z can be covered by the effect of the noise, and the factor 2 is needed to specify whether the point is moved to the right or to the left.

Hence to specify the word ω^n we do not need more than $AIC(\psi^n) + r_n$ bits of information. If we evaluate the mean of $AIC(\omega^n)$ over the measure μ_σ then the thesis follows from the definition of the probability q. $\square$

Theorem 5.9 ([Bo2]). *If $h_\mu(\epsilon)$ is the KS entropy relative to a partition Z of diameter ϵ of an unperturbed dynamical system (X, T, μ), for the ϵ-entropy $h_{\mu_\sigma}(\epsilon)$ of the system perturbed by an output noise $\{w_n\}$ with values in the interval $[-\sigma, \sigma]$ it holds*

$$h_{\mu_\sigma}(\epsilon) \leq h_\mu(\epsilon) + q \log \left(2 \left\lceil \frac{\sigma}{\epsilon} \right\rceil \right) + H(q) \ .$$

Proof. The thesis simply follows from Proposition 5.8, using equations (5.4) and (5.3) for $h_\mu(\epsilon)$ and $h_{\mu_\sigma}(\epsilon)$, and using equation (5.5). □

5.2.2 The case of dynamical noise

We now study the case of dynamical noise. We assume to have an unperturbed dynamical system (X, T, μ) and a dynamical perturbation due to a stochastic process $\{w_n\}$ of i.i.d. random variables with values in the interval $[-\sigma, \sigma]$, such that the points $x \in X$ follow orbits given by $x_{n+1} = T(x_n) + w_{n+1}$.

Let consider a partition Z of the space X in a finite number N of measurable sets $\{I_i\}$, such that the diameter of each set of the partition is $\epsilon = 1/N$. Let $\Omega = \{1, \dots, N\}^{\mathbb{N}}$ be the space of the symbolic orbits associated to the dynamical system, and denote by ω a single sequence in Ω'_σ, given by the measurements made on the system. That is ω is the symbolic orbit of a point $x \in X$ after the effect of the noise. To estimate the AIC of the subwords of ω from above, we show how to construct it from the knowledge of the noise and of the unperturbed dynamics.

Let x_0 denote the initial point of an orbit of the system. By the classical symbolic representation of orbits, there is a symbolic sequence $\psi^0 = (\psi^0_0, \psi^0_1, \dots) \in \Omega$, relative to the partition Z, associated to the point x_0 for the unperturbed dynamical system, that is $T^i(x_0) \in I_{\psi^0_i}$ for all $i \in \mathbb{N}$, and there is also the symbolic sequence $\omega = (\omega_0, \omega_1, \dots)$ that is given by the perturbed system. We have $\omega_0 = \psi^0_0$. In the case of dynamical noise, we have to consider the effect due to the fact that the action of the noise is not forgotten at each step. So we have to extend the knowledge of the unperturbed orbit to more than just one iteration. To find a good number of iterations to be stored, we use the approximation of the Kolmogorov-Sinai entropy $h_\mu(T, Z) = h_\mu(\epsilon)$ relative to the partition Z by the decreasing sequence $H_\mu(T^{-n}Z|Z_n)$ (see equation (1.26)). Hence for any $\delta > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$h_\mu(\epsilon) - \delta \leq H_\mu(T^{-n_0}Z|Z_{n_0}) \leq h_\mu(\epsilon) + \delta \ . \tag{5.7}$$

Let $\delta > 0$ be fixed, and find the corresponding integer n_0 according to equation (5.7). Assume that for the initial point x_0 of the orbit, we know the first n_0 digits of the sequence ψ^0. That is we assume to know the cylinder

$Z_{n_0}(x_0)$, where with this notation we mean the set of the partition Z_{n_0} containing x_0.

Let $\tilde{x}_1 = T(x_0)$ and $x_1 = \tilde{x}_1 + w_1$. As before we denote by ψ^1 the unperturbed symbolic orbit associated to the point x_1. We assume to know the action of the noise w_1 with respect to the partition Z_{n_0}, so we compare the strength of the noise σ with the diameter[1] ϵ_{n_0} of the partition Z_{n_0}. Moreover, to know $Z_{n_0}(\tilde{x}_1)$ we only need the symbol $\psi^0_{n_0}$, and to know $Z_{n_0}(x_1)$, once we have $Z_{n_0}(\tilde{x}_1)$, we only need the action of w_1 with respect to the partition Z_{n_0}. This is enough to obtain $\omega_1 = \psi^1_0$. So to obtain the first two symbols of the sequence ω it is enough the following information: $I(\psi^0_0, \dots, \psi^0_{n_0-1})$, $I(w_1|Z_{n_0})$ and $I(\psi^0_{n_0}|\psi^0_0, \dots, \psi^0_{n_0-1})$. We have used $I(\cdot|\cdot)$ to denote the conditional information.

At this point it is possible to iterate the argument with the sequence ψ^1 instead of the sequence ψ^0, hence to find the symbol $\omega_2 = \psi^2_0$ using the following information: $I(w_2|Z_{n_0})$, $I(\psi^1_0, \dots, \psi^1_{n_0-1})$ (which is given by $Z_{n_0}(x_1)$ and is indeed known) and $I(\psi^1_{n_0}|\psi^1_0, \dots, \psi^1_{n_0-1})$.

Iterating this argument we obtain the following proposition:

Proposition 5.10 ([Bo2]). *If the random variables w_n have values in the interval $[-\sigma, \sigma]$, then for any $\delta > 0$ there exists $n_0 \in \mathbb{N}$ and a constant $C > 0$ such that*

$$\mathbb{E}_{\mu_\sigma}[AIC(\omega^n)] \leq (n-1)\left[h_\mu(\epsilon) + \delta + q \log\left(2\left\lceil \frac{\sigma}{\epsilon_{n_0}} \right\rceil\right)\right] + H_{n-1}(q) + C$$

where $\lceil\cdot\rceil$ denotes the upper integer part of a real number, the probability q is defined as in Proposition 5.8 and $\omega^n = (\omega_0, \dots, \omega_{n-1})$.

Proof. To obtain the entropy on the right hand side we only have to use equation (5.7), and to remark that the conditional entropy $H_\mu(T^{-n_0}Z|Z_{n_0})$ is obtained as the integral over the space of all admissible sequences of the information function $I(\psi_{n_0}|\psi_0, \dots, \psi_{n_0-1})$. The constant C is the information $I(\psi^0_0, \dots, \psi^0_{n_0-1})$ needed to start the iteration.

For the part related to the noise we repeat the same argument of Proposition 5.8. Then the action of the noise can be bounded by

$$\mathbb{E}_P[r_n] = (n-1) q \log\left(2\left\lceil \frac{\sigma}{\epsilon_{n_0}} \right\rceil\right) + H_{n-1}(q)$$

bits of information. □

Let again $h_{\mu_\sigma}(\epsilon)$ be the ϵ-entropy of the perturbed dynamical system, hence it follows

[1] In this case we cannot ask the partition Z_{n_0} to be uniform, so we consider the diameter to be the smaller diameter of the intervals that make the partition Z_{n_0}.

Theorem 5.11 ([Bo2]). *If $h_\mu(\epsilon)$ is the ϵ-entropy of an unperturbed dynamical system (X, T, μ), for the ϵ-entropy $h_{\mu_\sigma}(\epsilon)$ of the system perturbed by a dynamical noise $\{w_n\}$ it holds: for any $\delta > 0$ there exists $n_0 \in \mathbb{N}$ such that*

$$h_{\mu_\sigma}(\epsilon) \le h_\mu(\epsilon) + \delta + q \log\left(2\left\lceil\frac{\sigma}{\epsilon_{n_0}}\right\rceil\right) + H(q)$$

5.2.3 Conclusions

At this point we can make some conclusions about the estimates of the KS entropy of a randomly perturbed system.

We recall the conditions we imposed in Section 5.1 on the data we study. So, given a fixed σ, we vary the diameter ϵ of the partition considered.

Let $\epsilon > \sigma$. In this case we can assume that for some ϵ we have $\Omega' = \Omega$. This implies (Propositions 5.4 and 5.6) that $h_{\mu_\sigma}(\epsilon) = h_\mu(\epsilon)$. When $\Omega' \subset \Omega$ we have that, fixed $\delta > 0$ such that $h_\mu(\epsilon) + \delta < -\log\epsilon$, if $\epsilon_{n_0} > \sigma$

$$-\log\epsilon - const \ \le h_{\mu_\sigma}(\epsilon) \le h_\mu(\epsilon) + \delta + q\log 2 + H(q) \ , \tag{5.8}$$

using Proposition 5.7 and Theorem 5.11. We recall that $H(q)$ takes its maximum value $\log 2$ for $q = 0.5$, and it converges to 0 as q goes to 0 or 1. Then if $\epsilon_{n_0} >> \sigma$ we have that q converges to 0, hence $(q\log 2 + H(q))$ converges to 0, and equation (5.8) becomes $h_{\mu_\sigma}(\epsilon) \le h_\mu(\epsilon) + \delta$. In particular this implies equation (5.1). If instead $\epsilon_{n_0} \le \sigma$ then we have the same conclusion as in the following case.

If $\epsilon < \sigma$ then it holds

$$\begin{aligned} -\log\epsilon - const \ &\le h_{\mu_\sigma}(\epsilon) \le \\ &\le \min\left(-\log\epsilon, h_\mu(\epsilon) + \delta + q\log\left(2\left\lceil\tfrac{\sigma}{\epsilon_{n_0}}\right\rceil\right) + H(q)\right) \ , \end{aligned} \tag{5.9}$$

using Propositions 5.5, 5.7 and Theorem 5.11. Again if $\epsilon_{n_0} << \sigma$ then q converges to 1, hence $H(q)$ converges to 0, and the estimate on the right hand side of equation (5.9) becomes $h_\mu(\epsilon) + \delta - \log\epsilon_{n_0} + const$.

From equations (5.8) and (5.9) it emerges that by computing the KS entropy relative to partitions, we obtain an indication of the smallness of the random perturbation present in the data we are studying. Indeed we expect to find a change in the behaviour of the curve $h_{\mu_\sigma}(\epsilon)$ as ϵ becomes comparable to σ. In Chapter 6 we apply this method to a perturbed dynamical system, using a compression algorithm to estimate the *AIC*.

