A concentration inequality for interval maps with an indifferent fixed point

J.-R. Chazottes, P. Collet
Centre de Physique Théorique
CNRS, Ecole Polytechnique
91128 Palaiseau Cedex, France
e-mail: chazottes@cpht.polytechnique.fr,
collet@cpht.polytechnique.fr
F. Redig
Mathematisch Instituut Universiteit Leiden
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
e-mail: redig@math.leidenuniv.nl
E. Verbitskiy
Philips Research, HTC 36 (M/S 2), 5656 AE Eindhoven The Netherlands
e-mail: evgeny.verbitskiy@philips.com

February 2, 2008

Abstract

For a map of the unit interval with an indifferent fixed point, we prove an upper bound for the variance of all observables of n variables $K : [0,1]^n \to \mathbb{R}$ which are componentwise Lipschitz. The proof is based on coupling and decay of correlation properties of the map. We then give various applications of this inequality to the almost-sure central limit theorem, the kernel density estimation, the empirical measure and the periodogram.

key-words: variance, componentwise Lipschitz observable, almost-sure central limit theorem, kernel density estimation, empirical measure, periodogram, shadowing, Kantorovich-Rubinstein theorem.

Contents

1 Introduction

Nowadays, concentration inequalities are a fundamental tool in probability theory and statistics. We refer the reader to, *e.g.*, [14, 11, 16, 17, 20]. In particular, they also turn out to be essential tools to develop a non-asymptotic theory in statistics, exactly as the central limit theorem and large deviations are known to play a central part in the asymptotic theory. Besides the non-asymptotic aspect of concentration inequalities, the crucial point is that they allow in principle to study random variables $Z_n = K(X_1, \ldots, X_n)$ that "smoothly" depend on the underlying random variables X_i, but otherwise can be defined in an indirect or a complicated way, and for which explicit computations can be very hard, even in the case where the X_i's are independent.

In the context of dynamical systems, central limit theorems and their refinements, large deviations, and other type of limit theorems have been proved, almost exclusively for Birkhoff sums of sufficiently "smooth" observables. But many natural observables are not Birkhoff sums. Let us just mention a typical example (see below for more examples), namely the so-called power spectrum, that is, the Fourier transform of the correlation function, whose estimator is the integral of the periodogram. This is a very complicated quantity from the analytic point of view. Besides the computational difficulties proper to each observable, one would like to have a systematic method to approach the questions of fluctuations of observables, instead of designing a particular method for each case.

A possible method is concentration inequalities. An additional difficulty comes in for dynamical systems, namely the fact that we loose independence, except in very special cases, and that the mixing properties of dynamical systems are not as nice as for stochastic processes encountered usually in probability theory, such as Markov chains, renewal processes, *etc.* So new approaches have to be proposed, based on typical tools of dynamical systems like the spectral gap (when it exists) of the transfer operator and the decay of correlations. The first concentration inequality in this context was obtained by Collet *et al.* [9] for uniformly expanding maps of the interval, without assuming the existence of a Markov partition. They obtained the so-called Gaussian concentration inequality (also called exponential concentration inequality) by bounding the exponential moment of any observable of n variables only assuming that it is componentwise Lipschitz. They deduced several applications (kernel density estimation, shadowing, etc). In the hope of proving concentration inequalities for more general dynamical systems, one can start with an inequality for the variance, leading to a polynomial concentration inequality. This was indeed done in [5] for a large class of non-uniformly hyperbolic systems modeled by a "Young tower with exponential return times" [21]. In [6], the authors of [5] showed the usefulness of this variance inequality (therein called "Devroye inequality") through various examples. Let us also mention another approach based on coupling [4, 8] that gives, *e.g.*, an altenative proof of the Gaussian deviation inequality in the case of uniformly expanding maps of the interval, and also used in the context of Gibbs random fields.

Regarding Birkhoff sums of "smooth" observables (*e.g.*, Hölder), central limit theorems and large deviation estimates have been proved both for systems modeled by a Young towers with exponential return-time tail mentioned above and those with a summable return-time tail, see, *e.g.*, [21, 22, 18, 19]. So, a natural question is to try to prove an inequality for the variance of any observable of n variables only assuming it is componentwise Lipschitz, as in [5], but relaxing the exponential decay of the return-time tail of Young towers [22]. This would give a way to analyze fluctuations of complicated observables, which are not Birkhoff sums. The simplest and classical example is a map of the unit interval with an indifferent fixed point. In this paper, we prove a variance inequality for the map $T(x) = x + 2^\alpha x^{1+\alpha}$ when $x \in [0, 1/2[$ and strictly expanding on $[1/2, 1]$, when α is small enough (Theorem 3). However, the proof verbatim applies to the class of maps with a unique indifferent fixed point considered in [13]. The major difference with the situation in [5, 9] is that the transfer operator has no spectral gap and that the decay of correlations is polynomial instead of being exponential. Therefore we develop a different approach based on decay of correlations. We need to control the covariance of C^0 functions and Lipschitz functions, which is done by H. Hu [13]. An important ingredient is coupling through the Kantorovich-Rubinstein duality theorem. At present, we are not able to construct explicitely a coupling for the backward process as the one constructed in [1] for uniformly expanding maps of the interval. This explicit coupling was used in [8] in order to prove the Gaussian concentration inequality. After proving the variance inequality, we show various applications of it, namely, to the almost-sure central limit theorem, the kernel density estimation, the empirical measure, the integrated periodogram and the shadowing.

The paper is organized as follows. Section 2 contains the necessary informations on the maps, while Section 3 contains our main result, namely the variance inequality. In Section 4 we give various applications of it. Section 5 contains the proof of the Devroye inequality.

2 The map and its properties

2.1 The map and the invariant measure

For the sake of definiteness, we consider the maps $T : [0, 1] \circlearrowleft$ such that on $[0, 1/2[$

$$T(x) = x + 2^\alpha x^{1+\alpha}$$

and such that $|T'x| > 1$ and $|T"(x)| < \infty$ on $[1/2, 1]$. In fact, all what follows is valid under the assumptions of H. Hu [13].

For $\alpha \in [0, 1[$, this map admits an absolutely continuous invariant probability measure $d\mu(x) = h(x)dx$, where $h(x) \sim x^{-\alpha}$ when x tends to 0.

We define the sequence of points x_ℓ by $x_0 = 1$, $x_1 = 1/2$ and for $\ell \geq 2$ $T(x_\ell) = x_{\ell-1}$ and $x_\ell < 1/2$. It is easy to verify that the sequence of intervals

$$I_\ell :=]x_{\ell+1}, x_\ell],$$

for $\ell = 0, 1, 2, \ldots$, is a Markov partition of the interval $]0, 1]$.

We have the behavior, see *e.g.* [13],

$$|I_\ell| \sim \ell^{-\frac{1}{\alpha}-1},\ x_\ell \sim \ell^{-\frac{1}{\alpha}}. \tag{1}$$

2.2 Decay of correlations

The covariance or correlation coefficient $\mathrm{Cov}_{v,w}(\ell)$ of two $L^2(\mu)$ functions $v, w : [0,1] \to \mathbb{R}$ is defined, as usual, by

$$\mathrm{Cov}_{v,w}(\ell) = \int v \circ T^\ell \; w \; d\mu - \int v \; d\mu \int w \; d\mu.$$

When $v = w$, we simply write Cov_v.

Various people established the (optimal) decay of correlations for the map T, namely $\mathrm{Cov}_{v,w}(\ell) \approx \ell^{-\frac{1}{\alpha}+1}$. In, *e.g.*, [22] this is proved for u, v both being Hölder. As it will turn out, we need the following estimate proved in [13]. There exists a constant $C > 0$ such that, for all $v \in C^0$ and w Lipschitz, we have the following decay:

$$|\mathrm{Cov}_{v,w}(\ell)| \leq C \, \|v\|_{C^0} \; \mathrm{Lip}(w) \; \gamma_\ell \tag{2}$$

where

$$\gamma_\ell := \ell^{-\frac{1}{\alpha}+1} \tag{3}$$

and where

$$\mathrm{Lip}(w) = \sup_{x \neq x'} \frac{|w(x) - w(x')|}{|x - x'|} \cdot$$

This follows from [13]. The fact that C *does not depend* on v, w is the consequence of Theorem B.1 in [6].

