Dynamical Systems and Stochastic Processes.

P. Collet
Centre de Physique Théorique
CNRS UMR 7644
Ecole Polytechnique
F-91128 Palaiseau Cedex (France)
e-mail: collet@cpht.polytechnique.fr

April 23, 2008

Abstract

This series of lectures is devoted to the study of the statistical properties of dynamical systems. When equipped with an invariant measure, a dynamical system can be viewed as a stochastic process. Many questions and results can be borrowed from probability theory and lead to many important results. On the other hand, the context of differential geometry often present in the phase space leads to a rich collection of important special questions and results.

1 Introduction.

A (classical) physical system is described by the set Ω of all its possible states, often called the phase space of the system. At a given time, all the properties of the system can be recovered from the knowledge of the instantaneous state $x \in \Omega$. The system is observed using the so called observables which are real valued functions on Ω. Most often the space of states Ω is a metric space (so we can speak of nearby states). In many physical situations there are even more structures on Ω: $\mathbf{R}^d$, Riemaniann manifolds, Banach or Hilbert spaces etc.

As a simple example one can consider a mechanical system with one degree of freedom. The state of the system at a given time is given by two real numbers: the position q and momentum p. The state space is therefore $\mathbf{R}^2$. A continuous material (solid, fluid etc) is characterized by the field of local velocities, pressure, density, temperature etc. In that case one often uses phase spaces which are Banach spaces.

As time goes on, the instantaneous state changes (unless the system is in a situation of equilibrium). The time evolution is a rule giving the change of the state with time. It comes in several flavors and descriptions summarized below.

i) Discrete time evolution. This is a map T from the state space Ω into itself producing the new state from the old one after one unit of time. If x_0 is the state of the system at time zero, the state at time one is $x_1 = T(x_0)$ and more generally the state at time n is given recursively by $x_n = T(x_{n-1})$. This is often written $x_n = T^n(x_0)$ with $T^n = T \circ T \circ \cdots \circ T$ (n-times). The sequence $\big(T^n(x_0)\big)$ is called the trajectory or the orbit of the initial condition x_0.

ii) Continuous time semi flow. This is a family $\big(\varphi_t\big)_{t\in\mathbf{R}^+}$ of maps of Ω satisfying

$$\varphi_0 = \mathrm{Id}\ , \qquad \varphi_s \circ \varphi_t = \varphi_{s+t}\ .$$

The set $\big(\varphi_t(x_0)\big)_{t\in\mathbf{R}^+}$ is called the trajectory (orbit) of the initial condition x_0. Note that if we fix a time step $\tau > 0$, and observe only the state at times $n\tau$ ($n \in \mathbf{N}$), we obtain a discrete time dynamical system given by the map $T = \varphi_\tau$.

iii) A differential equation on a manifold associated to a vector field $\vec{F}$

$$\frac{d\vec{x}}{dt} = \vec{F}(\vec{x})\ .$$

This is for example the case of a mechanical system in the Hamiltonian formalism. Under regularity conditions on $\vec{F}$, the integration of this equation leads to a semi-flow (and even a flow).

iv) There are other more complicated situations like non-autonomous systems (in particular stochastically forced systems), systems with memory, systems with delay, etc. but we will not consider them below.

A dynamical system is a set of states Ω equipped with a time evolution. If there is more structure on Ω, one can put more structure on the time evolution itself. For example in the case of discrete time, the map T may be measurable, continuous, differentiable etc.

Needless to say, dynamical systems abound in all domains of science. We already mentioned mechanical systems with a finite number of degrees of freedom and continuum mechanics, mechanical systems with many degrees of freedom are at the root of thermodynamics (and a lot of works have been devoted to the study of billiards which are related with simple models of molecules), one can also mention chemical reactions, biological and ecological systems, etc., even many random number generators turn out to be dynamical systems.

A useful general notion is that of equivalence (conjugation) of dynamical systems. Assume we have two dynamical systems (Ω_1, T_1) and (Ω_2, T_2). We say that they are conjugated if there is a bijection Φ from Ω_1 to Ω_2 such that

$$\Phi \circ T_1 = T_2 \circ \Phi \ .$$

It follows immediately that $\Phi \circ T_1^n = T_2^n \circ \Phi$, and a similar definition holds for (semi-)flows. Of course the map Φ may have supplementary properties: measurable, continuous etc.

Exercise 1.1. Show that the flows associated to two vector fields $\vec{X}$ and $\vec{Y}$ are conjugated by a diffeomorphism Φ if and only if for any point x

$$D\Phi_x \vec{X}(x) = \vec{Y}_{\Phi(x)} \ .$$

1.1 Examples.

We now give a list of examples to which we will refer many times later on. In each of them we first specify the phase space and then the time evolution. These examples appear often in the literature in equivalent (conjugated) forms.

Example 1.1. The phase space is the interval $\Omega = [0, 1]$, the time evolution is the map

$$f(x) = 2x \pmod 1 \ .$$

This is one of the simplest but, as we will see later on, it is one of the most chaotic dynamical system. Note that this map is not invertible.

Example 1.2. A generalisation of the above example is the class of piecewise expanding maps of the interval. The phase space is again the interval $\Omega = [0, 1]$. There is a finite sequence $a_0 < a_1 < \ldots < a_k = 1$ and for each interval $]a_j, a_{j+1}[$ $(0 \le j \le k-1)$ a monotone map f_j from $]a_j, a_{j+1}[$ to $[0, 1]$ which extends to a C^2 map on $[a_j, a_{j+1}]$. The time evolution is given by

$$f(x) = f_j(x) \quad \text{if} \quad x \in]a_j, a_{j+1}[\ .$$

One still has to decide the images of the points a_j, and we impose moreover the (uniform) expanding property:
There is an integer $m > 0$ and a constant $c > 1$ such that at any point where f^m is differentiable we have

$$\left| f^{m\prime} \right| \ge c \ .$$

The map $2x \pmod 1$ is of course an example of piecewise expanding maps of the interval. A piecewise expanding maps of the interval is said to have the Markov property if the closure of the image of any defining interval $]a_j, a_{j+1}[$ is a union of closures of such intervals. This is of course a topological notion but we will see later on that it has some connection with the notion of Markov chains in probability theory.

Exercise 1.2. Show that a finite iterate of a piecewise expanding map of the interval is also a piecewise expanding map of the interval.

Most random number generators turn out to be dynamical systems given by simple expressions. They generate a sequence of numbers which is meant to be a typical realisation of a sequence of independent drawings of a random variable. Since computers act on a finite set of rational numbers, the theory of random number generators is mostly devoted to the study of the arithmetic properties of the maps (for example large period, uniform distribution). Some popular examples are maps of the interval $\big[0, 2^{31} - 1\big]$ given by $T(x) = 16807x$ mod $\big(2^{31} - 1\big)$ or $T(x) = 48271x \mod \big(2^{31} - 1\big)$ which are among the best generators for 32 bits machines. We refer to [96], [105], and [114] for more on the subject and for quality tests.

Example 1.3. The previous dynamical systems have discontinuities, and many examples have been studied which are more regular, but often more difficult to analyse. One of the well known example is the one parameter family of quadratic maps. The phase space is the interval $\Omega = [-1, 1]$ and for a value of the parameter $\mu \in [0, 2]$, the time evolution is given by the map f_μ

$$f_\mu(x) = 1 - \mu x^2 \ .$$

There is a vast literature on these maps and on the more general case of unimodal maps.

Exercise 1.3. Prove that the invertible map $\Phi(x) = \sin(\pi x/2)$ on $[-1, 1]$ conjugates the map f_2 and the map $g(x) = 1 - 2|x|$.

Example 1.4. We now describe a more abstract example which is in some sense the general model for uniformly chaotic systems. Let $\mathcal{A}$ be a finite set (often called a finite alphabet). Let M be a matrix of zeros and ones of size $|\mathcal{A}| \times |\mathcal{A}|$ ($|\mathcal{A}|$ denotes the cardinality of $\mathcal{A}$). The matrix M is often called the incidence matrix. Recall that $\mathcal{A}^{\mathbf{Z}}$ is the set of all bi-infinite sequences $\underline{x} = \big(x_p\big)_{p \in \mathbf{Z}}$ of elements of $\mathcal{A}$. The phase space Ω is defined by

$$\Omega = \big\{ \underline{x} \in \mathcal{A}^{\mathbf{Z}} \,\big|\, M_{x_j, x_{j+1}} = 1 \; \forall \, j \in \mathbf{Z} \big\} \ .$$

The time evolution is the shift map $\mathcal{S}$ given by

$$\mathcal{S}\big(\underline{x}\big)_j = x_{j+1} \ .$$

Exercise 1.4. Show that $\mathcal{S}$ is an invertible map from Ω to itself (in particular describe the inverse map).

This dynamical system $\big(\Omega, \mathcal{S}\big)$ is called a sub-shift of finite type. When all the entries of the matrix M are equal to one, the dynamical system is called the full shift. If one uses $\mathbf{N}$ instead of $\mathbf{Z}$ in the definition of Ω, one speaks of a unilateral shift (which is not invertible).

Consider again the simple case of the map $2x$ (mod 1). A point of the interval $[0, 1]$ can be coded by its dyadic decomposition, namely

$$x = \sum_{j=1}^{\infty} \epsilon_j 2^{-j}$$

with $\epsilon_j = 0$ or 1. The decomposition is unique except for a countable set of points.

Exercise 1.5. Describe the countable set of points for which the decomposition is not unique.

It is easy to verify that the map $2x \pmod 1$ acts by shifting the dyadic sequence $\big(\epsilon_j\big)$. Therefore if we consider the alphabet with two symbols $\mathcal{A} = \{0, 1\}$, we have a conjugation between the dynamical system given by the map $2x \pmod 1$ on the unit interval and the unilateral full shift over two symbols except for a countable set.

Exercise 1.6. Explain why on most computers, for any initial condition in the interval $[0, 1]$ the orbit under the map $2x \pmod 1$ finishes (after 32 or 64 iterations) on the fixed point $x = 0$. Do the experiment.

This example can be easily generalised to the case of piecewise expanding Markov maps of the interval.

Exercise 1.7. Consider a piecewise expanding Markov map f of the interval with monotonicity and regularity intervals $]a_j, a_{j+1}[$, $0 \le j < k$ (see example 1.2). To each point x of the interval we associate the following code $\underline{\omega} \in \mathcal{A}^{\mathbf{N}}$on the alphabet of k symbols $\mathcal{A} = \{0, \ldots, k-1\}$ defined for $n = 0, 1, \ldots$ by

$$\omega_n = j \quad \text{if} \quad f^n(x) \in [a_j, a_{j+1}] \ .$$

Show that this gives a well defined and invertible coding except for a countable number of exceptional points (hint: use the expansivity property). Prove that this coding defines a conjugacy between the piecewise expanding Markov map f and a unilateral sub-shift of finite type (start by constructing the transition matrix M of zeros and ones).

Up to now, all the examples were one dimensional. We now give examples in two dimension.

Example 1.5. An historically important example in two dimension is the so called baker's map. The phase space is the square $[0, 1]^2$ and the map T is defined as follows

$$T(x, y) = \begin{cases} (2x, y/2) & \text{if } x \le 1/2 \\ (2x-1, (1+y)/2) & \text{if } x > 1/2 \end{cases} \tag{1}$$

Exercise 1.8. Show that except for a “small” set of exceptions the baker's map is invertible. To each point (x, y) of the phase space $[0, 1]^2$, we can associate a bi-infinite sequence $\underline{\omega} \in \{0, 1\}^{\mathbf{Z}}$ by

$$\omega_n = \begin{cases} 0 & \text{if } 0 \le T^n(x, y)_1 \le 1/2 \\ 1 & \text{otherwise.} \end{cases} \ .$$

Here $T^n(x, y)_1$ denotes the first coordinate of the point $T^n(x, y) \in [0, 1]^2$. Using this coding show that the baker's map is conjugated (except for a “small” set of points) to the full (bilateral) shift over two symbols.

There are of course many variants of this construction which is a particular case of a skew-product system. A dynamical system is called a skew-product if the phase space is a product space $\Omega = \Omega_1 \times \Omega_2$, and the time evolution T has the special form

$$T(x, y) = \big(T_1(x), T_2(x, y)\big) \ .$$

The dynamical system (Ω_1, T_1) is often called the base of the skew product, and the map T_2 the action in the fibers.

Another example of skew-product is the dissipative baker's map. The phase space is again the unit square $[0,1] \times [0,1]$. The map is for example given by

$$T(x,y) = \begin{cases} (2x, y/4) & \text{if } 0 \le x < 1/2 \\ (2(x-1/2), (2+y)/3) & \text{if } 1/2 \le x \le 1. \end{cases} \tag{2}$$

There are many variants of this example.

The above example is discontinuous, and we now give some differentiable examples.

Example 1.6. The phase space is the two torus $\Omega = \mathbf{T}^2$, and the map is given by

$$T(x,y) = \begin{pmatrix} 2x+y & \pmod{2\pi} \\ x+y & \pmod{2\pi} \end{pmatrix} .$$

This map is often called the cat map.

Exercise 1.9. Prove that this map is indeed a C^∞ map of the torus with a C^∞ inverse.

Example 1.7. A generalisation is given by the standard map of the torus

$$T(x,y) = \begin{pmatrix} 2x - y + a\sin(2\pi x) \pmod{1} \\ x \end{pmatrix} . \tag{3}$$

Exercise 1.10. Prove that the standard map of the torus is C^∞ with a C^∞ inverse.

Example 1.8. Another historically important example is the Hénon map. The phase space is $\Omega = \mathbf{R}^2$ and the two parameters (a and b) family of maps is given by

$$T(x,y) = \begin{pmatrix} 1 - ax^2 + y \\ bx \end{pmatrix} . \tag{4}$$

The "historical" values of the parameters are $a = 1.4$, $b = .3$ (see [74]).

Exercise 1.11. Prove that the Hénon map is invertible with a regular inverse if $b \neq 0$ (give the formula for the inverse).

Example 1.9. Another important example in three dimensions is given by the so called the solenoid. The phase space is $\Omega = \mathbf{R}^3$. In cylindrical coordinates (ρ, θ, z) the map is given by

$$T(\rho,\theta,z) = \begin{pmatrix} R + r\cos(\theta) + \lambda\rho \\ 2\theta \pmod{2\pi} \\ r\sin(\theta) + \lambda z \end{pmatrix}$$

where $0 < \lambda < 1/2$ and $R > r/\lambda > 0$.

Exercise 1.12. Show that for any $r' \in]r/(1-\lambda), r/\lambda[$, the solid torus $(\rho - R)^2 + z^2 < r'^2$ is map injectively into itself by T (see figure 1).

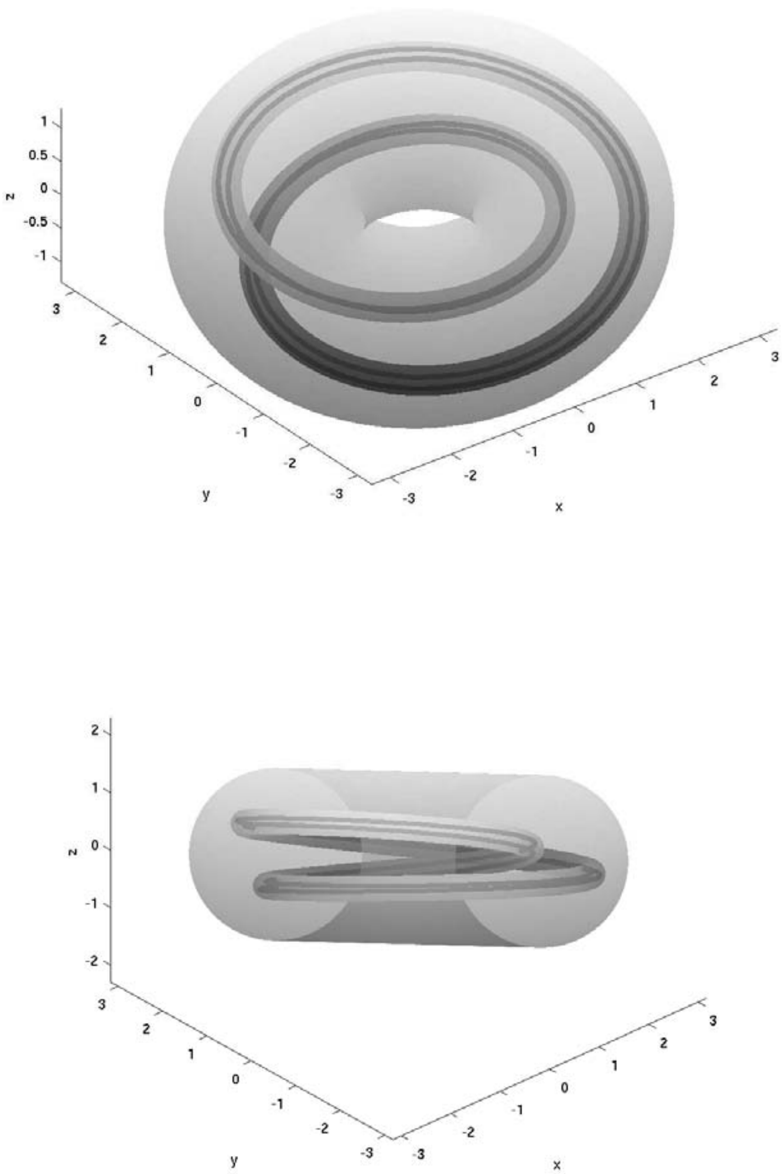

Figure 1: Three iterations of a torus under the solenoid map of example 1.9.

For the case of continuous time dynamical systems, given by vector fields, the most important example is the case of mechanical systems. We recall that the phase space is $\Omega = \mathbf{R}^{2d}$ (or some even dimensional manifold). The first d coordinates $q_1, \ldots, q_d$ are position coordinates, and the remaining d coordinates $p_1, \ldots, p_d$ are the conjugated momentum coordinates. A real valued function $H(\vec{q}, \vec{p})$ is given on the phase space and is called the Hamiltonian of the mechanical system. The time evolution is given by Hamilton's system of equations

$$\frac{d\,q_j}{dt} = \partial_{p_j} H \qquad \frac{d\,p_j}{dt} = -\partial_{q_j} H \qquad (j = 1 \ldots n)\ . \tag{5}$$

One can also add friction terms. Note that systems with time dependent forcing (non autonomous systems) do not fall in our present description, unless one can give a dynamical description of the forcing (this often leads to a skew product system).

One can also reduce a continuous time dynamical system to a discrete time map by the technique of Poincaré section (see figure 2 for the construction) or by looking at the time one map.

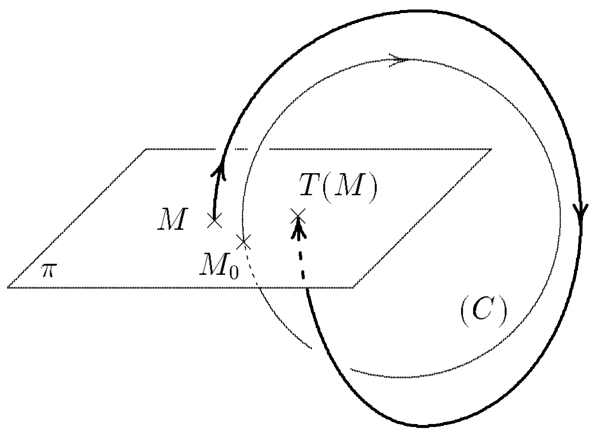

Figure 2: Construction of the Poincaré map around a cycle.

Similarly one can construct a continuous time system using the technique of suspension of a discrete time system. We refer the reader to the literature for more details on these constructions.

Example 1.10. An historically important example in dimension three is the Lorenz system. The phase space is $\Omega = \mathbf{R}^3$ and the vector field (depending on three parameters a, b and c) defining the time evolution is given by

$$\vec{X}(x, y, z) = \begin{pmatrix} \sigma(y - x) \\ -xy + rx - y \\ xy - bz \end{pmatrix}\ . \tag{6}$$

The historical values of the parameters are $\sigma = 10$, $r = 28$, $b = 8/3$, see [110]. There are many other continuous time dynamical systems from physical origin like for example the Navier Stokes equation. We will not attempt to list them here.

1.2 Invariant measures.

The basic goal of the study of dynamical systems is to understand the effect of time evolution on the state of the system, and primarily the long time behaviour of the state (although transient behaviours are sometimes very interesting and important in applications). There are basically two approaches to this problem. The first one can be called topological or geometric and studies objects like periodic orbits, attractors, bifurcations etc. The second approach can be called ergodic and studies the behavior of trajectories from a measure theoretic point of view. This is mostly interesting when there is a large variety of dynamical behaviours depending on the initial condition. One can for example discuss average quantities most often coming from the geometrical approach. In this course we will mostly deal with this second approach. Nevertheless, as we will often see below the interplay between the two approaches (geometric and ergodic) is particularly important and fruitful.

To introduce in more details the ergodic approach, let us consider the following question which is historically one of the first motivation of ergodic theory. Let A be subset of the phase space Ω (describing the states with a property of interest, for example that all the molecules in this room are in the left half). One would like to know for example in a long time interval $[0, N]$ (N large) how much time the system has spent in A, namely how often the state has the property described by A. Assume for simplicity we have a discrete time evolution. If χ_A denotes the characteristic function of the set A, the average time the system has spent in A over an interval $[0, N]$ starting in the initial state x_0 is given by

$$\mathcal{A}_N(x_0, A) = \frac{1}{N+1} \sum_{j=0}^{N} \chi_A\big(T^j(x_0)\big) \ . \tag{7}$$

It is natural to ask if this quantity has a limit when N tends to infinity. The answer may of course depend on A and x_0, but we can already make two important remarks. Assume the limit exists and denote it by $\mu_{x_0}(A)$.

First, it is easy to check that the limit also exists for $T(x_0)$ and also for any $y \in T^{-1}(x_0) = \big\{z \,|\, T(z) = x_0\big\}$. Moreover we have

$$\mu_{T(x_0)}(A) = \mu_y(A) = \mu_{x_0}(A) \ . \tag{8}$$

Second, the limit also exists if A is replaced by $T^{-1}(A)$ and has the same value, namely

$$\mu_{x_0}(T^{-1}(A)) = \mu_{x_0}(A) \ . \tag{9}$$

If one assumes that $\mu_{x_0}(A)$ does not depend on x_0 at least for Borel sets A (or some other sigma algebra but we will mostly consider the Borel sigma algebra below), one is immediately lead to the notion of invariant measure.

Definition 1.1. A measure μ on a sigma-algebra $\mathscr{B}$ is invariant by the measurable map T if for any measurable set A

$$\mu\big(T^{-1}(A)\big) = \mu(A) \ . \tag{10}$$

A similar definition holds for (semi-)flows.

Unless otherwise stated, when speaking below of an invariant measure we will assume it is a probability measure. We will denote by $(\Omega, T, \mathscr{B}, \mu)$ the dynamical system with state space Ω, discrete time evolution T, $\mathscr{B}$ is a sigma-algebra on Ω such that T is measurable with respect to $\mathscr{B}$ and μ is a measure on $\mathscr{B}$ invariant by T. As mentioned above $\mathscr{B}$ will most often be the Borel sigma-algebra and we will not mention it.

Exercise 1.13. Let (Ω, T) be a dynamical system and assume Ω is a metric space. Assume T is continuous, and let μ be a Borel measure. Show that μ is invariant if and only if

$$\int g \circ T \, d\mu = \int g \, d\mu \,,$$

for any continuous function g.

The goal of ergodic theory is to study systems $(\Omega, T, \mathscr{B}, \mu)$ and in particular their large time evolution. Assume now we have an observable g which we recall is a measurable function on the phase space Ω. We have seen above one such observable, namely the function χ_A which takes the value one if the state is in A (has the property described by A) and takes the value zero otherwise. $(\Omega, \mathscr{B}, \mu)$ is a probability space and therefore g is a random variable on this probability space. More generally $\big(g \circ T^n\big)$ is a discrete time, stationary, stochastic process. Therefore we can apply all the ideas and results of the theory of stochastic process to dynamical systems equipped with an invariant measure. As mentioned above and as we will see in more details below, this will be particularly interesting when done in conjunction with questions and concepts coming from the geometric approach.

Although any (stationary) stochastic process can be considered as a dynamical system (with phase space the set of all possible trajectories and transformation given by the time shift), there are however some important differences to mention. First of all, it is often the case that a dynamical system has many invariant measures. This raises the question of whether there is a more natural one. This is indeed the case from a physical point of view (the so called Physical measure when it exists, see below). However other invariant measures can be interesting from different point of views (measure of maximal entropy for example). In other words, in dynamical system theory sets of measure zero for one invariant measure may be important from some other point of view. There are other concepts like Hausdorff dimension and Hausdorff measure which can be interesting but may involve sets of measure zero. It is also important to mention that some sets of measure zero can indeed be "observed" (like for example in the multifractal formalism). It is worth mentioning that the choice of a particular invariant measure is related to the choice of an initial condition for a trajectory. We will come back to this point when discussing the ergodic theorem. The conclusion is that in dynamical systems, sets of measure zero should not be disregarded as systematically as in probability theory.

If μ_1 is an invariant measure for the dynamical system (Ω_1, T_1) and the map Φ is measurable, then the measure $\mu_2 = \mu_1 \circ \Phi^{-1}$ is an invariant measure for the dynamical system (Ω_2, T_2) conjugated by Φ.

We now discuss some simple examples of invariant measure. The simplest situation is when a system has a fixed point, namely a point in phase which does not move through time evolution. If we have a discrete time system given by a map T, this is a point ω of phase space such that $T(\omega) = \omega$. If we have a continuous time system given by a vector field $\vec{X}$, a fixed point (also called a stationary state) is a point ω of phase space such that $\vec{X}(\omega) = 0$. Such a point satisfies $\varphi_t(\omega) = \omega$ for any $t \in \mathbf{R}$ where φ_t is the flow integrating $\vec{X}$. It is easy to verify that if a system has a fixed point, the Dirac mass in this fixed point is an invariant measure, namely this measure satisfies equation (10). More generally, if we have a (finite) periodic orbit for a discrete time evolution, the average of the Dirac masses at the points of the orbit is an invariant measure.

Exercise 1.14. Find an infinite sequence of periodic orbits for the map of example 1.1.

The Lebesgue measure is also invariant by the map f of example 1.1. To prove this we compute $f^{-1}(A)$ for each measurable set $A \subset [0,1]$. It is easy to verify that (if $1 \notin A$)

$$f^{-1}(A) = \big(A/2\big) \cup \big(1/2 + A/2\big)$$

where

$$\big(A/2\big) = \big\{x \in [0,1/2] \,\big|\, 2x \in A\big\} \ ,$$

and

$$\big(1/2 + A/2\big) = \big\{x \in [1/2,1] \,\big|\, 2x - 1 \in A\big\} \ .$$

If λ denotes the Lebesgue measure, we have

$$\lambda\big(A/2\big) = \lambda\big(1/2 + A/2\big) = \lambda(A)/2$$

and since the intersection of the two sets $\big(A/2\big)$ and $\big(1/2 + A/2\big)$ is empty or reduced to the point $\{1/2\}$, equality (10) follows for $\mu = \lambda$ and any measurable set A. We will see below a generalisation of this idea.

Exercise 1.15. Prove that the Lebesgue measure on $[-1,1]$ is invariant for the map $g(x) = 1 - 2|x|$. Use exercise 1.3 to prove that the measure $dx/\sqrt{1-x^2}$ is invariant by the map $f_2(x) = 1 - 2x^2$ of example 1.3.

Exercise 1.16. Consider the middle third triadic Cantor set K. Recall that $K = \cap K_n$ where each K_n is a disjoint union of 2^n intervals, the intervals of K_{n+1} being obtained from those of K_n by dropping the (open) middle third subinterval. The construction is illustrated in figure 3.

Show that this is the set of points whose triadic representation does not contain any one. Define a measure ν by giving the weight 2^{-n} to any interval in K_n. Show that ν is invariant by the map $3x \pmod 1$.

Let p be a Markov transition matrix on a finite alphabet $\mathcal{A}$. Recall that this is a $\big|\mathcal{A}\big| \times \big|\mathcal{A}\big|$ matrix with non negative elements and such that for any $a \in \mathcal{A}$ we have

$$\sum_{b \in \mathcal{A}} p_{b,a} = 1 \ .$$

Such matrices are often denoted by $p(b|a)$. Let q be an eigenvector with non negative entries and eigenvalue one, namely for any $b \in \mathcal{A}$

$$q_b = \sum_{a \in \mathcal{A}} p_{b,a} q_a \ . \tag{11}$$

Let $\Omega = \mathcal{A}^{\mathbf{Z}}$ and consider the dynamical system on this phase space given by the shift. Given a finite sequence $x_r, \ldots, x_p$ ($r \le p \in \mathbf{Z}$) of elements of $\mathcal{A}$, often denoted by the short hand notation x_r^p, we denote by $C(x_r^p)$ the cylinder subset of Ω given by

$$C(x_r^p) = \left\{ \underline{y} \in \mathcal{A}^{\mathbf{Z}} \,\middle|\, y_j = x_j \ \ r \le j \le p \right\} \ .$$

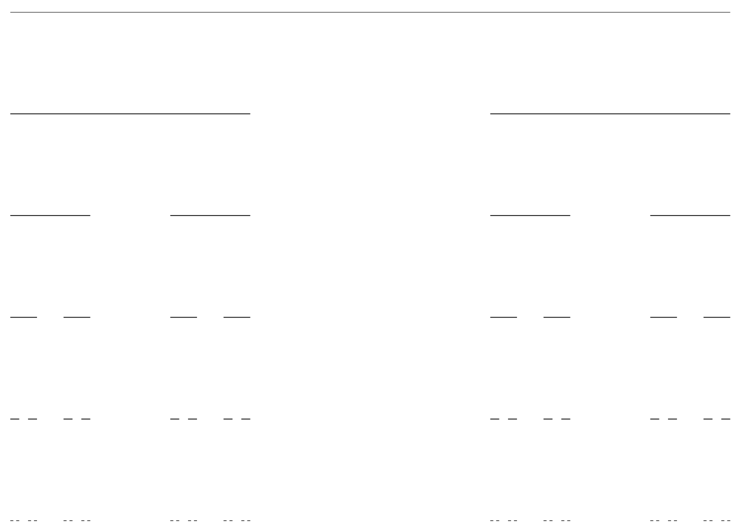

Figure 3: Recursive construction of the Cantor set.

We now define a measure μ on Ω by its value on any (finite) cylinder set:

$$\mu\big(C(x_r^p)\big) = q_{x_r} \prod_{j=r}^{p-1} p_{x_{j+1},x_j} \ .$$

It is easy to verify that this defines a measure on Ω, which is invariant by the shift.

Exercise 1.17. Prove this assertion.

This is of course nothing but a stationary Markov chain with transition probability p. In the particular case where for any $b \in \mathcal{A}$ the numbers $p_{b,a}$ do not depend on a we get a sequence of i.i.d. random variables on $\mathcal{A}$. This construction can be generalised in various ways and in particular to the construction of Gibbs measures on sub-shifts of finite type (see [131]).

Exercise 1.18. Let $p \in]0,1[$ and $q = 1-p$. Consider the phase space $\Omega = \{0,1\}^N$ equipped with the shift operator $\mathcal{S}$. Let μ_p be the infinite product measure with $\mu_p(0) = p$ (coin flipping). Using the conjugation of the shift and the map $f(x) = 2x \pmod 1$ (except for a countable set), show that one obtains an invariant measure (also denoted μ_p) for the map f. Show that the measure of any dyadic interval of length 2^{-n} is of the form $p^r q^{n-r}$ where $0 \le r \le n$ is given by the dyadic coding of the interval. Show that for $p = q = 1/2$ one gets the Lebesgue measure.

Exercise 1.19. Prove that the two dimensional Lebesgue measure is invariant for the baker's map (example 1.1), the cat map (example 1.6) and the standard map (example 1.7). Hint: compute the Jacobian.

1.3 The Perron-Frobenius operator.

It is natural at this point to ask how one finds the invariant measures of a dynamical system. Solving equation (10) is in general a non trivial task, in particular it is often the case that dynamical systems have an uncountable number of (inequivalent) invariant measures. One can then try to determine the invariant measures with some special properties. In this section we consider a map f of the unit interval $[0,1]$ and look for the invariant probability measures which are absolutely continuous with respect to the Lebesgue measure (a.c.i.p.m. for short). In other words, we are looking for measures $d\mu = hdx$, with h a non negative integrable function (of integral one) satisfying equation (10). This is equivalent to

$$\int_0^1 g(x)h(x)dx = \int_0^1 g(f(x))h(x)dx \ , \tag{12}$$

for any measurable (bounded) function g. Assume n f is piecewise monotone with finitely many pieces (see example 1.2), namely there is a sequence $a_0 = 0 < a_1 < \ldots < a_k = 1$ such that on each interval $[a_j, a_{j+1}]$ the map f is monotone and continuous. We have seen the case of piecewise expanding maps of the interval, but this is also the case for the quadratic family of example 1.3 and more generally for unimodal maps and for continuous maps with finitely many critical points. Let ψ_j denote the inverse of the map $f\big|_{[a_j,a_{j+1}]}$. This inverse exists since f is monotone on $[a_j, a_{j+1}]$, it is also a monotone map from $[f(a_j), f(a_{j+1})]$ to $[a_j, a_{j+1}]$. If the maps ψ_j are differentiable, we can perform a change of variables $y = f(x)$ in the right hand side of equation (12). More precisely

$$\begin{aligned}\int_0^1 g\big(f(x)\big)h(x)dx &= \sum_{j=0}^{k-1}\int_{a_j}^{a_{j+1}} g(f(x))h(x)dx \\ &= \sum_{j=0}^{k-1}\int_{f(a_j)}^{f(a_{j+1})} g(y)\big|\psi_j'(y)\big|h\big(\psi_j(y)\big)dy \ .\end{aligned} \tag{13}$$

Since this relation should hold for any bounded and measurable function g we conclude that h is the density of an a.c.i.p.m. for the map f if and only if $Ph = h$ where P is the so called Perron-Frobenius operator given by

$$Pg(x) = \sum_j \chi_{[f(a_j),f(a^{j+1})]}(x)g\big(\psi_j(x)\big)\big|\psi_j'(x)\big| \ . \tag{14}$$

In other words, h is the density of an a.c.i.p.m. for the map f if and only if it is an eigenvector of eigenvalue one for the operator P. Using the fact that the functions ψ_j are local inverses of f, the Perron-Frobenius operator is also given by

$$Pg(x) = \sum_{y,\, f(y)=x} \frac{g(y)}{\big|f'(y)\big|} \tag{15}$$

An immediate consequence of this formula is that for any measurable functions u and g,

$$P\big(u\,g\circ f\big)=gP\big(u\big)\ . \tag{16}$$

We mention without proof the following result for piecewise expanding maps of the interval. Before we state this result we recall that a function g on the interval is of bounded variation if there is a finite number $C>0$ such that for any strictly increasing sequence $0\le a_0<a_1<\ldots<a_n\le 1$ we have

$$\sum_{j=0}^{n-1}\big|g(a_j)-g(a_{j+1})\big|\le C\ .$$

The space of functions of bounded variations equipped with a norm given by the infimum of all these numbers C plus the sup norm is a Banach spec denoted below by **BV**.

Theorem 1.1. *Any piecewise expanding map of the interval has at least one absolutely continuous invariant probability measure with density in the space of functions of bounded variation.*

This Theorem due to Lasotta and Yorke is proved by investigating the spectral properties of the Perron-Frobenius operator. We refer to [80] for the details, references and consequences. We will also give some complements to this result in Theorem 3.3.

Equation (15) is very reminiscent of equation (11) for the invariant measures of a Markov chain. The next exercise develops this analogy.

Exercise 1.20. Consider a piecewise expanding Markov map of the interval with affine pieces. Show that there are a.c.i.p.m. with piecewise constant densities. Using the coding of exercise 1.7 show that this system is conjugated to a stationary Markov chain. Show that any stationary Markov chain with a finite number of states can be realised by a piecewise expanding Markov map of the interval with affine pieces.

Exercise 1.21. Let $p_1,\ p_2,\ldots$ be a sequence of positive numbers summing to one. Define an infinite sequence of intervals $I_1=[a_2,a_1],\ I_2=[a_3,a_2]\ \ldots,$ $I_j=[a_{j+1},a_j]\ \ldots$ where $a_1=1$ and

$$a_j=\sum_{l=j}^{\infty}p_l\ .$$

Define a map of the unit interval into itself by

$$f(x)=\begin{cases}\frac{x-a_1}{1-a_1} & \text{if } x\in I_1\ ,\\ a_j+(a_{j-1}-a_j)\frac{x-a_{j+1}}{a_j-a_{j+1}} & \text{if } x\in I_j \quad \text{for} \quad j>1\ .\end{cases}$$

Choose an initial condition at random with respect to the Lebesgue measure on the interval I_1 at generate the trajectory. Show that this dynamical system can be interpreted as a renewal process.

There is another version of relation (13) that will be useful later on. Let U be the Koopman operator defined on measurable functions by

$$Ug(x)=g(f(x))\ . \tag{17}$$

It is easy to verify that if μ is an invariant measure, U is an isometry of $L^2(d\mu)$. The equation (13) can now be written in the following form

$$\int_0^1 g_2 U\big(g_1\big) dx = \int_0^1 P\big(g_2\big) g_1 dx \tag{18}$$

for any pair g_1, g_2 of square integrable functions.

Although we have worked here explicitly with the Lebesgue measure, we mention that similar relations can be obtained for other reference measures. An important case is the case of Gibbs states on sub-shift of finite type (example 1.4). Recall that the phase space Ω is a shift invariant subspace of $\Theta = \mathcal{A}^{\mathbf{Z}}$ where $\mathcal{A}$ is a finite alphabet. For two elements $\underline{x}$ and $\underline{y}$ of Θ, denote by $\delta\big(\underline{x}, \underline{y}\big)$ the nearest position to the origin where these two sequences differ, namely

$$\delta\big(\underline{x}, \underline{y}\big) = \min\left\{|q| \,\middle|\, x_q \neq y_q\right\} .$$

For a given number $0 < \zeta < 1$ we define a distance d_ζ (denoted simply by d when there is no ambiguity in the choice of ζ) by

$$d_\zeta = \zeta^{\delta\left(\underline{x}, \underline{y}\right)} . \tag{19}$$

Exercise 1.22. Prove that d_ζ is a distance and that Θ is compact in this topology. Show that the phase space Ω (see example 1.4) of any sub-shift of finite type with alphabet $\mathcal{A}$ is closed in Θ. Prove that the shift map $\mathcal{S}$ on Ω is continuous (and even Hölder), with a continuous inverse.

Theorem 1.2. *Let ϕ be a real valued Hölder continuous function on Ω the phase space of sub-shift of finite type. Assume the incidence matrix M of the sub-shift is irreducible and aperiodic (in other words, there is an integer r such that all the entries of the matrix M^r are non zero). Then there is a unique probability measure μ invariant by the shift and a positive constant $\Gamma > 1$ such that for any cylinder set $C\big(x_q^p\big)$ $(q \leq p)$ and for any $\underline{y} \in C\big(x_q^p\big)$*

$$\Gamma^{-1} \leq \frac{\mu\big(C\big(x_q^p\big)\big)}{e^{-(p-q+1)P_\phi} e^{\sum_{j=q}^{p} \phi\left(\mathcal{S}^j(\underline{y})\right)}} \leq \Gamma , \tag{20}$$

where P_ϕ is the pressure defined by

$$P_\phi = \lim_{n\to\infty} \frac{1}{2n+1} \log\left(\sum_{x_{-n}^n,\, M_{x_j, x_{j+1}}=1} e^{\sum_{j=-n}^{n} \phi\left(\mathcal{S}^j(\underline{x})\right)}\right) . \tag{21}$$

In this last formula, $\underline{x}$ denote any point of Ω belonging to the cylinder set $C\big(x_q^p\big)$.

Exercise 1.23. Prove that the pressure does not depend on the choice of the point $\underline{x}$ in the cylinder set $C\big(x_q^p\big)$ (use the Hölder continuity of ϕ).

We refer to [20] or [131] for a proof of this result using a Perron-Frobenius operator and the relation with Gibbs states in statistical mechanics.

Exercise 1.24. Show that when $\phi\big(\underline{x}\big)$ depends only on x_0 and x_1 one gets all the Markov chains with finite states (start by constructing the incidence matrix).

1.4 attractors.

We mentioned above the interplay in the study of dynamical systems between the geometric and ergodic approach. One of the first example comes from the notion of attractor in dissipative systems. There are several possible definitions for an attractor and we will adopt the following one formulated for discrete time evolution. There is an analogous notion for continuous time evolution. We say that a subset A of the phase space is invariant by the map T if $T(A) \subset A$.

Definition 1.2. A (compact) invariant subset $\mathscr{A}$ of the phase space is an attracting set if there is a neighborhood V of $\mathscr{A}$ such that for any neighborhood U of $\mathscr{A}$, there is an integer n_U such that for any $n > n_U$, $T^n(V) \subset U$.

In particular, all orbits with initial condition in V accumulate on $\mathscr{A}$.

