On concentration, noise and entropy estimation in dynamical systems

Cesar Maldonado

Centre de Physique Théorique

THÈSE

en vue d'obtenir le titre de

Docteur

de l'École Polytechnique
Spécialité : Physique Théorique

par

Cesar MALDONADO

On concentration, noise and entropy estimation in dynamical systems

---

Directeur de Thèse: Jean-René Chazottes

Soutenue publiquement le 21 septembre 2012

Jury :

Rapporteurs :	1. Benoit Saussol	-	Université de Brest (PR)
	2. Sandro Vaienti	-	CPT, Université de Marseille (PR)
Examinateurs :	3. Stefano Galatolo	-	DMA Università di Pisa (Italy)
	4. Frank Redig	-	Universiteit Leiden (The Netherlands) (PR)
	5. Edgardo Ugalde	-	Instituto de Física, UASLP (Mexico) (PR)
Directeur :	6. Jean-René Chazottes	-	CPhT, École Polytechnique

Contents

Esta dedicatoria te espera al igual que tus padres...

Con mucho amor.

Acknowlegdements

First of all, I warmly thank my advisor Jean-René Chazottes for his patience. During this three years he carefully explained to me everything I didn't understand, which means, too many things. It has been a pleasure to work under his supervision.

I would like to express my gratitude to professors Benoît Saussol and Sandro Vaienti for accepting to be the *rapporteurs* of this thesis, also to professors Stefano Galatolo, Frank Redig and Edgardo Ugalde who accepted to be part of the jury.

I would like to thank Pierre Collet for discussions and comments, specially for those suggestions and ideas he gave me for my work on observational noise and the 'denoising' algorithm.

Another special mention goes to Stefano Galatolo for his warm hospitality at the *Dipartimento di Matematica Applicata* at *Università di Pisa*. I spent six wonderful weeks in Italy and much of that is due to Stefano's suggestions.

I am thankful to Jason Stover for his helpful comments on the denoising algorithm in which he worked during his PhD. Also I specially thank Marc Monticelli for the figures 5.1 and 5.2. As well, he incredibly improved an inefficient code I wrote for the denoising, and now is included in the xDim framework.

Since in more than one occasion helped me to solve the bureaucratic issues, I am grateful to all the people at the secretary of CPhT and at l'École doctorale.

I am indebted to Edgardo Ugalde, Gelasio Salazar, Valentin Afraimovich and Ricardo Guirado for their kind letters of recommendation. Without them, I could have never got the Monge scholarship.

Thanks to Ariel Grundwald, David Benzaquen, Eliahou Bouhanna and Sarah Llaneza.

Por último pero no por ello menos importante, unas líneas en español para que los más cercanos a mí me lean. Agradezco la compañia de mis amigos en Polytechnique, *Poncho*, *Rosa*, y al *el Jimmy*. Agradezco a Raúl Salgado quien me ayudo mucho al principio de mi estancia en Francia. Agradezco también a mis amigos Rigo Chavira, Pedro García, Chendo García, Omar Rico y Nehemías *Springfield* por su apoyo aún estando lejos.

Infinitas gracias a toda mi familia, su apoyo a lo largo del doctorado me fue indispensable, en especial a mis padres Alfredo y Blanca, a mis hermanos Alfredo y Fabián y sus respectivas familias. También a mis ahora, tres sobrinos, pero sobre todo al simio.

Y para terminar, quiero agradecer con todo mi corazón a Adriana Aguilar Hervert (léase, *un gusano*) por la confianza, la paciencia y el amor que me otorgó durante todo este tiempo, sobre todo durante la redacción de esta tésis ya que con su compañia hizo que todo el proceso fuera más dulce y llevadero. Adriana gracias por hacer que esto sea posible.

Abstract

This thesis is divided into three parts. In the first part we briefly describe the class of dynamical systems considered. We also give some known results on the study of fluctuations of observables in dynamical systems such as the central limit theorem, large deviations and concentration inequalities.

In the second part we study dynamical systems perturbed by observational noise. We prove that if a dynamical system satisfies a concentration inequality then the system with observational noise also satisfies a concentration inequality. We apply these inequalities to obtain fluctuation bounds for the auto-covariance function, the empirical measure, the kernel density estimator and the correlation dimension. Next, we study the work of S. Lalley on the problem of signal recovery. Given a time series of a chaotic dynamical system with observational noise, one can effectively eliminate the noise in average by using Lalley's algorithm. A chapter of this thesis is devoted to the proof of consistency of that algorithm. We end up the second part with a numerical quest for the best parameters of Lalley's algorithm.

The third part is devoted to entropy estimation in one-dimensional Gibbs measures. We study the fluctuations of two entropy estimators. The first one is based on the empirical frequencies of observation of typical blocks. The second is based on the time a typical orbit takes to hit an independent typical block. We apply concentration inequalities to obtain bounds on the fluctuation of these estimators.

Résumé

Cette thèse est divisée en trois parties. Dans la prèmiere partie nous décrivons les systèmes dynamiques que l'on considère tout au long de la thèse. Nous donnons aussi des résultats connus sur les fluctuations d'observables dans les systèmes dynamiques tels comme la théorème central limite, les grands déviations et les inégalités de concentration.

La deuxième partie de cette thèse est consacrée aux systèmes dynamiques perturbés par un bruit observationnel. Nous démontrons que si un système dynamique satisfait une inégalité de concentration alors le système perturbé satisfait lui aussi une inégalité de concentration adéquate. Ensuite nous appliquons ces inégalités pour obtenir des bornes sur la taille des fluctuations d'observables bruitées. Nous considérons comme observables la fonction d'auto-corrélation, la mesure empirique, l'estimateur à noyau de la densité de la mesure invariante et la dimension de corrélation. Nous étudions ensuite les travaux de S. Lalley sur le problème de débruitage d'une série temporelle. Etant donné une série temporelle générée par un système dynamique chaotique bruité, il est effectivement possible d'éliminer le bruit en moyenne en utilissant l'algorithme de Lalley. Un chapitre de cette thèse est consacré à la preuve de ce théorème. Nous finissons la deuxième partie avec une quête numérique pour les meilleurs paramètres de l'algorithme de Lalley.

Dans la troisième partie, nous étudions le problème de l'estimation de l'entropie pour des mesures de Gibbs unidimensionnelles. Nous étudions les propriétés de deux estimateurs de l'entropie. Le premier est basé sur les fréquences des blocs typiques observés. Le second est basé sur les temps d'apparition de blocs typiques. Nous appliquons des inégalités de concentrations pour obtenir un contrôle sur les fluctuations de ces estimateurs.

Part I

Generalities

Chapter 1

Considered systems

A dynamical system is defined by a set of 'states' called phase space that evolves in time. In this thesis we will consider as phase space a compact metric space X (usually $X \subset \mathbb{R}^d$). The passage of time is modeled by the successive iteration of a self map $T : X \to X$. Given an initial condition $x_0 \in X$, its orbit is the sequence $x_0, x_1 = Tx_0, x_2 = Tx_1 = T^2x_0, \ldots$, where, as usual, T^k denotes the k-fold composition of T with itself.

We are interested in the probabilistic properties of dynamical systems. The probability space is $(X, \mathscr{B}, \mu)$, where $\mathscr{B}$ is the Borel sigma-algebra. The probability measures of interest are those preserved under the transformation T. Their existence is assured by Kryloff-Bogoliuboff's theorem (see for instance [71]), defining a measure-preserving dynamical system (X, T, μ). The invariant measure μ is said to be ergodic with respect to T if for every measurable subset A satisfying $T^{-1}(A) = A$ then either $\mu(A) = 0$ or $\mu(A) = 1$. In the study of probabilistic properties of dynamical systems, an important issue is the description of the 'time averages' of functions $f : X \to \mathbb{R}$ called observables. A fundamental result is the Birkhoff's ergodic theorem.

Theorem 1.0.1 (Birkhoff's ergodic theorem)**.** *Let (X, T, μ) be a measure-preserving dynamical system. For any $f \in L^1(\mu)$, the limit*

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

exists μ-almost everywhere and in $L^1(\mu)$. Furthermore if μ is ergodic, then for any $f \in L^1(\mu)$,

$$\int \tilde{f} \mathrm{d}\mu = \int f \mathrm{d}\mu \quad \textit{and} \quad \tilde{f} \circ T = \tilde{f} \quad \mu\textit{-a.e.}$$

An idea that will be used repeatedly in this thesis is that one may interpret the orbits $(x, Tx, \ldots)$ as realizations of the stationary stochastic process defined by $X_n(x) = T^n x$. The finite-dimensional marginals of this process are the measures μ_n given by

$$d\mu_n(x_0, \ldots, x_{n-1}) = d\mu(x_0) \prod_{i=1}^{n-1} \delta_{x_i = Tx_{i-1}}. \tag{1.1}$$

Therefore, the stochasticity comes only from the initial condition. In view of this interpretation it is useful to write the previous theorem for an integrable stationary ergodic process (X_n), that is, $\frac{1}{n}\sum_{i=0}^{n-1} X_i \to \mathbb{E}[X_0]$ almost surely.

Next, the measure μ is said to be mixing with respect to T if for all measurable sets A, B,

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

When the system (X, T, μ) is sufficiently mixing, one may expect that the iterate $T^k x$ is more or less independent of x if k is large enough. In turn, the process $(X_n = T^n(\cdot))$ may behave like an independent process.

From the practical point of view one is interested in properties that can be 'observed', they are those that hold on positive Lebesgue measure sets. Invariant measures having densities are an example of this 'physically relevant' measures. Very often we will refer to the Hénon map as well as the Lozi map, these are examples of systems with a strange (or chaotic) attractor Λ which has zero Lebesgue measure. For maps having an attractor, all their invariant measures must be supported on Λ. Here is where the idea of Sinai-Ruelle-Bowen (SRB) measures comes into play, they are invariant measures compatible with Lebesgue when Lebesgue measure is not preserved. We refer to [75] for a survey on SRB measures, and [41] for an account on the development of the theory of chaotic attractors and their invariant measures.

1.1 Uniformly Hyperbolic dynamical systems

For later convenience we introduce the concept of hyperbolic dynamical systems. A straightforward way to model the sensitive dependence on initial conditions is by using the uniformly expanding property. Consider a differentiable map T on a compact metric space X. Let $C > 0$ and $\lambda > 1$ be constants, such that for all $x \in X$ and v in the tangent space at x and for all $n \in \mathbb{N}$

$$\|DT^n(x)v\| \geq C\lambda^n \|v\|.$$

Uniformly hyperbolic maps have the property that at each point x the tangent space is a direct sum of two subspaces E^u_x and E^s_x, one of them is expanded: $\|DT^n(x)v\| \geq C\lambda^n\|v\|$ for every $v \in E^u_x$ and is called unstable, while the other is contracted: $\|DT^n(x)v\| \leq C\lambda^{-n}\|v\|$ for every $v \in E^s_x$.

Example 1.1.1. *A famous example mapping $[0, 1)^2$ into it self, is the cat map. It is given by*

$$T(x, y) = (2x + y \pmod 1, x + y \pmod 1).$$

This map is area preserving.

Example 1.1.2. *Consider Smale's solenoid map, $T_S : \mathbb{R}^3 \to \mathbb{R}^3$ which maps the torus into itself:*

$$T_S(\phi, u, v) = (2\phi \mod 2\pi, \beta u + \alpha \cos(\phi), \beta v + \alpha \sin(\phi)),$$

where $0 < \beta < 1/2$ and $\beta < \alpha < 1/2$.

Example 1.1.3. *Take $[0,1]$ as state space. Fix a sequence $0 = a_0 < a_1 < \cdots < a_k = 1$, and consider for each interval (a_j, a_{j+1}) ($0 \leq j \leq k-1$) a monotone map $T_j : (a_j, a_{j+1}) \to [0,1]$. The map T on $[0,1]$ is given by $T(x) = T_j(x)$ if $x \in (a_j, a_{j+1})$. It is well known that when the map T is uniformly expanding, it admits an absolutely continuous invariant measure μ. It is unique under some mixing assumptions.*

1.2 Gibbs measures

A special case of uniformly hyperbolic systems are the Axiom A diffeomorphisms. For these systems it is possible to construct a Markov partition and use symbolic dynamics. Gibbs measures play a major role in the ergodic theory of Axiom A diffeomorphisms. In this section we briefly describe Gibbs measures for later convenience. For the complete details we remit the reader to [12].

We consider the set $\Omega = A^{\mathbb{N}}$ of infinite sequences $\underline{x}$ of symbols from the finite set A: $\underline{x} = x_0, x_1, \ldots$ where $x_j \in A$. We denote by σ the shift map on Ω: $(\sigma\underline{x})_i = x_{i+1}$, for all $i = 0, 1, \ldots$. The space (Ω, σ) is called the full-shift.
We equip Ω with the usual distance: fix $\theta \in (0,1)$ and for $\underline{x} \neq \underline{y}$, let $d_\theta(\underline{x}, \underline{y}) = \theta^N$ where N is the largest nonnegative integer with $x_i = y_i$ for every $0 \leq i < N$. (By convention, if $\underline{x} = \underline{y}$ then $N = \infty$ and $\theta^\infty = 0$, while if $x_0 \neq y_0$ then $N = 0$.) With this distance, Ω is a compact metric space.
For a given string $a_0^{k-1} = a_0, \ldots, a_{k-1}$ ($a_i \in A$), the set $[a_0^{k-1}] = \{\underline{x} \in \Omega : x_i = a_i, i = 1, \ldots, k-1\}$ is the cylinder of length k based on $a_0, \ldots, a_{k-1}$.
For a continuous function $f : \Omega \to \mathbb{R}$ and $m \geq 0$ we define

$$\mathrm{var}_m(f) := \sup\{|f(\underline{x}) - f(\underline{y})| : x_i = y_i,\ i = 0, \ldots, m\}\cdot$$

It is easy to see that $|f(\underline{x}) - f(\underline{y})| \leq C d_\theta(\underline{x}, \underline{y})$ if and only if $\mathrm{var}_m(f) \leq C\theta^m$, $m = 0, 1, \ldots$. Let

$$\mathscr{F}_\theta = \big\{f : f \text{ continuous},\ \mathrm{var}_m(f) \leq C\theta^m,\ m = 0, 1, \ldots,\ \text{for some } C > 0\big\}.$$

This is the space of Lipschitz functions with respect to the distance d_θ. For $f \in \mathscr{F}_\theta$ let $|f|_\theta = \sup\left\{\frac{\mathrm{var}_m(f)}{\theta^m} : m \geq 0\right\}$. We notice that $|f|_\theta$ is merely the least Lipschitz constant of f. Together with $\|f\|_\infty = \sup\{|f(\underline{x})| : \underline{x} \in \Omega\}$, this defines a norm on $\mathscr{F}_\theta$ by $\|f\|_\theta = \|f\|_\infty + |f|_\theta$.

Theorem 1.2.1 ([12])**.** *Consider the full-shift (Ω, σ) and let $\phi \in \mathscr{F}_\theta$. There exists a unique σ-invariant probability measure μ_ϕ on Ω for which one can find constants*

$C = C(\phi) > 1$ and $P = P(\phi)$ such that

$$C^{-1} \leq \frac{\mu_\phi\{\underline{y} : y_i = x_i,\ \forall i \in [0, m)\}}{\exp\left(-Pm + \sum_{k=0}^{m-1} \phi(\sigma^k \underline{x})\right)} \leq C.$$

for every $\underline{x} \in \Omega$ and $m \geq 1$. This measure is called the Gibbs measure of ϕ.

The constant P is the topological pressure of ϕ. There are various equivalent definitions of P. For instance,

$$P = \lim_{n\to\infty} \frac{1}{n} \log \sum_{a_0,\dots,a_{n-1}} e^{\sum_{k=0}^{n-1} \phi(\sigma^k \underline{x}^*)}$$

where, for each block $(a_0, \dots, a_{n-1}) \in A^n$, an arbitrary choice of $\underline{x}^* \in \Omega$ has been made such that $\underline{x}^* \in [a_0^{n-1}]$. We can always assume that $P = 0$ by considering the potential $\phi - P$ which yields the same Gibbs measure.

Next, for every continuous function ϕ in the full-shift, define the transfer operator $\mathscr{L}_\phi$ on the continuous functions, by

$$(\mathscr{L}_\phi f)(\underline{x}) = \sum_{\underline{y} \in \sigma^{-1}\underline{x}} e^{\phi(\underline{y})} f(\underline{y}).$$

The Ruelle's Perron-Frobenius theorem ([12]) states that, for $\phi \in \mathscr{F}_\theta$ there are $\lambda > 0$, a continuous function h with $h > 0$ and ν for wich $\mathscr{L}_\phi h = \lambda h$, $\mathscr{L}^* \nu = \lambda \nu$, $\int h \mathrm{d}\nu = 1$ and

$$\lim_{n\to\infty} \left\| \lambda^{-n} \mathscr{L}^n g - h \int g \mathrm{d}\nu \right\| = 0$$

for all continuous function g on the full-shift.

The Gibbs measure μ_ϕ satisfies the variational principle, namely

$$\sup \left\{ h(\eta) + \int \phi \mathrm{d}\eta : \eta \text{ shift-invariant} \right\} = h(\mu_\phi) + \int \phi \mathrm{d}\mu_\phi = P = 0.$$

More precisely, μ_ϕ is the unique shift-invariant measure reaching this supremum. In particular we have

$$h(\mu_\phi) = -\int \phi \mathrm{d}\mu_\phi. \tag{1.2}$$

1.3 Young towers

We are interested in a particular but large class of non-uniformly hyperbolic dynamical systems. Non-uniformity can be understood as follows. As before, we have at each $x \in X$ the splitting of the tangent space $T_x M$ into E^u_x and E^s_x. For almost every point $x \in X$ there exists constants $\lambda_1(x), \lambda_2 > 1$ and $C(x) > 0$ such that, for all $n \in \mathbb{N}$

$$\|DT^n(x) \cdot v\| \geq C(x)\lambda_1(x)^n \|v\| \quad \text{for every} \quad v \in E^u_x \quad \text{and}$$
$$\|DT^{-n}(x) \cdot v\| \leq C(x)\lambda_2(x)^{-n} \|v\| \quad \text{for every} \quad v \in E^s_x.$$

This means that, wherever the system presents expansiveness, the rate of expansiveness depends on the point. Even in the case when the set of points of 'no expansion' is of measure zero, one may in general loose the existence of the spectral gap.

In [73, 74], L.-S. Young developed a general scheme to study the probabilistic properties of a class of 'predominantly hyperbolic' dynamical systems. We briefly describe the construction of the towers.

Consider a map $T : X \to X$. The tower structure is constructed over a $Y \subset X$ called 'base', having $\mathrm{Leb}(Y) > 0$, and in which T is uniformly hyperbolic. For each $y \in Y$ let us define its return time into Y,

$$R(y) := \inf\{n \geq 1 \ : \ T^n y \in Y\}.$$

The function $R : Y \to \mathbb{Z}^+$ gives us the degree of non-uniformity, and it is everywhere well defined because of Poincaré's recurrence theorem. Define the map T_Y on the part of Y where R is finite,

$$T_Y(y) := T^{R(y)}(y).$$

From (Y, T^R) one constructs an extension $(\mathcal{Y}, F)$ which is called a Young tower. One can visualize a tower by writing that $\mathcal{Y} = \bigcup_{k=0}^{\infty} \mathcal{Y}_k$ where $\mathcal{Y}_k$ can be identified with the set $\{x \in Y \ : \ R(x) > k\}$, that is, the k-th floor of the tower. In particular $\mathcal{Y}_0$ is identified with Y. The dynamics is as follows: for each point $x \in \mathcal{Y}_0$ moves up the tower at each iteration of F until it reaches the top level after which it returns to the base. Moreover, F has a Markov partition $\{\mathcal{Y}_{0,j}\}$, if we let $R_j := R|_{\mathcal{Y}_{0,j}}$, the set $\mathcal{Y}_{R_j-1,j}$ is the top level of the tower above $\mathcal{Y}_{0,j}$. We assume for the sake of simplicity that $\gcd\{R_j\} = 1$. Let Leb^u denote the Lebesgue measure on the unstable direction.

Definition 1.3.1. *A non-uniformly hyperbolic dynamical system (X, T) is modeled by a Young tower, with exponential tails if there exists $\theta > 0$ such that*

$$\mathrm{Leb}^u(\{y \in Y : \ R(y) > n\}) = \mathscr{O}(e^{-\theta n}),$$

or with polynomial tails if there exists a real $\alpha > 0$ such that

$$\mathrm{Leb}^u(\{y \in Y : \ R(y) > n\}) = \mathscr{O}(n^{-\alpha}).$$

Some probabilistic properties of T are captured by the tails properties of R. The result of Young is the following

Theorem 1.3.1 ([73, 74])**.** *If a non-uniformly hyperbolic dynamical system is modeled by a Young tower with summable return time, i.e., if*

$$\int_Y R \ \mathrm{dLeb}^u = \sum_{n=1}^{\infty} \mathrm{Leb}^u(\{y \in Y : \ R(y) > n\}) < \infty,$$

then the system admits a SRB measure.

We include some examples of systems that can be modeled by Young towers.

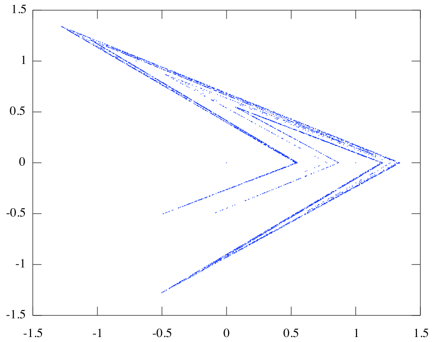


Figure 1.1: Simulation of the Lozi attractor for the parameters a=1.7 and b=0.5.

Example 1.3.1. *The Lozi map $T_L : \mathbb{R}^2 \to \mathbb{R}^2$ is given by*

$$T_L(u, v) = (1 - a|u| + v, bu), \qquad (u, v) \in \mathbb{R}^2.$$

For $a = 1.7$ and $b = 0.5$ one observes numerically an strange attractor. In [26] the authors constructed a SRB measure μ for this map. This map is included in the exponential case of Young's framework [73].

Example 1.3.2. *Another example of a system modeled by Young towers with exponential tails is the Hénon map $T_H : \mathbb{R}^2 \to \mathbb{R}^2$. It is defined by*

$$T_H(u, v) = \left(1 - au^2 + v, bu\right), \qquad (u, v) \in \mathbb{R}^2.$$

Where $0 < a < 2$ and $b > 0$ are some real parameters. It is known that there exists a set of parameters (a, b) of positive Lebesgue measure for which the map T_H has a topologically transitive attractor Λ, furthermore there exists a set $\Delta \subset \mathbb{R}^2$ with $\mathrm{Leb}(\Delta) > 0$ such that for all $(a, b) \in \Delta$ the map T_H admits a unique SRB measure supported on Λ ([8]).

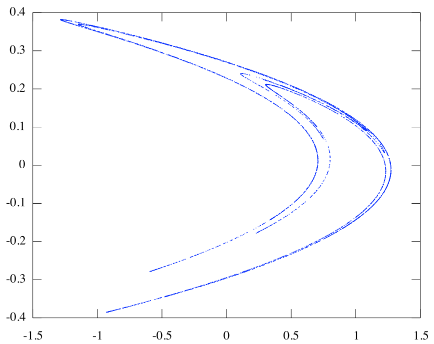


Figure 1.2: Simulation of the Hénon map for parameters a = 1.4, and b =0.3.

Example 1.3.3. *The Manneville-Pomeau map, is an example of non-uniformly hyperbolic dynamical system with polynomial tails. It is an expansive map in the interval except in zero where the slope is equal to 1 (i.e. zero is a neutral fixed point). Consider* $X = [0, 1]$*, the map is defined as follows,*

$$T_\alpha(x) = \begin{cases} x + 2^\alpha x^{1+\alpha} & \textit{if } \ x \in [0, 1/2) \\ 2x - 1 & \textit{if } \ x \in [1/2, 1), \end{cases}$$

where $\alpha \in (0, 1)$ *is a parameter. The greater the parameter, the larger the time takes to move away from the fixed point. Observe when* $\alpha = 0$*, the map is* $T_{\alpha=0}(x) = 2x \pmod 1$*, which is uniformly expanding. It is well known that there exists an absolutely continuous invariant measure* $d\mu(x) = h(x)dx$ *and* $h(x) \sim x^{-\alpha}$ *when* $x \to 0$*.*

Chapter 2

Fluctuations of observables in dynamical systems

Once we know that our dynamical system (X, T) admits a SRB measure μ, we may ask for its probabilistic properties. Given an observable $f : X \to \mathbb{R}$, let us consider the random variable $(X_k = f \circ T^k)$ on the probabilistic space (X, μ). Consider the ergodic sum $S_n f(x) = f(x) + f(Tx) + \cdots + f(T^{n-1}x)$ which is the partial sum of the process $(X_n; n \geq 0)$. Several probabilistic properties concern us, first of all, we are interested in determining the typical size of the fluctuations of $\frac{1}{n} S_n f(x)$ about $\int f \mathrm{d}\mu$. If the order of typical size of $S_n f - n \int f \mathrm{d}\mu$ is $\mathscr{O}(\sqrt{n})$, then the observable f is said to satisfy a central limit theorem with respect to the measure μ (or equivalently, with respect to the system (X, T, μ)). Secondly, we are interested in estimating the probability of deviation of the ergodic average of f from $\int f \mathrm{d}\mu$ up to some prescribed value t. More precisely, we would like to know the speed of convergence to zero of the following probability

$$\mu \left\{ x : \left| \frac{1}{n} S_n f(x) - \int f \mathrm{d}\mu \right| > t \right\},$$

for all $t > 0$ and for a large class of continuous observables f. In the probabilistic terminology this is stated as the *convergence in probability* of the ergodic averages towards its limit. If we expect that sufficiently chaotic dynamical systems behave as i.i.d. process, then that speed of convergence might be exponential. One also could have a convergence to some non-Gaussian distribution by scaling with some function different from $\sqrt{n}$.

From the point of view of applications it is important to determine the deviation probability of estimators of some quantities describing properties of the dynamical system. These quantities have the characteristic that they can be estimated using a single orbit. Generally the corresponding estimators are more complicated than ergodic sums. We consider general observables $K : X^n \to \mathbb{R}$. Evaluating K along an orbit up to the $n-1$ time step, one may ask if it is possible to find a positive function $b(n, t) < 1$ such that

$$\mu\left\{x \in X : \left|K(x,\dots,T^{n-1}x) - \int K(y,\dots,T^{n-1}y)\mathrm{d}\mu(y)\right| > t\right\} \leq b(n,t),$$

for all $n, t > 0$, with $b(n,t)$ depending on K. If $b(n,t)$ decreases rapidly with t and n, it means that $K(x, Tx, \dots, T^{n-1}x)$ concentrates around its expected value. As we shall see later on, one can achieve sufficiently fast decreasing functions $b(n,t)$, provided the function K is Lipschitz in each variable. This is the subject of concentration inequalities.

2.1 Central limit theorem

Given a dynamical system (X, T, μ) and a square integrable observable $f : X \to \mathbb{R}$, let us introduce the auto-covariance of the process $(f \circ T^k)$ by

$$\mathrm{Cov}_f(k) := \int f \cdot f \circ T^k \mathrm{d}\mu - \left(\int f \mathrm{d}\mu\right)^2,$$

for every $k \in \mathbb{N}_0$. More generally, the covariance of the observables f and g is given by

$$\mathrm{Cov}_{f,g}(k) := \int f \cdot g \circ T^k \mathrm{d}\mu - \int f \mathrm{d}\mu \int g \mathrm{d}\mu,$$

for every $k \in \mathbb{N}_0$.

Definition 2.1.1. *Let $f : X \to \mathbb{R}$ be an observable in $L^2(\mu)$. We say that f satisfies a central limit theorem with respect to (T, μ) if there exists $\sigma_f \geq 0$ such that*

$$\lim_{n\to\infty} \mu\left\{x \ : \ \frac{S_n f(x) - n\int f \mathrm{d}\mu}{\sqrt{n}} \leq t\right\} = \frac{1}{\sigma_f\sqrt{2\pi}} \int_{-\infty}^{t} e^{-s^2/2\sigma_f^2} \mathrm{d}s$$

for all $t \in \mathbb{R}$.

When $\sigma_f = 0$ we understand the right-hand side as the Heaviside function.

In the case of dynamical systems, due to the correlations the variance is defined as follows

$$\sigma_f^2 = \lim_{n\to\infty} \frac{1}{n} \int \left(S_n f - n\int f \mathrm{d}\mu\right)^2 \mathrm{d}\mu, \tag{2.1}$$

provided the limit exists. Using the invariance of the measure μ under T, we can write

$$\frac{1}{n}\int \left(S_n f - n\int f \mathrm{d}\mu\right)^2 \mathrm{d}\mu = \mathrm{Cov}_f(0) + 2\sum_{k=1}^{n-1} \frac{n-k}{n} \mathrm{Cov}_f(k).$$

If $\sum_{j=1}^{\infty} |\mathrm{Cov}_f(j)| < \infty$, then $\lim_{n\to\infty} \sum_{k=1}^{n-1} \frac{n-k}{n} \mathrm{Cov}_f(k) = \sum_{k=1}^{\infty} \mathrm{Cov}_f(k)$, which gives us

$$\sigma_f^2 = \mathrm{Cov}_f(0) + 2\sum_{k=1}^{\infty} \mathrm{Cov}_f(k).$$

Concerning to the systems modeled by Young towers with exponential tails, the result on the central limit theorem is the following.

Theorem 2.1.1 (Central limit theorem, [73]). *Let (X, T, μ) be a dynamical system modeled by a Young tower, where μ is its SRB measure. Let $f : X \to \mathbb{R}$ be a Hölder continuous observable. If $\int R^2 \mathrm{dLeb}^u < \infty$, then f satisfies the central limit theorem with respect to μ.*

2.2 Large deviations

Given a bounded i.i.d. process (X_n), it is known that $\mathbb{P}\{|n^{-1}(X_0+\cdots+X_{n-1})-\mathbb{E}(X_0)| > \delta\}$ decays exponentially with n (this result is called Cramér's theorem, [30]). More precisely one can prove that

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{\left|\frac{X_0+\cdots+X_{n-1}}{n} - \mathbb{E}(X_0)\right| > \delta\right\} = -I(\delta),$$

where I is the called *rate function*, which usually is strictly convex and it vanishes only at zero. It turns out that this rate function I is the Legendre transform of the cumulant generating function $\theta \mapsto \log \mathbb{E}(e^{\theta X_0})$. For a dynamical system (X, T, μ) and an observable f, the purpose is to prove that there exists a rate function $I_f : \mathbb{R} \to [0, \infty)$ such that

$$\lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log \mu \left\{x \in X : \frac{1}{n} S_n f(x) \in [a-\varepsilon, a+\varepsilon]\right\} = -I_f(a).$$

The standard route is to try to prove that the cumulant generating function

$$\Psi_f(z) = \lim_{n\to\infty} \frac{1}{n} \log \int e^{zS_n f} \mathrm{d}\mu$$

exists and is smooth for $z \in \mathbb{R}$ in an interval containing the origin.

The results for systems modeled by Young towers with exponential tails are the following.