2.3 Central limit theorem

Using [13, Proposition 5.2] and [15], we have a central limit theorem for Lipschitz observables when $0 < \alpha < 1/2$: for any v Lipschitz which is not of the form $h - h \circ T$ and such that $\int v d\mu = 0$, we have

$$\mu\left(\frac{1}{\sigma_v \sqrt{n}} \sum_{j=0}^{n-1} v \circ T^j \leq t \right) \xrightarrow{n \to \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\xi^2/2} \, d\xi, \;\; \forall t \in \mathbb{R}, \tag{4}$$

where

$$\sigma_v^2 = \mathrm{Cov}_v(0) + 2 \sum_{\ell=1}^{\infty} \mathrm{Cov}_v(\ell) > 0.$$

3 Variance inequality

Our main theorem is an upper-bound for the variance of any componentwise Lipschitz function.

We introduce the convenient notations

$$T_p^q(x) = (T^p(x), T^{p+1}(x), \dots, T^q(x)) \quad \text{and} \quad z_p^q = z_p, z_{p+1}, \dots, z_q$$

for $0 \le p \le q$. With this notation, if we take a function K of n variables, we write, *e.g.*, $K(z_1^j, z_j, z_{j+1}^n)$ for $K(z_1, z_2, \dots, z_n)$.

A real-valued function K on $[0,1]^n$ is said to be componentwise Lipschitz if, for all $1 \le j \le n$, the following quantities are finite:

$$\mathrm{Lip}_j(K) := \sup_{z_1, z_2, \dots, z_j, z_{j+1}, \dots, z_n} \sup_{z_j \neq \hat{z}_j} \frac{\left| K(z_1^{j-1}, z_j, z_{j+1}^n) - K(z_1^{j-1}, \hat{z}_j, z_{j+1}^n) \right|}{|z_j - \hat{z}_j|}.$$

Our main theorem reads as follows.

Theorem 3.1. *Let T be the map defined in Section 2. Then, for any $\alpha \in [0, 4 - \sqrt{15}[$, there exists $D = D(\alpha) > 0$ such that, for any componentwise Lipschitz function $K : [0,1]^n \to \mathbb{R}$, we have*

$$\int \left(K(T_0^{n-1}(x)) - \int K(T_0^{n-1}(y))\, d\mu(y) \right)^2 d\mu(x) \le D \sum_{j=1}^n (\mathrm{Lip}_j(K))^2. \quad (5)$$

(This inequality is called "Devroye inequality" in [5, 6].)

An application of Chebychev's inequality immediately yields the following concentration inequality.

Corollary 3.2. *Under the assumptions of Theorem 3, we have*

$$\mu\left(x \in [0,1] : \left| K(T_0^{n-1}(x)) - \int K(T_0^{n-1}(y))\, d\mu(y) \right| \ge t \right) \le \frac{D \sum_{j=1}^n \mathrm{Lip}_j(K)^2}{t^2}$$

for all $t > 0$.

Remark 3.3. *In our context, we cannot expect a Gaussian concentration bound. This would give a Gaussian concentration inequality incompatible with large deviation lower bounds obtained in [18] where, for a large class of Hölder observables v, it is proved that for $\epsilon > 0$ small enough*

$$\mu\left(\left\{ x \in [0,1] : \left| \frac{S_n v(x)}{n} - \int v d\mu \right| > \epsilon \right\} \right) \ge n^{-(\frac{1}{\alpha} - 1 + \delta)}$$

for any $\delta > 0$ and infinitely many n's, where $S_n v = v + v \circ T + \cdots + v \circ T^{n-1}$. This type of inequalities was also obtained in [10] under different conditions.

4 Some applications

We now give some applications of the variance inequality (5). We follow [6] where we obtained them in an abstract setting: therein we assumed that (X_k) was some real-valued, stationary, ergodic process satisfying (5), plus eventually an extra condition on the auto-covariance of Lipschitz observables, depending on each specific application. By (2) we have

$$\sum_{\ell=1}^{\infty} \mathrm{Cov}_v(\ell) \leq C'(\mathrm{Lip}(v))^2$$

where $C' = C\sum_\ell \gamma_\ell < \infty$. This condition will be sufficient to apply all the results from [6] that we will use.

The standing assumption in this section is that $0 < \alpha < 4 - \sqrt{15}$, so that Theorem 3 holds.

4.1 Almost-sure central limit theorem

For an observable v such that $\int v d\mu = 0$, define the sequence of weighted empirical (random) measures of the normalized Birkhoff sum by

$$\mathcal{A}_n(x) = \frac{1}{D_n} \sum_{k=1}^{n} \frac{1}{k}\, \delta_{S_k v(x)/\sqrt{k}}$$

where $D_n = \sum_{k=1}^{n} \frac{1}{k}$.

We say that the almost-sure central limit theorem holds if for μ almost every x, $\mathcal{A}_n(x)$ converges weakly to the Gaussian measure. In fact, we will prove a stronger statement, namely that the convergence takes place in the Kantorovich distance.

Let us recall that the Kantorovich distance between two probability measures μ_1 and μ_2 on $\mathbb{R}$ is defined by

$$\kappa(\mu_1, \mu_2) = \sup_{g\in\mathcal{L}} \int g(\xi)\, d\big(\mu_1 - \mu_2\big)(\xi) \tag{6}$$

where $\mathcal{L}$ denotes the set of real-valued Lipschitz functions on $\mathbb{R}$ with Lipschitz constant at most one.

We denote by $\mathscr{N}(0, \sigma_v^2)$ the Gaussian measure with mean zero and variance σ_v^2.

Theorem 4.1. *Let v be a Lipschitz function which is not of the form $h - h\circ T$ and assume that $\int v d\mu = 0$. Then, for μ almost every x, one has*

$$\lim_{n\to\infty} \kappa\left(\mathcal{A}_n(x), \mathscr{N}(0, \sigma_v^2)\right) = 0.$$

The theorem is an immediate application of Theorem 8.1 in [6] and (4).

Notice that this theorem immediately implies that for μ almost every x $\mathcal{A}_n(x)$ converges weakly to the Gaussian measure. The weak convergence is proved in [7] by another method (and not only for the present intermittent map). In [3], a speed of convergence in the Kantorovich distance was obtained for uniformly expanding maps of the interval using a Gaussian bound.

4.2 Kernel density estimation

We consider the sequence of regularized (random) empirical measures $\mathcal{H}_n(x)$ with densities (h_n) defined by

$$h_n(x;s) = \frac{1}{na_n}\sum_{j=1}^{n}\psi((s-T^j(x))/a_n)$$

where a_n is a positive sequence converging to 0 and such that na_n converges to $+\infty$, and ψ (the kernel) is a bounded, non-negative, Lipschitz continuous function with compact support whose integral equals 1. We are interested in the convergence in $L^1(ds)$ of this empirical density $h_n(x;\cdot)$ to the density $h(\cdot)$ of the invariant measure $d\mu(x) = h(x)dx$. This is nothing but the distance in total variation between $\mathcal{H}_n(x)$ and μ:

$$\mathrm{dist}_{\mathrm{TV}}(\mathcal{H}_n(x),\mu) = \int |h_n(x;s) - h(s)|\, ds.$$

Theorem 4.2. *Let ψ and a_n be as just described. Then, there exists a constant $C = C(\psi) > 0$ such that for any integer n and for any $t > C(a_n^{1-\alpha} + 1/(\sqrt{n}a_n^2))$, we have*

$$\mu\left(\{x\in[0,1] : \mathrm{dist}_{\mathrm{TV}}(\mathcal{H}_n(x),\mu) > t\}\right) \leq \frac{C}{t^2 n a_n^2}.$$

This theorem is a direct consequence of Theorem 6.1 in [6] (with $\tau = 1-\alpha$).

4.3 Empirical measure

The empirical measure associated to $x, Tx, \ldots, T^{n-1}x$ is the random measure on $[0,1]$ defined by

$$\mathcal{E}_n(x) = \frac{1}{n}\sum_{j=0}^{n-1}\delta_{T^j(x)}$$

where δ is the Dirac measure. From Birkhoff's ergodic theorem, for μ almost every x this sequence of random measures weakly converges to μ. We want to estimate the speed of this convergence with respect to the Kantorovich distance (6) (now used for probability measures on $[0,1]$).