Definition 1.3. We will say that a (compact) invariant subset $\mathscr{A}$ of the phase space is an attractor if it is an attracting set containing a dense orbit.

We refer to [129] and [73] for more on these notions. The basin of attraction of an attracting set $\mathscr{A}$ is the set of initial conditions whose orbit accumulate on $\mathscr{A}$. It follows immediately from this definition that if we have an attracting set $\mathscr{A}$, any invariant measure with support contained in the basin of this attracting set should have support in $\mathscr{A}$. Attractors may have complicated geometry in which case they are called strange attractors. We now give some pictures of attractors.

Figure 4 is a plot of the “attractor” for the Hénon map of example 1.8. This picture was obtained by starting with an initial condition at the origin $(0, 0)$, iterating a large number of times to be as near as possible to the attractor, and then plotting a certain number of the following iterations supposed to visit densely the attractor.

A subset V for checking that we have an attracting set and its first iterate are shown in picture 5 again for the Hénon map.

The attractor of the Lorenz system (6) is drawn in figure 6.

Remark 1.1.

1. The simplest attractors are the stable fixed points (stationary solutions) and the stable periodic orbits (stable invariant cycles).

2. If the dynamical system depends on a parameter, the attractor will also in general change with the parameter, not only quantitatively but also qualitatively as occurs for example in bifurcations. In figure 7 is drawn the attractor of the quadratic family of example 1.3 as a function of the parameter. One sees in particular the famous sequence of period doubling bifurcations.

3. A dynamical system may have several attractors in the same phase space.

4. Each attractor has its own basin: the set of initial conditions attracted by this attractor.

5. The boundaries of the basins are in general complicated invariant sets, repelling transversally (toward the attractors). A well known example is provided by the Newton method applied to the equation $z^3 = 1$. $z_{n+1} =$

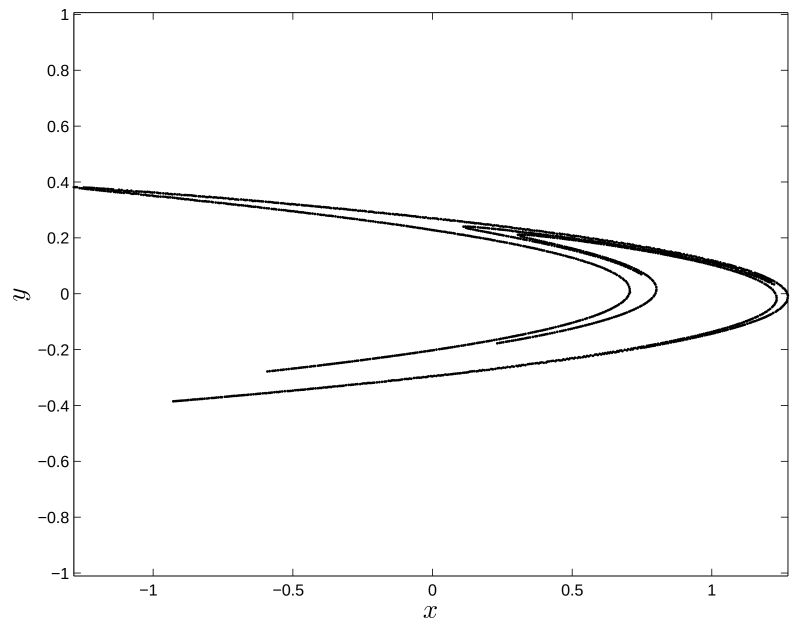

Figure 4: Attractor of the Hénon map, $a = 1.4$, $b = .3$.

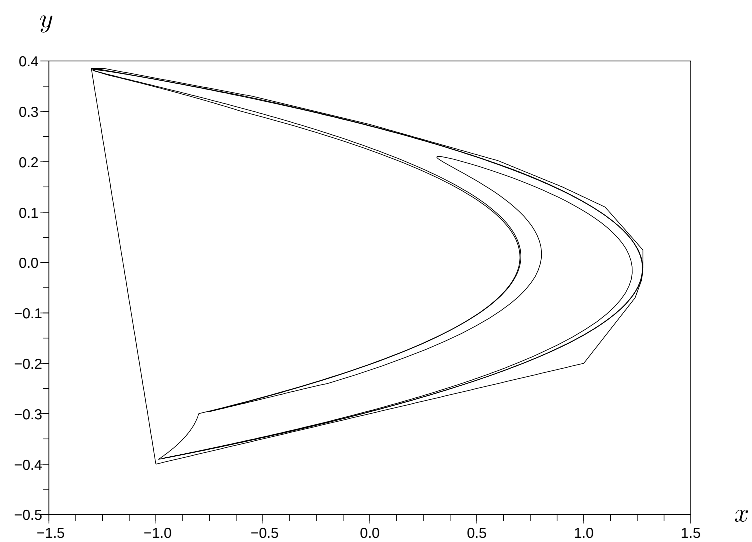

Figure 5: The neighborhood V and two of its iterates for the Hénon attractor, $a = 1.4$, $b = .3$.

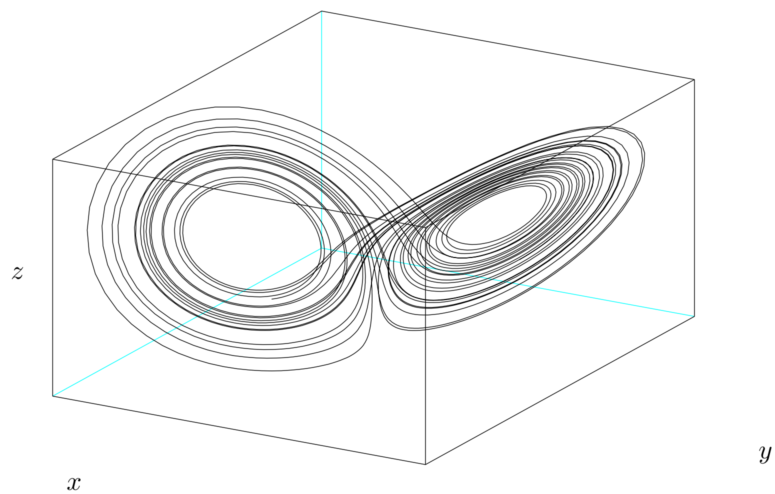

Figure 6: Attractor of the Lorenz system, $\sigma = 10$, $r = 28$, $b = 8/3$.

$f(z_n) = (z_n + 2/z_n^2)/3$. There are three stable fixed points (attractors): $1, j, \overline{j}$. Figure 8 is a drawing of the common boundary of the basins of attraction of these three fixed points.

6. The concept of attractor is suited for the study of dissipative systems. Volume preserving systems do not have attractors, like the mechanical systems without friction.

Exercise 1.25. Show that the attractor of the dissipative baker's map (2) is the product of a Cantor set by a segment (see figure 9). Show that there is an invariant measure which can be identified as the product of the Lebesgue measure (in the horizontal direction) with the Bernoulli (product) measure of parameters $(1/2, 1/2)$ (in the vertical direction).

We now recall the definition of an axiom A attractor (see [129]).

Definition 1.4. An attractor Λ for a diffeomorphims T on a phase space Ω is called an axiom A attractor if the tangent space $E(x)$ at any point $x \in \Lambda$ has a stable-unstable splitting

$$E(x) = E^u(x) \oplus E^s(x)$$

where $E^u(x)$ and $E^s(x)$ have the following property.

i) The fields of subspaces E^u and E^s are invariant, namely for any $x \in \Lambda$

$$DT_x E^s(x) = E^s(T(x)) \quad \text{and} \quad DT_x^{-1} E^u(x) = E^u(T^{-1}(x)) \; .$$

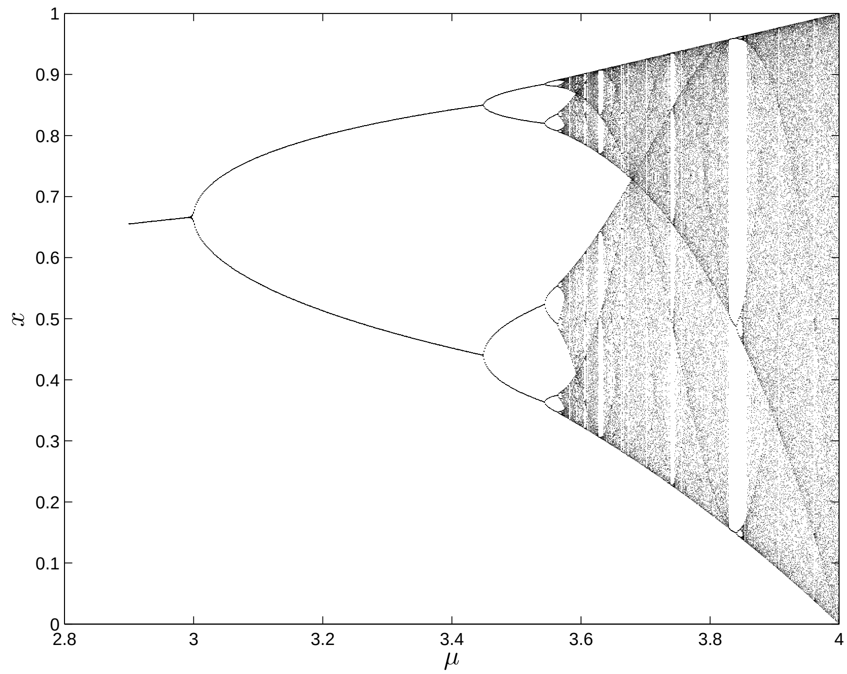

Figure 7: Attractor of the quadratic family as a function of the parameter.

ii) The splitting is uniformly hyperbolic. Namely, there are two positive constants C and $\rho < 1$ such that for any integer n and any $x \in \Lambda$ we have

$$\left\| DT_x^n \middle| E^s(x) \right\| \le C\rho^n \quad \text{and} \quad \left\| DT_x^{-n} \middle| E^u(x) \right\| \le C\rho^n \ .$$

We refer to [3], [4], [56], [73], [121] and [129] for general presentations, results and references.

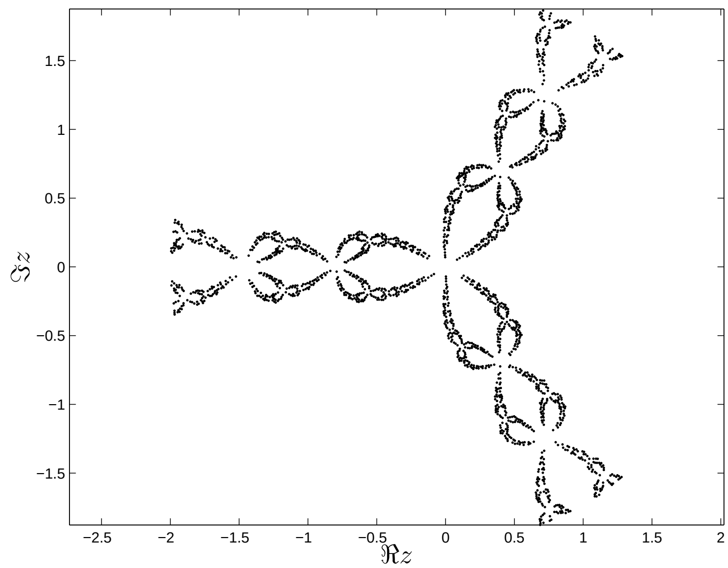

Figure 8: Intersection with the ball of radius 2 of the common boundary of the basins of attraction of the three fixed points of the Newton method applied to $z^3 = 1$.

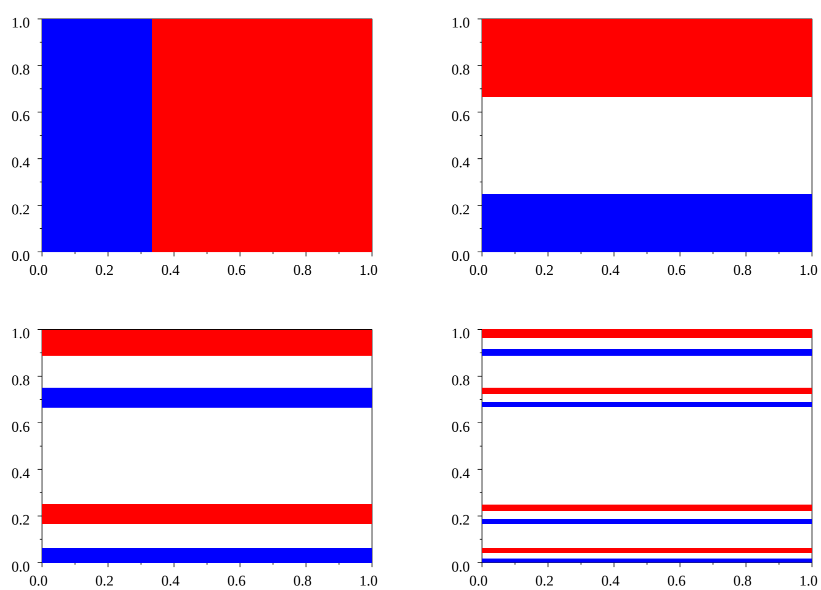

Figure 9: Several iterations of the square under the dissipative baker's map (2).

2 Ergodic Theory and Dynamical System Quantities.

2.1 The ergodic theorem.

We can now come back to the question of the asymptotic limit of the ergodic average (7). This is settled by the ergodic theorems. The ergodic theorem of Von-Neumann [151] applies in an L^2 context, while the Birkhoff ergodic theorem [14] applies almost surely (we refer to [99], [84] [92] and [123] for proofs and extensions and applications). We now state the Birkhoff ergodic theorem which generalises the law of large numbers in the non i.i.d. case.

Theorem 2.1. *Let $(\Omega, T, \mathscr{B}, \mu)$ be a dynamical system (recall that T is measurable for the sigma algebra $\mathcal{B}$ and the measure μ on $\mathcal{B}$ is T invariant). Then for any $f \in L^1(\Omega, \mathcal{B}, d\mu)$*

$$\mathcal{A}_N(x, f) = \frac{1}{N+1}\sum_{j=0}^{N} f\big(T^j(x)\big) \ .$$

converges when N tends to infinity for μ almost every x.

We now make several remarks about this fundamental result.

i) By (8), the set of points where the limit exists is invariant (and of full measure by Birkhoff's Theorem). Moreover, if we denote by $g(x)$ the limiting function (which exists for μ almost every x), it is invariant, namely $g(T(x)) = g(x)$.

ii) The set of μ measure zero where nothing is claimed depends on f and μ. One can often use a set independent of f (for example if $L^1(d\mu)$ is separable). We will comment below on the dependence on μ.

iii) The theorem is often remembered as saying that the time average is equal to the space average. This has to be taken with a grain of salt. As we will see below changing the measure may change drastically the exceptional set of measure zero and this can lead to completely different results for the limit.

iv) The set of initial conditions where the limit does not exist, although small from the point of view of the measure μ may be big from other points of views (see [12]).

The most interesting case is of course when the limit in the ergodic theorem is independent of the initial condition (except for a set of μ measure zero). This leads to the definition of ergodicity.

Definition 2.1. A measure μ invariant for a dynamical system $(\Omega, \mathcal{A}, T)$ is ergodic if any invariant function (*i.e.* any measurable function f such that $f \circ T = f$, μ almost surely) is μ almost surely constant.

There are two useful equivalent conditions. The first one is in terms of invariant sets. An invariant (probability) measure is ergodic if and only if

$$\mu\big(A \Delta T^{-1}(A)\big) = 0 \iff \mu(A) = 0 \quad \text{or} \quad \mu(A) = 1 \ . \tag{22}$$

Exercise 2.1. Prove this assertion

The second equivalent condition is in terms of the Birkhoff average. An invariant (probability) measure is ergodic if and only if for any $f \in L^1(d\mu)$

$$\lim_{N\to\infty} \frac{1}{N+1} \sum_{j=0}^{N} f\big(T^j(x)\big) = \int f \, d\mu \tag{23}$$

μ almost surely.

Remark 2.1.

i) In the condition (23), the limit exists μ almost surely by the Birkhoff ergodic Theorem (2.1). It is enough to require that the limit is μ almost surely constant since the constant has to be equal to the integral.

ii) In formula (23), the state x does not appear on the right hand side, but it is hidden in the fact that the formula is only true outside a set of measure zero.

iii) It often happens that a dynamical system $(\Omega, \mathcal{A}, T)$ has several ergodic invariant (probability) measures. Let μ and ν be two different ones. It is easy to verify that they are disjoint, namely one can find a set of measure one which is of measure zero for the other and vice versa. This explains why the ergodic theorem applies to both measures leading in general to different time averages.

iv) For non ergodic measures, one can use an ergodic decomposition (disintegration). We refer to [99] for more information. However in concrete cases this may lead to rather complicated sets.

v) In probability theory, the ergodic theorem is usually called the law of large numbers for stationary sequences.

vi) Birkhoff's ergodic theorem holds for semi flows (continuous time average). It also holds of course for the map obtained by sampling the semi flow uniformly in time. However non uniform sampling may spoil the result (see [133] and references therein).

vii) Simple cases of non ergodicity come from Hamiltonian systems with the invariant Liouville measure. First of all since the energy is conserved (the Hamiltonian is a non trivial invariant function), the system is not ergodic. One has to restrict the dynamics to each energy surface. More generally if there are other independent constants of the motion one should restrict oneself to lower dimensional manifolds. For completely integrable systems, one is reduced to a constant flow on a torus which is ergodic if the frequencies are incommensurable. It is also known that generic Hamiltonian systems are neither integrable nor ergodic (see [113]).

viii) Proving ergodicity of a measure is sometimes a very hard problem. Note that it is enough to prove (23) for a dense set of functions, for example the continuous functions.

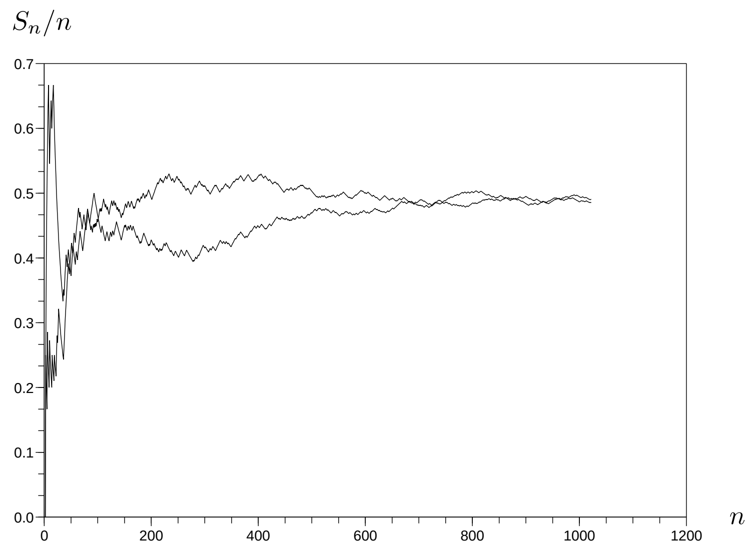

Figure 10: Evolution of S_n/n for two random initial conditions.

Some of the above remarks are illustrated in figures 10, 11 and 12. The dynamical system is the map $3x \pmod 1$ of the unit interval and the function f is the characteristic function of the interval $[1/2, 1]$. The figure shows S_n/n as a function of n for several initial conditions. The Lebesgue measure is invariant and ergodic and we expect that if we choose an initial condition uniformly at random on $[0, 1]$, then S_n/n should converge to $1/2$ (the average of f with respect to the Lebesgue measure). This is what we see in figure 10 where two different initial conditions were chosen at random according to the Lebesgue measure. Note that the convergence to the limit $1/2$ is not very fast. We will come back to this question later on. Figure 11 shows S_n/n as a function of n for the initial condition $x_0 = 7/80$. This point is periodic of period four, and its orbit has three points in the interval $[1/2, 1]$ (exercise: verify this statement). The invariant measure (the sum of the Dirac measures on the four points of the orbit) is ergodic and therefore, S_n/n converges to $3/4$ (the average of f with respect to the discrete invariant measure supported by the orbit).

Figure 11 was drawn using an atypical initial condition for which S_n/n will oscillate forever between two extreme values, although the oscillations will take longer and longer time.

Let $p \in]0, 1[$ with $p \neq 1/2$. Consider the probability measure μ_p on $\Omega = \{0, 1\}^{\mathbf{N}}$ of exercise 1.18. This is equivalent to flipping a coin with probability p to get head and probability $q = 1 - p$ to get tail. It is easy to prove that the measure μ_p is invariant and ergodic for the shift $\mathcal{S}$. In this case, ergodicity follows from the law of large numbers and the ergodic theorem. We will see later on some other methods of proving ergodicity using mixing.

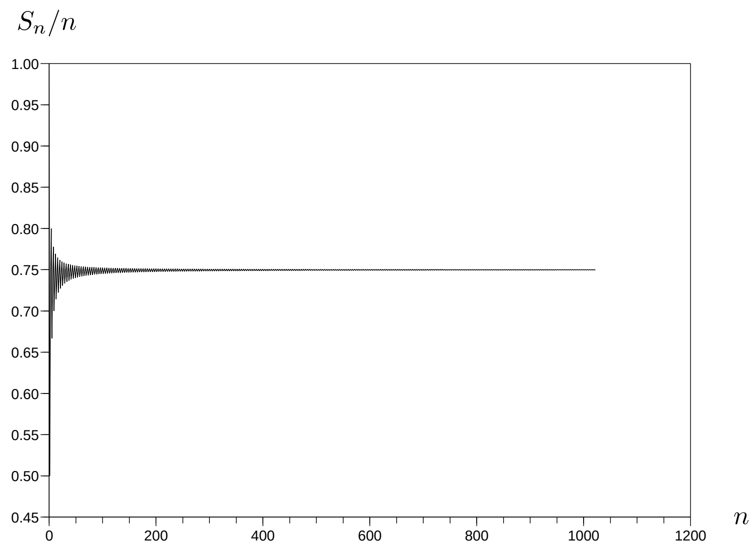

Figure 11: Evolution of S_n/n for a periodic initial condition.

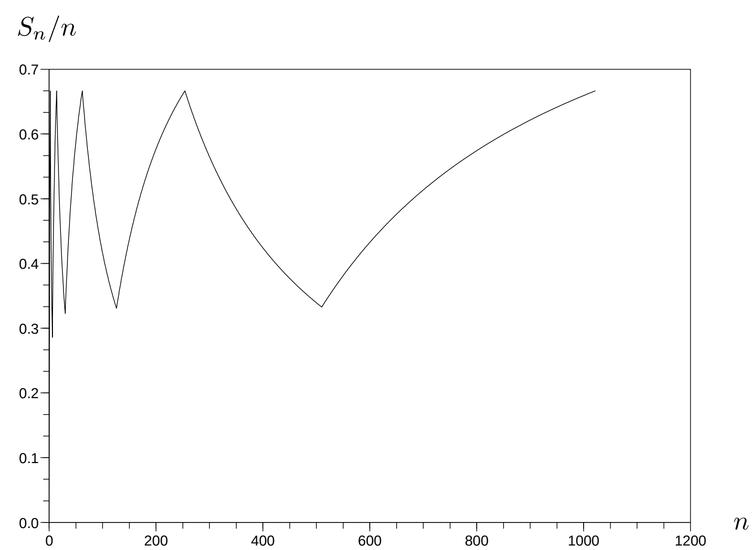

Figure 12: Evolution of S_n/n for an atypical initial condition.

An extension of the ergodic theorem is the so called sub-additive ergodic theorem which has many useful applications.

Theorem 2.2. *Let $(\Omega, \mathscr{B}, T)$ a measurable dynamical system and μ an ergodic invariant measure. Let (f_n) be a sequence of integrable functions satisfying for each $x \in \Omega$ and for any pair of integers m and n the inequality*

$$f_{n+m}(x) \le f_n(x) + f_m\big(T^n(x)\big) \ , \tag{24}$$

and such that

$$\liminf_{n\to\infty} \frac{1}{n} \int f_n d\mu > -\infty \ .$$

Then μ almost surely

$$\lim_{n\to\infty} \frac{f_n}{n} = \lim_{n\to\infty} \frac{1}{n} \int f_n d\mu \ .$$

We refer to [99] for a proof.

2.2 Physical measure.

As we have seen already several times, it is often the case that a dynamical system has different ergodic invariant measures. It is therefore natural to ask if there is one of these measures which is in some sense is more important than the others. From a physical point of view that we will not discuss here, the Lebesgue measure of the phase space is singled out. However if the dynamical system has a rather small attractor, for example a set of Lebesgue measure zero as often occurs in dissipative systems, this does not look very helpful. In such cases there is no invariant measure absolutely continuous with respect to the Lebesgue measure. The study of axiom A systems by Sinai, Ruelle and Bowen led however to a generalisation of the notion of ergodicity which makes use of the Lebesgue measure on the whole phase space.

Definition 2.2. Consider a dynamical system given by a map T on a phase space Ω which is a compact Riemanian manifold. An ergodic invariant measure μ for a map T is called a Physical measure if there is a subset U of Ω of positive Lebesgue measure such that for any continuous function g on Ω and any initial point $x \in U$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} g\big(T^j(x)\big) = \int g \, d\mu \ .$$

We emphasize again that the measure μ may be supported by an attractor of zero Lebesgue measure. For axiom A systems it is known that such measures always exit. They have some more properties and are particular cases of SRB measures (see below for the definition).

The idea behind the definition is that if the dynamical system has an attractor, this set should be transversally attracting. An orbit will therefore rapidly approach the attractor, and a continuous function will take almost the same value for a point on the attractor and a point nearby. One would then be able to apply the ergodic theorem for the ergodic measure μ (supported by the attractor) provided exceptional sets of μ measure zero would in some sense be projections of sets of Lebesgue measure zero in the phase space. This requires an additional technical property (the so-called absolute continuity of the projection along the stable manifolds).

We have already seen examples of Physical measures. The simplest case is of course when the system has an invariant measure absolutely continuous with respect to the Lebesgue measure. For the quadratic family of example (1.3) it is known that there is a unique Physical measure for all the parameters and a subset of positive measure of the parameter space (the interval $[0, 2]$) for which such a measure is absolutely continuous. We refer to [5] for recent works and references. Similarly, for the Hénon map of example (1.8) it has been shown that for a set of positive Lebesgue measure of parameter values, Physical measures exist (see [10], [11]). We refer to [63] for a general discussion.

Exercise 2.2. Show that the invariant measure of the dissipative baker's map constructed in exercise 1.25 is a Physical measure.

We mention that there are dynamical systems which do not possess Physical measures. We refer to [154] for more information.

2.3 Lyapunov exponents.

An important application of the sub-additive ergodic theorem is the proof of the Oseledec Theorem about Lyapunov exponents. We will see here the interplay between ergodic theory and differentiable structure on the phase space and the time evolution. We first explain the ideas in an informal way.

Complex dynamical behaviours are often called chaotic, but one would like to have a precise definition of this notion. It is often observed in these systems that the trajectories depend in a very sensitive way on the initial conditions. If one starts with two very nearby generic initial conditions, it is often observed in such systems that as time evolves, the distance between the orbits grow exponentially fast. On the contrary, the reader can compute what happens if one considers a simple harmonic oscillator which is an integrable (non chaotic) system. Namely for two nearby initial conditions with the same energy, the distance does not grow at all. The phenomenon of rapid growth of little errors in the initial condition is called sensitive dependence on initial conditions. This phenomenon is illustrated in figure 13 for the Lorenz model. On sees a plot of the x component as a function of time for two 'very' nearby initial conditions. After having staid together for a while, they suddenly separate macroscopically.

Since it is so important and so common in chaotic systems, one would like to have a quantitative measurement of the sensitive dependence on initial conditions.

Assume that the time evolution map T is differentiable (and the phase space Ω is a differentiable manifold, we will assume $\Omega = \mathbf{R}^d$ for simplicity). If we start with two nearby initial conditions x and $y = x + \vec{h}$ with $\vec{h}$ small, we would like to estimate the size of

$$T^n\big(x + \vec{h}\big) - T^n\big(x\big) \ ,$$

namely how the initially small error $\vec{h}$ evolves in time. As long as the error has not grown too much, we can expect to have a reasonable answer from the first

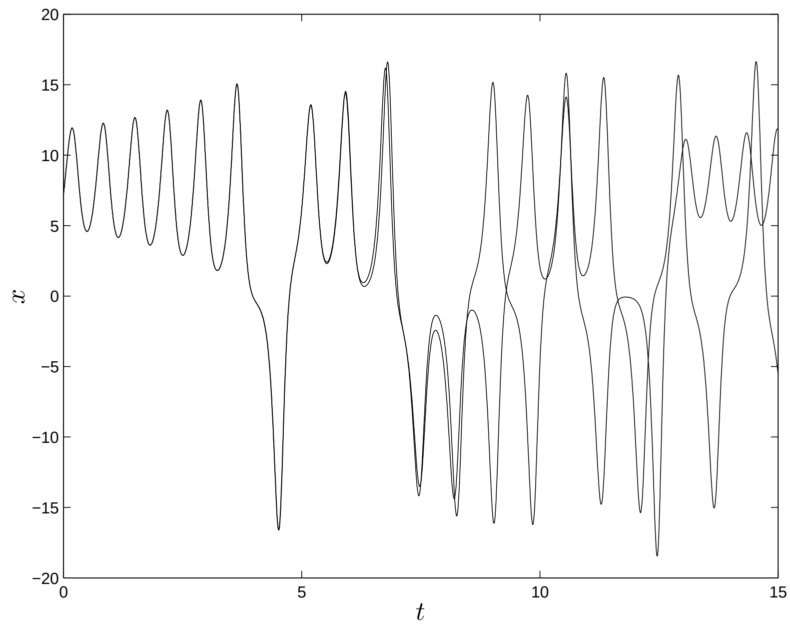

Figure 13: Sensitive dependence to the initial condition in the Lorenz model.

order Taylor formula, namely

$$T^n\big(x+\vec{h}\big) - T^n\big(x\big) = DT^n_x\vec{h} + \mathscr{O}(\vec{h}^2)\ ,$$

Where DT^n_x denotes the differential of the map T^n at the point x. We recall that from the chain rule we have

$$DT^n_x = DT_{T^{n-1}(x)}\cdots DT_x$$

where in general the matrices in this product do not commute. The fact that we have a product of n terms suggests that we look (test) for an exponential growth, namely that we take the logarithm, divide by n and take the limit $n \to \infty$. Since the matrices do not commute, the formulation of the Oseledec theorem for the general case will be slightly more sophisticated. We will start with some warming up examples.

First (easy) case: dimension one, the phase space is an interval ([-1,1]), the dynamics is given by a regular map f. One gets by Taylor's formula

$$f^n(x+h) = f^n(x) + f^{n\prime}(x)h + \mathcal{O}(h^2)\ .$$

By the chain rule ($f^n = f \circ f \circ \cdots \circ f$, n times), we have

$$f^{n\prime}(x) = \prod_{j=0}^{n-1} f'\big(f^j(x)\big)\ .$$

As explained above, this naturally suggests an exponential growth in n, and to look for the exponential growth rate per step (unit of time), namely

$$\frac{1}{n}\log\big|f^{n\prime}(x)\big| = \frac{1}{n}\sum_{j=0}^{n-1}\log\big|f'\big(f^j(x)\big)\big|\ .$$

We see appearing a temporal average and we can apply Birkhoff's ergodic Theorem 2.1. Let μ be an ergodic invariant measure such that the function $\log\left|f'\right|$ has an integrable modulus. Then except on a set of μ measure zero, we have convergence of the temporal average, and moreover

$$\lim_{n\to\infty}\frac{1}{n}\log\left|f^{n\prime}(x)\right| = \int \log\left|f'(\cdot)\right| d\mu \ .$$

This number is called the Lyapunov exponent of the measure μ for the transformation f. Here one should stress again the importance of the initial condition. There are many initial conditions for which the limit does not exist. For many other initial conditions, the limit exists, but take another value. For example the initial conditions typical of a different ergodic measure.

Exercise 2.3. On the phase space $\Omega = [0, 1]$, consider the map

$$T(x) = \begin{cases} 3x & \text{if } 0 \le x < 1/3 \\ 3(1-x)/2 & \text{if } 1/3 \le x \le 1. \end{cases} \tag{25}$$

whose graph is shown in figure 14.

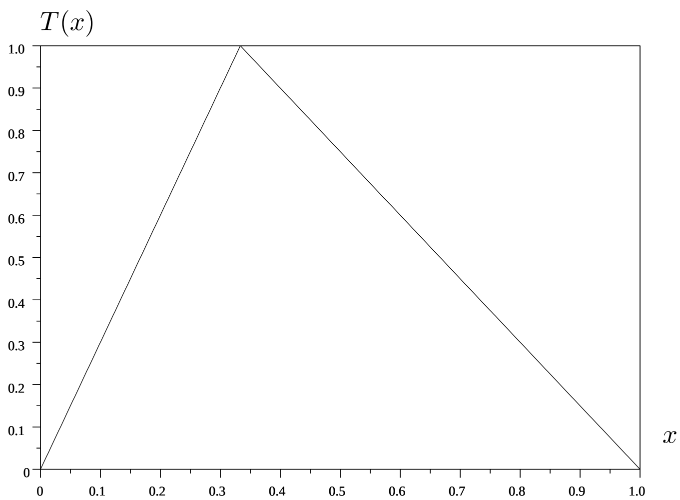

Figure 14: Graph of the map T given by 25.

Show that the Lebesgue measure is T invariant and ergodic, and that its Lyapunov exponent is equal to $\frac{1}{3}\log 3 + \frac{2}{3}\log(3/2)$. Show that the transformation T has a fixed point $x = 3/5$, that the Dirac measure in this point is also invariant and ergodic, and its Lyapunov exponent is equal to $\log(3/2)$. Show that for any ergodic invariant measure the Lyapunov exponent belongs to the interval $[\log(3/2), \log 3]$. Show that there is an uncountable set of invariant ergodic measures, all with different Lyapunov exponents (taking all the values between $\log 3/2$ and $\log 3$). Hint: use a conjugacy (coding) to the unilateral shift on two symbols, and then consider the product measures.

We now consider the next level of difficulty: dimension of the phase space larger than one. As a simple example we will use the dissipative baker's map T given in (2), whose attractor has been studied in exercise 1.25.

The differential of the dissipative baker's map is given by

$$DT_{(x,y)} = \begin{cases} \left(\begin{smallmatrix} 3 & 0 \\ 0 & 1/4 \end{smallmatrix}\right) & \text{if } 0 \le x < 1/3 \\ \left(\begin{smallmatrix} 3/2 & 0 \\ 0 & -1/3 \end{smallmatrix}\right) & \text{if } 1/3 \le x \le 1. \end{cases}$$

To estimate $DT^n_{(x,y)}$ we now have to perform a product of matrices. If we start from the initial point (x,y), with an initial error $\vec{h}$, we have after n iteration steps (using the chain rule) an error given (up to order $\mathscr{O}(\vec{h}^2)$) by

$$DT^n_{(x,y)}\vec{h} = DT_{T^{n-1}(x,y)}DT_{T^{n-2}(x,y)}\cdots DT_{(x,y)}\vec{h} \ .$$

In general matrices do not commute (hence one should be careful with the order in the product). However here they do commute (and they are even diagonal). Therefore, to obtain the product matrix, it is enough to take the product of the diagonal elements:

$$DT_{T^{n-1}(x,y)}DT_{T^{n-2}(x,y)}\cdots DT_{(x,y)} = \begin{pmatrix} \prod_{j=0}^{n-1} u\big(T^j(x,y)\big) & 0 \\ 0 & \prod_{j=0}^{n-1} v\big(T^j(x,y)\big) \end{pmatrix}$$

where

$$u(x,y) = 3\chi_{[0,1/3]}(x) + (3/2)\chi_{[1/3,1]}(x)$$

and

$$v(x,y) = (1/4)\chi_{[0,1/3]}(x) - (1/3)\chi_{[1/3,1]}(x) \ .$$

We can now take the log of the absolute value of each diagonal entry of the product and apply Birkhoff's ergodic theorem as in the one dimensional case. Since the functions u and v do not depend on y, the integral with respect to the Physical measure reduces to the integration with respect to the one dimensional Lebesgue measure. Therefore, for Lebesgue almost any (x,y) (two dimensional Lebesgue measure, recall the definition of Physical measure)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \log\big|u\big(T^j(x,y)\big| = (1/3)\log 3 + (2/3)\log(3/2) = \log\big((27/4)^{1/3}\big) \ .$$

Similarly

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \log\big|v\big(T^j(x,y)\big| = -(1/3)\log 4 - (2/3)\log(3) = \log\big((36)^{-1/3}\big) \ .$$

We conclude that

$$DT_{T^{n-1}(x,y)}DT_{T^{n-2}(x,y)}\cdots DT_{(x,y)} \approx \begin{pmatrix} (27/4)^{n/3} & 0 \\ 0 & (36)^{-n/3} \end{pmatrix} \ ,$$

and we now have to interpret this result.

If we take a vector $\vec{h} = \big(h_1, h_2\big)$, such that $h_1 \neq 0$, we see that the first component of $DT^n_{(x,y)}\vec{h}$ grows exponentially fast with growth rate $\log\big((27/4)^{1/3}\big)$, and this component dominates the other one which decreases exponentially fast at the rate $\log\left(36^{-1/3}\right)$ (< 0). In other words, almost any error grows exponentially fast with rate $\log\left((27/4)^{1/3}\right)$, this is the maximal Lyapunov exponent. But there is another Lyapunov exponent $\log\left(36^{-1/3}\right)$ (< 0) corresponding to special initial errors satisfying $h_1 = 0$. These errors do not grow but decay exponentially fast. This is similar to the diagonalisation of matrices but the interpretation is slightly more involved. We have to distinguish two subspaces. The first one is the entire space $E_0 = \mathbb{R}^2$. The second one is the subspace of codimension one $E_1 = \left\{(0, h_2)\right\} \subset E_0$. If $\vec{h} \in E_0 \backslash E_1 \backslash \{\vec{0}\}$, the initial error grows exponentially fast with rate $\log\left((27/4)^{1/3}\right)$. If $\vec{h} \in E_1$ the initial error decreases exponentially fast with rate $\log\left(36^{-1/3}\right)$.

We now come to the study of Lyapunov exponents in the general case of dynamical systems. This general case combines the two preceding ideas (product of matrices and ergodic theorem) together with a new fact: the subspaces E_0, E_1, etc. depend on the initial condition, they vary from point to point. Moreover the matrices appearing in the product do not commute (be careful with the order).

For the moment let us first consider the behaviour of the norm and define

$$f_n(x) = \log\left\|DT_x^n\right\| .$$

It is easy to verify that this sequence of functions satisfies the sub-additive inequality (24).

Exercise 2.4. Prove it.

Assume now that we have an ergodic invariant measure μ for T, then we can apply Theorem 2.2 (provided we can check the second assumption) and conclude that

$$\frac{1}{n}\log\left\|DT_x^n\right\|$$

converges μ almost surely. If this quantity is positive, we know that some initially small errors are amplified exponentially fast. If this quantity is negative or zero, we know that initial errors cannot grow very rapidly.

More generally, if we have a fixed vector $\vec{h}$, recall that

$$\left\|DT_x^n\vec{h}\right\|^2 = \left\langle \vec{h} \,\middle|\, \left(DT_x^n\right)^t DT_x^n \vec{h}\right\rangle$$

where $\langle \quad | \quad \rangle$ denotes the scalar product in $\mathbf{R}^d$. This suggests to study the asymptotic exponential growth of the matrix $\left(DT_x^n\right)^t DT_x^n$. The answer is provided by the following theorem due initially to Oseledec.

Theorem 2.3. *Let μ be an ergodic invariant measure for a diffeomorphism T of a compact manifold Ω. Then for μ almost every initial condition x, the sequence of symmetric non-negative matrices*

$$\left(\left(DT_x^n\right)^t DT_x^n\right)^{1/2n}$$

converges to a symmetric non-negative matrix Λ (independent of x). Denote by $\lambda_0 > \lambda_1 > \ldots > \lambda_k$ the strictly decreasing sequence of the logarithms of the eigenvalues of the matrix Λ (some of them may have non trivial multiplicity).

These numbers are called the Lyapunov exponents of the map T for the ergodic invariant measure μ. Except for a subset of the phase space of μ measure zero, for every point x there is a decreasing sequence of subspaces

$$\Omega = E_0(x) \supsetneq E_1(x) \supsetneq \cdots \supsetneq E_k(x) \supsetneq E_{k+1}(x) = \{\vec{0}\}$$

satisfying (μ almost surely) the following properties

$$DT_x E_j(x) = E_j\big(T(x)\big)$$

and for any $\vec{h} \in E_j(x) \backslash E_{j+1}(x)$ for some $0 \le j \le k$,

$$\lim_{n\to\infty} \frac{1}{n} \log \big\| DT_x^n \vec{h} \big\| = \lambda_j \ .$$

Remark 2.2.

1. Note that the theorem says that

$$\big\| DT_x^n \vec{h} \big\| \sim e^{n\lambda_j} \ .$$

   There may be large (sub-exponential) prefactors which can depend on $\vec{h}$ and x.

2. Positive Lyapunov exponents are obviously responsible for the Sensitive Dependence on Initial Conditions. Their corresponding “eigen” directions are tangent to the attractor.