Theorem 2.2.1 (Cumulant generating functions, [62]). *Let (X, T, μ) be a dynamical system modeled by a Young tower where μ is its SRB measure. Assume that $\mathrm{Leb}^u\{R > n\} = \mathscr{O}(e^{-an})$ for some $a > 0$. Let $f : X \to \mathbb{R}$ be a Hölder continuous observable such that $\int f \mathrm{d}\mu = 0$. Then there exist positive numbers $\eta = \eta(f)$ and $\zeta = \zeta(f)$ such that Ψ_f exists and is analytic in the strip*

$$\{z \in \mathbb{C} \ : \ |\mathrm{Re}(z)| < \eta,\ |\mathrm{Im}(z)| < \zeta\}.$$

In particular, $\Psi_f'(0) = \int f \mathrm{d}\mu = 0$ and $\Psi_f''(0) = \sigma_f^2$, which is the variance (2.1)*. Moreover, $\Psi_f(z)$ is strictly convex for real z provided $\sigma_f^2 > 0$.*

Theorem 2.2.2 (Large deviations, [62]). *Under the same assumptions as in the previous theorem, let I_f be the Legendre transform of Ψ_f, i.e. $I_f(t) = \sup_{z\in(-\eta,\eta)}\{tz - \Psi_f(z)\}$. Then for any interval $[a, b] \subset [\Psi_f'(-\eta), \Psi_f'(\eta)]$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mu \left\{x \in X \ : \ \frac{1}{n} S_n f(x) \in [a, b]\right\} = -\inf_{t\in[a,b]} I_f(t).$$

Remark 2.2.1. *For Gibbs measure on $A^{\mathbb{Z}}$, one has a complete description of large deviations (see e.g. [16]).*

2.3 Concentration Inequalities

2.3.1 Definitions & generalities

The aim is to quantify the probability of deviation of a general function K of n random variables $X_0, \dots, X_{n-1}$ about its expected value. Roughly speaking, the only two general requisites to be satisfied in order to have concentration inequalities are weak dependence between the random variables and the 'smoothness' of the observable K, more accurately, it is enough for K. The right notion of smoothness turns out to be a Lipschitz condition.

Let (X, d) be a metric space. A real-valued function of n variables $K : X^n \to \mathbb{R}$ is said to be *separately Lipschitz* if the quantity

$$\mathrm{Lip}_j(K) := \sup_{x_0,\dots,x_{n-1}} \sup_{x_j \neq x'_j} \frac{|K(x_0,\dots,x_j,\dots,x_{n-1}) - K(x_0,\dots,x'_j,\dots,x_{n-1})|}{d(x_j, x'_j)}$$

is finite for all $j = 0, \dots, n-1$.

Now, we give two definitions. The first one describes what we mean for a stochastic process to satisfy an exponential concentration inequality. The second definition is about weaker inequalities.

Definition 2.3.1 (Exponential concentration inequality)**.** *Consider a stochastic process $\{Z_0, Z_1, \dots,\}$ taking values on X. This process is said to satisfy an exponential concentration inequality if there exists a constant C such that, for any separately Lipschitz function of n variables $K(x_0, \dots, x_{n-1})$, one has*

$$\mathbb{E}\left[e^{K(Z_0,\dots,Z_{n-1}) - \mathbb{E}[K(Z_0,\dots,Z_{n-1})]}\right] \leq e^{C \sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2}. \tag{2.2}$$

Definition 2.3.2 (Polynomial concentration inequality)**.** *Consider a stochastic process $\{Z_0, Z_1, \dots,\}$ taking values on X. We say that this process satisfies a polynomial concentration inequality with moment $q \geq 2$ if there exists a constant C_q such that, for any separately Lipschitz function of n variables $K(x_0, \dots, x_{n-1})$, one has*

$$\mathbb{E}\left[|K(Z_0,\dots,Z_{n-1}) - \mathbb{E}[K(Z_0,\dots,Z_{n-1})]|^q\right] \leq C_q \left(\sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2\right)^{q/2}. \tag{2.3}$$

A special case of the last inequality is the variance inequality, when $q = 2$:

$$\mathrm{Var}(K(Z_0,\dots,Z_{n-1})) \leq C_2 \sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2. \tag{2.4}$$

One can obtain a general consequence of the previous concentration inequalities which gives us upper bounds for the deviation probabilities of $K(Z_0, \dots, Z_{n-1})$ from its expected value.

Corollary 2.3.1. *If the stationary process $\{Z_0, Z_1, \ldots\}$ satisfies the exponential concentration inequality* (2.2) *then for all $t > 0$ and for all $n \geq 1$ we have,*

$$\mathbb{P}\left\{|K(Z_0, \ldots, Z_{n-1}) - \mathbb{E}(K(Z_0, \ldots, Z_{n-1}))| > t\right\} \leq 2e^{\frac{-t^2}{4C\sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2}}. \tag{2.5}$$

If the process satisfies the polynomial concentration inequality (2.3) *for some $q \geq 2$, then we have for all $t > 0$ and for all $n \geq 1$,*

$$\mathbb{P}\left\{|K(Z_0, \ldots, Z_{n-1}) - \mathbb{E}(K(Z_0, \ldots, Z_{n-1}))| > t\right\} \leq \frac{C_q}{t^q}\Big(\sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2\Big)^{q/2}. \tag{2.6}$$

Proof. First observe that inequalities (2.2) and (2.3) are homogeneous. Let λ be a real parameter $\lambda > 0$. Using Markov's inequality we obtain from (2.2), for all $t > 0$

$$\begin{aligned}
&\mathbb{P}\{K(Z_0, \ldots, Z_{n-1}) - \mathbb{E}(K(Z_0, \ldots, Z_{n-1})) > t\} \\
&\qquad\qquad \leq e^{-\lambda t}\mathbb{E}\left(e^{\lambda(K(Z_0,\ldots,Z_{n-1}))-\mathbb{E}(K(Z_0,\ldots,Z_{n-1}))}\right) \\
&\qquad\qquad \leq e^{-\lambda t} e^{C\lambda^2\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2}.
\end{aligned}$$

The last inequality is minimized when $\lambda = t/\left(2C\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2\right)$, giving us

$$\mathbb{P}\{K(Z_0, \ldots, Z_{n-1}) - \mathbb{E}(K(Z_0, \ldots, Z_{n-1})) > t\} \leq e^{\frac{-t^2}{4C\sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2}}.$$

Applying this inequality to $-K$ and using an union bound one obtains (2.5). Analogously, applying Markov's inequality to (2.3) one gets (2.6) immediately. □

It is important to stress that the inequalities above are valid for every n.

2.3.2 Concentration inequalities for dynamical systems

Since chaotic dynamical systems may be seen as stochastic processes, one might expect the concentration inequalities defined in the previous section, also hold in the context of dynamical systems. This is indeed true. For instance, in [27] it was established an exponential concentration inequality for piecewise expanding maps on the interval. In [18], a polynomial concentration inequality was achieved for a system with indifferent fixed point. Here we include the main results in [21], namely, dynamical systems modeled by Young towers satisfy concentration inequalities, we will state this result formally. For a complete panorama we remit the reader to [17] .

As we already mention, dynamical systems modeled by Young towers satisfy concentration inequalities, they are either exponential or polynomial depending on the decay rate of their tails. The results are the following.

Theorem 2.3.1 (Exponential concentration inequality [21]). *Consider (X, T, μ), a dynamical system modeled by a Young tower with exponential tails. Then it satisfies an exponential concentration inequality: there exists a constant $C > 0$ which only depends on T, such that for any integer $n \geq 1$ and for any separately Lipschitz function K of n variables we have*

$$\int e^{K(x,\ldots,T^{n-1}x)-\int K(y,\ldots,T^{n-1}y)\mathrm{d}\mu(y)}\mathrm{d}\mu(x) \leq e^{C\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2}. \tag{2.7}$$

The systems of the examples 1.1.3, 1.3.1 and 1.3.2 in Chapter 1 are included in that framework.

Theorem 2.3.2 (Polynomial concentration inequality [21]). *Consider (X, T, μ), a dynamical system modeled by a Young tower. Assume that, for some $q \geq 2$, $\int R^q \mathrm{dLeb}^u < \infty$. Then it satisfies a polynomial concentration inequality with moment $2q-2$, i.e., there exists a constant $C_q > 0$ such that for any integer $n \geq 1$ and for any separately Lipschitz function K of n variables we have*

$$\int \left|K(x,\ldots,T^{n-1}x) - \int K(y,\ldots,T^{n-1}y)\mathrm{d}\mu(y)\right|^{2q-2}\mathrm{d}\mu(x) \leq C_q\Big(\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2\Big)^{q-1}. \tag{2.8}$$

The constant C_q depends only on q and on T

The prototypical example of system with polynomial tails is the Maneville-Pomeau map 1.3.3, this map satisfies this polynomial concentration inequality of order $q < \frac{2}{\alpha} - 2$ when $\alpha \in (0, 1/2)$ (see also [21]).

As for stochastic processes, an immediate consequence are the corresponding deviation inequalities.

Corollary 2.3.2 ([21]). *Under the same hypothesis as in theorem 2.3.1, for all $t > 0$ and any $n \in \mathbb{N}$*

$$\mu\left(\left|K(x,\ldots,T^{n-1}x) - \int K(y,\ldots,T^{n-1}y)\mathrm{d}\mu(y)\right| > t\right) \leq 2e^{-\frac{t^2}{4C\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2}}. \tag{2.9}$$

Under the hypothesis of theorem 2.3.2, for all $t > 0$ and any $n \in \mathbb{N}$,

$$\mu\left(\left|K(x,\ldots,T^{n-1}x) - \int K(y,\ldots,T^{n-1}y)\mathrm{d}\mu(y)\right| > t\right) \leq \frac{C_q}{t^{2q-2}}\Big(\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2\Big)^{q-1}. \tag{2.10}$$

Before we move to the next section let us make a remark. Consider the function $K_0(x_0, \ldots, x_{n-1}) = f(x_0) + \cdots + f(x_{n-1})$ where f is a Lipschitz observable. We clearly have $\mathrm{Lip}_i(K_0) = \mathrm{Lip}(f)$ for all $i = 0, \ldots, n-1$. When evaluated along the orbit segment

$x, \ldots, T^{n-1}x$ we get the ergodic sum $S_n f(x)$. Supposing we can apply inequality (2.9), one gets for all $t > 0$

$$\mu\left(\left|S_n f - n\int f \mathrm{d}\mu\right| > t\right) \leq e^{\frac{-t^2}{4Cn\mathrm{Lip}(f)^2}},$$

rescaling t by n, one can rewrite the expression above, having

$$\mu\left(\left|\frac{1}{n}S_n f - \int f \mathrm{d}\mu\right| > t\right) \leq 2e^{\frac{-nt^2}{4C\mathrm{Lip}(f)^2}},$$

for all $t > 0$. This inequality gives the same order in n as the large deviations description, although one looses accuracy we obtain estimates valid for any n. If we scale t by $\sqrt{n}$ we obtain

$$\mu\left(\left|\frac{1}{\sqrt{n}}\big(S_n f - \int f \mathrm{d}\mu\big)\right| > t\right) \leq 2e^{\frac{-t^2}{4C\mathrm{Lip}(f)^2}},$$

for all $t > 0$. This gives the same order in t as the central limit theorem. In this sense the result given by concentration inequalities are compatible with both, central limit theorem and large deviations but with the advantage that are applicable to general functions not only ergodic sums.

2.4 Some applications

In this section we apply concentration inequalities in the context of dynamical systems. We obtain deviation probabilities of some empirical estimators. The quantities we consider are the auto-convariance function, the empirical measure, the kernel density estimator and the correlation dimension. All of them have been studied previously in [19] and [21].

2.4.1 Auto-covariance function

Consider the dynamical system (X, T, μ) and a square integrable observable $f : X \to \mathbb{R}$. Assume that f is such that $\int f d\mu = 0$. We remind that the auto-covariance function of f is given by

$$\mathrm{Cov}(k) := \mathrm{Cov}_f(k) = \int f(x)f(T^k x)\mathrm{d}\mu(x). \tag{2.11}$$

In practice, one has a finite number of iterates of some μ-typical initial condition x, thus, what we may obtain from data is the empirical estimator of the auto-covariance function:

$$\widehat{\mathrm{Cov}}_n(k) := \frac{1}{n}\sum_{i=0}^{n-1} f(T^i x)f(T^{i+k}x).$$

From Birkhoff's ergodic theorem it follows that $\mathrm{Cov}(k) = \lim_{n\to\infty} \widehat{\mathrm{Cov}}_n(k)$ almost surely. Observe that the expected value of the estimator $\widehat{\mathrm{Cov}}_n(k)$ is exactly $\mathrm{Cov}(k)$, because of the invariance of μ and the assumption that $\int f \mathrm{d}\mu = 0$.

The following result gives us *a priori* theoretical bounds to the fluctuations of the estimator $\widehat{\mathrm{Cov}}_n$ around Cov for every n.

Theorem 2.4.1 ([21]). *Let (X, T, μ) be a dynamical system modeled by a Young tower with exponential tails, then there exists a constant $C > 0$ such that for all $t > 0$ and any $n, k \in \mathbb{N}$ we have that*

$$\mu\left(\left|\widehat{\mathrm{Cov}}_n(k) - \mathrm{Cov}(k)\right| > t\right) \leq 2\exp\left(-C\frac{t^2n^2}{n+k}\right).$$

If the system (X, T, μ) is a dynamical system modeled by a Young tower with L^q-tails, for some $q \geq 2$, then there exists a constant C_q such that, for all $t > 0$ and any integer $n, k \in \mathbb{N}$ we have

$$\mu\left(\left|\widehat{\mathrm{Cov}}_n(k) - \mathrm{Cov}(k)\right| > t\right) \leq \frac{C_q}{t^{2q-2}}\left(\frac{n+k}{n^2}\right)^{q-1}.$$

Proof. Choose the following observable of $n + k$ variables,

$$K(z_0, \ldots, z_{n+k-1}) := \frac{1}{n}\sum_{i=0}^{n-1} f(z_i)f(z_{i+k}).$$

In order to estimate the Lipschitz constant of K, consider $0 \leq l \leq n+k-1$ and change the value z_l to z'_l. Observe that the difference between $K(z_0, \ldots, z_l, \ldots, z_{n+k-1})$ and $K(z_0, \ldots, z'_l, \ldots, z_{n+k-1})$ is less than or equal to $\frac{1}{n}|f(z_{l-k})f(z_l) + f(z_l)f(z_{l+k}) - f(z_{l-k})f(z'_l) - f(z'_l)f(z_{l+k})|$, and so for every index l, we obtain

$$\mathrm{Lip}_l(K) \leq \sup_{z_0,\ldots,z_{n+k-1}} \sup_{z_l \neq z'_l} \frac{1}{n}\frac{|[f(z_l) - f(z'_l)][f(z_{l-k}) + f(z_{l+k})]|}{d(z_l, z'_l)} \leq \frac{2}{n}\mathrm{Lip}(f)\|f\|_\infty.$$

The exponential inequality follows immediately by applying inequality (2.9). Using analogously the inequality (2.10) for the polynomial case gives us the deviation probability of the auto-covariance function for dynamical systems with non-uniform Young towers as desired. □

2.4.2 Empirical measure

Given a μ-typical $x \in X$, define the empirical measure of the sequence $x, \ldots, T^{n-1}x$ as

$$\mathcal{E}_n(x) := \frac{1}{n}\sum_{i=0}^{n-1} \delta_{T^ix}, \tag{2.12}$$

where δ_y denotes the Dirac measure sitting at y. From Birkhoff's ergodic theorem it follows that the sequence of random measures $\{\mathcal{E}_n\}$ converges weakly to the T-invariant measure μ, almost surely. We want to quantify the speed of this convergence. For that purpose, we choose the Kantorovich distance on the set of probability measures of X, which is defined by

$$\kappa(\mu,\nu) := \sup_{g\in\mathcal{L}} \int g d\mu - \int g d\nu,$$

where μ and ν are two probability measures on X and $\mathcal{L}$ denotes the space of all real-valued Lipschitz functions on X with Lipschitz constant at most one.

We shall use the shorthand notation

$$\mathcal{K}_n(x) := \kappa(\mathcal{E}_n(x), \mu).$$

We have the following general bounds.

Theorem 2.4.2 ([21])**.** *If the system (X,T,μ) is modeled by a Young tower with exponential tails, with SRB measure μ, then there exists a constant $C>0$ such that, for any $n\in\mathbb{N}$ and for any $t>0$,*

$$\mu\left(\left|\mathcal{K}_n(x) - \int \mathcal{K}_n(y)\mathrm{d}\mu(y)\right| > \frac{t}{\sqrt{n}}\right) \leq 2\exp^{-Ct^2}.$$

Furthermore, if the system (X,T,μ) is modeled by a Young tower with L^q-tails, for some $q\geq 2$, then there exists a constant $C_q>0$ such that, for all $n\in\mathbb{N}$ and all $t>0$,

$$\mu\left(\left|\mathcal{K}_n(x) - \int \mathcal{K}_n(y)\mathrm{d}\mu(y)\right| > \frac{t}{\sqrt{n}}\right) \leq \frac{C_q}{t^{2q-2}}.$$

This result follows directly applying inequalities (2.7) and (2.8) to the following function of n variables

$$K(x_0,\ldots,x_{n-1}) = \sup\left\{\frac{1}{n}\sum_{i=0}^{n-1} g(x_i) - \int g\mathrm{d}\mu\right\},$$

where $g\in\mathcal{L}$. It is easy to see that the Lipschitz constants of K are bounded by $1/n$. Finding an upper bound to $\int \mathcal{K}_n \mathrm{d}\mu$ would provide an estimate on the deviation probability for $\mathcal{K}_n$ itself. Up to now, we can find a general good estimate only in dimension one.

Corollary 2.4.1 ([21])**.** *If the system (X,T,μ) is an one-dimensional dynamical system modeled by a Young tower with exponential tails, then there exist some constants $B,C>0$ such that, for any $n\in\mathbb{N}$ and for any $t>0$,*

$$\mu\left(\mathcal{K}_n(x) > \frac{t}{\sqrt{n}} + \frac{B}{n^{1/4}}\right) \leq e^{-Ct^2}.$$

If the system (X,T,μ) is modeled by a Young tower with L^q-tails, for some $q\geq 2$, then there exist some constant $B,C_q>0$ such that, for any $n\in\mathbb{N}$ and for any $t>0$,

$$\mu\left(\mathcal{K}_n(x) > \frac{t}{\sqrt{n}} + \frac{B}{n^{1/4}}\right) \leq \frac{C_q}{t^{2q-2}}.$$

This corollary follows at once from the next lemma.

Lemma 2.4.1 ([19]). *Let (X, T, μ) be an one-dimensional dynamical system. If there exists a constant $c > 0$ such that for every Lipschitz function $f : X \to \mathbb{R}$, the auto-covariance function $\mathrm{Cov}_f(k)$ satisfies that $\sum_{k=1}^{\infty} |\mathrm{Cov}_f(k)| \leq c\|f\|_{\mathrm{Lip}}^2$, then there exists a constant B such that*

$$\int \mathcal{K}_n(x) \mathrm{d}\mu(x) \leq \frac{B}{n^{1/4}}.$$

The proof of the preceding lemma is found in [19, Section 5]. It relies in the fact that in dimension one, it is possible to rewrite the Kantorovich distance using distribution functions. Then by an adequate Lipschitz approximation of the distribution function, the bound follows from the summability condition on the auto-covariance function.

2.4.3 Kernel density estimator for one-dimensional maps

In this section we consider the system (X, T, μ) to take values on a bounded subset of $\mathbb{R}$. We assume the measure μ to be absolutely continuous with density h. For a given trajectory of a randomly chosen initial condition x (according to μ), the empirical density estimator is defined by,

$$\widehat{h}_n(x, Tx, \ldots, T^{n-1}x; s) := \frac{1}{n\alpha_n} \sum_{j=0}^{n-1} \psi\left(\frac{s - T^j x}{\alpha_n}\right),$$

where $\alpha_n \to 0$ and $n\alpha_n \to \infty$ as n diverges. The kernel ψ is a bounded, non-negative Lipschitz function with bounded support and it satisfies $\int \psi(s) ds = 1$. We assume the following hypothesis.

Hypothesis 2.4.1. *The probability density h satisfies*

$$\int |h(s) - h(s - \sigma)| \mathrm{d}s \leq C' |\sigma|^{\beta}$$

for some $C' > 0$ and $\beta > 0$ and for every $\sigma \in \mathbb{R}$.

This assumption is indeed valid for maps on the interval satisfying the axioms of Young towers (see Appendix C. in [19]). For convenience, we present the following result on the L^1 convergence of the density estimator.

Theorem 2.4.3. *Let ψ be a kernel defined as above. If the system (X, T, μ) satisfies the exponential concentration inequality (2.2) and the hypothesis 2.4.1, then there exists a constant $C_\psi > 0$ such that for any integer $n \geq 1$ and every $t > C_\psi \left(\alpha_n^{\beta} + \frac{1}{\sqrt{n}\alpha_n^2}\right)$, we have*

$$\mu\left(\left\{\int \left|\widehat{h}_n(x_0, \ldots, T^{n-1}x; s) - h(s)\right| \mathrm{d}s > t\right\}\right) \leq e^{-\frac{t^2 n \alpha_n^2}{4D\mathrm{Lip}(\psi)^2}}.$$

Under the same conditions above, if the system satisfies the polynomial concentration inequality (2.3) *for some* $q \geq 2$*, then for any integer* $n \geq 1$ *and every* $t > C_\psi \left(\alpha_n^\beta + \frac{1}{\sqrt{n}\alpha_n^2}\right)$*, we obtain,*

$$\mu\left(\left\{\int \left|\widehat{h}_n(x_0,\dots,T^{n-1}x;s) - h(s)\right| \mathrm{d}s > t\right\}\right) \leq \frac{D}{t^q}\left(\frac{\mathrm{Lip}(\psi)}{\sqrt{n}\alpha_n}\right)^q.$$

For the proof of this statement see [21] or Theorem 6.1 in [19].

2.4.4 Correlation dimension

The correlation dimension $d_c = d_c(\mu)$ of the measure μ is defined by

$$d_c = \lim_{r \searrow 0} \frac{\log \int \mu(B_r(x))\mathrm{d}\mu(x)}{\log r},$$

whenever the limit exists. We denote by $\mathrm{Corr}(r)$ the *spatial correlation integral* which is defined by

$$\mathrm{Corr}(r) = \int \mu(B_r(x))\mathrm{d}\mu(x).$$

In [36] the authors defined a method to empirically determine the correlation dimension. As empirical estimator of $\mathrm{Corr}(r)$ they use the function

$$K_{n,r}(x_0,\dots,x_{n-1}) := \frac{1}{n^2}\sum_{i\neq j} H(r - d(x_i,x_j)),$$

where H is the Heaviside function. For large n they look at the power law behavior in r of $K_{n,r}(x,\dots,T^{n-1}x)$. It has been rigorously proved (see [61] and [67]) that

$$\mathrm{Corr}(r) = \lim_{n\to\infty} K_{n,r}(x,\dots,T^{n-1}x),$$

μ-almost surely at the continuity points of $\mathrm{Corr}(r)$. Observe that we are not allowed to immediately apply concentration inequalities since Heaviside function is not Lipschitz. The usual trick is the following: Consider any real-valued Lipschitz function ψ and define

$$K^\psi_{n,r}(x_0,\dots,x_{n-1}) := \frac{1}{n^2}\sum_{i\neq j}\psi\left(1 - \frac{d(x_i,x_j)}{r}\right).$$

Theorem 2.4.4 ([19])**.** *For any real-valued Lipschitz* ψ*, there exists a constant* $C > 0$ *such that for any* $r > 0$ *and any integer* n*, we have*

$$\mathrm{Var}(K^\psi_{n,r}) \leq \frac{C}{r^2 n}.$$

The proof is an straightforward application of the variance inequality (2.4). For a more detailed discussion of this application see [28, Section 9.6].

2.5 Concentration Inequalities for Gibbs measures

For the sake of completeness we include the proof of the exponential concentration inequality for Gibbs measures on the full-shift. We adopt the definitions of section 1.2.

Theorem 2.5.1 ([21])**.** *The system $(\Omega, \sigma, \mu_\phi)$ satisfies an exponential concentration inequality.*

Proof. Let us consider a separately Lipschitz function of n variables, $K : \Omega^n \to \mathbb{R}$. In order to lighten notations, we write x_i meaning a $\underline{x}^{(i)} \in \Omega$, and so, $K(x_0, \dots, x_{n-1})$ instead of $K\big(\underline{x}^{(0)}, \dots, \underline{x}^{(n-1)}\big)$.

One may interpret K as a function on $\Omega^{\mathbb{N}}$ depending only on the first n coordinates and having $\mathrm{Lip}_j(K) = 0$ for all $j \geq n$. The space $\Omega^{\mathbb{N}}$ is endowed with the measure μ_∞ limit of the measure μ_N (given by (1.1)) when $N \to \infty$.

Consider $\{\mathcal{F}_p\}$ to be a decreasing sequence of the σ-algebras of the events depending only on the coordinate $(x_j)_{j \geq p}$.

The trick is to write K as a sum of martingale differences with respect to $\mathcal{F}_p$. Let $K_p := \mathbb{E}(K \mid \mathcal{F}_p)$ and $D_p := K_p - K_{p+1}$. Note that the function D_p is $\mathcal{F}_p$-measurable and $\mathbb{E}(D_p \mid \mathcal{F}_{p+1}) = 0$. Observe also that $K - \mathbb{E}(K) = \sum_{p \geq 0} D_p$.

Using the Hoeffding inequality (see [50]) one obtain for any integer $P \geq 1$

$$\mathbb{E}\Big(e^{\sum_{p=0}^{P-1} D_p}\Big) \leq e^{\sum_{p=0}^{P-1} \sup|D_p|^2}.$$

The rest of the proof consists in getting a good bound on D_p, let us assume for the moment the following lemma.

Lemma 2.5.1. *There exist constants $C > 0$ and $\rho < 1$ such that, for any p, one has*

$$|D_p| \leq C \sum_{j=0}^{p} \rho^{p-j} \mathrm{Lip}_j(K). \tag{2.13}$$

Next, using the Cauchy-Schwarz inequality we have

$$\Big(\sum_{j=0}^{p} \rho^{p-j} \mathrm{Lip}_j(K)\Big)^2 \leq \Big(\sum_{j=0}^{p} \rho^{p-j} \mathrm{Lip}_j(K)^2\Big)\Big(\sum_{j=0}^{p} \rho^{p-j}\Big) \leq \frac{1-\rho^p}{1-\rho} \sum_{j=0}^{p} \mathrm{Lip}_j(K)^2.$$

We apply this bound to the right hand side of the inequality (2.13). Then summing over p one has $\sum_{p=0}^{P-1} \sup|D_p|^2 \leq C' \sum_j \mathrm{Lip}_j(K)^2$. Using the Hoeffding inequality at a fixed P and then letting P tend to infinity, one obtain

$$\mathbb{E}\Big(e^{\sum_{p \geq 0} D_p}\Big) \leq e^{C' \sum_j \mathrm{Lip}_j(K)^2},$$

which is the desired inequality. □

We continue with the proof of lemma 2.5.1.

Proof of lemma 2.5.1. The transfer operator $\mathscr{L}$ associated to the potential $\phi \in \mathscr{F}_\theta$ is given by

$$\mathscr{L}u(x) = \sum_{\sigma y = x} e^{\phi(y)} u(x).$$

Thus, $\mathscr{L}^k u(x) = \sum_{\sigma^k y = x} e^{S_k \phi(y)} u(y)$. On the set of preimages of $\underline{x}$ one may identify $\mathscr{L}$ as a Markov operator, whose transition probabilities are given by $e^{\phi(y)}$. Then one has

$$\begin{aligned} K_p(x_p, x_{p+1}, \ldots) =& \mathbb{E}(K \mid \mathcal{F}_p)(x_p, x_{p+1}, \ldots) = \mathbb{E}(K(X_0, \ldots, X_{p-1}, x_p, \ldots) \mid X_p = x_p) \\ =& \sum_{\sigma^p y = x_p} e^{S_p \phi(y)} K(y, \ldots, \sigma^{p-1} y, x_p, \ldots). \end{aligned}$$

To prove that D_p is bounded one make use of the following lemma,

Lemma 2.5.2. *One has*

$$\left| K_p(x_p, \ldots) - \int K(y, \ldots, \sigma^{p-1} y, x_p, \ldots) \mathrm{d}\mu_\phi(y) \right| \leq C \sum_{j=0}^{p-1} \mathrm{Lip}_j(K) \rho^{p-1-j},$$

where $C > 0$ and $\rho < 1$ depend only on (Ω, σ)

Notice that, in particular one obtain that $K(x_p, x_{p+1}, \ldots) - K_p(x'_p, x_{p+1} \ldots)$ is bounded from above by $C \sum_{j=0}^{p} \mathrm{Lip}_j(K) \rho^{p-j}$. Taking the average over the preimages x'_p of x_{p+1} one obtain the same bound for $D_p(x_p, x_{p+1}, \ldots)$ and that proves the lemma 2.5.1. $\square$

Proof of lemma 2.5.2. Let us fix a point $x_* \in \Omega$, one writes K_p in the following telescopic way using the transfer operator,

$$\begin{aligned} K_p(x_p, \ldots) =& \sum_{j=0}^{p} \sum_{\sigma^p(y) = x} e^{S_p \phi(y)} \big[K(y, \ldots, \sigma^j y, x_*, \ldots, x_*, x_p, \ldots) \\ & \qquad - K(y, \ldots, \sigma^{j-1} y, x_*, \ldots, x_*, x_p) \big] + K(x_*, \ldots, x_*, x_p, \ldots) \\ =& \sum_{j=0}^{p-1} \mathscr{L}^{p-j} f_j(x_p) + K(x_*, \ldots, x_*, x_p, \ldots), \end{aligned}$$

where

$$\begin{aligned} f_j(z) &= \sum_{\sigma^j y = z} e^{S_j \phi(y)} \big[K(y, \ldots, \sigma^j y, \underline{x}_*, \ldots, x_*, x_p, \ldots) \\ & \qquad - K(y, \ldots, \sigma^{j-1} y, x_*, \ldots, x_*, x_p, \ldots) \big] \\ &= \sum_{\sigma^j y = z} e^{S_j \phi(y)} H(y, \ldots, \sigma^j y). \end{aligned}$$

Since the transfer operator acts on functions of one variable we will eliminate the variables $x_0, \ldots, x_{p-1}$ one after another.

First notice that the function H is bounded by $\mathrm{Lip}_j(K)$, then $|f_j| \leq \mathrm{Lip}_j(K)$ since $\sum_{\sigma^j y=z} e^{S_j\phi(y)} = 1$. To estimate the Lipschitz norm of f_j, write

$$f_j(z) - f_j(z') = \sum_{\sigma^j y=z} \big(e^{S_j\phi(y)} - e^{S_j\phi(y')}\big) H(y, \ldots, \sigma^j y) \\ + \sum_{\sigma^j y=z} e^{S_j\phi(y')} \big[H(y, \ldots, \sigma^j y) - H(y', \ldots, \sigma^j y')\big],$$

where z and z' are two points in the partition element and their respective preimages y and y' are paired according to the cylinder of length j they belong to. A distortion control gives $|e^{S_j\phi(y)} - e^{S_j(y')}| \leq Ce^{S_j\phi(y)}d_\theta(z, z')$, for some constant $C > 0$, hence the first sum is bounded by $C\mathrm{Lip}_j(K)d_\theta(z, z')$. For the second sum, using successively the triangle inequality, one has

$$|H(y, \ldots, \sigma^j y) - H(y', \ldots, \sigma^j y')| \leq 2\sum_{i=0}^{j} \mathrm{Lip}_i(K)d_\theta(\sigma^i y, \sigma^i y') \leq \sum_{i=0}^{j} \mathrm{Lip}_i(K)\beta^{j-i}d_\theta(z, z').$$

Summing over the different preimages, one obtain that the Lipschitz norm of f_j is bounded by $C\sum_{i=0}^{j} \mathrm{Lip}_i(K)\beta^{j-i}$.