Moreover this method can also be thought of as a method to distinguish between purely stochastic systems and randomly perturbed dynamical systems. Indeed for randomly perturbed systems equations (5.8) and (5.9) suggest the presence of a change of behaviour for the ϵ-entropy at the level of the perturbation. This change should not appear in purely stochastic systems. However this aspect has to be examined further.

Chapter 6

Numerical experiments

We have so far presented the applications of the use of information measures to study dynamical systems. In particular, in the previous chapter, we have shown the application of our method to the study of real time series, obtained by real experiments, in which the measurements are affected by the presence of noise.

All the theory we have presented is based on the notion of *Algorithmic Information Content (AIC)*, introduced in Section 2.3, a particular example of information measure on the set $\mathcal{A}^*$, the set of the finite words built on a finite alphabet $\mathcal{A}$. An important feature of the concept of *AIC* is that it does not depend on the context from which the word is extracted. So, using the language of dynamical systems, the *AIC* of a symbolic word generated by a system does not depend on the invariant measure on the phase space. Hence, we have a measure of information depending on the single word. However, the *complexity* defined using the *AIC* still has the property of coinciding with the Kolmogorov-Sinai entropy for finite ergodic invariant dynamical systems (see Theorems 2.12 and 2.13). Hence, measuring the *AIC* of a "generic" (with respect to the ergodic invariant measure of the system) symbolic word of a system, we have important information on the system. Not only the KS entropy, but also information on the sensitivity to initial conditions, on the return time to a measurable set and on the decay of correlations (see Chapter 3).

However, as already stated in Section 2.3 and repeatedly remarked, the *AIC* is not a *computable function*, that is there cannot be an algorithm able to compute the *AIC* of any word in $\mathcal{A}^*$. Hence to apply our method, we cannot compute the *AIC* for a real word generated by a system. The solution to this problem is to search for an algorithm able to approximate well enough, for our aims, the behaviour of the *AIC*. We will use *compression algorithms*, and in particular a compression algorithm we developed, called *CASToRe*, *C*ompression *A*lgorithm *S*ensitive *To Re*gularity. The name is justified by the fact that the important property of the algorithm *CASToRe* is to be able to distinguish between chaotic and weakly chaotic dynamical systems.

In the next section we briefly introduce the theory of compression algorithms and the algorithm *CASToRe*. Then in Section 6.2, we show applications of *CASToRe* to time series obtained by some dynamical systems, including those studied in Chapter 3, and randomly perturbed dynamical systems, confirming the results of Chapter 5. This chapter is based on [Be].

6.1 Compression algorithms

We recall from Chapter 2 that, given a finite alphabet $\mathcal{A}$, an information function $I : \mathcal{A}^* \to \mathbb{N}$ is a function such that $I(s)$ is, intuitively, the length of the smallest message from which it is possible to reconstruct $s \in \mathcal{A}^*$. In Chapter 2 we recalled the notions of Shannon entropy, empirical entropy and Algorithmic Information Content. We now study the information function that is associated to an algorithm $\mathcal{Z}$, defined on $\mathcal{A}^*$, that gives as output an encoding $\mathcal{Z}(s)$ of a word $s \in A^*$. The word $\mathcal{Z}(s)$ is a word in $\{0,1\}^*$.

Let us suppose to have some recursive lossless (reversible) coding procedure $\mathcal{Z} : \mathcal{A}^* \to \{0,1\}^*$ (for example, the data compression algorithms that are in any personal computer). Since the coded word contains all the information that is necessary to reconstruct the original word, we can consider the length of the coded word as an approximate measure of the quantity of information that is contained in the original word. The information content of a word $s \in \mathcal{A}^*$ with respect to $\mathcal{Z}$ is defined as $I_{\mathcal{Z}}(s) = |\mathcal{Z}(s)|$.

First of all we assume that the complexity associated to the information content $I_{\mathcal{Z}}$ (see Definition 2.2) satisfies

$$K_{\mathcal{Z}}(\omega) := \limsup_{n\to\infty} \frac{I_{\mathcal{Z}}(\omega^n)}{n} = h_\mu \tag{6.1}$$

for μ-almost all sequences $\omega \in \Omega = \mathcal{A}^{\mathbb{N}}$, being μ an invariant ergodic probability measure on the dynamical system (Ω, τ) (where τ is the usual shift map) and h_μ the Shannon entropy of the system.

Of course not all the coding procedures are equivalent and give the same performances. For this reason we introduce the notion of *optimality* of an algorithm $\mathcal{Z}$, defined by comparing its compression ratio with the empirical entropy.

An algorithm $\mathcal{Z}$ is considered optimal if its compression ratio $|\mathcal{Z}(s)|/|s|$ is asymptotically less than or equal to $\frac{1}{l}\,\hat{H}_l(s)$ for all l. We also recall that, from the definition of AIC, for any algorithm $\mathcal{Z}$ it holds

$$\frac{AIC(s)}{|s|} \leq \frac{|\mathcal{Z}(s)|}{|s|} \tag{6.2}$$

Definition 6.1. A reversible coding algorithm $\mathcal{Z}$ is *optimal* if for all $l \in \mathbb{N}$ there is a function f_l, with $f_l(n) = o(n)$, such that for all finite words s

$$\frac{|\mathcal{Z}(s)|}{|s|} \leq \frac{1}{l}\ \hat{H}_l(s) + \frac{f_l(|s|)}{|s|}.$$

Many data compression algorithms that are used in applications are proved to be optimal.

Remark 6.2. *The universal coding algorithms $LZ77$ and $LZ78$ ([ZL1], [ZL2]) satisfy Definition 6.1. For the proof see [KM].*

The notion of optimality is not enough if we ask a coding algorithm to be able to reproduce the rate of convergence of the sequence $\frac{1}{l}\ \hat{H}_l(\omega^n)$ as $n \to \infty$ for sequences generated by ergodic weakly chaotic dynamical systems, for which (see Theorem 2.6)

$$\lim_{l\to\infty}\ \lim_{n\to\infty}\ \frac{1}{l}\ \hat{H}_l(\omega^n) = 0$$

Indeed, if in the positive entropy systems optimality implies that asymptotically $|\mathcal{Z}(\omega^n)|/n$ is equivalent to $\frac{1}{l}\ \hat{H}_l(\omega^n)$ (using equation (6.2) and the equivalence with the Shannon entropy for the complexities defined using the *AIC* and the empirical entropy), in the weakly chaotic systems it may happen that the asymptotic behaviour dominant in the right hand side of equation (6.1) is that of the function $\frac{f_l(n)}{n}$.

For example let us consider the string 0^n1 and the $LZ78$ algorithm. Then $\frac{1}{l}\ \hat{H}_l(0^n1)$ goes like $log(n)/n$ while $LZ78(0^n1)/n$ goes like $\frac{n^{1/2}log(n)}{n}$ (see also [Ar]). This implies that optimality is not sufficient to have a coding algorithm able to characterize null entropy sequences according to the rate of convergence of their entropy to 0. For this aim we need an algorithm having the same asymptotic behaviour of the empirical entropy. In this way even in the null entropy case our algorithm will provide a meaningful measure of the information. The following definition (from [KM]) is an approach to define optimality of a compression algorithm for the null entropy case.

Definition 6.3. A compression algorithm $\mathcal{Z}$ is called *asymptotically optimal* with respect to $\hat{H}_l$ if it is optimal and there are a function g_l with $g_l(n) = o(n)$ and a constant $\lambda > 0$ such that for all $s \in \mathcal{A}^*$ with $\hat{H}_l(s) \neq 0$ it holds

$$|\mathcal{Z}(s)| \leq \lambda \frac{|s|}{l}\ \hat{H}_l(s) + g_l(|\mathcal{Z}(s)|).$$

It is not trivial to construct an asymptotically optimal algorithm. For instance the well known Lempel-Ziv compression algorithms are not asymptotically optimal. $LZ78$ is not asymptotically optimal even with respect to

$\hat{H}_1$ ([KM]). In [KM] some examples are described of algorithms (LZ78 with RLE and LZ77) which are asymptotically optimal with respect to $\hat{H}_1$. But these examples are not asymptotically optimal for each $\hat{H}_l$ with $l \geq 2$. The asymptotic optimality of LZ77 with respect to $\hat{H}_1$ (Theorem 6.5) however is enough to prove (see Theorem 6.8) that LZ77 can estimate correctly the information of sequences generated by the Manneville maps.

The set of asymptotically optimal compression algorithms with respect to each $\hat{H}_l$ is not empty. In [Ga1] an example is given of a compression algorithm that is asymptotically optimal for each $\hat{H}_l$. The algorithm is similar to the Kolmogorov frequency coding algorithm which is also used in [Br]. This compression algorithm is not of practical use because of its computational complexity.

To our knowledge the problem of finding a fast asymptotically optimal compression algorithm is still open.

We now enter into details of how a compression algorithm works on sequences generated from a dynamical system. In particular we study the behaviour of $LZ77$ on sequences generated by the Manneville maps. Then we introduce the compression algorithm *CASToRe*.

6.1.1 The algorithm *LZ77*

The Ziv-Lempel compression scheme $LZ77$ with infinite window ([ZL1]) is the one from which almost all practical adaptive dictionary encoders derived. A dictionary of an input word is the set of words (i.e. group of consecutive symbols) in which the algorithm parses the input word.

The essence is that phrases (i.e. sets of consecutive words in the word to be encoded) are replaced with a pointer to where they have occurred earlier in the input word. Novel words and phrases can also be constructed from parts of earlier words.

In the $LZ77$ compression algorithm, the new word is defined as a pair (*pointer,symbol*). The pointer is referred to a phrase contained in the part of the input word which precedes the current position of the front end. As an example, let the alphabet $\mathcal{A}$ be the set $\{a_1, \dots, a_r\}$ and consider an input word $s \in \mathcal{A}^*$. As usual, $s^n = (s_1 \dots s_n)$ is the substring of s of length n and containing its first n symbols.

Consider some step of the coding procedure.

Suppose the first p symbols $(s_1 \dots s_p)$ have already been encoded. The dictionary now contains for example h words $\{e_1, \dots, e_h\}$. Thus, the current position of the front end is the $(p+1)^{th}$ site in the input word and the next word in the dictionary will be labelled as the $(h+1)^{th}$ word e_{h+1}.