3. Transversally to the attractor one gets contracting directions, namely negative Lyapunov exponents.

Exercise 2.5. Prove that if we conjugate the map T to a map T' by a diffeomorphism, and consider an image measure, then we get the same Lyapunov exponents.

If the map depends on a parameter, the Lyapunov exponents depend also in general on the parameter (one has also to decide on an ergodic invariant measure for each parameter). This is illustrated in figure 15 where the Lyapunov exponent of the quadratic family of example 1.3 is drawn as a function of the parameter for the Physical measure.

We refer to [8] and [9] for proofs and references.

2.4 Entropies

Often one does not have access to the points of phase space but only to some fuzzy approximation. For example if one uses a real apparatus which always has a finite precision. There are several ways to formalize this idea.

i) The phase space Ω is a metric space with metric d. For a given precision $\epsilon > 0$, two points at distance less than ϵ are not distinguishable.

ii) One gives a (measurable) partition $\mathcal{P}$ of the phase space, $\mathcal{P} = \big\{A_1, \cdots, A_k\big\}$ (k finite or not) , $A_j \cap A_l = \emptyset$ for $j \neq l$ and

$$\Omega = \bigcup_{j=1}^{k} A_j \ .$$

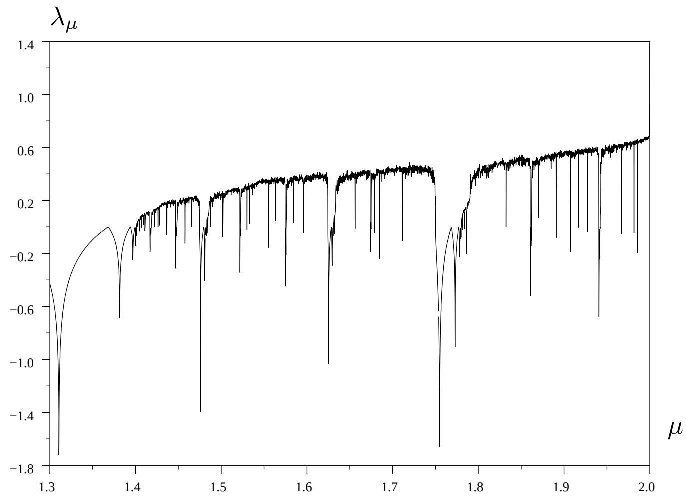

Figure 15: Lyapunov exponent as a function of the parameter for the quadratic family of example (1.3).

> Two points in the same atom of the partition $\mathcal{P}$ are considered indistinguishable. If there is a given measure μ on the phase space it is often useful to use partitions modulo sets of μ measure zero.

The notion of partition leads naturally to a coding of the dynamical system. This is a map Φ from Ω to $\{1, \cdots, k\}^{\mathbf{N}}$ given by

$$\Phi_n(x) = l \qquad \text{if} \qquad T^n(x) \in A_l \ .$$

If the map is invertible, one can also use a bilateral coding. If $\mathcal{S}$ denotes the shift on sequences, it is easy to verify that $\Phi \circ T = \mathcal{S} \circ \Phi$. In general $\Phi(\Omega)$ is a complicated subset of $\{1, \cdots, k\}^{\mathbf{N}}$, *i.e.* it is difficult to say which codes are admissible. There are however some examples of very nice codings like for Axiom A attractors (see [20] [92] and [131]).

Let $\mathcal{P}$ and $\mathcal{P}'$ be two partitions, the partition $\mathcal{P} \vee \mathcal{P}'$ is defined by

$$\mathcal{P} \vee \mathcal{P}' = \{A \cap B \ , \ A \in \mathcal{P} \ , \ B \in \mathcal{P}'\} \ .$$

If $\mathcal{P}$ is a partition,

$$T^{-1}\mathcal{P} = \{T^{-1}(A)\}$$

is also a partition. Recall that (even in the non invertible case) $T^{-1}(A) = \{x \ , \ T(x) \in A\}$.

A partition $\mathcal{P}$ is said to be generating if (modulo sets of measure zero)

$$\bigvee_{n=0}^{\infty} T^{-n}\mathcal{P} = \epsilon$$

with ϵ the partition into points. In this case the coding is injective (modulo sets of measure zero).

Exercise 2.6. For the map of the interval $3x \pmod 1$, show that

$$\mathcal{P} = \big\{[0,1/3],[1/3,1]\big\}$$

is a generating partition (modulo sets of measure zero).

We now come to the definition of entropies. There are two main entropies, the topological entropy and the so called Kolmogorov-Sinai entropy. Both measure how many different orbits one can detect through a fuzzy observation.

The topological entropy is defined independently of a measure. We will only consider here the case of a metric phase space. The topological entropy counts all the orbits modulo fuzziness. We say that two orbits of initial condition x and y respectively are ϵ (the precision) different before time n (with respect to the metric d) if

$$\sup_{0\le k\le n} d\big(T^k(x),T^k(y)\big) > \epsilon\ .$$

Intuitively the apparatus with precision ϵ has detected their difference. Let $N_n(\epsilon)$ be the maximum number of pairwise ϵ different orbits up to time n. In other words this is the maximum number of pairwise different films the dynamics can generate up to time n if two images differing at most by ϵ are considered identical.

The topological entropy is defined by

$$h_{\mathrm{top}} = \lim_{\epsilon\searrow 0}\limsup_{n\to\infty}\frac{1}{n}\log N_n(\epsilon)\ .$$

We show in figure 16 the topological entropy as a function of the parameter for the one parameter family f_μ of quadratic maps of example 1.3.

A positive topological entropy characterizes a large number of qualitatively different trajectories. It is an indication of chaos. A transverse homoclinic crossing already implies a positive topological entropy. But this chaos may occur in a very small, irrelevant, part of the phase space. For example one can have a stable periodic orbit whose basin is almost all the phase space and a fractal repeller of small dimension, zero volume, supporting all the positive topological entropy. This is what happens for example for (unimodal) maps of the interval with a stable periodic orbit of period 3, and has led to the famous statement that period three implies chaos (see [37]). As we have already mentioned, in general one does not observe all the trajectories but only the typical trajectories with respect to a measure (a Physical measure for example).

When an ergodic invariant measure μ is considered, the disadvantage of the topological entropy is that it measures the total number of (distinguishable) trajectories, including trajectories which have an anomalously small probability to be chosen by μ. It even often happens that these trajectories are much more numerous than the ones favored by μ. The Kolmogorov-Sinai entropy is then more adequate.

If $\mathcal{P}$ is a (measurable) partition, its entropy $H_\mu(\mathcal{P})$ with respect to the measure μ is defined by

$$H_\mu(\mathcal{P}) = -\sum_{A\in\mathcal{P}}\mu(A)\log\mu(A)\ .$$

This is Shannon's formula, and in communication theory one often uses the logarithm base 2 (see [137]).

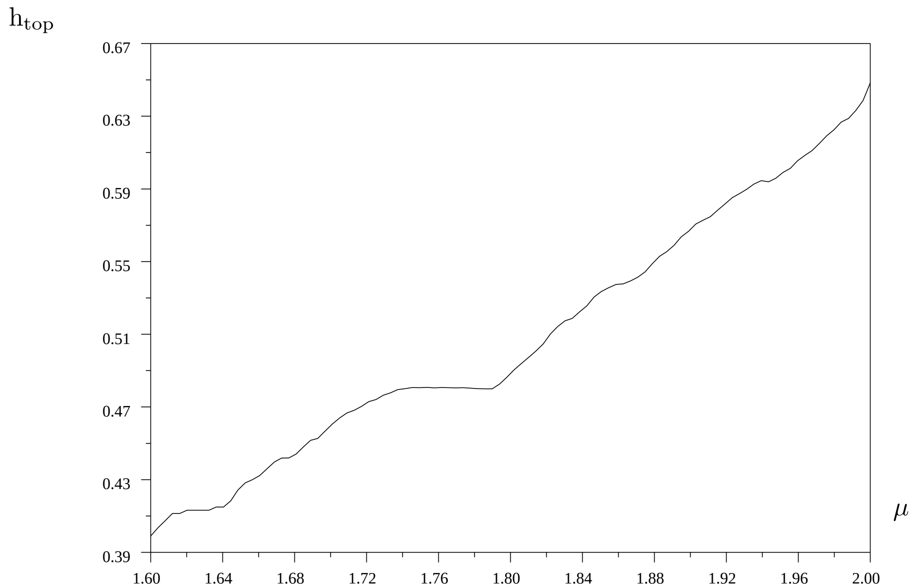

Figure 16: Topological entropy as a function of the parameter for quadratic family of example 1.3.

The entropy of the dynamical system with respect to the partition $\mathcal{P}$ and the (invariant, ergodic, probability) measure μ is defined by

$$H_\mu(T, \mathcal{P}) = \lim_{n \to \infty} \frac{1}{n} H_\mu(\vee_{j=0}^{n-1} T^{-j}\mathcal{P}) \ .$$

One can prove that the limit always exists. Finally the Kolmogorov-Sinai entropy is defined by

$$h_\mu(T) = \sup_{\mathcal{P}\ \text{finite or countable}} H_\mu(T, \mathcal{P}) \ .$$

If $\mathcal{P}$ is generating, and $h_\mu(T) < \infty$, then

$$h_\mu(T) = H_\mu(T, \mathcal{P}) \ .$$

One has also the general inequality

$$h_\mu(T) \le h_{\text{top}}(T) \ , \tag{26}$$

and moreover

$$h_{\text{top}}(T) = \sup_{\mu,\ T\ \text{ergodic}} h_\mu(T) \ .$$

However the maximum may not be reached. We refer to [138] or [123] for proofs and more results on the Kolmogorov Sinai entropy.

An important application of the entropy is Ornstein's isomorphism theorem for Bernoulli shifts, we refer to [118] for more information.

Another important application of the entropy is the Shannon-Mc Millan-Breiman theorem which in a sense counts the number of typical orbits for the measure μ.

Theorem 2.4. *Let $\mathcal{P}$ be a finite generating partition. For any $\epsilon > 0$ there is an integer $N(\epsilon)$ such that for any $n > N(\epsilon)$ one can separate the atoms of the partition $\mathcal{P}_n = \vee_{j=0}^{n-1} T^{-j}\mathcal{P}$ into two disjoint subsets*

$$\mathcal{P}_n = \mathcal{B}_n \cup \mathcal{G}_n$$

such that

$$\mu\left(\bigcup_{A\in\mathcal{B}_n} A\right) \le \epsilon$$

and for any $A \in \mathcal{G}_n$

$$e^{-n(h_\mu(T)+\epsilon)} \le \mu(A) \le e^{-n(h_\mu(T)-\epsilon)}$$

In other words, the atoms in $\mathcal{B}_n$ are in some sense the "bad" atoms (for the measure μ), but their total mass is small. On the other hand, the good atoms (those in $\mathcal{G}_n$) have almost the same measure, and of course their union gives almost the total weight. An immediate consequence is that

$$\left|\mathcal{G}_n\right| \asymp e^{n\,h_\mu(T)} ,$$

where $\left|\mathcal{G}_n\right|$ denotes the cardinality of the set $\mathcal{G}_n$. This is similar to the well known formula of Boltzmann in statistical mechanics relating the entropy to the logarithm of the (relevant) volume in phase space. There is also an obvious connection with the equivalence of ensembles. We refer to [99] and [103] for more information.

As mentioned before, it often happens that $\left|\mathcal{G}_n\right| \ll \left|\mathcal{B}_n\right|$. A simple example is given by the Bernoulli shift on two symbols. Let $p \in]0,1[$ with $p \neq 1/2$. Consider the probability measure μ_p on $\Omega = \{0,1\}^{\mathbf{N}}$ of exercise 1.18. This is equivalent to flipping a coin with probability p to get head and probability $q - 1 - p$ to get tail. It is easy to prove that the measure μ is invariant and ergodic for the shift $\mathcal{S}$. Recall that we have defined a distance d on Ω (and in fact a one parameter family of distances) in (19). It is easy to prove that $h_{\text{top}}(\mathcal{S}) = \log 2$.

Exercise 2.7. Prove this assertion.

The partition

$$\mathcal{P} = \Big\{\{x_0 = 0\,\}, \{x_0 = 1\,\}\Big\}$$

is generating. For the entropy one has $h_\mu(\mathcal{S}) = -p\log p - q\log q$. However for n large we get (since $p \neq 1/2$)

$$\left|\mathcal{G}_n\right| \asymp e^{n\,h_\mu(\mathcal{S})} \ll 2^n \asymp \left|\mathcal{B}_n\right| .$$

One can work this classical example in more details, and ask for a more precise asymptotic of the set of typical sequences among the sequences of zeros (tail) and ones (head) of length n. To count them we will classify them according the number m of ones (head). If m is given, there are $\binom{n}{m}$ sequences of length n with m ones and $n-m$ zeros. Of course

$$2^n = \sum_{m=0}^{n} \binom{n}{m}$$

but also

$$1 = \sum_{m=0}^{n} \binom{n}{m} p^m q^{n-m} \ .$$

Let us look more carefully at this well known identity. If a sum of $(n+1)$ positive terms is equal to 1, there are two extreme cases: all the terms are equal (to $1/(n+1)$), or one is equal to one and the others are zero. The present case is more like this second situation. Indeed, by Stirling's formula

$$\log\left(\binom{n}{m} p^m q^{n-m}\right) \approx$$

$$n\log n - m\log m - (n-m)\log(n-m) + m\log p + (n-m)\log q$$

and this quantity is maximal for $m = \overline{m} = p\,n$. It decreases very fast when m deviates from $\overline{m}$. For example, for $\left|m-\overline{m}\right| > 2.6\sqrt{pqn}$ we have

$$\sum_{|m-\overline{m}|>2.6\sqrt{pqn}} \binom{n}{m} p^m q^{n-m} \le .01$$

Hence there is an extremely small probability to observe a sequence with an m differing form $\overline{m}$ by more than $2.6\sqrt{pqn}$. When $m = \overline{m}$, $\log\left(p^m q^{n-m}\right)$ is equal to e^{-nh} with

$$h = -p\log p - q\log q \ ,$$

the entropy per flip. For m near $\overline{m}$

$$\binom{n}{m} p^m q^{n-m} \approx \frac{1}{2\pi\sqrt{pqn}} e^{-(m-\overline{m})^2/(2pqn)} \ .$$

Note that $p = 1/2$ gives the maximal entropy which is equal to $\log 2$. In this case all the sequences of outcome have the same probability. This measure is called the measure of maximal entropy (its entropy is equal to the topological entropy and this is the maximum as follows from inequality (26). All the trajectories are equally typical. This is why we assumed before $p \neq 1/2$.

Let us summarize (for $p \neq 1/2$). Out of the 2^n possible outcomes, the observed ones belong with overwhelming probability to a much smaller subset ($m \simeq \overline{m}$). In this subset, all sequences have about the same probability $p^{\overline{m}} q^{n-\overline{m}} = e^{-nh}$. Since these are the only sequences which weight in probability, their number is $N_p(n) \approx e^{nh}$ (the total probability is one). For the Kolmogorov-Sinai entropy of the independent coin flipping we have

$$\lim_{n\to\infty} \frac{1}{n} \log N_p(n) = h = -p\log p - q\log q \ . \tag{27}$$

and

$$N_p(n) \approx e^{nh} \ll 2^n$$

for $p \neq 1/2$. Said differently, one can separate the possible outcomes of length n in two categories: the good ones and the bad ones. This discrimination depends heavily on the measure. All bad outcomes together form a set of small measure (not only each one separately). The good outcomes have all about the same probability e^{-nh} and together account for almost all the weight (one). Their number is about e^{nh}. This is similar to the equivalence of ensembles in statistical mechanics (see [103]).

Another way to formulate the Shannon-McMillan-Breiman theorem is to look at the measure of cylinder sets. For a point $x \in \Omega$, let $C_n(x)$ be the atom of $\vee_{j=0}^{n-1} T^{-j}\mathcal{P}$ which contains x. In other words, $C_n(x)$ is the set of $y \in \Omega$ such that for $0 \leq j \leq n-1$, $T^j(x)$ and $T^j(y)$ belong to the same atom of $\mathcal{P}$ (the trajectories are indistinguishable up to time $n-1$ from the fuzzy observation defined by $\mathcal{P}$, they have the same code). Then for μ almost every x we have

$$h_\mu(T) = -\lim_{n\to\infty} \frac{1}{n} \log \mu\big(C_n(x)\big)\ .$$

A similar result holds using a metric instead of a partition. It is due to Brin and Katok, and uses the so called Bowen balls defined for $x \in \Omega$, the transformation T, $\delta > 0$ and an integer n by

$$B(x, T, \delta, n) = \left\{ y\ ,\ d\big(T^j(x), T^j(y)\big) < \delta \text{ for } j = 0, \cdots, n-1 \right\}\ .$$

These are again the initial conditions leading to trajectories indistinguishable (at precision δ) from that of x up to time $n-1$.

Theorem 2.5. *If μ is T ergodic, we have for μ almost any initial condition*

$$h_\mu(T) = \lim_{\delta\searrow 0} \liminf_{n\to\infty} -\frac{1}{n} \log \mu\big(B(x, T, \delta, n)\big)\ .$$

We refer to [26] for proofs and related results.

Another way to get the entropy was obtained by Katok [90] in the case of a phase space Ω with a metric d and a continuous transformation T. The formula is analogous to the formula for the topological entropy except that it refers only to typical orbits for a measure. Let A be a set, we denote by $N_n(A, \epsilon)$ the maximal number of ϵ different orbits before time n with initial condition in A. Recall that two orbits of initial condition x and y respectively are ϵ different before time n if

$$\sup_{0\leq k\leq n} d\big(T^k(x), T^k(y)\big) > \epsilon\ .$$

Note that the orbits are not required to stay in A but only to start in A. Let μ be an ergodic invariant probability measure for T. We define for $\delta \in]0, 1[$ the sequence of numbers $\tilde{N}_n(\mu, \delta, \epsilon)$ by

$$\tilde{N}_n(\mu, \delta, \epsilon) = \inf_{A,\ \mu(A)\geq 1-\delta} N_n(A, \epsilon)\ .$$

Theorem 2.6. *For any $\delta > 0$ we have*

$$h\big(\mu\big) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{\log \tilde{N}_n(\mu, \delta, \epsilon)}{n} = \lim_{\epsilon\to 0} \liminf_{n\to\infty} \frac{\log \tilde{N}_n(\mu, \delta, \epsilon)}{n}$$

We refer to [90] for the proof. We refer to [92], [86], [98], and references therein for more results.

A way to measure the entropy in coded systems was discovered by Ornstein and Weiss using return times to the first cylinder. The result was motivated by the investigation of the asymptotic optimality of Ziv's compression algorithms which has very popular implementations (gzip, zip etc.)

Let q be a finite integer, and assume the phase space of a dynamical system is a shift invariant subset of $\{1, \cdots, q\}^{\mathrm{N}}$.

As before, we denote the shift by $\mathcal{S}$. Let μ be an ergodic invariant measure. Let n be an integer and for $x \in \Omega$, define $R_n(x)$ as the smallest integer such that the first n symbols of x and $\mathcal{S}^{R_n(x)}(x)$ are identical.

Theorem 2.7. *For μ almost every x we have*

$$\lim_{n\to\infty} \frac{1}{n} \log R_n(x) = h_\mu(\mathcal{S}) \ .$$

We refer to [119] for the proof.

A metric version was recently obtained by Donarowicz and Weiss, using again Bowen balls, in the context of dynamical systems with a metric phase space. Define for a number $\delta > 0$, an integer n, and a point $x \in \Omega$ the integer

$$R(x, T, \delta, n) = \inf \left\{ l > 0 \ , \ T^l(x) \in B(x, T, \delta, n) \right\} .$$

In other words

$$R(x, T, \delta, n) = \inf \left\{ l > 0 \ , \ d\big(T^{l+j}(x), T^j(x)\big) < \delta \text{ for } j = 0, \cdots, n-1 \right\} .$$

Theorem 2.8. *For μ almost every x, we have*

$$\lim_{\delta \searrow 0} \limsup_{n\to\infty} \frac{1}{n} \log R(x, T, \delta, n) = h_\mu(T) \ .$$

We refer to [58] for the proof and related results.

2.5 SRB measures.

It is natural to ask if there is a connection between entropy and Lyapunov exponents. The first relation is an inequality due to D.Ruelle.

Theorem 2.9. *Let μ be an ergodic invariant measure. The entropy of μ is less than or equal to the sum of the positive Lyapunov exponents of μ.*

We will only give a rough idea of the proof. Take a ball B of radius ϵ (small) such that $\mu(B) > 0$. Iterate k times (k not too large such that $T^k(B)$ is still small). One gets an ellipsoid $T^k(B)$, and we cover it with balls of the same radius ϵ. In order to cover $T^k(B)$ by balls of radius ϵ we need at most $N = e^{k\Theta}$ balls, where Θ is the sum of the positive exponents. Indeed, in the direction of exponent j with $\lambda_j > 0$, we stretched from a size ϵ to a size $e^{k\lambda_j}\epsilon$, and we need now $e^{k\lambda_j}$ balls of radius ϵ to cover "in this direction".

We now interpret the construction. Let $b_1, \cdots, b_N$ the balls of radius ϵ used to cover $T^k(B)$. If two initial conditions in B manage to ϵ separate before time

k, they will land in two different balls b_j and b_l (forget re-encounters). If they do not separate they will land in the same ball b_k. Therefore

$$N \le e^{k\Theta} ,$$

and using the result of Katok 2.6 we get

$$h(\mu) \le \frac{\log N}{k} \le \Theta = \sum_{\lambda_i(\mu)>0} \lambda_i(\mu) .$$

One is then eager to know when equality holds. In order to explain the answer to this question, we first need some definitions.

Consider a negative Lyapunov exponent for an ergodic invariant measure μ and let x be a typical point for this measure. Then from Oseledec theorem we know that at the linear level, there are directions at x which are contracted exponentially fast. One may wonder if such a fact also holds for the complete dynamics and not only for the linearisation. This leads to the definition of the local stable manifold.

Definition 2.3. An (open) sub-manifold W of the phase space Ω is called a local stable manifold with exponent $\gamma > 0$ if there is a constant $C > 0$ such that for any x and y in W and any integer n we have

$$d\big(T^n(x), T^n(y)\big) \le Ce^{-\gamma n} .$$

Note that it follows immediately from the definition that if W is a local stable manifold with exponent γ, then for any integer p, $T^p(W)$ is also a local stable manifold with exponent γ. When T is invertible, one can define the local unstable manifolds as the local stable manifolds of the inverse. We refer to [78] for the general case and more details. A local stable manifold containing a point x is called a local stable manifold of x.

When we have a measure μ we can consider stable manifolds except on sets of measure zero. The idea being that the size of the stable manifold depends on the point x (or more interestingly the distance of x to the boundary of the manifold). The nice picture is when points without local stable manifolds (interpreted as local stable manifolds of width zero) are of measure zero, and when there is some coherence among the local stable manifolds through the dynamics. This leads naturally to the following definitions.

Definition 2.4. Given an invariant measure μ, for a map T a coherent field of local stable manifolds for μ with exponent $\gamma > 0$ is a collection $\big(W(x)\big)_{x\in\mathcal{A}}$ of local stable manifolds for μ with exponent γ indexed by a subset $\mathcal{A}$ of full measure and such that

1. For any x in $\mathcal{A}$, $x \in W(x)$.
2. For any x in $\mathcal{A}$, $T\big(W(x)\big) \subset W\big(T(x)\big)$.

Definition 2.5. Given an invariant measure μ, for an invertible map T, a coherent field of local unstable manifolds for μ with exponent $\gamma > 0$ is a collection $\big(W(x)\big)_{x\in\mathcal{A}}$ of local unstable manifolds for μ with exponent γ indexed by a subset $\mathcal{A}$ of full measure and such that

1. For any x in $\mathcal{A}$, $x \in W(x)$.
2. For any x in $\mathcal{A}$, $T^{-1}\big(W(x)\big) \subset W\big(T^{-1}(x)\big)$.

We now state an "integrated version" of the Oseledec theorem 2.3.

Theorem 2.10. *Let μ be an ergodic invariant measure for a diffeomorphims of a compact manifold T. Let $\lambda_1 > \ldots > \lambda_k > 0$, be the collection of positive Lyapunov exponents, and $\Omega = E_1 \supsetneq E_1 \supsetneq \cdots \supsetneq E_k$ the sequence of associated linear bundles in the Oseledec theorem. Then there is a set B of full measure such that for any ϵ satisfying*

$$0 < \epsilon < \inf_{1 \le j \le k-1} \big(\lambda_j - \lambda_{j+1}\big) \;,$$

and for each $x \in B$, there exists k nested sub-manifolds

$$W_k(x) \subsetneq W_{k-1}(x) \subsetneq \ldots \subsetneq W_1(x)$$

where for any j ($1 \le j \le k$), $W_j(x)$ is a $\lambda_j - \epsilon$ local unstable manifold of x.

A similar result holds for the negative exponents.

We refer to [122] and [132] for the proof and extension. One of the big difference with the case of uniformly hyperbolic systems like axiom A systems is that there is in general no uniform lower bound on the size of the local stable or unstable manifolds.

We now state the definition of SRB measures.

Definition 2.6. Let the phase space Ω of a dynamical system be a compact Riemannian manifold. Let the map T be a C^2 diffeomorphims. An ergodic invariant measure μ is said to be an SRB measure if it has a positive Lyapunov exponent and the conditional measures on the local unstable manifolds are absolutely continuous.

In other words, one can disintegrate μ as follows, where g is any continuous function

$$\int g(x) d\mu(x) = \int_{\mathcal{W}^u} d\nu(w) \int_w g(y) \rho_w(y) d_w y \tag{28}$$

where $\mathcal{W}^u$ is the set of local unstable manifolds, $d_w y$ is a Lebesgue measure on the local unstable manifold w, $\rho_w(y)$ is a non negative density supported by w, and ν is called the transverse measure.

We can now state an important theorem due initially to Pesin and generalised by Ruelle which complements the inequality in Theorem 2.9.

Theorem 2.11. *Let the phase space Ω of a dynamical system be a compact Riemannian manifold. Let the map T be a C^2 diffeomorphims. Let μ be an ergodic invariant measure. Then μ is an SRB measure if and only if the entropy of μ is equal to the sum of its positive Lyapunov exponents.*

In particular, in dimension one, an ergodic invariant measure is absolutely continuous if and only if it has a positive exponent and its entropy is equal to this exponent. In axiom A systems, Physical and SRB measures are the same. They have another property connected with stochastic perturbations that will be mentioned later on. We mention that there are dynamical systems which do not possess SRB measures. We refer to [154] for more information.

3 Statistical Properties of Dynamical Systems.

3.1 Convergence rate in the ergodic Theorem.

It is natural to ask how fast is the convergence of the ergodic average to its limit in Theorem 2.1. We have seen in figure 10 that this convergence can be slow. At this level of generality any kind of velocity above $1/n$ can occur. Indeed Halasz and Krengel have proved the following result (see [83] for a review).

Theorem 3.1. *Consider a (measurable) automorphism T of the unit interval $\Omega = [0,1]$, leaving the Lebesgue measure $d\mu = dx$ invariant.*

i)) For any increasing diverging sequence a_1, $a_2, \cdots$, with $a_1 \geq 2$, and for any number $\alpha \in]0,1[$, there is a measurable subset $A \in \Omega$ such that $\mu(A) = \alpha$, and

$$\left| \frac{1}{n} \sum_{j=0}^{n-1} \left(\chi_A \circ T^j - \mu(A) \right) \right| \leq \frac{a_n}{n}$$

μ almost surely, for all n.

ii)) For any sequence b_1, $b_2, \cdots$, of positive numbers converging to zero, there is a measurable subset $B \in \Omega$ with $\mu(B) \in]0,1[$ such that almost surely

$$\lim_{n\to\infty} \left| \frac{1}{nb_n} \sum_{j=0}^{n-1} \left(\chi_B \circ T^j - \mu(B) \right) \right| = \infty \ .$$

In spite of these negative results, there is however an interesting and somewhat surprising general theorem by Ivanov dealing with a slightly different question.

To formulate the result we first define the sequence of down-crossings for a non-negative sequence $(u_n)_{n\in\mathbf{N}}$. Let a and b be two numbers such that $0 < a < b$. For an integer $k \geq 0$ such that $u_k \leq a$, we define the first down crossing from b to a after k as the smallest integer $n_d > k$ (if it exists) such that

i)) $u_{n_d} \leq a$,

ii)) There exists at least one integer $k < j < n_d$ such that $u_j \geq b$.

Let now (n_l) be the sequence of successive down-crossings from a to b (this sequence may be finite and even empty). We denote by $N\big(a, b, p, (u_n)\big)$ the number of successive down-crossings from b to a occurring before time p for the sequence (u_n), namely

$$N(a, b, p, (u_n)) = \sup \left\{ l \,\middle|\, n_l \leq p \right\} \ .$$

Theorem 3.2. *Let $(\Omega, \mathcal{A}, T, \mu)$ be a dynamical system. Let f be a non negative observable with $\mu(f) > 0$. Let a and b be two positive real numbers such that $0 < a < \mu(f) < b$, then for any integer r*

$$\mu\left(\left\{ x \,\middle|\, N\left(a, b, \infty, \left(f\big(T^n(x)\big) \right) \right) > r \right\} \right) \leq \left(\frac{a}{b} \right)^r \ .$$

The interesting fact about this Theorem is that the bound on the right hand side is explicit and independent of the map T and of the observable f. We refer to [81], [38], [85] for proofs and extensions.

In order to get more precise information on the rate of convergence in the ergodic theorem, one has to make some hypothesis on the dynamical system and on the observable.

3.2 Mixing and decay of correlations

If one considers the numerator of the ergodic average, namely the ergodic sum

$$S_n(f)(x) = \sum_{j=0}^{n-1} f\big(T^j(x)\big) \tag{29}$$

this can be considered as a sum of random variables, although in general not independent. It is however natural to ask if there is something similar to the central limit theorem in probability theory. To have such a theorem, one has first to obtain the limiting variance. Assuming for simplicity that the average of f is zero, we are faced with the question of convergence of the sequence

$$\frac{1}{n}\int \Big(S_n(f)(x)\Big)^2 \, d\mu(x)$$

$$= \int f^2(x)\, d\mu(x) + 2\sum_{j=1}^{n-1} \frac{n-j}{n} \int f(x)\, f\big(T^j(x)\big)\, d\mu(x) \;.$$

Here we restrict of course the discussion to observables which are square integrable. This sequence may diverge when n tends to infinity. It may also tend to zero. This is for example the case if $f = u - u \circ T$ with $u \in L^2(d\mu)$. Indeed, in that case $S_n = u - u \circ T^n$ is of order one in L^2 and not of order $\sqrt{n}$ (see [83] for more details and references).

A quantity which occurs naturally from the above formula is the auto-correlation function $C_{f,f}$ of the observable f. This is the sequence defined by

$$C_{f,f}(j) = \int f(x)\, f\big(T^j(x)\big)\, d\mu(x) \;. \tag{30}$$

If this sequence belongs to l^1, the limiting variance exists and is given by

$$\sigma_f^2 = C_{f,f}(0) + 2\sum_{j=1}^{\infty} C_{f,f}(j) \;. \tag{31}$$

This shows that the decay of the auto-correlation function (30) is an important quantity. A natural generalization of the auto-correlation is the cross correlation (often called simply the correlation) between two square integrable observables f and g. This function (sequence) is defined by

$$C_{f,g}(j) = \int f(x)\, g\big(T^j(x)\big)\, d\mu(x) - \int f\, d\mu \int g\, d\mu \;. \tag{32}$$

The second term appears in the general case when neither f nor g has zero average. The case where f and g are characteristic functions is particularly interesting. If $f = \chi_A$ and $g = \chi_B$ (assuming $\mu(A)\mu(B) \neq 0$ otherwise $C_{\chi_A,\chi_B}(n) = 0$), we have

$$C_{\chi_A,\chi_B}(n) = \int_A \chi_B \circ T^n d\mu = \mu\big(A \cap T^{-n}(B)\big) = \mu\big(T^n(x) \in B \big| A\big)\mu(A) \ .$$

If for large time the system looses memory of its initial condition, it is natural to expect that $\mu\big(T^n(x) \in B \big| x \in A\big)$ converges to $\mu(B)$. This leads to the definition of mixing. We say that for a dynamical system $(\Omega, \mathscr{B}, T)$ the T invariant measure μ is mixing if for any measurable subsets A and B of Ω, we have

$$\lim_{n\to\infty} \mu\big(A \cap T^{-n}(B)\big) = \mu(A)\mu(B) \ . \tag{33}$$

Note that this can also be written (for $\mu(B) > 0$)

$$\lim_{n\to\infty} \mu\big(A \big| T^{-n}(B)\big) = \mu(A) \ ,$$

and if the map T is invertible

$$\lim_{n\to\infty} \mu\big(T^n(A) \big| B\big) = \mu(A) \ ,$$

In other words, if we have the information that at some time the state of the system is in the subset B, then if we wait long enough (more precisely asymptotically in time), this information is lost in the sense that we get the same statistics as if we had not imposed this condition. There are many other mixing conditions including various aspects and velocity of convergence. We refer to [59] for details.

Exercise 3.1. Show that mixing implies ergodicity.

Exercise 3.2. Let T be a map with a finite periodic orbit $x_0, \ldots, x_{k-1}$ ($T(x_r) = x_{r+1 \pmod k}$ for $0 \leq r < k$). Consider the invariant probability measure obtained by putting Dirac masses with weight $1/k$ at the points of the orbit. Show that this measure is ergodic but not mixing.

Exercise 3.3. Consider the map $2x \pmod 1$ and the ergodic invariant measure μ_p of exercise 1.18. Show that this dynamical system is mixing (hint: show formula (33) for any cylinder set and conclude by approximation).

As explained before, in order to show that the ergodic sum has a (normalized) limiting variance, we need to estimate the convergence rate in (33). For general measurable sets (or for square integrable functions in (32) this is often a hopeless task. It is often the case that the rate of convergence in (32) depends on the class of functions one considers.

Note that the correlation function (32) can also be written

$$C_{f,g}(j) = \int f(x)\, U^j g(x)\ d\mu(x) - \int f\ d\mu \int g\ d\mu \ ,$$

where U is the Koopman operator (17). In the case of maps of the interval, when the invariant probability measure μ is absolutely continuous with respect to the Lebesgue measure with density h the correlations can be expressed in term of the Perron-Frobenius operator using equality (18), namely

$$C_{f,g}(j) = \int \mathcal{L}^j f(x)\, g(x)\ d\mu(x) - \int f\ d\mu \int g\ d\mu \ , \tag{34}$$

with

$$\mathcal{L}f = \frac{1}{h}P\big(hf\big)\ . \tag{35}$$

Note that since $P\big(h\big) = h$ this formula is well defined Lebesgue almost everywhere even if h vanishes by setting $\mathcal{L}f = 0$ where $h = 0$. It is now obvious from (34) that information about the spectrum of $\mathcal{L}$ translates into information on the decay of correlations. Since $\mathcal{L}$ is conjugated to the Perron-Frobenius operator P, one often studies in more details this last operator. We recall that the relation between decay of correlations and spectrum of the transition matrix is also a well known fact in the theory of Markov chains.

Consider for example the map $T(x) = 2x \pmod 1$ of the interval $[0, 1]$ (example 1.1) with the Lebesgue measure as invariant measure. If f and g belong to L^2, one can use their Fourier series $\big(f_p\big)$ and $\big(g_p\big)$, namely

$$f(x) = \sum_{p\in\mathbf{Z}} f_p e^{2\pi ipx}$$

where

$$f_p = \int_0^1 e^{-2\pi ipy} f(y)dy\ . \tag{36}$$

One gets

$$C_{f,g}(j) = \sum_{p,q} f_p g_q \int_0^1 e^{ipx} e^{iq2^j x} dx - f_0 g_0 = \sum_{q\neq 0} f_{q2^j} g_q\ . \tag{37}$$

A first brutal estimate using Schwarz inequality gives

$$\big|C_{f,g}(j)\big| \le \|g\|_2 \left(\sum_{q\neq 0} \big|f_{q2^j}\big|^2\right)^{1/2}\ . \tag{38}$$

It follows easily from $f \in L^2$ using Parseval's inequality that the last term tends to zero when j tends to infinity. In other words, we have just proved that for the Lebesgue invariant measure the map $2x \pmod 1$ is mixing. Note that it is crucial in this argument that the sum extends on $q \neq 0$, this is exactly the role of the subtraction of the last term in definition (32), see also (37).

If f can be extended to an analytic function in a domain containing the interval $[0, 1]$, its Fourier coefficients decay exponentially fast. In this case, for any $g \in L^2$ we have a doubly exponential exponential decay of the correlations.

Exercise 3.4. Prove these two statements, the first one using formula (36), the second one using formula (38).

If we only have a power law bound on the decay of the Fourier coefficients of f, we can only conclude that the correlations have a power law decay. Note that the behaviour of the Fourier coefficients is related to the regularity properties of the function. For example if $f \in C^1$ we have for any $p \neq 0$

$$\big|f_p\big| \le \frac{1}{2\pi p}\|f'\|_2\ .$$

Exercise 3.5. Prove this estimate using integration by parts in formula (36).

Using the estimate (38), we conclude that for any $g \in L^2$ and any $f \in C^1$

$$\left|C_{f,g}(j)\right| \le \Gamma \|g\|_2 \|f'\|_2 2^{-j} \tag{39}$$

with

$$\Gamma = \frac{1}{2\pi}\left(\sum_{q\neq 0}\frac{1}{q^2}\right)^{1/2} .$$

By comparing the case of f analytic and $f \in C^1$ we see that the bounds we have obtained depend on the regularity properties of the function f. One may wonder if this is an artifact of our crude estimation method or if this is indeed the case. To see this, we can consider a particular example, namely, let for $p \neq 0$

$$f_p = \frac{1}{|p|^\alpha}$$

for some fixed $1/2 < \alpha < 1$. Using Parseval's inequality, it follows at once that the function f whose Fourier series is (f_p) belongs to L^2. We take $g_p = 0$ for any p except that $g_1 = g_{-1} = 1$. We get from (37)

$$C_{f,g}(j) = 2\, 2^{-j\alpha} .$$

We conclude by comparison with (39) that the function f does not belong to C^1 and its correlation decays indeed more slowly.

Exercise 3.6. Use the same technique of Fourier series to discuss the decay of correlations for the cat map of example 1.6.

The situation described above relating the rate of decay of correlations with some kind of regularity properties of the observables is quite frequent in the theory of dynamical systems. One looks for two Banach spaces of functions on the phase space $\mathcal{B}_1$ and $\mathcal{B}_2$ and a function $C_{\mathcal{B}_1,\mathcal{B}_2}(n)$ tending to zero when n tends to infinity such that for any $f \in \mathcal{B}_1$, for any $g \in \mathcal{B}_2$ and for any integer n

$$\left|C_{f,g}(n)\right| \le C_{\mathcal{B}_1,\mathcal{B}_2}(n)\|f\|_{\mathcal{B}_1}\|g\|_{\mathcal{B}_2} . \tag{40}$$

Note that it is not necessary for $\mathcal{B}_1$ and $\mathcal{B}_2$ to be contained in $L^2(d\mu)$, they only need to be in duality in some sense (see [108]).

As explained above, this kind of estimate may follow from adequate information on the spectral theory of a Perron-Frobenius operator. We mention here without proof the case of piecewise expanding maps of the interval (example 1.2).

Theorem 3.3. *Let T be a piecewise expanding map of the interval. The associated Perron-Frobenius operator P maps the space* **BV** *of functions of bounded variations into itself. In this space, the spectrum is composed of the eigenvalue one with finite multiplicity, a finite number (may be there are none) of eigenvalues of modulus one with finite multiplicity, and the rest of the spectrum is contained into a disk centered at the origin and of radius $\sigma < 1$. Any eigenvector of eigenvalue one is a linear combination of a finite number of non negative eigenvectors with support of disjoint interiors, which are finite union of intervals permuted by the map. The eigenvalues of modulus one different from one if they exist are rational multiples of 2π and therefore do not occur in the spectrum of a sufficiently high power of P.*

We refer to [80] for a proof of this theorem, to [43] [6] for more results and extension.

We can now consider some iterate of T and choose one of the absolutely continuous invariant measure with positive density h and support an interval $[a, b]$ (recall that a finite iterate of a piecewise expanding map of the interval is also a piecewise expanding map of the interval).

In that case, in the space $\mathbf{BV}([a, b])$ of functions of bounded variation with support equal to the support of h, the Perron-Frobenius operator P has one as a simple eigenvalue (with eigenvector h), and the rest of the spectrum contained in a disk centered at the origin and of radius $\sigma < 1$. In particular, for any $1 > \rho > \sigma$, there is a constant $\Gamma_\rho > 0$ such that for any $f \in \mathbf{BV}([a, b])$ and any integer n, we have

$$\big\|P^n f\|_{\mathbf{BV}([a,b])} \le \Gamma_\rho \rho^n \big\|f\|_{\mathbf{BV}([a,b])} \ . \tag{41}$$

In particular, this immediately implies that for any $f \in \mathbf{BV}([a, b])$ and $g \in L^1(dx)$, we have

$$\big|C_{f,g}(n)\big| \le \Gamma_\rho \rho^n \|f\|_{\mathbf{BV}([a,b])} \|g\|_{L^1} \ . \tag{42}$$

In other words, we have the estimate (40) with $\mathcal{B}_1 = \mathbf{BV}([a, b])$ and $\mathcal{B}_2 = L^1$.