Since the operator $\mathscr{L}$ possess a spectral gap on $\mathscr{F}_\theta$. i.e. there exist constants $C > 0$ and $\rho < 1$ such that $\|\mathscr{L}^k f - \int f\mathrm{d}\mu\|_{\mathscr{F}_\theta} \leq C\rho^k\|f\|_{\mathscr{F}_\theta}$. Using the previous inequality one obtains

$$\|\mathscr{L}^{p-j} f_j - \int f_j\mathrm{d}\mu\|_{\mathscr{F}_\theta} \leq C\rho^{p-j}\sum_{i=0}^{j} \mathrm{Lip}_i(K)\beta^{j-i}.$$

This bound implies a bound for the supremum. Assumig $\rho \geq \beta$ and putting together the previous bounds, one has

$$\begin{aligned}\Big|K_p(x_p, \ldots) - \sum_{j=0}^{p-1}\int f_j\mathrm{d}\mu - K(x_*, \ldots, x_*, x_p, \ldots)\Big| \\ \leq C\sum_{j=0}^{p-1}\rho^{p-j}\sum_{i=0}^{j}\mathrm{Lip}_i(K)\rho^{j-i} \leq C\sum_{j=0}^{p-1}\mathrm{Lip}_j(K)\rho^{p-j}(p-j) \\ \leq C'\sum_{j=0}^{p-1}\mathrm{Lip}_j(K)(\rho')^{p-j},\end{aligned}$$

for any $\rho' \in (\rho, 1)$. Computing the sum of the integrals of f_j one gets only the term $\int K(y, \ldots, \sigma^{p-1}y, x_p \ldots)\mathrm{d}\mu(y)$ giving us the desired expression. □

As a final remark of this chapter, one might ask if it is possible to estimate the potential ϕ associated to the measure μ_ϕ provided a single sequence $x_0 \cdots x_{n-1}$. This is indeed possible and it was proved by V. Maume-Deschamps in [57]. The approach in that paper is to use empirical probabilities as estimators of conditional probabilities.

In the case of Gibbs measures that result can be used to derive an estimation of the potential. Given a sequence x_0^{n-1} the empirical frequency of the word a_0^{k-1} for $k \leq n$ is defined by

$$\mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) := \frac{1}{n} \#\{0 \leq j \leq n : \tilde{x}_j^{j+k-1} = a_0^{k-1}\},$$

where $\underline{\tilde{x}} := x_0^{n-1} x_0^{n-1} \cdots$ is the periodic point with period n made from x_0^{n-1}. The result of interest is the following.

Theorem 2.5.2 ([57])**.** *Let μ_ϕ be a Gibbs measure and let the sequence x_0^{n-1} be produced by μ_ϕ. For some $\varepsilon \in (0,1)$ there exits $L > 0$ such that for all $t > 0$,*

$$\mu_\phi \left(\left| \frac{\mathcal{E}_{k,n}(a_0^{k-1}; x_0^{n-1})}{\mathcal{E}_{k-1,n}(a_1^{k-1}; x_1^{n-1})} - e^{\phi(\underline{x})} \right| > t \right) \leq 4e^{-Lt^2 n^{1-\varepsilon}} + 2e^{-Ln^{1-\varepsilon}},$$

for every $\underline{x} \in [a_0^{k-1}]$.

We remit the reader to [57] for the proof and more details.

Part II

Chaotic Dynamical Systems plus Noise

Chapter 3

Modeling dynamical systems with noise

Suppose we are given only the form of the model of a certain dynamical system, but the parameters of the model have to be determined from empirical observations. This is actually a problem of statistical inference called parameter estimation (we remit the reader to [58] for a review). At this point, another problem arises, 'empirical observations' are normally given by experimental data in the form of time series, and in practice all experimental data are corrupted by noise. Therefore it is important to include noise in realistic mathematical models.

In the literature one may find two main sorts of noise models. On one hand, the *dynamical noise* in which the sequence of 'states' are intrinsically corrupted by noise and thus the noise term evolves within the dynamics. And on the other hand, the so called *observational noise* (also called *measurement noise*), in which the perturbation is supposed to be generated by the observation process (measurement).

3.1 Dynamical noise

Let us first give an intuitive description of what dynamical noise is. Consider our discrete dynamical system (X, T, μ) and an initial condition $x_0 \in X$. For every $i \geq 0$, let

$$x_{i+1} = Tx_i + e_i,$$

where $(e_i)_{i \geq 0}$ is a stationary sequence of random variables given with common probability distribution P. To observe how the noise folds into the dynamics, take the next iteration $x_{i+2} = T(x_{i+1}) + e_{i+1} = T(x_i + e_i) + e_{i+1}$.

One can model rigorously the random perturbations of a transformation T using Markov chains. This is usually done in two approaches. On one hand, one considers transitions of the image $T(x)$ of some point x, according to a probability distribution. On the other hand, by letting T to depend on some parameter ω, which is chosen randomly at each iteration. The two approaches are called random noise and random maps respectively and they will be roughly described below. For complete details we refer the interested reader to [5, 9, 45, 54] for instance.

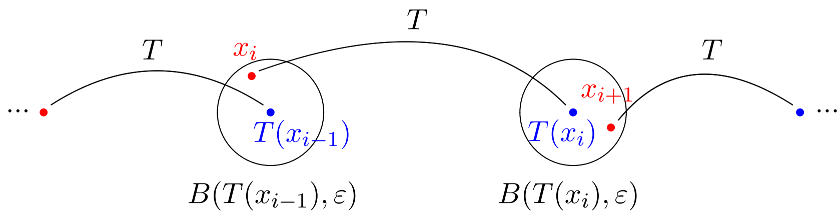
3.1.1 Random noise

Given the map T and a family $\{P(\,\cdot \mid x) : x \in X\}$ of probability measures on X, such that the support of $P(\,\cdot \mid x)$ is a subset of X containing $T(x)$. Define the random orbits as the sequences $(x_i)_{i\geq 1}$ where each x_{i+1} is a random variable with probability distribution $P(\,\cdot \mid x_i)$. This defines a Markov chain with state space X and transition probabilities $\{P(\,\cdot \mid x)\}_{x\in X}$.

Example 3.1.1 (Random jumps)**.** *Given $T : X \to X$ and $\varepsilon > 0$, define*

$$P_\varepsilon(A \mid x) := \frac{\mathrm{Leb}(A \cap B(T(x), \varepsilon))}{\mathrm{Leb}(B(T(x), \varepsilon))},$$

where Leb *is the Lebesgue measure on X and $B(y, r)$ is the ball centered in y with radius r. Then $P_\varepsilon(\,\cdot \mid x)$ is the normalized volume restricted to the ε–neigborhood of $T(x)$, defining a family of transition probabilities allowing the points to "jump" from $T(x)$ to any other point in the ball according to the uniform distribution.*



3.1.2 Random maps

In this context let the map T_0 be the original dynamical system. One may chose maps $T_1, T_2, \ldots, T_i$ independently but close to T_0 and randomly according to a probability distribution ν in the space of maps $\mathcal{T}(X)$ whose support is close to T_0 in the same topology. Consider the random orbits of the initial point x_0 be defined by

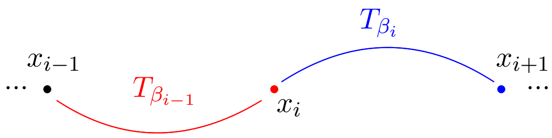
$$x_i := T_i \circ \cdots \circ T_1(x_0),$$

for every $i \geq 1$ and any $x_0 \in X$.

Example 3.1.2 (random β–transformations)**.** *For a real number $\beta \in (1, \infty)$, the β–transformation $T_\beta : [0, 1] \to [0, 1]$ is defined by $x \mapsto \beta x \bmod 1$. Fix a β_0, and consider β_i*

to be a random variable chosen according to a stationary stochastic process (in particular i.i.d.). Then the iteration is given by

$$x_{i+1} := T_{\beta_i} x_i = T_{\beta_i} \circ \cdots \circ T_{\beta_0} x_0.$$



In [13] J. Buzzi, proved that random Lasota-Yorke maps admits absolutely continuous SRB measures. For instance the β-transformations just described satisfy his conditions.

3.1.3 Abstract setting: Skew-maps

Both approaches, random noise and random maps can be placed into the abstract setting of random dynamical systems and skew-products. Let $(\Omega, \mathcal{B}, \mathbb{P}, \theta)$ be a given probabilistic space, which will be the model for the noise. Consider a measurable space $(X, \mathcal{F})$ and let $\mathbb{T}$ be the time set, usually $\mathbb{Z}_+$ or $\mathbb{Z}$. A random dynamical system on X over Ω is a map

$$\varphi : \mathbb{T} \times \Omega \times X \to X,$$

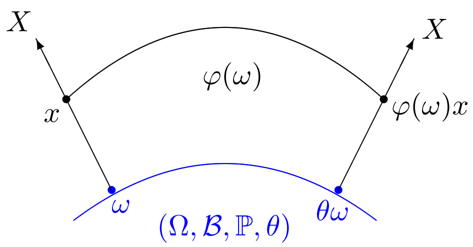
such that the triplet (n, ω, x) is mapped into $\varphi(n, \omega)x$ satisfaying

$$\varphi(0, \omega) = \mathrm{Id}_X \quad \text{for all} \quad \omega \in \Omega,$$

and the cocycle property,

$$\varphi(n + m, \omega) = \varphi(n, \theta^m \omega) \circ \phi(m, \omega),$$

for all $n, m \in \mathbb{T}$ and for every $\omega \in \Omega$, and such that for all $n \in \mathbb{T}$ the map $\varphi(n, \cdot) : \Omega \times X \to X$ is measurable. Corresponding to the random dynamical system φ, it is introduced the skew-product $\mathcal{S} : \Omega \times X \to \Omega \times X$ mapping (ω, x) into $(\theta\omega, \varphi(\omega)x)$. One can see $\mathcal{S}$ as a bundle map, as represented in the figure below.



The further properties and complete details are out of the purpose of this thesis and we refer the reader to [5, 45, 44, 9].

3.2 Observational noise

Consider the discrete dynamical system (X, T, μ). The noise process is modeled as bounded random variables ξ_n defined on a probability space $(\Omega, \mathcal{B}, P)$ and assuming values in X. Without loss of generality, we can assume that the random variables ξ_n are centered, i.e. have expectation equal to 0.
In most cases, the noise is small and it is convenient to represent it by the random variables $\varepsilon\xi_i$ where $\varepsilon > 0$ is the amplitude of the noise and ξ_i is of order one.

We introduce the following definition.

Definition 3.2.1 (Observed system)**.** *For every $i \in \mathbb{N} \cup \{0\}$ (or $i \in \mathbb{Z}$ if the map T is invertible), we say that the sequence of points $\{y_i\}$ given by*

$$y_i := T^i x + \varepsilon\xi_i,$$

is a trajectory of the dynamical system (X, T, μ) perturbed by the observational noise (ξ_n) with amplitude $\varepsilon > 0$. Hereafter we refer to it simply as the observed system.

A natural assumption on the noise is the following.

Standing assumption on noise:

1. (ξ_n) is independent of X_0 and $\|\xi_n\| \leq 1$;
2. The random variables ξ_i are independent.

Example 3.2.1. *Consider the Lozi map $T_L : \mathbb{R}^2 \to \mathbb{R}^2$ which is given by*

$$T_L(u, v) = (1 - a|u| + v, bu), \qquad (u, v) \in \mathbb{R}^2.$$

For $a = 1.7$ and $b = 0.5$ one observes numerically a strange attractor. Consider the uniform distribution on $B_1(0)$, the ball centered at zero with radius one, which is taken as the state space of the random variables ξ_i. Let us denote by x the vector (u, v). For $\varepsilon > 0$, the observed system is given by $y_i = T_L^i x + \varepsilon\xi_i$.

Example 3.2.2. *Consider the Hénon map $T_H : \mathbb{R}^2 \to \mathbb{R}^2$ defined as*

$$T_H(u, v) = \big(1 - au^2 + v, bu\big), \qquad (u, v) \in \mathbb{R}^2.$$

Where $0 < a < 2$ and $b > 0$ are some real parameters. The state space of the random variables is $B_1(0)$ and consider the uniform distribution as in the previous example. The observed system is given by $y_i = T_H^i x + \varepsilon\xi_i$ provided the initial condition $x = (u, v)$ in the basin of attraction of the Hénon map.

Example 3.2.3. *For the Manneville-Pomeau map defined as in example 1.3.3, the observed sequence is defined by $y_i = T_\alpha^i(x) + \varepsilon\xi_i$. The random variables ξ_i are uniformly distributed in X. One identifies the $[0, 1]$ with the unit circle to avoid leaks.*

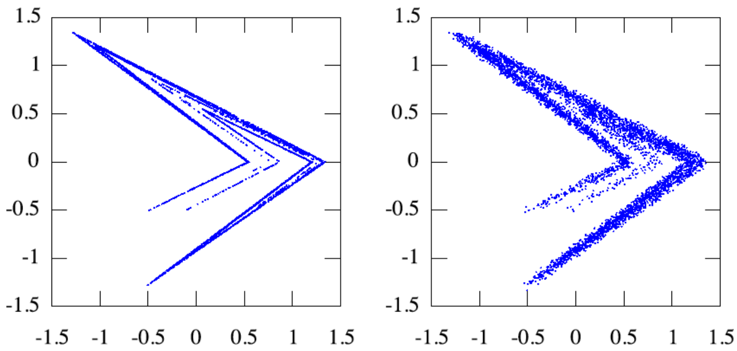


Figure 3.1: Left. A simulation of the Lozi map for the parameters a=1.7 and b=0.5. Right. A simulation of the observed Lozi map with observational noise whose magnitude is bounded by $\varepsilon = 0.06$.

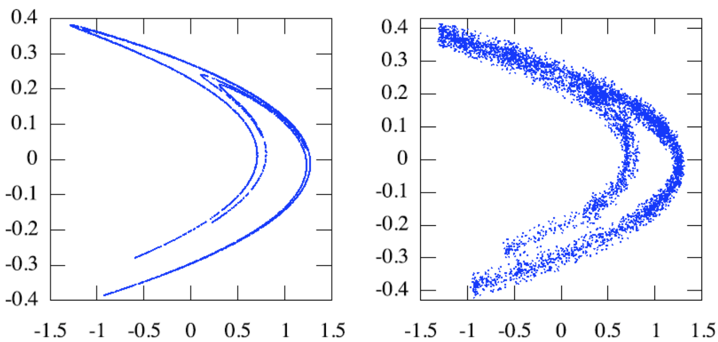


Figure 3.2: Top. A simulation of the Hénon map for the classical parameters a=1.4 and b=0.3. Bottom. A simulation of the Hénon map with observational noise with $\varepsilon = 0.04$.

As a final remark, we point out that from the observations one is not able to distinguish between dynamical and observational noise. An interesting study has been done in the physical context in [70]. We do not address this problem here but it would be an interesting problem to study mathematically.

Chapter 4

Fluctuation bounds for chaos plus noise

In this chapter we give and prove concentration inequalities for systems perturbed by observational noise (observed systems). We also apply these inequalities to obtain bounds on fluctuation of some estimators. The results here presented can be found in [53].

Suppose that we are given with a finite 'sample' $y_0, \ldots, y_{n-1}$. The sample is generated by a dynamical system perturbed by an observational noise. Consider a general observable $K(y_0, \ldots, y_{n-1})$. We are interested in estimating the fluctuations of K and its convergence properties as n grows. It is important to quantifying the effect of the noise in that estimates.

Our main tool to provide such an estimates are concentration inequalities.

4.1 Concentration inequalities for Chaos plus noise

After fixing some notations and conventions, we give and prove the result of this section.

We recall that P is the common distribution of the random variables ξ_i. The expected value with respect to a measure ν is denoted by $\mathbb{E}_\nu$. Recall the expression (1.1) for the measure μ_n. Hence in particular

$$\begin{aligned}\mathbb{E}_{\mu_n}(K) &= \int \cdots \int K(x_0, \ldots, x_{n-1}) \mathrm{d}\mu_n(x_0, \ldots, x_{n-1}) \\ &= \int K(x, \ldots, T^{n-1}x) \mathrm{d}\mu(x).\end{aligned}$$

Next, we denote by $\mu_n \otimes P^n$ the product of the measures μ_n and P^n, where P^n stands for $P \otimes \cdots \otimes P$ (n times). The expected value of $K(y_0, \ldots, y_{n-1})$ is denoted by

$$\mathbb{E}_{\mu_n \otimes P^n}(K) := \int K(x + \varepsilon\xi_0, \ldots, T^{n-1}x + \varepsilon\xi_{n-1}) \mathrm{d}\mu(x) \mathrm{d}P(\xi_0) \cdots \mathrm{d}P(\xi_{n-1}).$$

Theorem 4.1.1. *If the original system (X, T, μ) satisfies the exponential inequality (2.2), then the observed system satisfies an exponential concentration inequality. For any $n \geq 1$, it is given by*

$$\mathbb{E}_{\mu_n \otimes P^n}\left(e^{K(y_0,\ldots,y_{n-1})-\mathbb{E}_{\mu_n\otimes P^n}(K(y_0,\ldots,y_{n-1}))}\right) \leq e^{D(1+\varepsilon^2)\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2}, \tag{4.1}$$

Furthermore, if the system (X, T, μ) satisfies the polynomial concentration inequality (2.3) with integer moment $q \geq 2$, then the observed system satisfies a polynomial concentration inequality with the same moment. For any $n \geq 1$, it is given by

$$\mathbb{E}_{\mu_n \otimes P^n}\left(|K(y_0,\ldots,y_{n-1}) - \mathbb{E}_{\mu_n\otimes P^n}(K(y_0,\ldots,y_{n-1}))|^q\right) \leq D_q(1+\varepsilon)^q\Big(\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2\Big)^{q/2}. \tag{4.2}$$

Observe that one recovers the corresponding concentration inequalities for the original dynamical system when ε vanishes.

Remark 4.1.1. *Our proof works provided the noise process satisfies a concentration inequality. Indeed a bounded i.i.d process satisfies the exponential concentration inequality* (2.2) *(see e.g. [50]). It also satisfies* (2.3) *for all $q \geq 2$, see e.g. [10] for more details. Although we do not need independence, it is reasonable to model the observational perturbations in this manner. Nevertheless, one can slightly modify the proof to get the result valid for correlated perturbations.*

Proof of theorem 4.1.1. First let us fix the noise $\{\xi_j\}$ and let $\bar{\xi} := (\xi_0, \xi_1, \ldots, \xi_{n-1})$. Introduce the auxiliary observable

$$\widetilde{K}_{\bar{\xi}}(x_0,\ldots,x_{n-1}) := K(x_0+\varepsilon\xi_0,\ldots,x_{n-1}+\varepsilon\xi_{n-1}).$$

Since the noise is fixed, it is easy to see that $\mathrm{Lip}_j(\widetilde{K}_{\bar{\xi}}) = \mathrm{Lip}_j(K)$ for all j.
Notice that $\widetilde{K}_{\bar{\xi}}(x,\ldots,T^{n-1}x) = K(x+\varepsilon\xi_0,\ldots,T^{n-1}x+\varepsilon\xi_{n-1}) = K(y_0,\ldots,y_{n-1})$. Next we define the observable $F(\xi_0,\ldots,\xi_{n-1})$ of n variables on the noise, as follows,

$$F(\xi_0,\ldots,\xi_{n-1}) := \mathbb{E}_{\mu_n}(\widetilde{K}_{\bar{\xi}}(x,\ldots,T^{n-1}x)).$$

Observe that, $\mathrm{Lip}_j(F) \leq \varepsilon \mathrm{Lip}_j(K)$.
Now we prove inequality (4.1). Observe that is equivalent to prove the inequality for

$$\mathbb{E}_{\mu_n\otimes P^n}\left(e^{\widetilde{K}_{\bar{\xi}}(x,\ldots,T^{n-1}x)-\mathbb{E}_{\mu_n\otimes P^n}(\widetilde{K}_{\bar{\xi}}(x,\ldots,T^{n-1}x))}\right).$$

Adding and subtracting $\mathbb{E}_{\mu_n}(\widetilde{K}_{\bar{\xi}}(x,\ldots,T^{n-1}x))$ and using the independence between the noise and the dynamical system, we obtain that the expression above is equal to

$$\mathbb{E}_{P^n}\left(e^{F(\xi_0,\ldots,\xi_{n-1})-\mathbb{E}_{P^n}(F(\xi_0,\ldots,\xi_{n-1}))}\mathbb{E}_{\mu_n}\left(e^{\widetilde{K}_{\bar{\xi}}(x,\ldots,T^{n-1}x)-\mathbb{E}_{\mu_n}(\widetilde{K}_{\bar{\xi}}(x,\ldots,T^{n-1}x))}\right)\right).$$

Since in particular, i.i.d. bounded processes satisfy the exponential concentration inequality (see Remark 4.1.1 above), we may apply (2.2) first to the dynamical system and then to the noise, yielding

$$\mathbb{E}_{P^n}\left(e^{F(\xi_0,\ldots,\xi_{n-1})-\mathbb{E}_{P^n}(F(\xi_0,\ldots,\xi_{n-1}))}\mathbb{E}_{\mu_n}\left(e^{\widetilde{K}_{\overline{\xi}}(x,\ldots,T^{n-1}x)-\mathbb{E}_{\mu_n}(\widetilde{K}_{\overline{\xi}}(x,\ldots,T^{n-1}x))}\right)\right)$$
$$\leq e^{C\sum_{j=0}^{n-1}\mathrm{Lip}_j(\widetilde{K}_{\overline{\xi}})^2}e^{C'\sum_{j=0}^{n-1}\mathrm{Lip}_j(F)^2}\leq e^{D(1+\varepsilon^2)\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2},$$

where $D:=\max\{C,C'\}$.
Next, we prove inequality (4.2) similarly. We use the binomial expansion after the triangle inequality with $\mathbb{E}_{\mu_n}(\widetilde{K}_{\overline{\xi}}(x,\ldots,T^{n-1}x))$. Using the independence between the noise and the dynamics, we get

$$\begin{aligned}&\mathbb{E}_{\mu_n\otimes P^n}(|K(y_0,\ldots,y_{n-1})-\mathbb{E}_{\mu_n\otimes P^n}(K(y_0,\ldots,y_{n-1}))|^q)\\ &\leq\sum_{p=0}^{q}\binom{q}{p}\mathbb{E}_{\mu_n}(|\widetilde{K}_{\overline{\xi}}(x,\ldots,T^{n-1}x)-\mathbb{E}_{\mu_n}(\widetilde{K}_{\overline{\xi}}(x,\ldots,T^{n-1}x))|^p)\times\\ &\qquad\mathbb{E}_{P^n}\left(|F(\xi_0,\ldots,\xi_{n-1})-\mathbb{E}_{P^n}(F(\xi_0,\ldots,\xi_{n-1}))|^{q-p}\right).\end{aligned}\tag{4.3}$$

We proceed carefully using the polynomial concentration inequality. The terms corresponding to $p=1$ and $p=q-1$ have to be treated separately. For the rest we obtain the bound

$$\sum_{\substack{p=0\\p\neq1,q-1}}^{q}\binom{q}{p}C_p\Big(\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2\Big)^{p/2}\times C'_{q-p}\Big(\varepsilon^2\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2\Big)^{\frac{q-p}{2}}.$$

For the case $p=1$, we apply Cauchy-Schwarz inequality and (2.3) for $q=2$ to get

$$\mathbb{E}_{\mu_n}\left(|\widetilde{K}_{\overline{\xi}}(x,\ldots,T^{n-1}x)-\mathbb{E}_{\mu_n}(\widetilde{K}_{\overline{\xi}}(x,\ldots,T^{n-1}x))|\right)\leq\sqrt{C_2}\Big(\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2\Big)^{1/2}.$$

If $q=2$, we proceed in the same way for the second factor in the right hand side of (4.3). The case $p=q-1$ is treated similarly. Finally, putting this together and choosing adequately the constant D_q we obtain the desired bound. □

Next we obtain an estimate of deviation probability of the observable K from its expected value.

Corollary 4.1.1. *If the system (X,T,μ) satisfies the exponential concentration inequality, then for the observed system $\{y_i\}$, for every $t>0$ and for any $n\geq1$ we have,*

$$\mu_n\otimes P^n\big(|K(y_0,\ldots,y_{n-1})-\mathbb{E}_{\mu_n\otimes P^n}(K)|\geq t\big)\leq2\exp\left(\frac{-t^2}{4D(1+\varepsilon^2)\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2}\right).\tag{4.4}$$

If the system (X, T, μ) *satisfies the polynomial concentration inequality with moment* $q \geq 2$*, then the observed system satisfies for every* $t > 0$ *and for any* $n \geq 1$*,*

$$\mu_n \otimes P^n\big(|K(y_0, \ldots, y_{n-1}) - \mathbb{E}_{\mu_n \otimes P^n}(K)| > t\big) \leq \frac{D_q}{t^q}(1+\varepsilon)^q \left(\sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2\right)^{q/2}. \quad (4.5)$$

The proof is completely analogous to that of corollary 2.3.1, but using the concentration inequalities for perturbed systems instead of those of the original ones.

4.2 Applications

4.2.1 Dynamical systems

Concentration inequalities are available for the class of non-uniformly hyperbolic dynamical systems modeled by Young towers ([21]). Actually, systems with exponential tails satisfy an exponential concentration inequality and if the tails are polynomial then the system satisfies a polynomial concentration inequality. The examples given in section 2 are included in that class of dynamical systems. We refer the interested reader to [73] and [75] for more details on systems modeled by Young towers. Here we consider dynamical systems satisfying either the exponential or the polynomial concentration inequality. We apply our result of concentration in the setting of observed systems to empirical estimators of the auto-covariance function, the empirical measure, the kernel density estimator and the correlation dimension.

4.2.2 Auto-covariance function for Chaos plus noise

Consider the dynamical system (X, T, μ), a square integrable observable $f : X \to \mathbb{R}$. Assume that f is such that $\int f d\mu = 0$ and the perturbed itinerary $y_0, \ldots, y_{n-1}$. Define the perturbed empirical estimator of the auto-covariance function (2.11), as follows

$$\widetilde{\mathrm{Cov}}_n(k) := \frac{1}{n} \sum_{i=0}^{n-1} f(y_i) f(y_{i+k}). \quad (4.6)$$

We are interested in quantifying the influence of noise on the correlation. We give a bound on the probability of the deviation of the perturbed empirical estimator from the covariance function.

Theorem 4.2.1. *Let* $\widetilde{\mathrm{Cov}}_n(k)$ *be given by* (4.6)*. If the dynamical system* (X, T, μ) *satisfies the exponential inequality* (2.2) *then for all* $t > 0$ *and for any integer* $n \geq 1$ *we have*

$$\mu_n \otimes P^n\left(\left|\widetilde{\mathrm{Cov}}_n(k) - \mathrm{Cov}(k)\right| > t + 2a_f\varepsilon\right) \leq 2\exp\left(\frac{-t^2}{64Da_f^2(1+\varepsilon^2)}\left(\frac{n^2}{n+k}\right)\right) + 2\exp\left(\frac{-t^2}{16Ca_f^2}\left(\frac{n^2}{n+k}\right)\right),$$

where $a_f = \mathrm{Lip}(f)\|f\|_\infty$*,* C *and* D *are the constants appearing in* (2.2) *and* (4.1) *respectively. If the system satisfies the polynomial inequality with moment* $q \geq 2$*, then for all* $t > 0$ *and any integer* $n \geq 1$ *we have*

$$\mu_n \otimes P^n \left(\left| \widetilde{\mathrm{Cov}}_n(k) - \mathrm{Cov}(k) \right| > t + 2a_f\varepsilon \right) \leq (2^q D_q(1+\varepsilon)^q + C_q) \left(\frac{2a_f}{t} \right)^q \left(\frac{n+k}{n^2} \right)^{q/2},$$

where C_q *and* D_q *are the constants appearing in* (2.3) *and* (4.2) *respectively.*

Proof. To prove this assertion we will use an estimate of

$$\mu_n \otimes P^n \left(\left| \widetilde{\mathrm{Cov}}_n(k) - \widehat{\mathrm{Cov}}_n(k) \right| > t + \mathbb{E}_{\mu_n \otimes P^n} \left(\left| \widetilde{\mathrm{Cov}}_n(k) - \widehat{\mathrm{Cov}}_n(k) \right| \right) \right).$$

First let us write $x_i := T^i x$, and observe that by adding and subtracting $f(x_i + \varepsilon\xi_i)f(x_{i+k})$, the quantity $|\widetilde{\mathrm{Cov}}_n(k) - \widehat{\mathrm{Cov}}_n(k)|$ is less than or equal to

$$\frac{1}{n} \sum_{i=0}^{n-1} \left| f(x_i + \varepsilon\xi_i)[f(x_{i+k} + \varepsilon\xi_{i+k}) - f(x_{i+k})] + [f(x_i + \varepsilon\xi_i) - f(x_i)]f(x_{i+k}) \right|,$$

which leads us to the following estimate,

$$\mathbb{E}_{\mu_n \otimes P^n} \left(\left| \widetilde{\mathrm{Cov}}_n(k) - \widehat{\mathrm{Cov}}_n(k) \right| \right) \leq 2\varepsilon \mathrm{Lip}(f)\|f\|_\infty. \tag{4.7}$$

For a given realization of the noise $\{e_i\}$, consider the following observable of $n+k$ variables

$$K(z_0, \ldots, z_{n+k-1}) := \frac{1}{n} \sum_{i=0}^{n-1} \left(f(z_i + \varepsilon e_i) f(z_{i+k} + \varepsilon e_{i+k}) - f(z_i)f(z_{i+k}) \right).$$

For every $0 \leq l \leq n-1$, one can easily obtain that

$$\mathrm{Lip}_l(K) \leq \frac{4}{n} \mathrm{Lip}(f)\|f\|_\infty.$$

In the exponential case, from the inequality (4.4) and the bound (4.7) on the expected value of K, we obtain that

$$\mu_n \otimes P^n \left(\left| \widetilde{\mathrm{Cov}}_n(k) - \widehat{\mathrm{Cov}}_n(k) \right| > t + 2\varepsilon a_f \right) \leq 2 \exp \left(\frac{-t^2}{64 D a_f^2 (1+\varepsilon^2)} \left(\frac{n^2}{n+k} \right) \right).$$

Using proposition 2.4.1, a union bound and an adequate rescaling, we get the result. In order to prove the polynomial inequality, proceed similarly applying (4.5). □

4.2.3 Empirical measure

Consider the observed itinerary $y_0, \ldots, y_{n-1}$ and define the observed empirical measure by

$$\widetilde{\mathcal{E}}_n := \frac{1}{n}\sum_{i=0}^{n-1}\delta_{y_i}.$$

Observe that this measure is well defined on X. Again Birkhoff's ergodic theorem implies that almost surely

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1} g(y_i) = \int\int g(x+\xi)\mathrm{d}\mu(x)\mathrm{d}P(\xi),$$

for every continuous function g. More precisely, this convergence holds for a set of μ-measure one of initial conditions for the dynamical system (X,T) and a set of measure one of noise realizations (ξ_i) with respect to the product measure $P^{\mathbb{N}}$.

We want to estimate the speed of convergence of the observed empirical measure. As in the case without noise, we study the convergence rate of the empirical measure by the fluctuations of the Kantorovich distance of the empirical measure and μ around its expected value. In this case, we consider the Kantorovich distance of the observed empirical measure to the measure μ. The statement is the following.

Proposition 4.2.1. *If the system (X,T,μ) satisfies the exponential concentration inequality* (2.2)*, then for all $t>0$ and any integer $n\geq 1$,*

$$\mu_n\otimes P^n\left(\kappa(\widetilde{\mathcal{E}}_n,\mu) > t + \mathbb{E}_{\mu_n\otimes P^n}\left(\kappa(\widetilde{\mathcal{E}}_n,\mu)\right)\right) \leq e^{-\frac{t^2 n}{4D(1+\varepsilon^2)}}.$$

If the system satisfies the polynomial concentration inequality (2.3) *with moment $q\geq 2$, then for all $t>0$ and any integer $n\geq 1$,*

$$\mu_n\otimes P^n\left(\kappa(\widetilde{\mathcal{E}}_n,\mu) > t + \mathbb{E}_{\mu_n\otimes P^n}\left(\kappa(\widetilde{\mathcal{E}}_n,\mu)\right)\right) \leq \frac{D_q(1+\varepsilon)^q}{t^q}\frac{1}{n^{q/2}}.$$

Using the following separately Lipschitz function of n variables,

$$K(z_0,\ldots,z_{n-1}) := \sup_{g\in\mathcal{L}}\left[\frac{1}{n}\sum_{i=0}^{n-1} g(z_i) - \int g d\mu\right].$$

It is easy to check that $\mathrm{Lip}_j(K)\leq \frac{1}{n}$, for every $j=0,\ldots,n-1$. The proposition follows from the concentration inequalities (4.4) and (4.5).