Theorem 4.3. *There exists a positive constant C such that for all $t>0$ and $n \geq 1$, we have*

$$\mu\left(\left\{x \in [0,1] : \kappa(\mathcal{E}_n(x),\mu) > t + \frac{C}{n^{1/4}}\right\}\right) \leq \frac{D}{nt^2}.$$

This is an immediate consequence of Theorem 5.2 in [6].

4.4 Integrated periodogram

Let v be an $L^2(\mu)$ observable and assume, for the sake of simplicity, that $\int v d\mu = 0$. We recall (see, *e.g*, [2]) that the raw periodogram (of order n) of the process $(v \circ T^k)$ is the random variable

$$I_n^v(\omega;x) = \frac{1}{n}\left|\sum_{j=1}^{n} e^{-ij\omega} \ \big(v(T^j(x))\big)\right|^2$$

where $\omega \in [0,2\pi]$. The spectral distribution function of order n (integral of the raw periodogram of order n) is given by

$$J_n^v(\omega;x) = \int_0^\omega I_n^v(s;x)\ ds\,.$$

Let $\hat{C}_v(\omega)$ be the Fourier cosine transform of the auto-covariance of v, namely

$$\hat{C}_v(\omega) = \sum_{k=0}^{\infty} \cos(\omega k)\ \mathrm{Cov}_v(k+1)\,.$$

We will denote by $J^v(\omega)$ the following quantity

$$J^v(\omega) = \int_0^\omega (2\hat{C}_v(s) - \mathrm{Cov}_v(0))\ ds = \mathrm{Cov}_v(0)\ \omega + 2\sum_{k=1}^{\infty}\frac{\sin(\omega k)}{k}\ \mathrm{Cov}_v(k)\,.$$

Theorem 4.4. *Let v be a Lipschitz observable. Then there exists a positive constant $C = C(v)$ such that for any $n \geq 1$, one has*

$$\int \Big(\sup_{\omega\in[0,2\pi]} \big|\tilde{J}_n^v(\omega;x) - J^v(\omega)\big|\Big)^2 d\mu(x) \leq C\,\frac{(1+\log n)^{4/3}}{n^{2/3}}.$$

This theorem is a direct application of Theorem 3.1, and the remark just after it, in [6].

4.5 Shadowing and mismatch

Let A be a set of initial conditions with positive measure. If $x \notin A$, we can ask how well we can approximate the orbit of x by an orbit starting from an initial condition in A.

We can measure the average quality of "shadowing" by the following quantity:

$$\mathcal{Z}_A(x) = \frac{1}{n} \inf_{y \in A} \sum_{j=0}^{n-1} |T^j(x) - T^j(y)|.$$

Theorem 4.5. *Let A be a subset of positive measure. Then, for all $n \geq 1$, for all $t > 0$, one has*

$$\mu\left(\left\{x \in [0,1] : \mathcal{Z}_A(x) \geq \frac{1}{n^{\frac{1}{3}}}\left(t + \frac{2^{\frac{4}{3}} D^{\frac{1}{3}}}{\mu(A)}\right)\right\}\right) \leq \frac{D}{n^{\frac{1}{3}} t^2}.$$

We can also look at the number of mismatch at a given precision: for $\epsilon > 0$, let

$$\mathcal{Z}'_{A,\epsilon}(x) = \frac{1}{n} \inf_{y \in A} \operatorname{Card}\{0 \leq j \leq n-1 \,:\, |T^j(x) - T^j(y)| > \epsilon\}.$$

Theorem 4.6. *Let A be a subset of positive measure. Then, for all $n \geq 1$, for all $t > 0$, for any $\epsilon > 0$, one has*

$$\mu\left(\left\{x \in [0,1] : \mathcal{Z}'_{A,\epsilon}(x) \geq \frac{1}{\epsilon^{\frac{2}{3}} n^{\frac{1}{3}}}\left(t + \frac{2^{\frac{4}{3}} D^{\frac{1}{3}}}{\mu(A)}\right)\right\}\right) \leq \frac{D}{\epsilon^{\frac{2}{3}} n^{\frac{1}{3}} t^2}.$$

Theorem 4.5 is a direct application of Theorem 7.1 in [6] whereas Theorem 4.6 is a direct application of Theorem 7.2 in [6].

5 Proof of Theorem 3.1

5.1 First telescoping

Let $(X_n)_{n \in \mathbb{N}_0}$ be the stationary process where X_0 is distributed according to μ and $X_i = T(X_{i-1})$ for $i \geq 1$. The expectation in this process is denoted by $\mathbb{E}$. We abbreviate $X_i^j := (X_i, X_{i+1}, \ldots, X_j)$ for $i \leq j$. We denote by $\mathcal{F}_i^n$ the sigma-field generated by $X_i, X_{i+1}, \ldots, X_n$ for $i \leq n$ and by convention $\mathcal{F}_{n+1}^n = \{\emptyset, [0,1]\}$, the trivial sigma-field. We then have the following telescoping identity (martingale difference decomposition):

$$\begin{aligned} K(X_0,\ldots,X_{n-1}) - \mathbb{E}\left(K(X_0,\ldots,X_{n-1})\right) &= \sum_{i=1}^{n} \mathbb{E}\left(K\middle|\mathcal{F}_{i-1}^{n-1}\right) - \mathbb{E}\left(K\middle|\mathcal{F}_{i}^{n-1}\right) \\ &=: \sum_{i=1}^{n} \mathcal{V}_i. \end{aligned}$$

The measurable function $\mathbb{E}\left(K\middle|\mathcal{F}_{i-1}^{n-1}\right)$ is a function of $X_{i-1},\ldots,X_{n-1}$ only. When evaluated along an orbit segment $T_0^{n-1}(x)$, it takes the value

$$\begin{aligned} \mathbb{E}\left(K\middle|\mathcal{F}_{i-1}^{n-1}\right)\left(T_0^{n-i}(x)\right) &= \mathbb{E}\left(K(X_0,\ldots,X_{n-1})\middle|X_{i-1}^{n-1} = T_0^{n-i}(x)\right) \\ &= \sum_{y:T^{i-1}(y)=x} \frac{h(y)}{h(x)|(T^{i-1})'(y)|}\, K\left(T_0^{i-2}(y), T_0^{n-i}(x)\right). \end{aligned}$$

To obtain the second equality, notice that the reversed process $(X_{n-i})_{i=0,\ldots,n}$ is a Markov chain with transition probability kernel

$$\mathbb{P}(X_0 = dz\big|X_1 = x) = \sum_{y:T(y)=x} \frac{h(y)}{h(x)|T'(y)|}\,\delta(y-z) \tag{7}$$

and similarly

$$\mathbb{P}(X_0 = dz\big|X_k = x) = \sum_{y:T^k(y)=x} \frac{h(y)}{h(x)|(T^k)'(y)|}\,\delta(y-z).$$

The identity (7) follows at once from Bayes formula and the identity $\mathbb{P}(X_1 = x|X_0 = z) = \delta(x - T(z))$.