The method of Perron-Frobenius operators has also been used to prove the decay of correlations for Gibbs states on sub-shifts of finite type (see [20] and [131]), and for some non uniformly hyperbolic dynamical systems (see [155] and references therein). The relation between the rate of decay of correlations and other quantities is not simple. We refer to [50] for some results.

Another important method is the method of coupling borrowed from probability theory. We refer to [7] for the case of piecewise expanding maps of the interval, [24] for the case of chains with complete connexions which generalize the case of Gibbs states on sub-shifts of finite type, to [25], and to [153] for non uniformly hyperbolic dynamical systems.

We first recall that if we have two measures μ on a space X and ν on a space Y, a coupling between μ and ν is a measure on the product space such that its marginals are μ and ν respectively. Let T is a piecewise expanding map on the interval with an a.c.i.p.m. μ with a non-vanishing density h. In this situation we can define a Markov chain on the interval by the following formula for its transition probability

$$p\big(y|x\big) = \frac{h(y)}{h(x)} \delta\big(x - T(y)\big) = \frac{1}{h(x)} \sum_{z,\, T(z)=x} \frac{h(z)}{\big|T'(z)\big|} \delta(y - z) \ . \tag{43}$$

Exercise 3.7. Show that the integral over y is equal to one.

This transition probability is nothing but the kernel of the operator $\mathcal{L}$ defined in formula (35). This is a rather irregular kernel, it only charges a finite number of points, namely the points y such that $T(y) = x$. These are called the preimages of x under the map T. Nevertheless, for piecewise expanding maps of the interval the following result has been proven in [7].

Theorem 3.4. *For any piecewise expanding map on the interval with an a.c.i.p.m. μ with a non-vanishing density, there are three positive constants C, $\rho_1 < 1$ and $\rho_2 < 1$, and for any pair of points x_1 and x_2 in the interval, there is a coupling*

$\mathbf{P}_{x_1,x_2}$ *between the two Markov chains* $\big(X_n^1\big)$ *and* $\big(X_n^2\big)$ *defined by the kernel (43) starting in* x_1 *and* x_2 *respectively and such that for any* $n \in \mathbf{n}$

$$\mathbf{P}_{x_1,x_2}\left(\left|X_n^1 - X_n^2\,\right| \ > \rho_1^n\right) \le C\rho_2^n \ .$$

The statement in the original paper [7] is in fact stronger but we will only use the above form.

Note that the constants C, ρ_1 and ρ_2 do not depend on x_1 and x_2. Note also the difference with the coupling for Markov chains on finite state space. Here because of the singularity of the kernel, the realisations of $\big(X_n^1\big)$ and $\big(X_n^2\big)$ will in general not meet. However they can approach each other fast enough. The condition that h does not vanish is not restrictive. We have seen already that this is first connected with ergodicity (see Theorem 3.3). Once ergodicity is ensured, the vanishing of h can be discussed in details. We refer to [7] and [29] for the details.

We now show the application of this theorem to the proof of the decay of correlations. Le g be a function of bounded variation. We have from the definition of the Markov chains $\big(X_n^1\big)$ and $\big(X_n^2\big)$ and from the definition of the coupling

$$\big(\mathcal{L}^n g\big)(x_1) - \big(\mathcal{L}^n g\big)(x_2) = \mathbf{E}\big(g\big(X_n^1\big)\big) - \mathbf{E}\big(g\big(X_n^2\big)\big) = \mathbf{E}_{x_1,x_2}\Big(\big(g\big(X_n^1\big)\big) - \big(g\big(X_n^2\big)\big)\Big)\ .$$

Assume now g is a Lipschitz function with Lipschitz constant L_g, then we get by applying Theorem 3.4

$$\begin{aligned}
&\Big|\mathbf{E}_{x_1,x_2}\Big(\big(g\big(X_n^1\big)\big) - \big(g\big(X_n^2\big)\big)\Big)\Big| \\
&\le \mathbf{E}_{x_1,x_2}\Big(\chi_{\left|X_n^1 - X_n^2\right| \le \rho_1^n}\Big|\big(g\big(X_n^1\big)\big) - \big(g\big(X_n^2\big)\big)\Big|\Big) \\
&+\mathbf{E}_{x_1,x_2}\Big(\chi_{\left|X_n^1 - X_n^2\right| > \rho_1^n}\Big|\big(g\big(X_n^1\big)\big) - \big(g\big(X_n^2\big)\big)\Big|\Big) \\
&\le L_g\rho_1^n + 2\ C\ L_g\rho_2^n\ .
\end{aligned}$$

We now consider a correlation $C_{u,v}(n)$ between a Lipschitz function u and an integrable function v. We have easily using formula (34) and the invariance of μ

$$C_{u,v}(n) = \int\int\Big(\mathcal{L}^n u(x_1)v(x_1) - \mathcal{L}^n u(x_1)v(x_2)\Big)\,d\mu(x_1)\ d\mu(x_2)\ ,$$

We can now use the above estimate to get

$$|C_{u,v}(n)| \le \big(L_u\rho_1^n + 2\ C\ L_u\rho_2^n\big)\int |v|\ d\mu\ ,$$

which is the exponential decay of correlations as in formulas (40) and (42).

We explain here very shortly the idea of coupling on the particularly simple case of the map $T(x) = 2x \pmod 1$. In that case, the transition probability is given by

$$p\big(y|x\big) = \frac{1}{2}\sum_{z\,,\,2z=x \pmod 1}\delta(y-z) = \delta(2y-x)\ . \tag{44}$$

Let $\mathcal{P}$ be the partition of the interval $\mathcal{P} = \big\{[0,1/2],]1/2,1[$. It is easy to verify that, except for a countable number of points x, any of the 2^n atoms of the partition $\vee_{j=1}^n T^{-j}\mathcal{P}$ contains one and only one preimage of order n of a point x (the points y such that $T^n(y) = x$).

Exercise 3.8. Use the coding with the full unilateral shift on two symbols to prove this assertion (see example 1.2).

In other words for two points x and x' we have a coupling at the topological level of their preimages by putting together the preimages which are in the same atom of $\vee_{j=1}^{n} T^{-j}\mathcal{P}$. We can now write the coupling on the trajectories of the Markov chain. For any pair of points x and x' in $[0,1]$, for any integer n, and for any pair of points $\big(y_1, \ldots, y_n\big)$ and $\big(y'_1, \ldots, y'_n\big)$ of $[0,1]^2$ we define

$$\mu_{x,x'}\left(\big(y_1, \ldots, y_n\big), \big(y'_1, \ldots, y'_n\big)\right)$$

$$= \sum_{I \in \vee_{j=1}^{n} T^{-j}\mathcal{P}} \chi_I\big(y_n\big)\chi_I\big(y'_n\big) \prod_{l=1}^{n} \delta\big(2y_l - y_{l-1}\big) \prod_{l=1}^{n} \delta\big(2y'_l - y'_{l-1}\big)$$

where for convenience of notations we have defined $y_0 = x$ and $y_{0'} = x'$.

Exercise 3.9. Show that the above formula defines a measure on $\big([0,1]\times[0,1]\big)^{\mathbf{N}}$ (use Kolmogorof's Theorem, see for example [70]). Show that except for a countable number of points x and x' of the interval, this measure is a coupling between of the two Markov chains with transition kernel (44) and initial points x and x'. Show that this coupling satisfies the conclusions of Theorem 3.4.

The more general cases use basically the same ideas with the complication that the transition probability is not constant. We refer to the above mentioned references for the details.

3.3 Central limit Theorem.

In the case where the asymptotic standard deviation σ_f given by formula (31) is non zero, we say that the central limit theorem holds if

$$\lim_{n\to\infty} \mu\left(\left\{x \;\middle|\; \frac{S_n(f)(x)}{\sigma_f\sqrt{n}} \le t\right\}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2} du \;. \tag{45}$$

We emphasize that this kind of result has been only established for certain classes of dynamical systems and observables. We refer to [54] and [111] for reviews and [153] for recent results.

There are two methods which were mostly used up to now for proving this result. One is based on a result of Gordin (see [71]) which reduces the problem to the known central limit theorem for martingales. The other method can be used when a Perron-Frobenius operator with adequate spectral properties is available. We illustrate this idea for the case of piecewise expanding maps of the interval (see example 1.2 for the definition and 3.3 for the spectral result). Let f be a real function on the unit interval with bounded variations and zero average with respect to an ergodic and mixing a.c.i.p.m. with density h. According to a theorem of Paul Levy (see for example [13] or [66]), if $\sigma_f > 0$, in order to prove (45), it is enough to prove convergence of the Fourier transforms of the laws, namely that for any real θ

$$\lim_{n\to\infty} \int e^{i\theta S_n(f)(x)/\sqrt{n}} h(x)dx = e^{-\sigma_f^2\theta^2/2} \;. \tag{46}$$

To establish this result, we first define a family of operators P_v $(v \in \mathbf{BV})$ by

$$P_v u(x) = P\left(e^v u\right) \;. \tag{47}$$

Using several times relations (16) and (18), it follows that the integral in the left hand side of equation (46) is equal to

$$\int P^n_{if\theta/\sqrt{n}}\, hdx\ .$$

For large n, the function $i\theta f/\sqrt{n}$ is small and therefore we expect the operator $P_{if\theta/\sqrt{n}}$ to be in some sense near the operator P. This is indeed the case in the following sense. In the space of functions of bounded variations, it is easy to verify that for any complex number z of modulus smaller than one

$$\big\|P - P_{zf}\|_{\mathbf{BV}} \le \mathcal{O}(1)|z|\ .$$

Therefore, if $|z|$ is small enough, using Theorem 3.3 and the analytic perturbation theory around simple eigenvalues (see [88]) one can establish the following result.

Theorem 3.5. *There is a positive number η and two positive numbers Γ_1 and $\rho_1 < 1$ such that for any complex valued function v of bounded variation satisfying $\|v\|_{\mathbf{BV}} < \eta$ operator P_v has a unique simple eigenvalue $\lambda(v)$ of modulus larger than ρ_1 with eigenvector h_v and eigenform α_v (satisfying $\alpha_v\big(h_v\big) = 1$). $\lambda(v)$, h_v and α_v are analytic in v in the ball of functions of bounded variation centered at the origin and of radius η. Moreover, $\lambda(0) = 1$, $h_0 = h$, and α_0 is the integration against the Lebesgue measure. The rest of the spectrum of P_v (in the space of functions of bounded variations) is contained in the disk of radius ρ_1 and we have for any positive integer m the estimate*

$$\big\|P^m_v \Pi_v\| \le \Gamma_1 \rho_1^m\ ,$$

where Π_v is the spectral projection on $\lambda(v)$ given by

$$\Pi_v(g) = \alpha_v(g)\ h_v\ .$$

We can now apply this result with $v = i\theta f/\sqrt{n}$ and $m = n$ because the estimates are uniform. We get for n large enough such that $|\theta|/\sqrt{n} < \eta$

$$P^n_{i\theta f/\sqrt{n}} h = \lambda(i\theta f/\sqrt{n})^n h_{i\theta f/\sqrt{n}} \alpha_{i\theta/\sqrt{n}}(h) + \mathcal{O}(1)\rho_1^n\ .$$

A simple perturbation theory computation (see [88] or [35]) shows that since f has average zero

$$\lambda(zf) = 1 + \frac{z^2}{2\sigma_f^2} + \mathcal{O}\big(|z|^3\big)\ ,$$

and the result (46) follows.

For some particular classes of dynamical systems some more precise results have been established.

For Gibbs states over sub-shifts of finite type (see example 1.4 and 20), a Berry-Esseen inequality has been obtained by Coelho and Parry [34]. They proved that for such dynamical systems, for any Hölder continuous observable f (of zero average and with $\sigma_f > 0$), there is a constant $C > 0$ such that for any integer $n > 0$

$$\left|\mu\left(\left\{x \;\middle|\; \frac{S_n(f)(x)}{\sigma_f\sqrt{n}} \le t\right\}\right) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^t e^{-u^2/2}du\right| \le \frac{C}{\sqrt{n}}\ .$$

Analogous results were also obtained in the non uniformly hyperbolic case in [72].

One can also study more globally the sequence of random variables $S_n(f)/\sqrt{n}$ (with f of zero average and $\sigma_f > 0$).

Recall that the Brownian motion $B_{\cdot}$ is the unique continuous time Gaussian process with independent increments, defined for for $t \geq 0$, with $B_0 = 0$, with B_t of zero average for any $t > 0$ and such that

$$\mathbf{E}\big(B_t, B_s\big) = \min\big\{t, s\big\} \ .$$

Let $\big(\xi_n\big)$ be the sequence of random functions defined by

$$\xi_n(t) = \frac{1}{\sigma_f\sqrt{n}}\sum_{j=0}^{[nt]} f\circ T^j + \frac{1}{\sigma_f\sqrt{n}}\big(nt-[nt]\big) f\circ T^{[nt]+1} \ , \tag{48}$$

where [ ] denotes the integer part. This is a random sequence of continuous functions, the randomness comes from the choice of the initial condition with respect to the invariance measure. If we drop the last term, we obtain a random sequence of piecewise constant functions.

Exercise 3.10. Prove that for any fixed integer n, the function ξ_n defined in (48) is continuous.

There are several classes of dynamical systems where one can show that this process converges weakly to a Brownian motion (even without the last term). We state here the case of piecewise expanding map of the interval. We refer to [54] for more results and references.

Theorem 3.6. *Let T be a piecewise expanding map of the interval and μ a mixing absolutely continuous invariant probability measure with a non vanishing density h. Let f be a Lipschitz continuous function on the interval with zero μ average and with nonzero standard deviation σ_f. Then the sequence of process $\big(\xi_n(\,\cdot\,)\big)$ defined in (48) converges weakly to the Brownian motion.*

This theorem also holds without the last term in equation (48). We will give later on a proof of this theorem using exponential estimates.

There are many other results around the central limit theorem, we will just mention here a few. They have been proven under various hypothesis, and we refer the reader to the literature for the details. We consider as before the ergodic sum S_n of an observable f with zero average and finite (non zero) variance σ_f, and satisfying some adequate hypothesis.

The law of iterated logarithm says that almost surely,

$$\limsup_{n\to\infty} \frac{S_n}{\sigma_f\sqrt{2n\log\log n}} = 1 \ . \tag{49}$$

We refer to [54] and [27] for precise hypothesis and proofs.

The almost sure invariance principle (see [124] for a general approach) establishes that there is another probability measure wit a Brownian motion B_t and a sequence of random variables $\big(\tilde{S}_n\big)$ with the same joint distribution as $\big(\tilde{S}_n\big)$ such there is an integer valued random variable N almost surely finite and a constant $\delta > 0$ such that for any $n > N$

$$\big|\tilde{S}_n - \sigma_f B_n\big| \leq n^{-\delta+1/2} \ .$$

See in particular [54], [158], [27], [116] for the hypothesis and the proofs. The case of the map $3x \pmod 1$ of the unit interval with the Lebesgue invariant measure and the observable $f(x) = \chi_{[0,1]-1/2}$ is illustrated in figure 17.

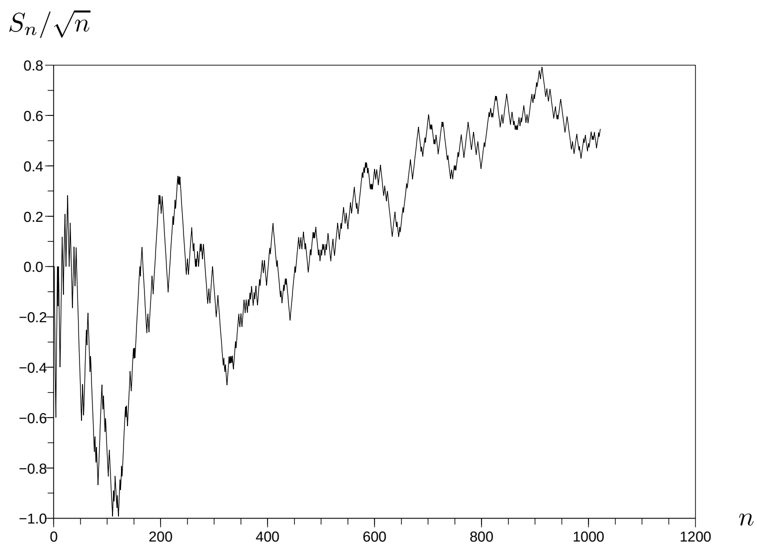

Figure 17: Evolution of $S_n(f)/\sqrt{n}$ as a function of n.

One can also prove almost sure central limit theorems, for example that for any real s, the sequence of random variables

$$\sum_{k=1}^{n} \frac{1}{k}\, \theta\left(s - \frac{S_k}{\sigma_f \sqrt{k}}\right)$$

converges almost surely to

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s} e^{-u^2/2} du \; .$$

We refer to [30] for the hypothesis, proofs and references. An analog of the Erdös-Renyi asymptotic is also proven in [30].

Moderate deviations have been treated in [53].

3.4 Large deviations.

Beyond the central limit theorem and its variants and corrections, one can wonder about the probability that an ergodic sum deviates from its limit. In other words, if μ is an ergodic invariant measure and$f$ an integrable observable with zero average, one can ask for the probability of a large deviation, namely for $t > 0$ this is the quantity

$$\mu\left(\left\{x \,\middle|\, S_n(f) > nt\right\}\right) \ . \tag{50}$$

One can look for a similar quantity for $t < 0$ (or equivalently change f to $-f$). This quantity often decays exponentially fast with n, and it is therefore convenient to define the large deviation function $\varphi(t)$ by

$$\varphi(t) = -\lim_{n\to\infty}\frac{1}{n}\log\mu\left(\left\{x \,\middle|\, S_n(f) > nt\right\}\right) \ . \tag{51}$$

Of course the limit may not exist and one can take the limsup or the liminf. In order to investigate this quantity, one can use the analogy with statistical mechanics (see [103] and [131]). There one first computes another quantity called the pressure function and defined in the present context (for $z \in \mathbf{C}$) by

$$\Xi(z) = \lim_{n\to\infty}\frac{1}{n}\log\left(\int e^{zS_n(f)}d\mu\right) \ . \tag{52}$$

As we will see below it is often convenient to take for z a complex number. Of course the limit may not exist. We now explain informally the relation between the large deviation function φ and the pressure. From the definition of the pressure (52), we have

$$\int e^{zS_n(f)}d\mu \approx e^{n\Xi(z)} \ .$$

On the other hand, if z is real we have from the definition of φ (51) and Chebychev's inequality

$$\int e^{zS_n(f)}d\mu \geq e^{nzt}\mu\left(\left\{x \,\middle|\, S_n(f) > nt\right\}\right) \approx e^{nzt}e^{-n\varphi(t)} \ .$$

We conclude immediately from these two relations that for any real z

$$\varphi(t) \geq zt - \Xi(z) \ .$$

Since this should hold for any real z, we conclude that

$$\varphi(t) \geq \sup_{z\in\mathbf{R}}\big(zt - \Xi(z)\big) \ .$$

The right hand side of this inequality is called the Legendre transform of the pressure. This argument can be made rigorous under certain hypothesis on the pressure (basically differentiability which in statistical mechanics correspond to the absence of phase transition). One can also establish under some hypothesis that the equality holds. We refer to [125], [103] and [64] for the details. We now explain briefly how to prove the existence of the limit in the simple case of piecewise expanding maps of the interval. We consider the case of μ the a.c.i.p.m. with density h and an observable f of bounded variation. Using the properties of the Perron-Frobenius operator, in particular formulas (16) and (17) we get

$$\int e^{zS_n(f)}d\mu = \int e^{zS_n(f)}h dx = \int P^n_{zf}(h) \tag{53}$$

where P_{zf} is the operator defined in (47).

Exercise 3.11. Prove this equality.

Intuitively the integral should behave like the eigenvalue $\lambda(z)$ of largest modulus of the operator P_{zf}. For values of z of modulus not too large, Theorem 3.5 shows that the peripheral spectrum is reduced to a simple eigenvalue and we get immediately

$$\Xi(z) = \log \lambda(z) \ .$$

Differentiability of the pressure can be derived similarly, and as explained before this allows to obtain the large deviation function in a neighborhood of the origin (even analyticity of the pressure follows from the analyticity properties of $\lambda(z)$ and the fact that this function does not vanish near the origin). For the case of (piecewise regular) Markov maps of the interval and for Gibbs states over sub-shifts of finite type (with Hölder potentials), one can show that for Hölder continuous functions f, the limit in (52) exists for any real z and is regular. In those cases one can construct the large deviation function for all t (more precisely between the supremum and the infimum of the observable f). We refer to [20] and [131] for a detailed study of these cases. We refer to [157] for a general approach to the large deviations of dynamical systems.

3.5 Exponential estimates.

The large deviation results are very precise but as we saw above they require proving that the pressure exist. This is often a difficult problem and one would prefer to obtain less precise estimates at a lower cost. The concentration phenomena has been known for a long time (in some sense already to Gibbs and others) and was revived ten year ago by Talagrand mostly in the context of independent random variables (see [146], and [106]). Several results in the non independent case have been obtained since and it turns out that one can prove such results for some classes of dynamical systems. To be more precise I will now state a definition. As usual, we consider a discrete time dynamical system given by a map T in a phase space Ω, equipped with a metric d, and with an ergodic invariant measure μ. I will assume that Ω is a metric space. A real valued function K on Ω^n will be said componentwise Lipschitz if for any $1 \le j \le n$, the constant $L_j(K)$ defined by

$$L_j(K) =$$

$$\sup_{x_1,\ldots,x_n,y} \frac{\big|K\big(x_1,\ldots,x_{j-1},x_j,x_{j+1},\ldots,x_n\big) - K\big(x_1,\ldots,x_{j-1},y,x_{j+1},\ldots,x_n\big)\big|}{d(x_j,y)} \tag{54}$$

is finite. In other words, we assume the function to be Lipschitz in each component with a uniformly bounded Lipschitz constant.

Definition 3.1. We say that the measure μ satisfies the exponential inequality for the map T if there are two constants $C_1 > 0$ and $C_2 > 0$ such that for any integer $n > 0$, and for any componentwise Lipschitz function K of n variables we have

$$\int e^{K\left(x,T(x),\ldots T^{n-1}(x)\right) - \mathbf{E}(K)} d\mu(x) \le C_1 e^{C_2 \sum_{j=1}^{n} L_j(K)^2} \tag{55}$$

where

$$\mathbf{E}\big(K\big) = \int K\big(x, T(x), \ldots T^{n-1}(x)\big) d\mu(x) \ .$$

Remark 3.1. There are several important facts to emphasize about this definition.

i) Comparing with the definition of the pressure (52) one sees that we have on the left hand side the same kind of integral with $K = S_n(f)$.

ii) However we require here an estimation for a larger class of functions, not only ergodic sums.

iii) On the other hand, contrary to the case of the pressure, we require only an upper bound, not the existence of a precise limit.

iv) It is also worse emphasizing that the estimate is assumed for any n, not only as a limit.

v) One often exploits the exponential inequality by using a Chebychev inequality (see examples below).

As was mentioned before, there is a relation between this exponential estimate and concentration. This relation is somewhat analogous to the relation between pressure and large deviation function explained in the previous section. As an example we consider a sub-shift of finite type (see example 1.4) on the finite alphabet $\mathcal{A}$. We consider on the phase space a metric d_ζ (see 19) denoted below by d. For a fixed integer $p > 0$, let A be a measurable subset of Ω^p such that

$$\alpha = \mu\left(\left\{x \,\middle|\, \big(x, T(x), \ldots, T^{p-1}(x)\big) \in A\right\}\right) > 0.$$

For a given $\epsilon > 0$, denote by B_ϵ the neighborhood of A, given by

$$B_\epsilon = \left\{ \big(x_1, \ldots, x_p\big) \,\middle|\, \exists \big(y_1, \ldots, y_p\big) \in A \text{ such that } \frac{1}{p} \sum_{j=1}^{p} d\big(x_j, y_j\big) \leq \epsilon \right\} \ .$$

Let K_A be the function defined on Ω^p by

$$K_A\big(x_1, \ldots, x_p\big) = \inf_{(y_1, \ldots, y_p) \in A} \ \frac{1}{p} \sum_{j=1}^{p} d\big(x_j, y_j\big) \ .$$

This function measures the average distance to A of the different components of the vector $\big(x_1, \ldots, x_p\big)$. Note that with this definition, we have $B_\epsilon = \{K_A \leq \epsilon\}$. It is easy to verify that K_A is a componentwise Lipschitz function with all the Lipschitz constants L_j equal to $1/p$.

Exercise 3.12. Prove this assertion.

Assume the measure μ satisfies the exponential inequality (55). Then we get for any real number β

$$\int e^{\beta K_A \left(x, T(x), \ldots, T^{p-1}(x)\right)} d\mu(x) \leq C_1 e^{\beta \mathbf{E}(K_A)} e^{\beta^2 C_2 / p} \ .$$

By Chebyschev's inequality, we immediately derive

$$\mu\big(\big\{x \,\big|\, K_A\big(x, T(x), \ldots, T^{p-1}(x)\big) > \epsilon\big\}\big) \le C_1 e^{-\beta\epsilon} e^{\beta \mathbf{E}(K_A)} e^{\beta^2 C_2/p} \ . \tag{56}$$

It remains to estimate $\mathbf{E}\big(K_A\big)$. For the particular case at hand this can be done easily. Observe that $-\beta K_A$ is also componentwise Lipschitz, and we can apply the exponential inequality to this function. Equivalently we can change β into $-\beta$. We get

$$\int e^{-\beta K_A\left(x, T(x), \ldots, T^{p-1}(x)\right)} d\mu(x) \le C_1 e^{-\beta \mathbf{E}(K_A)} e^{\beta^2 C_2/p} \ .$$

We now observe that if $\big(x, T(x), \ldots, T^{p-1}(x)\big)$ belongs to A, then

$$K_A\big(x, T(x), \ldots, T^{p-1}(x)\big) = 0 \ .$$

Therefore we have

$$\alpha \le \int e^{-\beta K_A\left(x, T(x), \ldots, T^{p-1}(x)\right)} d\mu(x) \le C_1 e^{-\beta \mathbf{E}(K_A)} e^{\beta^2 C_2/p} \ ,$$

which implies

$$e^{\beta \mathbf{E}(K_A)} \le C_1 \alpha^{-1} e^{\beta^2 C_2/p} \ .$$

Combining with equation (56) we get

$$\mu\big(\big\{x \,\big|\, K_A\big(x, T(x), \ldots, T^{p-1}(x)\big) > \epsilon\big\}\big) \le C_1^2 \alpha^{-1} e^{2\beta^2 C_2/p} e^{-\beta\epsilon} \ .$$

since this is true for any real β we can take the optimal value

$$\beta = \frac{\epsilon p}{4 C_2}$$

and obtain the bound

$$\mu\big(B^c_{epsilon}\big) = \mu\big(\big\{x \,\big|\, K_A\big(x, T(x), \ldots, T^{p-1}(x)\big) > \epsilon\big\}\big) \le C_1^2 \alpha^{-1} e^{-\epsilon^2 p/(8C_2)} \ .$$

We can now see the concentration phenomena. Even if α is small (for example equal to $1/2$), and for a fixed (small) ϵ, if p is large enough ($p \gg \epsilon^{-2} \log \alpha^{-1}$, the set B^c_ϵ has a very small measure. This was illustrated by Talagrand by saying that in large dimension (p large), sets of measure $1/2$ (here A) are big (in the sense that a small neighborhood B_ϵ has almost full measure). Different methods have been used in the non independent case like information inequalities, coupling and extensions of the Perron-Frobenius operator.

We give another useful application to the supremum of ergodic sums (29). Let g be a Lipschitz function on the phase space with zero average, and consider the sequence of its ergodic sums $\big(S_n\big)$ 29. One is often interested in a bound for the probability that $\sup_{1\le j\le n} S_j$ is large. This can be obtained from an exponential estimate. We first use again Pisier inequality, namely for a fixed number β to be chosen adequately later on we have

$$\mathbf{E}\left(e^{\beta \sup_{1\le j\le n} S_j}\right) \le \sum_{j=1}^{n} \mathbf{E}\left(e^{\beta S_j}\right) \ .$$

We now use the exponential estimate to bound the right hand side. For this purpose consider the function

$$K\big(x_1, \dots, x_j\big) = \sum_{m=1}^{j} g(x_m) \ .$$

If $\big(y_1, \dots, y_j\big)$ is another sequence of points in the phase space with $y_l = x_l$ for any $l \neq m$, we have

$$\left|\sum_{m=1}^{j} g(x_m) - \sum_{m=1}^{j} g(y_m)\right| \leq L_g \big|x_j - y_j\big|$$

where L_g is the Lipschitz constant of g. Therefore, for any $j \leq n$ we have from the exponential inequality (55)

$$\mathbf{E}\left(e^{\beta S_j}\right) \leq C_1 e^{C_2 j \beta^2 L_g^2} \ .$$

We conclude that for any $\beta > 0$

$$\mathbf{E}\left(e^{\beta \sup_{1\leq j\leq n} S_j}\right) \leq n C_1 e^{C_2 n \beta^2 L_g^2} \ .$$

From Chebychev's inequality, we get

$$\mathbf{P}\Big(\sup_{1\leq j\leq n} S_j > s\Big) \leq e^{-\beta s}\mathbf{E}\left(e^{\beta \sup_{1\leq j\leq n} S_j}\right) \leq n C_1 e^{-\beta s} e^{C_2 n \beta^2 L_g^2} \ .$$

We can now optimise over β to get

$$\mathbf{P}\Big(\sup_{1\leq j\leq n} S_j > s\Big) \leq n C_1 e^{-s^2/(4 n C_2 L_g^2)} \ .$$

Note that this implies that $\sup_{1\leq j\leq n} S_j$ is unlikely to be much larger than $\sqrt{n \log n}$ which is essentially the scale of the law of iterated logarithm (49). We refer to ([30]) and references therein for more results in this direction.

As another application of exponential estimates we give a proof of Theorem 3.6.

Proof of Theorem 3.6. We first recall that to prove that the sequence of processes $\big(\xi_n\big)$ defined in (48), namely

$$\xi_n(t) = \frac{1}{\sigma_f \sqrt{n}} \sum_{j=0}^{[nt]} f \circ T^j + \frac{1}{\sigma_f \sqrt{n}} \big(nt - [nt]\big) f \circ T^{[nt]+1} \ ,$$

converges to the Brownian motion, we have to prove two things (see [70] or [13]).

i) All the marginal distributions $\mathbf{P}\big(\xi_n(t_1) > s_1, \dots, \xi_n(t_k) > s_k\big)$ converge to those of the Brownian motion. This can be proven using the same technique as for the proof of the central limit theorem. We leave it as an exercise to the reader.

ii) The sequence $\big(\xi_n\big)$ is tight.

Recall (see [70] or [13]) that tightness follows if we can prove that there exists three positive numbers α, γ and H such that for any $t_1 \geq 0$ and $t_2 \geq 0$ we have for any integer n

$$\mathbf{E}\Big(\big|\xi_n(t) - \xi_n(t')\big|^{\alpha}\Big) \leq H\,|t-t'|^{1+\gamma}\,.$$

We can of course assume $t' > t$, and we observe that if $t' - t > 1/n$,

$$\big|\xi_n(t) - \xi_n(t')\big| \leq \left|\frac{1}{\sigma_f\sqrt{n}}\sum_{[nt]\leq j\leq [nt']} f\circ T^j\right| + \frac{4|t-t'|^{1/2}L_f}{\sigma_f}$$

where L_f is the Lipschitz constant of f. On the other hand, it is left as an exercise to the reader to check that if $0 < t' - t \leq 1/n$, then

$$\big|\xi_n(t) - \xi_n(t')\big| \leq \frac{4|t-t'|^{1/2}L_f}{\sigma_f}\,.$$

Therefore it is enough to prove that for some $\alpha > 2$, $\gamma > 0$ and $H > 0$, we have

$$\mathbf{E}\left(\left|\frac{1}{\sigma_f\sqrt{n}}\sum_{[nt]\leq j\leq [nt']} f\circ T^j\right|^{\alpha}\right) \leq H\,|t-t'|^{1+\gamma}\,. \tag{57}$$

This is where we will use the exponential inequality. For a fixed integer $p \geq 0$, define the function K_p of p variables $y_1, \ldots, y_p$, by

$$K_p(y_1, \ldots, y_p) = \frac{1}{\sigma_f\sqrt{n}}\sum_{j=1}^{p} f(y_j)\,.$$

It is left to the reader to verify that for any integer $1 \leq j \leq p$, the Lipschitz constant of the function K_p satisfies

$$L_j(K_p) \leq \frac{L_f}{\sigma_f\sqrt{n}}\,.$$

We now observe that (recall that $t' > t$)

$$\frac{1}{\sigma_f\sqrt{n}}\sum_{[nt]\leq j\leq [nt']} f\big(T^j(x)\big) = K_{[nt']-[nt]+1}\big(T^{[nt]}(x), \ldots, T^{[nt']}(x)\big)\,.$$

We can now apply the exponential inequality (55) to conclude that for any real β we have for $t' - t > 1/n$

$$\mathbf{E}\left(e^{\beta\left(\sum_{[nt]\leq j\leq [nt']} f\circ T^j\right)}\right) \leq C_1\; e^{C_2\beta^2 L_f^2([nt']-[nt])/(n\sigma_f^2)} \leq C_1\; e^{2C_2\beta^2 L_f^2(t'-t)/\sigma_f^2}\,.$$

Since the same inequality holds with β changed into $-\beta$ we get

$$\mathbf{E}\left(\cosh\left(\beta\sum_{[nt]\leq j\leq [nt']} f\circ T^j\right)\right) \leq C_1\; e^{2C_2\beta^2 L_f^2(t'-t)/\sigma_f^2}\,.$$

Since the Taylor series of the hyperbolic cosine has only positive terms we get for any integer k

$$\frac{\beta^{2k}}{2k!}\mathbf{E}\left(\left(\sum_{[nt]\le j\le[nt']} f\circ T^j\right)^{2k}\right)\le C_1\; e^{2C_2\beta^2L_f^2(t'-t)/\sigma_f^2}\;,$$

or in other words

$$\mathbf{E}\left(\left(\sum_{[nt]\le j\le[nt']} f\circ T^j\right)^{2k}\right)\le C_1\;2k!\;\beta^{-2k}\;e^{2C_2\beta^2L_f^2(t'-t)/\sigma_f^2}\;.$$

If we optimise over the choice of β we get

$$\mathbf{E}\left(\left(\sum_{[nt]\le j\le[nt']} f\circ T^j\right)^{2k}\right)\le (t'-t)^kM_k$$

where

$$M_k=C_1\;2k!\;\inf_{s>0}s^{-2k}\;e^{2C_2s^2L_f^2/\sigma_f^2}\;.$$

We can take for example $k=2$, and this proves inequality (57) for $\alpha=4$ and $\gamma=1$. □

The same technique can be used to prove Theorem 3.6 without the last term in equation (48). One has to use the Skorokhod topology (see [13] or [70]). We leave the details to the interested reader.

We will see later on some other important consequences of the exponential inequality (55). We now give some idea how it can be proved for piecewise expanding maps of the interval using coupling.

Theorem 3.7. *For a piecewise expanding map T of an interval with a mixing a.c.i.p.m. with non vanishing density h, the exponential inequality (55) holds.*

Proof. The first step in the proof is a doubling argument which is often used in the case of the variance. From Jensen inequality, we have

$$e^{-\mathbf{E}(K)}\le\int e^{-K(y,T(y),\ldots T^{n-1}(y))}h(y)\;dy$$

therefore

$$\int e^{K(x,T(x),\ldots T^{n-1}(x))-\mathbf{E}(K)}h(x)\;dx$$

$$\le\int\int e^{K(x,T(x),\ldots T^{n-1}(x))-K(y,T(y),\ldots T^{n-1}(y))}h(x)\;h(y)\;dx\;dy\;.$$

We now perform a change of variables and use $T^{n-1}(x)$ and $T^{n-1}(y)$ as the integration variables, as in formula (13). As in formula (1.1) we can define

$$p\big(y\big|x\big)=\frac{h(y)}{h(x)}\delta(x-T(y))=\frac{1}{h(x)}\sum_{z\;,\;T(z)=x}\frac{h(z)}{\big|T'(z)\big|}\delta(y-z)\;.$$

We now have

$$\int\int e^{K\left(x,T(x),\ldots T^{n-1}(x)\right)-K\left(y,T(y),\ldots T^{n-1}(y)\right)}h(x)\;h(y)\;dx\;dy\;=$$

$$\int\int e^{K\left(x_1,\ldots,x_n\right)-K\left(y_1,\ldots y_n\right)}\prod_{j=1}^{n-1}p\left(x_j\middle|x_{j+1}\right)\prod_{j=1}^{n-1}p\left(y_j\middle|y_{j+1}\right)\,h\left(x_n\right)\,h\left(y_n\right)\,dx_n\,dy_n\ . \tag{58}$$

We will now make use of the following general inequality due to Hoeffding (see for example [57]). Let ν be a probability measure on a space Ω and g a bounded function. Then

$$\int_\Omega e^{g-\mathbf{E}(g)}d\nu\le e^{\mathrm{Osc}(g)^2/8} \tag{59}$$

where the oscillation of g is defined by

$$\mathrm{Osc}(g)=\max_{x\in\Omega}g(x)-\min_{y\in\Omega}g(y)\ .$$

Exercise 3.13. Prove the Hoeffding inequality with the exponent 1/2 instead of 1/8. Hint: prove and use that the function $l(t)=\log\int\exp\big(t(g-\mathbf{E}(g)\big)d\nu$ is convex, and estimate from above its second derivative. The constant 1/8 is optimal, see [57]) for a proof and references.