As a consequence of the proposition 4.2.1 and the lemma 2.4.1, we obtain the following result.

Theorem 4.2.2. *Assume that the system (X,T,μ) satisfies the assumptions of lemma 2.4.1. Let $\widetilde{\mathcal{E}}_n$ be the observed empirical measure. If the system satisfies the exponential inequality* (2.2) *then there exists a $B>0$ for all $t>0$ and for all $n\geq 1$ we have that*

$$\mu_n\otimes P^n\left(\kappa(\widetilde{\mathcal{E}}_n,\mu) > \frac{t+B}{n^{1/4}} + \varepsilon\right) \leq e^{-\frac{t^2\sqrt{n}}{4D(1+\varepsilon^2)}}.$$

If the system satisfies the polynomial inequality (2.3) *with moment* $q \geq 2$*, then for all* $t > 0$ *and for all* $n \geq 1$ *we obtain*

$$\mu_n \otimes P^n \left(\kappa(\widetilde{\mathcal{E}}_n, \mu) > \frac{t+B}{n^{1/4}} + \varepsilon \right) \leq \frac{D_q(1+\varepsilon)^q}{t^q} \frac{1}{n^{q/4}}.$$

Proof. Clearly $\mathbb{E}_{\mu_n \otimes P^n}(\kappa(\widetilde{\mathcal{E}}_n, \mu)) \leq \mathbb{E}_{\mu_n \otimes P^n}(\kappa(\widetilde{\mathcal{E}}_n, \mathcal{E})) + \mathbb{E}_{\mu_n \otimes P^n}(\kappa(\mathcal{E}, \mu))$. A straightforward estimation yields

$$\mathbb{E}_{\mu_n \otimes P^n}(\kappa(\widetilde{\mathcal{E}}_n, \mathcal{E})) \leq \int \sup_{g \in \mathcal{L}} \left[\frac{1}{n} \sum_{i=0}^{n-1} \mathrm{Lip}(g) \varepsilon \|\xi_i\| \right] \mathrm{d}\mu_n \otimes P^n \ \leq \ \varepsilon.$$

We obviously have $\mathbb{E}_{\mu_n \otimes P^n}\left(\kappa(\mathcal{E}, \mu)\right) = \mathbb{E}_{\mu_n}\left(\kappa(\mathcal{E}, \mu)\right)$. Using the exponential estimate of proposition 4.2.1 and lemma 2.4.1 we obtain, for any $t > 0$,

$$\mu_n \otimes P^n \left(\kappa(\widetilde{\mathcal{E}}_n, \mu) \geq t + \varepsilon + \frac{B}{n^{1/4}} \right) \leq \exp\left(\frac{-t^2 n}{4D(1+\varepsilon^2)} \right).$$

Rescaling adequately we get the result. For the polynomial case, one uses the polynomial estimate of proposition 4.2.1. □

4.2.4 Kernel density estimator for one-dimensional maps plus noise

In section 2.4.3 we considered the system (X, T, μ) taking values on a bounded subset of $\mathbb{R}$, here, in order to avoid 'leaking' problems, we assume $X = \mathbb{S}^1$.

Given the observed trajectory $\{y_j\}$, let us define the perturbed empirical density estimator by

$$\widetilde{h}_n(y_0, \ldots, y_{n-1}; s) := \frac{1}{n\alpha_n} \sum_{j=0}^{n-1} \psi\left(\frac{s - y_j}{\alpha_n} \right).$$

where $\alpha_n \to 0$ and $n\alpha_n \to \infty$ as n diverges. Again, the function ψ is a kernel, which means that it is a bounded, non-negative Lipschitz function with bounded support satisfying $\int \psi(s) ds = 1$.

The result in the case of observed systems is the following.

Theorem 4.2.3. *If* (X, T, μ) *satisfies the hypothesis 2.4.1 and the exponential concentration inequality, then there exists a constant* $C_\psi > 0$ *such that, for all* $t > C_\psi \left(\alpha_n^\beta + \frac{1}{\sqrt{n}\alpha_n^2} \right)$ *and for any integer* $n \geq 1$*,*

$$\mu_n \otimes P^n \left(\int \left| \widetilde{h}_n(y_0, \ldots, y_{n-1}; s) - h(s) \right| \mathrm{d}s > t + \mathrm{Lip}(\psi) \frac{\varepsilon}{\alpha_n^2} \right) \leq \exp\left(-\frac{n\alpha_n^4 t^2}{R(1+\varepsilon^2)} \right),$$

where $R := 4D\mathrm{Lip}(\psi)^2$.
If the system satisfies the hypothesis 2.4.1 and the polynomial concentration inequality, then for all $t > C_\psi \left(\alpha_n^\beta + \frac{1}{\sqrt{n}\alpha_n^2} \right)$ *and for any integer* $n \geq 1$*, we have*

$$\mu_n \otimes P^n \left(\int \left| \widetilde{h}_n(y_0, \ldots, y_{n-1}; s) - h(s) \right| \mathrm{d}s > t + \mathrm{Lip}(\psi) \frac{\varepsilon}{\alpha_n^2} \right) \leq D_q \left(\frac{(1+\varepsilon)\mathrm{Lip}(\psi)}{t\sqrt{n}\alpha_n^2} \right)^q.$$

The parameter β is the same constant appearing as in the hypothesis 2.4.1.

Proof. Consider the following observable of n variables,

$$K(z_0,\dots,z_{n-1}) := \int \Big|\frac{1}{n\alpha_n}\sum_{j=0}^{n-1}\psi\left(\frac{s-z_j}{\alpha_n}\right) - h(s)\Big|\mathrm{d}s.$$

It is straightforward to obtain that $\mathrm{Lip}_l(K) \leq \frac{\mathrm{Lip}(\psi)}{n\alpha_n^2}$, for every $l = 0,\dots,n-1$. Next, we need to give an upper bound for the expected value of the observable K, first

$$\mathbb{E}_{\mu_n\otimes P^n}(K) \leq \int\Big(\int\Big|\frac{1}{n\alpha_n}\sum_{j=0}^{n-1}\Big[\psi\Big(\frac{s-y_j}{\alpha_n}\Big) - \psi\left(\frac{s-x_j}{\alpha_n}\right)\Big]\Big|\mathrm{d}s\Big)\mathrm{d}\mu_n\otimes P^n$$
$$+\int\Big(\int\Big|\frac{1}{n\alpha_n}\sum_{j=0}^{n-1}\psi\Big(\frac{s-x_j}{\alpha_n}\Big) - h(s)\Big|\mathrm{d}s\Big)\mathrm{d}\mu_n.$$

Subsequently we proceed on each part. For the first one we get

$$\int\Big(\int\Big|\frac{1}{n\alpha_n}\sum_{j=0}^{n-1}\Big[\psi\Big(\frac{s-y_j}{\alpha_n}\Big) - \psi\Big(\frac{s-x_j}{\alpha_n}\Big)\Big]\Big|\mathrm{d}s\Big)d\mu_n\otimes P^n$$
$$\leq \int\Big(\frac{1}{n\alpha_n}\sum_{j=0}^{n-1}\frac{\mathrm{Lip}(\psi)\varepsilon}{\alpha_n}\Big)\mathrm{d}\mu_n\otimes P^n \;\leq\; \mathrm{Lip}(\psi)\frac{\varepsilon}{\alpha_n^2}.$$

For the second part, there exist some constant C_ψ such that

$$\int\Big(\int\Big|\frac{1}{n\alpha_n}\sum_{j=0}^{n-1}\psi\Big(\frac{s-x_j}{\alpha_n}\Big) - h(s)\Big|\mathrm{d}s\Big)\mathrm{d}\mu_n \leq C_\psi\left(\alpha_n^\beta + \frac{1}{\sqrt{n}\alpha_n^2}\right).$$

The proof of this statement is found in [19, Section 6]. We finish the proof applying (4.4) and (4.5), respectively. □

4.2.5 Correlation dimension

Recall the definition of the correlation dimension and the spatial correlation integral given in section 2.4.4.

In the case of observed systems, let us consider the observed sequence $y_0,\dots,y_{n-1}$, and define the estimator of $\mathrm{Corr}(r)$ for observed systems, as follows

$$\widetilde{K}_{n,r}(y_0,\dots,y_{n-1}) := \frac{1}{n^2}\sum_{i\neq j}H(r-d(y_i,y_j)).$$

Once again, since $\widetilde{K}_{n,r}(y_0,\dots,y_{n-1})$ is not a Lipschitz function we cannot apply directly concentration inequalities. Then we apply the usual trick replacing H by a

Lipschitz continuous function ϕ and then define the new estimator

$$\widetilde{K}^{\phi}_{n,r}(y_0,\ldots,y_{n-1}) := \frac{1}{n^2}\sum_{i\neq j}\phi\left(1-\frac{d(y_i,y_j)}{r}\right). \tag{4.8}$$

The result of this section is the following estimate on the variance of the estimator $\widetilde{K}^{\phi}_{n,r}$.

Theorem 4.2.4. *Let ϕ be a Lipschitz continuous function. Consider the observed trajectory $y_0,\ldots,y_{n-1}$ and the function $\widetilde{K}^{\phi}_{n,r}(y_0,\ldots,y_{n-1})$ given by (4.8). If the system (X,T,μ) satisfies the polynomial concentration inequality with $q=2$, then for any integer $n\geq 1$,*

$$\mathrm{Var}(\widetilde{K}^{\phi}_{n,r}) \leq D_2 \mathrm{Lip}(\phi)^2(1+\varepsilon)^2\frac{1}{r^2 n},$$

where $\mathrm{Var}(Y) := \mathbb{E}(Y^2)-\mathbb{E}(Y)^2$ *is the variance of* Y.

The proof follows the lines of section 4 in [19], and by applying the inequality (4.2) with $q=2$ and noticing that $\mathrm{Lip}_l(\widetilde{K}^{\phi}_{n,r}) \leq \frac{\mathrm{Lip}(\phi)}{rn}$ for every $l=0,\ldots,n-1$.

Chapter 5

Signal Recovery from Chaos plus Noise

5.1 Noise reduction methods

In the context of time series analysis, an important problem is that of signal recovery (also called signal separation or noise reduction). Its importance relies in the fact that in practice every data signal is contaminated by some 'other unwanted sources' that we simply call *noise*. The problem of signal recovery can be stated as follows: we observe some signal which contains a noise component, but we want it as clean as possible. Is it possible to recover the signal of interest? and in the affirmative case, then how? The answer of the first question depends on the nature of the noise and the source of the signal of interest. And, for the 'how', many methods have been proposed. There is a large amount of bibliography on this problem, see for instance [2, 43, 66], and references therein.

The classical approach to separate the noise from the wanted signal is using frequency filters. That method is effectively applicable when the wanted signal has some kind of periodicity ([2]). In real experiments many times the signal of study is produced by some nonlinear source, making those filters not very efficient. In the nineties a great effort was done by physicists in order define better methods to recover a noised signal, making them applicable to nonlinear phenomena. We will briefly mention some of these methods.

To delimit the problem, first one has to fix:

a) The nature of the systems we observe.

b) The nature of the perturbing noise.

For the moment, assume that the observed signal is a discrete signal coming from a chaotic dynamics, since the methods we will describe in this section do not directly take advantage of the dynamics, chaoticity is the only required property. Later on we will make a more precise statement on the nature of the dynamics to be considered, in order to get some rigorous result on the consistency of the method. In order to fix the nature of the noise, recall the two types of noise defined in chapter 3. Given that dynamical noise interacts within the dynamics, it is more complicated to deal with. Moreover for standard computer simulation (no rigorous computations) of chaotic dynamical systems perturbed with a small amount of dynamical noise very often diverges in a few iterations (in the order of 10^3 iterations). Thus we restrict to the observational model of noise described in section 3.2. The existing implemented methods have been tested for Gaussian as well as identically distributed and bounded noise. In both cases numerical simulations exhibits effective cleansing of data (see for instance [35, 64, 65]).

Let us be more precise. Let $x_0, x_1, \ldots, x_n$ be the first n iterates of the initial condition x_0 under the chaotic map T (i.e. $x_i = T^i x_0$ for $i \geq 0$). Assume for the sake of definiteness that T is invertible. Commonly the state space is $\mathbb{R}^d$. For instance, the Hénon map (which is a map in $\mathbb{R}^2$, see example 1.3.2) is taken as the common example in most of the references cited in this chapter. Next, instead of the sequence $x_0, x_1, \ldots, x_n$ one 'observes' to the sequence $y_0, y_1, \ldots, y_n$, given by the relation

$$y_i := x_i + e_i.$$

Here e_i denotes the noise component which is independent of x_i and is distributed according to some probability distribution P. We further make some precise assumptions on the noise according to the method in turn. For example, most of the existing noise reduction methods were tested on systems with Gaussian noise, and with identically distributed and bounded noise, but only in the latter case rigorous results on the convergence of the algorithm are available.

With respect to the information used while implemented, existing methods may be classified in three different types:

1. The dynamics is known ([38, 31]).
2. A clean signal of the dynamical system is known and is used as reference to clean up the noisy signals provided they are produced by the same dynamical system ([55]).
3. Blind methods. There is no *a priori* knowledge of the dynamical system nor even a single clean signal is known. The methods in this case require further assumptions on the systems and on the noise. Their efficiency depends not only on the method but also on the system and on the noise. ([15, 63, 65, 35, 48]).

Obviously the methods in the third case are the most interesting ones, since in 'real' situations one does not know *a priori* the dynamics in a explicit way neither even a single clean orbit is provided. In the following we only consider blind methods.

5.1.1 Local projective maps

In this class of methods the geometry of the signal is taken into account (and is the most important part) to reduce the noise. They use the fact that noise spreads out the observations in all directions, in opposition to the map, that follows some precise direction. The idea of Cawley-Hsu [15] and Sauer [63], is that noise can be reduced by projecting the observations onto the subspace spanned by a suitable collection of singular vectors at each point on the 'perturbed attractor'. The subspace is defined by taking a local piece of the 'perturbed attractor'. A variation of this idea was implemented by Schreiber and Grassberger in [65] and Grassberger and his group, in [35]. They impose some constraints using also the time ordering of the time series. A complete survey of the methods can be found in [46] and for a numerical comparative study check [35]. The Hénon map is the common example treated by all these methods, but is not the only example, for instance, the Ikeda map[1], the Lorentz attractor and experimental data showing fractal behavior are also considered. Although the numerical results seem to be quite impressive, the consistency properties remain largely unknown for those methods and thus there is a lack of rigorous results.

5.1.2 Schreiber-Lalley method

In reference [64], Schreiber defined a simple algorithm for noise reduction. Independently, the same idea was used later by Lalley in [47]. That is why we call this method the 'Schreiber-Lalley method'. This algorithm is the only one that exploits directly the assumption on the chaoticity of the map and is not only based in the geometry of the data as the previously mentioned methods. There are several advantages of this algorithm with respect to the previous ones, the simplicity, the ease of implementation, no knowledge of the map or a clean signal is needed and on the top of that, up to now, is the only one for which rigorous results are available for its consistency.

The idea in [64] is the following: Each iteration y_i is replaced by the average value of this coordinate over points in a suitably chosen neighborhood. To define such a neighborhood, first fix positive integers, k and l and take the vector $w_i = (y_{i-k}, \ldots, y_{i+l})$. Further, choose a radius δ for the neighborhoods. For each value y_i find the set $A_i(\delta)$ of all neighbors y_j for which

$$\sup\{|y_{j-k} - y_{i-k}|, \ldots, |y_{j+l} - y_{i+l}|\} < \delta,$$

that is, all segments of the trajectory which are close during a time from k iterations in the past to l iterations in the future. Next, replace the coordinate y_i by its mean value in $A_i(\delta)$,

$$\widehat{x}_i := \frac{1}{|A_i(\delta)|} \sum_{j \in A_i(\delta)} y_j.$$

This method produces another time series which reduces magnitude of the noise component. This is done without using artificial assumptions on the geometry of the attractor but using the fact that points which remain close in time should be close enough in the state space, given the chaotic nature of the map.

[1]The Ikeda map is a 2-dimensional map whose iterates are defined by $x_{n-1} = 1 + u(x_n \cos t_n - y_n \sin t_n)$, $y_{n+1} = u(x_n \sin t_n + y_n \cos t_n)$, and $t_n = 0.4 - \frac{6}{1+x_n^2+y_n^2}$, where $u \in (0,1)$ is a parameter. For some values of u the Ikeda map has a chaotic attractor.

There are some minor differences between Schreiber and Lalley's implementations. Lalley fixes the radius of the neighborhoods up to 3 times the magnitude of the noise (which is assumed to be bounded and uniformly distributed). And the integers k and l are both made equal to some function of n the size of the sampled signal. The substantial difference is that Lalley is restricted to bounded i.i.d. noise, because he proved the consistency of the algorithm in that case. Furthermore, he proved that if we consider unbounded noise (even Gaussian), then it is impossible to consistently recover a clean signal. We need to recall the notion of homoclinic pair of points. A pair of points x, x' is said to be a homoclinic pair if they satisfy $\sum_{n=-\infty}^{\infty} \|T^n x - T^n x'\| < \infty$. Next, the negative result is stated as follows.

Theorem 5.1.1 ([47, 48, 49]). *If the density of the noise is Gaussian and the dynamical system admits strongly homoclinic pairs of points, then there is no sequence of functions* $\psi_n(y_{-n}, y_{-n+1}, \ldots, y_n)$ *such that for all orbits* $x_n = T^n x$,

$$\lim_{n\to\infty} \mathbb{P}(|\psi_n(y_{-n}, \ldots, y_n) - x| > t) = 0$$

for any $t > 0$.

Actually, this result is proved for a more general class of densities with unbounded support which includes the Gaussian case.

With this result, Lalley gives a clear distinction between the results achieved by heuristic implementations, such as those obtained by the 'german school', and the rigorous results.

In the following section we describe in a more precise manner the algorithm defined by Lalley and we give the results on its consistency.

5.2 Lalley's Algorithm for Recovery of a Signal

In the present section we assume the time series $y_0, \ldots, y_{n-1}$ $(n \geq 1)$ to be an observed system (see definition 3.2.1), given by

$$y_i = x_i + \varepsilon\xi_i, \tag{5.1}$$

where $x_i = T^i x$ is the orbit of x, given a chaotic dynamical system T with an attractor Λ and ξ_i is an observational noise satisfying to be independent and independent if x_0. We assume the noise absolute magnitude is bounded by the constant ε.

5.2.1 The smoothing algorithm

The algorithm takes as an input a time series $y_0, y_1, \ldots, y_{n-1}$ given by (5.1). It defines for every i an estimator $\widehat{x}_i$ of the true iterate of the dynamics x_i. So, the algorithm gives as output another time series $\{\widehat{x}_i\}$ called *recovered system* (or *cleaned signal*). The idea is to define the estimators by an average of a well chosen set points taken from the noisy time series $\{y_i\}$. The main steps of Lalley's algorithm are the following.

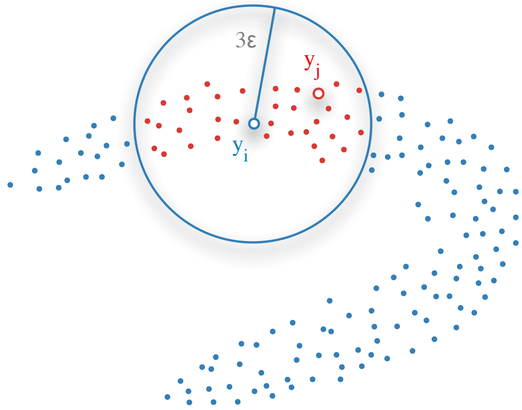


Figure 5.1: Schematic time series, the 3ε neighborhood of y_i is taken and all the y_j inside are pictured in red.

1. For each element y_i, look for all other points y_j in the time series that are close to y_i up to a distance equals to 3ε (see figure 5.1 below).

2. Define an integer $\tau \geq 1$. We further make assumptions on this number depending on the length of the sample.

3. For every $i = \tau, \ldots, n - \tau$, define the following set of indices

$$A_n(i, \tau) := \{j \ : \ |y_{j+r} - y_{i+r}| \leq 3\varepsilon \ \text{ for } \ |r| \leq \tau\}\,.$$

   Essentially this is the set of indices whose respective iterate y_j remains 'close' to y_i during τ iterations from the past and to the future (see figure 5.2).

4. Finally, define the estimator $\widehat{x}_i$ for every $i = \tau, \ldots, n - \tau$, as follows

$$\widehat{x}_i := \frac{1}{|A_n(i, \tau)|} \sum_{j \in A_n(i,\tau)} y_j, \tag{5.2}$$

   for the rest of the indices, define $\widehat{x}_i := y_i$.

5.2.2 Consistency of Lalley's algorithm

The consistency of the algorithm described above has been proved for Axiom A systems in [47], and for a more general class of dynamical systems in [48] and [49], assuming some degree of expansiveness. In those articles the authors proved consistency for almost every realization of i.i.d. noise independent of the map. Here we include the most general result available, for which we need to introduce the following definitions and a proposition.

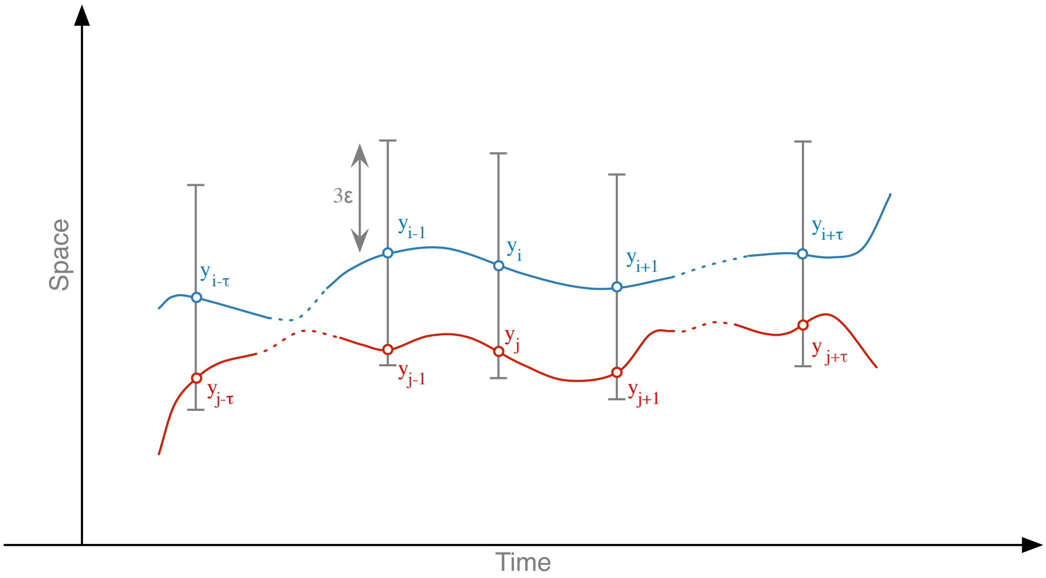


Figure 5.2: The itinerary of y_i and y_j are depicted. In this case y_j remains close to y_i in time, thus is contained in the set $A_n(i, \tau)$.

Definition 5.2.1. *A map $T : X \to X$ is said to be expansive if there exists $\Delta > 0$ such that for every pair of points $x, x' \in X$ with $x \neq x'$,*

$$\sup_{n \in \mathbb{Z}} |T^n x - T^n x'| > \Delta.$$

The constant Δ is called a separation threshold for T.

Definition 5.2.2. *The horizon of separation $H(\alpha)$ of the map T with separation threshold Δ, is a function $H_T(\alpha) : (0, \infty) \to \mathbb{N}_0$ that for all $\alpha > 0$ is defined by*

$$H_T(\alpha) := \sup\{s(x, x') \ : \ |x - x'| \geq \alpha\},$$

where $s(x, x') := \min\{|s| \ : \ |T^s(x) - T^s(x')| > \Delta\}$, is the minimal time of separation. Furthermore, T has a finite horizon of separation is for all $\alpha > 0$ is satisfies that $H_T(\alpha) < +\infty$.

Proposition 5.2.1. *If the map T has a finite horizon of separation then the inverse function $H^{-1}(n)$ tends monotonically to zero when n diverges.*

Proof. The monotonicity of H^{-1} follows from that of the function H itself. If $H^{-1}(n) \geq \alpha_0 > 0$ for all n, then for some $\alpha < \alpha_0$ we have that $H(\alpha) > H(\alpha_0)$, and given that $H(\alpha_0) \geq H(H^{-1}(n)) = n$ for all n, then we have that $H(\alpha) = \infty$. $\square$

The result on the consistency of the algorithm is the following.

Theorem 5.2.1 ([49]). *Let T be a expansive map with separation threshold $\Delta > 0$ and finite horizon of separation. Assume $\varepsilon \leq \Delta/5$. If $\tau \to \infty$ and $\tau/\log n \to 0$, then*

$$\frac{1}{n-2\tau}\sum_{i=\tau}^{n-\tau}|\widehat{x}_{i,n} - x_i| \to 0 \quad \text{when } n \to \infty$$

with probability one for every initial condition x in the basin of attraction of the map.

The idea is to use the assumption that the noise is observational, and so, the observed system is $y_j = x_j + \varepsilon\xi_j$. Write

$$\widehat{x}_i = \frac{1}{|A_n(i,\tau)|}\sum_{j\in A_n(i,\tau)} y_j = \frac{1}{|A_n(i,\tau)|}\sum_{j\in A_n(i,\tau)} x_j + \frac{1}{|A_n(i,\tau)|}\sum_{j\in A_n(i,\tau)} \varepsilon\xi_j, \tag{5.3}$$

then it is sufficient to have for every i that the cardinality of $A_n(i,\tau)$ is large enough, and use the assumption of the 'chaoticity' of the system which assures that if $j \in A_n(i,\tau)$ then $|x_i - x_j|$ is small. This is done by means of the proposition 5.2.1. Next, one uses a type of large numbers law to prove that $\frac{1}{|A_n(i,\tau)|}\sum_{j\in A_n(i,\tau)}\varepsilon\xi_j$ is small, when $n \to \infty$. Observe that this is not immediate because $A_n(i,\tau)$ is a stochastic set depending on the random variables ξ_j. The trick is to rearrange the indices in a proper way such that the summing set is independent of the ξ's.

Proof of theorem 5.2.1. Let us write the difference $|\widehat{x}_i - x_i|$ using the definition of the estimates.

$$\begin{aligned}|\widehat{x}_i - x_i| =& \Big|x_i - \frac{1}{|A_n(i,\tau)|}\sum_{j\in A_{n(i,\tau)}} y_j\Big| \\ \leq& \frac{1}{|A_n(i,\tau)|}\sum_{j\in A_n(i,\tau)}|x_i - x_j| + \frac{1}{|A_n(i,\tau)|}\Big|\sum_{j\in A_n(i,\tau)}\varepsilon\xi_j\Big|.\end{aligned} \tag{5.4}$$

Note that the first term in the right hand side of the inequality above controls the bias of the estimator and the second term controls the stochastic variation.

Next we proceed by proving that both terms in the above inequality goes to zero. For the first term we make use of the following lemma.

Lemma 5.2.1. *Let $\tau \leq n$ be such that $\tau = \tau(n) \to \infty$ as $n \to \infty$, then*

$$\frac{1}{|A_n(i,\tau)|}\sum_{j\in A_n(i,\tau)}|x_i - x_j| \to 0 \quad \text{when} \quad n \to \infty.$$

In order to proof the convergence to zero of the second term in (5.4) it will be useful to introduce some additional notation.

For every $m \geq 1$ and $1 \leq \tau \leq \frac{n}{2}$ define the set

$$\mathcal{I}_n(m,\tau) = \{i \ : \ |A_n(i,\tau)| \geq m \ \text{ and } \ \tau \leq i \leq n-\tau\}.$$

This set collects those indices whose respective set A_n has at least m elements. In other words, if $k \in \mathcal{I}$ then the estimate $\widehat{x}_k$ is defined by an average of at least m entries y_j. Once introduced the set $\mathcal{I}_n$ we use the following small lemma.

Lemma 5.2.2. *For every fixed* $\beta \in (\frac{1}{2}, 1)$*, one has that*

$$\frac{1}{n-2\tau}\sum_{i=\tau}^{n-\tau}\frac{1}{|A_n(i,\tau)|}\Big|\sum_{j\in A_n(i,\tau)}\varepsilon\xi_j\Big| \leq \max_{j\in\mathcal{I}_n(n^\beta,\tau)}\frac{1}{|A_n(i,\tau)|}\Big|\sum_{j\in A_n(i,\tau)}\varepsilon\xi_j\Big| + \frac{\Delta}{5(n-2\tau)}\sum_{i=\tau}^{n-\tau}\mathbb{1}\big\{|A_n(i,\tau)| \leq n^\beta\big\}.$$

Observe that the previous lemma only rewrites the second summand in (5.4) and splits it into two parts. In order to prove the theorem it suffices to show that both terms above vanishes as n goes to infinity. This is indeed true and it is given by the following two lemmata. First, for the case of the set of indices whose cardinality is 'large' we have a precise upper bound.

Lemma 5.2.3. *Consider a real number* $t > 0$*, if* $tm \geq \frac{2\Delta}{5}(2\tau+1)$*, then*

$$\mathbb{P}\left\{\max_{i\in\mathcal{I}_n(m,\tau)}|V_n(i,\tau)| > t\right\} \leq 2n\exp\left(\frac{-t^2m^2}{2n\left(\frac{\Delta}{5}\right)^2(2\tau+1)^2}\right),$$

where $|V_n(i,\tau)| = \frac{1}{|A_n(i,\tau)|}\left|\sum_{j\in A_n(i,\tau)}\varepsilon\xi_j\right|$.

And for those sets of indices with 'small' cardinality we use the following lemma.

Lemma 5.2.4. *If* $\tau(n) = o(\log n)$ *then for every* $\gamma > 0$*, one has*

$$\frac{1}{n}\sum_{j=0}^{n-1}\mathbb{1}\left\{|A_n(i,\tau)| \leq n^{1-\gamma}\right\} \to 0 \quad \textit{when} \;\; n \to \infty.$$

We continue by taking $m = n^\beta$ in the bound given in lemma 5.2.3, yielding

$$\mathbb{P}\left\{\max_{i\in\mathcal{I}_n(n^\beta,\tau)}|V_n(i,\tau)| > t\right\} \leq 2n\exp\left(\frac{-t^2n^\eta}{2\left(\frac{\Delta}{5}\right)^2(2\tau+1)^2}\right),$$

with $\eta = 2\beta - 1$. If $\xi > 0$ then we apply Borel-Cantelli lemma assuring that

$$\max_{i\in\mathcal{I}_n(n^\beta,\tau)}|V_n(i,\tau)| \to 0$$

with probability one, as $n \to \infty$. This is satisfied whenever $\beta > 1/2$, and that finishes the proof of the theorem. □

5.2.3 Proofs of lemmas

Here we collect the proofs of the previous lemmas.