The algorithm selects this new word as the longest word which can be obtained by adding a single character $\tilde{a}$ chosen in the alphabet $\mathcal{A}$ to a phrase ρ contained in the substring $(s_1 \dots s_p)$. Hence, the word e_{h+1} has as a prefix the phrase ρ followed by $\tilde{a}$ as an ending symbol $(e_{h+1} = \rho\ \tilde{a})$.

Once the new word e_{h+1} has been found, the algorithm outputs a binary encoding of the triplet $(S_{h+1}, l_{h+1}, \tilde{a})$ where S_{h+1} is the starting position of the prefix ρ of the new word in $(s_1 \ldots s_p)$, l_{h+1} is the length of the new word e_{h+1} and the symbol $\tilde{a}$ from the alphabet is the last character of e_{h+1}.

The following example shows how the algorithm $LZ77$ encodes the input stream

$$s = (aababbbbaababba)$$

Let $\mathcal{A} = \{a, b\}$ be the source alphabet. The output is the binary encoding of the following triplets. The first column is the dictionary index number of the codeword, whose triplet is showed in the same line, second column. For an easier reading, we add a third column which shows each encoded word in the original stream s, but we remark that it is not contained in the output file:

1	$(1, 1, 'a')$	$[a]$
2	$(1, 2, 'b')$	$[ab]$
3	$(2, 3, 'b')$	$[abb]$
4	$(5, 3, 'a')$	$[bba]$
5	$(2, 5, 'a')$	$[ababba]$

Now we recall some results from [KM], concerning the optimality of the $LZ77$ algorithm.

Lemma 6.4 ([KM]). *Let t denote the number of words in which the algorithm $LZ77$ parses the word s^n. Set $m = n - N$ where $N = \max_{i=1,\ldots,r} n_i$, where n_i is the number of occurrences of the symbol $a_i \in \mathcal{A}$. Then $t \leq 2(m+1)$.*

Theorem 6.5 ([KM]). *Let t denote the number of words in which the algorithm $LZ77$ parses the word s^n.*

(i) For all $l \geq 1$ it holds

$$\begin{aligned} I_{LZ77}(s^n) = |LZ77(s^n)| \;\leq\; & n \, \tfrac{\hat{H}_l(s^n)}{l} + 3\, t \log_2\left(\tfrac{n}{t}\right) + \\ & + O\Big((l-1)\, t + t \log_2 \log_2\left(\tfrac{n}{t}\right)\Big) \,. \end{aligned} \tag{6.3}$$

(ii) The algorithm $LZ77$ is optimal and for all $n \geq 1$, for all $s \in \mathcal{A}^n$ and for all $l \geq 1$ it holds

$$\frac{I_{LZ77}(s^n)}{n} \leq \frac{\hat{H}_l(s^n)}{l} + O\left(\frac{\log_2 \log_2(n)}{\log_2(n)} + \frac{l-1}{\log_2(n)}\right). \tag{6.4}$$

(iii) The algorithm $LZ77$ is asymptotically optimal with respect to $\hat{H}_1$ with constant $\lambda = 8$ and for any string s^n such that $\hat{H}_1(s^n) \neq 0$ it holds

$$I_{LZ77}(s^n) \leq \; 8\, n\, \hat{H}_1(s^n) + \; O\Big(t \log_2 \log_2\left(\frac{n}{t}\right)\Big) \,. \tag{6.5}$$

6.1.2 *LZ77* on the Manneville map

Now we are ready to prove the following theorem, which links the information content obtained via the algorithm $LZ77$ to the AIC on the symbolic orbits of the Manneville maps $T_z(x) = x + x^z \pmod 1$ for $x \in [0,1]$ and $z > 1$ (see Section 4.3.1). The importance of this result is that we obtain a result about the behaviour of a compression algorithm on a weakly chaotic system, showing that this algorithm is able to approximate the asymptotic behaviour of the AIC on the symbolic sequences generated by the system. This fact is an important step towards the application of compression algorithms in the characterization of weakly chaotic systems.

In order to state and prove our results, we recall some useful lemmas coming from probability theory.

Lemma 6.6 (Jensen's Inequality). *Let I be a closed interval in $\mathbb{R}$ and $u : I \to \mathbb{R}$ be a convex function, continuous at the endpoints of I. If M is a random variable which takes its values in I, then $\mathbb{E}[u \circ M] \geq u(\mathbb{E}[X])$.*

Lemma 6.7. *If M and N are two real random variables such that $M \geq 0$ and $N \geq 0$, then*

$$\mathbb{E}[\max\{M,N\}] \geq \max\{\mathbb{E}[M], \mathbb{E}[N]\}$$

Theorem 6.8 ([Be]). *Consider the dynamical system $([0,1], \mathcal{T}_z)$ driven by the Manneville map $\mathcal{T}_z(x) = x + x^z (mod\ 1)$, with $z > 1$. Let $\tilde{x} \in (0,1)$ be such that $\mathcal{T}_z(\tilde{x}) = 1$. Consider the partition $\alpha = \{[0,\tilde{x}], (\tilde{x},1]\}$ of the unit interval $[0,1]$. If ω is a symbolic orbit drawn from the Manneville map, with respect to the partition α, then there exist two positive constants c_1, c_2 such that*

$$\begin{aligned} \mathbb{E}[I_{LZ77}(\omega^n)] &\sim n && \text{if } z < 2 \\ c_1\ n^p \leq \mathbb{E}[I_{LZ77}(\omega^n)] &\leq c_2\ n^p \log_2(n) && \text{if } z > 2 \end{aligned} \tag{6.6}$$

for n big enough, where $p = \frac{1}{z-1}$ and the average is made according to the Lebesgue measure on the interval.

Proof. We will study the two cases separately.

If $z < 2$: a result of [GW1] shows that for the expectation value of the AIC of a symbolic orbit of the Manneville map with $z < 2$, with respect to the Lebesgue measure on the interval, it holds that

$$\mathbb{E}[AIC(\omega^n)] \sim n\ .$$

Since it is $AIC(\omega^n) \leq I_{LZ77}(\omega^n) \leq O(n)$, then in this case we have that $\mathbb{E}[I_{LZ77}(\omega^n)] \sim n$.

If $z > 2$: From Theorem 6.5 (*iii*), we know that the algorithm LZ77 is asymptotically optimal with respect to $\hat{H}_1$, that is

$$I_{LZ77}(\omega^n) \leq \; 8\; n\; \hat{H}_1(\omega^n) \; + \; O\Big(t \log_2 \log_2 \left(\frac{n}{t}\right)\Big) \; , \qquad (6.7)$$

where t is the number of words in the LZ77 parsing of ω^n. Thus, for any sequence ω^n, it holds

$$\mathbb{E}[I_{LZ77}(\omega^n)] \leq \; 8\; n\; \mathbb{E}[\hat{H}_1(\omega^n)] \; + \; \mathbb{E}\Big[t \log_2 \log_2 \left(\frac{n}{t}\right)\Big] \; . \qquad (6.8)$$

First, we will prove that $n\; \mathbb{E}[\hat{H}_1(\omega^n)]$ is bounded by $O(n^p \log_2(n))$ with $p = \frac{1}{z-1}$. Then, we will give an estimate for $\mathbb{E}[t \log_2 \log_2(\frac{n}{t})]$, so completing the proof.

By definition, $\hat{H}_1(\omega^n) = -(\frac{N_n}{n} \log_2(\frac{N_n}{n}) + (1 - \frac{N_n}{n}) \log_2(1 - \frac{N_n}{n}))$ where N_n is the defined as the number of occurrences of the event $\mathcal{E} = \{\textit{passage through } (\tilde{x}, 1]\}$.

In [Fel1] it has been proved that if $z > 2$ then $\mathbb{E}[N_n] \sim n^p$.

Therefore, for the first order empirical entropy of a symbolic orbit drawn by the Manneville map with respect to the Lebesgue measure, we can apply Jensen's inequality and obtain:

$$\begin{aligned}
\mathbb{E}[\hat{H}_1(\omega^n)] \;\; & \leq -\left(\mathbb{E}\left[\tfrac{N_n}{n} \log_2\left(\tfrac{N_n}{n}\right)\right] + \mathbb{E}\left[\left(1 - \tfrac{N_n}{n}\right) \log_2\left(1 - \tfrac{N_n}{n}\right)\right]\right) \leq \\
& \leq -\Bigg(\mathbb{E}\left[\tfrac{N_n}{n}\right] \; \log_2\left(\; \mathbb{E}\left[\tfrac{N_n}{n}\right]\right) + \\
& \qquad\qquad + \left(1 - \mathbb{E}\left[\tfrac{N_n}{n}\right]\right) \; \log_2\left(1 - \mathbb{E}\left[\tfrac{N_n}{n}\right]\right)\Bigg) \sim \\
& \sim -\left(n^{p-1} \log_2\left(n^{p-1}\right) + \left(1 - n^{p-1}\right) \log_2\left(1 - n^{p-1}\right)\right) .
\end{aligned}$$

Consequently, we can easily verify that

$$\begin{aligned}
8\; n\; \mathbb{E}[\hat{H}_1(\omega^n)] \;\; & \leq \; - \; 8\; n^p \log_2\left(n^{p-1}\right) - 8\; \left(n - n^p\right) \log_2(1 - n^{p-1}) = \\
& = \; 8\; n^p \log_2\left(n^{1-p} - 1\right) - \; 8\; n\; \log_2\left(1 - n^{p-1}\right) .
\end{aligned}$$

For the fact that $p < 1$, it holds

$$
\begin{aligned}
& 8\ n^p \log_2 \Big(n^{1-p} - 1\Big) - \ 8\ n\ \log_2 \Big(1 - n^{p-1}\Big) \sim \\
\sim\ & n^p \Big((1-p)\log_2(n) - 1\Big)\ = \qquad (6.9) \\
=\ & O(n^p \log_2(n))\ .
\end{aligned}
$$

Now we will prove that, for ω^n a symbolic orbit of the Manneville map with $z > 2$, if t is the number of words in the LZ77 parsing of ω^n, it holds

$$\mathbb{E}\Big[t \log_2 \log_2 \Big(\frac{n}{t}\Big)\Big] = O(n^p \log_2 \log_2(n))\ . \qquad (6.10)$$

We apply Lemma 6.4, with alphabet $\mathcal{A} = \{0, 1\}$ where 1 is the symbol associated to the event $\mathcal{E} = \{$*passage through* $[\tilde{x}, 1]\}$, which appears in ω^n with mean probability $\frac{N_n}{n} = n^{p-1}$ and 0 is the symbol associated to the event *not* $\mathcal{E}$.