We will now use recursively Hoeffding's inequality in the right hand side of equation (58). We first use for ν the measure

$$d\nu\left(x_1,y_1\right)=p\left(dx_1\middle|x_2\right)\,p\left(dy_1\middle|y_2\right)\ .$$

We get immediately for fixed $x_2,\ldots,x_n$ and $y_2,\ldots,y_n$ using Hoeffding's inequality

$$\int e^{K\left(x_1,\ldots,x_n\right)-K\left(y_1,\ldots y_n\right)}d\nu\left(x_1,y_1\right)\le e^{K_1\left(x_2,\ldots,x_n\right)-K_1\left(y_2,\ldots,y_n\right)}e^{Osc_1(K)^2/4}$$

where

$$K_1\left(y_2,\ldots,y_n\right)=\int K\left(x_1,\ldots,x_n\right)p\left(dx_1\middle|x_2\right)$$

and

$$\mathrm{Osc}_1(K)=\sup_{x_2,\ldots,x_n}\left(\max_{x_1\in\Omega}K\left(x_1,\ldots,x_n\right)-\min_{x_1\in\Omega}K\left(x_1,\ldots,x_n\right)\right)\ .$$

We leave to the reader to apply iteratively this argument and show that

$$\int e^{K\left(x,T(x),\ldots T^{n-1}(x)\right)-\mathbf{E}(K)}h(x)\,dx\le e^{\sum_{j=1}^n\mathrm{Osc}_j(K_j)^2/4}\ ,$$

where $K_1=K$, for $2\le l\le n-1$ the function K_l is defined by

$$K_l\left(x_l,\ldots,x_n\right)=\int K\left(x_1,\ldots,x_n\right)\prod_{r=1}^{l-1}p\left(dx_r\middle|x_{r+1}\right)\ ,$$

and Osc_j denotes the oscillation with respect to x_j, namely

$$\mathrm{Osc}_j(K_j)=\sup_{x_{j+1},\ldots,x_n}\left(\max_{x_j\in\Omega}K_j\left(x_j,\ldots,x_n\right)-\min_{x_j\in\Omega}K_j\left(x_j,\ldots,x_n\right)\right)\ .$$

In the last step, one uses Hoeffding's inequality with respect to the measure $h\left(x_n\right)\,h\left(y_n\right)\,dx_n\,dy_n$. This sequence of estimates is similar to the proof of

Azuma's inequality, see [57]. We now estimate each term in the exponential using the coupling of Theorem 3.4. We have for any u, v and $x_{l+1},\ldots,x_n$ in the interval

$$K_l\big(u,x_{l+1},\ldots,x_n\big)-K_l\big(v,x_{l+1},\ldots,x_n\big)$$

$$=\int d\mu_{u,v}\Big(\big(x_1,\ldots,x_{l-1}\big),\big(y_1,\ldots,y_{l-1}\big)\Big)$$

$$\Big[K\big(x_1,\ldots,x_{l-1},u,x_{l+1},\ldots,x_n\big)-K\big(y_1,\ldots,y_{l-1},v,x_{l+1},\ldots,x_n\big)\Big]$$

$$=\int d\mu_{u,v}\Big(\big(x_1,\ldots,x_{l-1}\big),\big(y_1,\ldots,y_{l-1}\big)\Big)$$

$$\begin{array}{ll}\sum_{r=0}^{l-2}\Big[& K\big(x_1,\ldots,x_r,x_{r+1},y_{r+2},\ldots,y_{l-1},u,x_{l+1},\ldots,x_n\big)\\ \quad - & K\big(x_1,\ldots,x_r,y_{r+1},y_{r+2},\ldots,y_{l-1},v,x_{l+1},\ldots,x_n\big)\ \Big]\end{array}$$

where in this last expression the indices out of range should be left out. This immediately implies

$$\big|K_l\big(u,x_{l+1},\ldots,x_n\big)-K_l\big(v,x_{l+1},\ldots,x_n\big)\big|$$

$$\le\sum_{r=0}^{l-2}L_{r+1}(K)\int\int\big|x_{r+1}-y_{r+1}\big|d\mu_{u,v}\Big(\big(x_1,\ldots,x_{l-1}\big),\big(y_1,\ldots,y_{l-1}\big)\Big)\ .$$

Using Theorem Theorem 3.4, we get that there are two positive constants C' and $0<\rho<1$ such that for any $0\le r\le l-2$, and for any u and v in the interval we have

$$\int\int\big|x_{r+1}-y_{r+1}\big|d\mu_{u,v}\Big(\big(x_1,\ldots,x_{l-1}\big),\big(y_1,\ldots,y_{l-1}\big)\Big)\le C'\rho^{l-r}\ ,$$

and this implies for any $1\le j\le n$

$$\mathrm{Osc}_j(K_j)\le C'\sum_{q=1}^{j}L_q\rho^{j-q}\ .$$

Therefore by Schwarz inequality

$$\sum_{j=1}^{n}\mathrm{Osc}_j(K_j)^2\le {C'}^2\sum_{j=1}^{n}\left(\sum_{q=1}^{j}L_q\rho^{j-q}\right)^2$$

$$\le {C'}^2\sum_{j=1}^{n}\sum_{q=1}^{j}L_q^2\rho^{j-q}\sum_{p=1}^{j}\rho^{j-p}\le {C'}^2(1-\rho)^{-2}\sum_{q=1}^{n}L_q^2\ .$$

$\square$

4 Measurements.

In this section we will explain some of of the methods developed for the measurement of various interesting quantities for dynamical systems and more generally for stochastic processes. These methods can be applied to the output of numerical simulations or experimental data. They are usually targeted to a (large) unique realisation of the process. More precisely, the system is observed through an observable g (a real or vector valued function on the phase space) and for an initial condition x one has a recording of the sequence

$$\big(X_n\big)_{0\le n\le N} = \big(g\big(T^n(x)\big)\big)_{0\le n\le N} \ . \tag{60}$$

This is somewhat different from the usual statistical setting where one assumes to have a sequence of independent observations. In the sequel we will always assume that an ergodic and invariant measure μ has been selected and that the in initial condition x is typical for this measure. This is of course hard to verify in a numerical situation and even impossible in an experimental setting, this is why the SRB assumption is relevant. In any case, we are exactly in the situation, common in statistics, of having the recording of the observation of a finite piece of one trajectory of a stochastic process. We will assume for simplicity that the phase space is $\mathbf{R}^d$ although the methods can be easily extended to the case of differentiable manifolds.

4.1 Correlation function and power spectrum.

One of the easiest thing to determine from the data is the (auto) correlation function. It follows immediately from the ergodic Theorem and definition (30) that for μ almost every x and for any square observable g (with zero average)

$$C_{g,g}(k) = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} g\big(T^j(x)\big)g\big(T^{j+k}(x)\big) \ . \tag{61}$$

In figure 18 we show the correlation function computed with this formula for the map $3x \pmod 1$ of the unit interval, and the observable $x-1/2$. The figure shows two applications of formula (61), one with $n=1024$ terms, the other one with $n=262144$ terms.

In practice, the measure μ is often unknown and one cannot subtract to g its average in order to get a function with zero average. However one can use again the ergodic theorem to estimate the average of g, namely

$$\int g d\mu = \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} g\big(T^j(x)\big) \ .$$

For particular classes of dynamical systems, the convergence rate can be estimated using various techniques, we refer to [31] for some examples using the exponential estimate or De Vroye inequality.

Another quantity often used in practice is the power spectrum W_g (see [91]). This is the Fourier transform of the correlation function, namely

$$W_g(u) = \sum_{k\ge 0} C_{g,g}(k)e^{iku} \ .$$

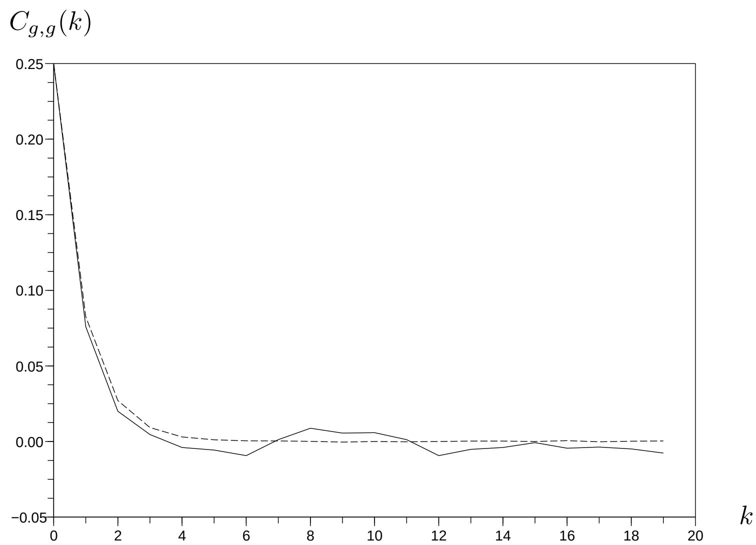

Figure 18: Correlation function for the Lebesgue measure, the map $3x \pmod 1$, and the observable $x - 1/2$. The solid line is for $n = 1024$ and the dash line for $n = 262144$ in formula (61).

If $< g >$ denotes the average of g, using the ergodic theorem one is lead to guess that the so-called periodogram $I_n(u, x)$ defined by

$$I_n(u, x) = \frac{1}{n} \left| \sum_{k=0}^{n-1} \big(g\big(T^k(x)\big)\big) - < g > e^{iku} \right|^2 \tag{62}$$

is a good approximation. Unfortunately, although it is well known that for any u

$$\lim_{n\to\infty} \int I_n(u, x) d\mu(x) = W_g(u) \ ,$$

the periodogram does not converge μ almost everywhere in x (it converges to a distribution valued process). This is illustrated in figure 19 which shows the periodogram (62) of the function $\chi_{[0,1/2]} - 1/2$ for the map $3x \pmod 1$ of the unit interval (and random initial condition for the Lebesgue measure).

One can prove under adequate conditions that the right quantity to look at is the integrated periodogram J_n given by

$$J_n(u, x) = \int_0^u \frac{1}{n} \left| \sum_{k=0}^{n-1} \big(g\big(T^k(x)\big)\big) - < g > e^{iks} \right|^2 ds \ . \tag{63}$$

Namely, for any u, and for μ almost every x we have

$$\lim_{n\to\infty} J_n(u, x) = \int_0^u W_g(s) ds \ .$$

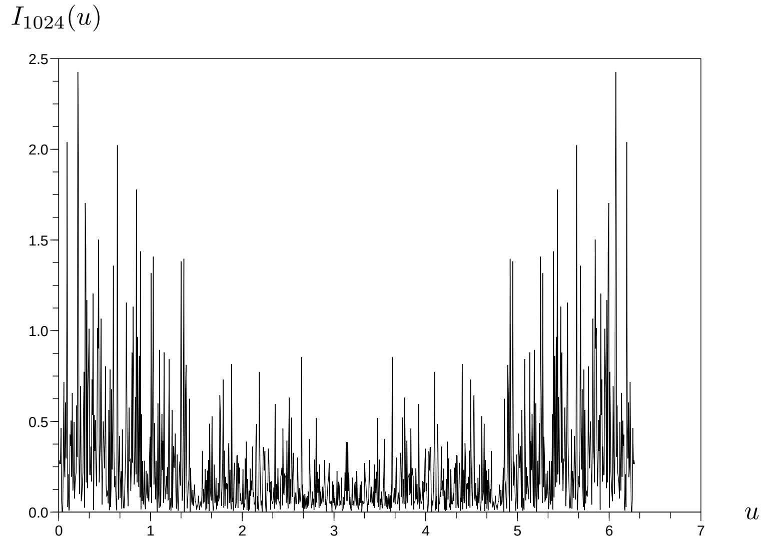

Figure 19: Periodogram of the function $\chi_{[0,1/2]} - 1/2$, with $n = 1024$ in formula (62).

Figure 20 shows the integrated periodogram for of the function $\chi_{[0,1/2]} - 1/2$ for the map $3x \pmod 1$ of the unit interval (and random initial condition for the Lebesgue measure).

We refer to [31] for the case where the average of g has to be estimated from the data and for (uniform) estimates on the convergence rate under some assumptions on the dynamical system using an exponential or a De Vroye inequality. Other algorithms use local averages of the periodogram. There is also an algorithm of Burg based on the maximum entropy principle. We refer to [109] for other results.

One can see in the power spectrum picks of periodic components and fat spectrum from noise of dynamical (chaotic) or experimental origin. For example for the map $x \to 1 - 1.7x^2$ of the interval $[-1, 1]$, figure 21 shows the correlation function and the power spectrum.

The information correlation functions are also useful in analysing time series of dynamical systems. If U and V are two independent random variables we have obviously

$$\mathbf{P}\big(U = a,\ V = b\big) = \mathbf{P}\big(U = a\big)\ \mathbf{P}\big(V = b\big)\ .$$

To test if the probability $\mathbf{P}\big(U = a,\ V = b\big)$ is a product one uses the relative entropy between this measure and the product measure

$$D = \sum_{a,b} \mathbf{P}\big(U = a\big)\mathbf{P}\big(V = b\big) \log \frac{\mathbf{P}\big(U = a,\ V = b\big)}{\mathbf{P}\big(U = a\big)\mathbf{P}\big(V = b\big)}\ .$$

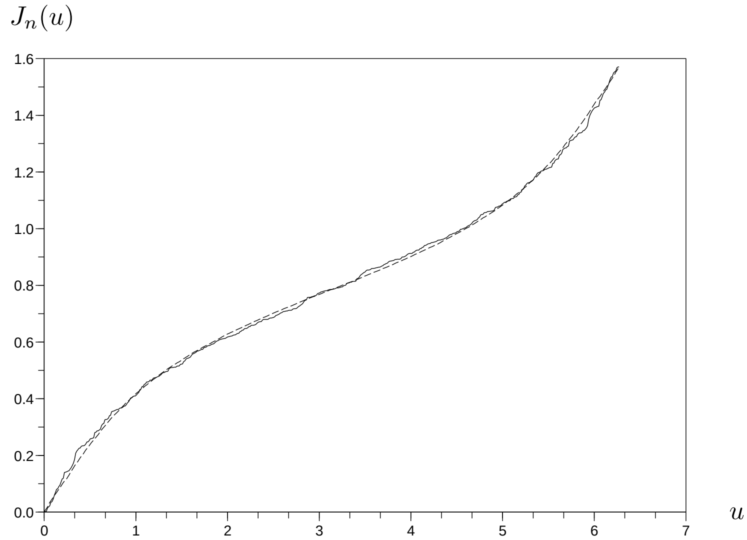

Figure 20: Integrated periodogram of the function $\chi_{[0,1/2]} - 1/2$ for the map $3x \pmod 1$. The dash curve corresponds to $n = 32768$ in formula (63) and the full curve to $n = 512$.

One can also use the symmetrical expression

$$\sum_{a,b} \mathbf{P}\big(U = a,\ V = b\big) \log \frac{\mathbf{P}\big(U = a\big)\mathbf{P}\big(V = b\big)}{\mathbf{P}\big(U = a,\ V = b\big)} \ .$$

In practice, U is taken as the value of an observable at the initial time, and V its value at a later time $k : V = U \circ T^k$. The ergodic theorem is used to compute the probabilities (measure), it is simpler to use observables taking only finitely many values. For the map $x \to 1.8x^2$ on the phase space $[-1, 1]$ we take for example the characteristic function of the interval $[0, 1] : U = \chi_{[0,1]}$. Therefore

$$\mathbf{P}(U = 1) = \lim_{N \to \infty} \frac{1}{N} \sum_{j=0}^{N-1} \chi_{[0,1]}\big(T^j(x)\big) \ ,$$

and if $V = U \circ T^k$

$$\mathbf{P}(U = 1, V = 1) = \lim_{N \to \infty} \frac{1}{N-k} \sum_{j=0}^{N-k-1} \chi_{[0,1]}\big(T^j(x)\big)\chi_{[0,1]}\big(T^{j+k}(x)\big) \ .$$

Figure 22 shows the information correlation for the map $x \to 1 - 1.8x^2$ of the interval $[-1, 1]$.

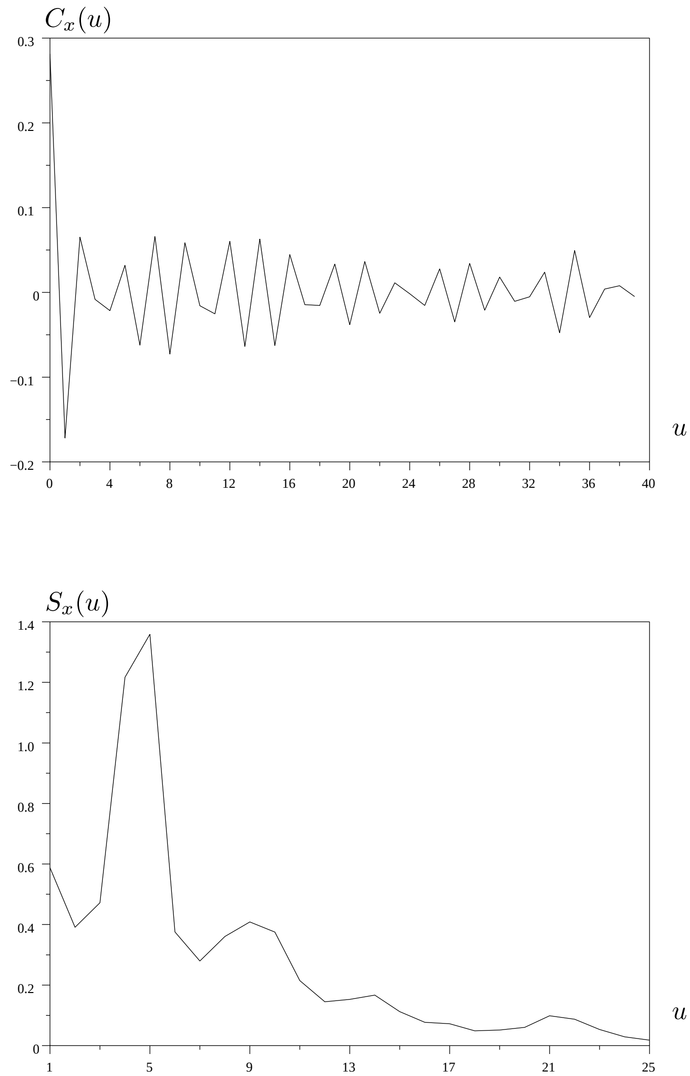

Figure 21: correlation function and power spectrum for the map $1 - 1.7x^2$.

4.2 Lyapunov exponents

There are three main difficulties in the measurement of Lyapunov exponents. Recall that from Oseledec's Theorem 2.3, we have to compute $\left(M_n^t M_n\right)^{1/2n}$ where

$$M_n = A_n \dots A_1 \ .$$

We also remind the reader that in general the order matters in this product.

i) For n large, one may wonder what is the most efficient way to compute M_n given that this matrix will in general grow exponentially fast with n (if

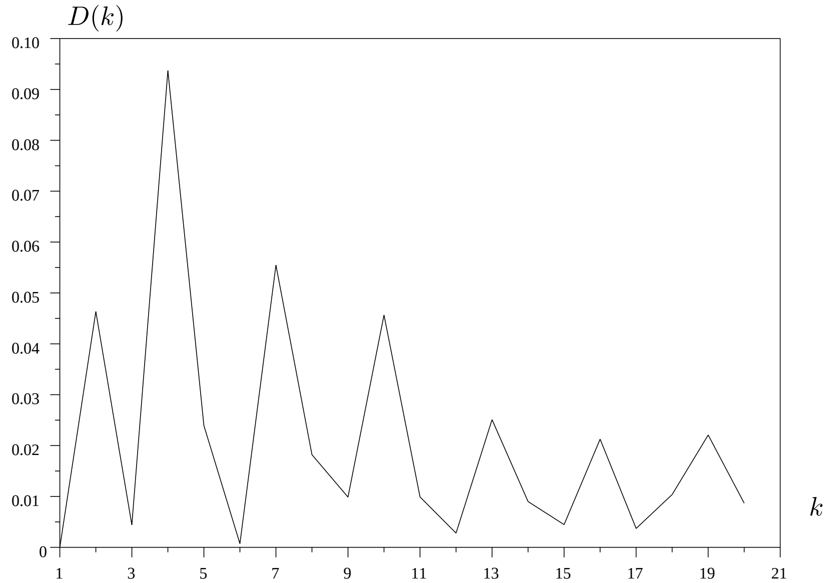

Figure 22: Information correlation for the map $1 - 1.8x^2$.

there is a positive Lyapunov exponent).

ii) If the phase space is known but not the transformation, how to estimate $A_n = DT_{x_{n-1}}$?

iii) Reconstruct the information in the phase space if it is unknown (for example in experimental situations).

We will deal with the three questions one after the other, but we first explain a simpler algorithm for the determination of the maximal exponent λ_0. The idea is that two typical nearby initial conditions x and y should separate exponentially fast with that rate (namely for typical x and y we should have $x - y \notin E_1(x)$, see 2.3). Recall also that in practice we want to use the (finite) orbit of a point x, namely

$$\mathcal{O}_n(x) = \big\{x, T(x), \ldots, T^{n-1}(x)\big\} \ .$$

The algorithm consists in computing for a small fixed $\epsilon > 0$ and several integers s the quantity

$$L_N(s, \epsilon, x) = \frac{1}{N-s} \sum_{j=0}^{N-s} \log \left(\frac{1}{\big|\mathcal{U}_\epsilon\big(T^j(x)\big)\big|} \sum_{j \in \mathcal{U}_\epsilon\left(T^j(x)\right)} \big\|T^{j+s}(x) - T^s(y)\big\| \right)$$

where $\mathcal{U}_\epsilon\big(T^j(x)\big)$ is the set of points y in the orbit $\mathcal{O}_n(x)$ at a distance less than ϵ of $T^j(x)$. The idea is that each term in the sum should be of order $\epsilon e^{s\lambda_0}$. Hence for s not too small, one expects to see a linear behaviour as a function of s with a slope giving a good approximation of λ_0. However if s is too large, $\epsilon e^{s\lambda_0}$ is larger than the size of the attractor, and we should observe a saturation. The summation in the formula serves to eliminate fluctuations, and it is advisable to compare the results for different values of ϵ.

Figure 23 shows the graph of $L_{2000}(s, \epsilon, x)$ as a function of s for the Hénon map (see example 1.8) and an initial condition chosen at random according to the Lebesgue measure.

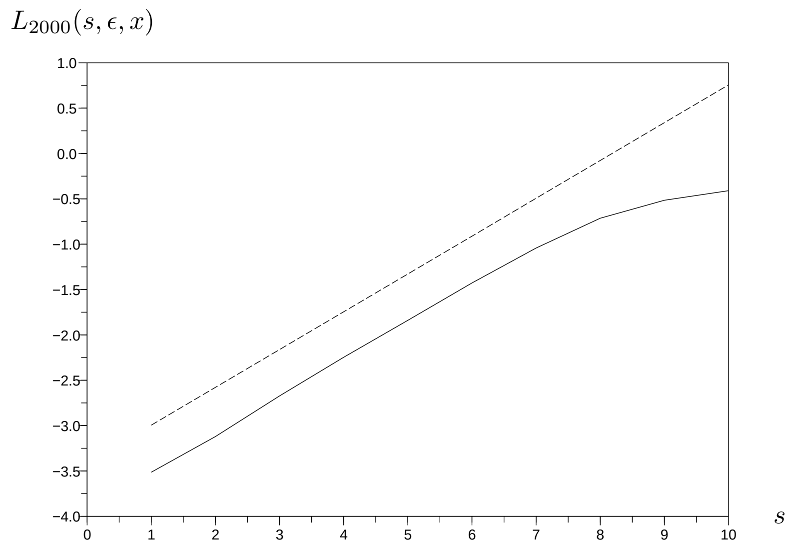

Figure 23: The function $s \mapsto L_{2000}(s, \epsilon, x)$ for the Hénon map (full line) and the theoretical slope (dash line).

We now discuss the general case where one would like to compute several (all) Lyapunov exponents. This is often numerically difficult, in particular if there are many exponents. Several methods have been proposed, we will briefly explain one of them based on the **QR** decomposition. Recall that any real matrix M can be written as a product $M = QR$ with Q a real orthogonal matrix and R a real upper triangular matrix (see [79]). This leads to a convenient recursive computation. Assume we have the **QR** decomposition of the matrix M_n, namely $M_n = Q_n R_n$ (Q_n orthogonal, R_n upper triangular). It follows from Theorem 2.3 that the logarithms of the diagonal elements of R_n (the eigenvalues) divides by n converge to the Lyapunov exponents. Since $M_{n+1} = A_{n+1} M_n$, we have $M_{n+1} = A_{n+1} Q_n R_n$. Perform now the **QR** decomposition of the matrix $A_{n+1} Q_n$, namely $A_{n+1} Q_n = Q_{n+1} W_n$ with Q_{n+1} orthogonal and W_n upper triangular. We then have $M_{n+1} = Q_{n+1} W_n R_n$ and since $W_n R_n$ we can take it as R_{n+1}, an obtain a **QR** decomposition of M_{n+1}. In practice it is therefore enough to keep at each step the matrix Q_n and to accumulate the logarithm of the absolute values of the diagonal elements (eigenvalues) of the matrices W_n. Note that since one accumulates logarithms instead of multiplying numbers, overflows (or underflows) are less likely. Since the matrix Q_n is orthogonal it has no risk of overflow.

This method is illustrated in figure 24 for the Hénon map (see example 1.8) with the parameters $a = 1.4$, $b = .3$. The figure shows how the result evolves when varying the number of iterations. The top curve corresponds to the positive exponent, the bottom curve to the negative one. In the case of the Hénon map, the determinant of the differential of the map is constant and equal to $-b$. Therefore, the sum of the exponents should be equal to $\log|b|$. This sum minus $\log|b|$ is the dash curve in figure 24.

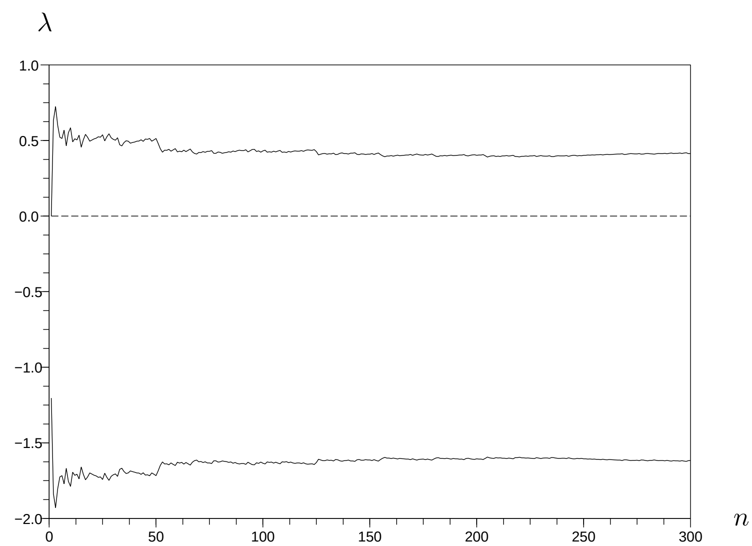

Figure 24: The two Lyapunov exponents for the Hénon map (solid lines) and their sum minus $\log|b|$, as a function of the number of iterations used in the algorithm.

Remark 4.1. Some simple facts help sometime in the computation or in checking the numerical accuracy.

i) If the map is invertible, the inverse map has the opposite exponents (for the same measure).

ii) If the volume element of the phase space is preserved by the dynamics (Jacobian of the map is of modulus one, or the vector field is divergence free), the sum of the exponents is zero. This generalises immediately to the case of constant Jacobian as in the Hénon map.

iii) For symplectic maps, if λ is a Lyapunov exponent, $-\lambda$ is also a Lyapunov exponent.

iv) For continuous time dynamical systems one can make a discrete time sampling and use the above technique. There is also a direct **QR** method for continuous time.

It is often the case that the differential of the map DT_x is not known explicitly. In that case one can try to use Taylor's formula. There are however two difficulties. One already mentioned is that we assumed we only have available one finite orbit $\mathcal{O}_n$, the other one is that in practice the data are often corrupted by some noise (numerical or experimental). In other words, instead of having a sequence $x_1, \ldots, x_n$ with $T\big(x_j\big) = x_{j+1}$, one observes (records) $y_1, \ldots, y_n$ with

$$y_j = x_j + \epsilon_j \ . \tag{64}$$

It is often assumed that the ϵ_j form a sequence of independent Gaussian random variables. This is a convenient assumption but not always verified.

The idea of one of the algorithm to estimate the differential is as follows. Let $\mathcal{O}_n$ be an orbit of length n. Let x be a point where we want to estimate DT_x (x need not belong to $\mathcal{O}_n$). Let $\mathscr{U}_x$ be a small neighborhood of x. Assume we have a point x_l of $\mathcal{O}_n$ in $\mathscr{U}_x$, this means that x_l is near x. By Taylor's formula we have

$$x_{l+1} = T\big(x_l\big) = T(x) + DT_x\big(x_l - x\big) + \mathscr{O}\big(\|x_l - x\|^2\big) \ .$$

Neglecting the quadratic term, this can be written

$$x_{l+1} = c + A\big(x_l - x\big) \tag{65}$$

where $c = T(x)$ is a vector in the phase space $\mathbf{R}^d$ and $A = DT_x$ is a $d \times d$ matrix with real entries. If we know c (if x is in the orbit $\mathcal{O}_n$), we only need to recover A and therefore we need d^2 independent equations for the entries. These equations come from relations (65) if we have at least d points of the orbit $\mathcal{O}_n$ inside the neighborhood $\mathscr{U}_x$, since each relation (65) leads to d linear equations (we may need more points if some of these equations are not independent). If c is unknown (if x is not in the orbit $\mathcal{O}_n$), we can at the same time recover $c = T(x)$ and $A = DT_x$ but we need at least $d+1$ points of the orbit inside the neighborhood $\mathscr{U}_x$. In the presence of noise, we have using equation (64)

$$\begin{aligned} y_{l+1} = \epsilon_{l+1} + T\big(x_l\big) = \epsilon_{l+1} + T\big(y_l - \epsilon_l\big) \\ = \epsilon_{l+1} + T(x) + DT_x\big(y_l - x\big) - DT_x\epsilon_l + \mathscr{O}\big(\|y_l - x\|^2\big) + \mathscr{O}\big(\|\epsilon_l\|^2\big). \end{aligned}$$

Neglecting the quadratic terms, we have with the previous notations

$$y_{l+1} = c + A\big(y_l - x\big) + \epsilon_{l+1} - DT_x\epsilon_l \ .$$

We are now in the classical problem of estimating an affine transformation in the presence of noise. Note that from our assumptions, ϵ_{l+1} and $DT_x\epsilon_l$ are independent and Gaussian with zero average. To solve this estimation problem, one can for example use a least square algorithm. In the case where $T(x)$ is known this leads to

$$A = \mathrm{argmin}_B \sum_{y_j \in \mathscr{U}_x} \left\| y_{j+1} - T(x) - B\big(y_j - x\big)\right\|^2 \ .$$

If $T(x)$ is not known, one can use the algorithm

$$(c, A) = \text{argmin}_{(b,B)} \sum_{y_j \in \mathscr{U}_x} \left\| y_{j+1} - b - B\big(y_j - x\big) \right\|^2 \ ,$$

to estimate $T(x)$ and DT_x at the same time.

Remark 4.2.

i) This and related methods have been proposed to perform noise reduction since we recover the deterministic quantity $T(x)$ out of the noisy data $y_1, \ldots, y_n$.

ii) This algorithm performs a prediction of the future orbit of the initial condition x. Having predicted $T(x)$ one can repeat the algorithm at this new point. We refer to [89] for more on this method and references.

Remark 4.3. The choice of the size of the neighborhood $\mathscr{U}_x$ is important and relates to two competing constraints.

i) The neighborhood $\mathscr{U}_x$ should be small enough so that quadratic terms are indeed negligible. Some authors have proposed to take the quadratic corrections into account but the numerics becomes of course much heavier. If $\mathscr{U}_x$ is too large the non linearities are not negligible and fake exponents may appear. These spurious exponents are often multiples of other ones. If one observes this phenomenon of having an exponent multiple of another one, it is advisable to diminish the size of the neighborhood and see if this relation persists.

ii) If the neighborhood $\mathscr{U}_x$ is too small, the statistics in the least square may become poor, there may be even less than the d points needed to solve the problem in the deterministic case.

iii) In practice one should use neighborhoods of various sizes and compare the results.

iv) We refer to [130] for the theory and to [62] [68], [89], [149] and [150] for more details on the numerical implementation and discussion of the results.

4.3 Reconstruction.

Contrary to the case of numerical simulations, when one deals with experimental data, the phase space is not known. For partial differential equations, for example, the phase space is often infinite dimensional but one has a finite dimensional attractor. As explained above, the system is observed through the time evolution of a real (or vector valued) observable, namely one has a recording of a real (or vector valued) sequence (60). The initial condition x is of course unavailable and the observation may be corrupted by noise. What can be done in this apparently adverse context? A reconstruction method has been developed by F.Takens which is based on the shift. One first fixes an integer d for the dimension of the space where the attractor and the dynamics will be reconstructed (we will discuss the best choice later on). From the data (60) one constructs a sequence of d dimensional vectors $Z_0, Z_1, \ldots, Z_{N-d}$ by

$$Z_0 = \begin{pmatrix} X_{d-1} \\ \vdots \\ X_0 \end{pmatrix}, \; Z_1 = \begin{pmatrix} X_d \\ \vdots \\ X_1 \end{pmatrix}, \cdots, \; Z_{N-d} = \begin{pmatrix} X_N \\ \vdots \\ X_{N-d+1} \end{pmatrix} . \tag{66}$$

In other words, we define a map Φ from the phase space Ω to $\mathbf{R}^d$ by

$$\Phi(x) = \begin{pmatrix} g\big(T^{d-1}(x)\big) \\ \vdots \\ g(x) \end{pmatrix} . \tag{67}$$

If the observable g is regular, Φ is regular but in general not invertible. This map (semi) conjugates the time evolution T and the shift $\mathcal{S}$, namely $\mathcal{S} \circ \Phi = \Phi \circ T$. Indeed

$$Z_{n-1} = \begin{pmatrix} X_{n+d-2} \\ \vdots \\ X_{n-1} \end{pmatrix} = \begin{pmatrix} g\big(T^{n+d-2}(x)\big) \\ \vdots \\ g\big(T^{n-1}(x)\big) \end{pmatrix}$$

and therefore

$$\mathcal{S}Z_{n-1} = \mathcal{S}\Phi\big(T^{n-1}(x)\big) = \begin{pmatrix} X_{n+d-1} \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} g\big(T^{n+d-1}(x)\big) \\ \vdots \\ g\big(T^n(x)\big) \end{pmatrix}$$

$$= \begin{pmatrix} g\big(T^{n+d-2}\big(T(x)\big)\big) \\ \vdots \\ g\big(T^{n-1}\big(T(x)\big)\big) \end{pmatrix} = \Phi\big(T^{n-1}\big(T(x)\big)\big) .$$

Remark 4.4. One cannot expect to get information outside the attractor. Indeed, after a transient time (which may be quite short), the orbit of the initial condition x is very near to the attractor.

There are well known results about the reconstruction of manifolds. A simplified formulation is as follows. A natural question is: given a manifold H of dimension d, in a space of dimension n (n could be infinite), what is the minimal dimension m such that one can reconstruct accurately H in $\mathbb{R}^m$? More precisely, find a map Φ from H to $\mathbb{R}^m$ which is regular, injective and with the inverse map (defined only on the image) regular (and injective). This is called an embedding.

An important theorem of Whitney says that this is always possible if $m \geq 2d + 1$, note that this number does not depend on n (see for example [77]).

Another important (and more difficult) theorem of Nash says that a compact (bounded) Riemannian manifold (i.e. a hyper-surface with a metric) can be embedded in $\mathbb{R}^m$ with its metric if $m \geq d(3d + 11)/2$.

Figure 25 is an example of a curve ($d = 1$) in $\mathbb{R}^3$ ($n = 3$). Many projections have a double point but some do not. For those we have an embedding in $\mathbb{R}^2$.

There is a result of Mané analogous to Whitney's theorem for the projections (which extends to fractal sets). If K is a set of dimension D (not necessarily

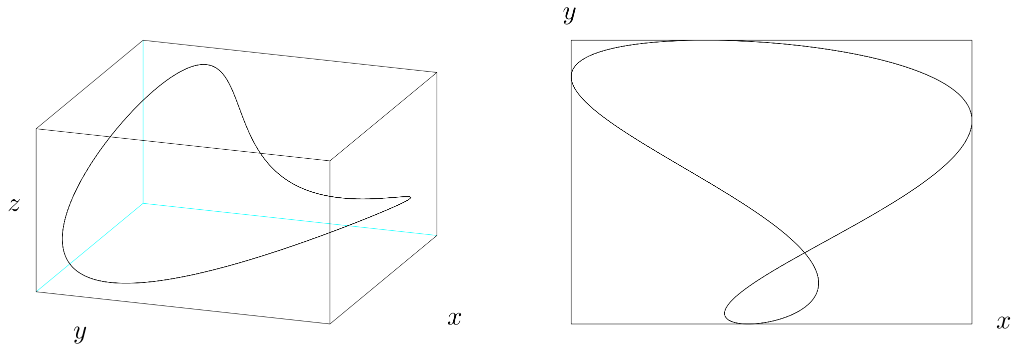

Figure 25: A projection with a double point.

integer), almost all projections over subspaces of (integer) dimension $m > 2D+1$ are injective (they separate the points). The difference with Whitney's theorem is that here we use projections, Whitney's theorem says there is an embedding. In the preceding example of a curve one gets $m = 4$

Unfortunately, Whitney's theorem, Nash theorem, Mané's theorem for Cantor sets etc. are not very useful in the present context because we require a mapping of the special form (67). Such special mappings have sometimes some unpleasant behaviours that should be avoided. For example if a and b are two fixed points in Ω (namely $T(a) = a$ and $T(b) = b$) such that $g(a) = g(b)$, then it follows immediately from the definition (67) that $\Phi(a) = \Phi(b)$, hence Φ is not injective. Similar anomalies occur for periodic orbits and this lead F.Takens to consider the generic case for g **and** T.

Remark 4.5. A similar reconstruction approach can be used in the case of continuous time, and similar difficulties occur. Assume that the flow φ_t has a periodic orbit C (cycle) which is stable and of (finite) period $\tau > 0$. Assume also that C is a C^0 sub-manifold. Define a discrete time evolution T by $T = \varphi_\tau$. Then every point of C is a fixed point of T. Since C is a closed curve and g is continuous, $g(C)$ is a segment and $\Phi(C)$ is a segment of the diagonal in $\mathbf{R}^d$. Hence there are certainly at least two points of C which are identified through Φ even if g is not constant on C.

We now state Takens reconstruction theorem.

Theorem 4.1. *For a generic transformation T with an attractor $\mathscr{A}$ of (Hausdorff) dimension D, and for a generic real valued observable g, the map (67) is an embedding of $\mathscr{A}$ in $\mathbf{R}^d$ if $d > 2D$.*

Remark 4.6.

i) We will see later on that attractors may have non integer dimension.

ii) In the original work generic was in the sense of Baire's second category (countable intersection of open dense sets).

iii) The same result has been proven for other genericity conditions for example of probabilistic nature (this has to be defined with some care because the spaces of maps and observables are in general infinite dimensional, see [134]).

iv) The genricity conditions avoid the problems we have mentioned above for the maps of the form (67). One can list a number of conditions so that Theorem 4.1 holds, they are however hard to check in practice. Theorem 4.1 states that these conditions are generic.

v) Recall that symmetry properties for example of the observable g may break genericity.

In practice, one does not know a-priori the dimension of the attractor, and therefore the dimension d of the reconstruction space. One tries several values for d and apply various criteria to select an adequate one. One of them is the false neighbors criteria. If d is too small, then by accident two far away points a and b of the phase space may be mapped to nearby points by Φ (recall the example of the projection of a three dimensional curve on a plane in figure 25, very often the projected curve self intersect even if the original curve does not). However this accidental effect should not persist for the iterates by the map T of the points a and b. At least some of the first few iterates should be reasonably far away one from the other. In practice, one selects a number $\delta > 0$ much smaller than the size of the image of the attractor, and considers all the pairs of points at a distance less than δ. For each such pair, one looks at the image pair, and if for one pair one gets a large distance, then the reconstruction dimension is suspected to be too small.

Remark 4.7. Note that by continuity of Φ and T, if we consider two nearby initial conditions, their first few iterates should be close. Therefore if we have a pair of points at distance less than δ, the images should not be far away. If they are we have the above problem of a too small embedding dimension.

In principle there is no problem in using a large reconstruction dimension. However the statistics of the different measurements will be poorer. We will see later on when discussing the measurement of dimension another criteria to select a reconstruction dimension. It is wise to use as many criteria as available (see also the Kaplan-Yorke formula later).

We refer to [142], [134], [95] and [139] for proofs and references and to [69], citeks fore more on the implementation of these techniques.

For partial differential equations, there is a reconstruction involving the (physical) space. Consider a non linear PDE in a bounded space domain $\mathcal{D}$. For example

$$\partial_t A = \Delta A + A - A\big(|A|^2 - 1\big)$$

with A complex, and some boundary condition on $\partial\mathcal{D}$. The phase space Ω is infinite dimensional. For example the space of regular functions (with square summable Laplacian). There is a well defined semi-flow (but going backward in time is mostly explosive). One can prove that there is a finite dimensional attracting set (in the infinite dimensional phase space). The dimension of this attracting set is proportional to the size of the domain $\mathcal{D}$ (provided there is some uniformity on the boundary condition). Moreover there is an inertial manifold.

This is an invariant manifold (hyper-surface) of finite (integer) dimension which attracts transversally all the orbits. In particular it contains the attracting set (see [147]). Points of the attractor can be distinguished by a (fine enough) discrete set in $\mathcal{D}$ (see [82]). In other words, there is a number $\eta > 0$, such that if a set of points M in $\mathcal{D}$ does not leave any hole of size η, then if two functions A_1 and A_2 of the attractor are equal at each point of M, they are equal everywhere in $\mathcal{D}$. This result is known under certain conditions on the equation and for regular domains. It is still an open problem for the Navier Stokes equation, even in dimension two. The number η depends on the equation, and in particular of the size of the coefficients. The cardinality of M is proportional to the volume.

4.4 Measuring the Lyapunov exponents.

One applies first the technique explained above to reconstruct the system. Note that because of the special form of the reconstruction, the tangent (reconstructed) map has a special form

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & a_d \end{pmatrix}$$

Remark 4.8.

i) We have seen that fake exponents may appear as multiple of true exponents if the nonlinearities are not negligible.

ii) In the reconstruction, fake exponents can also appear due to the numerical noise in the transverse directions of the reconstructed attractor.

iii) The reconstruction dimension is often larger than the dimension of the attractor, and the negative Lyapunov exponents are polluted by noise (also fake zero exponents may appear).

iv) These fake exponents are in general unstable by time inversion (if the map is invertible one can look for the exponents of the inverse). Also it is wise to explore changes in the various parameters of the algorithms.

v) It is in general difficult to determine several exponents, especially the negative ones.

vi) For the positive exponents, one can try to reduce the dimension to the number of positive exponents. For this purpose, one can use only some components of the reconstruction. For example if one uses a reconstruction dimension $d = pq$ with p and q integers, one can consider the subspaces $\big(g(x), g(T^{q-1}(x)), \cdots, g(T^{pq-1}(x))\big)$, when trying to measure the p largest exponents.

We refer to [2], [89] and [135] for more results and references.