Since $\mathcal{F}_i^{n-1} \subset \mathcal{F}_j^{n-1}$ for $i \geq j$, we have the orthogonality property

$$\mathbb{E}(\mathcal{V}_i\mathcal{V}_j) = 0 \quad \text{for } i \neq j,$$

and hence

$$\mathbb{E}\left(K - \mathbb{E}(K)\right)^2 = \sum_{i=1}^{n} \mathbb{E}(\mathcal{V}_i^2).$$

The function $\mathcal{V}_i$ is $\mathcal{F}_{i-1}^{n-1}$-measurable and

$$\mathcal{V}_i(T_0^{n-i}(x)) = \mathbb{E}\left(K\middle|X_{i-1}^{n-1} = T_0^{n-i}(x)\right) - \mathbb{E}\left(K\middle|X_i^{n-1} = T_1^{n-i}(x)\right).$$

Hence, by Cauchy-Schwarz inequality,

$$\begin{aligned} \mathcal{V}_i^2(T_0^{n-i}(x)) &\leq \int \mathbb{P}(X_{i-1} = dx'\big|X_i^{n-1} = T_1^{n-i}(x)) \\ &\quad \times \left[\mathbb{E}\left(K\middle|X_{i-1}^{n-1} = T_0^{n-i}(x)\right) - \mathbb{E}\left(K\middle|X_{i-1}^{n-1} = (x', T_0^{n-i}(x))\right)\right]^2 \\ &= \sum_{x':T(x')=T(x)} \frac{h(x')}{h(T(x))|T'(x')|}\,\left[\mathbb{E}\left(K\middle|X_{i-1} = x\right) - \mathbb{E}\left(K\middle|X_{i-1} = x'\right)\right]^2. \end{aligned}$$

For x, x' such that $T(x) = T(x')$, let

$$M_i(x,x') := \mathbb{E}\left(K \middle| X_i = x\right) - \mathbb{E}\left(K \middle| X_i = x'\right)$$

and μ_x^i denote the conditional distribution of $X_0, \ldots, X_{i-1}$ given that $X_i = x$. By using the Lipschitz property of K one gets

$$|M_i(x,x')| \leq 2\mathrm{Lip}_i(K) + \left| \int K(z_0^{i-1}, 0, T_1^{n-i-1}(x))\,(d\mu_x^i(z_0^{i-1}) - d\mu_{x'}^i(z_0^{i-1})) \right|$$

and one obtains

$$\begin{aligned}
\mathbb{E}(\mathcal{V}_i^2) &= \int \sum_{x':T(x')=T(x)} \frac{h(x')}{h(T(x))} \frac{d\mu(x)}{|T'(x')|}\, M_i^2(x,x') \\
&\leq 8(\mathrm{Lip}_i(K))^2 + 2\int \sum_{x':T(x')=T(x)} \frac{h(x')}{h(T(x))} \frac{d\mu(x)}{|T'(x')|} \\
&\quad \times \left[\int K(z_0^{i-1}, 0, T_1^{n-i-1}(x))\,(d\mu_x^i(z_0^{i-1}) - d\mu_{x'}^i(z_0^{i-1})) \right]^2 .
\end{aligned}$$

Let us further abbreviate

$$\Gamma_i(x,x') := \int K(z_0^{i-1}, 0, T_1^{n-i-1}(x))\,(d\mu_x^i(z_0^{i-1}) - d\mu_{x'}^i(z_0^{i-1})). \tag{8}$$

We then obtain

$$\begin{aligned}
\mathbb{E}\left(K - \mathbb{E}K\right)^2 &\leq 8\sum_{i=1}^{n} (\mathrm{Lip}_i(K))^2 \\
&\quad + 2\sum_{i=1}^{n} \int d\mu(x) \sum_{x':T(x')=T(x)} \frac{h(x')}{h(T(x))|T'(x')|}\, \Gamma_i^2(x,x').
\end{aligned} \tag{9}$$

(Observe that $\Gamma_k(x,x) = 0$.)

5.2 Second telescoping

Our aim is now to further estimate the quantity $\Gamma_i(x,x')$ by using a second telescoping where the decay of correlations (2) can be used.

Let

$$\Psi_k(x) := \mathbb{E}\left(K(X_0, \ldots, X_{k-1}, 0, T_1^{n-k-1}(x)) \middle| X_k = x\right).$$

With this notation (8) reads

$$\Gamma_k(x,x') = \Psi_k(x) - \Psi_k(x'). \tag{10}$$

The idea is now to telescope the Ψ_k's by introducing an independent copy $(Y_i)_{i\in\mathbb{N}_0}$ of the process $(X_i)_{i\in\mathbb{N}_0}$. We write

$$\begin{aligned}\Psi_k(x) \;&=\; \sum_{p=1}^{k}\mathbb{E}\left[K(X_0^{p-1},Y_p^{k-1},0,T_1^{n-k-1}(x))-K(X_0^{p-2},Y_{p-1}^{k-1},0,T_1^{n-k-1}(x))\Big|X_k=x\right]\\ &\quad+\mathbb{E}\left(K(Y_0,\ldots,Y_{k-1},0,T_1^{n-k-1}(x))\right) \qquad (11)\end{aligned}$$

where now $\mathbb{E}$ denotes expectation both with respect to the random variables X and Y, and where we make the convention that, if Y_i^j (resp. X_i^j) occurs with $j<i$, then Y (resp. X) is simply not present.

Combining (10) and (11), we obtain

$$\Gamma_k(x,x')=\sum_{p=1}^{k}\int\omega_{p-1}(z_0^{p-1},T_1^{n-k-1}(x))(d\mu_x^{p-1,k-p+1}(z)-d\mu_{x'}^{p-1,k-p+1}(z))$$

where $\mu_x^{p-1,k-p+1}$ is the conditional distribution of $X_0,\ldots,X_{p-1}$ given $X_k=x$, and where

$$\begin{aligned}&\omega_{p-1}(z_0^{p-1},T_1^{n-k-1}(x)):=\\ &\mathbb{E}\left(K(z_0^{p-1},Y_p^{k-1},0,T_1^{n-k-1}(x))-K(z_0^{p-2},Y_{p-1}^{k-1},0,T_1^{n-k-1}(x))\right), \qquad (12)\end{aligned}$$

where the expectation is taken with respect to Y. Observe that

$$\left|\omega_{p-1}(z_0^{p-1},T_1^{n-k-1}(x))\right|\leq \mathrm{Lip}_p(K).$$

We now define the distance

$$d_p(z_0^p,\hat z_0^p):=\inf\left(2\mathrm{Lip}_{p+1}(K),\sum_{j=0}^{p}\mathrm{Lip}_{j+1}(K)|z_j-\hat z_j|\right).$$

Without loss of generality, we assume $\inf_j \mathrm{Lip}_j(K)>0$. Hence, equipped with the distance d_p, $[0,1]^{p+1}$ is a complete, separable, metric space. From (12) it follows that

$$\sup_{z_0^{p-1}\neq\hat z_0^{p-1},x_1^{n-k-1}}\frac{|\omega_{p-1}(z_0^{p-1},x_1^{n-k-1})-\omega_{p-1}(\hat z_0^{p-1},x_1^{n-k-1})|}{d_{p-1}(z_0^{p-1},\hat z_0^{p-1})}\leq 1,$$

i.e., for each fixed x, the function $z_0^{p-1}\mapsto\omega_{p-1}(z_0^{p-1},T_1^{n-k-1}(x))$ is Lipschitz with respect to the d_{p-1} distance, with Lipschitz norm less than or equal to one.

Denote by $\mathbf{c}_{x,x'}^{p,q}(z_0^p,\hat z_0^p)$ the Kantorovich-Rubinstein coupling, associated with the distance d_p, of the measures $\mu_x^{p,q}$ and $\mu_{x'}^{p,q}$ (cf [12, Theorem 11.8.2, p. 421]).

For this coupling we thus have

$$\int d_p(z_0^p, \hat{z}_0^p)\ d\mathbf{c}_{x,x'}^{p,k-p}(z_0^p, \hat{z}_0^p) = \sup_{f:\mathrm{Lip}_{d_p}(f)\leq 1} \Big(\int f\ d\mu_x^{p,k-p} - \int f\ d\mu_{x'}^{p,k-p} \Big).$$

Hence, by the definition of the distance d_p and the Kantorovich-Rubinstein duality theorem [12], one gets

$$\begin{aligned}
|\Gamma_k(x,x')| &= \left| \sum_{p=0}^{k-1} \int \omega_p(z_0^p, T_1^{n-k-1}(x))\ (d\mu_x^{p,k-p}(z_0^p) - d\mu_{x'}^{p,k-p}(z_0^p)) \right| \\
&= \left| \sum_{p=0}^{k-1} \int \left[\omega_p(z_0^p, T_1^{n-k-1}(x)) - \omega_p(\hat{z}_0^p, T_1^{n-k-1}(x))\right]\ d\mathbf{c}_{x,x'}^{p,k-p}(z_0^p, \hat{z}_0^p) \right| \\
&\leq \sum_{p=0}^{k-1} \int d_p(z_0^p, \hat{z}_0^p)\ d\mathbf{c}_{x,x'}^{p,k-p}(z_0^p, \hat{z}_0^p) \\
&= \sum_{p=0}^{k-1} \left(\sup_{f:\mathrm{Lip}_{d_p}(f)\leq 1} \Big(\int f\ d\mu_x^{p,k-p} - \int f\ d\mu_{x'}^{p,k-p} \Big) \right). \qquad (13)
\end{aligned}$$

In order to estimate Γ_k, we will now exploit the fact that for $k-p$ "large" the measure $\mu_x^{p,k-p}$ is "close" to the invariant measure μ. More precisely, passing from $\mu_x^{p,0}$ to $\mu_x^{p,k-p}$ involves $k-p$ iterations of the normalised Perron-Frobenius operator.