Proof of lemma 5.2.1. Since $\varepsilon \leq \frac{\Delta}{5}$ by hypothesis then $A_n(i,\tau) \subset \{j \ : \ |y_{j+r} - y_{i+r}| \leq \frac{3\Delta}{5} \text{ for } |r| \leq \tau\}$. By definition we have that $j \in A_n(i,\tau)$ implies that $|y_{j+r} - y_{i+r}| \leq \frac{3\Delta}{5}$ for every $|r| \leq \tau$, and so

$$|x_{j+r} - x_{i+r}| \leq |y_{j+r} - y_{i+r}| + |\varepsilon\xi_{i+r} - \varepsilon\xi_{j+r}| \leq \frac{3\Delta}{5} + \frac{2\Delta}{5} = \Delta,$$

which means that the time of separation of the pair x_i, x_j is at least equal to τ (i.e. $s(x_i, x_j) \geq \tau$). In particular,

$$\sup\{s(x,x') \ : \ |x - x'| \geq |x_i - x_j|\} \geq \tau,$$

that is, $H(|x_i - x_j|) \geq \tau$, then $H^{-1}(\tau) \geq |x_i - x_j|$ for every $j \in |A_n(i,\tau)|$. This allows us to write

$$\frac{1}{|A_n(i,\tau)|} \sum_{j \in A_n(i,\tau)} |x_i - x_j| \leq \frac{1}{|A_n(i,\tau)|} \sum_{j \in A_n(i,\tau)} H^{-1}(\tau) = H^{-1}(\tau).$$

Finally, we use the proposition 5.2.1 and that finishes the proof. □

Proof of lemma 5.2.2. For n and τ fixed, we can rearrange the indices into two sets: those indices i for which $|A_n(i,\tau)| > n^\beta$ (that is $i \in \mathcal{I}_n(n^\beta, \tau)$) and those that belong to the complement. Thus we have

$$\sum_{j \in A_n(i,\tau)} \varepsilon\xi_j = \sum_{j \in A_n(i,\tau)} \varepsilon\xi_j \left(\mathbb{1}\left\{|A_n(i,\tau)| > n^\beta\right\} + \mathbb{1}\left\{|A_n(i,\tau)| \leq n^\beta\right\} \right).$$

Consider first the contribution of the noise from those indices in $\mathcal{I}_n(n^\beta, \tau)$.

$$\frac{1}{|A_n(i,\tau)|} \Big| \sum_{j \in A_n(i,\tau)} \varepsilon\xi_j \Big| \mathbb{1}\left\{|A_n(i,\tau)| > n^\beta\right\} \leq \max_{i \in \mathcal{I}_n(n^\beta,\tau)} \frac{1}{|A_n(i,\tau)|} \Big| \sum_{j \in A_n(i,\tau)} \varepsilon\xi_j \Big|.$$

For the case of those indices belonging to the complement of $\mathcal{I}_n$ we have the following trivial bound, using only that $\varepsilon \leq \frac{\Delta}{5}$ and $\|\xi_j\| \leq 1$, for all j.

$$\frac{1}{|A_n(i,\tau)|} \Big| \sum_{j \in A_n(i,\tau)} \varepsilon\xi_j \Big| \mathbb{1}\left\{|A_n(i,\tau)| \leq n^\beta\right\} \leq \frac{\Delta}{5} \cdot \mathbb{1}\{|A_n(i,\tau)| \leq n^\beta\}.$$

Taking the average over all the indices $\tau \leq i \leq n - \tau$, gives us the desired expression. □

Proof of lemma 5.2.4. The lemma only makes use of the compactness of Λ. There exist a finite set $S \subset \Lambda$ such that

$$\max_{u\in\Lambda}\min_{v\in S}|u-v| \leq \frac{\Delta}{10}.$$

That defines a $\left(\frac{\Delta}{10}\right)$-net. Let $S^{2\tau+1}$ be the set of all sequences $\underline{s} = (s_{-\tau}, \ldots, s_{\tau})$, where $s_i \in S$, for every i. For every $x \in \Lambda$ there exists a sequence $\underline{s} \in S^{2\tau+1}$ that codifies the $2\tau+1$ steps of the itinerary of x, by satisfying

$$\max_{|i|\leq\tau}|T^i(x) - s_i| \leq \frac{\Delta}{10}.$$

For every $\underline{s} \in S^{2\tau+1}$, define the set

$$J_n(\underline{s}) = \left\{ j \ : \ 0 \leq j \leq n \quad \text{and} \quad \max_{|i|\leq\tau}|T^{i+j}(x) - s_i| \leq \frac{\Delta}{10} \right\}.$$

In words, $J_n(\underline{s})$ contains all the indices that have the same code $\underline{s}$. Observe that every $j = \tau, \ldots, n-\tau$, is contained in at least one set $J_n(\underline{s})$. Moreover, if $j_1, j_2 \in J_n(\underline{s})$, then

$$\max_{|i|\leq\tau}|x_{j_1+i} - x_{j_2+i}| \leq \frac{\Delta}{5} \quad \text{and} \quad \max_{|i|\leq\tau}|y_{j_1+i} - y_{j_2+i}| \leq \frac{3\Delta}{5}.$$

Therefore $j_1 \in A_n(j_2,\tau)$ and $j_2 \in A_n(j_1,\tau)$. Taking $j \in J_n(\underline{s})$, for every $j' \in J_n(\underline{s})$ we have that $j' \in A_n(j,\tau)$, it follows that if $j \in J_n(\underline{s})$ then $|J_n(\underline{s})| \leq |A_n(j,\tau)|$. Fix $0 < \gamma < 1$, then

$$\sum_{j=0}^{n-1} \mathbb{1}\left\{|A_n(j,\tau)| \leq n^{1-\gamma}\right\} \leq \sum_{j=0}^{n-1}\sum_{\underline{s}} \mathbb{1}\left\{|A_n(j,\tau)| \leq n^{1-\gamma}\right\} \cdot \mathbb{1}\left\{j \in J_n(\underline{s})\right\}.$$

Since $|A_n(j,\tau)| \leq n^{1-\gamma}$ and $j \in J_n(\underline{s})$ then $|J_n(\underline{s})| \leq n^{1-\gamma}$, and so,

$$\begin{aligned}\sum_{j=0}^{n-1} \mathbb{1}\left\{|A_n(j,\tau)| \leq n^{1-\gamma}\right\} &\leq \sum_{j=0}^{n-1}\sum_{\underline{s}} \mathbb{1}\left\{|J_n(\underline{s})| \leq n^{1-\gamma}\right\} \cdot \mathbb{1}\left\{j \in J_n(\underline{s})\right\} \\ &\leq \sum_{\underline{s}}|J_n(\underline{s})| \cdot \mathbb{1}\left\{|J_n(\underline{s})| \leq n^{1-\gamma}\right\} \\ &\leq n^{1-\gamma}|S^{2\tau+1}|.\end{aligned}$$

Choose $\tau(n) = \frac{\gamma}{4}\frac{\log n}{\log|S|}$ to obtain $|S^{2\tau+1}| = |S|n^{\gamma/2} = o(n^{\gamma/2})$. Then,

$$\sum_{j=0}^{n-1} \mathbb{1}\left\{|A_n(j,\tau)| \leq n^{1-\gamma}\right\} \leq |S|n^{1-\gamma/2} = o(n).$$

Therefore $\frac{1}{n}\sum_{j=0}^{n-1} \mathbb{1}\left\{|A_n(j,\tau)| \leq n^{1-\gamma}\right\} \to 0$, when $n \to \infty$, and this concludes. $\square$

Proof of lemma 5.2.3. For convenience we introduce some new notation. Define $U_n(i,\tau) := \sum_{j\in A_n(i,\tau)} \varepsilon\xi_j$. Observe that U_n is a sum of random variables over the sets A_n which are not independent from ξ_j. In view of the dependence of A_n on the ξ's, McDiarmid's theorem can not be applied directly. The idea is to modify the sets and arrange them in the proper way in order to may use the McDiarmid's theorem (see e.g. [50]).

We make use of the following sub-lemma.

Lemma 5.2.5. *If $H(\Delta/5) \le \tau < n/2$, for $i \in [\tau, n-\tau]$, then*

$$U_n(i,\tau) = \sum_{j=\tau}^{n-\tau} \varepsilon\xi_j \mathbb{1}\left\{|x_j - x_i| \le \frac{\Delta}{5}\right\} \prod_{1\le|s|\le\tau} \mathbb{1}\left\{|y_{i+s} - y_{j+s}| \le \frac{3\Delta}{5}\right\}. \tag{5.5}$$

Proof. By definition of the set $A_n(i,\tau)$, one can write

$$U_n(i,\tau) = \sum_{j\in A_n(i,\tau)} \varepsilon\xi_j = \sum_{j=\tau}^{n-\tau} \varepsilon\xi_j \prod_{|s|\le\tau} \mathbb{1}\left\{|y_{j+s} - y_{i+s}| \le \frac{3\Delta}{5}\right\}.$$

On one hand, if $|x_i - x_j| \le \Delta/5$ then it is clear that $|y_j - y_i| \le 3\Delta/5$, and therefore

$$\prod_{|s|\le\tau} \mathbb{1}\left\{|y_{j+s} - y_{i+s}| \le \frac{3\Delta}{5}\right\} = \mathbb{1}\left\{|x_i - x_j| \le \frac{\Delta}{5}\right\} \prod_{1\le|s|\le\tau} \mathbb{1}\left\{|y_{j+s} - y_{i+s}| \le \frac{3\Delta}{5}\right\}.$$

On the other hand, if $|x_i - x_j| > \frac{\Delta}{5}$ and suppose that $\prod_{1\le|s|\le\tau} \mathbb{1}\left\{|y_{j+s} - y_{i+s}| \le \frac{3\Delta}{5}\right\} = 1$, then $|x_{i+s} - x_{j+s}| \le \Delta$ for $|s| \le \tau$, which implies that $s(x_i, x_j) \ge \tau$.
That is $H(|x_i - x_j|) \ge \tau$, which is a contradiction, since $H(|x_i - x_j|) < H(\frac{\Delta}{5}) \le \tau$ by hypothesis. This establishes the equality above. □

Using the equation (5.5), we define the following quantity, which does not take into account those indices j that make $A_n(i,\tau)$ depend on ξ_j. Let us define

$$\widetilde{U} = \left(\sum_{j=\tau}^{i-\tau-1} + \sum_{j=i+\tau+1}^{n-\tau}\right) \varepsilon\xi_j \cdot \mathbb{1}\left\{|x_i - x_j| \le \frac{\Delta}{5}\right\} \prod_{1\le|s|\le\tau} \mathbb{1}\left\{|y_{i+s} - y_{j+s}| \le \frac{3\Delta}{5}\right\},$$

that is, $\widetilde{U}$ excludes the indices $j = i-\tau, \dots, i+\tau$. When $i \le 2\tau$ the first sum is equal to zero and when $i \ge n - 2\tau$, then the second sum is equal to zero. Then $|U_n(i,\tau) - \widetilde{U}| \le (2\tau+1)\frac{\Delta}{5}$, and as ξ_j is independent of the other products in the j-th summand, $\mathbb{E}(\widetilde{U}) = 0$.

Suppose for the moment that the values $\xi_{j-\tau}, \dots, \xi_{i+\tau}$ are fixed. This implies that $y_{i-\tau}, \dots, y_{i+\tau}$ are fixed as well. Now $\widetilde{U}$ is a function of $n - 4\tau - 1$ independent random variables $\Xi := \{\xi_j \ : \ j = \tau, \dots, i-\tau-1, i+\tau+1, \dots, n-\tau\}$. Identify $\widetilde{U} = f(\Xi)$, clearly $\mathrm{Lip}_j(\widetilde{U}) \le (2\tau+1)\frac{\Delta}{5}$. Note that $\mathbb{E}(\widetilde{U} \mid \xi_{i-\tau}^{i+\tau}) = \mathbb{E}(\widetilde{U}) = 0$.

Now we are able to use the McDiarmid's inequality ([50]), obtaining for every $t > 0$,

$$\mathbb{P}\left\{|\widetilde{U}| > t \mid \xi_{i-\tau}^{i+\tau}\right\} \leq 2\exp\left(\frac{-2t^2}{n\left(\frac{\Delta}{5}\right)^2(2\tau+1)^2}\right).$$

By definition of conditional probability and using the independence of the ξ's, one has

$$\mathbb{P}\{|\widetilde{U}| > t\} = \int \mathbb{P}\{|\widetilde{U}| > t \mid \xi_{i-\tau}^{i+\tau}\}\mathrm{d}\mathbb{P}(\xi_{i-\tau}^{i+\tau}) \leq 2\exp\left(\frac{-2t^2}{n\left(\frac{\Delta}{5}\right)^2(2\tau+1)^2}\right). \tag{5.6}$$

From the fact that $|U_n(i,\tau) - \widetilde{U}| \leq (2\tau+1)\frac{\Delta}{5}$, we obtain that

$$\mathbb{P}\left\{|U_n(i,\tau)| > t\right\} \leq \mathbb{P}\left\{\left|\widetilde{U}\right| > t - (2\tau+1)\frac{\Delta}{5}\right\}.$$

Let $t' = t - (2\tau+1)\frac{\Delta}{5}$ and use (5.6), then

$$\begin{aligned}\mathbb{P}\left\{|U_n(i,\tau)| > t\right\} \leq \mathbb{P}\left\{|\widetilde{U}| > t'\right\} &\leq 2\exp\left(\frac{-2\left(t-(2\tau+1)\frac{\Delta}{5}\right)^2}{n\left(\frac{\Delta}{5}\right)^2(2\tau+1)^2}\right)\\ &\leq 2\exp\left(\frac{-2t^2}{n\left(\frac{\Delta}{5}\right)^2(2\tau+1)^2} + \frac{4t}{n\frac{\Delta}{5}(2\tau+1)}\right),\end{aligned}$$

in particular, for $t \geq \frac{8\Delta}{15}(2\tau+1)$, clearly

$$\mathbb{P}\left\{|U_n(i,\tau)| > t\right\} \leq 2\exp\left(\frac{-t^2}{2n\left(\frac{\Delta}{5}\right)^2(2\tau+1)^2}\right). \tag{5.7}$$

Recalling the definition of $\mathcal{I}_n(m,\tau)$, if $i \in \mathcal{I}_n(m,\tau)$ then $|A_n(i,\tau)| \geq m$, and so we have

$$\mathbb{P}\left\{\max_{i\in\mathcal{I}_n(m,\tau)}|V_n(i,\tau)| > t\right\} \leq \mathbb{P}\left\{\max_{i\in\mathcal{I}_n(m,\tau)}|U_n(i,\tau)| > tm\right\}.$$

Using classical bounds we obtain,

$$\begin{aligned}\mathbb{P}\left\{\max_{i\in\mathcal{I}_n(m,\tau)}|U_n(i,\tau)| > tm\right\} &\leq \sum_{i\in\mathcal{I}_n(m,\tau)}\mathbb{P}\left\{|U_n(i,\tau)| > tm\right\}\\ &\leq n\cdot\max_{i\in\mathcal{I}_n(m,\tau)}\mathbb{P}\left\{|U_n(i,\tau)| > tm\right\}.\end{aligned}$$

Finally, for $tm > \frac{8\Delta}{15}(2\tau+1)$ applying (5.7) yields

$$\mathbb{P}\left\{\max_{i\in\mathcal{I}_n(m,\tau)}|V_n(i,\tau)| > t\right\} \leq 2n\exp\left(\frac{-t^2m^2}{2n\left(\frac{\Delta}{5}\right)^2(2\tau+1)^2}\right),$$

which finishes the proof of the lemma 5.2.3. □

5.3 Convergence rate of Lalley's Algorithm for Axiom A diffeomorphisms

The rate of convergence of Lalley's algorithm for denoising time series can be explicitly given at least for the case of systems whose expanding rate is uniform. In this section we give the proof of J. Stover ([69]) for the rate of convergence of the consistent signal recovered by Lalley's algorithm in the special case of Axiom A diffeomorphisms.

The result for the rate of convergence is the following.

Theorem 5.3.1 ([69]). *Let the time series $\{x_i\}$ be generated by an Axiom A diffeomorfism, and $\widehat{x}_i$ be the estimator* (5.2) *defined by Lalley's algorithm, if $\tau = \tau(n) = o(\log n)$ then for any $\alpha \in (0, \frac{1}{2})$,*

$$\mathbb{P}\left\{\max_{\tau \le i \le n-\tau} |\widehat{x}_i - x_i| > \frac{1}{n^\alpha} \quad \text{i.o.}\right\} = 0.$$

The abbreviation i.o. stands for 'infinitely often'.

Remark 5.3.1. *The proof shows that $\mathbb{P}\{\max_{\tau \le i \le n-\tau} |\widehat{x}_i - x_i| > n^{-\alpha}\}$ is bounded by stretched exponential in n for n large enough.*

The proof of this theorem is based in the following three lemmas. The first one uses the uniform hyperbolicity of the system, roughly said, two close point will separate exponentially in time.

Lemma 5.3.1. *For Axiom A systems, given $x_i = T^i x$ there exists a constant $C > 0$ such that if $j \in A_n(i, \tau)$, then*

$$|x_i - x_j| \le e^{-C\tau}.$$

The second one is an argument of compactness, is just the same as in lemma 5.2.4 (its proof was already given in the previous section).

Lemma 5.3.2. *For every $\beta > 0$, all sufficiently large n and all integers $i = \tau, \dots, n - \tau$,*

$$\mathbb{P}\{|A_n(i, \tau)| \le n^{1-4\beta}\} \le e^{-n^\beta}.$$

The third lemma gives a control on the deviation probability of the stochastic terms averaged over the set of indices $A_n(i, \tau)$. Let us recall the notation $V_n(i, \tau) := \frac{1}{|A_n(i,\tau)|}\sum_{j \in A_n(i,\tau)} \varepsilon\xi_j$, which is the second summand in (5.3).

Lemma 5.3.3. *For any $\alpha \in (0, \frac{1}{2})$, for $i = \tau, \dots, n - \tau + 1$ and β sufficiently small, one has*

$$\mathbb{P}\left\{\max_{\tau \le i \le n-\tau+1} |V_n(i, \tau)| > \frac{1}{n^\alpha}\right\} \le 2n(2\tau + 1)\exp\left(-\frac{n^{1-2\beta-2\alpha}}{2(2\tau+1)^2\varepsilon^2}\right) + n\varepsilon(2\tau + 1)e^{-n^{\beta/4}}$$

The proof of this lemma is similar to that of lemma 5.2.3, with a little difference. We will give it after the proof of the theorem.

Proof of Theorem 5.3.1. The proof is in essence as that of theorem 5.2.1. We start with the inequality (5.4), as in the case of the proof of theorem 5.2.1. For the case of Axiom A systems we use the uniform hyperbolicity of the map given by the lemma 5.3.1. Combining with lemma 5.3.2 one can get that for any $x_i \in \Lambda$ and $j \in A_n(i,\tau)$, one has

$$\mathbb{P}\Big\{ \max_{\tau \le i \le n-\tau+1} \frac{1}{|A_i|}\Big| \sum_{j \in A_i} x_i - x_j \Big| > \frac{1}{n^{1-4\beta}} \quad \text{i.o.} \Big\} = 0.$$

Next, using lemma 5.3.3, for all $\alpha \in (0, \frac{1}{2})$ we have that the probability

$$\mathbb{P}\Big\{ \max_{\tau \le i \le n-\tau+1} |V_n(i,\tau)| > \frac{1}{n^{\alpha}} \Big\}$$

is absolutely summable, thus applying Borel-Cantelli lemma the proof is finished.

□

Proof of lemma 5.3.3. The proof of this lemma consists in finding independent sets (of indices) and to separate those with large cardinality from those with small cardinality. Let us write A_i instead of $A_n(i,\tau)$ since n and therefore τ are kept fix until the end of the proof.

Next, for each i define A_i^* as the set of indices j such that $|i-j| \le 2\tau$. Note that $|A_i^*| \le 4\tau + 1 = o(\log n)$, so when $|A_i| > n^{1-\beta}$, for some $\beta < 1$, then the indices in A_i^* do not affect the average $\frac{1}{|A_i|} \sum_{j \in A_i} \varepsilon \xi_j$.

For every i and each $j = 1, \ldots, 2\tau+1$ define A_i^j as the set of indices l such that $l \cong j$ (mod $2\tau + 1$), that is $l \notin A_i^*$. The sets $A_i^*, A_i^1, \ldots, A_i^{2\tau+1}$ are pairwise disjoint and

$$A_i = A_i^* \cup \left(\bigcup_{j=1}^{2\tau+1} A_i^j \right).$$

The event $\{l \in A_i^j\}$ is uniquely determine by y_{i+r} and y_{l+r} for $|r| \le \tau$. The set A_i^j define a collection of independent vectors ξ_l if $l \in A_i^j$. The event $\{l \in A_i^j\}$ is neither affected by the value of ξ_l, since if $|y_{l+r} - y_{i+r}| < 3\varepsilon$ where $l \in A_i^j$, then $|y_l - y_i| < 3\varepsilon$ no matters the value of ξ_l and ξ_i given that the magnitude of the noise is ε.

For each i, we construct a partition of the sets A_i^j in $\mathcal{I}$ and $\mathcal{J}$, where $\mathcal{I}$ contains all the indices $*$ and j such that $|A_i^j| < n^{1-\beta}$ and $\mathcal{J}$ contains the rest. If $l \in A_i^j$, then A_i^j and ξ_l are independent one each other, as we saw, then we may use the following lemma.

Lemma 5.3.4 (Hoeffding's inequality [11])**.** *Let $S_n = \zeta_1 + \ldots + \zeta_n$ be the sum of n independent bounded random variables such that ζ_i falls in the interval $[a_i, b_i]$ with probability one. Then for any $t > 0$ we have*

$$\mathbb{P}\{S_n - \mathbb{E}(S_n) \ge t\} < \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2} \right).$$

Applying this lemma to $\zeta_i = \varepsilon\xi_i$, and taking $t = \frac{n^{1-\alpha}}{2\tau+1}$, for $\alpha > 0$, for any $j \in \mathcal{J}$ we obtain

$$\mathbb{P}\Big\{\Big|\sum_{j\in A_i^j} \varepsilon\xi_j\Big| > \frac{n^{1-\alpha}}{2\tau+1} \;\Big|\; |A_i^j| > n^{1-\beta}\Big\} \leq 2\exp\left(-\frac{n^{1-2\alpha+\beta}}{2(2\tau+1)^2\varepsilon^2}\right). \tag{5.8}$$

Next, when $|A_i|^{1-\alpha} > n^{1-\beta}$ and

$$\left|\sum_{l\in A_i^*} \varepsilon\xi_l + \sum_{l\in A_i^1} \varepsilon\xi_l + \cdots + \sum_{l\in A_i^{2\tau+1}} \varepsilon\xi_l\right| = \left|\sum_{l\in A_i} \varepsilon\xi_l\right| > |A_i|^{1-\alpha}$$

both hold true, then there exists some $j \in \mathcal{J}$ such that $|\sum_{l\in A_i^j} \varepsilon\xi_l| > \frac{n^{1-\beta-\alpha}}{2\tau+1}$. We have as well,

$$\Big\{\Big|\sum_{l\in A_i} \varepsilon\xi_l\Big| > |A_i|^{1-\alpha}\Big\} \subset \{|A_i| \leq n^{1-\beta}\} \cup \Big\{\bigcup_{j\in\mathcal{J}} \Big\{\Big|\sum_{l\in A_i^j} \varepsilon\xi_l\Big| > \frac{n^{1-\beta-\alpha}}{2\tau+1}\Big\}\Big\}.$$

Since $|\mathcal{J}| \leq 2\tau+1$, using the union bound and (5.8), for $j \in \mathcal{J}$

$$\begin{aligned}\mathbb{P}\Big\{\Big|\sum_{l\in A_i} \varepsilon\xi_l\Big| > |A_i|^{1-\alpha} \;\Big|\; |A_i| > n^{1-\beta}\Big\} &\leq (2\tau+1)\max_{j\in\mathcal{J}} \mathbb{P}\Big\{\Big|\sum_{l\in A_i^j} \varepsilon\xi_l\Big| > \frac{n^{1-\beta-\alpha}}{2\tau+1}\Big\} \\ &\leq 2(2\tau+1)\exp\left(-\frac{n^{1-2\beta-2\alpha}}{2(2\tau+1)^2\varepsilon^2}\right).\end{aligned}$$

Using lemma 5.3.2 for those $j \in \mathcal{I}$, for which the contribution is small, we get,

$$\mathbb{P}\Big\{|V_n(i,\tau)| > \frac{1}{n^\alpha}\Big\} \leq 2(2\tau+1)\exp\left(-\frac{n^{1-2\beta-2\alpha}}{2(2\tau+1)^2\varepsilon^2}\right) + \varepsilon(2\tau+1)e^{-n^{\beta/4}}.$$

Observe that the right hand side of the previous inequality goes exponentially fast to zero if $\alpha \in (0, \frac{1}{2})$ and for β sufficiently small. Using once more, the union bound, we obtain the desired result. $\square$

Chapter 6

Numerical simulations

In chapter 5 we described Lalley's algorithm for signal recovery, and we also showed that it indeed eliminates the noise. The obvious next step is to implement it. In [48] the author gives no further details on the simulations, that is why, in this chapter we present some results on the implementation of Lalley's "Denoising algorithm" for chaotic dynamical systems. This is an ongoing work in collaboration with Marc Monticelli, who is developing an interactive version of the algorithm for the package for numerical simulations xDim.

6.1 Implementation

Recall we are given a time series of the form $y_i = x_i + \varepsilon\xi_i$. Where the x_i's are produced by a chaotic dynamical system. We assume that computational rounding do not affect considerably the dynamics (that is, we assume that the noise is purely observational). We simulate the noise by considering the function `unifrnd` from the package `statistics` in Matlab. The noise is bounded by one, and so the amplitude of the noise is given by ε. We remind also that the estimator of the true orbit is denoted by $\widehat{x}_i$ and is defined by

$$\widehat{x}_i := \frac{1}{|A_i(\delta,\tau)|} \sum_{j\in A_i(\delta,\tau)} y_j,$$

where

$$A_i(\delta,\tau) := \{j \ : \ |y_{j+r} - y_{i+r}| \leq \delta \ \text{ for } \ |r| \leq \tau\}.$$

The algorithm depends on two parameters. The first one is the length of the window τ which is the time two close points remain close up to a distance δ (as in figure 5.2). The second is δ, the radius of the ball considered as neighborhood (see Fig. 5.1). We remark that in the proof, $\delta = 3\varepsilon$ and so $\delta \leq \frac{3\Delta}{5}$ (since $\varepsilon \leq \frac{\Delta}{5}$), where Δ is the separation threshold of the map. Since *a priori* we do not know the map, one cannot give precisely the value of Δ. In this study we vary δ in a very coarse manner, from ε to 6ε.

Next, in order to quantify the efficiency of the algorithm we consider three quantities. The first one is the mean distance between the estimated orbit and the original one. We refer to it as the mean remaining error, and it is given by

$$e_1 := \frac{1}{n}\sum_{i=0}^{n-1} d(\widehat{x}_i, x_i),$$

where d is the euclidean distance and n is the length of the series. The second one is considered in [43, section 10.3.4, page 187], is given as follows

$$e_2 := \sqrt{\frac{1}{n}\sum_{i=0}^{n-1} d(\widehat{x}_i, x_i)^2}.$$

And the third one is the ratio between the total squared error before and after the implementation of the algorithm, it is given by

$$r := \sqrt{\frac{\sum_i d(y_i, x_i)^2}{\sum_i d(\widehat{x}_i, x_i)^2}}.$$

This quantity is a signal-to-noise ratio and it is expected to increase as the performance of the algorithm improves. This ratio is used in [64] and [35], for instance[1].

Remark 6.1.1. *Although the algorithm does not rely on the previous knowledge of the clean orbit or the precise form of the dynamics, all these quantifiers do make use the true orbit.*

We implemented the algorithm in the straightforward way, that is, a $\mathscr{O}(n^2)$-step algorithm, which is caused by the neighbor search. One can run this algorithm for a series of length 5000 in a few seconds, in a 2GHz processor.

We present results for the noise reduction on the Hénon and Lozi maps with the 'classical' parameters as in examples 1.3.1 and 1.3.2.

6.2 Results for the Hénon map

To start our numerical investigation for the optimal parameters δ and τ for the denoising algorithm applied on the Hénon map, we chose $\delta = 3\epsilon$. We consider 10,000 iterations of the observed Hénon map. We remind that τ is not completely free, since the algorithm works rigorously for $\tau \leq \log(n)$. In this case we can test τ for all integers between 1 to 9 (since $\lfloor \log(10,000) \rfloor = 9$). In figure 6.1 we show for three different amplitudes of noise, the remaining error e_1 and e_2 after the application of the algorithm for all possible values of τ.

Next, let us chose $\varepsilon = 0.05$. We check the algorithm for every possible value of τ and some values of δ. One observes that for small τ (say $\tau = 1, 2$) the smoothing is made over sets $A_i(\delta, \tau)$ containing too many points and so defining a bad estimator. This is

[1] There it is denoted by r_0.

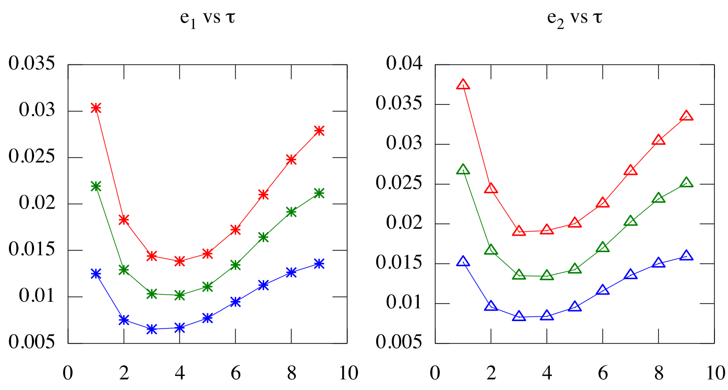


Figure 6.1: For 10,000 iterations, and $\delta = 3\varepsilon$. Left. We plot the error e_1 for each τ, we used continuous lines for clarity. The blue line are the result for $\varepsilon = 0.03$, the green one shows the results for $\varepsilon = 0.05$ and the red one for $\varepsilon = 0.07$. Right. We plot the mean squared error versus τ, with the same color convention for the amplitude of the noise.

caused because for small τ's one does not take advantage of the chaotic nature of the dynamics. In figure 6.2 we plot a 'curve' for each τ from 3 to 9, showing the mean remaining error e_1 after the implementation of the algorithm at six values of δ. We vary δ as $k\varepsilon$ for $k = 1, \ldots, 6$, this is of course a very corse search. An natural improvement of the results displayed in figure 6.2 would be varying in a finer manner δ as a function of ε in the interval $[2\varepsilon, 4\varepsilon]$ for example, where our plot shows the local minimum.

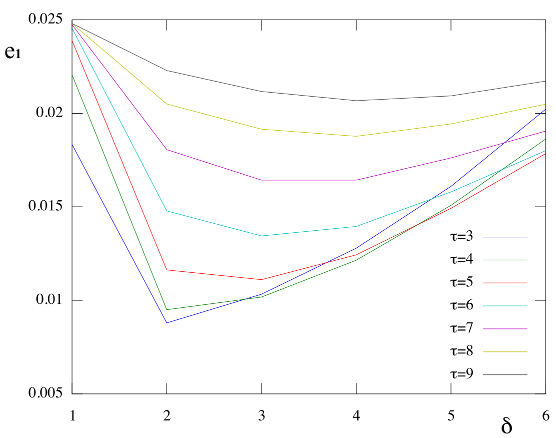


Figure 6.2: This plot shows the remaining error e_1 after the application of the denoising algorithm. The x-axis shows the value of δ as a multiple of ε. We took 10,000 iterations with a noise whose amplitude is bounded by $\varepsilon = 0.05$. The optimal δ is bigger as τ increases. The best results are obtained for $\tau = 3, 4$ with $\delta \in [2\varepsilon, 3\varepsilon]$.

Similar qualitative results are valid for $\varepsilon = 0.03$ and $\varepsilon = 0.07$. In figure 6.3, we show the remaining error e_1 after the application of the algorithm with 10,000 iterations, we only show for $\tau = 3, 4$ and varying δ as before. For 10,000 iterations the best results are obtained with $\tau = 3$ and $\delta \in [2\varepsilon, 3\varepsilon]$. We expect that, as the number of iterations increases the optimal τ should also increase, since we expect to have more close points.

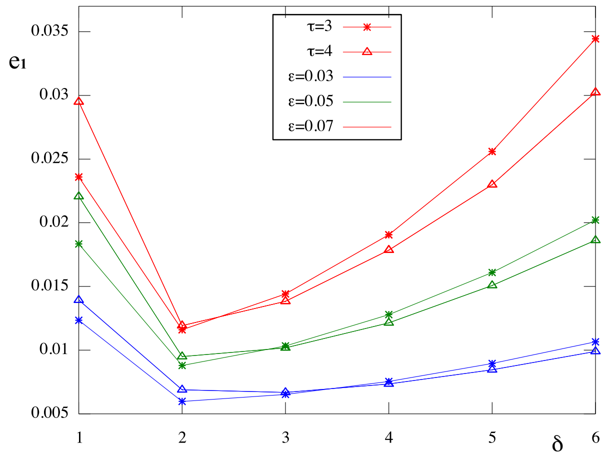


Figure 6.3: The remaining error e_1 with the best values of τ for 10,000 iterations and as a function of δ.

