A concentration inequality for interval maps with an indifferent fixed point

J.-R. CHAZOTTES†, P. COLLET†, F. REDIG‡ and E. VERBITSKIY§

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

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 separately Lipschitz. The proof is based on coupling and decay of correlation properties of the map. We also present applications of this inequality to the almost-sure central limit theorem, the kernel density estimation, the empirical measure and the periodogram.

1. *Introduction*

Nowadays, concentration inequalities are a fundamental tool in probability theory and statistics; we refer the reader to [**11, 14, 16, 17, 20**], for instance. In particular, concentration inequalities have turned out to be essential tools in developing a non-asymptotic theory in statistics, exactly as the central limit theorem and large deviations have played a central role in the asymptotic theory. Besides the non-asymptotic aspect of concentration inequalities, a crucial point is that they allow us, in principle, to study random variables $Z_n = K(X_1, \ldots, X_n)$ that, while depending 'smoothly' on the underlying random variables X_i, can be defined only in an indirect or 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 deviation estimates, and other types of limit theorems have been proved almost exclusively for Birkhoff sums of sufficiently 'smooth' observables. However, many natural observables are not Birkhoff sums. For now, let us just mention a typical example

(see below for further examples), namely the so-called power spectrum, i.e. 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 overcoming the computational difficulties pertaining to each observable, one would like to develop a systematic method for dealing with the question of fluctuations of observables, instead of designing a particular method for each case.

A possible approach is to use concentration inequalities. We note that an additional difficulty arises for dynamical systems, namely the fact that independence is lost except in very special cases, and the mixing properties of dynamical systems are not as nice as those of the stochastic processes usually encountered in probability theory, such as Markov chains and renewal processes. Therefore, new methods are needed, preferably based on typical dynamical systems tools such as the spectral gap (when it exists) of the transfer operator and the decay of correlations. The first concentration inequality in the dynamical systems context was obtained by Collet *et al* [9] for uniformly expanding maps of the interval, without assuming the existence of a Markov partition. Collet *et al* derived the so-called Gaussian concentration inequality (also called the exponential concentration inequality) by bounding the exponential moment of any observable of *n* variables, assuming only that the observable is separately Lipschitz; they also deduced several applications (kernel density estimation, shadowing, etc). In the hope of proving concentration inequalities for more general dynamical systems, one might start with an inequality for the variance, which then leads to a polynomial concentration inequality. This route was indeed taken 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 the 'Devroye inequality') through various examples. Let us also mention another approach based on coupling (see, e.g., [4, 8]), which gives an alternative proof of the Gaussian concentration inequality in the case of uniformly expanding maps of the interval, and has also been used in the context of Gibbs random fields.

As regards Birkhoff sums of 'smooth' (e.g. Hölder) observables, central limit theorems and large deviation estimates have been proved both for systems that can be modeled by a Young tower with exponential return-time tail, as mentioned above, and for systems with a summable return-time tail; see, e.g., [18, 19, 21, 22]. A natural question, therefore, is to try to prove an inequality for the variance of any observable of *n* variables, assuming only that the observable is separately Lipschitz, as in [5], but relaxing the exponential decay of the return-time tail of Young towers [22]. This would provide a way of analyzing fluctuations of complicated observables which are not Birkhoff sums. The classical and simplest 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 defined by $T(x) = x + 2^\alpha x^{1+\alpha}$ when $x \in [0, 1/2[$ and which is strictly expanding on $[1/2, 1]$, where α is small enough (Theorem 3.1). However, the proof applies verbatim to the class of maps with a unique indifferent fixed point considered in [13]. The major difference from the situation in [5, 9] is that the transfer operator has no spectral gap and the decay of correlations is polynomial rather than exponential. Thus, we develop a different approach based on decay of correlations. We need to control the covariance of C^0 functions and Lipschitz functions, which was

done by Hu in [**13**]. An important ingredient is coupling through the Kantorovich–Rubinstein duality theorem. At present, we are not able to construct explicitly a coupling for the backward process like 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. In addition to proving the variance inequality, we shall show various applications of it: to the almost-sure central limit theorem, the kernel density estimation, the empirical measure, the integrated periodogram, and shadowing.

The paper is organized as follows. Section 2 summarizes the information we need about the maps, while §3 contains our main result, namely the variance inequality for separately Lipschitz observables. In §4 we give various applications of the main theorem and, finally, in §5 the proof of the variance inequality is presented.

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 $|T'(x)| > 1$, $|T''(x)| < \infty$ on $[1/2, 1]$. In fact, all of what follows will be valid under the assumptions of Hu [**13**].