5.3 Distortion and correlation estimates

We now proceed by estimating the final expression in (13).

Define, as usual, the normalised Perron-Frobenius operator

$$\begin{aligned}
\mathscr{L} w(x) &= \mathbb{E}\left(w(X_0)\middle| X_1 = x\right) \\
&= \int w(y)\mathbb{P}(X_0 = dy | X_1 = x) \\
&= \sum_{u:T(u)=x} \frac{h(u)}{h(x)|T'(u)|}\ w(u).
\end{aligned}$$

By the Markov property of the reversed process we have

$$\mathscr{L}^k w(x) = \mathbb{E}\left(w(X_0)\middle| X_k = x\right) = \int w(y)\mathbb{P}(X_0 = dy | X_k = x).$$

For f a function of $(p+1)$ variables, define

$$\begin{aligned} f_p(x) &= \mathbb{E}\left(f(X_0,\ldots,X_p)\big|X_p=x\right) \\ &= \sum_{u:T^p(u)=x} \frac{h(u)}{h(x)|(T^p)'(u)|}\, f(T_0^p(u)). \end{aligned}$$

We then have

$$\begin{aligned} \int f\, d\mu_x^{p,k-p} &= \mathbb{E}\left(f(X_0,\ldots,X_p)\big|X_k=x\right) \\ &= \int \mathbb{E}\left(f(X_0,\ldots,X_p)\big|X_p=y\right)\, \mathbb{P}(X_p=dy\big|X_k=x) \\ &= \mathscr{L}^{k-p} f_p(x). \end{aligned}$$

The next three lemmas will be useful.

Lemma 5.1. *Let f be such that $\mathrm{Lip}_{d_p}(f) \le 1$. Then, for any $y, \tilde{y} \in I_\ell$ and any $m \ge 0$, we have*

$$|(\mathscr{L}^m f_p)(y) - (\mathscr{L}^m f_p)(\tilde{y})| \le C_p\, \frac{|y-\tilde{y}|}{y} \tag{14}$$

where

$$C_p = \mathcal{O}(1) \sum_{j=0}^{p} \frac{\mathrm{Lip}_{j+1}(K)}{(p-j+1)^{1/\alpha}}.$$

Proof. Observe that it is enough to prove the lemma in the case where f vanishes at some point. The general case follows by adding a constant. Without loss of generality, we can assume that $\mathscr{L}^m f_p)(y) \le \mathscr{L}^m f_p)(\tilde{y})$. Indeed, the opposite case would lead to the same estimate because there exists a constant $C > 0$ such that $y/\tilde{y} \le C$, for all $y, \tilde{y} \in I_\ell$ and all ℓ.

Since f vanishes at some point and $\mathrm{Lip}_{d_p}(f) \le 1$, we have $|f_p(T_0^p(\cdot))|/\mathrm{Lip}_{p+1}(K) \le 2$. We also have $|\mathscr{L}^m f_p(T_0^p(\cdot))|/\mathrm{Lip}_{p+1}(K) \le 2$. Now we use the inequality

$$1 + \frac{3(a-b)}{5} \le \frac{1+a}{1+b}$$

for all a, b such that $-2/3 \le b \le a \le 2/3$. Therefore,

$$\left|\frac{(\mathscr{L}^m f_p)(y)}{3\mathrm{Lip}_{p+1}(K)} - \frac{(\mathscr{L}^m f_p)(\tilde{y})}{3\mathrm{Lip}_{p+1}(K)}\right| \le \frac{5}{3}\left(\frac{(\mathscr{L}^m f_p)(y) + 3\mathrm{Lip}_{p+1}(K)}{(\mathscr{L}^m f_p)(\tilde{y}) + 3\mathrm{Lip}_{p+1}(K)} - 1\right)\cdot \tag{15}$$

We have

$$\frac{(\mathscr{L}^m f_p)(y)+3\mathrm{Lip}_{p+1}(K)}{(\mathscr{L}^m f_p)(\tilde y)+3\mathrm{Lip}_{p+1}(K)} \le \frac{h(\tilde y)}{h(y)} \sup_{z,\tilde z} \frac{h(z)}{h(\tilde z)}\, \frac{T^{(p+m)'}(\tilde z)}{T^{(p+m)'}(z)}\, \frac{f(T_0^p(z))+3\mathrm{Lip}_{p+1}(K)}{f(T_0^p(\tilde z))+3\mathrm{Lip}_{p+1}(K)} \tag{16}$$

where the supremum is taken over the pairs $(z,\tilde z)$ of pre-images of y and $\tilde y$ whose iterates lie in the same atoms of the Markov partition until $p+m$. To estimate this, we use the bounds

$$\frac{h(\tilde y)}{h(y)} \le 1 + C\frac{|y-\tilde y|}{y}, \quad \frac{T^{(p+m)'}(\tilde z)}{T^{(p+m)'}(z)} \le 1 + C\frac{|y-\tilde y|}{y}$$

proved in [13]: the first one follows from the fact that h belongs to the space $\mathcal{G}$ [13, p. 502] whereas the second one is [13, Proposition 2.3 (ii)]. We also use the bounds

$$\frac{|z-\tilde z|}{z} \le C\frac{|y-\tilde y|}{y}\,, \tag{17}$$

and

$$|T^j(z)-T^j(\tilde z)| \le \frac{C}{(p-j+1)^{1/\alpha}}\;\frac{|T^p(z)-T^p(\tilde z)|}{T^p(z)} \tag{18}$$

which are proved in the appendix (Lemmas 5.6 and 5.7).

Therefore, using (18), we get

$$\begin{aligned}\frac{f(T_0^p(z))+3\mathrm{Lip}_p(K)}{f(T_0^p(\tilde z))+3\mathrm{Lip}_{p+1}(K)} &\le 1+\mathcal{O}(1)\frac{|f(T_0^p(z))-f(T_0^p(\tilde z))|}{\mathrm{Lip}_{p+1}(K)}\\ &\le 1+\frac{\mathcal{O}(1)}{\mathrm{Lip}_{p+1}(K)}\sum_{j=0}^{p}\frac{\mathrm{Lip}_{j+1}(K)}{(p-j+1)^{1/\alpha}}\frac{|T^p(z)-T^p(\tilde z)|}{T^p(z)}.\end{aligned}$$

Using (17) and all the previous bounds in (16), we obtain

$$\frac{(\mathscr{L}^m f_p)(y)+3\mathrm{Lip}_{p+1}(K)}{(\mathscr{L}^m f_p)(\tilde y)+3\mathrm{Lip}_{p+1}(K)} \le 1+\frac{\mathcal{O}(1)}{\mathrm{Lip}_{p+1}(K)}\frac{|y-\tilde y|}{y}\sum_{j=0}^{p}\frac{\mathrm{Lip}_{j+1}(K)}{(p-j+1)^{1/\alpha}}.$$

This inequality together with (15) completes the proof of the lemma. ☐

Lemma 5.2. *Let f be such that $\mathrm{Lip}_{d_p}(f)\le 1$. Then for any $q\ge 0$ we have*

$$\int \left|\mathscr{L}^q\left(f_p - \int f_p\; d\mu\right)\right|\; d\mu \le D_p\; \gamma_q^{\frac{1-\alpha}{3}}$$

where

$$D_p = \mathcal{O}(1)\; C_p\,.$$

Proof. Let $M > 0$ be an integer and $\epsilon > 0$ to be fixed later on. Recall the notation $I_\ell =]x_{\ell+1}, x_\ell]$. For $\ell \le M$, we define the sequence of functions f_p^ℓ, each vanishing outside I_ℓ, given by

$$f_p^\ell(x) := \begin{cases} \frac{x-x_{\ell+1}}{\epsilon|I_\ell|} f_p(x_{\ell+1} + \epsilon|I_\ell|) \quad \text{for} \quad x \in [x_{\ell+1}, x_{\ell+1} + \epsilon|I_\ell|] \\ \\ f_p(x) \quad \text{for} \quad x \in [x_{\ell+1} + \epsilon|I_\ell|, x_\ell - \epsilon|I_\ell|] \\ \\ \frac{x_\ell - x}{\epsilon|I_\ell|} f_p(x_\ell - \epsilon|I_\ell|) \quad \text{for} \quad x \in [x_\ell - \epsilon|I_\ell|, x_\ell] \,. \end{cases}$$