Following the proof of the consistency of the algorithm, one must perceive the fall of the remaining error as we increase the number of iterations. This is actually shown in figure 6.4, where we plot both, e_1 and e_2 versus the number of iterations, we apply the algorithm with parameters $\delta = 3\varepsilon$ and $\tau = 4$ for a series with 1,000, 5,000, 10,000, 50,000 and 100,000 iterations respectively. Another evidence of the good performance of the denoising algorithm is that the distribution of the distances $d(\widehat{x}_i, x_i)$ should change from the uniform distribution towards a delta centered at zero, in the perfect case. Obviously one does not expect to get a delta but a Poissonian distribution. We show in figure 6.5 that this is indeed the case by plotting histograms for the distribution of $d(\widehat{x}_i, x_i)$. It would be interesting to carry out further analysis in statistics.

The main purpose is to recover the attractor from the given perturbed object. In figure 6.6, we show an example of the attractor perturbed with observational noise and the recovered signal after the application of the algorithm using 100,000 iterations[2] with parameters $\delta = 3\varepsilon$ and $\tau = 4$.

Finally we collect the information in Table I, we show the resulting three quantities e_1, e_2 and r for the algorithm applied to the Hénon map.

[2]We plot only 10,000 points, from iteration 45,000 to 54,999, since images are quite heavy with 100,000 iterations.

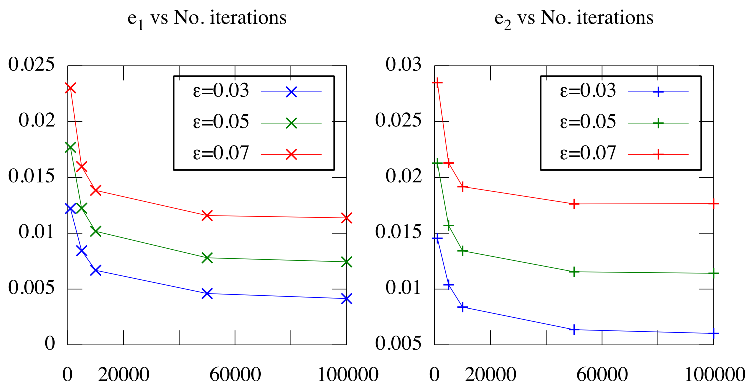


Figure 6.4: The remaining error decreases as the iterations considered for the smoothing increases. Left: We plot e_1 versus the number of iterations. Right: Shows e_2 versus the number of iterations. Both plots are result of the denoising algorithm with parameters $\delta = 3\varepsilon$ and $\tau = 4$.

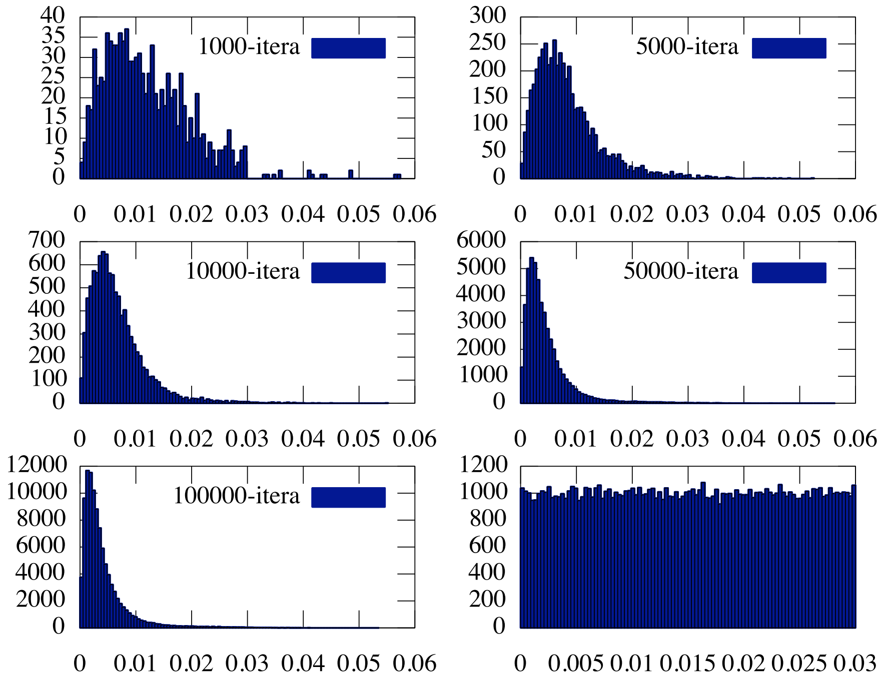


Figure 6.5: Histograms of the distance $d(\widehat{x}_i, x_i)$. We apply the denoising algorithm with $\delta = 3\varepsilon$ and $\tau = 4$.

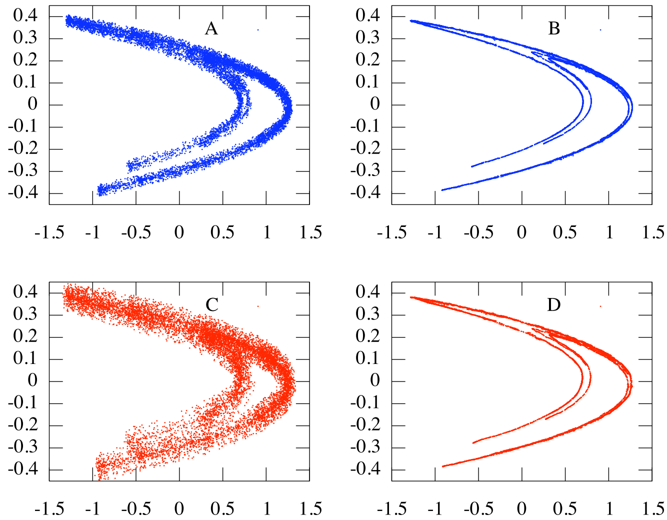


Figure 6.6: A. Hénon attractor with observational noise with amplitude bounded by $\varepsilon = 0.03$. B. The recovered attractor after the application of the denoising algorithm with 100,000 iterations. C. Hénon attractor with observational noise with amplitude bounded by $\varepsilon = 0.07$. D. The recovered attractor after the application of the denoising algorithm 100,000 iterations.

Table I. Hénon map, $\delta = 3\varepsilon$, $\tau = 4$

Iterations	Noise Amplitude	e_1	e_2	r
1000	$\varepsilon = 0.03$	0.012207	0.014540	1.1740
	$\varepsilon = 0.05$	0.017689	0.021278	1.3371
	$\varepsilon = 0.07$	0.023026	0.028501	1.3976
5000	$\varepsilon = 0.03$	0.0084413	0.010391	1.6580
	$\varepsilon = 0.05$	0.012242	0.015692	1.8297
	$\varepsilon = 0.07$	0.015979	0.021296	1.8876
10000	$\varepsilon = 0.03$	0.0066832	0.0083862	2.0554
	$\varepsilon = 0.05$	0.010178	0.013430	2.1391
	$\varepsilon = 0.07$	0.013842	0.019174	2.0976
50000	$\varepsilon = 0.03$	0.0046040	0.0063687	2.7140
	$\varepsilon = 0.05$	0.0078003	0.011547	2.4948
	$\varepsilon = 0.07$	0.011587	0.017627	2.2881
100000	$\varepsilon = 0.03$	0.0041491	0.0060286	2.8743
	$\varepsilon = 0.05$	0.0074330	0.011409	2.5314
	$\varepsilon = 0.07$	0.011378	0.017650	2.2908

6.3 Results for Lozi map

We proceed with a similar numerical analysis for the Lozi map. We search for the best parameters δ and τ using 10,000 iterations if the observed Lozi map with a noise whose amplitude is bounded by 0.05. This is shown in figure 6.7. Similar results are valid for different values of ε for instance 0.03 and 0.07 as considered before. For $\tau = 1$ and 2, the algorithm introduces more error while doing the smoothing and for $\tau = 8$ and 9 almost no correction is made.

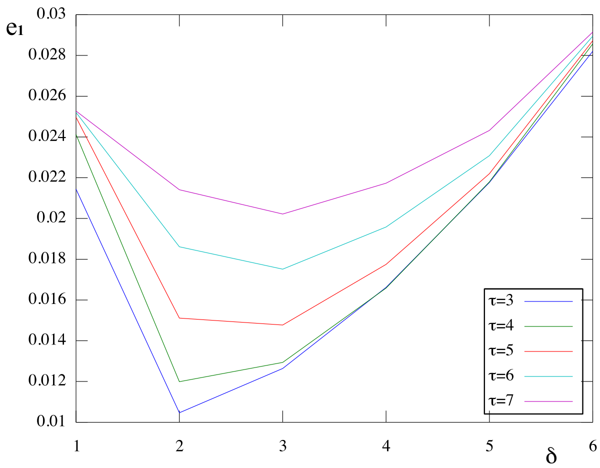


Figure 6.7: The remaining error e_1 after the application of the denoising algorithm using 10,000 iterations. In the x-axis are the values of δ as an integer multiple of ε. In this plot, the amplitude of the noise is bounded by 0.05. The best results are obtained for $\tau = 3$ and $\delta \in [2\varepsilon, 3\varepsilon]$.

Since for 10,000 iterations the best parameters are $\tau = 3$ and $\delta \in [2\varepsilon, 3\varepsilon]$, we fix ourselves to this parameters, although (as for Hénon map one expects to obtain better results using 100,000 with bigger τ's). In figure 6.8 we plot the remaining errors e_1 and e_2 after the application of the algorithm. We show the histograms of the distribution of the distance $d(\widehat{x}_i, x_i)$ using the algorithm with 1000, 5000, 10000, 50000 and 100,000 iterations, see figure 6.9. Observe that the results are little worse that for the Hénon map, we think this is caused by the error included by the same algorithm for points close to the non-differentiable point of Lozi map, and also probably because Lozi map is 'less expansive' than Hénon map, property which enhances the good performance of the algorithm.

In figure 6.10 we show two simulations of the observed Lozi map with noise whose magnitude is bounded by 0.03 and 0.07 respectively, and their corresponding reconstructed 'attractor' after the application of the algorithm[3]. Finally, we collect the results

[3]We used a series with 100,000 iterations but we only plot 10,000 points, from iteration 45,000 to 54,999, since pictures are quite heavy otherwise.

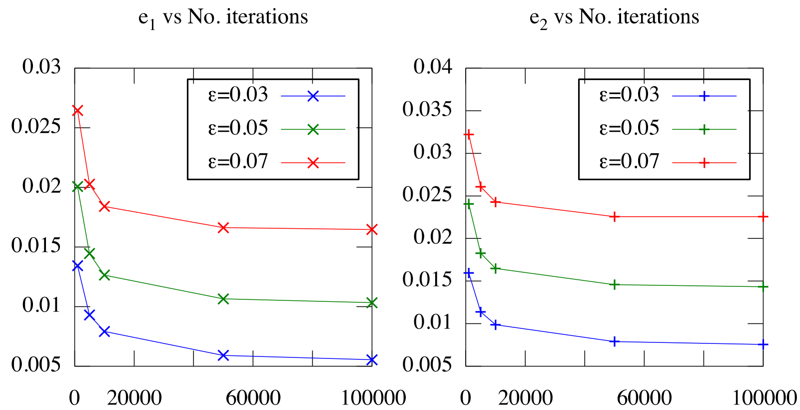


Figure 6.8: Denoising algorithm applied with parameters $\delta = 3\varepsilon$ and $\tau = 3$. Left: We plot e_1 versus the number of iterations. Right: The plot shows e_2 versus the number of iterations.

for the three quantities e_1, e_2 and r for the algorithm with parameters $\delta = 3\varepsilon$ and $\tau = 3$ applied to the Lozi map.

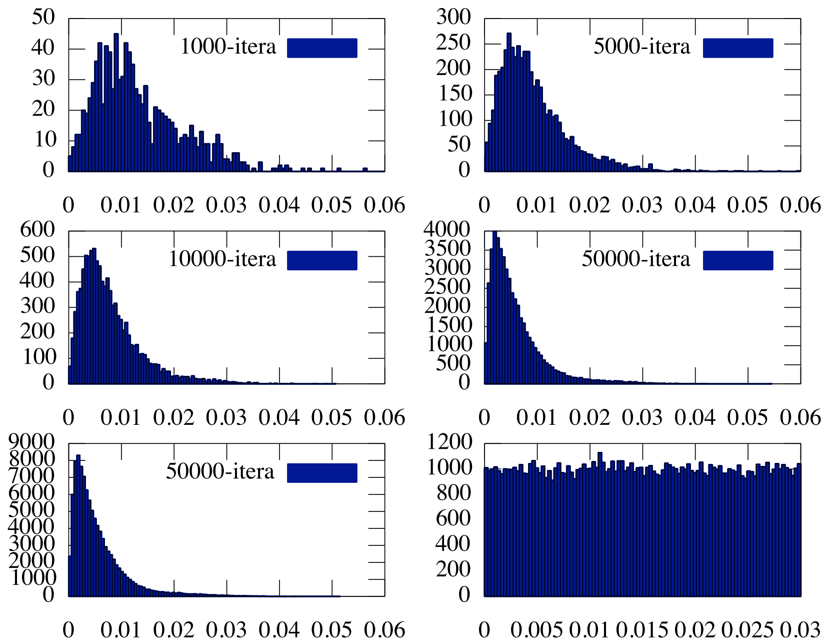


Figure 6.9: Histograms of the distance $d(\widehat{x}_i, x_i)$. The denoising algorithm applied with parameters $\delta = 3\varepsilon$ and $\tau = 3$, in all cases. The last histogram shows $d(y_i, x_i)$.

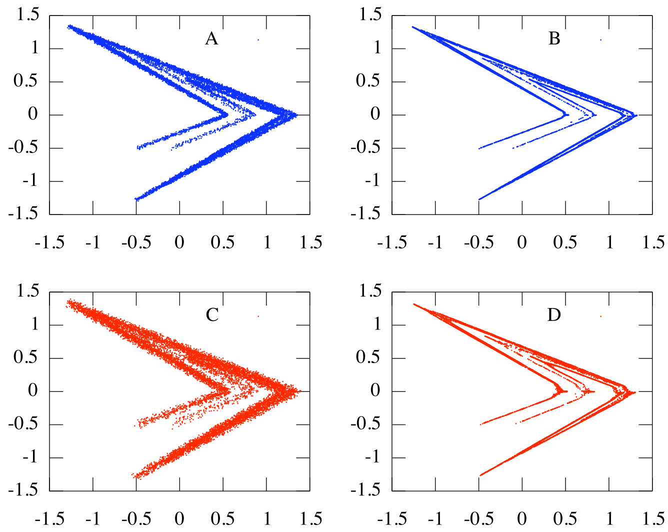


Figure 6.10: A. Observed Lozi attractor with $\varepsilon = 0.03$. B. The recovered attractor after the application of the denoising algorithm with 100,000 iterations. C. Observed Lozi attractor with $\varepsilon = 0.07$. D. The recovered attractor after the application of the denoising algorithm 100,000 iterations.

Table II. Lozi map, $\delta = 3\varepsilon$, $\tau = 3$

Iterations	Noise Amplitude	e_1	e_2	r
1000	$\varepsilon = 0.03$	0.013430	0.015948	1.0899
	$\varepsilon = 0.05$	0.020072	0.024051	1.2045
	$\varepsilon = 0.07$	0.026461	0.032221	1.2587
5000	$\varepsilon = 0.03$	0.0093033	0.011372	1.5452
	$\varepsilon = 0.05$	0.014456	0.018264	1.6035
	$\varepsilon = 0.07$	0.020272	0.026082	1.5720
10000	$\varepsilon = 0.03$	0.0079149	0.0098613	1.7754
	$\varepsilon = 0.05$	0.012642	0.016474	1.7713
	$\varepsilon = 0.07$	0.018401	0.024278	1.6827
50000	$\varepsilon = 0.03$	0.0059072	0.0079032	2.1985
	$\varepsilon = 0.05$	0.010651	0.014578	1.9865
	$\varepsilon = 0.07$	0.016623	0.022559	1.7971
100000	$\varepsilon = 0.03$	0.0055369	0.0075546	2.2918
	$\varepsilon = 0.05$	0.010326	0.014335	2.0130
	$\varepsilon = 0.07$	0.016475	0.022564	1.7904

6.4 Simulations on xDim

The present version of xDim[4] contains an interactive version of the denoising algorithm. We show here an example of simulation on the Hénon map. In this framework one can interactively change the amplitude of the noise, see figure 6.11, and the parameters of the denoising algorithm τ and δ, giving the results of the recovered signal in real time. We show an example in figure 6.12. This implementation is still under development.

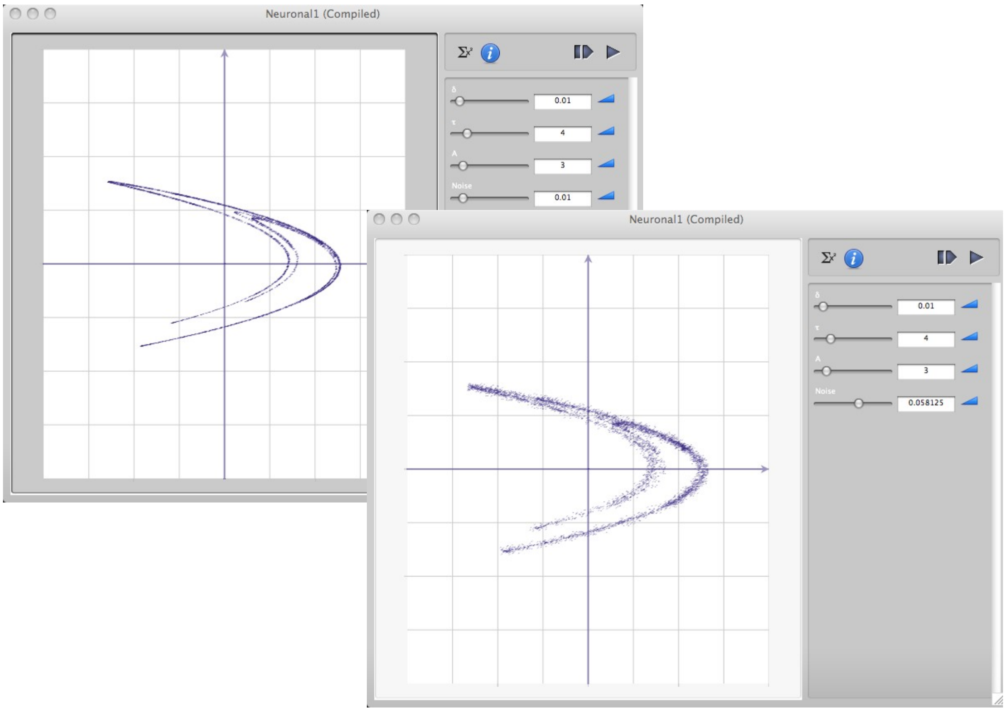


Figure 6.11: A window of xDim showing a simulation of the Hénon attractor and the observed Hénon attractor. The amplitude of the noise can be modified using an slide bar.

[4]xDim is a Macintosh based application for numerical simulations of dynamical systems. It is under development carried out by Marc Monticelli. For further information see, http://math.unice.fr/∼monticel/Marc_Monticelli/Activites.html

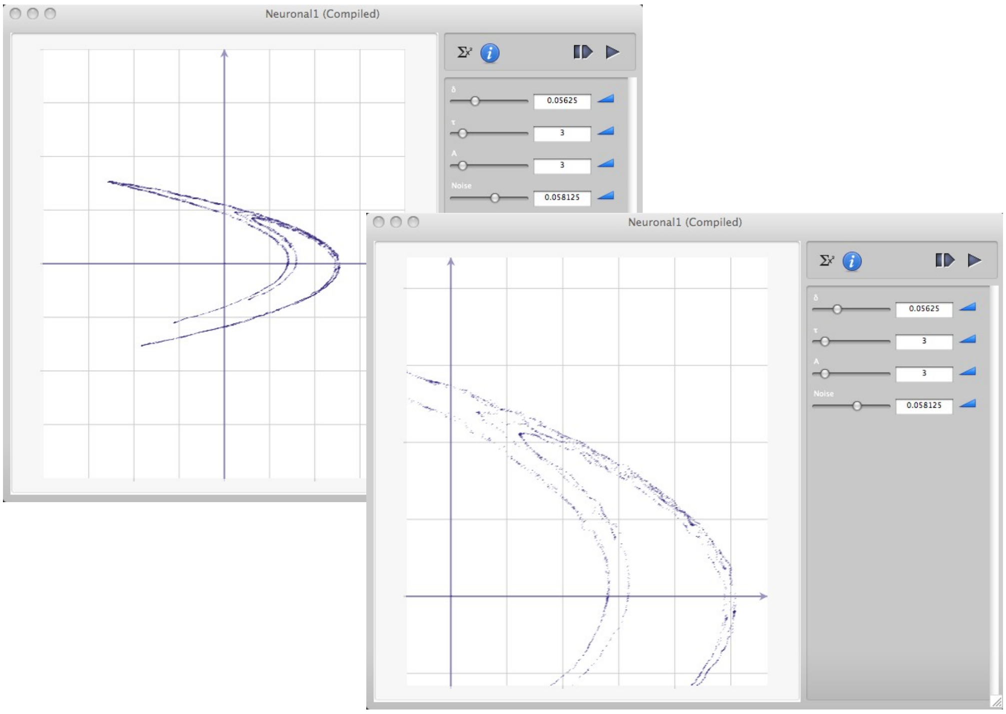


Figure 6.12: The recovered attractor. One can adjust τ and δ using slide bars and see interactively the resulting image. As shown, one can also zoom the image.

Figure 6.13: Parameter's slide bars for δ, τ and the maximum amplitude of the noise.

Part III

Entropy Estimation of one-dimensional Gibbs measures

Chapter 7

Study of fluctuations for entropy estimation

7.1 Introduction

Assume that we have a sequence $x_0, x_1, \dots$ of outcomes of an unknown discrete ergodic process ν. A natural and extensively studied problem is that of consistently estimating the entropy $h(\nu)$. One may find several quantities that converge to the entropy. Thus we may use them as entropy estimators, either theoretically or empirically. Some of those quantities are, for instance, the empirical entropies, return times, hitting and waiting times.

Once the convergence is established for a given entropy estimator, the following natural question arises: How fast does the estimate converge to the true entropy? This question opens the study for the fluctuations of the entropy estimates. In the present chapter we focus on the fluctuations of entropy estimates. We present the respective results for two of the main approaches: The so called 'plug-in' estimator (empirical entropy), as well as conditional empirical entropy, and the hitting time entropy estimator. We discuss two type of results which constitute the classical scheme to study the fluctuations, the central limit theorem and large deviations principle.

We begin by fixing some notations and recalling some definitions. Let A be a finite set which is called *alphabet*, with cardinality $|A| \geq 2$. Recall that $A^{\mathbb{N}_0}$ denotes the set of all infinite sequences $\underline{x} = \{x_i\}_{i \in \mathbb{N}_0}$ of elements of the alphabet A. Let us denote the sample $x_0, x_1, \dots, x_{n-1}$ by x_0^{n-1}. In the same way, we write a_0^{n-1} for the word $a_0 a_1 \cdots a_{n-1}$. Recall that the symbol $[a_i^j]$ stands for the cylinder set $[a_i^j] := \{\underline{x} \in A^{\mathbb{N}_0} : x_i^j = a_i^j\}$.

For every $k \geq 1$ the k-block Shannon entropy of ν is defined as

$$H_k(\nu) := -\sum_{a_0^{k-1}} \nu([a_0^{k-1}]) \log \nu([a_0^{k-1}]).$$

If we denote by ν_k the k-marginal of ν, then ν_k defines a probability measure on A^k

given by $\nu_k(a_0^{k-1}) = \nu([a_0^{k-1}])$, so we may write

$$H_k(\nu) = H_k(\nu_k) := -\sum_{a_0^{k-1}} \nu_k(a_0^{k-1}) \log \nu_k(a_0^{k-1}).$$

The k-block Shannon entropy gives us the average amount of information of the measure contained in a word of length k.

Let $\nu_k(a_{k-1} \mid a_0^{k-2})$ be the conditional probability $\nu_k(a_0^{k-1})/\sum_b \nu_k(a_0^{k-2}b)$. Then, for every $k \geq 2$ the k-block conditional entropy, is defined as follows,

$$h_k(\nu) := -\sum_{a_0^{k-1}} \nu([a_0^{k-1}]) \log \frac{\nu([a_0^{k-1}])}{\nu(a_0^{k-2})} = h_k(\nu_k) := -\sum_{a_0^{k-1}} \nu_k(a_0^{k-1}) \log \nu_k(a_{k-1} \mid a_0^{k-2}).$$

It is know that the following relation holds (see for instance [68]),

$$h_k(\nu) = H_k(\nu) - H_{k-1}(\nu), \quad k \geq 1, \tag{7.1}$$

where by convention we set $H_0(\nu) := 0$. If ν is a stationary measure, then

$$\lim_{k\to\infty} h_k(\nu) = \lim_{k\to\infty} \frac{H_k(\nu)}{k} = h(\nu),$$

where $h(\nu)$ is the (Shannon-Kolmogorov-Sinai) entropy of ν.

Standing assumption: From now on, ϕ is a Hölder continuous potential and μ_ϕ is its unique Gibbs measure.

7.2 Empirical entropies

The most natural procedure to estimate the entropy of a probability measure ν is to take the empirical distribution of the k-blocks as a estimate of the measure ν_k and then calculate the Shannon entropy of that estimate. Let us formalize this last statement. Denote the empirical frequency of the word a_0^{k-1} in the sample $x_0, x_1, \ldots, x_{n-1}$ by

$$\mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) := \frac{1}{n} \#\{0 \leq j \leq n : \tilde{x}_j^{j+k-1} = a_0^{k-1}\},$$

where $\underline{\tilde{x}} := x_0^{n-1} x_0^{n-1} \cdots$ is the periodic point with period n made from x_0^{n-1}. This trick makes $\mathcal{E}_k(\cdot\,; x_0^{n-1})$ a locally shift-invariant probability measure on A^k.

For any ergodic measure ν, there is a set of ν-measure one of $\underline{x}$'s such that for every $k \geq 1$

$$\lim_{n\to\infty} \mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) = \nu([a_0^{k-1}]).$$

The k-block empirical entropy is defined as

$$\widehat{H}_k(x_0^{n-1}) := H_k(\mathcal{E}_k(\cdot\,; x_0^{n-1})).$$

From Birkhoff's ergodic theorem one has that $\widehat{H}_k(x_0^{n-1})/k$ converges to $H_k(\nu)/k$, for ν-almost every $\underline{x}$. Which in turn, converges to h as $k \to \infty$. That is, one has that the following double limit

$$\lim_{k\to\infty}\lim_{n\to\infty}\frac{\widehat{H}_k(x_0^{n-1})}{k} = h(\nu).$$

A natural question is: Can we take both limits at once by an adequate definition of $k(n)$? This is indeed possible as shown by D.Ornstein and B.Weiss in [59] (see also theorem II.3.5 in [68]). Here we include it.

Theorem 7.2.1 ([59])**.** *If ν is an ergodic measure of entropy $h(\nu) > 0$, if $k(n) \to \infty$ and $k(n) \le \frac{1}{h(\nu)}\log n$, then*

$$\lim_{n\to\infty}\frac{\widehat{H}_{k(n)}(x_0^{n-1})}{k(n)} = h(\nu) \quad \text{for } \nu - \text{almost-every } \underline{x}.$$

In particular, if $k(n) \sim \log_{|A|} n$ then the convergence above holds for any ergodic measure ν with alphabet A, while if $k(n) \sim \log\log n$, then it holds for any finite-alphabet ergodic process.

Note that since $h(\nu) \le \log|A|$ we can always take $k(n) \le \frac{1}{\log|A|}\log n$.

Analogously, based on conditional entropy, one can define an empirical estimator of the k-block conditional entropy, as follows,

$$\widehat{h}_k(x_0^n) := h_k(\mathcal{E}_k(\cdot\,; x_0^{n-1})).$$

Theorem 7.2.2. *Let ν be an ergodic measure of positive entropy $h(\nu)$, if $k(n) \to \infty$ and $k(n) \le \frac{1-\epsilon}{h(\nu)}\log n$, for any $\epsilon \in (0,1)$, then*

$$\lim_{n\to\infty}\widehat{h}_{k(n)}(x_0^{n-1}) = h(\nu), \quad \text{for } \nu - \text{almost every } \underline{x}.$$

In particular, we can take $k(n) \le \frac{1-\varepsilon}{\log|A|}\log n$.

The proof is based on the theorem 7.2.1 and general properties of the entropy.

Proof. Let $(k(n))$ be a sequence of positive integers such that $k(n) \to \infty$ as $n \to \infty$ and

$$k(n) \le (1-\epsilon)\frac{1}{h}\log n, \quad \epsilon \in (0,1)\ .$$

In order to lighten the notation, we will drop the dependence in the sample since it is fixed. By h we denote de entropy of ν. We also omit systematically the integer part symbol.

By the relation (7.1) and the theorem 7.2.1, we have

$$\lim_{n\to\infty}\frac{1}{\frac{1}{h}\log n}\widehat{H}_{\frac{1}{h}\log n} = \lim_{n\to\infty}\frac{1}{\frac{1}{h}\log n}\sum_{i=0}^{\frac{1}{h}\log n - 1}\widehat{h}_i = h, \qquad \nu - \text{almost surely.}$$

Also using 7.1, one may write that

$$\begin{aligned}\frac{\left(1-\frac{k(n)}{\frac{1}{h}\log n}\right)}{\frac{1}{h}\log n-k(n)}\sum_{i=k(n)}^{\frac{1}{h}\log n-1}\widehat{h}_i =& \frac{1}{\frac{1}{h}\log n}\sum_{i=0}^{\frac{1}{h}\log n-1}\widehat{h}_i-\left(\frac{k(n)}{\frac{1}{h}\log n}\right)\frac{1}{k(n)}\sum_{i=0}^{k(n)-1}\widehat{h}_i\\ =& \frac{1}{\frac{1}{h}\log n}\widehat{H}_{\frac{1}{h}\log n}-\left(\frac{k(n)}{\frac{1}{h}\log n}\right)\frac{1}{k(n)}\widehat{H}_{k(n)}.\end{aligned}$$

Using theorem 7.2.1 one get ν-almost surely that for any $\delta > 0$ there exists an integer n_0 such that for all $n > n_0$

$$\frac{\left(1-\frac{k(n)}{\frac{1}{h}\log n}\right)}{\frac{1}{h}\log n-k(n)}\sum_{i=k(n)}^{\frac{1}{h}\log n-1}\widehat{h}_i = h\left(1-\frac{k(n)}{\frac{1}{h}\log n}\right)\pm\delta\left(1+\frac{k(n)}{\frac{1}{h}\log n}\right).$$

Hence,

$$\frac{1}{\frac{1}{h}\log n-k(n)}\sum_{i=k(n)}^{\frac{1}{h}\log n-1}\widehat{h}_i = h\pm\delta\frac{1+\frac{k(n)}{\frac{1}{h}\log n}}{1-\frac{k(n)}{\frac{1}{h}\log n}}.$$

Using the hypothesis on $k(n)$, we obtain that,

$$\frac{1}{\frac{1}{h}\log n-k(n)}\sum_{i=k(n)}^{\frac{1}{h}\log n-1}\widehat{h}_i = h+\frac{2\delta}{\epsilon},$$

since $\epsilon > 0$ is fixed, and $\delta > 0$ is arbitrary,

$$\lim_{n\to\infty}\frac{1}{\frac{1}{h}\log n-k(n)}\sum_{i=k(n)}^{\frac{1}{h}\log n-1}\widehat{h}_i = h \qquad \nu\text{-almost surely.}$$

Finally, notice that

$$\frac{1}{\frac{1}{h}\log n-k(n)+1}\sum_{i=k(n)}^{\frac{1}{h}\log n-1}\widehat{h}_i\le\widehat{h}_{k(n)}\le\frac{1}{k(n)}\sum_{i=0}^{k(n)-1}\widehat{h}_i,$$

by the monotonicity of the conditional entropy.