Thus, $m = n - \max\{n_0, n_1\}$ and $t \leq 2(m+1) = 2(n - \max\{n_0, n_1\} + 1)$.

Thanks to Lemma 6.7, we obtain the following estimates:

$$
\begin{aligned}
\mathbb{E}[t] &\leq\ 2\ \ (\mathbb{E}[m] + 1) = \\
&=\ 2\ \ (n + 1 - \mathbb{E}[\max\{n_0, n_1\}]) \leq \\
&\leq\ 2\ \ (n + 1 - \max\{\mathbb{E}[n_0], \mathbb{E}[n_1]\}) = \qquad (6.11) \\
&=\ 2\ \ (n + 1 - \max\{n - n^p, n^p\}) = \\
&=\ 2\ \ (n^p + 1)\ .
\end{aligned}
$$

Hence the inequality (6.10) easily follows.

From the estimates (6.9) and (6.10) together with (6.8), it follows that

$$\mathbb{E}[I_{LZ77}(\omega^n)] = O(n^p \log_2(n))\ .$$

Finally, since the AIC is the ideal information content of a string, it is $I_{LZ77}(\omega^n) \geq AIC(\omega^n)$ and the same inequality relates the expectation values.

In Section 4.3.1, it has been proved that $\mathbb{E}[AIC(\omega^n)] \geq c_1\ n^p$. This completes the proof. $\square$

6.1.3 The algorithm *CASToRe*

We have created and implemented a particular compression algorithm we called *CASToRe* which is a modification of the well known $LZ78$ algorithm ([ZL2]). Its theoretical advantages with respect to $LZ78$ are shown in [Ar], [BM]: it is a sensitive measure of the information content of low entropy sequences. That's why it is called *CASToRe*: *C*ompression *A*lgorithm, *S*ensitive *To Re*gularity.

In Section 6.1.4, we show that the information $I_{\mathcal{Z}}$ of a constant word s, originally with length n, is $\Psi(n) = 4 + 2\log(n+1)[\log(\log(n+1)) - 1]$, if the algorithm $\mathcal{Z}$ is *CASToRe*. The theory predicts that the best possible information for a constant word of length n is $AIC(s) = \log(n) + const$. In [Ar], it is shown that the algorithm $LZ78$ encodes a constant n digits long word into a word with length of the order $n^{\frac{1}{2}}$ bits; so, we cannot expect that $LZ78$ is able to distinguish a sequence whose information grows like n^{α} ($\alpha < \frac{1}{2}$) from a constant or periodic one.

On the other hand, we showed that the algorithm $LZ77$ is able to distinguish different behaviours for weakly chaotic systems. In fact in Section 6.2 we show that $LZ77$ and *CASToRe* have the same behaviour on weakly chaotic systems. The two algorithms are different on their internal running and the reason to choose *CASToRe* is based on the computational time it needs.

Now we briefly describe the internal running of *CASToRe*.

As the algorithm $LZ77$, the algorithm *CASToRe* is based on an adaptive dictionary. One of the basic differences in the coding procedure is that the algorithm $LZ77$ splits the input words in overlapping words, while the algorithm *CASToRe* (as already the algorithm $LZ78$) parses the input word in non-overlapping words.

At the beginning of encoding procedure, the dictionary contains only the alphabet. In order to explain the principle of encoding, let's consider a step h within the encoding process, when the dictionary already contains h words $\{e_1, \ldots, e_h\}$.

The new word is defined as a pair (*prefix pointer*,*suffix pointer*). The two pointers are referred to two (not necessarily different) words ρ_p and ρ_s chosen among the ones contained in the current dictionary as follows. First, the algorithm reads the input stream starting from the current position of the front end, looking for the longest word ρ_p matching the stream. Then, we look for the longest word ρ_s such that the joint word $\rho_p\rho_s$ matches the stream. The new word e_{h+1} which will be added to the dictionary is then $e_{h+1} = \rho_p\rho_s$.

The output file contains an ordered sequence of the binary encoding of the pairs (i_p, i_s) such that i_p and i_s are the dictionary index numbers corresponding to the prefix word ρ_p and to the suffix word ρ_s, respectively. The pair (i_p, i_s) is referred to the new encoded word e_{h+1} and has its own index number i_{h+1}.

The following example shows how the algorithm *CASToRe* encodes the input stream

$$s = (abcababccabb)$$

Let the source alphabet be $\mathcal{A} = \{a, b, c\}$.

The output is the binary encoding of the following pairs contained in the second column. The first column is the dictionary index number of the encoded word in the dictionary which is showed in the same line, second column. For an easier reading, we add a third column which shows each encoded word in the original stream s, but it is not contained in the output file:

	First, the dictionary is being loaded	
1	$(0, \, 'a \, ')$	$[a]$
2	$(0, \, 'b \, ')$	$[b]$
3	$(0, \, 'c \, ')$	$[c]$
	Then, the encoding procedure starts	
4	$(1, 2)$	$[ab]$
5	$(3, 4)$	$[cab]$
6	$(4, 3)$	$[abc]$
7	$(5, 3)$	$[cabc]$

We remark that this coding procedure, which pairs words already in the dictionary to create a new word, is similar to the procedure that can be found in the work [Gr2], which seems to be able to give a very precise entropy estimation, detecting very long range correlations in the English language.

6.1.4 *CASToRe* on periodic orbits

Before showing the results of the applications of $LZ77$ and *CASToRe* on some chaotic dynamical systems, we analyze the behaviour of *CASToRe* on periodic orbits.

We first study the case of period $p = 1$ (constant sequences). If we have an input stream given by

$$\omega = (aaaaaaaaaaaaaa.............) \tag{6.12}$$

the algorithm *CASToRe* gives a codification

$$\begin{array}{lcll}
1 & : & (0, a) & [a] \\
2 & : & (1, 1) & [aa] \\
3 & : & (2, 2) & [aaaa] \\
4 & : & (3, 3) & [aaaaaaaa] \\
 & & \dots & \\
k & : & (k-1, k-1) & [a \dots a]
\end{array}$$

As before, the first column shows the dictionary index number, the second column shows the code word and the third column (which is not contained in the output file) shows the corresponding encoded word in the input sequence ω.

So, if we compute the information $I(\omega^n)$[1], as a function of the length n of the sequence to be compressed, we can immediately deduce that $I(\omega^n) \sim \log_2 n$. Indeed a much more accurate estimate is obtained in the following theorem.

Theorem 6.9 ([BM]). *Let ω be a constant sequence as in (6.12), then the information function $I(\omega^n)$ obtained applying the compression algorithm CASToRe is approximated by the function $\Psi(n)$:*

$$\Psi(n) = 4 + 2\log_2(n+1)\left(\log_2(\log_2(n+1)) - 1\right) \tag{6.13}$$

where the approximation is given by the identification of the integer part $[\log_2(\log_2(n))]$ with its real value $\log_2(\log_2(n))$.

Proof. We notice that the algorithm *CASToRe* acts on ω in such a way that at the end of each term of the codification, the length n of the codified word is of the form $n = 2^h - 1$. Then we look for the value of $I(2^h - 1)$. We can easily obtain that if $k = [\log_2 h]$, where $[\cdot]$ denotes the integer part, then

$$\begin{aligned} I(2^h - 1) = \; & 2 + \textstyle\sum_{i=0}^{k-1} 2(2^{i+1} - 2^i)\log_2(2^{i+1}) + 2(n - 2^k)\log_2(2^{k+1}) \\ = \; & 2 + \textstyle\sum_{i=0}^{k-1}(i+1)2^{i+1} + 2(n - 2^k)\log_2(2^{k+1}) \\ = \; & 4 + (k-1)2^{k+1} + 2(n - 2^k)\log_2(2^{k+1}) \end{aligned} \tag{6.14}$$

where we used

$$\sum_{j=1}^{r} j2^j = (r-1)2^{r+1} + 2$$

At this point it is enough to approximate k with its real value $\log_2 h$, and substitute to h its value $\log_2(n+1)$ to obtain equation (6.13). The theorem is proved. □

We now look at sequences ω with prime period $p > 1$. From the point of view of the information content, we expect that periodic strings are not so different from constant strings. Indeed we find that the algorithm *CASToRe* is able to recognize periodic strings as constant strings after having encoded n_p symbols, depending on p. Then, the $I(\omega^n)$ must behave as in the constant case for $n > n_p$, indeed it is given by $\Psi(n)$ plus a constant term C_ω that depends on the particular sequence ω. This term C_ω can be estimated as the amount of information obtained from ω at the n_p-th symbol. We prove the following theorem

[1] Using *CASToRe* as compression algorithm in the definition of information measure.

Theorem 6.10 ([BM]). *Let ω be a periodic sequence with prime period $p > 1$, then there exists a constant n_p, depending on ω, such that, when the algorithm CASToRe reaches the n_p-th term of the sequence, it starts codifying the sequence ω as if it were a constant sequence. The constant n_p has as upper bound*

$$K_p = p^2 \left[2 + \frac{3}{2} \log_2 p + \frac{1}{2} \log_2 |\mathcal{A}| \right] \tag{6.15}$$

where $|\mathcal{A}|$ is the number of symbols in the alphabet used to construct the sequence ω. Note that K_p depends only on the period p.

Proof. Let us assume that $|\mathcal{A}| = r$. Since ω has prime period p, we can look at it as a sequence of subwords, each one p symbols long; they will be called p-subwords.

Once the codification has started and the first p characters have been read, the last encoded word will be t symbols long and will end at some k-th (mod p) site in the second p-subword. Then, after at most $(p+1)$ subsequent reading of contiguous p-subwords, the algorithm will arrive at the end of an already known word; this word will finish at some h-th (mod p) site of the $(p+2)$-th p-subword of s.

Let us assume that the number of words needed by the algorithm to codify the fragment starting from the h-th symbol of a previous p-subword to its correspondent belonging to the $(p+2)$-th p-subword of ω is m; we will indicate the collection of that m words by $\mathcal{F}(h)$. Let us analyse the different cases.

First case: $m = 2^c$. The algorithm continues to process the remaining part of ω, codifying the next fragment $\mathcal{F}(h)$ by coupling two words each time. So, after c steps only *one* word will be necessary and later the algorithm will follow the period.