We have the identity

$$\mathscr{L}^q \left(f_p - \int f_p \, d\mu \right) =$$

$$\sum_{\ell=0}^{M} \mathscr{L}^q \left(f_p^\ell - \int f_p^\ell \, d\mu \right) + \mathscr{L}^q \left(f_p - \sum_{\ell=0}^{M} f_p^\ell \right) - \mathscr{L}^q \left(\int f_p \, d\mu - \sum_{\ell=0}^{M} \int f_p^\ell \, d\mu \right) .$$

The decay of correlations (2) gives us

$$\begin{aligned} \int \left| \mathscr{L}^q \left(f_p^\ell - \int f_p^\ell \, d\mu \right) \right| \, d\mu &= \sup_{u : \|u\|_{C^1} \le 1} \int u \; \mathscr{L}^q \left(f_p^\ell - \int f_p^\ell \, d\mu \right) d\mu \\ &\le \frac{C_p}{\epsilon |I_\ell|} \, \gamma_q, \end{aligned}$$

since $|f_p| \le C_p$ and using Lemma 5.1 with $m = 0$.

On the other hand, we have

$$\int \left| f_p - \sum_{\ell=0}^{M} f_p^\ell \right| \, d\mu \le 2C_p \sum_{\ell=0}^{M} \epsilon \, |I_\ell| h_\ell + C_p \sum_{\ell=M+1}^{\infty} |I_\ell| h_\ell \,.$$

The optimal bound is obtained with

$$\epsilon = \gamma_q^{\frac{1}{2}} M^{1+\frac{1}{2\alpha}} \,, \; M = \gamma_q^{-\frac{\alpha}{3}} \,.$$

The Lemma follows. □

Lemma 5.3. *Let f be such that $\mathrm{Lip}_{d_p}(f) \le 1$. Then for any $q \ge 0$ and $\ell \ge 0$ we have*

$$\left\| \mathscr{L}^q \left(f_p(x) - \int f_p \, d\mu \right) \right\|_{L^\infty(I_\ell)} \le \Delta(\ell, q; f_p)$$

where

$$\Delta(\ell, q; f_p) := \begin{cases} \frac{2}{|I_\ell|} \int_{I_\ell} |g_{q,f_p}| dx & \text{if} \quad C_p \, |I_\ell|^2 \le x_\ell \int_{I_\ell} |g_{q,f_p}| \, dx \\ \\ 2\sqrt{\frac{C_p \int_{I_\ell} |g_{q,f_p}| dx}{x_\ell}} & \text{otherwise} \end{cases}$$

where

$$g_{q,f_p} = \mathscr{L}^q \left(f_p - \int f_p \, d\mu \right).$$

Proof. By (14) we have

$$|g_{q,f_p}(y) - g_{q,f_p}(y')| \le C_p \, \frac{|y - y'|}{x_\ell}$$

for $y, y' \in I_\ell$.

Hence, if we let $J \subseteq I_\ell$ and $y \in J$, then we have

$$|g_{q,f_p}(y)| \le \frac{1}{|J|} \int_J |g_{q,f_p}(y')| \, dy' + \frac{1}{|J|} \int_J |g_{q,f_p}(y) - g_{q,f_p}(y')| \, dy' \le$$

$$\frac{1}{|J|} \int_J |g_{q,f_p}| dx + C_p \, \frac{|J|}{x_\ell} \, .$$

The first case follows by taking $J = I_\ell$. In the second case, we take J such that

$$|J| = \sqrt{\frac{x_\ell}{C_p} \int_{I_\ell} |g_{q,f_p}| dx} \le |I_\ell| \, .$$

The lemma is proved. □

Now return to (9). We have to estimate

$$\int d\mu(x) \sum_{x' : x \ne x', T(x) = T(x')} \frac{h(x')}{h(T(x)) |T'(x')|} \Gamma_k^2(x, x') =: S_1(k) + S_2(k) \tag{19}$$

where

$$S_1(k) := \sum_{m=1}^{\infty} \int_{I_m} d\mu(x) \sum_{x' : x \ne x', T(x) = T(x')} \frac{h(x')}{h(T(x)) |T'(x')|} \Gamma_k^2(x, x')$$

and

$$S_2(k) := \int_{I_0} d\mu(x) \sum_{x' : x \ne x', T(x) = T(x')} \frac{h(x')}{h(T(x)) |T'(x')|} \Gamma_k^2(x, x'),$$

where the intervals I_ℓ form the Markov partition defined in Subsection 2.1.

We have the following lemmas.

Lemma 5.4. *Let*

$$Q_k := \sum_{\ell=1}^{\infty} |I_\ell| \sup_{x\in I_\ell, x'\in I_0 : T(x)=T(x')} \Gamma_k^2(x,x')\,.$$

Then there exists a constant $B>0$ such that for any k

$$S_1(k) \le B\ Q_k$$

and

$$S_2(k) \le B\ Q_k\,.$$

Proof. We first observe that, if $m\ge 1$, $x\in I_m$, $T(x)=T(x')$, and $x\neq x'$, then $x'\in I_0$. Next, using [13, Lemma 4.4 (iv)] and the fact that h is bounded on I_0, we get

$$G := \sup_{m\ge 1}\sup_{x\in I_m}\sup_{x'\in I_0} \frac{h(x')h(x)}{h(T(x))|T'(x')|} < \infty\,.$$

The bound on $S_1(k)$ follows immediately.

For the bound on $S_2(k)$, we have

$$S_2(k) = \sum_{m\ge 1}\int_{I_0} d\mu(x)\ \chi_{I_m}(x') \sum_{x' : x\neq x', T(x)=T(x')} \frac{h(x')}{h(T(x))}\frac{\Gamma_k^2(x,x')}{|T'(x')|}.$$

Note that the term corresponding to $m=0$ is absent because $x\neq x'$. Observe that

$$G' := \sup_{m\ge 1}\sup_{x'\in I_m}\sup_{x\in I_0, T(x)=T(x')} \frac{h(x')h(x)}{h(T(x))|T'(x')|} < \infty$$

and there is a constant $C>0$ such that for any $m\ge 1$

$$|\{x\in I_0 | T(x)\in T(I_m)\}| \le C\ |I_m|\,.$$

The lemma follows. $\square$

Lemma 5.5. *Assume that $\alpha\in[0,4-\sqrt{15}[$. Then there exists a constant $H>0$ such that*

$$\sum_{k=1}^{n} Q_k \le H\ \sum_{j=1}^{n} (\mathrm{Lip}_j(K))^2\,.$$

Proof. Observe that

$$\sup_{x\in I_0, x'\in I_m} |\Gamma_k(x,x')| \le$$

$$\sum_{p=0}^{k-1} \sup_{f:\mathrm{Lip}_{d_p}(f)\le 1} \Delta(0,k-p,f_p) + \sum_{p=0}^{k-1} \sup_{f:\mathrm{Lip}_{d_p}(f)\le 1} \Delta(m,k-p,f_p)$$

$$= \Sigma(0,k) + \Sigma(m,k)$$

where

$$\Sigma(j,k) := \sum_{p=0}^{k-1} \sup_{f:\mathrm{Lip}_{d_p}(f)\le 1} \Delta(j, k-p, f_p)$$

for $j \ge 0$.