4.5 Dimensions.

Because of the sensitive dependence on initial condition and the compactness of the attractors, these sets have often a complicated (fractal) structure. One way to capture (measure) this property is to determine their dimension. There are many definitions for the dimension of a set. Two are most often used in dynamical system theory: the Hausdorff dimension and the box counting dimension (also called capacity). They are not equivalent, the Hausdorff dimension is more satisfactory from a theoretical point of view, the box counting dimension (or box dimension) is easier to compute numerically. Recall (see [87], [65] or [115]) that the Hausdorff dimension of a set A is defined as follows. Let $B_r(x)$ denote the ball of radius r centered in x. Let $B_{r_j}\big(x_j\big)$ be a sequence of balls covering A, namely

$$A \subset \bigcup_j B_{r_j}\big(x_j\big)\ .$$

For a numbers $d > 0$ and $\epsilon > 0$, define

$$\mathcal{H}_d(\epsilon, A) = \inf_{A\subset\cup_j B_{r_j}(x_j)\,,\,\sup_j r_j\le\epsilon} \ \sum_j r_j^d\ . \tag{68}$$

This is obviously a non increasing function of ϵ which may diverge when ϵ tends to zero. Moreover, if the limit when $\epsilon \searrow 0$ is finite for some d, it is equal to zero for any larger d. This limit is also non increasing in d. Moreover, if it is finite and non zero for some d, it is infinite for any smaller d. The Hausdorff dimension of A, denoted below by $d_{\mathrm{H}}(A)$ is defined as the infimum of the positive numbers d such that the limit vanishes (for this special d the limit may be zero, infinite, finite or does not exist), namely

$$d_{\mathrm{H}}(A) = \inf\left\{d\,\Big|\,\lim_{\epsilon\to 0}\mathcal{H}_d(\epsilon, A) = 0\right\}$$

This is also the supremum of the set of d positive numbers such that the limit is infinite.

The box counting dimension of a subset A of $\mathbf{R}^n$ is defined as follows. For a positive number r, let $N_A(r)$ be the smallest number of balls of radius r needed to cover A. If for (all) small r we have $N_A(r) \approx r^{-d}$, we say that the box dimension of A is d, and we denote it by $d_{\mathrm{Box}}(A)$.

Exercise 4.1. Show that for any (bounded) set A, $d_{\mathrm{H}}(A) \le d_{\mathrm{Box}}(A)$ (there are sets for which the inequality is strict).

Exercise 4.2. Show that for the triadic Cantor set $d_{\mathrm{H}}(A) = d_{\mathrm{Box}}(A) = \log 2/\log 3$.

We refer to [87], [65] and [115] for other definition of dimensions and the main properties.

A difficulty is that in general one does not have access to the whole attractor. As explained at the beginning of this chapter, it is often the case that one knows only one trajectory which is not necessarily uniformly spread over all the attractor, even though it may be dense. One can however assume that this trajectory is typical for a certain measure μ (for example an SRB measure), and this leads to the definition of the dimension of a probability measure.

Definition 4.1. The dimension of a probability measure is the infimum of the dimensions of the (measurable) sets of measure one.

Remark 4.9. In practice one should of course say which dimension is used (Hausdorff, box counting etc.) Note also that the infimum may not be attained. If a measure is supported by an attractor, its dimension is of course at most the dimension of the attractor (the dimension of a measure is at most the dimension of its support). This inequality may be strict.

The dimension of a measure is somehow related with the behaviour of the measure of small balls as a function of the diameter. We start by a somewhat informal argument. Let μ be a measure, and let $d_{\text{Box}}(\mu)$ be its box counting dimension. This roughly means that we need about $N_A(r) \approx r^{-d_{\text{Box}}(\mu)}$ balls of radius r to cover (optimally) the "smallest" set A of full measure. Let $B_1,\ B_2, \ldots,\ B_{N_A(r)}$ be such balls. In order to have as few balls as possible, they should be almost disjoint, and to simplify the argument we will indeed assume that they are disjoint. Since the set A is of full measure we have

$$1 = \mu(A) = \mu\left(\cup_{j=1}^{N_A(r)} B_j\right) = \sum_{j=1}^{N_A(r)} \mu\big(B_j\big)\ .$$

In the simplest homogeneous case, all the balls have the same measure, and we get for $1 \le j \le N_A(r)$

$$\mu\big(B_j\big) = \frac{1}{N_A(r)} \approx r^{d_{\text{Box}}(\mu)}\ .$$

A rigorous argument leads to a lower bound for the Hausdorff dimension.

Lemma 4.2. *Assume there are two constants $C > 0$ and $\delta > 0$ such that a probability measure μ on $\mathbf{R}^n$ satisfies for any $x \in \mathbf{R}^n$ and any $r > 0$ the inequality*

$$\mu\left(B_r(x)\right) \le C r^{\delta}\ .$$

Then

$$d_{\text{H}}(\mu) \ge \delta\ .$$

Proof. Let A be a set of measure one, and let $B_{r_j}\big(x_j\big)$ be any sequence of balls covering A. Since

$$A \subset \bigcup_j B_{r_j}\left(x_j\right)\ .$$

we have (since the balls are not necessarily disjoint)

$$1 = \mu(A) \le \sum_j \mu\big(B_{r_j}\left(x_j\right)\big)\ .$$

Therefore from the assumption of the theorem, we have

$$1 \le C \sum_j r_j^{\delta}\ .$$

In other words, for any $\epsilon > 0$ and any set A of full measure we have

$$\mathcal{H}_\delta(\epsilon, A) \ge \frac{1}{C}\ .$$

The theorem follows immediately from the definition of the Hausdorff dimension. □

We refer to [87] for more on this so called Frostman lemma, and in particular for the converse.

It is natural to ask if there are relations between the dimension of a measure and the other quantities like entropy and Lyapunov exponents. In dimension two, L.S.Young found an interesting relation.

Theorem 4.3. *For a regular invertible map of a compact manifold and an ergodic invariant probability measure μ with two exponents satisfying $\lambda_1 > 0 > \lambda_2$, we have*

$$d_{\mathrm{H}}(\mu) = h(\mu)\left(\frac{1}{\lambda_1} + \frac{1}{\big|\lambda_2\big|}\right) .$$

In the case of an SRB measure, we have under the hypothesis of the above Theorem and using Theorem 2.11, $h(\mu) = \lambda_1$, hence

$$d_{\mathrm{H}}(\mu) = 1 + \frac{\lambda_1}{\big|\lambda_2\big|} . \tag{69}$$

In the particular case of the Hénon map (see example 1.8), $\lambda_1 + \lambda_2 = \log b$ and hence for an SRB measure

$$d_{\mathrm{H}}(\mu) = 1 + \frac{\lambda_1}{\big|\log b - \lambda_1\big|} .$$

We refer to [155] for a proof of the Theorem, but we now give an intuitive argument for this result. From the Katok result 2.6, we have to count the maximal number of typical trajectories ϵ distinct on a time interval of length k. For that purpose, we will make the counting at an intermediate time $0 < k_1 < k$ (k_1 optimally chosen later). At time k_1, we cover a set of positive measure by balls of radius $\epsilon' \ll \epsilon$ (ϵ' will be optimally chosen below). From k_1 to k the balls of radius ϵ' are stretched. Therefore

$$\epsilon = \epsilon' e^{k_2\lambda_1}$$

where $k_2 = k - k_1$ (one looks at the worse case: separation takes place at the last time k).

From k_1 to 0, the balls of radius ϵ' are stretched by the inverse map, hence

$$\epsilon = \epsilon' e^{-k_1\lambda_2} .$$

Therefore

$$\epsilon = \epsilon' e^{-k_1\lambda_2} = \epsilon' e^{k_2\lambda_1}$$

which implies $k_1 = k\lambda_1/\big(\lambda_1 - \lambda_2\big)$, from which we derive ϵ'. However the number of trajectories we are looking for is the minimal number of balls of radius ϵ' necessary to make a set of (almost) full measure. Since the balls have measure ${\epsilon'}^{d_{\mathrm{Box}}(\mu)}$, we need ${\epsilon'}^{-d_{\mathrm{Box}}(\mu)}$ of them. The maximum number of pairwise ϵ separated trajectories is therefore

$${\epsilon'}^{-d_{\mathrm{Box}}(\mu} = \epsilon^{-d} e^{d_{\mathrm{Box}}(\mu) k_2\lambda_1} \approx e^{kh(\mu)} ,$$

and therefore

$$h(\mu) = \frac{d_{\mathrm{Box}}(\mu)\ k_2\ \lambda_1}{k} = \frac{d_{\mathrm{Box}}(\mu)\ \lambda_1\ \big|\lambda_2\big|}{\lambda_1 + \big|\lambda_2\big|} .$$

Another relation between dimension and Lyapunov exponents is the so called Kaplan-Yorke formula. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ be the decreasing sequence of Lyapunov exponents **with multiplicity** (in other words the number of terms is equal to the dimension). Let

$$k = \max\left\{ i \,\middle|\, \lambda_1 + \ldots + \lambda_i \geq 0 \right\} .$$

Kaplan and Yorke defined the Lyapunov dimension by

$$d_{\mathrm{L}}(\mu) = k + \frac{\left|\lambda_{k+1}\right|}{\lambda_1 + \ldots + \lambda_k} .$$

They conjectured that $d_{\mathrm{L}}(\mu) = d_{\mathrm{H}}(\mu)$ (the Kaplan Yorke formula) but counter examples were discovered later on. Note however that in the hypothesis of Theorem 4.3 the equality holds. In the general case, Ledrappier proved the following inequality.

Theorem 4.4. *For any ergodic invariant measure μ we have $d_{\mathrm{H}}(\mu) \leq d_{\mathrm{L}}(\mu)$.*

We now give a rough idea for the "proof" of the Kaplan-Yorke formula. Take a ball B of radius ϵ (small) in the attractor. Its measure is about $\approx \epsilon^{d_{\mathrm{Box}}}(\mu)$. If one iterates this ball k times, (ϵ small, k not too large), one obtains an ellipsoid $\mathcal{E} = T^k(B)$, elongated in the unstable directions, contracted in the stable directions. We now cover this ellipsoid $\mathcal{E}$ by balls of radius

$$\epsilon' = e^{k\lambda_j}\epsilon$$

with j chosen (optimally) later on but such that $\lambda_j < 0$. A ball of radius ϵ' has measure ${\epsilon'}^{d_{\mathrm{H}}(\mu})$, and since the measure μ is invariant, we get

$$\epsilon^{d_{\mathrm{H}}(\mu)} = \mu\big(B\big) = \mu\big(\mathcal{E}\big) = N\, {\epsilon'}^{d_{\mathrm{H}}(\mu)}$$

where N is the number of balls of radius ϵ' necessary to cover the ellipsoid $\mathcal{E} = T^k(B)$. We now evaluate N. In the direction of the exponent λ_1, the original ball is stretched ($\lambda_1 > 0$), and its size becomes $\epsilon\exp(\lambda_1 k)$. We need therefore

$$\frac{\epsilon e^{\lambda_1 k}}{\epsilon'}$$

balls to cover $\mathcal{E}$ in that direction. The same argument holds in any direction with $l < j$ (with multiplicities). In the direction of λ_j one needs only one ball, which also covers in all directions ($l > j$) since they are contracted. Summarizing, since

$$N = \prod_{l=1}^{j} \frac{\epsilon e^{\lambda_l k}}{\epsilon'}$$

and using $\epsilon^{d_{\mathrm{H}}(\mu)} = N{\epsilon'}^{d_{\mathrm{H}}(\mu)} = N\epsilon^{d_{\mathrm{H}}(\mu)}e^{k\lambda_j d_{\mathrm{H}}(\mu)}$, we obtain

$$\epsilon^{d_{\mathrm{H}}(\mu)-j} = {\epsilon'}^{d_{\mathrm{H}}(\mu)-j}e^{kS_{j-1}} = \epsilon^{d_{\mathrm{H}}(\mu)-j}e^{k\lambda_j(d_{\mathrm{H}}(\mu)-j)}e^{kS_{j-1}}$$

with $S_p = \lambda_1 + \cdots + \lambda_p$. This must be true for any integer k, hence

$$S_{j-1} + \lambda_j(d_{\mathrm{H}}(\mu) - j) = 0 .$$

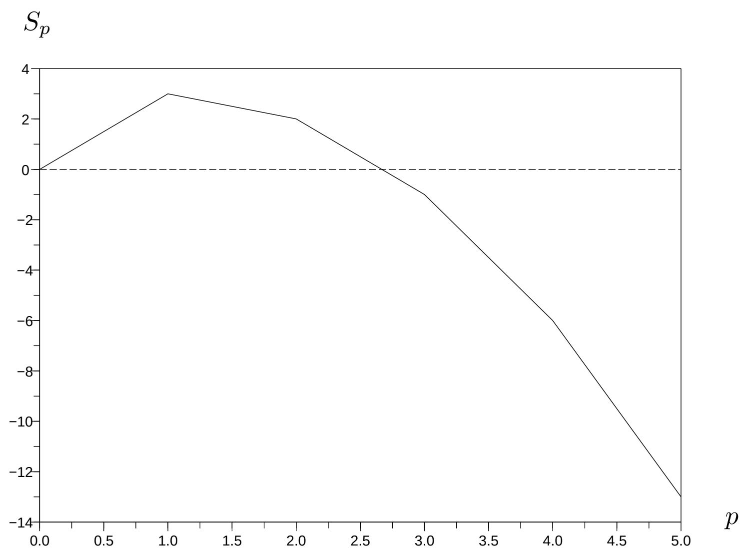

Figure 26: Piecewise linear interpolation of the numbers S_p as a function of p.

It seems that we have only one equation for two unknowns j and $d_{\mathrm{H}}(\mu)$. Nevertheless there is only one solution! To see this it is convenient to plot the piecewise linear interpolation of S_p as a function of p. Between $p = k$ and $p = k + 1$ the linear interpolation is $S_k + \lambda_{k+1}(x - k)$. d_{L} is the unique zero of this function, see figure 26 We now come back to the question of measuring the dimension of an ergodic invariant measure μ. The idea is the following. Assume that for most points x in the support of the measure

$$\mu\big(B_r(x)\big) \approx r^{d_{\mathrm{H}}(\mu)} \ ,$$

then obviously

$$d_{\mathrm{H}}(\mu) \approx \frac{\log \mu\big(B_r(x)\big)}{\log r} \ .$$

Remark 4.10. There are cases where the above assumption fails, see [107].

In order to increase the statistics, a better formula would be something like

$$d_{\mathrm{H}}(\mu) = \lim_{r \searrow 0} \int \frac{\log \mu\big(B_r(x)\big)}{\log r} d\mu(x) \ .$$

Numerical simulations performed on several systems suggested that this quantity converges slowly to its limit, moreover it is not easy to compute from a numerical time series (one finite orbit). For these reasons people have adopted another definition. Consider the function of $r > 0$

$$C_2(r) = \int \mu\big(B_r(x)\big) \, d\mu(x) \ . \tag{70}$$

The first guess is that this quantity should again behave roughly as $r^{d_{\mathrm{H}}(\mu)}$, the integrand does. It turns out that it behaves like $r^{d_2(\mu)}$ with $d_2(\mu) \leq d_{\mathrm{H}}(\mu)$. This is because of Jensen's inequality (the logarithm of the integral is in general larger than the integral of the logarithm). The number $d_2(\mu)$ is often called the correlation dimension.

Exercise 4.3. Consider the map $2x \pmod 1$ and the ergodic invariant measure μ_p of exercise 1.18. Show that for $p \neq 1/2$, $d_2(\mu) \neq d_{\mathrm{H}}$.

The quantity $C_2(r)$ is statistically rather stable and easy to compute from a numerical time series as we will see now.

Assume that we have a trajectory in the phase space $x_1, x_2, \ldots$ (an orbit). We will apply twice the ergodic theorem to obtain a (asymptotic) formula for the function $C_2(r)$.

By the ergodic theorem 2.1 we have almost surely

$$\int \mu\big(B_r(x)\big)\, d\mu(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mu\big(B_r(x_j)\big) \ .$$

We can now use again the ergodic theorem to determine $\mu\big(B_r(x_j)$. Let θ denote the Heaviside function on $\mathbf{R}$ (the function equal to one on the positive real numbers and equal to zero on the negative ones). We have

$$\mu\big(B_r(x_j)\big) = \lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} \theta\big(r - d(x_k, x_j)\big)$$

where d is the distance on the phase space. We can now put these two formulas together and obtain the following algorithm

$$C_2(r) = \lim_{N\to\infty} \frac{1}{N^2} \sum_{p=1,\, q=1}^{p=N,\, q=N} \theta\big(r - d(x_p, x_q)\big) \ . \tag{71}$$

We mention a rigorous result in the context of the above formal manipulation.

Theorem 4.5. *Formula (71) holds almost surely point where $C_2(r)$ is continuous.*

We refer to [112] and [136] for a proof.

We now discuss some of the the practical issues in using formula (71) to measure the dimension. To give a first idea, we show in figure 27 the function $C_2(r)$ for the Hénon map (see example 1.8).

Remark 4.11. There are several practical difficulties in measuring $d_2(\mu)$.

i) Strong local correlations may occur. For example if one samples a flow $x_j = \phi_{j\tau}\big(x_0\big)$. If τ is too small, by continuity there is a strong correlation between x_j and x_{j+1}. To avoid these problems, the summation is extended only on those p and q with $|p-q| > K$, K large enough, of order $1/s$ where s is an estimate of the decorrelation time. This time scale can be obtained by looking at the signal (for example the change of side in the Lorenz system), by looking at the decay of correlation function or by other means.

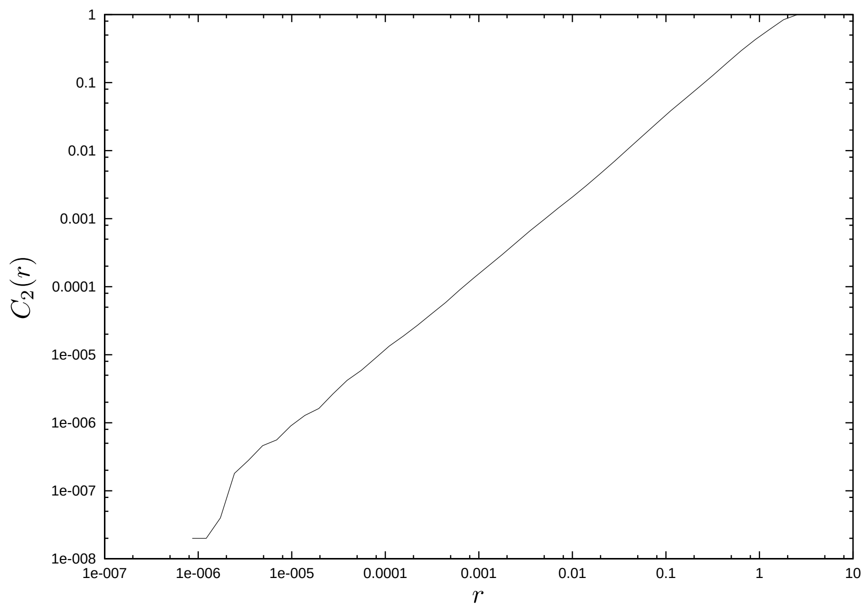

Figure 27: The function $C_2(r)$ for the Hénon map.

ii) Another difficulty is the presence of noise. A white noise corresponds to an attractor of infinite dimension. If the noise has amplitude ϵ, its influence will only be noticeable if r is of the order ϵ or smaller. Above this value, one expects that the noise averages out. One can hope for a power law behaviour of $C_2(r)$ only for a range of values of r which must be

a) small enough for the (dynamical) fluctuations of $\log \mu\big(B_r(x)\big)/\log r$ to be small,

b) large enough to be above the noise.

iii) One also uses other kernels than the Heaviside functions, like for example a Gaussian.

For some "good" systems there is a theoretical bound on the fluctuations of C_2. In the general case one expects worse results.

For some systems where exponential or De Vroye estimates are available, the standard deviation of $C_2(r, N)$ can be bounded by $N^{-1/2}r^{-1}$. If one wants to see the power law behaviour above the dynamical fluctuations, one needs

$$r^{d_2(\mu)} \gg \frac{1}{N^{1/2} r}$$

That is to say

$$r \gg \frac{1}{N^{1/(2d_2(\mu)+2)}} \qquad \text{or} \qquad N \gg \frac{1}{r^{2d_2(\mu)+2}}$$

If one has an a-priori idea on the value of $d_2(\mu)$, one can guess a value for N. For the Hénon map this gives $r \gg N^{-.21}$, which leads to $N \sim 60000$ for $r > .1$.

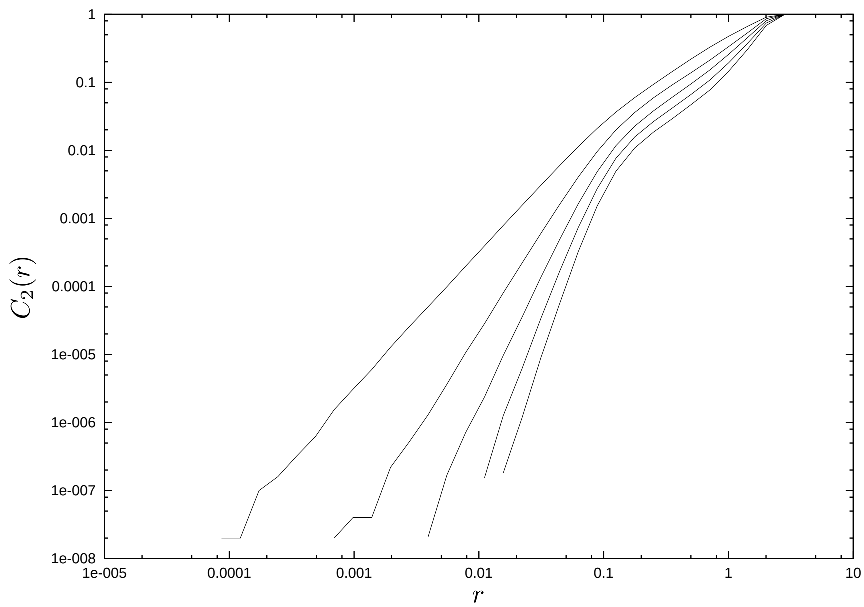

Figure 28: The function $C_2(r)$ for the x component of the Hénon map and several reconstruction dimensions with noise of standard deviation .05.

These bounds are often pessimistic (they are for the worse cases). But they may give an order of magnitude.

We refer to [18], [19], [21], [52], [93], [94], [140], [141], [144], [145], [117] for more results and references.

From a practical point of view, one looks in the graph of $C_2(r)$ for a range of intermediate values of r where this function looks like a power of r. Logarithmic scales are convenient in a first approach.

Except in numerical simulations, one works with reconstructed attractors. It is wise to look at the results for different values of the reconstruction dimension. If the observed dimension (slope of the log-log plot of $C_2(r)$) is equal to the reconstruction dimension, this means that the dimension of the attractor has not yet been reached. One should increase the reconstruction dimension. It is useful to plot the measured dimension as a function of the reconstruction dimension to look for a saturation when a reconstruction dimension is large enough to obtain an embedding. Also one should verify the coherence of the different estimates of the reconstruction dimension: false neighbors etc. The influence of the noise at small r is clearly visible in figure 28 which shows the function $C_2(r)$ for the x component of the Hénon map and several reconstruction dimensions with noise of standard deviation .05.

Estimating the slope of the function $C_2(r)$ is not an easy task, especially if the power law is only valid on a restricted range (limited by dynamical fluctuations at small $-\log r$ and by noise at large $-\log r$). Several statistical works have dealt with this problem using linear interpolation, maximum likelihood, derivative estimates etc. Here is one of the proposed formula (with no claim that it is universally better than others).

$$d_2(\mu) = \frac{-N_s}{\sum_{i=1}^{N_s} \log(r_i) + N_l \log r_l}$$

where the power law range is $[r_l, r_u]$, and it contains N_s points $r_1, \cdots, r_{N_s}$, while the range $[0, r_l]$ contains N_l points. Statistical studies of this estimator or of similar ones allow to construct confidence intervals.

4.6 Multifractal

Even if the functions $x \to \log \mu\big(B_r(x)\big)/\log r$ converge almost surely when r tends to zero, this quantity can show for non zero r interesting (and even measurable) fluctuations (recall that $B_r(x)$ is the ball centered in x of radius r). The study of these fluctuations is called the multifractal analysis. More precisely, for any $\alpha \geq 0$, let

$$\mathcal{E}_\alpha^+ = \left\{ x \,\middle|\, \limsup_{r\to 0} \frac{\log \mu\big(B_r(x)\big)}{\log r} = \alpha \right\} ,$$

and

$$\mathcal{E}_\alpha^- = \left\{ x \,\middle|\, \liminf_{r\to 0} \frac{\log \mu\big(B_r(x)\big)}{\log r} = \alpha \right\} .$$

One would like to characterise the "size" of these sets. This can be done for example using another measure (or family of measures), or using a dimension. The so called multifractal spectrum is usually defined a the function $\alpha \to d_{\mathrm{H}}\big(\mathcal{E}_\alpha^+\big)$ (or $\alpha \to d_{\mathrm{H}}\big(\mathcal{E}_\alpha^+\big)$).

Under suitable hypothesis, one can relate these quantities to some large deviation functions. We refer to [44], [61], [127], [152] [143], [148] for more results and references. We will explain here how this works on a simple example. Consider the dissipative Baker map (2). As we have seen in exercise 1.25, the attractor is the product of a segment by a Cantor set $\mathcal{K}$. This Cantor set can be constructed as follows. Consider the two contractions f_1 and f_2 of the interval given by

$$f_1(y) = \frac{y}{4} \quad \text{and} \quad f_2(y) = \frac{2+y}{3} .$$

The Cantor set is obtained by applying all the infinite products of f_1 and f_2 to the interval. Namely, one defines for each n a set $\big(\mathcal{K}_n\big)$ consisting of 2^n disjoint intervals, and constructed recursively as follows. First $\mathcal{K}_0 = [0, 1]$, and then for any $n \geq 1$,

$$\mathcal{K}_{n+1} = \big\{ f_1\big(I\big) \,\big|\, I \in \mathcal{K}_n \big\} \cup \big\{ f_2\big(I\big) \,\big|\, I \in \mathcal{K}_n \big\} . \tag{72}$$

The transverse measure ν of the SRB measure (see formula 28) satisfies $\nu(I) = 2^{-n}$ for any $I \in \mathcal{K}_n$.

Exercise 4.4. Show that this construction uniquely defines the measure ν.

Exercise 4.5. Consider the map f of the unit interval defined by

$$f(x) = \begin{cases} 4x & \text{for } 0 \leq x \leq 1/4 , \\ 12(x - 1/4)/5 & \text{for } 1/4 < x < 2/3 , \\ 3(x - 2/3) & \text{for } 2/3 \leq x \leq 1 . \end{cases}$$

Show that the Cantor set $\mathcal{K}$ is invariant by f. Show that f_1 and f_2 are two inverse branches of f. Show that ν is an ergodic invariant measure for this map (one can use a coding).

We will need a large deviation estimate for the sizes of the intervals in each $\mathcal{K}_n$. In the present situation, this is easy since the maps f_1 and f_2 have constant slope. Indeed, we have for any integer $0 \le p \le n$

$$\left|\left\{I \in \mathcal{K}_n \,\Big|\, |I| = 4^{-p}\, 3^{p-n}\right\}\right| = \begin{pmatrix} n \\ p \end{pmatrix} .$$

For later purposes it is convenient to use the Stirling approximation. Let φ be the function defined for $s \in [-\log 4, -\log 3]$ by

$$\varphi(s) = \frac{s + \log_2 3}{\log_2(4/3)} \log_2 \left(-\frac{s + \log_2 3}{\log_2(4/3)}\right) - \frac{s + \log_2 4}{\log_2(4/3)} \left(\frac{s + \log_2 4}{\log_2(4/3)}\right) ,$$

then for $s \in [-\log_2 4, -\log_2 3]$

$$\left|\left\{I \in \mathcal{K}_n \,\Big|\, |I| \approx 2^{ns}\right\}\right| \approx 2^{n\,\varphi(s)} . \tag{73}$$

Exercise 4.6. Consider the sequence of functions

$$Z_n(\beta) = \sum_{I \in \mathcal{K}_n} |I|^{\beta} .$$

Compute $Z_n(\beta)$, and derive that

$$F(\beta) = \lim_{n \to \infty} \frac{1}{n} \log_2 Z_n(\beta) = \log_2 \left(4^{-\beta} + 3^{-\beta}\right) .$$

Show that φ is the Legendre transform of F. Prove relation (73) using steepest descent. Show that $\varphi(s)$ is concave, and its maximum is attained at $s = -\log_2 \sqrt{12}$.

We can now state the result.

Theorem 4.6. *For any* $\alpha \in [1/\log_2 4, 1/\log_2 3]$ *we have*

$$d_{\mathrm{H}}\big(\mathcal{E}_\alpha^{-}\big) = d_{\mathrm{H}}\big(\mathcal{E}_\alpha^{+}\big) = \alpha\, \varphi(-1/\alpha) .$$

Proof (*sketch*). We will only give a proof for $\alpha \le \alpha_c = 1/\log_2 \sqrt{12}$. We refer the reader to the literature for the other range of α. Let

$$\tilde{\mathcal{E}_\alpha} = \left\{ x \,\Bigg|\, \alpha \text{ is an accumulation point of } \frac{\log \mu\big(B_r(x)\big)}{\log r} \text{ when } r \to 0 \right\} .$$

Note that $\mathcal{E}_\alpha^{\pm} \subset \tilde{\mathcal{E}_\alpha}$. To construct the upper bound for $d_{\mathrm{H}}\big(\tilde{\mathcal{E}_\alpha}\big)$ we will use a particular covering by balls. Indeed, let $x \in \tilde{\mathcal{E}_\alpha}$. Then by definition, there exists an infinite sequence $\big(r_j\big)$ tending to zero such that

$$\lim_{j \to \infty} \frac{\log \mu\big(B_{r_j}(x)\big)}{\log r_j} = \alpha .$$

Since $x \in \mathcal{K}$, for any j we can find an integer n_j such that there is an interval $I \in \mathcal{K}_{n_j}$ satisfying

$$\mathcal{K} \bigcap B_{r_j}(x) \subset I$$

and an interval $J \in \mathcal{K}_{n_j+2}$ such that

$$J \subset B_{r_j}(x) \ .$$

Exercise 4.7. Prove this statement, for example by drawing a local picture picture around x of $\mathcal{K}_n$.

In particular, we have

$$2^{-n_j-2} \le \mu\big(B_{r_j}(x)\big) \le 2^{-n_j} \ .$$

From the condition $x \in \tilde{\mathcal{E}_\alpha}$, we derive

$$2^{-n_j} \approx \mu\big(B_{r_j}(x)\big) \approx r_j^{\alpha} \ ,$$

and therefore

$$r_j \approx 2^{-n_j/\alpha} \ .$$

This implies

$$\big|I\big| \lessapprox 2^{-n_j/\alpha}$$

that for any fixed N we have

$$\tilde{\mathcal{E}_\alpha} \subset \bigcup_{n=N}^{\infty} \ \bigcup_{I\in\mathcal{K}_n \ |I|\le 2^{-n/\alpha}} |I| \ .$$

If we assume $\alpha \le \alpha_c = 1/\log_2\sqrt{12}$, we have $1\alpha \ge \log_2\sqrt{12}$ and therefore $s = -1\alpha \le -\log_2\sqrt{12}$.

If we use the above intervals I as a covering for $\tilde{\mathcal{E}_\alpha}$, we get from the large deviation estimate (73) and the definition (68) that for any $d > \alpha\,\varphi(-1/\alpha)$

$$\mathcal{H}_d\big(3^{-N},\tilde{\mathcal{E}_\alpha}\big) \le \sum_{n=N}^{\infty} 2^{-n\,d/\alpha} 2^{n\,\varphi(-1/\alpha)} \ ,$$

which implies

$$\lim_{N\to\infty} \mathcal{H}_d\big(3^{-N},\tilde{\mathcal{E}_\alpha}\big) = 0 \ ,$$

and therefore

$$d_{\mathrm{H}}\big(\tilde{\mathcal{E}_\alpha}\big) \le \alpha\,\varphi(-1/\alpha) \ . \tag{74}$$

We now prove a lower bound on $d_{\mathrm{H}}((\mathcal{E}_\alpha^- \cap \mathcal{E}_\alpha^+)$. For this purpose we will use Frostman Lemma 4.2 and we start by constructing a measure with suitable properties. This can be done using a coding and the measure appears as a Bernoulli measure, we will however construct it directly. We will in fact construct a one parameter family of measures ν_p, $p \in]0,1[$ (and we define as usual $q = 1 - p$). The construction is recursive using the sequence of intervals in $\mathcal{K}_n$. The measure μ_p will be a probability measure and we therefore set $\nu_p\big([0,1]\big) = 1$. Assume now that $\nu_p(I)$ has already been defined for any $I \in \mathcal{K}_n$. As we have seen in the recursive construction (72), the intervals in $\mathcal{K}_{n+1}$ are obtained from those of $\mathcal{K}_n$ by applying f_1 or f_2. We now define ν_p on the intervals of $\mathcal{K}_{n+1}$ by

$$\nu_{n+1}(J) = \begin{cases} p\,\nu_n(I) & \text{if } J = f_1(I)\;, \\ q\,\nu_n(I) & \text{if } J = f_2(I)\,. \end{cases}$$

We leave to the reader that this recursive construction defines a Borel measure on $[0,1]$ with support $\mathcal{K}$. In particular, $\mu = \nu_{1/2}$. Consider now the map f of exercise (4.5). It is easy to verify that its Lyapunov exponent for the measure ν_p is equal to

$$\lambda_p = p\log 4 + q\log 3\;.$$

We define a set $A \subset \mathcal{K}$ by

$$A = \left\{ x \,\middle|\, \lim_{r\to 0} \frac{\log \nu_p\big(B_r(x)\big)}{\log r} = \frac{-p\log p - q\log q}{\lambda_p} \right\}\;.$$

It follows from the ergodic Theorem (2.1) and the Shannon Mc-Millan Breiman Theorem (2.4) that $\nu_p(A) = 1$.

Exercise 4.8. Prove this relation using formula (27).

On the other hand, we see that if we choose p such that

$$\alpha = \frac{-\lambda_p}{\log 2}\;,$$

since $A \subset \mathcal{E}^-_\alpha \cap \mathcal{E}^+_\alpha$, we have

$$\nu_p\left(\mathcal{E}^-_\alpha \cap \mathcal{E}^+_\alpha\right) = 1\;.$$

Therefore, if

$$p = -\frac{\alpha + \log_2 3}{\log_2(4/3)}$$

we get using Frostman Lemma 4.2

$$d_{\mathrm{H}}(\big(\mathcal{E}^-_\alpha \cap \mathcal{E}^+_\alpha\big) \geq \frac{-p\log p - q\log q}{\lambda_p} = \alpha\,\varphi(-1/\alpha)\;.$$

Since this quantity is equal to the upper bound (74), the result follows. □

Exercise 4.9. Write a complete proof of the above result by putting all the necessary ϵ and δ.

The maximal value for the Hausdorff dimension of $\mathcal{E}^\pm_\alpha$ is the dimension of the invariant set. On the other hand, if α is equal to the Lyapunov exponent (divided by $\log 2$), then $\mathcal{E}^\pm_\alpha$ have full measure. This implies that the curves $\alpha \to \mathcal{E}^\pm_\alpha$ should be below the first diagonal and tangent to it in one point. This is illustrated in figure 29 which shows the graph of $d_{\mathrm{H}}\big(\mathcal{E}^-_\alpha\big) = d_{\mathrm{H}}\big(\mathcal{E}^+_\alpha\big)$ when α varies. This figure was produce with the two maps

$$f_1(y) = \frac{y}{29} \quad\text{and}\quad f_2(y) = \frac{2+y}{3}\;. \tag{75}$$

so that one can see more clearly the curves.

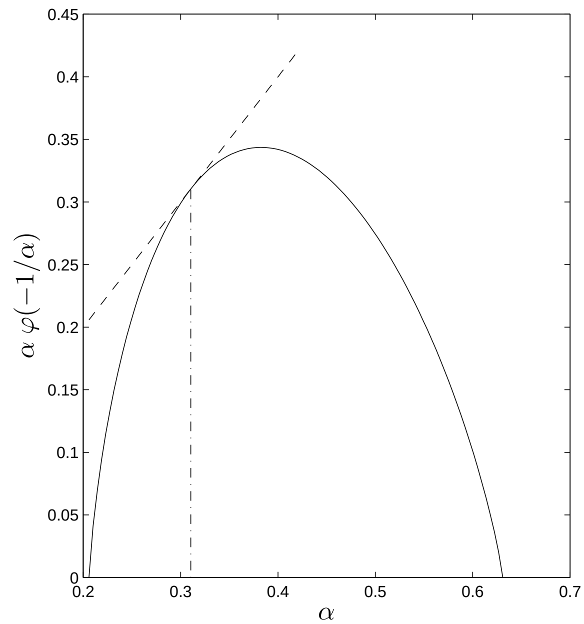

Figure 29: Hausdorff dimension of the sets $\mathcal{E}_\alpha^\pm$ as a function of α for the maps (75) .

4.7 Entropy

To measure the entropy one uses the Brin-Katok formula (2.5). As in the case of the correlation dimension, in order to compute $\mu\big(V(x,\epsilon,n)\big)$ we apply the ergodic theorem. For μ almost all z,

$$\mu\big(V(x,\epsilon,n)\big) = \lim_{N\to\infty}\frac{1}{N}\sum_{j=0}^{N-1}\chi_{V(x,\epsilon,n)}\big(T^j(z)\big)\ ,$$

$$= \lim_{N\to\infty}\frac{1}{N}\sum_{j=0}^{N-1}\prod_{l=0}^{n-1}\theta\big(\epsilon - d\big(T^l(x),T^{l+j}(z)\big)\big)\ .$$

This formula is analogous to the formula used to compute the correlation dimension (71) and similar algorithms are used in practice. One then uses the Brin-Katok formula (2.5), namely

$$h_\mu(T) = \lim_{\delta\searrow 0}\liminf_{n\to\infty} -\frac{1}{n}\log\mu\big(V(x,T,\delta,n)\big)\ .$$

4.8 Measuring the invariant measure

As we have seen, it is often the case that one does not have access to the phase space but only to an image of it. In particular, as was emphasised several times, we may only have a time series corresponding to the observation of the system through an observable g, namely a finite sequence of real numbers $y_1,\ldots,y_n$

satisfying

$$y_j = g\big(T^{j-1}(x)\big)$$

for an initial condition x on the phase space which in general is unknown. In particular, we only have an image μ_g of the invariant measure μ through g, namely

$$\mu_g = \mu \circ g^{-1} \ .$$

In order to determine μ_g, one can use the empirical distribution function

$$F_{n,g}(t) = \frac{1}{n}\sum_{j=1}^{n} \theta\big(t - y_j\big) = \frac{1}{n}\sum_{j=1}^{n} \theta\big(t - g\big(T^{j-1}(x)\big)\big) \ .$$

If the measure μ is ergodic, one concludes immediately from Birkhoff's ergodic theorem (2.1) that for μ almost every x, this sequence of random functions converges to the distribution function of μ_g, $F_g(t) = \mu_g(]-\infty, t]$. In the case of independent random variables, this is the Glivenco Cantelli theorem (see for example [17]). Again in the case of in dependent random variables (under some minor hypothesis), the fluctuations are known. More precisely, a celebrated theorem of Kolmogorov establishes that the L^∞ norm of the difference between the empirical distribution function and the true one, multiplied by $n^{1/2}$ converges in law to a random variable whose distribution is known explicitly(see [17]). This leads to the well known Kolmogorov Smirnov non parametric test (see [17]). One may wonder if similar results hold for dynamical systems. For maps of the interval equipped with the absolutely continuous invariant measure, and the observable $g(x) = x$ (which is enough in that case), this question was considered in [47] where it was proven that the process

$$\sqrt{n}\big(F_{n,g}(t) - F_g(t)\big)$$

converges to a Gaussian bridge (in general not a Brownian bridge even for the map $2x \pmod 1$). From tis result one can derive powerful non parametric tests. We refer to [46] for the details. In [31] this question was also considered for some non uniformly hyperbolic systems.

We derive here for the case of piecewise expanding maps of the interval a general estimate following from the exponential estimate (see also [47]).

Theorem 4.7. *Let T be a dynamical system on a phase space Ω, and let μ be an ergodic invariant measure satisfying the exponential estimate (55). Then there are two positive constants Γ_1 and Γ_2 such that for any real valued Lipschitz observable g with $\|g\|_{L^\infty} \le 1$ such that $g^*\mu$ is absolutely continuous with bounded density, we have*

$$\mathbf{P}\Big(\sup_t \big|F_{n,g}(t) - F_g(t)\big| > sn^{-1/4}\Big) \le\le \Gamma_1\sqrt{n}\; s^{-2} e^{\Gamma_2 s^4}.$$

Remark 4.12. The assumption $\|g\|_{L^\infty} \le 1$ is only for convenience, one can rescale the result by replacing g by $g/\|g\|_{L^\infty}$.