Second case: $m = 2^{c_1} q$, where q is odd. The algorithm first behaves c_1 times as described above, and after just q words will be needed to process the following $\mathcal{F}(h)$. Let us call $A_1, \ldots, A_q$ those known words. After one step, the algorithm *CASToRe* will have built the new words $B_1 = (A_1, A_2), \ldots, B_{\frac{q-1}{2}} = (A_{q-2}, A_{q-1})$, and $B_{\frac{q-1}{2}+1} = (A_q, B_1)$. Then, the algorithm will reach the period after c_2 steps, where $c_2 = \min\{c \in \mathbb{N} : 2^c > q+1\}$.

Third case: m is odd. The algorithm will behave as in the second case, with $c_1 = 0$.

So we have proved the existence of the constant n_p. Let us estimate its upper bound, K_p. It holds:

$$n_p \leq p(p+2)[1 + C]$$

where $C = \min\{c \in \mathbb{N} : 2^c > m + 1\}$. Now, we estimate m. It does hold $m \leq (p+2) \times S$ where S is the number of words necessary to cover up a p-subword. Since r different symbols are available in the alphabet $\mathcal{A}$, then it must be $S \leq N$ where N is the smaller integer such that it holds $p \leq \sum_{i=1}^{N} ir^i$.

Since we have

$$p \leq \frac{r-(N+1)r^{N+1}+Nr^{N+2}}{(r-1)^2} = \sum_{i=1}^{N} ir^i$$

and we are interested in an estimation of N, let us look for the smallest integer N such that

$$\log_r\left(\frac{p(r-1)^2}{r}-1\right) \leq 2\ \log_r N$$

Let N_0 the real solution of the following equation:

$$N_0 = \sqrt{\frac{(r-1)^2 p}{r}-1}$$

Then the integer N we looked for is $O(\sqrt{rp})$ and it holds $m \leq (p+2)\sqrt{rp}$. So, we have $C \leq 1+\log_2(m+1)$. Finally, if we approximate $p+2$ with p, we obtain

$$K_p \leq p^2\left[2+\frac{3}{2}\log_2 p+\frac{1}{2}\log_2 r\right]$$

and the theorem is proved. □

6.2 Results

In this section we show some numerical experiments. We consider some examples of fully chaotic and weakly chaotic dynamical systems and we measure the information content of the generated symbolic orbits with respect to some partition.

To measure the information content we used the two different data compression algorithms *LZ77* with infinite window and *CASToRe*, whose internal running has been presented in the previous section. These two algorithms seems to be suitable for the compression of null entropy sequences, and they are fast enough to allow the compression of long sequences (we managed to compress trajectories of $O(10^7)$ symbols). From the computational point of view, whereas *LZ77* requires a big amount of RAM (Random Access Memory), since it needs to retain the entire word already encoded and the entire dictionary built, the algorithm *CASToRe* only remembers the dictionary which is implemented in a tree structure. Hence, the computation time is a distinguishing feature between the two algorithms: *CASToRe* can compress a $O(10^7)$-symbols word in few seconds.

As we will see the results agree with theoretical predictions when they are available or with other numerical results which can be found in the literature.

It is worth to remark that, even if the two compression schemes are basically different, the two algorithms give a behavior of the information content of the same order in all the numerical experiments we performed.

6.2.1 Chaotic systems

The Arnold cat map [Be]

The Arnold cat map is an example of a two-dimensional hyperbolic toral automorphism, that is the projection on the two-torus $\mathbb{R}^2/\mathbb{Z}^2$ of a linear map of $\mathbb{R}^2$, represented by a matrix M with integer elements and determinant one and real eigenvalues λ and $1/\lambda$, different from 1. The Arnold cat map is specified by the matrix

$$M = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

with $\lambda = (3+\sqrt{2})/2$. It follows that the Kolmogorov-Sinai entropy, with respect to the Lebesgue measure m_2, of a two-dimensional hyperbolic toral automorphism is given by the logarithm of the modulus of the eigenvalue bigger than 1 (see [Arn]). Hence, in this case

$$h_{m_2} = \log_2 \frac{3+\sqrt{2}}{2} \sim 1.388\ldots$$

Our computations give the same result. We only show the results for the compression algorithm *CASToRe*, since in this case no evident differences can be appreciated. Studying the behaviour of the information content with respect to the length of the compressed word, we expect to find a straight line with angular coefficient equal to the entropy of the dynamical system, when working with a generating partition. In Figure 6.1 it is represented the information function for three different choices of the partition. The dotted line is obtained with a partition of the square $[0,1) \times [0,1)$ in two horizontal strips. The solid line is obtained with a partition in four equal squares along the axis, and the dashed line is obtained with a partition in four squares along the eigendirections associated to the matrix M. The angular coefficients of the lines are 0.98 for the dotted line, 1.56 for the solid line and 1.37 for the dashed line, showing that the first two partitions considered are not able to simulate the whole complexity of the system, whereas the last one can be considered to be a generating partition.

The Froeschlé map [Be]

This map was studied in [Fr] as an example of a symplectic map for which it is possible to find the integrable and non-integrable initial conditions. The

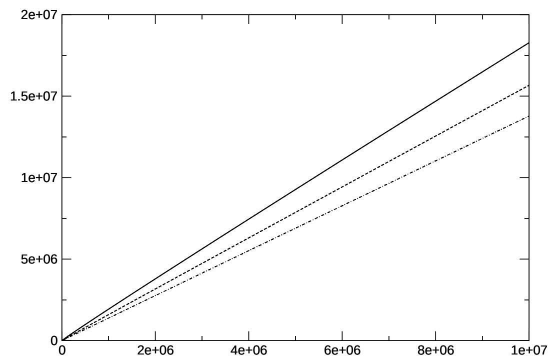

Figure 6.1: *The information content of the Arnold cat map for three different partitions. The dashed line is the information content of the generating partition. For the partitions used see the description in the text.*

map is defined on the two-torus $\mathbb{R}^2/(2\pi\mathbb{Z})^2$ by

$$\begin{aligned} x_{n+1} &= x_n + a \sin y_n \\ y_{n+1} &= x_n + y_n + a \sin y_n \end{aligned}$$

with $a = 1.3$. We studied the behaviour of the information content for orbits generated by two different initial conditions, one corresponding to the regular zone, and the other to the irregular zone. Both orbits have been studied with *CASToRe* and with *LZ77*. In the regular zone, the initial condition is given by the point $(0, 2.5)$. From the results in Figure 6.2 (dotted curves), one can see that both the compression algorithms give indication of regularity by an increasing of the information content of the order of a logarithm (this behaviour is visible clearly in the inset in the two figures, where the information content is plotted on a log-linear scale). The two compression algorithms also give a strong indication of full chaos for the irregular orbit, generated by the initial condition $(2, 0)$. In Figure 6.2, the information content is a straight line (solid and dotted lines, with partitions in two vertical strips and in four equal squares) whose angular coefficients give an indication of the value 0.40 and 0.44, respectively, of the Kolmogorov-Sinai entropy with respect to the measure associated to the initial point. These results are in agreement with those of [CP].

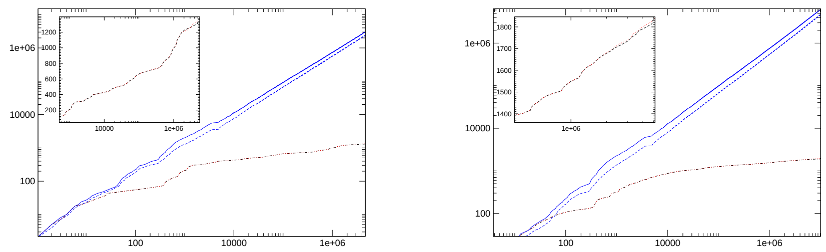

Figure 6.2: *Information content for the Froeschlé map. On the left the results are obtained using CASToRe, on the right using LZ77. In both the pictures, the solid and dashed lines are for the full chaotic orbit, and the dotted curves are for the regular orbit. In the insets only the regular orbit is plotted in a log-linear scale.*

6.2.2 Weakly chaotic systems

The logistic map at the chaos threshold [BM]

The plan of is to apply *CASToRe* to symbolic words generated by the logistic map T_λ (see Section 4.2) for different values of λ that approximate the chaos threshold λ_∞. In particular we consider two approximating sequences of values of λ: the sequence of values $\lambda_j < \lambda_\infty$ at which the period doubling bifurcations occur and the sequence of values $\mu_j > \lambda_\infty$ at which the inverse tangent bifurcations occur. Thus we obtain approximations for the information content at the chaos threshold from below and from above, and the two coincide. We can then say to have obtained an approximation for the information content computed with *CASToRe* for the logistic map at the chaos threshold T_{λ_∞}.

We remark that the symbolic words are obtained from the orbits of the logistic map considering the partition $Z = \{I_0, I_1\}$ of $[0,1]$ given by $I_0 = [0, 1/2)$ and $I_1 = [1/2, 1]$.

In Section 6.1.4 we have analysed the behaviour of *CASToRe* on periodic sequences, obtaining that the information content $I(\omega^n)$ of a periodic sequence ω behaves with respect to the length n of the encoded word as

$$I(\omega^n) = \Psi(n) + C_s,$$

where $\Psi(n)$ is given by equation (6.13) and C_s is a constant depending on the sequence ω (Theorem 6.10).

We start considering the logistic map for values of the parameter λ in the sequence $(\lambda_j)_{j\in\mathbb{N}}$ given by the period doubling bifurcations, for which values the logistic map is periodic. The sequence $(\lambda_j)_{j\in\mathbb{N}}$ converges to the chaos threshold value λ_∞ from below and can be generated by the equation (4.19) (Section 4.2).