□

7.2.1 Central limit theorem for empirical entropies

Here we present the central limit theorem result for the conditional empirical entropy for the case of Gibbs measures.

Theorem 7.2.3 ([33]). *For any sequence $(k(n))_{n\in\mathbb{N}}$ of positive integers such that $k(n) < \frac{\log n}{2\log|A|}$ and $k(n) > \frac{(1+\varepsilon)}{2}\frac{\log n}{\log\theta^{-1}}$, for some $0<\varepsilon<1$. If $\sigma_\phi^2 > 0$, then for every $t>0$,*

$$\lim_{n\to\infty}\mu_\phi\Big\{\frac{\sqrt{n}}{\sigma_\phi}\left(\widehat{h}_{k(n)} - h(\mu_\phi)\right)\le t\Big\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}\exp(-s^2/2)\mathrm{d}s.$$

Proof. Let $\widehat{\Delta}_k$ be defined as follows,

$$\widehat{\Delta}_k(x_0^{n-1}) := -\sum_{a_0^{k-1}\in A^k}\mathcal{E}_k(a_0^{k-1};x_0^{n-1})\log\frac{\mathcal{E}_k(a_0^{k-1};x_0^{n-1})}{\mathcal{E}_{k-1}(a_0^{k-2};x_0^{n-1})} + \sum_{a_0^{k-1}\in A^k}\mathcal{E}_k(a_0^{k-1};x_0^{n-1})\log\frac{\mu_\phi([a_0^{k-1}])}{\mu_\phi([a_1^{k-1}])}.$$

We make use of the following decomposition of the conditional empirical entropy.

Lemma 7.2.1. *We have*

$$\widehat{h}_{k(n)}(x_0^{n-1}) = \frac{1}{n}\sum_{j=0}^{n-1}(-\phi(\sigma^j\underline{x})) + \widehat{\Delta}_{k(n)}(x_0^{n-1}) + \mathcal{O}(\theta^{k(n)}). \tag{7.2}$$

Furthermore

$$\left|\mathbb{E}\left(\widehat{\Delta}_{k(n)}\right)\right| \le \frac{M|A|^{k(n)}}{n}, \tag{7.3}$$

where $M>0$.

Let us for the moment assume the previous lemma, we will give its proof further (its proof can be deduced from the proof of Theorem 2.1 in [33]). Using the decomposition of $\widehat{h}_k$ given above, we write

$$\sqrt{n}\left(\widehat{h}_k - h(\mu_\phi)\right) = \frac{1}{\sqrt{n}}\left(\sum_{j=0}^{n-1}-\phi(\sigma^j\underline{x}) - nh(\mu_\phi)\right) + \sqrt{n}\widehat{\Delta}_k + \sqrt{n}\mathcal{O}(\theta^k).$$

It is known that Hölder continuous functions ψ satisfy the central limit theorem with variance $\sigma_\psi > 0$ with respect to the Gibbs measure μ_ϕ(see for instance [60]). We have for all $t>0$,

$$\lim_{n\to\infty}\mu_\phi\left\{\frac{1}{\sqrt{n}}\Big(\sum_{j=0}^{n-1}-\phi(\sigma^j\underline{x}) - nh(\mu_\phi)\Big)\le t\right\} = \frac{1}{\sigma_\phi\sqrt{2\pi}}\int_{-\infty}^{t}e^{-s^2/2\sigma_\phi^2}\mathrm{d}s.$$

Now, it is enough to show that $\sqrt{n}\widehat{\Delta}_k$ and $\sqrt{n}\mathcal{O}(\theta^k)$ converge to zero in probability.

Using the Markov inequality, the bound on the expectation of $\widehat{\Delta}_k$ given by (7.3), if $k(n) = \frac{\log n}{2\log|A|}$ we obtain that $\sqrt{n}\mathbb{E}(\widehat{\Delta}_{k(n)}) \le n^{-1/2}M|A|^{\log_{|A|}n^{1/2}} = M$. Finally for the last term, it converges to zero as a sequence of real numbers if $k(n) > \frac{(1+\varepsilon)}{2}\frac{\log n}{\log\theta^{-1}}$ for every $0<\varepsilon<1$. □

Proof of Lemma 7.2.1. We start with the following identity:

$$\widehat{h}_k(x_0^{n-1}) = -\sum_{a_0^{k-1}\in A^k} \mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) \log \frac{\mathcal{E}_k(a_0^{k-1}; x_0^{n-1})}{\mathcal{E}_{k-1}(a_0^{k-2}; x_0^{n-1})}$$

$$= -\sum_{a_0^{k-1}\in A^k} \mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) \log \frac{\mu_\phi([a_0^{k-1}])}{\mu_\phi([a_1^{k-1}])} + \widehat{\Delta}_k(x_0^{n-1}), \tag{7.4}$$

where

$$\widehat{\Delta}_k(x_0^{n-1}) := -\sum_{a_0^{k-1}\in A^k} \mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) \log \frac{\mathcal{E}_k(a_0^{k-1}; x_0^{n-1})}{\mathcal{E}_{k-1}(a_0^{k-2}; x_0^{n-1})} +$$

$$\sum_{a_0^{k-1}\in A^k} \mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) \log \frac{\mu_\phi([a_0^{k-1}])}{\mu_\phi([a_1^{k-1}])} =$$

$$-\sum_{a_0^{k-1}\in A^k} \mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) \log \frac{\mathcal{E}_k(a_0^{k-1}; x_0^{n-1})}{\mu_\phi([a_0^{k-1}])} + \tag{7.5}$$

$$\sum_{a_0^{k-1}\in A^k} \mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) \log \frac{\mathcal{E}_{k-1}(a_0^{k-2}; x_0^{n-1})}{\mu_\phi([a_1^{k-1}])} =$$

$$-H_k(\mathcal{E}_k(\cdot\,; x_0^{n-1}) \mid \mu_\phi) + H_{k-1}(\mathcal{E}_{k-1}(\cdot\,; x_0^{n-1}) \mid \mu_\phi), \tag{7.6}$$

where

$$H_k(\eta \mid \mu_\phi) = \sum_{a_0^{k-1}\in A^k} \eta([a_0^{k-1}]) \log \frac{\eta([a_0^{k-1}])}{\mu_\phi([a_0^{k-1}])}$$

is the k-block relative entropy of η with respect to μ_ϕ. The second term in (7.5) is equal to $H_{k-1}(\mathcal{E}_{k-1}(\cdot\,; x_0^{n-1}) \mid \mu_\phi)$ because of the following two facts. First, $\sum_{a_0\in A} \mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) = \mathcal{E}_{k-1}(a_1^{k-1}; x_0^{n-1})$. This is because $\mathcal{E}_k(\cdot\,; x_0^{n-1})$ is a locally shift-invariant probability measure on A^k. Second, $\sum_{a_{k-1}\in A} \mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) = \mathcal{E}_{k-1}(a_0^{k-2}; x_0^{n-1})$, because the family $(\mathcal{E}_k(\cdot\,; x_0^{n-1}))_{k=1,2,\ldots}$ is consistent.

The quantity $|\widehat{\Delta}_k(x_0^{n-1})|$ is bounded above by $(M|A|^k)/n$ according to [33, formula (4.16)], where $M > 0$ is a constant.

Now we deal with the first term in (7.4). We first introduce the function

$$\phi_k(\underline{y}) := \log \frac{\mu_\phi([y_0^{k-1}])}{\mu_\phi([y_1^{k-1}])}$$

which is a locally constant function on cylinders of length k. It is easy to verify that $\|\phi - \phi_k\|_\infty \le |\phi|_\theta \theta^k$ (see e.g. [60, Prop. 3.2 p. 37]). We get that

$$-\sum_{a_0^{k-1}\in A^k} \mathcal{E}_k(a_0^{k-1}; x_0^{n-1}) \log \frac{\mu_\phi([a_0^{k-1}])}{\mu_\phi([a_1^{k-1}])} = \frac{1}{n}\sum_{j=0}^{n-1}(-\phi(\sigma^j \underline{x})) + \mathcal{O}(\theta^k).$$

The proof of the lemma is complete. □

Next, the following result concerns to the k-block empirical entropy.

Theorem 7.2.4 ([33]). *Let $(q(n))_{n\in\mathbb{N}}$ be a sequence of real positive numbers. For every sequence $(k(n))_{n\in\mathbb{N}}$ of positive integers such that $k(n)\to\infty$ as $n\to\infty$ and such that $k(n)<(1-\varepsilon)\frac{\log n}{\log|A|}$, for some $0<\varepsilon<1$, the following statements hold:*

a) If $\lim_{n\to\infty}\dfrac{q(n)}{k(n)}=0$, then for every $t>0$

$$\lim_{n\to\infty}\mu_\phi\Big\{q(n)\Big(\frac{1}{k(n)}\widehat{H}_{k(n)}-h(\mu_\phi)\Big)>t\Big\}=0.$$

b) If $\lim_{n\to\infty}\dfrac{q(n)}{k(n)}=\alpha$, for some $0<\alpha<\infty$, then

$$\lim_{n\to\infty}\mu_\phi\Big\{q(n)\Big(\frac{1}{k(n)}\widehat{H}_{k(n)}-h(\mu_\phi)\Big)>\alpha\sum_{k=0}^{\infty}(h_k-h(\mu_\phi))\Big\}=0.$$

c) If $\lim_{n\to\infty}\dfrac{q(n)}{k(n)}=+\infty$, then for any $r\in\mathbb{R}$

$$\lim_{n\to\infty}\mu_\phi\left\{q(n)\left(\frac{1}{k(n)}\widehat{H}_{k(n)}-h(\mu_\phi)\right)<r\right\}=0.$$

This means that no scaling produce asymptotic normality, except for the i.i.d case with the common scaling $\sqrt{n}$, which was proved in [42].

The two results above were proved in [33] in the setting of g-measures[1]. As we already shown, one may adapt their proofs for the case of Gibbs measures. In fact it is know that g-measures can be interpreted as a one-dimensional Gibbs measure if the variations go to 0 exponentially fast ([32]).

7.2.2 Large deviations

Besides the central limit theorem, one might ask if the conditional entropy estimator satisfy a large deviations principle. Indeed, in the present section we include the results obtained by J.-R. Chazottes and D. Gabrielli in [20] related to that question. In that article the authors provide a result for g-measures. Here we state that result for Gibbs measures with Hölder continuous potentials.

[1] g-measures or chains of infinite order are stationary process in which at each step the probability governing the choice of a new symbol depends on the entire past in a continuous way.

Theorem 7.2.5 ([20],[16]). *Let μ_ϕ be a Gibbs measure with Hölder continuous potential ϕ which is not the measure of maximal entropy. Assume that $k(n) \to \infty$ as $n \to \infty$ and satisfies*

$$k(n) \leq \frac{1-\varepsilon}{\log |A|} \log n, \quad \textit{for some } 0 < \varepsilon < 1.$$

Then the $k(n)$-block conditional empirical entropy $\widehat{h}_{k(n)}(x_0^{n-1})$ satisfies the following large deviations principle:
for every closed set $C \subset \mathbb{R}$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_\phi\{x_0^{n-1} \ : \ \widehat{h}_{k(n)}(x_0^{n-1}) \in C\} \leq -\inf\{I(u) \ : \ u \in C\},$$

for every open set $O \in \mathbb{R}$

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_\phi\{x_0^{n-1} \ : \ \widehat{h}_{k(n)}(x_0^{n-1}) \in O\} \geq -\inf\{I(u) \ : \ u \in O\},$$

where the function I is defined by

$$I(u) = \begin{cases} \inf\{h(\nu|\mu_\phi) \ : \ \nu \in \mathscr{M}_\sigma(\Omega), \ h(\nu) = u\} & u \in [0, \log |A|] \\ +\infty & \textit{otherwise}, \end{cases}$$

where $\mathscr{M}_\sigma(\Omega)$ is the set of shift-invariant probability measures on the full-shift.
The same large deviations principle holds if we replace $\widehat{h}_{k(n)}(x_0^{n-1})$ by $\frac{\widehat{H}_{k(n)}(x_0^{n-1})}{k(n)}$.

In fact, one knows that the function $I(u)$ is strictly convex on the interval $[h_\infty, \log(|A|)]$, where $h_\infty := \lim_{\beta\to\infty} h(\mu_{\beta\phi})$. Since $I(u)$ is convex, one knows that it is the Legendre transform of the cumulant generating function of $\widehat{h}_{k(n)}$ and $\widehat{H}_{k(n)}/k(n)$. We recall that

$$\begin{aligned} \mathcal{H}(q) &= \lim_{n\to\infty} \frac{1}{n} \log \int e^{nq\widehat{h}_{k(n)}(x_0^{n-1})} \mathrm{d}\mu_\phi(\underline{x}) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \int e^{nq(\widehat{H}_{k(n)}(x_0^{n-1})/k(n))} \mathrm{d}\mu_\phi(\underline{x}), \end{aligned}$$

is the cumulant generating function of both $\widehat{h}_{k(n)}$ and $\widehat{H}_{k(n)}/k(n)$. The calculation of the Legendre transform of $I(u)$ is given by the following proposition.

Proposition 7.2.1 ([16]). *Let μ_ϕ be a Gibbs measure with Hölder continuous potential ϕ which is not the measure of maximal entropy. The Legendre transform of the function I is given by*

$$\mathcal{H}(q) = \begin{cases} (q+1)P\left(\frac{\phi}{q+1}\right) & \textit{for } q \geq -1 \\ \sup\left\{ \int \phi \mathrm{d}\nu \ : \ \nu \in \mathscr{M}_\sigma(\Omega)\right\} & \textit{for } q \leq -1. \end{cases}$$

These results on large deviations with the previous ones about the C.L.T., give a picture of the fluctuations for the empirical entropy estimates.

7.3 Hitting times

There exists a connection between hitting times and entropy for some classes of ergodic process. First let us introduce the corresponding definitions. The (first) hitting time of $\underline{x}$ to a cylinder $[a_0^{n-1}]$ is defined as follows

$$\tau_{[a_0^{n-1}]}(\underline{x}) := \inf\{j \geq 1 \ : \ x_j^{j+n-1} = a_0^{n-1}\}.$$

Given $\underline{x}, \underline{y} \in \Omega$, the waiting time is

$$W_n(\underline{x}, \underline{y}) := \tau_{[x_0^{n-1}]}(\underline{y}) = \inf\{j \geq 1 \ : \ y_j^{j+n-1} = x_0^{n-1}\}.$$

This quantity is just the first time one sees the n first symbols of $\underline{x}$ appearing in $\underline{y}$. In the case of irreducible Markov chains, Wyner and Ziv in [72] have shown that the quantity $\frac{1}{n} \log W_n(\underline{x}, \underline{y})$ converges to h in probability. Here we include the same result of convergence in the case of Gibbs measures, as before ϕ is a Hölder continuous potential and μ_ϕ its unique associated Gibbs measure.

Theorem 7.3.1 ([24]). *Let W_n be defined as above, and μ_ϕ be the unique Gibbs measure associated to ϕ. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log W_n(\underline{x}, \underline{y}) = h(\mu_\phi) \qquad \mu_\phi \otimes \mu_\phi\text{-almost surely.}$$

Where $\otimes$ denotes the direct product measure. The meaning of this result is that if we pick up a pair $\underline{x}, \underline{y}$ at random and independent of one another, then the time needed for orbit of $\underline{y}$ to hit the cylinder $[x_0^{n-1}]$ is typically of order $\exp(nh(\mu_\phi))$. This result is valid under the more general assumption that the process is weak Bernoulli. For full details on the properties of this quantities see [68]. The proof that Gibbs measure are weak Bernoulli process can be found in [12].

7.3.1 Central limit theorem for hitting times

As a matter of fact a central limit theorem holds for Hölder continuous observables with respect to Gibbs measure associated to Hölder continuous potentials (see [60]). Let us recall the definition (2.1) of the variance,

$$\sigma_\phi^2 := \lim_{n\to\infty} \frac{1}{n} \int \left(S_n\phi - n \int \phi \mathrm{d}\mu_\phi\right)^2 \mathrm{d}\mu_\phi.$$

If $\sigma_\phi^2 > 0$ and $\int \phi \mathrm{d}\mu_\phi = 0$, then for every $t \in \mathbb{R}$

$$\lim_{n\to\infty} \mu_\phi \left\{ \frac{-S_n\phi - nh(\mu_\phi)}{\sigma_\phi \sqrt{n}} < t \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp(-\xi^2/2) \mathrm{d}\xi.$$

This fact is used to show that $\log W_n(\underline{x}, \underline{y})$ satisfies a central limit theorem around the entropy adequately rescaled.

Theorem 7.3.2 ([24]). *Assume that μ_ϕ is a Gibbs measure with Hölder continuous potential which is not the measure of maximal entropy. Then*

$$\lim_{n\to\infty} \mu_\phi \otimes \mu_\phi \left\{ \frac{\log W_n - nh(\mu_\phi)}{\sigma\sqrt{n}} < t \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp(-\xi^2/2)\mathrm{d}\xi.$$

Moreover

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \int (\log W_n - h(\mu_\phi))^2 \mathrm{d}\mu_\phi \otimes \mu_\phi.$$

7.3.2 Large deviations

Now we give the large deviations description for the hitting times. Here we include the results in [24]. One wants to calculate the following function for all $q \in \mathbb{R}$ as

$$\mathcal{W}(q) := \lim_{n\to\infty} \frac{1}{n} \log \int W_n^q(\underline{x}, \underline{y}) \mathrm{d}\mu_\phi \otimes \mu_\phi(\underline{x}, \underline{y}),$$

provided the limit exists.

Before giving the result, fix the notation $b_n \sim c_n$ which means that $\max\{b_n/c_n, c_n/b_n\}$ is bounded from above by a constant.

Theorem 7.3.3 ([24]). *Assume that μ_ϕ is a Gibbs measure with Hölder continuous potential ϕ. Then*

$$\int W_n^q(\underline{x}, \underline{y}) \mathrm{d}\mu_\phi \otimes \mu_\phi(\underline{x}, \underline{y}) \sim \sum_{a_0^{n-1}} \mu_\phi([a_0^{n-1}])^{1-q}, \quad \textit{for } q > -1$$

and

$$\int W_n^q(\underline{x}, \underline{y}) \mathrm{d}\mu_\phi \otimes \mu_\phi(\underline{x}, \underline{y}) \sim \sum_{a_0^{n-1}} \mu_\phi([a_0^{n-1}])^{2}, \quad \textit{for } q < -1.$$

Furthermore, as a consequence

$$\mathcal{W}(q) = \begin{cases} P((1-q)\phi), & \textit{for } q \geq -1, \\ P(2\phi), & \textit{for } q \leq 1. \end{cases}$$

Note that $q \mapsto \mathcal{W}(q)$ is continuous but not differentiable at $q = -1$. Indeed the right derivative at -1 of $\mathcal{W}$ is equal to $-\int \phi \mathrm{d}\mu_{2\phi} > 0$.

Hence one has the following large deviations result.

Theorem 7.3.4 ([24]). *Assume that μ_ϕ is a Gibbs measure with Hölder continuous potential which is not the measure of maximal entropy. Then for $u \geq 0$ we have*

$$\lim_{n\to\infty} \frac{1}{n} \log \mu \otimes \mu \left\{ \frac{1}{n} \log W_n > h(\mu) + u \right\} = \inf_{q > -1} \{-(h(\mu) - u)q + \mathcal{W}(q)\},$$

and for $u \in (0, u_0)$*, where* $u_0 := |\lim_{q\downarrow -1} \mathcal{W}'(q) - h(\mu)|$,

$$\lim_{n\to\infty} \frac{1}{n} \log \mu \otimes \mu \left\{ \frac{1}{n} \log W_n < h(\mu) - u \right\} = \inf_{q<-1} \{-(h(\mu) - u)q + \mathcal{W}(q)\}.$$

Note that $u_0 > h(\mu)$ and since $\mathcal{W}'(0) = h(\mu)$, then the previous result study the large fluctuations of $\log \frac{W_n}{n}$ above and below $h(\mu)$.

Chapter 8

Concentration bounds for entropy estimation

The present chapter is devoted to the study of fluctuations of the 'plug-in' and the hitting time entropy estimators in the spirit of concentration inequalities. As we already explained in previous chapters, concentration inequalities enable us to complete the qualitative picture of the fluctuations of Lipschitz observables, because they are valid for every n, where n is the length of the given sample. In particular in this chapter we apply concentration inequalities to the problem of entropy estimation, specifically to the empirical entropy and the hitting time estimators. Except form [4] (in the i.i.d. case), this is the first time that concentration inequalities are used to prove fluctuation bounds for entropy estimators. The results we present in this chapter can be found in [22].

Throughout this chapter $\phi \in \mathscr{F}_\theta$ and μ_ϕ is its unique Gibbs measure. Recall that Gibbs measures satisfy the exponential concentration inequality (2.7), and as a consequence one obtains a deviation probability and a bound for the variance of K (see chapter 2).

For later convenience we write the following particular case of the deviation probability applied to ergodic sums of a Lipschitz function $f : \Omega \to \mathbb{R}$. We have that

$$\mu_\phi \left\{ \underline{x} : \frac{1}{n}\big(f(\underline{x}) + \cdots + f(\sigma^{n-1}\underline{x})\big) - \int f \mathrm{d}\mu_\phi \geq t \right\} \leq e^{-Bnt^2} \tag{8.1}$$

for every $t > 0$ and for every $n \geq 1$, where $B := (4D|f|_\theta^2)^{-1}$.

8.1 'Plug-in' estimator

The convergence properties of the 'plug-in' estimator of the entropy has been studied before in [4] in the case of a i.i.d. process, although they consider a countable alphabet. There the authors use the Azuma's inequality as their main tool. Despite of the restriction of a finite alphabet our result generalizes that of [4] in the sense that our process is Gibbsian.

Theorem 8.1.1. *Let D be the constant appearing in the exponential concentration inequality* (2.7)*. For every $\alpha \in (0,1)$, $t > 0$ and $n \geq 2$ one has*

$$\mu_\phi \left\{ \left| \frac{\widehat{H}_{k(n)}}{k(n)} - \int \frac{\widehat{H}_{k(n)}}{k(n)} \mathrm{d}\mu_\phi \right| \geq t \right\} \leq 2 \exp\left(- \frac{n^{1-\alpha} t^2}{16 D (\log n)^2} \right)$$

provided that $k(n) \leq \frac{\alpha}{2 \log |A|} \log n$.

Moreover for every $n \geq 2$

$$\int \left(\frac{\widehat{H}_{k(n)}}{k(n)} - \int \frac{\widehat{H}_{k(n)}}{k(n)} \mathrm{d}\mu_\phi \right)^2 \mathrm{d}\mu_\phi \leq 8D \frac{(\log n)^2}{n^{1-\alpha}}.$$

Proof. Given any integer $k \geq 1$, consider the function $\tilde{K} : A^n \to \mathbb{R}$ defined by

$$\tilde{K}(s_0, \ldots, s_{n-1}) = \widehat{H}_k(s_0^{n-1}).$$

Since our function $\tilde{K}$ is defined on A^n instead of Ω^n, $\mathrm{Lip}_j(K)$ has to be replaced by $\delta_j(\tilde{K})$, the oscillation at the j-th coordinate, where

$$\delta_j(\tilde{K}) = \sup_{a_0, \ldots, a_{n-1}} \sup_{a_j \neq b_j} \left| \tilde{K}(a_0, \ldots, a_{j-1}, a_j, a_{j+1}, \ldots, a_{n-1}) - \tilde{K}(a_0, \ldots, a_{j-1}, b_j, a_{j+1}, \ldots, a_{n-1}) \right|. \tag{8.2}$$

We estimate the $\delta_j(\tilde{K})$'s. We claim that

$$\delta_j(\tilde{K}) \leq 2k|A|^k \frac{\log n}{n}, \quad \forall j = 0, \ldots, n-1.$$

Indeed, given any string a_0^{k-1}, the change of one symbol in s_0^{n-1} can decrease $\mathcal{E}(a_0^{k-1}; s_0^{n-1})$ by at most k/n. It is possible that another string gets its frequency increased, and this can be at most by k/n. This is the worst case. We then use the fact that for any pair of positive integers l and k such that $l + k \leq n$, one has

$$\left| \left(\frac{l}{n} \right) \log \left(\frac{l}{n} \right) - \left(\frac{l+k}{n} \right) \log \left(\frac{l+k}{n} \right) \right| \leq \frac{k}{n} \log n.$$

The claim follows by summing up this bound for all strings, which gives the factor $|A|^k$. Finally, taking $k(n) \leq \frac{\alpha}{2 \log |A|} \log n$, with $\alpha \in (0,1)$, and using the consequences of (2.7) on the deviation probability and the variance, we get the desired bounds. □

8.2 Conditional empirical entropy

It is natural to seek for a concentration bound for the empirical entropy not about its expectation, but about $h(\mu_\phi)$, the entropy of the Gibbs measure. To have good control on this expectation, it turns out that a better estimator is the conditional empirical entropy. To define it, we need to recall a few definitions and facts.

For a shift-invariant measure ν and $k \geq 2$, let

$$h_k(\nu) = H_k(\nu) - H_{k-1}(\nu) = -\sum_{a_0^{k-1}} \nu([a_0^{k-1}]) \log \frac{\nu([a_0^{k-1}])}{\nu([a_0^{k-2}])}.$$

It is well-known that $\lim_{k\to\infty} h_k(\nu) = h(\nu)$ (see for instance [68]).

The k-block conditional empirical entropy is

$$\hat{h}_k(x_0^{n-1}) = h_k(\mathcal{E}_k(\cdot; x_0^{n-1})).$$

When ν is ergodic, one can prove (see Theorem 7.2.2) that, if $k(n) \to \infty$ and $k(n) \leq \frac{(1-\epsilon)}{\log|A|} \log n$, for any $\epsilon \in (0,1)$, then

$$\lim_{n\to\infty} \hat{h}_{k(n)}(x_0^{n-1}) = h(\nu), \quad \text{for } \nu - \text{almost every } \underline{x}.$$

We have the following result.

Theorem 8.2.1. *Assume that $\theta < |A|^{-1}$. There exist strictly positive constants c, γ, Γ, ξ such that for every $t > 0$ and for every n large enough*

$$\mu_\phi \left\{ \left| \hat{h}_{k(n)} - h(\mu_\phi) \right| \geq t + \frac{c}{n^\gamma} \right\} \leq 2 \exp\left(-\frac{\Gamma n^\xi t^2}{(\log n)^4} \right)$$

provided that $k(n) < \frac{\log n}{2 \log|A|}$.

Remark 8.2.1. *From the proof we have $\gamma = 1/\big(1 + \frac{\log|A|}{\log(\theta^{-1})}\big)$, $\xi = 1 - 2/\big(1 + \frac{\log(\theta^{-1})}{\log|A|}\big)$ and $\Gamma = (\log|A|)^2/16D$.*

Proof. Let us denote the expectation by $\mathbb{E}$.

By definition $\hat{h}_k = \widehat{H}_k - \widehat{H}_{k-1}$. If we let $\tilde{K}'(s_0, \ldots, s_{n-1}) = \hat{h}_k(s_0^{n-1})$, we estimate $\delta_j(\tilde{K}')$ by $2\delta_j(\tilde{K})$.
We now estimate the expectation of $\hat{h}_{k(n)}$. We use lemma 7.2.1.

Now substract $h(\mu_\phi)$ and take the expectation on both sides of (7.2), to get, using (1.2),

$$\mathbb{E}\big(\hat{h}_{k(n)}\big) - h(\mu_\phi) = \mathbb{E}\big(\widehat{\Delta}_{k(n)}\big) + \mathcal{O}(\theta^{k(n)}).$$

We now take $k(n) = q \log n / \log|A|$, where $0 < q < 1$ has to be determined. Choosing $q = 1/\big(1 + \frac{\log\theta^{-1}}{\log|A|}\big)$ we easily get that

$$|\mathbb{E}\big(\hat{h}_{k(n)}\big) - h(\mu_\phi)| \leq \frac{c}{n^\gamma}, \tag{8.3}$$

where $c > 0$ is some constant and $\gamma = 1/\big(1 + \frac{\log|A|}{\log(\theta^{-1})}\big)$.

To end the proof, we apply the deviation probability and use (8.3). For the exponent ξ in the statement of the theorem be strictly positive, one must have $q < 1/2$, which is equivalent to the requirement that $\theta < |A|^{-1}$. □

8.3 Hitting time estimator

Given $\underline{x}, \underline{y} \in \Omega$, let

$$W_n(\underline{x}, \underline{y}) = \inf\{j \geq 1 : y_j^{j+n-1} = x_0^{n-1}\}.$$

Under suitable mixing conditions on the shift-invariant measure ν, we have that,

$$\lim_{n\to\infty} \frac{1}{n} \log W_n(\underline{x}, \underline{y}) = h(\nu), \quad \text{for } \nu \otimes \nu - \text{almost every } (\underline{x}, \underline{y}),$$

where the symbol $\otimes$ denotes the direct product. In particular, as we already saw in theorem 7.3.1, when ν is a Gibbs measure, then the result above holds true [24]. We have the following concentration bounds for the hitting-time estimator.

Theorem 8.3.1. *There exist constants $C_1, C_2 > 0$ and $t_0 > 0$ such that, for every $n \geq 1$ and every $t > t_0$,*

$$(\mu_\phi \otimes \mu_\phi)\left\{(\underline{x}, \underline{y}) : \frac{1}{n} \log W_n(\underline{x}, \underline{y}) > h(\mu_\phi) + t\right\} \leq C_1 e^{-C_2 n t^2} \tag{8.4}$$

and

$$(\mu_\phi \otimes \mu_\phi)\left\{(\underline{x}, \underline{y}) : \frac{1}{n} \log W_n(\underline{x}, \underline{y}) < h(\mu_\phi) - t\right\} \leq C_1 e^{-C_2 n t}. \tag{8.5}$$

Let us notice that the upper tail estimate behaves differently than the lower tail estimate as a function of t. This asymmetric behavior also shows up in the large deviation asymptotics [24].

Let us sketch the strategy to prove Theorem 8.3.1. We cannot apply directly our concentration inequality to the random variable W_n for the following basic reason. Given $\underline{x}$ and $\underline{y}$, the first time that one sees the first n symbols of $\underline{x}$ in $\underline{y}$ is $W_n(\underline{x}, \underline{y})$ and assume it is finite. If we make $\underline{y}'$ by changing one symbol in $\underline{y}$, we have a priori no control on $W_n(\underline{x}, \underline{y})$ which can be arbitrarily larger than $W_n(\underline{x}, \underline{y})$ and even infinite. Of course, this situation is not typical, but we are forced to use the worst case to apply our concentration inequality. Roughly, we proceed as follows. We obviously have $\log W_n = \log(W_n \mu_\phi([X_0^{n-1}])) - \log \mu_\phi([X_0^{n-1}])$. On the one hand, we use a sharp approximation of the law of the random variables $W_n \mu_\phi([X_0^{n-1}])$ by an exponential law proved in [1]. On the other hand, by the Gibbs property, $\log \mu_\phi([x_0^{n-1}]) \approx \phi(\underline{x}) + \cdots + \phi(\sigma^{n-1}\underline{x})$ and we can the inequality (8.1) with $f = \phi$.