Proof. The difficulty here is that the function θ entering the definition of the empirical distribution $F_{n,g}$ is not Lipschitz. In order to apply nevertheless the exponential inequality, we will sandwich the function θ between two Lipschitz functions. For a positive number γ (small in the application), we define the function θ_γ by

$$\theta_\gamma(s) = \begin{cases} 0 & \text{if } s < 0 \\ s/\gamma & \text{if } 0 \le s \le \gamma \\ 1 & \text{if } s \ge \gamma\ . \end{cases}$$

We now define the random function $F_{n,\gamma,g}$ by

$$F_{n,\gamma,g}(t) == \frac{1}{n}\sum_{j=1}^{n} \theta_\gamma\big(t - g\circ T^{j-1}\big)\ .$$

It is easy to verify that for any real s we have

$$\theta_\gamma(s) \le \theta(s) \le \theta_\gamma(s+\gamma)$$

which immediately implies for any t

$$F_{n,\gamma,g}(t) \le F_{n,g}(t) \le F_{n,\gamma,g}(t+\gamma)\ . \tag{76}$$

Note that $F_{n,\gamma,g}(t)$ is now a Lipschitz function in t with Lipschitz constant γ^{-1}. Therefore, if $\delta > 0$ is a (small) number to be optimally chosen later on, we have

$$\left|\sup_t \Big(F_{n,\gamma,g}(t) - F_{\gamma,g}(t)\Big) - \sup_p \Big(F_{n,\gamma,g}(p\delta) - F_{\gamma,g}(p\delta)\Big)\right| \le \gamma^{-1}\delta\ , \tag{77}$$

where

$$F_{\gamma,g}(t) = \mathbf{E}\big(\theta_\gamma\big(t - g(\,\cdot\,)\big)\big)\ .$$

This estimate is of course interesting only if $\gamma^{-1}\delta < 1$.

We now consider the function $K_{n,\gamma,g}$ of n variables $z_1,\ldots,z_n$ given by

$$K_{n,\gamma,g}(t) = \frac{1}{n}\sum_{j=1}^{n} \theta_\gamma\big(t - g(z_j)\big)\ .$$

This function is obviously componentwise Lipschitz, and it is easy to verify that for each variable its Lipschitz constant defined in (54) is bounded by

$$L_j\big(K_{n,\gamma,g}\big) \le \frac{\gamma^{-1}L_g}{n}\ ,$$

where L_g is the Lipschitz constant of the observable g. Therefore, from Pisier's inequality, we get for any real number β

$$\mathbf{E}\left(e^{\beta \sup_{-\delta^{-1}\le p\le \delta^{-1}(1+\gamma)} \left(F_{n,\gamma,g}(p\delta) - F_{\gamma,g}(p\delta)\right)}\right)$$

$$\le \sum_{-\delta^{-1}\le p\le\delta^{-1}(1+\gamma)} \mathbf{E}\left(e^{\beta\left(F_{n,\gamma,g}(p\delta) - F_{\gamma,g}(p\delta)\right)}\right)\ ,$$

where as we have seen already several times, the expectation is taken over the measure

$$d\mu\big(z_1\big)\prod_{j=1}^{n-1}\delta\big(z_{j+1} - T(z_j)\ .$$

Using the exponential inequality (55) we get

$$\mathbf{E}\left(e^{\beta \sup_{-\delta^{-1}\le p\le \delta^{-1}(1+\gamma)}\left(F_{n,\gamma,g}(p\delta)-F_{\gamma,g}(p\delta)\right)}\right)$$

$$\le \delta^{-1}(2+\gamma)C_1 e^{C_2\beta^2\gamma^2 L_g^2/n}\ .$$

If ϵ_1 is a (small) positive number, we have using Chebychev's inequality

$$\mathbf{P}\left(\sup_{-\delta^{-1}\le p\le \delta^{-1}(1+\gamma)}\left(F_{n,\gamma,g}(p\delta)-F_{\gamma,g}(p\delta)\right)>\epsilon_1\right)\le \delta^{-1}(2+\gamma)C_1 e^{-\beta\epsilon_1}e^{C_2\beta^2\gamma^{-2}L_g^2/n}\ ,$$

and this inequality holds for any real β. Using the estimate (77) this implies for any $\epsilon_1>0$

$$\mathbf{P}\left(\sup_t\left(F_{n,\gamma,g}(t)-F_{\gamma,g}(t)\right)>\epsilon_1+\gamma^{-1}\delta\right)\le \delta^{-1}(2+\gamma)C_1 e^{-\beta\epsilon_1}e^{C_2\beta^2\gamma^{-2}L_g^2/n}\ .$$

Let the function $\omega_g(\gamma)$ be defined by

$$\omega_g(\gamma)=\sup_t\left(F_g(t)-F_{\gamma,g}(t)\right)\ .$$

We then have immediately from inequality (76)

$$\mathbf{P}\left(\sup_t\left(F_{n,\gamma,g}(t)-F_{\gamma,g}(t)\right)>\epsilon_1+\gamma^{-1}\delta+\omega_g(\gamma)\right)\le \delta^{-1}L_g C_1 e^{-\beta\epsilon_1}e^{C_2\beta^2\gamma^{-2}L_g^2/n}\ .$$

If we assume

$$\omega_g(\gamma)\le C_1\gamma\ ,$$

the optimal choice of δ and γ in the quantity

$$\epsilon_1+\gamma^{-1}\delta+\omega_g(\gamma)$$

are (within constant factors)

$$\gamma=\delta^{1/2}=\epsilon_1\ .$$

The optimal choice of β (again within constant factors) is now

$$\beta=n\epsilon_1^3\ ,$$

and we get for some positive constants Γ_1 and Γ_2

$$\mathbf{P}\Big(\sup_t\big|F_{n,g}(t)-F_g(t)\big|>sn^{-1/4}\Big)\le \Gamma_1\sqrt{n}\ s^{-2}e^{\Gamma_2 s^4}.$$

□

Although this estimate is asymptotically weaker than the convergence in law to a Gaussian bridge as in the Komomogorov Smirnov test, it allows however to obtain rather strong tests.

There are also results on kernel density estimates in the case of absolutely continuous invariant measures, see [31], [22] and [23], [126].

5 Other probabilistic results.

There are a number of questions from probabilistic origin that have been investigated in the class of stochastic processes generated by dynamical systems. We describe some of the results below.

5.1 Entrance and recurrence times.

One of the oldest question in ergodic theory is the problem of entrance time in a (small) set. For a (measurable) subset A of the phase space Ω, one defines the (first) entrance time $\tau_A(x)$ of the trajectory of x in the set A by

$$\tau_A(x) = \inf\left\{n > 0 \,\middle|\, T^n(x) \in A\right\} .$$

For x in A this number is called the return time. Note that the function τ_A is a measurable function on the phase space with values in the integers.

Historically, this quantity appeared in Boltzman ideas about ergodic theory, when he asked about the time it would take for all the molecules of a room to concentrate in only half of the available volume. His idea was that events of small probability occur only on a very large time scale.

One of the first few general results around these questions is the Poincaré recurrence theorem which dates from the same period of heated discussions about the foundations of statistical mechanics.

Theorem 5.1. *Let T be a transformation on a phase space Ω, and μ be an invariant probability measure for T. Then for each measurable set A of positive measure, there is a finite integer $n \ge 1$ such that $A \cap T^n(A) \neq \emptyset$ and moreover $\mu\big(A \cap T^n(A)\big) > 0$.*

Proof. The proof is by contradiction. Assume there exists a measurable set A with $\mu(A) > 0$ and such that for any n we have $\mu\big(A \cap T^n(A)\big) = 0$. By the invariance of the measure we have for any integer n,

$$\mu\big(T^{-n}(A) \cap A\big) \subset \mu\Big(T^{-n}\big(A \cap T^n(A)\big)\Big) = \mu\big(A \cap T^n(A)\big) = 0 .$$

Therefore, since for any integer m, we have

$$T^{-m}\big(A \cap T^{-n}(A)\big)\big) = T^{-m}(A) \cap T^{-n-m}(A) ,$$

we conclude that for any integers m and n

$$\mu\big(T^{-m}(A) \cap T^{-n-m}(A)\big) = 0; .$$

Let

$$B_p = \cup_{n=0}^{p} T^{-n}(A) ,$$

since the sets $T^{-n}(A)$ have an intersection of measure zero and the same measure $\mu(A)$, we have

$$\mu\big(B_p\big) = p\ \mu(A) .$$

This is only possible as long as $p\ \mu(A) \le 1$ since A is of positive measure. Hence we have a contradiction with our assumptions on A, and the theorem is proved. ∎

Remark 5.1. Note that it follows from the proof of the theorem that $A \cap T^n(A) \neq \emptyset$ (and in fact $\mu\big(A \cap T^n(A)\big) > 0$) for at least one n such that $1 \leq n \leq 1/\mu(A)$.

The next important result follows from the ergodic theorem, and says that the entrance time is a well defined function.

Theorem 5.2. *If A is such that $\mu(A) > 0$, and μ is ergodic, then τ_A is almost surely finite.*

Proof. The proof is by contradiction. Let B be a measurable set of positive measure where $\tau_A = \infty$. We have for almost every x in B

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \chi_A\big(T^n(x)\big) = 0 \ .$$

On the other hand by ergodicity, this quantity should be almost surely equal to $\mu(A) > 0$, hence a contradiction. $\square$

Exercise 5.1. Show that almost surely

$$\tau_A(x) = 1 + \sum_{j=1}^{\infty} \prod_{m=1}^{j-1} \chi_{A^c}\big(T^j(x)\big) \ .$$

The ergodic theorem tells us that if one considers a large time interval [0,S], and a (measurable) set A of positive measure, then the number of times a typical trajectory has visited A (the event A has occurred) is of the order $\mu(A)S$. In other words, $1/\mu(A)$ is a time scale associated to the set A, and one may wonder if this time scale is related to the entrance time. There is a general theorem in this direction due to Mark Kac.

Theorem 5.3. *Let T be a transformation on a phase space Ω, and μ be an ergodic invariant probability measure for T. Then for each measurable set A of positive measure,*

$$\int_A \tau_A d\mu = 1.$$

Proof. We give a proof in a particular case, and refer to the literature (for example [99]) for the general case. Using exercise 5.1 it immediately follows that almost surely

$$\chi_{A^c}\big(T(x)\big)\tau_A\big(T(x)\big) = \tau_A(x) - 1 = \chi_A(x)\tau_A(x) + \chi_{A^c}(x)\tau_A(x) - 1 \ .$$

If we now assume that the function τ_A is integrable, the result follows by integrating both sides of this equality and using the invariance of the measure. $\square$

There is also a general lower bound on the expectation of τ_A.

Theorem 5.4. *For any measurable set A with $\mu(A) > 0$ we have*

$$\mathbf{E}(\tau_A) \geq \frac{1}{2\mu(A)} - 3/2.$$

Note that the bound is not very good for $\mu(A) > 1/5$ since by definition $\tau_A \geq 1$.

Proof. Recall that

$$\mathbf{E}\big(\tau_A\big) = \sum_{p=1}^{\infty} \mathbf{P}\big(\tau_A \geq p\big) \ .$$

We now estimate from below each term in the sum. We have

$$\mathbf{P}\big(\tau_A \geq p\big) = \int d\mu(x) \prod_{j=1}^{p-1} \chi_{A^c}\big(T^j(x)\big) = \int d\mu(x) \prod_{j=1}^{p-1} \Big(1 - \chi_A\big(T^j(x)\big)\Big) \ .$$

It is left to the reader to verify that

$$\prod_{j=1}^{p-1} \Big(1 - \chi_A\big(T^j(x)\big)\Big) \geq 1 - \sum_{j=1}^{p-1} \chi_A\big(T^j(x)\big) \ .$$

Therefore for any $p \geq 1$ we have

$$\mathbf{P}\big(\tau_A \geq p\big) \geq 1 - (p-1)\mu(A) \ .$$

Note that this bound is not very useful if $p > 1 + 1/\mu(A)$. Let q be a positive integer to be chosen optimally later on. We have

$$\mathbf{E}\big(\tau_A\big) \geq \sum_{p=1}^{q} \mathbf{P}\big(\tau_A \geq p\big) \geq \sum_{p=1}^{q} \big(1 - (p-1)\mu(A)\big) = q - \frac{q(q-1)\mu(A)}{2} \ .$$

Since this estimate is true for any integer q, we can look for the largest right hand side. We take $q = [1/\mu(A)]$ which is near the optimum and get the estimate using $1/\mu(A) - 1 \leq q \leq 1/\mu(A)$. ☐

In general one cannot say much more on the random variable τ_A without making some hypothesis on the dynamical system or on the set A. We now give an estimate on the tail of the distribution of τ_A in the case of dynamical systems with decay of correlations like in (40).

Theorem 5.5. *Assume the dynamical system defined by the map T and equipped with the ergodic invariant measure μ satisfies the estimate (40) with $C_{\mathcal{B}_1,\mathcal{B}_2}(n)$ decaying exponentially fast, namely there are two constants $C > 0$ and $0 < \rho < 1$ such that for any integer n*

$$C_{\mathcal{B}_1,\mathcal{B}_2}(n) \leq C\rho^n \ .$$

Let A be a measurable set such that $0 < \mu(A) < 1$, and such that the characteristic function of A belongs to $\mathcal{B}_1$. Assume also that $\mathcal{B}_2 = L^1(d\mu)$. Then

$$\mathbf{P}\big(\tau_A > n\big) \leq e^{\gamma_A \, n \, \mu(A)/\log \mu(A)} \ ,$$

where γ_A depends only on the $\mathcal{B}_1$ norm of the characteristic function of A.

Note that $\log \mu(A)$ is negative and therefore the upper bound decays exponentially fast with n. This estimate says that τ_A is unlikely to be much larger than $-(\log \mu(A))/\mu(A)$. Note also that in this estimate there is no restriction on the set A except that its characteristic function belongs to $\mathcal{B}_1$. All these assumptions are satisfied in the case of piecewise expanding maps of the interval when A is an interval and μ the Lebesgue measure (see Theorem 3.3). In this case $\mathcal{B}_1$ is the set of functions of bounded variations, and the norm of the characteristic function of an interval is independent of the interval.

Proof. We have from the definition

$$\mathbf{P}\big(\tau_A > n\big) = \int \prod_{j=1}^{n} \chi_{A^c} \circ T^j(x)\, d\mu(x)\ .$$

Let k be an integer to be fixed later on, and let $m = [n/k]$. We have obviously

$$\mathbf{P}\big(\tau_A > n\big) \le \int \prod_{l=1}^{m} \chi_{A^c} \circ T^{lk}(x)\, d\mu(x)\ .$$

Using (40) with our hypothesis (and the invariance of the measure) we get

$$\mathbf{P}\big(\tau_A > n\big) \le\le \int \big(1-\chi_A\big) \prod_{l=1}^{m-1} \chi_{A^c} \circ T^{lk}(x)\, d\mu(x) \le$$

$$\Big(1-\mu(A) + C\rho^k \big\|\chi_A\big\|_{\mathcal{B}_1}\Big) \int \prod_{l=1}^{m-1} \chi_{A^c} \circ T^{lk}(x)\, d\mu(x)\ .$$

We now choose k as the smallest integer such that

$$C\rho^k \big\|\chi_A\big\|_{\mathcal{B}_1} \le \frac{1}{2}\mu(A)\ .$$

Iterating the estimate we obtain

$$\mathbf{P}\big(\tau_A > n\big) \le \big(1-\mu(A)/2\big)^m \le e^{-m\,\mu(A)/2}\ ,$$

and the result follows. $\square$

In the case of small sets A a lot of works have been devoted to the study of asymptotic laws. One can expect to see emerging something similar to the exponential law. We only mention here the case of piecewise expanding maps of the interval.

Theorem 5.6. *Let T be a piecewise expanding map of the interval, and μ a mixing a.c.i.p.m. with density h. There is a set B of full measure such that if A_n be a sequence of intervals of length tending to zero and accumulating to a point $b \in B$, then the sequence of random variables $\mu\big(A_n\big)\tau_{A_n}$ converges in law to an exponential random variable of parameter one. In other words, for any fixed number $s > 0$ we have*

$$\lim_{n\to\infty} \mathbf{P}\big(\tau_{A_n} > s/\mu\big(A_n\big)\big) = e^{-s}\ .$$

We give below the sketch of a proof of this theorem modulo a technical point for which we refer the reader to the literature. This proof is based on an idea of Kolmogorov for proving the central limit theorem in the i.i.d. case (see [16]).

Proof. In order to alleviate the notation, we will denote by ϵ the positive number $\mu(A)$. Let s be a positive number and let $n = [s/\epsilon]$. We have obviously

$$\mathbf{P}\big(\tau_A > n\big) = (1-\epsilon)^n + \sum_{q=0}^{n-1} \Big((1-\epsilon)^{n-q-1}\mathbf{P}\big(\tau_A > q+1\big) - (1-\epsilon)^{n-q}\mathbf{P}\big(\tau_A > q\big)\Big) .$$

If ϵ tends to zero, the first term converges to e^{-s}, and we only need to prove that the second term converges to zero. Since for $q \geq 1$

$$\mathbf{P}\big(\tau_A > q\big) = \mathbf{E}\big(\prod_{j=0}^{q} \chi_{A^c} \circ T^j\big) ,$$

we have from the invariance of the measure

$$(1-\epsilon)^{n-q-1}\mathbf{P}\big(\tau_A > q+1\big) - (1-\epsilon)^{n-q}\mathbf{P}\big(\tau_A > q\big)$$

$$= (1-\epsilon)^{n-q-1}\left(\mathbf{E}\Big(\prod_{j=0}^{q+1}\chi_{A^c}\circ T^j\Big) - (1-\epsilon)\mathbf{E}\Big(\big(\chi_A + \chi_{A^c}\big)\prod_{j=1}^{q+1}\chi_{A^c}\circ T^j\Big)\right)$$

$$= (1-\epsilon)^{n-q-1}\left(\epsilon\,\mathbf{E}\Big(\prod_{j=0}^{q+1}\chi_{A^c}\circ T^j\Big) - (1-\epsilon)\mathbf{E}\Big(\chi_A\prod_{j=1}^{q+1}\chi_{A^c}\circ T^j\Big)\right)$$

$$= (1-\epsilon)^{n-q-1}\left(\epsilon\,\mathbf{E}\Big(\prod_{j=1}^{q+1}\chi_{A^c}\circ T^j\Big) - \mathbf{E}\Big(\chi_A\prod_{j=1}^{q+1}\chi_{A^c}\circ T^j\Big)\right) + (1-\epsilon)^{n-q-1}\mathcal{O}(1)\epsilon^2 , \tag{78}$$

where the last term comes from the estimation

$$0 \leq \epsilon\mathbf{E}\Big(\chi_A\prod_{j=1}^{q+1}\chi_{A^c}\circ T^j\Big) \leq \epsilon\mathbf{E}\Big(\chi_A\Big) = \epsilon^2 .$$

Note that the modulus of the whole term is less than $2\epsilon+\epsilon^2$ (drop the product of the characteristic functions). Let k_A be an integer such that $A \cap T^j(A) = \emptyset$ for $j = 1, \ldots, k_A$ (it is enough to assume that the intersection has measure zero). Therefore for $q > k_A$ we have

$$\mathbf{E}\Big(\chi_A\prod_{j=1}^{q+1}\chi_{A^c}\circ T^j\Big) = \mathbf{E}\Big(\chi_A\prod_{j=k_A+1}^{q+1}\chi_{A^c}\circ T^j\Big) .$$

For the other term, we use Bonferoni's inequality, namely

$$1 - \sum_{j=1}^{k_A}\chi_A\circ T^j \leq \prod_{j=1}^{k_A}\chi_{A^c}\circ T^j \leq 1 - \sum_{j=1}^{k_A}\chi_A\circ T^j + \sum_{1\leq r\neq s\leq k_A}\chi_A\circ T^r\chi_A\circ T^s .$$

Exercise 5.2. Prove these inequalities.

This implies immediately for $q > k_A$

$$\left|\mathbf{E}\Big(\prod_{j=1}^{q+1}\chi_{A^c}\circ T^j\Big) - \mathbf{E}\Big(\prod_{j=k_A+1}^{q+1}\chi_{A^c}\circ T^j\Big)\right| \le \epsilon k_A^2\ .$$

We can now use the decay of correlations in the form of equation (40) for $\chi_A \in \mathcal{B}_1$ and $\prod_{j=k_A+1}^{q+1}\chi_{A^c}\circ T^j \in \mathcal{B}_2$. For example for a mixing a.c.i.p.m. of a piecewise expanding maps of the interval we can take for A an interval which belongs the space of functions of bounded variations ($\mathcal{B}_1$) and for $\mathcal{B}_2$ it is enough to take L^∞. We get

$$\left|\epsilon\mathbf{E}\,\Big(\prod_{j=k_A+1}^{q+1}\chi_{A^c}\circ T^j\Big) - \mathbf{E}\Big(\chi_A\prod_{j=k_A+1}^{q+1}\chi_{A^c}\circ T^j\Big)\right| \le C_{\mathcal{B}_1,L^\infty}(k_A)\ \|\chi_A\|_{\mathcal{B}_1}\ .$$

Combining all the above estimates we get for $n-1 \ge q > k_A$

$$\Big|(1-\epsilon)^{n-q-1}\mathbf{P}\big(\tau_A > q+1\big) - (1-\epsilon)^{n-q}\mathbf{P}\big(\tau_A > q\big)\Big|$$

$$\le (1-\epsilon)^{n-q-1}\left(\epsilon^2(1+k_A^2) + C_{\mathcal{B}_1,L^\infty}(k_A)\big\|\chi_A\|_{\mathcal{B}_1}\right)\ .$$

On the other hand, as was already observed, we have immediately from equation (78)

$$\Big|(1-\epsilon)^{n-q-1}\mathbf{P}\big(\tau_A > q+1\big) - (1-\epsilon)^{n-q}\mathbf{P}\big(\tau_A > q\big)\Big| \le 2\epsilon\ .$$

Summing over q (from 1 to k_A using the rough bound, and from k_A to n using the more elaborated one) we get for $n > k_A$

$$\left|\sum_{q=0}^{n-1}\Big((1-\epsilon)^{n-q-1}\mathbf{P}\big(\tau_A > q+1\big) - (1-\epsilon)^{n-q}\mathbf{P}\big(\tau_A > q\big)\Big)\right|$$

$$\le \mathcal{O}(1)\Big(\epsilon k_A + \epsilon k_A^2 + \epsilon^{-1}C_{\mathcal{B}_1,L^\infty}(k_A)\big\|\chi_A\|_{\mathcal{B}_1}\Big)\ .$$

The theorem follows for any sequence $\big(A_n\big)$ of sets for which the right hand side of the above estimate trends to zero. For a mixing a.c.i.p.m of a piecewise expanding map of the interval, $\big\|\chi_A\|_{\mathbf{BV}} = 2$, and $C_{\mathcal{B}_1,L^\infty}(k_A)$ decays exponentially fast in k_A (recall $\mathcal{B}_1 = \mathbf{BV}$ and see estimate (42). It is therefore enough to ensure that $k_A = \mathcal{O}(1)(-\log\mu(A))$. There turns out to be many such intervals. We refer to [41] and exercise 5.3 for the details. □

Exercise 5.3. Consider the full shift on two symbols $\{0,1\}$ and the product measure $(1/2, 1/2)$ (or the map $2x$ (mod 1) with the Lebesgue measure). Let $\mathcal{C}_{n,k}$ $(1 \le k/ \le n-1)$ be the set of cylinders C of length n (starting at position one) such that $C \cap \mathcal{S}C \ne \emptyset$ (recall that $\mathcal{S}$ is the shift). For $C \in \mathcal{C}_{n,k}$ show that the last $n-k$ symbols are determined by the k first ones. Show that $\mu\big(\mathcal{C}_{n,k}\big) \le 2^{k-n}$ $\big(\big|\mathcal{C}_{n,k}\big| \le 2^k\big)$. For a cylinder set C of length n, define for $1 \le q \le n-1$ the function

$$f_{C,\,q} = \chi_C\prod_{j=q}^{n-1}\big(1 - \chi_C\circ S^j\big)\ .$$

Show that $f_{c,q}$ is equal to zero if $C \in \mathcal{C}_{n,k}$ for some $q \le k \le n-1$ and is equal to χ_C otherwise. Derive that

$$\mu\left(\bigcap_{k=q}^{n-1} \mathcal{C}_{n,k}^c\right) = \int \sum_C f_{C,q}\, d\mu.$$

Show that (see Bonferoni inequality)

$$f_{C,\,q} \ge \chi_C - \sum_{j=q}^{n-1} \chi_C \chi_C \circ S^j \ .$$

Conclude that if $C \in \mathcal{C}_{n,k}$ for some $q \le k \le n-1$, then

$$\int f_{C,\,q}\, d\mu \ge \mu(C) - n2^{-n-q} \ .$$

Using the previous estimate of $\mu(C)$ for $C \in \mathcal{C}_{n,k}$ with $k < n/2$, show that there is a constant $c > 0$ such that for any n

$$\mu\left(\bigcup_{k=1}^{n-1} \mathcal{C}_{n,k}\right) \le c\, n\, 2^{-n/2} \ .$$

For $k = 2m$, let K be a cylinder of length m which does not belong to $\mathcal{C}_{m,j}$ for any $1 \le j < m$. Show that the cylinder $K \cap \mathcal{S}^m K$ belongs to $\mathcal{C}_{k,m}$. Show that $\mu\big(\mathcal{C}_{k,k/2}\big) \ge 2^{-k/2}$.

Periodic orbits of small period prevent k_A to be large, for example if A contains a fixed point we have $A \cap T(A) \ne \emptyset$. Those sets which recur too fast prevent in some sense enough loss of memory to ensure a limiting exponential law. Some results can nevertheless be obtained in these situations, we refer to [75], and [32] for some examples. By looking at the successive entrance times one can under adequate hypothesis prove that the limiting process when the set becomes smaller and smaller is a Poisson point process (or a marked Poisson point process). We refer to [75], and [32], [35], [40], [42], [76], [1], [55] for more details, results and references. The approach of two typical trajectories, namely the entrance time in a neighborhood of the diagonal for the product system leads to a marked Poisson process at least for piecewise expanding maps of the interval. We refer to [32] for details and to [97] for related results.

An interesting case of set A is the first cylinder of length n of a coded sequence. We recall that the return time is related to the entropy of the system by the Theorem of Ornstein Weiss 2.7. We refer to [119], [39], [97] and [30] for more results on this case and the study of the fluctuations.

The possible asymptotic laws for entrance times are discussed in [60] and [104].

5.2 Number of visits to a set.

Given a measurable set A of positive measure, it is natural to ask how many times a typical orbit visits A during a time interval $[0, S]$ (it is convenient to allow S to be any positive real number). This number of visits denoted by $N_A[0,S]$ is obviously given by

$$N_A[0,S](x) = \sum_{0\le j\le S} \chi_A\big(T^j(x)\big) \tag{79}$$

for the orbit of the point x. Birkhoff's ergodic theorem 2.1 tells us that if one considers a large time interval [0,S], and a (measurable) set A of positive measure, then the number of times a typical trajectory has visited A (the event A has occurred) is of the order $\mu(A)S$. In other words, $1/\mu(A)$ is a time scale associated to the set A. For a fixed A, we have discussed at length the behaviour of $N_A[0,S]$ for large S (see section 3). Another type of asymptotic is concerned with the case where $\mu(A)$ is small. In that case, on a "large" interval of time of scale $1/\mu(A)$ one expects to see only a few occurrences of the event A. It turns out that for certain classes of dynamical systems, this can be quantified in a precise way. It follows for example that if the process of successive entrance times converges in law to a Poisson point process, the number of visits suitably scaled converges in law to a Poisson random variable.

5.3 Records.

Another interesting quantity to look at is the distance of a typical finite orbit to a given point or the minimal distance between two finite orbits. Consider a dynamical system with a metric d on the phase space. Assume a point x_0 has been chosen once for all. One can look at the successive distances to x_0 of the orbit of x, namely $d\big(x_0,x\big),\ldots,d\big(x_0,T^n(x)\big)$. This is a sequence of positive numbers and the record is the smallest one. This number depends of course on x_0, n and x, and we define it more precisely by

$$R_{n,x_0}(x) = -\inf_{1\le j\le n} \log d\big(x_0,T^j(x)\big)\ .$$

The logarithm is a convenient way to transform a small number into a large one but other means can be used to do that (for example the inverse). The logarithm is more convenient to formulate the results. If we consider a measure μ on the phase space for the distribution of x (for example an invariant measure), then R_{n,x_0} becomes a random variable. The law of R_{n,x_0} is related to entrance times. Indeed, it follows at once from the definition that if $R_{n,x_0}(x)>s$, then $d\big(x_0,T^j(x)\big)>e^{-s}$ for $0\le j\le n$. In other words

$$\big\{x \,\big|\, R_{n,x_0}(x)>s\big\} = \Big\{x \,\big|\, \tau_{B_{e^{-s}}(x_0)}(x)>n\Big\}\bigcap B_{e^{-s}}(x_0)^c\ .$$

This allows to pass from results on entrance times to records and vice versa. We now give a simple estimate on the record related to the capacitary dimension.

Definition 5.1. Let μ be a measure on a metric space Ω with a metric d, and let x_0 be a point in Ω. The capacitary dimension d_{Cap} of μ at x_0 is defined by

$$d^{x_0}_{\mathrm{Cap}}(\mu) = \sup\left\{\beta\ge 0 \,\middle|\, \int \frac{1}{d(x,x_0)^{\beta}}\ d\mu(x)<\infty\right\}\ .$$

We refer to [87] and [115] for more results on this quantity. Note that it can be infinite (for example if x_0 does not belong to the support of μ).

Theorem 5.7. *For a fixed x_0 and an invariant measure μ we have*

$$\limsup_{n\to\infty} \frac{1}{\log n}\mathbf{E}\big(R_{n,x_0}\big) \le \frac{1}{d_{\mathrm{Cap}}^{x_0}(\mu)} \ .$$

This result is somewhat similar to the general estimate (5.4).

Proof. We first apply Pisier inequality, namely for a fixed number $\beta > 0$ we have

$$e^{\beta R_{n,x_0}(x)} \le \sum_{j=0}^{n} e^{-\beta \log d\left(T^j(x),x_0\right)} = \sum_{j=0}^{n} \frac{1}{d\big(T^j(x),x_0\big)^{\beta}} \ .$$

Integrating over μ and using the invariance of this measure we get

$$\mathbf{E}\left(e^{\beta R_{n,x_0}}\right) \le n \int \frac{1}{d\big(T^j(x),x_0\big)^{\beta}}\ d\mu(x) \ .$$

This bound is only interesting if $\beta < d_{\mathrm{Cap}}^{x_0}(\mu)$ which is the condition for the right hand side to be finite. Using Jensen's inequality we get

$$e^{\beta \mathbf{E}\left(R_{n,x_0}\right)} \le n \int \frac{1}{d\big(T^j(x),x_0\big)^{\beta}}\ d\mu(x) \ ,$$

and taking the log we obtain

$$\mathbf{E}\left(R_{n,x_0}\right) \le \frac{\log n}{\beta} + \frac{1}{\beta}\log\left(\int \frac{1}{d\big(T^j(x),x_0\big)^{\beta}}\ d\mu(x)\right) \ .$$

This implies for any $\beta < d_{\mathrm{Cap}}^{x_0}(\mu)$

$$\limsup_{n\to\infty} \frac{1}{\log n}\mathbf{E}\big(R_{n,x_0}\big) \le \frac{1}{\beta}$$

and the result follows. $\square$

It is in general more difficult to obtain a lower bound.

We observe that if we know that the entrance time in a ball $B_r(x_0)$ of radius r around a point x_0 converges in law when $r \to 0$ (after suitable normalisation) to an exponential distribution, then we can often say something on the asymptotic distribution of the record (suitably normalised). Namely, if there is a sequence Λ_r such that

$$\lim_{r\to\infty} \mathbf{P}\big(\tau_{B_r(x_0)} > \Lambda_r s\big) = e^{-s}$$

then we get

$$\begin{aligned}\mathbf{P}\big(\tau_{B_r(x_0)} > \Lambda_r s\big) &= \mathbf{P}\left(-\log\left(d\big(T^j(x),x_0\big)\right) < -\log r \;\middle|\; 1 \le j \le \Lambda_r s\right)\\ &= \mathbf{P}\left(R_{[\Lambda_r s],x_0} < -\log r\right) = e^{-s} \ . \end{aligned} \tag{80}$$

where we have assumed that the record does not occur at the initial time. Assume now, as it is often the case that

$$\Lambda_r \approx r^{-a} \ .$$

Note that if we are in a regular situation,

$$\mu\big(B_r(x_0)\big) \approx r^{d_{\mathrm{H}}(\mu)}$$

and

$$\Lambda_r \approx \frac{1}{\mu\big(B_r(x_0)\big)} \approx r^{-d_{\mathrm{H}}(\mu)} \;.$$

It follows from (80) by defining $n = [\Lambda_r s]$, that

$$\lim_{n\to\infty} \mathbf{P}\left(R_{n,x_0} < \frac{1}{a}\log n + v\right) = e^{-e^{-av}}$$

which is Gumble's law (see [67]). We refer to [36] for more details.

5.4 Quasi invariant measures.

Another notion related to entrance times is that of dynamical systems with holes. Let A be a measurable subset of the phase space Ω and imagine that the trajectory under the map T is killed when it enters A for the first time (the system leaks out of A). Various interesting questions arise for the trajectories which have survived up to a time n. In particular the set of initial conditions which never enter A (in other words which survive for ever) is given by

$$\mathscr{R} = \bigcap_{j=0}^{\infty} T^{-j}\big(A^c\big) \;.$$

It often occurs that this set is small, for example of measure zero for some interesting measure. A simple example is given by the one parameter family of maps of the interval $[0, s]$ (for $s > 1$).

$$f_s(x) = \begin{cases} s(1 - |1 - 2x|) & \text{for } 0 \le x \le 1 \\ x & \text{for } 1 < x \le s \end{cases} \tag{81}$$

The image of the interval $]1/2s, 1 - 1/2s[$ falls outside the interval $[0, 1]$, see figure 30.

Exercise 5.4. Show that for the above maps f_s and $A =]1, s]$, the set $\mathscr{R}$ is a Cantor set of Lebesgue measure zero.

It may sometimes occur that the set $\mathscr{R}$ is empty. From now on we will always assume that this is not the case.

Assume now we have a given measure μ on the phase space Ω. We will denote by Ω_n the set of initial conditions which have survived up to time n (in other words whose trajectory has not met A), namely

$$\Omega_n = \bigcap_{j=0}^{n} T^{-j}\big(A^c\big) \;.$$

This is obviously the subset of $A^c a$ where $\tau_A > n$. As mentioned above, it is often the case that $\mu\big(A_n\big)$ tends to zero when n tends to infinity.

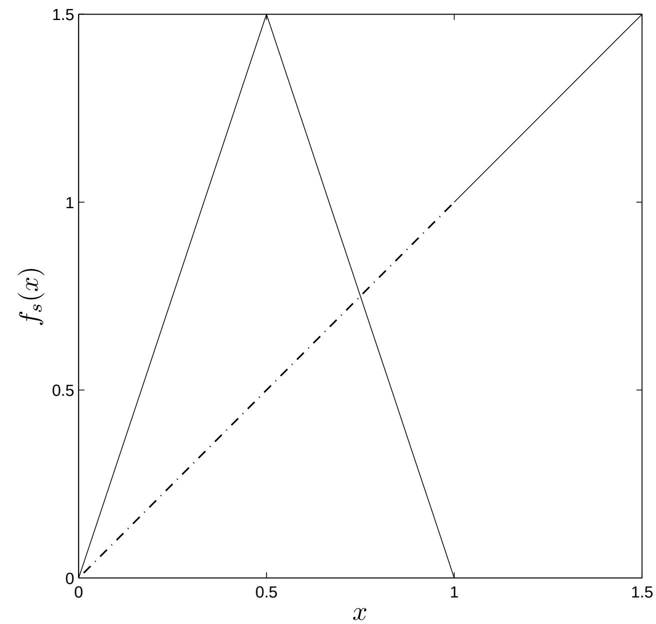

Figure 30: The graph of the map $x \to f_{3/2}(x)$ of formula (81).

One can look at the restriction of μ to these initial conditions, namely define a sequence of probability measures $\big(\mu_n\big)$ by

$$\mu_n(B) = \frac{\mu\left(B \cap \Omega_n\right)}{\mu\left(\Omega_n\right)} \ ,$$

which is just the conditional measure on the set of initial conditions which have not escaped up to time n (here and in the sequel we assume that for any n). The limiting measure if it exists is supported by the set $\mathscr{R}$. Such measures have been studied extensively in particular in the case of complex dynamics (see [28]), and more generally in the case of repulsors. We refer to [48], [33] and references therein.

Another interesting notion was introduce by Yaglom and is often called the Yaglom limit. It is connected with the notion of quasi invariant measure (or quasi invariant distribution). Consider the initial conditions which have survived up to time n, namely which belong to Ω_n. The n^{th} iterate of many of these initial conditions are points which are on the verge of escaping. Can one say something about their distribution? This leads to consider a sequence $\big(\nu_n\big)$ of probability measures given by

$$\nu_n(B) = \frac{\mu\left(T^{-n}(B) \cap \Omega_n\right)}{\mu\left(\Omega_n\right)} \ .$$

Again one can ask if there is a limit for this sequence (the so called Yaglom limit). It turns out that these measures satisfy an interesting relation.

Lemma 5.8. *For any measurable set B and any integer n we have*

$$\nu_n\big(T^{-1}(A^c \cap B)\big) = \frac{\nu\big(\Omega_{n+1}\big)}{\nu\big(\Omega_n\big)}\nu_{n+1}(B) \ .$$

Exercise 5.5. Give a proof of this lemma.

Iterating this relation, one gets for any integers n and p

$$\nu_n\big(\Omega_p \cap T^{-p}(B)\big) = \frac{\nu\big(\Omega_{n+p}\big)}{\nu\big(\Omega_n\big)}\nu_{n+p}(B) \ .$$

In particular, if the ratio $\nu\big(\Omega_{n+1}\big)/\nu\big(\Omega_n\big)$ converges to a number $\lambda > 0$, and the sequence $\big(\nu_n\big)$ converges in an adequate sense to a measure ν, this measure satisfies the relation

$$\nu\big(T^{-1}(A^c \cap B)\big) = \rho\, \nu(B) \ . \tag{82}$$

By analogy with equation (10) such a measure is called a quasi invariant measure.

Note that by taking $B = A^c$ we get

$$\rho = \frac{\nu\big(T^{-1}(A^c)\big)}{\nu\big(A^c\big)} \ .$$

Iterating equation (82) one gets

$$\nu\big(\Omega_p \cap T^{-p}(B)\big) = \rho^p \nu(B) \ .$$

Taking $B = \Omega$ and using that ν is a probability measure we get

$$\nu\big(\Omega_p\big) = \rho^p \ .$$

In other words, since

$$\nu\big(\Omega_p\big) = \nu\big(\tau_A > p\big)$$

we see that in the measure ν, the entrance time is exactly exponentially distributed.

In the other direction, one can show the following theorem.

Theorem 5.9. *Let ν_0 be a probability measure on Ω_0. Assume that*

$$\limsup_{n\to\infty} \frac{\log \nu_0\big(\Omega_n\big)}{n} < 0 \ .$$

Then there exists a quasi invariant measure.

We refer to [45] for a proof. Note that the quasi invariant measure whose existence is ascertained by the theorem may not be unique and need not have simple relations to the measure ν_0. We refer to [45] and [33] for details and references.

As an example, we come back to the case of the map f_s of example (81), with the set $A =]1, s]$. Let ν be the measure on $[0, s]$ which is the Lebesgue measure on the interval $[0, 1]$ and vanishes on $]1, s]$. In order to check that ν is a quasi invariant measure, it is enough to take $B \in A^c = [0,1]$ and by definition (82) to verify that

$$\frac{\nu\big(T^{-1}(B)\big)}{\nu\big(B\big)}$$

does not depend on B. Since ν is a Borel measure, it is enough to verify this assumption for finite unions of intervals. Moreover one readily checks that if this relation holds for any interval (contained in $[0,1]$), it will also hold for any finite union of intervals.

If $B = [a,b]$ ($0 \le a < b \le 1$), we have at once

$$T^{-1}(B) = [a/(2s), b/(2s)] \cup [1 - b/(2s), 1 - a/(2s)] \;.$$

Therefore

$$\nu\big(T^{-1}(B)\big) = \frac{(b-a)}{s}$$

and

$$\frac{\nu\big(T^{-1}(B)\big)}{\nu\big(B\big)} = \frac{1}{s}$$

which does not depend on B.

5.5 Stochastic perturbations.

As we already mentioned several times, in concrete experiments (and to some extent also in numerical simulations) one obtains a unique record of a (finite piece of) trajectory of a dynamical (or its image through an observable), often corrupted by noise. There are many models for this effect and we will only deal here with the simplest case. We will assume that the phase space is contained in $\mathbf{R}^d$ and that there is a sequence of i.i.d random variables ξ_0, ξ_1, $\ldots$ such that instead of observing the orbit of an initial condition x, one observes sequence of point $\big(x_n\big)$ in the phase space given by

$$x_{n+1} = T(x_n) + \epsilon\, \xi_n \;, \tag{83}$$

where ϵ is a fixed parameter. The process $\big(x_n\big)$ is called a stochastic perturbation of the dynamical system T. In most cases, the noise $\big(\xi_n\big)$ is small and it is convenient to normalize its amplitude by this parameter ϵ. For example one can impose that the variance of the ξ_n is equal to one. In experimental situation this noise reflects the fact that extra effects neglected in the description of the experiment constantly perturb the system. A good example are temperature fluctuations. Note also that formula (83) describes a non-autonomous situation since the map also depends on time through the sequence $\big(\xi_n\big)$.