We generated the terms of the sequence $(\lambda_j)_{j\in\mathbb{N}}$ (where the differences are given up to 38 decimal digits) and iterated the logistic map 10^7 times. So we had words of 10^7 symbols. When studying $I_j(\omega^n)$ for each λ_j, where ω^n is the symbolic orbit of the logistic map with parameter $\lambda = \lambda_j$, we first tried to obtain the constant terms C_j. We have obtained an increasing sequence of values C_j (for j that tends to infinity), and our results about the compression of periodic words by the algorithm *CASToRe* (see Theorem 6.10) are supported by the computations of the following limits

$$\lim_{n\to\infty} \frac{I_j(\omega^n) - C_j}{\Psi(n)} = 1 \quad \forall j \in \mathbb{N} \tag{6.16}$$

As expected, equation (6.16) tells us that, after the compression of a long enough subword of our sequence ω, we obtain the same information function as for constant sequences, simply translated by a constant. If we denote by n_j the number of symbols that the algorithm has to process before reaching that behaviour, we have $n_j \to \infty$ as $j \to \infty$. So we have that $C_j \to \infty$ as $j \to \infty$.

The approximation of the information content $I_j(\omega^n)$ with a function $\Psi(n)$ of the form given by (6.13) plus a constant term is not very good for periodic sequences with high period. This is because of the poorness of the upper bound of our computations: indeed, we know that this approximation is accurate for values of $n > n_j$ and there are sequences whose n_j is bigger than 10^7, the length of our orbits. In our case the period p_j of the sequences ω_j is already big enough for $j = 12$, indeed $p_{12} = 2^{12} = 4098$, so $n_{12} \approx 16 \times 10^6$. This means that we have to look for another way of approximation to the information function for values of n lower than n_j. At the bifurcation points it is well known from experimental results that the orbit separation is polynomial and, thanks to the results in [Ga3], we can say that the *AIC* is at most logarithmic. So, we can expect the information content to behave like

$$I_j(\omega^n) \sim \begin{cases} S_j(n)\Psi(n) & \text{for } n < n_j \\ C_j + \Psi(n) & \text{for } n > n_j \end{cases} \qquad j \in \mathbb{N} \tag{6.17}$$

with C_j as given before. To find the functions $S_j(n)$, we simply have to compute the fraction

$$S_j(n) = \frac{I_j(\omega^n)}{\Psi(n)}, \quad j \in \mathbb{N} \; . \tag{6.18}$$

First of all, we deduce from equation (6.16) that $\lim_{n\to\infty} S_j(n) = 1$ for all j. Moreover the numerical experiments show that the sequence of functions

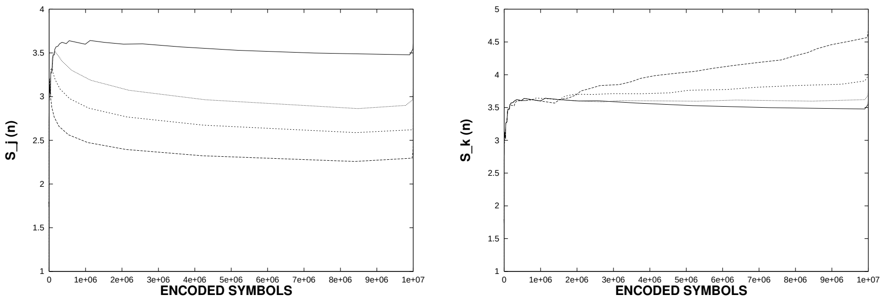

Figure 6.3: *On the left one can see the behaviour of some of the functions $S_j(n)$ (dashed lines), that shows the monotonicity of these functions. The solid line is the function $S_\infty(n)$ and one can see that it is an upper bound for the functions $S_j(n)$. On the right there is the same picture but for the functions $S_k(n)$ that approximate $S_\infty(n)$ from above.*

$(S_j(n))_{j\in\mathbb{N}}$ is pointwise increasing in j and bounded from above (see Figure 6.3 on the left). Then we define the bounding function $S_\infty(n)$ as

$$S_\infty(n) = \lim_{j\to\infty} \Big(S_j(n)\chi_{[0,n_j]}(n) + \chi_{[n_j,\infty]}(n) \Big)$$

for each $n \in \mathbb{N}$, where $\chi_{[a,b]}(n)$ denotes the characteristic function of the real interval $[a, b]$.

In order to establish the behaviour of the information function $I_\infty(\omega^n)$ at the chaos threshold, we applied the same method as before, building up a sequence $(\mu_k)_{k\in\mathbb{N}}$ of parameters approximating λ_∞ from above.

In the interval $[\lambda_\infty, 4]$, the logistic map has a general chaotic behaviour, except from narrow ranges (called *periodic windows*) of parameters for which the map is periodic. Inside each window, a new period doubling sequence can be identified, that leads the map to chaos. The behaviour becomes periodic from chaotic via the unfolding of a saddle-node (see Section 4.3.2). The values of the parameter at which the saddle-nodes occur converge to λ_∞ from above and can be generated using the following relation:

$$\mu_k = \lambda_\infty + \frac{c}{\delta^k}$$

where c is a suitable constant and δ is the Feigenbaum constant.

Using the same definition of $S_k(n)$ given in equation (6.18), we find a sequence of functions such that $\lim_{n\to\infty} S_k(n) = +\infty$ for all $k \in \mathbb{N}$, due to the chaotic behaviour of the map at μ_k.

As we can see in Figure 6.3 on the right, the numerical evidence is that the sequence $(S_k(n))_{k\in\mathbb{N}}$ is point-wise decreasing in k and the same function $S_\infty(n)$ found in the previous subsection is a lower bound for the sequence. Then we have

$$\lim_{k\to\infty} S_k(n) = S_\infty(n)$$

for each n.

We can now draw some conclusions. First of all, we can say that for all $n \in \mathbb{N}$ the information function $I_j(\omega^n)$ of the logistic map with parameter $\lambda = \lambda_j$ is

$$I_j(\omega^n) \sim S_j(n)\Psi(n)\chi_{[0,n_j]}(n) + (\Psi(n) + C_j)\chi_{[n_j,\infty]}(n) \quad \forall\, j \in \mathbb{N}.$$

We recall that n_j is the number of symbols which the algorithm *CASToRe* has to encode in order to recognize the periodicity of the j-th symbolic orbit.

As it has been showed before, until the first n_j symbols have been processed, the algorithm only identifies an at most logarithmic information content; later, the behaviour of the information function follows the results of Theorem 6.9.

For what concerns the information function $I_k(\omega^n)$, when the parameter is $\lambda = \mu_k$, we simply have that

$$I_k(s^n) \sim S_k(n)\Psi(n) \quad \forall\, k \in \mathbb{N}$$

Moreover, the two limits from above and from below give that

$$\lim_{j\to\infty} I_j(\omega^n) = \lim_{k\to\infty} I_k(\omega^n) = I_\infty(\omega^n) = S_\infty(n)\Psi(n) \quad \forall\, n \in \mathbb{N} \qquad (6.19)$$

From the numerical approximation of $S_\infty(n)$ from above and below, we conjecture that $S_\infty(n)$ is an increasing bounded function of n and there exists a constant $S_\infty = \lim_{n\to\infty} S_\infty(n)$. Numerical estimates give a value of S_∞ of more or less 3.5.

Finally, we make an important remark: the feature (6.19) of $I_\infty(\omega^n)$ (and in particular of $S_\infty(n)$) is typical of the logistic map and is not simply due to the fact that we are considering periodic orbits with period going to infinity. Indeed, from the behaviour of the function $S_\infty(n)$, we deduce that $\max_n S_\infty(n) < 3.7$. From (6.19), it follows that for $n = 10^3$, for example, $I_\infty(\omega^n)$ is less than approximately 148. But we can construct periodic sequences with period less than 100, such that the information function after 10^3 symbols is greater than 148. We can consider, for example, a sequence σ with period given by all the possible combinations of two symbols in pieces of length from 1 to 4. This sequence σ has period $98 << 2^7$, but the algorithm *CASToRe* for $n = 734$ gives $I(\sigma^n) = 356$, much greater than 148.

The Manneville maps [Be]

We now use the compression algorithms to measure the information content of symbolic orbits drawn from the Manneville maps $T_z(x) = x + x^z (mod\ 1)$ for different values of the parameter z (see Section 4.3.1).

Let us consider the partitions Z_1 and Z_2 of the interval $X = [0,1]$ in 2 or 4 subintervals. The partition Z_1 is given by the two intervals $I_0 = [0, \tilde{x}_z)$

and $I_1 = [\tilde{x}_z, 1)$, where $\tilde{x}_z$ is the solution of $T_z(x) = 1$. The partition Z_2 is a refinement of Z_1: Z_2 is obtained splitting in two equal parts each interval of Z_1. We denote by $AIC(x, Z_i, n), i = 1, 2$ the AIC of a n-long symbolic orbit of the Manneville maps with initial condition x, with respect to the partition Z_i.

By the results exposed in Section 4.3.1, we have that the mean value of $AIC(x, Z_i, n)$, with respect to the Lebesgue measure, on the initial conditions of the orbit is expected to be $\mathbb{E}[AIC(x, Z_i, n)] \sim n^p$, with $p = \frac{1}{z-1}$ for $z > 2$, and $\mathbb{E}[AIC(x, Z_i, n)] \sim n$ for $z < 2$. Moreover (by the results of Section 6.1.2) the same result holds for the information as it is measured by the algorithm $LZ77$. We verified numerically this statement, and the result is shown in Figure 6.4.

If we use the information content as it is measured by *CASToRe* the numerical result is also close to the previous one. This confirms the theoretical results and proves that the methods relative to the use of compression algorithms to estimate the information content are experimentally reliable.

We considered a set of one hundred initial points, then we generated the relative 10^7-long orbits and we applied the compression algorithms to the associated symbolic strings s (with respect both to the partition $\mathbb{Z}_1$ and Z_2).

In Table 6.1 we show the results. The first column indicates the partition to which the results are referred. The second column is the value of the parameter z which drives the dynamics of the system. The last column gives the results of the theory for the exponent p of the asymptotic behaviour of $AIC(x, Z_i, n) \sim n^p$. The third and the fourth columns show the experimental results. The shown number is the average $\bar{p}$ of the exponents of one hundred different orbits. The initial conditions of the orbits are chosen randomly with respect to the Lebesgue measure.

Symbols	z	*LZ77*	*CASToRe*	Theoretical value
4	2.5	0.64	0.64	0.66
2	2.5	0.64	0.64	0.66
4	3	0.49	0.43	0.5
2	3	0.47	0.48	0.5
4	4	0.27	0.25	0.33
2	4	0.32	0.28	0.33

Table 6.1: *Theoretical and experimental results for the information content of the Manneville maps*

In Figure 6.4 are plotted several examples of the behavior of the information content $I_{\mathcal{Z}}$ when $\mathcal{Z} = LZ77$ (on the right) or $\mathcal{Z} = CASToRe$ (on the left) and for different values of the parameter z. The scale is bi-logarithmic, so that the power laws become straight lines and the exponent p of the expected power law is the slope of the correspondent straight line.