By Lemma 5.2

$$\sup_{f:\mathrm{Lip}_{d_p}(f)\le 1} \int |g_{q,f_p}| d\mu \le \mathcal{O}(1)\ C_p\ \gamma_q^{(1-\alpha)/3}\,.$$

Both cases of Lemma 5.3 lead to the bound

$$\sup_{f:\mathrm{Lip}_{d_p}(f)\le 1} \Delta(0, k-p, f_p) \le \mathcal{O}(1)\ C_p\ \gamma_{k-p}^{(1-\alpha)/6}\,.$$

Since $\alpha \in [0, 4-\sqrt{15}[$, we have

$$\sum_q \gamma_q^{(1-\alpha)/6} < \infty$$

and Young's inequality yields

$$\sum_{k=1}^{n} \sum_m |I_m|\ \Sigma(0,k)^2 \le \mathcal{O}(1)\ \sum_p C_p^2 \le \mathcal{O}(1)\ \sum_{j=1}^{n} (\mathrm{Lip}_j(K))^2\,.$$

We now bound

$$\sum_\ell |I_\ell|\ \Sigma(\ell,k)^2\,.$$

By Lemma 5.3 we get

$$\sum_\ell |I_\ell|\ \Sigma(\ell,k)^2 \le A_1(k) + A_2(k)$$

where

$$A_1(k) := 8 \sum_\ell |I_\ell|\ \left(\sum_{p=0}^{k-1} \sup_{f:\mathrm{Lip}_{d_p}(f)\le 1}\ \sqrt{\frac{C_p \int_{I_\ell} |g_{k-p,f_p}(x)| dx}{x_\ell}} \right)^2$$

and

$$A_2(k) := 8 \sum_\ell |I_\ell|\ \left(\sum_{p=0}^{k-1} \frac{1}{|I_\ell|}\ \sup_{f:\mathrm{Lip}_{d_p}(f)\le 1}\ \int_{I_\ell} |g_{k-p,f_p}(x)| dx \right)^2\,.$$

Observe that

$$\int_{I_\ell} |g_{k-p,f_p}(x)|dx \leq \frac{\mathcal{O}(1)}{\ell} \int_{I_\ell} |g_{k-p,f_p}|d\mu \leq \frac{\mathcal{O}(1)}{\ell} \, C_p \, \gamma_{k-p}^{\frac{1-\alpha}{3}}$$

since $h_{|I_\ell} \sim \ell$ and by using Lemma 5.2. Hence,

$$A_1(k) \leq \mathcal{O}(1) \sum_\ell \frac{|I_\ell|}{\ell x_\ell} \left(\sum_{p=0}^{k-1} C_p \, \gamma_{k-p}^{\frac{1-\alpha}{6}} \right)^2$$

which implies, as above,

$$\sum_{k=1}^{n} A_1(k) \leq \mathcal{O}(1) \sum_{j=1}^{n} (\mathrm{Lip}_j(K))^2 \, .$$

We now bound $A_2(k)$. By Cauchy-Schwarz inequality, for any $\delta > 0$, we have

$$A_2(k) \leq \mathcal{O}(1) \sum_\ell |I_\ell| \sum_{p=0}^{k-1} \frac{1}{|I_\ell|^2} \sup_{f:\mathrm{Lip}_{d_p}(f)\leq 1} \left(\int_{I_\ell} |g_{k-p,f_p}(x)|dx \right)^2 (k-p)^{1+\delta} \, .$$

Observe that, if $\mathrm{Lip}_{d_p}(f) \leq 1$, then

$$\|g_{q,f_p}\|_{L^\infty} \leq \mathrm{Lip}_{p+1}(K) \, .$$

Indeed,

$$f_p(x) - \int f_p \; d\mu = \int d\mu_x^{p,0}(z_0^p) \; f(z_0^p) - \int d\mu(y) \int d\mu_y^{p,0}(\xi_0^p) \; f(\xi_0^p) =$$

$$\int d\mu_x^{p,0}(z_0^p) \int d\mu(y) \int d\mu_y^{p,0}(\xi_0^p) \; (f(z_0^p) - f(\xi_0^p)) \, ,$$

and we use the fact that $\mathrm{Lip}_{d_p}(f) \leq 1$ and that $\mathscr{L}$ has L^∞-norm equal to one. This implies that, for any $0 < \sigma < 2$,

$$A_2(k) \leq \mathcal{O}(1)\times$$

$$\sum_\ell \frac{1}{|I_\ell|} \sum_{p=0}^{k-1} \sup_{f:\mathrm{Lip}_{d_p}(f)\leq 1} \left(\int_{I_\ell} |g_{k-p,f_p}(x)|dx \right)^\sigma (k-p)^{1+\delta} (\mathrm{Lip}_{p+1}(K))^{2-\sigma} |I_\ell|^{2-\sigma} \, .$$

Using again $h_{|I_\ell} \sim \ell$, Lemma 5.2 and the fact that $\mathrm{Lip}_p(K) \leq \mathcal{O}(1) C_p$, we get

$$A_2(k) \leq \mathcal{O}(1) \sum_\ell \frac{|I_\ell|^{1-\sigma}}{\ell^\sigma} \sum_{p=0}^{k-1} C_p^2 \, \gamma_{k-p}^{\frac{1-\alpha}{3}\sigma} \, (k-p)^{1+\delta} \, .$$

Since $\alpha \in [0, 4-\sqrt{15}[$, there exist $0<\sigma<1$ and $\delta>0$ such that

$$\sum_q \gamma_q^{\frac{1-\alpha}{3}\sigma}\; q^{1+\delta} < \infty$$

then,

$$\sum_{k=1}^{n} A_2(k) \leq \mathcal{O}(1)\; \sum_{j=1}^{n} (\mathrm{Lip}_j(K))^2 \,.$$

This ends the proof of Lemma 5.5. □

5.4 End of the proof

We now conclude the proof of Theorem 3.1. By (9) and (19), we have

$$\begin{aligned}
\mathbb{E}\left(K-\mathbb{E}K\right)^2 &\leq 8\sum_{i=1}^{n}(\mathrm{Lip}_i(K))^2 \\
&\quad +2\sum_{i=1}^{n}\int d\mu(x) \sum_{x':T(x')=T(x)} \frac{h(x')}{h(T(x))|T'(x')|}\,\Gamma_k^2(x,x') \\
&= 8\sum_k(\mathrm{Lip}_k(K))^2 + 2\sum_k (S_1(k)+S_2(k))\,.
\end{aligned}$$

The theorem now follows from Lemmas 5.4 and 5.5.

Appendix

In this appendix we prove the inequalities (17) and (18) used in the proof of Lemma 5.1. We recall that the map T is defined in Section 2.

Lemma 5.6. *There exists a constant $C>0$ such that for any integer $m\geq 1$ and any pair of points z, $\tilde{z}$ such that for $0\leq j\leq m$, $T^j(z)$ and $T^j(\tilde{z})$ belong to the same atom of the Markov partition. Then one has*

$$\frac{|z-\tilde{z}|}{z} \leq C\frac{|T^m(z)-T^m(\tilde{z})|}{T^m z}.$$

Proof. We start by proving the following inequality:

$$|T^{m'}(z)| \geq C_0\left(\frac{T^m(z)}{z}\right)^{1+\alpha} \tag{20}$$

where $C_0 > 0$ is independent of m and z.

There are two cases.

If $z \geq 1/2$, the inequality is true provided that $C_0 \leq 2^{-(1+\alpha)}$.

Now consider the case $z < 1/2$. We define an integer $q \leq m$ as follows. If $T^j(z) \leq 1/2$ for $j = 0, 1, \ldots, m-1$ then we take $q = m$. Otherwise q is the smallest integer sucht that $T^q(z) \geq 1/2$. Since $z < 1/2$, there is an integer $\ell \geq 1$ such that $z \in I_\ell$. Moreover T^q is a diffeomorphism from I_ℓ to $I_{\ell-q}$. From the distortion lemma, see *e.g.* [13, Proposition 2.3], we get

$$|T^{q'}(z)| \geq C_1 \left(\frac{I_{\ell-q}}{I_\ell}\right)^{1+\alpha}$$

where $C_1 > 0$ is independent of q and z. From (1) it follows that

$$|T^{q'}(z)| \geq C_2 \left(\frac{T^q(z)}{z}\right)^{1+\alpha}$$

where $C_2 > 0$ is independent of q and z. If $q = m$ then (20) is proved with $C_0 = \min(C_2, 2^{-(1+\alpha)})$. If $q < m$ then we observe that

$$|T^{m'}(z)| = |T^{(m-q)'}(T^q(z))||T^{q'}(z)| \geq C_2 \left(\frac{T^q(z)}{z}\right)^{1+\alpha}$$

because $|T^{(m-q)'}| \geq 1$. Since $T^q(z) \geq 1/2$, we obtain

$$|T^{m'}(z)| \geq \frac{C_2}{2^{1+\alpha}} \frac{1}{z^{1+\alpha}} \geq \frac{C_2}{2^{1+\alpha}} \left(\frac{y}{z}\right)^{1+\alpha}.$$

This finishes the proof of inequality (20).