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

From, for example, [**13**], we have the behavior

$$|I_\ell| \sim \ell^{-1/(\alpha-1)}, \quad x_\ell \sim \ell^{-1/\alpha}. \tag{1}$$

2.2. *Decay of correlations.*

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

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

When $v = w$, we simply write Cov_v for the coefficient.

Various researchers have established the (optimal) decay of correlations for the map T, that is, $\text{Cov}_{v,w}(\ell) \approx \ell^{-(1/\alpha)+1}$. In [**22**], this was proved for v, w both being Hölder. It turns out that we shall need the following estimate proved in [**13**]: there exists a number $C > 0$ such that, for all $v \in C^0$ and w Lipschitz, we have the decay

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

where

$$\gamma_\ell := \ell^{-1/(\alpha+1)} \tag{3}$$

and

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

In [13], C is a function of w; but, in fact, C can be taken to be constant as a consequence of an abstract result [6, Theorem B.1]. Very often, this uniform estimate makes a natural appearance in proofs; see, for example, [22].

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

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

where

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

3. *Variance inequality*

Our main theorem gives an upper bound for the variance of any separately Lipschitz function.

We introduce the convenient notation

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

for $0 \leq p \leq q$. If we take a function K of n variables, in this notation we can write $K(z_1^{j-1}, z_j, z_{j+1}^n)$ for $K(z_1, z_2, \ldots, z_n)$.

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

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

Our main theorem is stated as follows.

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

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

(This inequality is referred to as the '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 3.1,*

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

for all $t > 0$.

Let us make several comments on the theorem.

We expect (5) to be true for $\alpha \in [0, 1/2[$. However, the present proof does not seem to allow us to go beyond $4 - \sqrt{15}$. Recall that for $\alpha \in [0, 1/2[$, we have the central limit theorem (and stronger versions of it).

Instead of using (2), we could have worked with the decay of correlations for L^∞ and Hölder functions proved in [22, Theorem 5]. However, the present proof would lead to a smaller interval of α's in the theorem.

In our context, we cannot expect a Gaussian concentration bound, i.e. a bound for the exponential moment of K. This would give a Gaussian concentration inequality which is incompatible with the large deviation lower bounds obtained in [18] where, for a large class of Hölder observables v, it was proved that, given $\epsilon > 0$ small enough,

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

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 inequality has also been obtained in [10] under different conditions.

Despite the previous comment, one expects to be able to control moments higher than the second one; in principle, this could be done in the spirit of [4, Theorem 3].

4. *Some applications*

We now give some applications of the variance inequality (5). We follow [6] where we obtained these results in an abstract setting: therein we assumed that (X_k) was some real-valued, stationary, ergodic process satisfying (5) together with, eventually, an extra condition on the auto-covariance of Lipschitz observables, which depended 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 for applying all the results from [6] that we shall use.

The standing assumption in this section is that $0 < \alpha < 4 - \sqrt{15}$, so that Theorem 3.1 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 $D_n = \sum_{k=1}^n (1/k)$.

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.

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

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

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

We denote by $\mathcal{N}(0, \sigma_v^2)$ 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*

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

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

Observe that this theorem immediately implies that for μ-almost every x, $\mathcal{A}_n(x)$ converges weakly to the Gaussian measure. The weak convergence was proved in [7] via 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 by means of 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 which converges to 0 and is such that na_n converges to $+\infty$, while 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 none other than the distance in total variation between $\mathcal{H}_n(x)$ and μ, namely

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

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

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

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

4.3. *Empirical measure.* The empirical measure associated with $x, Tx, \ldots, T^{n-1}x$ 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 converges weakly 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 $n \geq 1$,*

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

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

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

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

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 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*

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

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

4.5. *Shadowing and mismatch.* Let A be a set of initial conditions with positive measure. If $x \notin A$, we might ask how well the orbit of x can be approximated by an orbit starting from an initial condition in A.

We can measure the average quality of 'shadowing' by the 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 $n \geq 1$ and all $t > 0$, one has*