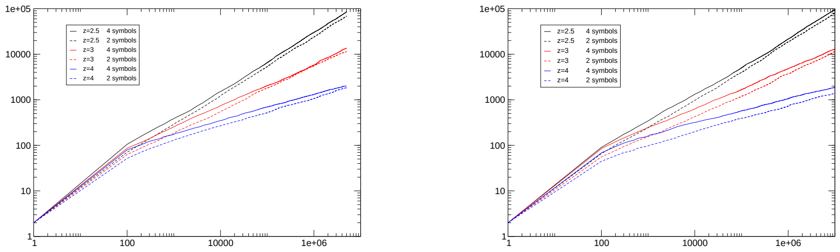

Figure 6.4: *The plot of the information content for the Manneville maps: each graph represents the behaviour of the information plotted versus the number of steps (in log-log scale). Note that in log-log scale power laws become straight lines. On the left we plot the information as it is measured by CASToRe, on the right we plot the information as it is measured by LZ77.*

Intermittency in the logistic map

In Section 4.3.2 we showed some results about the behaviour of unimodal maps near the unfolding of a *saddle-node*. Theorem 4.35 studies the asymptotic behaviour of the relative frequency Φ_U of visit to a neighbourhood U of a neutrally stable periodic point. As it appears clear from the approach to intermittency in the Manneville maps, the information content of symbolic orbits of an intermittent system is related to the number of passages of the orbits from the *laminar* region of the system, that is the neighbourhood of the neutrally stable fixed point. Hence if we study the behaviour of the *AIC* for the logistic map near the unfolding of a saddle-node, we expect to find a relation with the behaviour of the function Φ_U. We now show the numerical experiments.

We considered the logistic map $T_\lambda(x) = \lambda x(1-x)$, on the unit interval $X = [0,1]$, near the saddle-node bifurcation that generates the period 3 window. The bifurcation happens at the value $\bar\lambda = 3.82842676\dots$. For values $\lambda < \bar\lambda$ the map is chaotic with positive KS entropy with respect to a measure μ absolutely continuous with respect to m, and for $\lambda > \bar\lambda$ it is periodic. We studied the behaviour for values of λ approaching $\bar\lambda$ from below.

We used the compression algorithm *CASToRe* to approximate the *AIC* of the orbits. The input words are 10^6 symbols long on the alphabet $\mathcal{A} = \{0,1\}$, obtained by the partition $Z = \{[0,1/2),[1/2,1]\}$ of X.

By the results of Theorem 4.35, we expect that, for $\lambda < \bar\lambda$, the complexity $K(\omega,\lambda)$ (see Definition 2.2) of the symbolic orbits $\omega \in \mathcal{A}^{\mathbb{N}}$ generated by T_λ is positive, and it converges to 0 as λ approaches $\bar\lambda$. Moreover, we expect that the convergence has the same asymptotic behaviour of the convergence of $\Phi_U(\lambda)$ to 1. From Theorem 4.35, we have $\Phi_U(\lambda) \sim 1 - C\sqrt{\lambda}$, for a positive constant C.

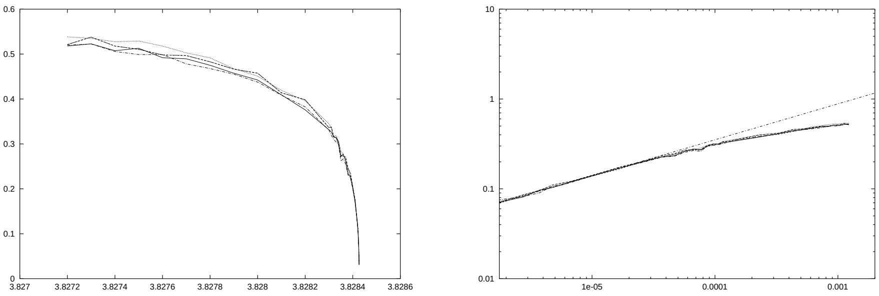

Figure 6.5: *The behaviour of the complexity for the logistic map near the saddle-node bifurcation that generates the period 3 window. As the parameter* λ *approaches the value of the bifurcation, the complexity decreases to* 0 *as a power law with exponent* 0.4*.*

In Figure 6.5 on the left, it is plotted the complexity of five different symbolic orbits. On the x axis there is the value of the parameter λ, and on the y axis the quantity $\frac{I(\omega^n)}{n}$, the approximation of the complexity, where $n = 10^6$ and $I(\omega^n)$ is measured using *CASToRe*. It is evident the convergence of the complexity to 0. To study the order of this convergence we plotted on the right the same data in bi-logarithmic scale, with on the x axis the value $|\lambda - \bar{\lambda}|$. The straight line is the function $f(\lambda) = 14\ \lambda^{0.4}$. Hence we find a good agreement with the expected results.

6.2.3 Randomly perturbed dynamical systems

In Chapter 5 we have proved that the Algorithmic Information Content can be used for randomly perturbed dynamical systems to obtain information on the size of the random perturbation.

Using *CASToRe* we have analyzed an example of randomly perturbed dynamical system. We have chosen the logistic map

$$T : x \mapsto \lambda x(1-x)$$

on the interval $[0,1]$ with $\lambda = 4$. To this map we have added a dynamical noise given by independent random variables $\{w_n\}$, uniformly distributed on the interval $[-\sigma, \sigma]$.

Keeping σ fixed we have varied the diameter of the partition from 0.5 to 0.004=1/250, and we have computed the complexity of the 10^6-long symbolic orbits of the perturbed system relative to the different partitions. For the interpretation of the results, we refer to Section 5.2.3.

In Figure 6.6 we show the results. It is a plot of the complexity of the symbolic orbits versus the diameter of the partition, plotted on the x axis.

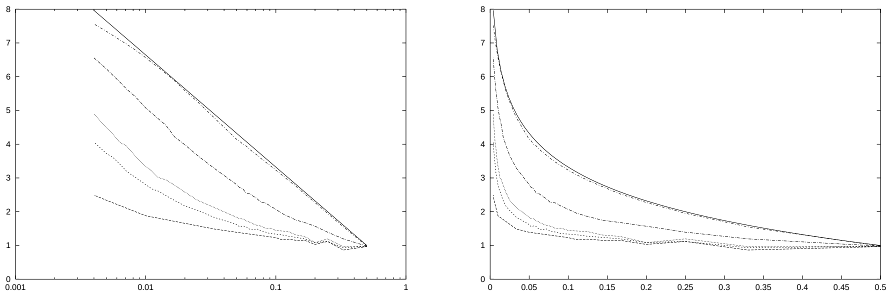

Figure 6.6: *The complexity for different values of σ with respect to the diameter of the partition.*

On the left there is the log-linear version, whereas on the right there is the bi-linear version. The solid line is the upper bound $-\log \epsilon$, and the other curves are the empirical curves which correspond to the following values of σ: 0.5, 0.1, 0.02, 0.01, 0.001 from the upper to the lower, respectively. It is clear from the figure that when the diameter of the partition is comparable to σ there is a change in the slope of the curves. This change agrees with the theoretical result in equations (5.8) and (5.9).

When the diameter of the partition is high with respect to the size of the noise, then we have a good approximation of the unperturbed KS entropy of the system, since the partition with diameter 0.5 is generating. In Figure 6.6 it is possible to see that the complexity of the symbolic orbits for partitions with diameter close to 0.5 is close to 1 (the KS entropy[2] of the logistic map) for the empirical curves with $\sigma < 0.1$.

[2]The KS entropy is $\log 2$, but we use logarithms in base 2 since we measure binary information contents.

Bibliography

[Aa] J.Aaronson, "An Introduction to Infinite Ergodic Theory", American Mathematical Society, 1997

[Ab] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, L.S. Tsimring, *The analysis of observed chaotic data in physical systems*, Rev. Mod. Phys. **65** (1993), 1331–1392

[Al1] P.Allegrini, M.Barbi, P.Grigolini, B.J.West, *Dynamical model for DNA sequences*, Phys. Rev. E **52** (1995), 5281–5297

[Al2] P.Allegrini, P.Grigolini, B.J.West, *A dynamical approach to DNA sequences*, Phys. Lett. A **211** (1996), 217–222

[ALP] J.F.Alves, S.Luzzatto, V.Pinheiro, *Markov structures and decay of correlations for non-uniformly expanding dynamical systems*, math.DS/0205191 (2002)

[Ar] F.Argenti, V.Benci, P.Cerrai, A.Cordelli, S.Galatolo, G.Menconi, *Information and dynamical systems: a concrete measurement on sporadic dynamics*, Chaos Solitons Fractals, **13** (2002), 461–469

[Arg] J.Argyris, I.Andreadis, *On the influence of noise on the largest Lyapunov exponent of attractors of stochastic dynamic systems*, Chaos Solitons Fractals **9** (1998), 959–963

[Arn] V.I.Arnold, A.Avez, "Ergodic problems of classical mechanics", Benjamin, 1968

[Ba] V.Baladi, "Positive Transfer Operators and Decay of Correlations", World Scientific, 2000

[Be] V.Benci, C.Bonanno, S.Galatolo, G.Menconi, M.Virgilio, *Dynamical systems and computable information*, Disc. Cont. Dyn. Syst. - B, to appear (2003)

[Bl] F.Blume, *Possible rates of entropy convergence*, Ergod. Th. Dyn. Syst. **17** (1997), 45–70

[Bo1] C.Bonanno, *The Manneville map: topological, metric and algorithmic entropy*, math.DS/0107195 (2001)

[Bo2] C.Bonanno, *The Algorithmic Information Content for randomly perturbed systems*, publications of Dipartimento di Matematica, University of Pisa (2002)

[BG] C.Bonanno, S.Galatolo, *Orbit complexity for the Manneville maps*, work in preparation

[BGI] C.Bonanno, S.Galatolo, S.Isola, *Recurrence and algorithmic information*, work in preparation

[BM] C.Bonanno, G.Menconi, *Computational information for the logistic map at the chaos threshold*, Disc. Cont. Dyn. Syst. - B **2** (2002), 415–431