To prove the lemma, we first observe that if $T^m(z) \leq z$ then

$$|z - \tilde{z}| \leq |T^m(z) - T^m(\tilde{z})| \leq \frac{z}{T^m(z)} |T^m(z) - T^m(\tilde{z})|$$

because the modulus of the T' is larger than or equal to one. The remaining case is when $T^m(z) > z$. We observe that

$$|T^m(z) - T^m(\tilde{z})| = \left| \int_z^{\tilde{z}} T^{m'}(\xi) d\xi \right| = \int_z^{\tilde{z}} |T^{m'}(\xi)| d\xi \geq \tilde{C} \left(\frac{T^m(z)}{z}\right)^{1+\alpha}$$

where we used again the distortion estimates ([13, Proposition 2.3]), (1), the monotonicity of T^m, and where $\tilde{C} > 0$ is independent $m, z, \tilde{z}$. This immediately implies

$$\frac{|z - \tilde{z}|}{z} \leq \frac{1}{\tilde{C}} \left(\frac{z}{T^m(z)}\right)^{\alpha} \frac{|T^m(z) - T^m(\tilde{z})|}{T^m z}.$$

The Lemma is proved. □

Lemma 5.7. *There exists a constant $C > 0$ such that for any integer $m \geq 1$ and any pair of points z, $\tilde{z}$ such that for $0 \leq j \leq m$, $T^j(z)$ and $T^j(\tilde{z})$ belong to the same atom of the Markov partition. Then one has*

$$|z - \tilde{z}| \leq \frac{C}{(m+1)^{1/\alpha}} \frac{|T^m(z) - T^m(\tilde{z})|}{T^m(z)}.$$

Proof. Observe that if $T^m(z) < \frac{1}{(m+1)^{1/\alpha}}$ then the estimate follows at once from the fact that the modulus of the derivative of T is larger than or equal to one. So we now assume that $T^m(z) \geq \frac{1}{(m+1)^{1/\alpha}}$. Let $\ell \geq 0$ be the integer such that $T^m z \in I_\ell$. There exists a unique z_* in $I_{\ell+m}$ such that $T^m(z_*) = T^m(z)$. Since z_* is the closest T^{-m}-preimage of $T^m(z)$ to the neutral fixed point 0, one can easily show that there is a constant $c > 0$ such that, for any m and z, one has $|T^{m'}(z)| \geq c|T^{m'}(z_*)|$.

As in the proof of the previous lemma, we use the distortion estimates ([13, Proposition 2.3]) and (1) to obtain

$$|T^{m'}(z)| \geq c' \frac{(\ell+m)^{1+\frac{1}{\alpha}}}{\ell^{1+\frac{1}{\alpha}}}$$

From the distortion estimates ([13, Proposition 2.3]) we get, using that $T^m(z) \in I_\ell$,

$$|T^m(z) - T^m(\tilde{z})| \geq \mathcal{O}(1)|T^{m'}(z)|\, |z - \tilde{z}| \geq \mathcal{O}(1)|T^m(z)|\, \frac{(\ell+m)^{1+\frac{1}{\alpha}}}{\ell}\, |z - \tilde{z}|.$$

This can be rewritten as

$$|z-\tilde{z}| \leq \mathcal{O}(1) \frac{\ell}{(\ell+m)^{1+\frac{1}{\alpha}}} \frac{|T^m(z) - T^m(\tilde{z})|}{|T^{m'} z|} \leq \mathcal{O}(1) \frac{1}{(1+m)^{\frac{1}{\alpha}}} \frac{|T^m(z) - T^m(\tilde{z})|}{|T^{m'}(z)|}.$$

The proof of the lemma is complete. ∎

References

[1] A. D. Barbour, R. M. Gerrard, G. Reinert. Iterates of expanding maps. Prob. Th. Rel. & Fields **116** (2000), no. 2, 151–180.

[2] P.J. Brockwell, R.A. Davis. *Time Series : Theory and Methods*. 2nd ed., Springer, Berlin, 1991.

[3] J.-R. Chazottes, P. Collet. Almost-sure central limit theorems and the Erdös-Rényi law for expanding maps of the interval. Ergod. Th. & Dynam. Syst. **25** (2005) 419–441.

[4] J.-R. Chazottes, P. Collet, C. Kuelske, F. Redig. Concentration inequalities for random fields via coupling. Prob. Th. & Rel. Fields **137** (2007) 201–225.

[5] J.-R. Chazottes, P. Collet, B. Schmitt. Devroye inequality for a class of non-uniformly hyperbolic dynamical systems. Nonlinearity **18** (2005) 2323–2340.

[6] J.-R. Chazottes, P. Collet, B. Schmitt. Statistical consequences of the Devroye inequality for processes. Applications to a class of non-uniformly hyperbolic dynamical systems. Nonlinearity **18** (2005) 2341–2364.

[7] J.-R. Chazottes, S. Gouëzel. On almost-sure versions of classical limit theorems for dynamical systems. Probab. Th. & Rel. Fields **138** (2007) 195–234.

[8] P. Collet. Variance and exponential estimates via coupling. Bull. Braz. Math. Soc. **37** (2006) 461–475.

[9] P. Collet, S. Martínez, B. Schmitt. Exponential inequalities for dynamical measures of expanding maps of the interval. Prob. Th. & Rel. Fields **123** (2002) 301–22.

[10] J. Dedecker, C. Prieur. Some unbounded functions of intermittent maps for which the central limit theorem holds. Preprint, 2007 (`http://arxiv.org/abs/0712.2726`).

[11] L. Devroye, G. Lugosi. Combinatorial Methods in Density Estimation. Springer, New York, 2000

[12] R. M. Dudley. Real analysis and probability. Revised reprint of the 1989 original. Cambridge Studies in Advanced Mathematics **74**. Cambridge University Press, Cambridge, 2002.

[13] H. Hu. Decay of correlations for piecewise smooth maps with indifferent fixed points. Ergodic Theory Dynam. Systems **24** (2004), no. 2, 495–524.

[14] M. Ledoux. *The concentration of measure phenomenon.* Mathematical Surveys and Monographs **89**, American Mathematical Society (2001).

[15] C. Liverani. Central limit theorem for deterministic systems. International Conference on Dynamical Systems (Montevideo, 1995), 56–75, Pitman Res. Notes Math. Ser., 362, Longman, Harlow, 1996.

[16] P. Massart. *Concentration inequalities and model selection*. École d'été de Probab. de Saint-Flour XXXIII - 2003. Lecture Notes in Mathematics **1896**, Springer, 2007

[17] C. McDiarmid. Concentration. Probabilistic methods for algorithmic discrete mathematics, 195–248, Algorithms Combin., **16** (1998), Springer.

[18] I. Melbourne, M. Nicol. Large deviations for nonuniformly hyperbolic systems. To appear in Trans. Amer. Math. Soc. (2007).

[19] L. Rey-Bellet, L.-S. Young. Large deviations in nonuniformly hyperbolic dynamical systems. To appear in Ergodic Theory and Dynamical Systems.

[20] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Etudes Sci. Publ. Math. **81** (1995) 73–205. New concentration inequalities in product spaces. Inventiones Math. **126** (1996) 505–563. A new look at independence. Ann. Probab. **24** (1996) 1–34.

[21] L.-S. Young. Statistical properties of dynamical systems with some hyperbolicity. Ann. of Math. (2) **147** (1998), no. 3, 585–650.

[22] L.-S. Young. Recurrence times and rates of mixing. Isr. J. Math. **110**, 153-188 (1999).