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

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

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

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

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

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

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 shall use the abbreviation $X_i^j := (X_i, X_{i+1}, \ldots, X_j)$ for $i \leq j$, and denote by $\mathcal{F}_i^n$ the sigma-field generated by $X_i, X_{i+1}, \ldots, X_n$ for $i \leq n$; by convention, $\mathcal{F}_{n+1}^n = \{\emptyset, [0, 1]\}$, the trivial sigma-field. We then have the following telescoping identity (martingale difference decomposition):

$$K(X_0, \ldots, X_{n-1}) - \mathbb{E}(K(X_0, \ldots, X_{n-1})) = \sum_{i=1}^{n} \mathbb{E}(K|\mathcal{F}_{i-1}^{n-1}) - \mathbb{E}(K|\mathcal{F}_i^{n-1})$$

$$=: \sum_{i=1}^{n} \mathcal{V}_i.$$

The measurable function $\mathbb{E}(K|\mathcal{F}_{i-1}^{n-1})$ 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

$$\mathbb{E}(K|\mathcal{F}_{i-1}^{n-1})(T_0^{n-i}(x)) = \mathbb{E}(K(X_0, \ldots, X_{n-1})|X_{i-1}^{n-1} = T_0^{n-i}(x))$$

$$= \sum_{y:T^{i-1}(y)=x} \frac{h(y)}{h(x)|(T^{i-1})'(y)|} K(T_0^{i-2}(y), T_0^{n-i}(x)).$$

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

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

and similarly,

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

The identity (7) follows at once from Bayes formula together with 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

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

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

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

Hence, by the Cauchy–Schwarz inequality,

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

For x and x' such that $T(x) = T(x')$, define

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

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

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

and thus obtain

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

Furthermore, let us use the abbreviation

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

We then obtain

$$\mathbb{E}(K - \mathbb{E}K)^2 \le 8 \sum_{i=1}^{n} (\mathrm{Lip}_i(K))^2$$

$$+ 2 \sum_{i=1}^{n} \int d\mu(x) \sum_{x':T(x')=T(x)} \frac{h(x')}{h(T(x))|T'(x')|} \Gamma_i^2(x, x'). \tag{9}$$

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

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

Let

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

With this notation, (8) reads

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

The idea is 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

$$\Psi_k(x) = \sum_{p=1}^{k} \mathbb{E}[K(X_0^{p-1}, Y_p^{k-1}, 0, T_1^{n-k-1}(x))$$
$$- K(X_0^{p-2}, Y_{p-1}^{k-1}, 0, T_1^{n-k-1}(x)) | X_k = x]$$
$$+ \mathbb{E}(K(Y_0, \dots, Y_{k-1}, 0, T_1^{n-k-1}(x))), \tag{11}$$

where now $\mathbb{E}$ denotes expectation with respect to both random variables X and Y, and we make the convention that if Y_i^j (respectively X_i^j) occurs with $j < i$, then Y (respectively 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 $X_0, \dots, X_{p-1}$ given $X_k = x$, and

$$\omega_{p-1}(z_0^{p-1}, T_1^{n-k-1}(x)) := \mathbb{E}[K(z_0^{p-1}, Y_p^{k-1}, 0, T_1^{n-k-1}(x))$$
$$- K(z_0^{p-2}, Y_{p-1}^{k-1}, 0, T_1^{n-k-1}(x))], \tag{12}$$

with the expectation being taken with respect to Y. Observe that

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

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 that $\inf_j \mathrm{Lip}_j(K) > 0$. 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 have

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

Hence, from the definition of the distance d_p and the Kantorovich–Rubinstein duality theorem [**12**], one gets

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

In order to estimate Γ_k, we will 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 normalized Perron–Frobenius operator.

5.3. *Distortion and correlation estimates.* We now proceed by estimating the final expression in (13).

Define, as usual, the normalized Perron–Frobenius operator

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

By the Markov property of the reversed process, we have

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

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

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

We then have

$$\int f \, d\mu_x^{p,k-p} = \mathbb{E}(f(X_0, \ldots, X_p)|X_k = x)$$
$$= \int \mathbb{E}(f(X_0, \ldots, X_p)|X_p = y) \, \mathbb{P}(X_p = dy|X_k = x)$$
$$= \mathcal{L}^{k-p} f_p(x).$$

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 $m \ge 0$, we have*

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

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 upon adding a constant. Without loss of generality, we can assume that $(\mathcal{L}^m f_p)(y) \le (\mathcal{L}^m f_p)(\tilde{y})$. Indeed, the opposite case would actually 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 that $|f_p(T_0^p(\cdot))|/\mathrm{Lip}_{p+1}(K) \le 2$. We also have $|\mathcal{L}^m f_p(T_0^p(\cdot))|/\mathrm{Lip}_{p+1}(K) \le 2$. Now we use the inequality

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

which holds for all a, b such that $-2/3 \le b \le a \le 2/3$. This gives

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

and we have

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

where the supremum is taken over 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

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

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

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

and

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

which are proved in Appendix A (Lemmas A.1 and A.2).