[BK] M.Brin, A.Katok, *On local entropy*, in "Geometric Dynamics", Lecture Notes in Mathematics **1007**, 30–38, Springer-Verlag, 1983

[Br] A.A.Brudno, *Entropy and the complexity of the trajectories of a dynamical system*, Trans. Moscow Math. Soc. **2** (1983), 127–151

[BLS] H.Bruin, S.Luzzatto, S.van Strien, *Decay of correlations in one-dimensional dynamics*, math.DS/0208114 (2002)

[BGP] M.Buiatti, P.Grigolini, L.Palatella, *Nonextensive approach to the entropy of symbolic sequences*, Physica A **268** (1999), 214–224

[CI1] M.Campanino, S.Isola, *Statistical properties of long return times in type I intermittency*, Forum Math. **7** (1995), 331–348

[CI2] M. Campanino, S. Isola, *Infinite invariant measures for non-uniformly expanding transformations of* $[0,1]$*: weak law of large numbers with anomalous scaling*, Forum Math. **8** (1996), 71–92

[CI3] M. Campanino, S. Isola, *On the invariance principle for non-uniformly expanding transformations of* $[0,1]$, Forum Math. **8** (1996), 475–484

[Ce] M.Cencini, M.Falcioni, E.Olbrich, H.Kantz, A.Vulpiani, *Chaos or noise: difficulties of a distinction*, Phys. Rev. E **62** (2000), 427–437

[Cha] G.J.Chaitin, "Information, Randomness and Incompleteness. Papers on Algorithmic Information Theory", World Scientific, 1987

[Chu] K.L.Chung, "Markov chains with stationary transition probabilities", Springer Verlag, Berlin-New York, 1967

[CP] A.Cohen, I.Procaccia, *Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems*, Phys. Rev. A **31** (1985), 1872–1882

[CE] P.Collet, J.P.Eckmann, "Iterated Maps on The Interval as Dynamical Systems", Birkhäuser, 1980

[CF] P.Collet, P.Ferrero, *Some limit ratio theorems related to a real endomorphism in case of a neutral fixed point*, Ann. Inst. H. Poinc. **52** (1990), 283–301

[CGS] P.Collet, A.Galves, B.Schmitt, *Unpredictability of the occurrence time of a long laminar period in a model of temporal intermittency*, Ann. Inst. H. Poinc. **57** (1992), 319–331

[CFS] I.P.Cornfeld, S.V.Fomin, Ya.G.Sinai, "Ergodic Theory", Springer-Verlag, 1982

[DS] W.de Melo, S.van Strien, "One-Dimensional Dynamics", Springer-Verlag, 1993

[De] R.L.Devaney, "An Introduction to Chaotic Dynamical Systems", Addison Wesley, 1987

[ER] J.P.Eckmann, D.Ruelle, *Ergodic theory of chaos and strange attractors*, Rev. Mod. Phys. **57** (1985), no. 3, 617–656

[FSS] V.V. Fedorenko, A.N. Sarkovskii, J. Smital, *Characterization of weakly chaotic maps of the interval*, Proc. Am. Math. Soc. **110** (1990), 141–148

[Fei] M.J.Feigenbaum, *Quantitative universality for a class of nonlinear transformation*, J. Stat. Phys. **19** (1978), 25–52

[Fel1] W.Feller, *Fluctuation theory of recurrent events*, Trans. Am. Math. Soc. **67** (1949), 98–119

[Fel2] W.Feller, "An introduction to probability theory and its applications" vol. 1, Wiley, New York, 1971

[Fr] C.Froeschlé, *A Numerical Study of the Stochasticity of Dynamical Systems with two Degrees of Freedom*, Astron. Astrophys. **9** (1970), 15–23

[Ga1] S.Galatolo, "Information, Complexity and Weak Chaos in Dynamical Systems. Theory and Measurement Methods", Ph.D. Dissertation, University of Pisa, 2001

[Ga2] S.Galatolo, *Orbit complexity by computable structures*, Nonlinearity **13** (2000), no. 5, 1531–1546

[Ga3] S.Galatolo, *Complexity, initial condition sensitivity, dimension and weak chaos in dynamical systems*, Nonlinearity **16** (2003), 1219–1238

[GW1] P.Gaspard, X.J.Wang, *Sporadicity: Between periodic and chaotic dynamical behaviors*, Proc. Natl. Acad. Sci. USA **85** (1988), 4591–4595

[GW2] P.Gaspard, X.J.Wang, *Noise, chaos, and* (ϵ,τ)*-entropy per unit time*, Phys. Rep. **235** (1993), 291–343

[Go] S.Gouëzel, *Sharp polynomial estimates for the decay of correlations*, math.DS/0202147 (2002)

[Gr1] P.Grassberger, *Toward a quantitative theory of self-generated complexity*, Int. J. Theor. Phys. **25** (1986), 907–938

[Gr2] P.Grassberger, *Data compression and entropy estimates by non-sequential recursive pair substitution*, physics/0207023 (2002)

[GP1] P.Grassberger, I.Procaccia, *Estimation of the Kolmogorov entropy from a chaotic signal*, Phys. Rev. A **28** (1983), 2591–2593

[GP2] P.Grassberger, I.Procaccia, *Characterization of strange attractors*, Phys. Rev. Lett. **50** (1983), 346–349

[HY] A.J.Homburg, T.Young, *Intermittency in families of unimodal maps*, Ergod. Th. Dyn. Syst. **22** (2002), 203–225

[Is1] S.Isola, *Renewal sequences and intermittency*, J. Stat. Phys. **97** (1999), 263–280

[Is2] S.Isola, *On systems with finite ergodic degree*, mp_arc 00-466 (2000)

[IP] S.Isola, A.Politi, *Universal encoding for unimodal maps*, J. Stat. Phys. **61** (1990), 263–291

[Ka] A.G.Kachurovskii, *Rates of convergence in ergodic theorems*, Russian Math. Surv. **51** (1996), 653–703

[KH] A.Katok, B.Hasselblatt, “Introduction to the Modern Theory of Dynamical Systems,” Cambridge University Press, 1995

[Ki] Y.Kifer, “Random perturbations of dynamical systems”, Birkhäuser, 1988

[Kol] A.N.Kolmogorov, *Three approaches to the quantitative definition of information*, Problems of Information Transmission **1** (1965), 1–7

[KM] S.R.Kosaraju, G.Manzini, *Compression of low entropy strings with Lempel-Ziv algorithms*, SIAM J. Comput. **29** (2000), 893–911

[Kr1] U.Krengel, *Entropy of conservative transformations*, Z. Wahrs. verw. Geb. **7** (1967), 161–181

[Kr2] U.Krengel, *On the speed of convergence in the ergodic theorem*, Monatsh. Math. **86** (1978), 3–6

[Krie] W.Krieger, *On entropy and generators of measure-preserving transformations*, Trans. Amer. Math. Soc. **149** (1970), 453–464

[LV] M.Li, P.Vitanyi, "An Introduction to Kolmogorov Complexity and Its Applications", Springer-Verlag, 1993

[Li] C.Liverani, *Rigorous numerical investigation of the statistical properties of piecewise expanding maps. A feasability study*, Nonlinearity **14** (2001), 463–490

[LSV] C.Liverani, B.Saussol, S.Vaienti, *A probabilistic approach to intermittency*, Ergod. Th. Dyn. Syst. **19** (1999), 671–685

[Ma] R.Mañé, "Ergodic Theory and Differentiable Dynamics", Springer-Verlag, 1987

[Mann] P.Manneville, *Intermittency, self-similarity and $1/f$ spectrum in dissipative dynamical systems*, J. Physique **41** (1980), 1235–1243

[Or] D.S.Ornstein, *Bernoulli shifts with the same entropy are isomorphic*, Adv. in Math. **4** (1970), 337–352

[Pa] W.Parry, *Intrinsic Markov Chains*, Trans. Am. Math. Soc. **112** (1964), 55–66

[Pet] K.Petersen, "Ergodic Theory", Cambridge University Press, 1983

[RS] A.M.Rockett, P.Szüsz, "Continued Fractions", World Scientific, 1992

[Sa] O.Sarig, *Subexponential decay of correlations*, Inv. Math. **150** (2002), 629–653

[STV] B.Saussol, S.Troubetzkoy, S.Vaienti, *Recurrence, dimensions and Lyapunov exponents*, J. Stat. Phys. **106** (2002), 623–634

[SD] C.Schittenkopf, G.Deco, *Identification of deterministic chaos by an information theoretic measure of the sensitive dependence on the initial conditions*, Physica D **110** (1997), 173–181

[SK] T.Schreiber, H.Kantz, *Noise in chaotic data: diagnosis and treatment*, Chaos **5** (1995), 133–142

[Sh] C.E.Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423, 623–656

[Th] M.Thaler, *Transformations on* $[0,1]$ *with infinite invariant measures*, Israel J. Math. **46** (1983), 67–96

[Ti] E.C.Titchmarsh, "The theory of functions", Oxford University Press, 1939

[TL] T.I.Toth, L.S.Liebovitch, *Models of ion channels kinetics with chaotic subthreshold behaviour*, Z. Angew. Math. Mech. **76** (1996), 523–524

[TPZ] C.Tsallis, A.R.Plastino, W.M.Zheng, *Power-law sensitivity to initial conditions – new entropic representation*, Chaos Solitons Fractals **8** (1987), 885–891

[Vi] M.Viana, "Stochastic Dynamics of Deterministic Systems", Col. Bras. Mat. notes, 1997

[W1] H.White, "On the Algorithmic Complexity of Trajectories of Points in Dynamical Systems", Ph.D. dissertation Univ. of North Carolina at Chapel Hill, 1991

[W2] H.White, *Algorithmic complexity of points in a dynamical system*, Ergod. Th. Dyn. Syst. **13** (1993), 807–830

[Yo] L.S.Young, *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188

[ZL1] J.Ziv, A.Lempel, *A universal algorithm for sequential data compression*, IEEE Trans. Inform. Th. **23** (1977), 337–342

[ZL2] J.Ziv, A.Lempel, *Compression of individual sequences via variable rate coding*, IEEE Trans. Inform. Th. **24** (1978), 530–536

[Zw] R.Zweimüller, *Ergodic properties of infinite measure preserving interval maps with indifferent fixed points*, Ergod. Th. Dyn. Syst. **20** (2000), 1519–1549