Proof. We first prove (8.4). We obviously have

$$
\begin{aligned}
&(\mu_\phi\otimes\mu_\phi)\left\{(\underline{x},\underline{y}) : \frac{1}{n}\log W_n(\underline{x},\underline{y}) > h(\mu_\phi)+t\right\}\\
&= (\mu_\phi\otimes\mu_\phi)\left\{(\underline{x},\underline{y}) : \frac{1}{n}\log W_n(\underline{x},\underline{y}) + \frac{1}{n}\log\mu_\phi([x_0^{n-1}]) - \frac{1}{n}\log\mu_\phi([x_0^{n-1}]) - h(\mu_\phi) > t\right\}\\
&\le (\mu_\phi\otimes\mu_\phi)\left\{(\underline{x},\underline{y}) : \frac{1}{n}\log\left[W_n(\underline{x},\underline{y})\mu_\phi([x_0^{n-1}])\right] > \frac{t}{2}\right\}\\
&\quad + \mu_\phi\left\{\underline{x} : -\frac{1}{n}\log\mu_\phi([x_0^{n-1}]) - h(\mu_\phi) > \frac{t}{2}\right\}\\
&=: T_1+T_2.
\end{aligned}
$$

We first derive an upper bound for T_2.
We use the Gibbs property, inequality (8.1) applied to $f=-\phi$ and (1.2) to get

$$
\begin{aligned}
T_2 &\le \mu_\phi\left\{-\frac{1}{n}\left(\phi+\cdots+\phi\circ\sigma^{n-1}\right) - h(\mu_\phi) > \frac{t}{2} - \frac{1}{n}\log C\right\}\\
&\le e^{-Bnt^2}
\end{aligned}
$$

for every t larger than $2\log C$.
We now derive an upper bound for T_1. To this end we apply the following result which we state as a lemma. It is an immediate consequence of Theorem 1 in [1].

Lemma 8.3.1 ([1])**.** *Let*

$$\tau_{[a_0^{n-1}]}(\underline{y}) := \inf\left\{j\ge 1 : y_j^{j+n-1} = a_0^{n-1}\right\}.$$

There exist strictly positive constants $C, c, \lambda_1, \lambda_2$, with $\lambda_1 < \lambda_2$, such that for every $n\in\mathbb{N}$, every string a_0^{n-1}, there exists $\lambda(a_0^{n-1})\in[\lambda_1,\lambda_2]$ such that

$$\left|\mu_\phi\left\{\underline{y} : \tau_{[a_0^{n-1}]}(\underline{y}) > \frac{u}{\lambda(a_0^{n-1})\mu_\phi([a_0^{n-1}])}\right\} - e^{-u}\right| \le Ce^{-cu}$$

for every $u>0$.

By definition and using the previous lemma we get

$$
\begin{aligned}
T_1 &= \sum_{a_0^{n-1}} \mu_\phi([a_0^{n-1}])\ \mu_\phi\left\{\underline{y} : \tau_{[a_0^{n-1}]}(\underline{y})\mu_\phi([a_0^{n-1}]) > e^{nt/2}\right\}\\
&\le C'\ e^{-c'e^{nt/2}}
\end{aligned}
$$

for some $c', C' > 0$.
Since the bound for T_1 is (much) smaller than the bound for T_2, we can bound T_1+T_2 by a constant times e^{-Bnt^2}. This yields (8.4).

We now turn to the proof of (8.5). We have

$$
\begin{aligned}
&(\mu_\phi\otimes\mu_\phi)\left\{(\underline{x},\underline{y}) : \frac{1}{n}\log W_n(\underline{x},\underline{y}) < h(\mu_\phi)-t\right\}\\
&=(\mu_\phi\otimes\mu_\phi)\left\{(\underline{x},\underline{y}) : -\frac{1}{n}\log W_n(\underline{x},\underline{y}) -\frac{1}{n}\log\mu_\phi([x_0^{n-1}])+\frac{1}{n}\log\mu_\phi([x_0^{n-1}])+h(\mu_\phi)>t\right\}\\
&\leq(\mu_\phi\otimes\mu_\phi)\left\{(\underline{x},\underline{y}) : -\frac{1}{n}\log\left[W_n(\underline{x},\underline{y})\mu_\phi([x_0^{n-1}])\right]>\frac{t}{2}\right\}\\
&\quad+\mu_\phi\left\{\underline{x} : \frac{1}{n}\log\mu_\phi([x_0^{n-1}])+h(\mu_\phi)>\frac{t}{2}\right\}\\
&=T_1'+T_2'.
\end{aligned}
$$

Proceeding as for T_2 (applying inequality (8.1) with $f=\phi$) we obtain the upper bound

$$
\begin{aligned}
T_2' &\leq \mu_\phi\left\{\frac{1}{n}\left(\phi+\cdots+\phi\circ\sigma^{n-1}\right)-\int\phi\mathrm{d}\mu_\phi>\frac{t}{2}-\frac{1}{n}\log C\right\}\\
&\leq e^{-Bnt^2}
\end{aligned}
$$

for some $C" > 0$ and for every $t > 2\log C$.

To bound T_1' we use the following lemma (Lemma 9 in [1]).

Lemma 8.3.2 ([1])**.** *For any $v>0$ and for any a_0^{n-1} such that $v\mu([a_0^{n-1}])\leq 1/2$, one has*

$$
\lambda_1\leq-\frac{\log\mu\left\{\tau_{[a_0^{n-1}]}>v\right\}}{v\mu([a_0^{n-1}])}\leq\lambda_2,
$$

where λ_1,λ_2 are the constants appearing in Lemma 8.3.1.

The previous lemma implies that

$$
\mu\left\{\tau_{[a_0^{n-1}]}\mu([a_0^{n-1}])<v\right\}\leq 1-e^{-v\lambda_2}\leq\lambda_2 v
$$

provided that $v\mu([a_0^{n-1}])\leq 1/2$. Taking $v=e^{-nt/2}$ it follows that

$$
\begin{aligned}
T_1' &= \sum_{a_0^{n-1}}\mu_\phi([a_0^{n-1}])\ \mu_\phi\left\{\underline{y} : \tau_{[a_0^{n-1}]}(\underline{y})\mu_\phi([a_0^{n-1}])<e^{-nt/2}\right\}\\
&\leq\lambda_2\ e^{-nt/2}.
\end{aligned}
$$

This inequality holds if $e^{-nt/2}\mu([a_0^{n-1}])\leq 1/2$, which is the case for any $n\geq 1$ if $t\geq 2\log 2$.

Inequality (8.5) follows from the bound we get for $T_1'+T_2'$. But the bound for T_1' is bigger than the one for T_2', whence the result.

The proof of the theorem is complete. □

Chapter 9

Perspectives & future work

In this chapter, we point out some open problems we intend to work on.

9.1 Deviation probability on the Alves-Viana map

The Alves-Viana map $T : \mathbb{S}^1 \times I \to \mathbb{S}^1 \times I$ is given by

$$T_{AV}(\omega, x) := (16\omega, a_0 + \varepsilon \sin(2\pi\omega) - x^2),$$

where $a_0 \in (1, 2)$, ε is small enough and I is a compact subinterval of $(-2, 2)$ such that T maps $\mathbb{S}^1 \times I$ into its interior.

S. Gouëzel proved ([34]) that in the Alves-Viana map the speed of decay of the correlations of Hölder functions is $\mathscr{O}\big(e^{-c\sqrt{n}}\big)$, which implies the central limit theorem. He used the strategy of building Young towers and some combinatorial techniques in the construction of the partition that allow to obtain better estimates that those obtainable by directly applying the result of [74].

In [3] the authors obtained a deviation probability for ergodic sums of Hölder functions in the Alves-Viana map. That is, for every ϕ Hölder continuous there exists $\tau = \tau(\phi) > 0$ and for all $t > 0$ exists $C = C(\phi, t) > 0$ such that

$$\mu\left(\left|\frac{1}{n}\sum_{i=0}^{n-1} \phi \circ T_{AV}^i - \int \phi \mathrm{d}\mu\right| > t\right) \leq Ce^{-\tau n^{1/5}}.$$

The exponent $1/5$ comes from a relation involving the corresponding exponent of decay of correlations, which is $e^{-c\sqrt{n}}$ as proved by Gouëzel.

One expect that is possible to obtain deviation probabilities in the Alves-Viana map of separately Lipschitz observables using the techniques developed in [21].

We present the following conjecture.

Conjecture 1. *Consider the Alves-Viana map, and let μ be its absolutely continuous invariant measure, then there exist constants M and c (depending only on the map) such* *that for any Lipschitz function K of n variables and for all $t > 0$*

$$\mu\left(\left|K(x, Tx, \ldots, T^{n-1}x) - \int K \mathrm{d}\mu\right| > t\right) \leq M \exp\Big(-ct^{\eta} \Big/ \big(\sum_{i=0}^{n-1} \mathrm{Lip}_i(K)^2\big)^{\eta/2}\Big),$$

for some $\eta \in (0, 2/5)$.

We conjecture $\eta \in (0, 2/5)$, to be consistent with the result in [3].

This would be a result giving a stretched exponential inequality and it would contribute to 'complete' the global picture of concentration inequalities in non-uniformly dynamical systems, since there are results for exponential concentration, as well as polynomial.

9.2 Concentration inequalities for random dynamical systems

Consider for instance the case of additive dynamical noise we barely described in section 3.1. Take $X \subset \mathbb{R}^d$ and let (ξ_n) be a sequence of X-valued random variables modeling the noise. The observed system is given by

$$x_{i+1} = T(x_i) + \varepsilon\xi_i,$$

where $\varepsilon > 0$ is the magnitude of the noise and is assumed to be small. As we saw in example 3.1.1 the process $\{x_n\}$ is a Markov chain. Assume μ_ε to be an invariant measure for the chain. Next, assuming the original system has a SRB measure μ, then one expect that it is the zero-noise limit of μ_ε. Indeed this was established for Axiom A systems and certain non-uniformly hyperbolic systems (see [7, 29] and [75] for a survey).

We believe that it concentration inequalities hold for random dynamical systems which are stochastically stable. In the case they actually do, this would bring quantitative information on, for instance, the distance between the empirical measure of the process $\{x_n\}$ and the SRB measure μ as a function of n and ε.

9.3 Convergence rate of Lalley's denoising algorithm for non-uniformly hyperbolic maps

In section 5.3 we presented a result on the convergence rate of Lalley's algorithm for Axiom A diffeomorfisms. Actually, in his Ph.D. thesis, J. Stover ([69]) worked in simulations of that algorithm. He tested it for the Smale's solenoid and the Hénon map. Nonetheless there is no proof on the convergence rate for non-uniformly hyperbolic systems.

This is an open problem that in principle it reduces (very roughly speaking) to exhibit a sequence of positive numbers β_n such that the average

$$\frac{1}{|A_i|} \sum_{j \in A_i} |x_i - x_j| \leq \beta_n,$$

where β_n converges to zero as $n \to \infty$. The rest of the bounds used to prove the consistency of the algorithm are explicit. We think that in the case of non-uniformly hyperbolic dynamical systems modeled by Young towers with exponential tails it is possible to show a polynomial rate of convergence.

9.4 More entropy estimators

We already shown some results on concentration bounds for entropy estimators in Gibbs measures. Besides the 'plug-in' estimator and hitting-times there exist several quantities that converge to the entropy.

A natural extension of our work would be finding concentration bounds for other estimators, we point out three of them.

9.4.1 Hochman's Estimator

Given a μ_ϕ-typical sequence $x_0 x_1 \cdots x_{n-1}$, the frequency of recurrence of the first k-block, is given by

$$F_{k,n(k)} := \frac{1}{n(k)} \# \left\{ 2 \leq j \leq n(k) \ : \ x_0^{k-1} = x_j^{j+k-1} \right\}.$$

By Birkhoff's ergodic theorem this quantity converges to $\mu_\phi(x_0^k)$. The Hochman's estimator[1] is defined by

$$\mathfrak{F}_{k,n(k)} := -\frac{1}{k} \log F_{k,n(k)}.$$

Using the Ornstein-Weiss and the Shannon-McMillan-Breiman's theorems, for an ergodic stationary process, if $n(k)$ is such that $e^{-hk} n(k) \to \infty$ (as $k \to \infty$) then

$$\lim_{k \to \infty} \mathfrak{F}_{k,n(k)} = h,$$

almost surely.

Just as for the previous estimators, we would like to know how it fluctuates.

9.4.2 Non-Sequential Recursive Pair Substitution

Let us barely describe the so-called non-sequential recursive pair substitution method (NSRPS). Assume that we are given with an infinite sequence $x_0 x_1 \cdots$ produced by a stationary source, where each x_i belongs to a finite alphabet A. The method substitutes all the non-overlapping apparitions of the pair of maximal frequency with a new symbol. This produces a new sequence with a new alphabet. The method is applied recursively.

Rigorous results were given in [6], it states that ergodic processes becomes Markovian in the limit. There the entropy can be calculated. We remit the reader to [6] for precise statements and proofs. In [14], the NSRPS method was numerically tested and compared with the empirical estimator and return times for several chaotic maps.

[1] We decided to name it Hochman's estimator, since it was defined by M. Hochman in [40] on a very general setting, and it also can be found in a simplified way in [39].

There are many open questions about this method, one of them is giving an estimate on its fluctuations.

9.4.3 Return times

We include in the list of problems as future work an open problem concerning the return times that was pointed out in [24].

Let $x_0 x_1 \cdots$ be a sequence produce by a stationary and ergodic source.The initial block of n symbols is denoted by x_0^{n-1}. The first recurrence time of the initial k-block is

$$R_k = \min\left\{ j \geq 1 \ : \ x_0^{n-1} = x_j^{j+n-1} \right\}.$$

An almost sure version of the Wyner-Ziv theorem ([72]) was proved by D. Ornstein and B. Weiss, it states that

$$\lim_{n\to\infty} \frac{1}{n} \log R_n = h, \quad \mu\text{-almost surely.}$$

Later, in [25] the authors studied the fluctuations of the return times for Gibbs measures with Hölder potentials. They established a central limit theorem for $\log R_n$ and the following large deviations result: There is a number $u_0 > 0$ such that for any $u \in [0, u_0)$ we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_\phi \left(\frac{1}{n} \log R_n > h + u \right) = -\mathcal{I}(h+u),$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_\phi \left(\frac{1}{n} \log R_n < h - u \right) = -\mathcal{I}(h-u),$$

where $\mathcal{I}$ is the Legendre transform of the function $F(q) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{a_0^{n-1}} (\mu_\phi[a_0^{n-1}])^{q+1}$.

Going further, one is interested in the following function of $q \in \mathbb{R}$ (provided the limit exists),

$$\mathcal{R}(q) := \lim_{n\to\infty} \frac{1}{n} \log \int R_n^q(\underline{x}) \mathrm{d}\mu(\underline{x}).$$

In [24] the authors obtained that for a Gibbs measure μ_ϕ which is not of maximal entropy, one has for all $u \geq 0$ that

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_\phi \left(\frac{1}{n} \log R_n > h(\mu) + u \right) = \inf_{q \geq 0} \{ -(h(\mu)+u)q + \mathcal{R}(q) \}.$$

The open problem is to calculate $\lim_{n\to\infty} \frac{1}{n} \log \int R_n^q \mathrm{d}\mu_\phi$ for all $q \in \mathbb{R}$ ([24]). One expects to have the same large deviations as for $\log W_n$. In section 7.3.2, we already mentioned the corresponding result for the waiting times ([24]). Thus the conjecture for the return times is that

$$\lim_{n\to\infty} \frac{1}{n} \log \int R_n^q \mathrm{d}\mu_\phi = \begin{cases} P((1-q)\phi) & \text{if } q > -1 \\ P(2\phi) & \text{if } q \leq -1. \end{cases}$$

This conjecture is supported by the numerical study in [37].

9.4.4 Estimation of entropy production

A way to quantify the entropy production or measuring the irreversibility of a process from a single trajectory was defined in [51]. Assume the process $\{X_n : n \in \mathbb{Z}\}$ to be Gibbsian. The space is $A^{\mathbb{Z}}$, where A is a finite alphabet. Consider a Hölder continuous potential $\phi : A^{\mathbb{Z}} \to \mathbb{R}$. Given a block x_1^n, its time-reverse its denoted by $x_n^1 = x_n x_{n-1} \cdots x_1$. From the time-reversion one may define the corresponding reversed potential ϕ^R and its Gibbs measure μ_{ϕ^R}. The entropy production of the process up to time n is defined as

$$\dot{\mathrm{S}}_n(\underline{x}) := \log \frac{\mu_\phi[x_1^n]}{\mu_\phi[x_n^1]} = \log \frac{\mu_\phi[x_1^n]}{\mu_{\phi^R}[x_1^n]}.$$

It is known that

$$\lim_{n\to\infty} \frac{\dot{\mathrm{S}}_n(\underline{x})}{n} = h(\mu_\phi \mid \mu_{\phi^R}) =: \mathrm{MEP} \qquad \mu_\phi - \text{almost surely},$$

where $h(\mu_\phi \mid \mu_{\phi^R})$ is the relative entropy of the measure μ_ϕ with respect to μ_{ϕ^R} and it is defined as the *mean entropy production*. In [23] it is proved the convergence, the central limit theorem and a large deviation principle for two estimators of the mean entropy production. These estimators are based in the hitting times studied in chapters 7 and 8.

The two estimators of the MEP are

$$\dot{\mathcal{S}}_n^H(\underline{x}) := \log \frac{\tau_{[x_n^1]}(\underline{x})}{\tau_{[x_1^n]}(\underline{x})} \qquad \text{and} \qquad \dot{\mathcal{S}}_n^W(\underline{x}, \underline{y}) := \log \frac{\tau_{[x_n^1]}(\underline{y})}{\tau_{[x_1^n]}(\underline{y})},$$

where $\tau_{[x_1^n]}(\underline{x}) := \inf\{j \geq 1 \ : \ \sigma^j \underline{x} \in [x_1^n]\}$. What one obtains is that if the process is not reversible $\tau_{[x_n^1]} \gg \tau_{[x_1^n]}$ typically and hence the estimators are typically positive. For this estimators one has the almost-sure convergence.

$$\lim_{n\to\infty} \frac{\dot{\mathcal{S}}_n^H}{n} = \mathrm{MEP} \quad \mu_\phi\text{–almost surely} \quad \text{and}$$

$$\lim_{n\to\infty} \frac{\dot{\mathcal{S}}_n^W}{n} = \mathrm{MEP} \quad \mu_\phi \times \mu_\phi\text{–almost surely.}$$

As we already said these estimators satisfy a central limit theorem and a large deviation principle. We are interested in providing concentration bounds for the estimators of the MEP.

9.5 Concentration inequality for Gibbs measures with non Hölder potentials

In section 2.5 we presented the proof for the exponential concentration inequality for Gibbs measures associated to a Hölder continuous potential ϕ. In that setting one has also the central limit theorem and a large deviation principle.

We are interested in the natural extension for non-Hölder potentials. In particular those satisfying $\sum_{n=0}^{\infty} n\mathrm{var}_n\phi < \infty$. For this class of potentials, in her thesis ([56]), V. Maume-Deschamps established the central limit theorem for Lipschitz functions with respect to the metric given by

$$d_0(\underline{x}, \underline{y}) := \sup_{\underline{x},\underline{y}\in[a_0^{n-1}]} \sup_{k\in\mathbb{N}} \sup_{a_0^{k-1}} \left| \sum_{i=0}^{k-1} \phi \circ \sigma^i(a_0^{k-1}\underline{x}) - \phi \circ \sigma^i(a_0^{k-1}\underline{y}) \right| .$$

On the other hand, in the case of potentials satisfying $\sum_{n=0}^{\infty} \mathrm{var}_n\phi < \infty$ one has a large deviation principle (see for instance [52]). As we have seen by the end of section 2.3.2 that the concentration inequalities are compatible with the central limit theorem and large deviations so, one might expect that Gibbs measures with this class of potentials enjoy also concentration inequalities. If so, this would give an example of system satisfying an exponential concentration inequality without spectral gap for the transfer operator.

Bibliography

[1] Miguel Abadi. Sharp error terms and necessary conditions for exponential hitting times in mixing processes. *Ann. Probab.*, 32(1A):243–264, 2004.

[2] Henry D. I. Abarbanel. *Analysis of observed chaotic data*. Institute for Nonlinear Science. Springer-Verlag, New York, 1996.

[3] José F. Alves, Jorge M. Freitas, Stefano Luzzatto, and Sandro Vaienti. From rates of mixing to recurrence times via large deviations. *Adv. Math.*, 228(2):1203–1236, 2011.

[4] András Antos and Ioannis Kontoyiannis. Convergence properties of functional estimates for discrete distributions. *Random Structures Algorithms*, 19(3-4):163–193, 2001. Analysis of algorithms (Krynica Morska, 2000).

[5] Ludwig Arnold. *Random dynamical systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998.

[6] Dario Benedetto, Emanuele Caglioti, and Davide Gabrielli. Non-sequential recursive pair substitution: some rigorous results. *J. Stat. Mech. Theory Exp.*, (9):P09011, 21 pp. (electronic), 2006.

[7] Michael Benedicks and Marcelo Viana. Random perturbations and statistical properties of Hénon-like maps. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 23(5):713–752, 2006.

[8] Michael Benedicks and Lai-Sang Young. Sinaĭ-Bowen-Ruelle measures for certain Hénon maps. *Invent. Math.*, 112(3):541–576, 1993.

[9] Rabi Bhattacharya and Mukul Majumdar. *Random dynamical systems*. Cambridge University Press, Cambridge, 2007. Theory and applications.

[10] Stéphane Boucheron, Olivier Bousquet, Gábor Lugosi, and Pascal Massart. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560, 2005.

[11] Stéphane Boucheron, Gábor Lugosi, and Olivier Bousquet. Concentration inequalities. In Olivier Bousquet, Ulrike von Luxburg, and Gunnar Rtsch, editors, *Advanced*

Lectures on Machine Learning, volume 3176 of *Lecture Notes in Computer Science*, pages 208–240. Springer Berlin / Heidelberg, 2004. 10.1007/978-3-540-28650-9-9.

[12] Rufus Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, revised edition, 2008. With a preface by David Ruelle, Edited by Jean-René Chazottes.

[13] Jérôme Buzzi. Absolutely continuous S.R.B. measures for random Lasota-Yorke maps. *Trans. Amer. Math. Soc.*, 352(7):3289–3303, 2000.

[14] Lucio M. Calcagnile, Stefano Galatolo, and Giulia Menconi. Non-sequential recursive pair substitutions and numerical entropy estimates in symbolic dynamical systems. *J. Nonlinear Sci.*, 20(6):723–745, 2010.

[15] Robert Cawley and Guan-Hsong Hsu. Local-geometric-projection method for noise reduction in chaotic maps and flows. *Phys. Rev. A (3)*, 46(6):3057–3082, 1992.

[16] J.-R. Chazottes. *Fluctuations et Théorèmes Limites pour des Systèmes Dynamiques et des Measures de Gibbs*. Université Paris XI, Orsay, 2010. HDR.

[17] J.-R. Chazottes. Fluctuations of observables in dynamical systems: from limit theorems to concentration inequalities. In *Nonlinear Dynamics: New Directions. Dedicated to Valentin Afraimovich on the occasion of his 65th birthday*. To appear, 2011.

[18] J.-R. Chazottes, P. Collet, F. Redig, and E. Verbitskiy. A concentration inequality for interval maps with an indifferent fixed point. *Ergodic Theory Dynam. Systems*, 29(4):1097–1117, 2009.

[19] J.-R. Chazottes, P. Collet, and B. Schmitt. Statistical consequences of the Devroye inequality for processes. Applications to a class of non-uniformly hyperbolic dynamical systems. *Nonlinearity*, 18(5):2341–2364, 2005.

[20] J.-R. Chazottes and D. Gabrielli. Large deviations for empirical entropies of g-measures. *Nonlinearity*, 18(6):2545–2563, 2005.

[21] J.-R. Chazottes and S. Gouëzel. Optimal concentration inequalities for dynamical systems. *arXiv: 1111.0849. preprint*, To appear in Comm. Math. Phys., 2011.

[22] J-R. Chazottes and C. Maldonado. Concentration bounds for entropy estimation of one-dimensional gibbs measures. *Nonlinearity*, 24(8):2371, 2011.

[23] J.-R. Chazottes and F. Redig. Testing the irreversibility of a Gibbsian process via hitting and return times. *Nonlinearity*, 18(6):2477–2489, 2005.

[24] J.-R. Chazottes and E. Ugalde. Entropy estimation and fluctuations of hitting and recurrence times for Gibbsian sources. *Discrete Contin. Dyn. Syst. Ser. B*, 5(3):565–586, 2005.

[25] P. Collet, A. Galves, and B. Schmitt. Repetition times for Gibbsian sources. *Nonlinearity*, 12(4):1225–1237, 1999.

[26] P. Collet and Y. Levy. Ergodic properties of the Lozi mappings. *Comm. Math. Phys.*, 93(4):461–481, 1984.

[27] P. Collet, S. Martinez, and B. Schmitt. Exponential inequalities for dynamical measures of expanding maps of the interval. *Probab. Theory Related Fields*, 123(3):301–322, 2002.

[28] Pierre Collet and Jean-Pierre Eckmann. *Concepts and results in chaotic dynamics: a short course*. Theoretical and Mathematical Physics. Springer-Verlag, Berlin, 2006.

[29] William Cowieson and Lai-Sang Young. SRB measures as zero-noise limits. *Ergodic Theory Dynam. Systems*, 25(4):1115–1138, 2005.

[30] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[31] J. Doyne Farmer and John J. Sidorowich. Optimal shadowing and noise reduction. *Phys. D*, 47(3):373–392, 1991.

[32] R. Fernández and G. Maillard. Chains with complete connections and one-dimensional Gibbs measures. *Electron. J. Probab.*, 9:no. 6, 145–176 (electronic), 2004.

[33] D. Gabrielli, A. Galves, and D. Guiol. Fluctuations of the empirical entropies of a chain of infinite order. *Math. Phys. Electron. J.*, 9:Paper 5, 17 pp. (electronic), 2003.

[34] Sébastien Gouëzel. Decay of correlations for nonuniformly expanding systems. *Bull. Soc. Math. France*, 134(1):1–31, 2006.

[35] Peter Grassberger, Rainer Hegger, Holger Kantz, Carsten Schaffrath, and Thomas Schreiber. On noise reduction methods for chaotic data. *Chaos*, 3(2):127–141, 1993.

[36] Peter Grassberger and Itamar Procaccia. Measuring the strangeness of strange attractors. *Phys. D*, 9(1-2):189–208, 1983.

[37] Nicolai Hadyn, José Luevano, Giorgio Mantica, and Sandro Vaienti. Multifractal properties of return time statistics. *Phys. Rev. Lett.*, 88:224502, May 2002.

[38] Stephen M. Hammel. A noise reduction method for chaotic systems. *Phys. Lett. A*, 148(8-9):421–428, 1990.

[39] Michael Hochman. Recurrence densities for stationary processes. Unpublished notes.

[40] Michael Hochman. Return times, recurrence densities and entropy for actions of some discrete amenable groups. *J. Anal. Math.*, 100:1–51, 2006.

[41] Brian R. Hunt, Judy A. Kennedy, Tien-Yien Li, and Helena E. Nusse, editors. *The theory of chaotic attractors.* Springer-Verlag, New York, 2004. Dedicated to James A. Yorke in commemoration of his 60th birthday.

[42] Marius Iosifescu. Sampling entropy for random homogeneous systems with complete connections. *Ann. Math. Statist*, 36:1433–1436, 1965.

[43] Holger Kantz and Thomas Schreiber. *Nonlinear time series analysis.* Cambridge University Press, Cambridge, second edition, 2004.

[44] Yuri Kifer. *Random perturbations of dynamical systems*, volume 16 of *Progress in Probability and Statistics.* Birkhäuser Boston Inc., Boston, MA, 1988.

[45] Yuri Kifer and Pei-Dong Liu. Random dynamics. In *Handbook of dynamical systems. Vol. 1B*, pages 379–499. Elsevier B. V., Amsterdam, 2006.

[46] Eric J. Kostelich and Thomas Schreiber. Noise reduction in chaotic time-series data: a survey of common methods. *Phys. Rev. E (3)*, 48(3):1752–1763, 1993.

[47] Steven P. Lalley. Beneath the noise, chaos. *Ann. Statist.*, 27(2):461–479, 1999.

[48] Steven P. Lalley. Removing the noise from chaos plus noise. In *Nonlinear dynamics and statistics (Cambridge, 1998)*, pages 233–244. Birkhäuser Boston, Boston, MA, 2001.

[49] Steven P. Lalley and A. B. Nobel. Denoising deterministic time series. *Dyn. Partial Differ. Equ.*, 3(4):259–279, 2006.

[50] Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2001.

[51] Christian Maes. The fluctuation theorem as a Gibbs property. *J. Statist. Phys.*, 95(1-2):367–392, 1999.

[52] Christian Maes and Evgeny Verbitskiy. Large deviations and a fluctuation symmetry for chaotic homeomorphisms. *Comm. Math. Phys.*, 233(1):137–151, 2003.

[53] Cesar Maldonado. Fluctuation bounds for chaos plus noise in dynamical systems. *Journal of Statistical Physics*, 148:548–564, 2012. 10.1007/s10955-012-0553-3.

[54] Ph. Marie. *Propriétés Statistiques des Systèmes Dynamiques Déterministes et Aléatoires.* PhD thesis, Université d'Aix-Marseille II, 2009.

[55] P. F. Marteau and H. D. I. Abarbanel. Noise reduction in chaotic time series using scaled probabilistic methods. *J. Nonlinear Sci.*, 1(3):313–343, 1991.

[56] V. Maume-Deschamps. *Propriètés de mélange pour des systèmes dynamiques markoviens.* PhD thesis, Université de Bourgogne, Dijon, 1998.

[57] V. Maume-Deschamps. Exponential inequalities and estimation of conditional probabilities. In *Dependence in probability and statistics*, volume 187 of *Lecture Notes in Statist.*, pages 123–140. Springer, New York, 2006.

[58] K. McGoff, S. Mukherjee, and N. Pillai. Statistical inference for dynamical systems: A review. *preprint.*, http://arxiv.org/abs/1204.6265, 2012.

[59] Donald S. Ornstein and Benjamin Weiss. How sampling reveals a process. *Ann. Probab.*, 18(3):905–930, 1990.

[60] William Parry and Mark Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque*, (187-188):268, 1990.

[61] Ya. B. Pesin. On rigorous mathematical definitions of correlation dimension and generalized spectrum for dimensions. *J. Statist. Phys.*, 71(3-4):529–547, 1993.

[62] Luc Rey-Bellet and Lai-Sang Young. Large deviations in non-uniformly hyperbolic dynamical systems. *Ergodic Theory Dynam. Systems*, 28(2):587–612, 2008.

[63] Tim Sauer. A noise reduction method for signals from nonlinear systems. *Phys. D*, 58(1-4):193–201, 1992. Interpretation of time series from nonlinear systems (Warwick, 1991).

[64] Thomas Schreiber. Extremely simple nonlinear noise-reduction method. *Phys. Rev. E*, 47:2401–2404, Apr 1993.

[65] Thomas Schreiber and Peter Grassberger. A simple noise-reduction method for real data. *Phys. Lett. A*, 160(5):411–418, 1991.

[66] Thomas Schreiber and Holger Kantz. Noise in chaotic data: Diagnosis and treatment. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):133–142, 1995.

[67] Regis J. Serinko. Ergodic theorems arising in correlation dimension estimation. *J. Statist. Phys.*, 85(1-2):25–40, 1996.

[68] P. C. Shields. *The ergodic theory of discrete sample paths*, volume 13 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 1996.

[69] J. H. Stover. *Filtering and Estimation of Noise-contaminated Chaotic Time Series.* PhD thesis, Purdue University, West Lafayette, Indiana, 2000.

[70] Marek Strumik, Wiesław M. Macek, and Stefano Redaelli. Discriminating additive from dynamical noise for chaotic time series. *Phys. Rev. E*, 72:036219, Sep 2005.

[71] Peter Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.

[72] Aaron D. Wyner and Jacob Ziv. Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inform. Theory*, 35(6):1250–1258, 1989.

[73] Lai-Sang Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2)*, 147(3):585–650, 1998.

[74] Lai-Sang Young. Recurrence times and rates of mixing. *Israel J. Math.*, 110:153–188, 1999.

[75] Lai-Sang Young. What are SRB measures, and which dynamical systems have them? *J. Statist. Phys.*, 108(5-6):733–754, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.