Therefore, using (18), we get

$$\frac{f(T_0^p(z)) + 3\operatorname{Lip}_p(K)}{f(T_0^p(\tilde{z})) + 3\operatorname{Lip}_{p+1}(K)} \leq 1 + \mathcal{O}(1)\frac{|f(T_0^p(z)) - f(T_0^p(\tilde{z}))|}{\operatorname{Lip}_{p+1}(K)}$$

$$\leq 1 + \frac{\mathcal{O}(1)}{\operatorname{Lip}_{p+1}(K)}\sum_{j=0}^{p}\frac{\operatorname{Lip}_{j+1}(K)}{(p - j + 1)^{1/\alpha}}\frac{|T^p(z) - T^p(\tilde{z})|}{T^p(z)}.$$

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

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

This inequality, together with (15), completes the proof of the lemma. $\qquad\square$

LEMMA 5.2. *Let f be such that $\operatorname{Lip}_{d_p}(f) \leq 1$. Then, for any $q \geq 0$, we have*

$$\int\left|\mathcal{L}^q\left(f_p - \int f_p \, d\mu\right)\right| d\mu \leq D_p\gamma_q^{(1-\alpha)/3},$$

where

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

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

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

We have the identity

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

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

The decay of correlations (2) gives us

$$\int \left|\mathcal{L}^q\left(f_p^\ell - \int f_p^\ell\, d\mu\right)\right| d\mu = \sup_{u:\|u\|_{C^1}\leq 1} \int u\,\mathcal{L}^q\left(f_p^\ell - \int f_p^\ell\, d\mu\right) d\mu$$

$$\leq \frac{C_p}{\epsilon|I_\ell|}\gamma_q$$

since $|f_p| \leq C_p$, by using Lemma 5.1 with $m = 0$.

On the other hand, we have

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

The optimal bound is obtained with

$$\epsilon = \gamma_q^{1/2} M^{1+(1/2\alpha)}, \quad M = \gamma_q^{-\alpha/3},$$

and the result follows.   □

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

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

where

$$\Delta(\ell, q; f_p) := \begin{cases} \dfrac{2}{|I_\ell|}\displaystyle\int_{I_\ell} |g_{q,f_p}|\, dx & \text{if } C_p|I_\ell|^2 \leq x_\ell \displaystyle\int_{I_\ell} |g_{q,f_p}|\, dx, \\[3ex] 2\sqrt{\dfrac{C_p \int_{I_\ell} |g_{q,f_p}|\, dx}{x_\ell}} & \text{otherwise,} \end{cases}$$

with

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

Proof. By (14), we have

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

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

Hence, if we let $J \subseteq I_\ell$ and $y \in J$, then it follows that

$$
\begin{aligned}
|g_{q,f_p}(y)| &\leq \frac{1}{|J|} \int_J |g_{q,f_p}(y')| \, dy' + \frac{1}{|J|} \int_J |g_{q,f_p}(y) - g_{q,f_p}(y')| \, dy' \\
&\leq \frac{1}{|J|} \int_J |g_{q,f_p}| \, dx + C_p \frac{|J|}{x_\ell}.
\end{aligned}
$$

The first case is obtained upon taking $J = I_\ell$. In the second case, choose J such that

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

Thus the lemma is proved. $\qquad \square$

Now let us return to (9). We need to estimate

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

where

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

and

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

with the intervals I_ℓ forming the Markov partition defined in §2.1.

We have the following two 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,

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

and

$$
S_2(k) \leq B Q_k.
$$

Proof. First, observe that if $m \geq 1$, $x \in I_m$, $T(x) = T(x')$ 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

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

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

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

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

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

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

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

$$|\{x \in I_0 \mid T(x) \in T(I_m)\}| \leq C|I_m|.$$

The result follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

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

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

Proof. Observe that

$$\sup_{x \in I_0, x' \in I_m} |\Gamma_k(x, x')| \leq \sum_{p=0}^{k-1} \sup_{f : \mathrm{Lip}_{d_p}(f) \leq 1} \Delta(0, k - p, f_p)$$

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

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

where

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

for $j \geq 0$.

By lemma 5.2,

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

Both cases of Lemma 5.3 lead to the bound

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

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

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

and Young's inequality yields

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

We now bound

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

From Lemma 5.3 we get

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

where

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

and

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

Observe that

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

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

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

which implies, as above,

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

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

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

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

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

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

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

$$\times (k-p)^{1+\delta} (\text{Lip}_{p+1}(K))^{2-\sigma} |I_\ell|^{2-\sigma}.$$

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

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

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

$$\sum_q \gamma_q^{((1-\alpha)/3)\sigma} q^{1+\delta} < \infty,$$

and then

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

This ends the proof of Lemma 5.5. □

5.4. *Completing the proof of Theorem 3.1.* We now conclude the proof of our main theorem. By (9) and (19), we have

$$\mathbb{E}(K - \mathbb{E}K)^2 \leq 8 \sum_{i=1}^n (\text{Lip}_i(K))^2$$

$$+ 2 \sum_{i=1}^n \int d\mu(x) \sum_{x': T(x') = T(x)} \frac{h(x')}{h(T(x))|T'(x')|} \Gamma_k^2(x, x')$$

$$= 8 \sum_k (\text{Lip}_k(K))^2 + 2 \sum_k (S_1(k) + S_2(k)).$$

The theorem then follows from Lemmas 5.4 and 5.5.