We will always assume below that the average of ξ_n is equal to zero. This is a natural assumption, for example the experimental noises are often white noises of zero average. If the average m of ξ_n is not zero, one can subtract it from the noise and add it to the transformation, namely replace ξ_n by $\xi_n - m$ and T by the map $x \to T(x) + \epsilon\, m$. In formula (83) this produces the same sequence $\big(x_n\big)$.

In the presence of noise there are two basic questions.

i) Can one recover to some extent the true trajectory out of the noisy one?

ii) Can one recover the statistical properties?.

Concerning the first question we have already mentioned the existence of denoising algorithms (see remark 4.2). We also recall that the shadowing lemma allows to construct a true orbit in the vicinity of a noisy one (see [20]). The more ambitious complete reconstruction of a trajectory out of a noisy perturbation (of infinite length) was discussed in several papers by Lalley and Nobel (see [100], [101], and [102]). We refer to the literature for the results

Concerning the second question, namely recovering the statistical properties of the dynamical system out of noisy data, we first observe that due to the hypothesis of independence of the ξ_n in (83), the sequence $\big(x_n\big)$ defined in formula (83) is a Markov process. Assume moreover that the random variables ξ_n have a density ρ with respect to the Lebesgue measure. Then the transition probability is given by

$$p_\epsilon\big(x_{n+1}\,\big|\,x_n\big) = \frac{1}{\epsilon}\,\rho\left(\frac{x_{n+1}-T\big(x_n\big)}{\epsilon}\right)\ . \tag{84}$$

Exercise 5.6. Prove this formula.

Formally, when ϵ tends to zero, the right hand side converges to $\delta\big(x_{n+1} - T\big(x_n\big)\big)$ (where δ is the Dirac distribution) and we recover the original dynamical system. This limiting process is however quite singular and requires some care.

It follows at once from formula (84) that in the case where the noise ξ_n has a density, any invariant measure for the Markov chain $\big(x_n\big)$ is absolutely continuous (there may be several invariant measures if the chain is not recurrent).

A natural question is what are the accumulation points of the invariant measures of the chain when the amplitude ϵ of the noise tends to zero. One should of course as usual avoid the phenomenon of escape of mass to infinity, although several works have been devoted to this interesting question in the context of quasi stationary measures, see [51], [49] and [128].

This can be done for example by imposing a compact phase space (in that case some correction to the noise has to be made near the boundary).

Proposition 5.10. *Assume the phase space is compact, that the map T of the dynamical system is continuous and the noise is bounded. Then any (weak) accumulation point of invariant measures of the Markov chain when the amplitude ϵ of the noise tends to zero is an invariant measure of T.*

Proof. Let $\big(\epsilon_n\big)$ be a sequence tending to zero, and assume we have an associated sequence $\big(\mu_{\epsilon_n}\big)$ of invariant measures (μ_{ϵ_n} is an invariant measure of the chain p_{ϵ_n}) which converges weakly to a measure μ. To show that μ is invariant, it is enough to prove (see exercise 1.13) that for any continuous function g we have

$$\int g\circ T\; d\mu = \int g\; d\mu\ .$$

From the invariance of μ_{ϵ_n}, we have for any fixed continuous function g

$$\mathbf{E}\left(\int g\big(T(x)+\epsilon_n\xi\big)\; d\mu_{\epsilon_n}(x)\right) = \int g(x)\; d\mu_{\epsilon_n}(x)\ ,$$

where the expectation is over ξ. Since the phase space is compact, g is uniformly continuous, and therefore since the noise is bounded and ϵ_n tends to zero, for any $\eta > 0$ we can find an integer N such that for any $n > N$ we have

$$\left|g\big(T(x)+\epsilon_n\xi\big) - g\big(T(x)\big)\right| < \eta \,.$$

This implies for any $n > N$

$$\left|\int g\big(T(x)\big)\, d\mu_{\epsilon_n} - \int g(x)\, d\mu_{\epsilon_n}(x)\right| < \eta \,,$$

and the result follows. □

This leads naturally to the following definition.

Definition 5.2. The invariant measure μ of a dynamical system is stochastically stable (with respect to the noise $\big(\xi_n\big)$) if it is an accumulation point of invariant measures of the stochastic processes (83) when the amplitude ϵ of the noise tends to zero.

Even if the random variables ξ_n have a density (and we have seen that in this case the invariant measures of the chain are absolutely continuous), the accumulation points may not be absolutely continuous. This is necessarily the case if the map T has only attracting sets of Lebesgue measure zero. The nearest we have seen to an absolutely continuous invariant measure is an SRB measure (see definition 2.6) whose conditional measures on the local unstable manifolds are absolutely continuous. The stochastic stability of SRB measures has been indeed proven for axiom A systems and in some non uniformly hyperbolic cases (see [156] and references therein).

Theorem 5.11. *Let (Ω, T) be an axiom A dynamical system. Let Λ be an attractor of T with a basin of attraction U. Let (x_n) be a stochastic perturbation as in (83) with a bounded noise with continuous density. Then for ϵ small enough there is a unique invariant measure μ_ϵ for the process (83) supported in U, and this measure converges weakly when ϵ tends to zero to the SRB measure.*

We will not prove this theorem, instead we will prove the stochastic stability of a.c.i.p.m. for piecewise expanding map of the circle. The theory is analogous to the case of the interval, and in particular Theorem 3.3 applies. Considering the circle in spite of the interval avoids unnecessary complications (which do not carry any interesting concepts). The problem is that on the interval if we use formula (83) for the definition of the stochastic perturbation, even if ξ_n is bounded, if the point $T(x_n)$ is too near to the boundary, the point x_{n+1} can be outside of the interval. This may be an interesting leaking problem somewhat related to the questions of section 5.4 but we do not want to add this supplementary effect here.

Theorem 5.12. *Let f be a piecewise expanding map of the circle with a unique mixing absolutely continuous invariant measure with density h. Assume the noise ξ_n has a density ρ which is continuous and with compact support. Then for any $\epsilon > 0$ small enough, the stochastic process*

$$\theta_{n+1} = f(\theta_n) + \epsilon\xi_n \pmod{2\pi} \tag{85}$$

has a unique invariant measure μ_ϵ which is absolutely continuous with density h_ϵ (which is a continuous function). When ϵ tends to zero, μ_ϵ converges weakly to $h\,dx$ and more precisely h_ϵ converges to h in the Banach space of functions of bounded variations.

Proof. Since the support of ρ is bounded, there is a number $\epsilon_0 > 0$ such that for any $\epsilon \in [0, \epsilon_0[$ we have almost surely and for any integer n the bound $\big|\xi_n\big| < \pi$. In particular, since for such an ϵ the function $\rho(x/\epsilon)$ vanishes outside the interval $]-\pi, \pi[$, we can consider this function as a function defined on the circle. We will only consider below amplitudes ϵ of the noise belonging to the interval $\epsilon \in [0, \epsilon_0[$. Let R_ϵ be the operator defined on integrable functions of the circle by

$$R_\epsilon g(\theta) = \frac{1}{\epsilon} \int \rho\left(\frac{\theta - \phi}{\epsilon}\right) g(\phi) d\phi \;.$$

Using Young's Theorem, it follows at once that this operator is bounded in $L^1(d\,\theta)$ and $L^\infty(d\,\theta)$ with norm one. It is also easy to verify that it is bounded in the Banach space $\mathbf{BV}$ also with norm one.

Exercise 5.7. Prove this last statement using the equivalent formula

$$R_\epsilon g(\theta) = \frac{1}{\epsilon} \int \rho\big(\psi/\epsilon\big) g(\theta - \psi) d\psi \;.$$

It is also easy to verify that R_ϵ tends strongly to the identity in the Banach spaces $L^1(d\,\theta)$, $L^\infty(d\,\theta)$ and $\mathbf{BV}$. Namely for any $g \in \mathcal{B}$ we have

$$\lim_{\epsilon \to 0} \big\|R_\epsilon(g) - g\big\|_{\mathcal{B}} = 0 \;.$$

Exercise 5.8. Prove this assertion. Hint: observe that

$$R_\epsilon(g)(\theta) - g(\theta) = \frac{1}{\epsilon} \int \rho\big(\psi/\epsilon\big)\big(g(\theta - \psi) - g(\theta)\big) d\psi \;,$$

and approximate g by piecewise continuous functions.

As we have explained before, the transition kernel of the stochastic process (85) is given by (84), and therefore any invariant measure $d\mu_\epsilon$ has a density h_ϵ given by

$$h_\epsilon(\theta) = \frac{1}{\epsilon} \int \rho\left(\frac{\theta - f\big(\phi\big)}{\epsilon}\right) d\mu_\epsilon(\phi) \;.$$

In particular, this function h_ϵ should satisfy the equation

$$h_\epsilon(\theta) = \frac{1}{\epsilon} \int \rho\left(\frac{\theta - f\big(\phi\big)}{\epsilon}\right) h_\epsilon(\phi)\; d\phi \;.$$

This relation implies immediately that h_ϵ should be continuous and with support the whole circle.

Exercise 5.9. Prove these statements (use that ρ is continuous).

Performing a change of variables as in equation (13), one obtains immediately that h_ϵ should satisfy the equation

$$h_\epsilon = R_\epsilon P h_\epsilon \;,$$

where P is the Perron-Frobenius operator (14). Had R_ϵ converged in norm to the identity, we could have used perturbation theory to finish the proof. This is however not true.

Exercise 5.10. Prove that in the three Banach spaces $\mathcal{B}$ above, for any $\epsilon > 0$

$$\big\|R_\epsilon - I\big\|_{\mathcal{B}} = 2 \ .$$

In order to control the perturbation in a weaker sense, we recall that from Theorem 3.3 we can write $P = P_0 + Q$ where P_0 is a rank one operator, and Q is an operator with spectral radius $\sigma < 1$. Therefore

$$R_\epsilon P = R_\epsilon P_0 + R_\epsilon Q \ ,$$

and here again, $R_\epsilon P_0$ is an operator of rank one. We now analyse the resolvent of this operator. We first observe that the rank one operator $R_\epsilon P_0$ has eigenvalue one. Indeed, we have (see Theorem 3.3)

$$\big(P_0 g\big)(x) = h(x) \int g(y)\, dy \ ,$$

and therefore since R_ϵ preserves the integral

$$\int \big(R_\epsilon P_0 g\big)\ (x) dx = \int \big(P_0 g\big)(x)\ dx = \int g(y)\, dy \ .$$

In other words, the integration with respect to the Lebesgue measure is a left eigenvector of eigenvalue one. This is not surprising and reflects the fact that the transition kernel is a probability. By a similar argument one also checks that for any integrable function g

$$\int \big(R_\epsilon Q g\big)\ (x) dx = 0 \ ,$$

namely $P_0 R_\epsilon Q = 0$. Since $R_\epsilon P_0$ is an operator of rank one with eigenvalue one, it is a projection.

Exercise 5.11. Prove directly that $\big(R_\epsilon P_0\big)^2 = R_\epsilon P_0$.

Therefore, for any complex number z with $z \neq 0, 1$ we have

$$\big(R_\epsilon P_0 - z\big)^{-1} = \frac{R_\epsilon P_0}{1 - z} + \frac{I - R_\epsilon P_0}{z} \ .$$

This implies that, for $z \neq 0, 1$, we can write

$$R_\epsilon P - z = \big(R_\epsilon P_0 - z\big)\Big(I - \big(R_\epsilon P_0 - z\big)^{-1} R_\epsilon Q\Big)$$

$$= \big(R_\epsilon P_0 - z\big)\left(I - \frac{R_\epsilon P_0}{1 - z} R_\epsilon Q - \frac{I - R_\epsilon P_0}{z} R_\epsilon Q\right) \ .$$

As we have seen above, $P_0 R_\epsilon Q = 0$, and we get

$$R_\epsilon P - z = \big(R_\epsilon P_0 - z\big)\left(I - \frac{R_\epsilon Q}{z}\right) \ .$$

Assume for the moment $\left\|Q\right\|_{\mathbf{BV}} < 1$ (recall that it is only the spectral radius which is known to be smaller than one). We can try to obtain the resolvent of $R_\epsilon P$ by using a Neuman series for the inverse of the last factor in the above formula. In other words, we have to investigate the convergence of the sum of operators

$$K_\epsilon(z) = \sum_{j=0}^{\infty}\left(\frac{R_\epsilon Q}{z}\right)^j \ . \tag{86}$$

This series converge for any complex number z such that $|z| \ge \upsilon(\epsilon) = \left\|R_\epsilon Q\right\|_{\mathbf{BV}}$. In particular, if $|z| > \upsilon(\epsilon)$ and $z \neq 1$, the operator $R_\epsilon P - z$ is invertible, and its inverse is given by

$$(R_\epsilon P - z)^{-1} = K_\epsilon(z)\left(\frac{R_\epsilon P_0}{1-z} + \frac{I - R_\epsilon P_0}{z}\right) \ .$$

In particular, if $\upsilon(\epsilon) < 1$, we conclude that the operator $R_\epsilon P$ has one as a simple eigenvalue, and this is the only point in the spectrum outside the disk $|z| \le \upsilon(\epsilon)$. We can therefore express the (rank one) spectral projection P_ϵ on this eigenvalue one by the formula (see [88])

$$P_\epsilon = \frac{1}{2\pi i}\int_{|z-1|<(1-\upsilon(\epsilon))/2}(R_\epsilon P - z)^{-1}\,dz =$$

$$\frac{1}{2\pi i}\int_{|z-1|<(1-\upsilon(\epsilon))/2}K_\epsilon(z)\left(\frac{R_\epsilon P_0}{1-z} + \frac{I - R_\epsilon P_0}{z}\right)dz \ .$$

The function $P_\epsilon h$ is proportional to h_ϵ since P_ϵ is a rank one projection on this function. Moreover, since h is of integral one and the integration with respect to the Lebesgue measure is a left eigenvector of P_ϵ of eigenvalue one, we have

$$\int \big(P_\epsilon h\big)(x)\ dx = \int h(x)\ dx = 1 \ .$$

This implies immediately

$$P_\epsilon h = h_\epsilon \ .$$

Using the above formula for P_ϵ and the fact that $P_0 h = h$ we get from Cauchy's formula

$$h_\epsilon = \frac{1}{2\pi i}\int_{|z-1|<(1-\upsilon(\epsilon))/2}K_\epsilon(z)R_\epsilon h\ \frac{dz}{1-z} = K_\epsilon(1)R_\epsilon h \ .$$

We now observe that

$$h_\epsilon = K_\epsilon(1)R_\epsilon h = R_\epsilon h + \sum_{n=0}^{\infty}\big(R_\epsilon Q\big)^n R_\epsilon Q R_\epsilon h \ .$$

Since R_ϵ converges strongly to the identity in $\mathbf{BV}$ and since $Qh = 0$, we have

$$\lim_{\epsilon\to 0} Q R_\epsilon h = 0 \ .$$

Since we have assumed that $\left\|Q\right\|_{\mathbf{BV}} < 1$, and since we have established that $\left\|R_\epsilon\right\|_{\mathbf{BV}} = 1$, this implies

$$\lim_{\epsilon\to 0}\sum_{n=0}^{\infty}\big(R_\epsilon Q\big)^n R_\epsilon Q R_\epsilon h = 0 \ .$$

It then follows again from the strong convergence of R_ϵ to the identity that

$$\lim_{\epsilon \to 0} h_\epsilon = h \ .$$

If $\left\|Q\right\|_{\mathbf{BV}} \geq 1$, since the spectral radius of Q is strictly smaller than one, there is an equivalent norm where Q is a strict contraction (see [88]). One equips $\mathbf{BV}$ with this new equivalent norm, and repeats the above argument. The details are left to the reader. $\square$

We refer to [156] and [15] for references and more results on this subject.

References

[1] M.Abadi.Exponential approximation for hitting times in mixing processes. Math. Phys. Electron. J. **7**, 19, (2001).

[2] H.Abarbanel. *Analysis of Observed Chaotic Data*. Springer 1996.

[3] V.Afraimovich, S.B.Hsu. *Lectures on Chaotic Dynamical Systems*. AMS, International Press 2002.

[4] T.Alligood, T.Sauer, J.Yorke. *Chaos: an Introduction to Dynamical Systems*. Springer 1996.

[5] A.Avila, C.Moreira. Statistical properties of unimodal maps: the quadratic family. Ann. of Math. **161**, 831-881 (2005).

[6] V. Baladi, *Positive transfer operators and decay of correlations*, Advanced series in nonlinear dynamics **16**, World Scientific, 2000.

[7] A.Barbour, R.Gerrard, D.Reinert. Iterates of expanding maps. Probab. Theory Related Fields **116** 51-180 (2000).

[8] L.Barreira, Ya.Pesin. Lyapunov exponents and smooth ergodic theory. University Lecture Series, 23. American Mathematical Society, Providence, RI, 2002. Lectures on Lyapunov exponents and smooth ergodic theory. Appendix A by M. Brin and Appendix B by D. Dolgopyat, H. Hu and Pesin. Proc. Sympos. Pure Math., 69, Smooth ergodic theory and its applications (Seattle, WA, 1999), 3-106, Amer. Math. Soc., Providence, RI, 2001.

[9] L.Barreira. Pesin theory. *Encyclopaedia of Mathematics*, Supplement I, Kluwer, 1997, pp. 406-411.

[10] M.Benedicks, L.Carleson. The dynamics of the Hénon map. Ann. of Math. **133**, 73-169 (1991).

[11] M.Benedicks, L.-S.Young. Sinaĭ-Bowen-Ruelle measures for certain Hénon maps. Invent. Math. **112**, 541-576 (1993).

[12] L.Barrera, J.Schmeling. Sets of "non typical" points have full topological entropy and full Hausdorff dimension. Israel J. Math. **116**, 29-70 (2000).

[13] P.Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York 1968.

[14] G.D.Birkhoff. Proof of the ergodic theorem. Proc. Nat. Acad. Sci. USA, **17**, 656-660 (1931).

[15] M.Blank, G.Keller. Stochastic stability versus localization in one-dimensional chaotic dynamical systems. Nonlinearity **10**, 81-107 (1997).

[16] A.Borovkov. Kolmogorov and boundary value problems of probability theory. Russian Math. Surveys **59**, 91-102 (2004).

[17] A.Borovkov. *Mathematical statistics*. Gordon and Breach Science Publishers, Amsterdam, 1998.

[18] S.Borovkova. Estimation and prediction for nonlinear time series. Thesis Groningen 1998.

[19] S.Borovkova, R.Burton and H.Dehling. Consistency of the Takens estimator for the correlation dimension. Ann. Appl. Prob. **9** 376-390 (1999).

[20] R.Bowen. *Equilibrium States and Ergodic Theory of Anosov Diffeomorphisms*. Lecture Notes in Mathematics **470**. Springer-Verlag, Berlin Heidelberg New York 1975.

[21] D.Bosq, D.Guegan, G.Leorat. Statistical estimation of the embedding dimension of a dynamical system. Internat. J. Bifur. Chaos Appl. Sci. Engrg. **9**, 645-656 (1999).

[22] D.Bosq, D.Guegan. Nonparametric estimation of the chaotic function and the invariant measure of a dynamical system. Statist. Probab. Lett. **25**, 201-212 (1995).

[23] D.Bosq. Optimal asymptotic quadratic error of density estimators for strong mixing or chaotic data. Statist. Probab. Lett. **22**, 339-347 (1995).

[24] X.Bressaud, R.Fernndez, A.Galves. Speed of $\overline{d}$-convergence for Markov approximations of chains with complete connections. A coupling approach. Stochastic Process. Appl. **83** 127-138 (1999). X.Bressaud, R.Fernndez, A.Galves. Decay of correlations for non-Hölderian dynamics. A coupling approach. Electron. J. Probab. **4** (1999).

[25] X.Bressaud, C.Liverani. Anosov diffeomorphisms and coupling. Ergodic Theory Dynam. Systems, **22**, 129-152 (2002).

[26] M.Brin, A.Katok. On local entropy. In *Geometric dynamics*. Lecture Notes in Math., **1007**, 30-38 Springer, Berlin, 1983.

[27] A.Broise. *Transformations dilatantes de l'intervalle et théorèmes limites. Etudes spectrales d'opérateurs de transfert et applications*. Astérisque **238** 1996.

[28] H.Brolin. Invariant sets under iteration of rational functions. Ark. Mat. **6** 103-144 (1965).

[29] J.Buzy. Specification on the interval. Trans. Mare.Math. Soc. **349**, 2737-2754 (1997).

[30] J.-R.Chazottes, P.Collet. Almost sure limit theorems for expanding maps of the interval. Ergodic Theory & Dynamical Systems **25**, 419-441 (2005).

[31] J.-R.Chazottes, P.Collet, B.Schmitt. Devroye inequality for a class of non-uniformly hyperbolic dynamical systems. Nonlinearity **18**, 2323-2340 (2005). Statistical consequences of the Devroye inequality for processes. Applications to a class of non-uniformly hyperbolic dynamical systems. Nonlinearity **18**, 2341-2364 (2005).

[32] Z.Coelho, P.Collet Asymptotic limit law for the close approach of two trajectories in expanding maps of the circle. Probab. Theory Related Fields **99**, 237-250 (1994).

[33] N.Chernov, R.Markarian, S.Troubetzkoy. Conditionally invariant measures for Anosov maps with small holes. Ergodic Theory Dynam. Systems **18**, 1049-1073 (1998). Invariant measures for Anosov maps with small holes. Ergodic Theory Dynam. Systems **20**, 1007-1044 (2000).

[34] Z.Coelho, W.Parry. Central limit asymptotics for shifts of finite type. Israel J. Math. **69** 235-249 (1990).

[35] P.Collet. Ergodic properties of maps of the interval. In *Dynamical Systems*. R.Bamon, J.-M.Gambaudo & S.Martínez éditeurs, Hermann, Paris 1996.

[36] P.Collet. Statistics of closest returns for some non uniformly hyperbolic systems. Ergodic Theory & Dynamical Systems, **21**, 401-420 (2001).

[37] P.Collet, J.-P.Eckmann. *Interated Maps on the Interval as Dynamical Systems*. Birkhäuser, Basel Boston Stuttgart, 1980.

[38] P.Collet, J.-P.Eckmann. Oscillations of Observables in 1-Dimensional Lattice Systems. Math. Phys. Elec. Journ. **3**, 1-19 (1997).

[39] P.Collet, A.Galves, B.Schmitt.Fluctuations of Repetition Times for Gibbsian Sources. Nonlinearity **12**, 1225-1237 (1999).

[40] 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. Poincaré **57**, 319-331 (1992).

[41] P.Collet, A.Galves. Asymptotic distribution of entrance times for expanding maps of the interval. In *Dynamical Systems and Applications*. R.P.Agarwal editor, World Scientific 1995.

[42] P.Collet, A.Galves. Statistics of close visits to the indifferent fixed point of an interval map. J. Stat. Phys. **72**, 459-478 (1993).

[43] P.Collet, S.Isola. Essential spectrum in C^k of expanding Markoff maps of the interval. Commun. Math. Phys. **139**, 551 (1991).

[44] P.Collet, J.Lebowitz, A.Porzio. The dimension spectrum of some dynamical systems. J. Stat. Phys. **47** 609-644 (1987).

[45] P.Collet, S.Martínez, V.Maume. On the existence of conditionally invariant probability measures in dynamical systems. Nonlinearity **13**, 1263-1274 (2000). Nonlinearity **17**, 1985-1987 (2004).

[46] P.Collet, S.Martínez, B.Schmitt. Asymptotic distribution of tests for expanding maps of the interval. Ergodic Theory Dynam. Systems **24**, 707-722 (2004).

[47] P.Collet, S.Martínez, B.Schmitt. Exponential inequalities for dynamical measures of expanding maps of the interval. Probab. Theory Related Fields **123**, 301-322 (2002).

[48] P.Collet, S.Martínez and B.Schmitt. The Yorke-Pianigiani measure and the asymptotic law on the limit Cantor set of expanding systems Nonlinearity **7**, 1437-1443 (1994). The Pianigiani-Yorke Measure for Topological Markov Chains. Isr. Journal of Math. **97**, 61-71 (1997).

[49] P.collet, S.Martínez. Diffusion coefficient in transient chaos. Nonlinearity **12**, 445-450 (1999).

[50] P.Collet, J.-P.Eckmann. Lyapunov multipliers and decay of correlations in dynamical systems. Journal Stat. Phys. **115**, 217-254 (2004).

[51] P.collet, S.Martínez and B.Schmitt. On the Enhancement of Diffusion by Chaos, Escape Rates and Stochastic Instability. Trans. Amer. Math. Soc. **351**, 2875-2897 (1999).

[52] C.Cutler. A theory of correlation dimension for stationary time series. Philos. Trans. Roy. Soc. London Ser. A **348**, 343-355 (1994).

[53] A.Dembo, O.Zeitouni Moderate deviations for iterates of expanding maps. In *Statistics and control of stochastic processes*, 1-11, World Sci. Publishing, River Edge, NJ, 1997.

[54] M.Denker.The central limit theorem for dynamical systems. In *Dynamical systems and ergodic theory*. Banach Center Publ., **23**, 33-62 PWN, Warsaw, 1989.

[55] M.Denker, M.Gordin, A.Sharova. A Poisson limit theorem for toral automorphisms. Illinois J. Math. **48**, 1-20 (2004).

[56] R.Devanney. *An Introduction to Chaotic Dynamical Systems*. Addison-Wesley 1989.

[57] L. Devroye, *Exponential inequalities in nonparametric estimation*. Nonparametric functional estimation and related topics (Spetses, 1990), 31-44, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 335, Kluwer Acad. Publ., Dordrecht, 1991.

[58] T.Donarowicz, B.Weiss. Entropy theorems along the time when x visits a set. Illinois Journ. Math. **48**, 59-69 (2004).

[59] P.Doukhan. *Mixing. Properties and examples*. Lecture Notes in Statistics, **85**. Springer-Verlag, New York, 1994.

[60] F.Durand, A.Maass. Limit laws of entrance times for low-complexity Cantor minimal systems. Nonlinearity **14**, 683-700 (2001).

[61] J.-P.Eckmann, I.Procaccia. Fluctuations of dynamical scaling indices in nonlinear systems. Phys. Rev. **A 34** 659-661 (1986).

[62] J.-P. Eckmann, S.Kamphorst, D.Ruelle and S.Ciliberto. Lyapunov exponents from time series. Phys. Rev. A, **34**, 4971-4979 (1985).

[63] J.-P. Eckmann and D.Ruelle. Ergodic theory of chaos and strange attractors. Rev. Mod. Phys. **57**, 617-656 (1985).

[64] R.S.Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer, Berlin 1985.

[65] K.Falconer. *The Geometry of Fractal Sets*. Cambridge University Press 1995. Fractal geometry. Mathematical foundations and applications. John Wiley & Sons, Ltd., Chichester, 1990.

[66] W.Feller. *An introduction to Probability Theory and its Applications I, II.* John Wiley & Sons, New York, 1966.

[67] J.Galambos. The asymptotic theory of extreme order statistics. Second edition. Robert E. Krieger Publishing Co., Inc., Melbourne, FL, 1987.

[68] K.Geist, U.Parlitz and W.Lauterborn. Comparison of different methods for computing Lyapunov exponents. Progress of Theoretical Physics, **83**, 875-893 (1990).

[69] J.Gibbson, D.Farmer, M.Casdagli and S.Eubank. An analytical approach to practical state space reconstruction. Physica D **57**, 1-30 (1992).

[70] I.Gikhman, A.Skorokhod. *Introduction to the theory of random processes.* Dover Publications, Inc., Mineola, NY, 1996.

[71] M.Gordin. The central limit theorem for stationary processes. (Russian) Dokl. Akad. Nauk SSSR **188**, 739-741 (1969). Homoclinic approach to the central limit theorem for dynamical systems. In *Doeblin and modern probability (Blaubeuren, 1991).* Contemp. Math. **149**, 149-162 Amer. Math. Soc., Providence, RI, 1993.

[72] S.Gouëzel. Berry-Esseen theorem and local limit theorem for non uniformly expanding maps Ann. Inst. H. Poincaré Prob. Stat. to appear.

[73] J.Guckenheimer, P.Holmes. *Nonlinear Oscillations, Dynamical Systems and Bifurcation of Vector Fields.* Springer 1983.

[74] M.Hénon. A two-dimensional mapping with a strange attractor. Comm. Math. Phys. **50**, 69-77 (1976) .

[75] M.Hirata. Poisson law for Axiom A diffeomorphisms. Ergodic Theory Dynamical Systems **13**, 533-556 (1993)

[76] M.Hirata, B.Saussol, S.Vaienti. Statistics of return times: a general framework and new applications. Comm. Math. Phys. **206**, 33-55 (1999).

[77] M.Hirsch. Differential topology. Graduate Texts in Mathematics, 33. Springer-Verlag, New York, 1994.

[78] M.Hirsch, C.Pugh, M.Shub. *Invariant manifolds.* Lecture Notes in Mathematics, Vol. 583. Springer-Verlag, Berlin-New York, 1977

[79] R.Horn, C.Johnson. *Topics in matrix analysis.* Cambridge University Press, Cambridge, 1991.

[80] F.Hofbauer, G.Keller. Ergodic properties of invariant measures for piecewise monotonic transformations. Math. Zeit. **180**, 119-140 (1982).

[81] V.Ivanov. Geometric properties of monotone fluctuations and probabilities of random fluctuations. Siberian Math. Journal **37**, 102-129 (1996). Oscillation of means in the ergodic theorem. Doklady Mathematics **53**, 263-265 (1996).

[82] D.Jones, E.Titi. Upper bounds on the number of determining modes, nodes, and volume elements for the Navier-Stokes equations. Indiana Univ. Math. J. **42**, 875-887 (1993).

[83] A.Kachurovskii. The rate of convergence in the ergodic theorems. Russian Math. Surveys **51**, 73-124 (1996).

[84] S.Kalikow. *Outline of Ergodic Theory.* `http://www.math.umd.edu/ djr/kalikow.html`

[85] S.Kalikov, B.Weiss. Fluctuations of the ergodic averages. Illinois J. Math. **43**, 480-488 (1999).

[86] A.Katok, B.Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press 1996.

[87] J.-P.Kahane. *Some random series of functions*. Cambridge University press, Cambridge 1985.

[88] T.Kato. Perturbation theory for linear operators. Classics in Mathematics. Springer-Verlag, Berlin, 1995.

[89] H.Kantz, T.Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press 1997.

[90] A.Katok. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. Inst. Hautes Etudes Sci. Publ. Math. **51**, 137-173 (1980).

[91] S.Kay, S.L.Marple. Spectrum Analysis-A Modern Perspective. Proceedings of the IEEE, **69**, 1380-1419 (1981).

[92] M.Keane. Ergodic theory and subshifts of finite type. In *Ergodic theory, symbolic dynamics, and hyperbolic spaces*, 35-70. Oxford Sci. Publ., Oxford Univ. Press, New York, 1991.

[93] G.Keller. A new estimator for information dimension with standard errors and confidence interval. Stoch. Proc. and Appl. **71**, 187-206 (1997).

[94] G.Keller, R.Sporer Remarks on the linear regression approach to dimension estimation. In *Stochastic and spatial structures of dynamical systems* (Amsterdam, 1995), 17-27, Konink. Nederl. Akad. Wetensch. Verh. Afd. Natuurk. Eerste Reeks, 45, North-Holland, Amsterdam, 1996.

[95] M.Kennel, H.Abarbanel. False neighbors and false strands: A reliable minimum embedding dimension algorithm. Phys. Rev. E **66**, 1-18 (2002).

[96] Knuth, Donald E. The art of computer programming. Vol. 2. Seminumerical algorithms. Addison-Wesley Publishing Co., Reading, Mass., 1981.

[97] I.Kontoyiannis Asymptotic recurrence and waiting times for stationary processes. J. Theoret. Probab. **11**, 795-811 (1998).

[98] I.Kornfeld́, Y.Sinai, S.Fomin. *Ergodic Theory*. Springer-Verlag New-York, 1980.

[99] U.Krengel. *Ergodic Theorems*. Walter de Gruyter, Berlin, New York 1985.

[100] S.Lalley, B.Nobel. Denoising deterministic time series. Preprint (2002).

[101] S.Lalley. Beneath the noise, chaos. Ann. Stat. **27**, 233-244 (2001).

[102] S.Lalley, A.Nobel. Indistinguishability of absolutely continuous and singular distributions. Statist. Probab. Lett. **62**, 145-154(2003).

[103] O.Lanford. Entropy and equilibrium states in classical statistical mechanics. In *Statistical mechanics and Mathematical Problems*. Lecture Notes in Physics **20**, 1-113, Springer-Verlag, Berlin 1973.

[104] Y.Lacroix. Possible limit laws for entrance times of an ergodic aperiodic dynamical system. Israel J. Math. **132**, 253-263 (2002).

[105] P. L'Ecuyer. Uniform Random Number Generation. Annals of Operations Research, **53**, 77-120 (1994). See also P. L'Ecuyer.*Random Number Generation* draft for a chapter of the forthcoming Handbook of Computational Statistics, J. E. Gentle, W. Haerdle, and Y. Mori, eds., Springer-Verlag, 2004.

[106] M.Ledoux. *The concentration of measure phenomenon*. Mathematical Surveys and Monographs, **89**. American Mathematical Society, Providence, RI, 2001.

[107] F.Ledrappier, M.Misiurewicz. Dimension of invariant measures for maps with exponent zero. Ergodic Theory Dynam. Systems **5**, 595-610 (1985).

[108] C.Liverani Decay of correlations. Ann. of Math. **142**, 239-301 (1995).

[109] A.Lopes, S.Lopes. Convergence in distribution of the periodogram of chaotic processes. Stoch. Dyn. **2**, 609-624 (2002). Parametric estimation and spectral analysis of piecewise linear maps of the interval. Adv. in Appl. Probab. **30**, 757-776 (1998).

[110] E.Lorenz. Deterministic Nonperiodic Flow. Journal of the Atmospheric Sciences. **20**, 130-141 (1963)

[111] S.Luzzatto. Stochastic-like behaviour in nonuniformly expanding maps. In: *Handbook of Dynamical Systems* **1B** (Hasselblat & Katok ed.), Elsevier, 2005.

[112] A.Manning, K.Simon. A short existence proof for correlation dimension. J. Stat. Phys. **90**, 1047-1049 (1998).

[113] L.Markus, K.Meyer. *Generic Hamiltonian dynamical systems are neither integrable nor ergodic*. Memoirs of the American Mathematical Society, **144**. American Mathematical Society, Providence, R.I., 1974.

[114] G.Marsaglia.The mathematics of random number generators. The unreasonable effectiveness of number theory (Orono, ME, 1991), 73-90, Proc. Sympos. Appl. Math., 46, Amer. Math. Soc., Providence, RI, 1992.

[115] P.Mattila. Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability. Cambridge Studies in Advanced Mathematics, 44. Cambridge University Press, Cambridge, 1995.

[116] I. Melbourne,M. Nicol. Almost sure invariance principle for nonuniformly hyperbolic systems. Commun. Math. Phys. **260**, 131-146 (2005).

[117] E.Olofsen, J.Degoede and R.Heijungs. A maximum likelihood approach to correlation dimension and entropy estimation. Bull. Math. Biol. **54**, 45-58 (1992).

[118] D.Ornstein. *Ergodic Theory, Randomness, and Dynamical Systems*. Yale University Press, New Haven London 1974.

[119] D.Ornstein, B.Weiss. Entropy and data compression schemes. stationary random fields. IEEE Trans. Information Theory **39**, 78-83 (1993).

[120] E.Ott. *Chaos in Dynamical Systems*. Cambridge University Press 1993.

[121] E. Ott, T.Sauer, J.Yorke. *Coping With Chaos*. Wiley 1994.

[122] J.Pesin. Families of invariant manifolds that correspond to nonzero characteristic exponents. Izv. Akad. Nauk SSSR Ser. Mat. **40**, 1332-1379 (1976).

[123] K.Petersen. *Ergodic Theory*. Cambridge University Press, Cambridge 1983. See also `http://www.math.unc.edu/Faculty/petersen`. Easy and nearly simultaneous proofs of the ergodic theorem and maximal ergodic theorem. `arXiv:math.DS/0004070`

[124] W.Philipp, W.Stout. *Almost sure invariance principles for partial sums of weakly dependent random variables*. Memoirs of the AMS, **161**, 1975.

[125] D.Plachky, J.Steinebach. A theorem about probabilities of large deviations with an application to queuing theory. Periodica Mathematica **6**, 343-345 (1975).

[126] C.Prieur. Density estimation for one-dimensional dynamical systems. ESAIM Probab. Statist. **5**, 51-76 (2001).

[127] D.Rand. The singularity spectrum $f(\alpha)$ for cookie-cutters. Ergodic Theory Dynam. Systems **9**, 527-541 (1989).

[128] K.Ramanan, O.Zeitouni. The quasi-stationary distribution for small random perturbations of certain one-dimensional maps. Stochastic Process. Appl. **84**, 25-51 (1999).

[129] D.Ruelle. *Elements of differentiable dynamics and bifurcation theory*. Academic Press, Inc., Boston, MA, 1989.

[130] D.Ruelle. *Chaotic Evolution and Strange Attractors: The Statistical Analysis of Time Series for Deterministic Nonlinear Systems*. Cambridge University Press 1989.

[131] D.Ruelle. *Thermodynamic formalism*. Addison-Wesley, Reading, 1978.

[132] D.Ruelle. Ergodic theory of differentiable dynamical systems. Inst. Hautes Etudes Sci. Publ. Math. **50**, 27-58 (1979).

[133] K.Reinhold. A smoother ergodic average. Illinois Journ. Math. **44**, 843-859 (2000).

[134] T.Sauer, J.Yorke, M.Casdagli. Embedology. J. Stat. Phys. **65**, 579-616 (1991).

[135] T.Schreiber. Interdisciplinary applications of nonlinear time series analysis. Phys. Rep. **308** 1-64 (1999).

[136] R.Serinko. Ergodic theorems arising in correlation dimension estimation. J. Stat. Phys. **85**, 25-40 (1996).

[137] C.Shannon. A mathematical theory of communication. Bell System Technical Journal, **27**, 379-423 and 623-656 (1948). available at `http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html`

[138] Y.Sinai. *Introduction to Ergodic Theory*. Princeton University Press, Princeton 1976.

[139] J.Stark. Analysis of time series. In Modeling Uncertainty, University of Cambridge Program for Industry, 1994.

[140] J.Schouten, F.Takens and C. van den Bleek. Estimation of the dimension of noisy attractor. Phys. Rev. E **50**, 1851-1861 (1994).

[141] J.Schouten, F.Takens and C. van den Bleek. Maximum likelihood estimation of the entropy of an attractor. Phys. Rev. E **49**, 126-129 (1994).

[142] F.Takens. Detecting strange attractors in turbulence. In *Dynamical Systems in Turbulence*. D.Rand and L.S. Young editors. lecture Notes in Mathematics **898**, Springer 1981.

[143] F.Takens, E.Verbitski. Generalized entropies: Rényi and correlation integral approach. Nonlinearity **11**, 771-782 (1998).

[144] F.Takens. Estimation of dimension and order of time series. In *Nonlinear dynamical systems and chaos* (Groningen, 1995), 405-422, Progr. Nonlinear Differential Equations Appl., **19**, Birkhäuser, Basel, 1996.

[145] F.Takens. The analysis of correlation integrals in terms of extremal value theory. Boletim Soc. Brasileira Math. **29**, 197-228 (1998).

[146] M.Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Etudes Sci. Publ. Math. **81**, 73-205 (1995).

[147] R.Temam. *Infinite-dimensional dynamical systems in mechanics and physics*. Applied Mathematical Sciences, **68**. Springer-Verlag, New York, 1997.

[148] J.Theller. Estimating fractal dimension. J. Opt. Soc. Am. **A7**, 1055-1073 (1990).

[149] TISEAN available at `http://www.mpipks-dresden.mpg.de/~tisean`

[150] TSTOOL available at `http://www.physik3.gwg.de/tstool`

[151] J.von-Neumann. Proof of the quasi ergodic hypothesis. Proc. Nat. Acad. Sci. USA, **18**, 70-82 (1932).

[152] Vuhl E.Vul, Ya.Sinaĭ, K.Khanin. Feigenbaum universality and thermodynamic formalism. Uspekhi Mat. Nauk **39**, 3-37 (1984).

[153] L.S.Young. Recurrence times and rates of mixing. Israel J. Math. **110**, 153-188 (1999).

[154] L.S. Young. What are SRB measures and which dynamical systems have them? J. Stat. Phys. **108** 733-754 (2002).

[155] Statistical properties of dynamical systems with some hyperbolicity. Ann. of Math. **147**, 585-650 (1998).

[156] L.-S.Young. Stochastic stability of hyperbolic attractors. Ergodic Theory Dynam. Systems **6**, 311-319 (1986).

[157] L.-S.Young. Large deviations in dynamical systems. Trans. Amer. Math. Soc. **318**, 525-543 (1990).

[158] K.Ziemian. Almost sure invariance principle for some maps of an interval. Ergodic Theory Dynam. Systems **5**, 625-640 (1985).