A. *Appendix*

In this appendix we prove the inequalities (17) and (18) used in the proof of Lemma 5.1. Recall the map T defined in §2.

LEMMA A.1. *There exists a constant $C > 0$ such that for any integer $m \geq 1$ and any pair of points $z, \tilde{z}$ with $T^j(z)$ and $T^j(\tilde{z})$ belonging to the same atom of the Markov partition whenever $0 \leq j \leq m$, 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 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 holds 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 take $q = m$; otherwise take q to be the smallest integer such 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, 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 derivative of T is greater than or equal to one. It remains to consider $T^m(z) > z$. In this case, 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 have again used the distortion estimates [13, Proposition 2.3], the behaviour (1), and the monotonicity of T^m; here $\tilde{C} > 0$ is independent of m, z and $\tilde{z}$. This immediately implies that

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

hence the lemma is proved. $\qquad\square$

LEMMA A.2. *There exists a constant $C > 0$ such that for any integer $m \geq 1$ and any pair of points z, $\tilde{z}$ with $T^j(z)$ and $T^j(\tilde{z})$ belonging to the same atom of the Markov partition whenever $0 \leq j \leq m$, 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 $T^m(z) < 1/(m+1)^{1/\alpha}$, then the estimate follows at once from the fact that the modulus of the derivative of T is greater than or equal to one. So we now assume that $T^m(z) \geq 1/(m+1)^{1/\alpha}$. Let $\ell \geq 0$ be the integer for which $T^m z \in I_\ell$. There exists a unique z_* in $I_{\ell+m}$ such that $T^m(z_*) = T^m(z)$. Since z_* is the T^{-m}-preimage of $T^m(z)$ that is closest 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 make use of the distortion estimates [13, Proposition 2.3] and (1) to obtain

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

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

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

This can be rewritten as

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

The proof of the lemma is therefore complete. □

REFERENCES

[1]   A. D. Barbour, R. M. Gerrard and G. Reinert. Iterates of expanding maps. *Probab. Theory Related Fields* **116**(2) (2000), 151–180.

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

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

[4]   J.-R. Chazottes, P. Collet, C. Kuelske and F. Redig. Concentration inequalities for random fields via coupling. *Probab. Theory Related Fields* **137** (2007), 201–225.

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

[6]   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** (2005), 2341–2364.

[7]   J.-R. Chazottes and S. Gouëzel. On almost-sure versions of classical limit theorems for dynamical systems. *Probab. Theory Related 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 and B. Schmitt. Exponential inequalities for dynamical measures of expanding maps of the interval. *Probab. Theory Related Fields* **123** (2002), 301–322.

[10]   J. Dedecker and C. Prieur. Some unbounded functions of intermittent maps for which the central limit theorem holds. *Preprint*, 2007. http://arxiv.org/abs/0712.2726. *Latin Amer. J. Probab. Math. Stat.* to appear.

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

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

[13]   H. Hu. Decay of correlations for piecewise smooth maps with indifferent fixed points. *Ergod. Th. & Dynam. Sys.* **24**(2) (2004), 495–524.

[14] M. Ledoux. *The Concentration of Measure Phenomenon (Mathematical Surveys and Monographs, 89)*. American Mathematical Society, Providence, RI, 2001.

[15] C. Liverani. Central limit theorem for deterministic systems. *International Conference on Dynamical Systems (Montevideo, 1995) (Pitman Research Notes in Mathematics Series, 362)*. Longman, Harlow, 1996, pp. 56–75.

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

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

[18] I. Melbourne and M. Nicol. Large deviations for nonuniformly hyperbolic systems. *Trans. Amer. Math. Soc.* **360**(12) (2008), 6661–6676.

[19] L. Rey-Bellet and L.-S. Young. Large deviations in nonuniformly hyperbolic dynamical systems. *Ergod. Th. & Dynam. Sys.* **28**(2) (2008), 587–612.

[20] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Inst. Hautes Études Sci.* **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**(3) (1998), 585–650.

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