*Ergod. Th. & Dynam. Sys.* (2016), **36**, 2384–2407 © *Cambridge University Press, 2015*
doi:10.1017/etds.2015.19

# Concentration inequalities for sequential dynamical systems of the unit interval

ROMAIN AIMINO† and JÉRÔME ROUSSEAU‡

† *Dipartimento di Matematica, II Università di Roma (Tor Vergata), Via della Ricerca Scientifica, 00133 Roma, Italy (e-mail: aimino@mat.uniroma2.it)*

‡ *Departamento de Matemática, Universidade Federal da Bahia, Av. Ademar de Barros s/n, 40170-110 Salvador, BA, Brazil (e-mail: jerome.rousseau@ufba.br)*

(*Received* 24 *June* 2014 *and accepted in revised form* 10 *February* 2015)

*Abstract*. We prove a concentration inequality for sequential dynamical systems of the unit interval enjoying an exponential loss of memory in the BV norm and we investigate several of its consequences. In particular, this covers compositions of $\beta$-transformations, with all $\beta$ lying in a neighborhood of a fixed $\beta_\star > 1$, and systems satisfying a covering-type assumption.

*Contents*

1 Introduction 2384
2 Sequential dynamical systems of the unit interval 2386
3 A concentration inequality 2391
4 Applications 2397
5 Large deviations and empirical measure: an alternative approach 2401
Acknowledgements 2404
A Appendix. Proof of a formula on conditional expectations 2404
References 2405

## 1. *Introduction*

Sequential dynamical systems consist of a composition of different maps from a space into itself, and model discrete-time non-autonomous dynamics, where the law of evolution is changing, possibly slowly, along the time. Such systems have been studied either from a topological point of view, with for instance the introduction of the notion of topological entropy [**29**] and Devaney's chaos [**17**], or from an ergodic theoretic point of view:

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

Berend and Bergelson [**8**] have developed the notions of ergodicity and mixing, and Kawan [**28**] defined a measure theoretic entropy.

It is important to notice that, in general, there is no common invariant measure for all maps, even though there is often a natural probability measure on the space, such as the Lebesgue measure, allowing us to look at random processes that arise from observations. But this lack of invariance of the measure and the fact that maps change with the time imply that the processes are non-stationary, adding difficulties to their study.

Up to now, the study of the statistical properties remains incomplete and the emphasis has been mainly put on the property of loss of memory, which is the generalization of the classical notion of decay of correlations to sequential systems: if we start from two probability densities, both belonging to a class of smooth enough functions, and if we let them evolve with the dynamics, they will be attracted one from each other if the system is sufficiently mixing. This leads to the idea that all smooth densities are attracted indifferently by the same moving target in the space of densities, whence the name of loss of memory. Many works have been devoted to the study of the speed of loss of memory. Among them, we can cite [**38**] for smooth expanding maps, [**38**, **20**] for one-dimensional piecewise expanding maps, [**23**] for multidimensional piecewise expanding maps, [**40**] for two-dimensional Anosov systems, [**41**] for billiards with moving convex scatterers, [**36**] for one-dimensional open systems, and [**1**] for maps of the interval with a neutral fixed point, the latter reference dealing with systems enjoying a polynomial loss of memory, whereas all the other examples exhibit an exponential speed of decorrelation.

Beyond the loss of memory, only the central limit theorem has been studied in [**5**, **6**] for some classes of hyperbolic maps, and in [**20**, **37**] for one-dimensional piecewise expanding maps. It is worth noting that Conze and Raugi [**20**] also investigated the strong law of large numbers and dynamical Borel–Cantelli lemmas. Very recently, the almost sure invariance principle has been studied in [**24**].

Despite this relative sparseness of the literature on statistical properties for non-autonomous systems, Stenlund [**40**] has written: 'Much of the statistical theory of stationary dynamical systems can be carried over to sufficiently chaotic non-stationary systems.'

Following this philosophy, we address in this paper the question of proving concentration inequalities for sequential dynamical systems. The main object of concentration inequalities, a well-known subject in probability theory, is to estimate the deviation from the mean of random variables of the form $Y_n = K(X_0, \ldots, X_{n-1})$ which depends in a smooth way of the variables $X_n$, even though the dependence could be complicated or implicit. Such observables arise naturally in statistical applications. Concentration inequalities are then an extension of large deviation inequalities for ergodic sums to more general observables. It should be noted that they are also non-asymptotic, in contrast to standard large deviation principles; however, they do not give precise asymptotics, with sharp bounds. We refer the reader to [**10**, **31**, **34**, **35**] among others for reviews of this field in a pure probabilistic setting.

Concentration inequalities were brought into a dynamical context in 2002 by Collet *et al* [**19**], who proved such an inequality for one-dimensional piecewise expanding maps of the interval. More precisely, they showed that if $T$ is a piecewise $C^2$ map of the

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

unit interval which is uniformly expanding and topologically mixing, then the process $(T^n)_n$, defined on the probability space $([0, 1], \mu)$, where $\mu$ is the unique absolutely continuous $T$-invariant measure, satisfies a concentration inequality, provided the density of $\mu$ is bounded away from 0. Subsequently, a lot of works have followed, aiming to extend this result to non-uniformly expanding/hyperbolic systems [**13**, **12**, **16**] or to random dynamical systems [**33**, **2**]. A rather complete description of the situation for systems modeled by a Young tower with an exponential or polynomial tail of the return time is given in [**16**, **22**].

In this paper, we focus on the class of sequential dynamical systems described by Conze and Raugi [**20**]: they consist of compositions of piecewise expanding maps $(T_n)$ of the unit interval that enjoy an exponential loss of memory in the BV norm. Under this decorrelation assumption, together with a minoration hypothesis on the evolution of the Lebesgue measure $m$ under the dynamics, which was shown to hold in [**20**] for compositions of $\beta$-transformations with all $\beta$ very close, we prove that the process $(T_n \circ \cdots \circ T_1)_n$ defined on the probability space $([0, 1], m)$ satisfies an exponential concentration inequality and we discuss some of its consequences.

*Outline of the paper.* In §2, we describe the class of sequential systems introduced by Conze and Raugi [**20**] and discuss in more detail our assumptions. In §3, we state and prove our main result: an exponential concentration inequality for sequential dynamical systems. In §4, we give several applications of concentration to the empirical measure, the shadowing and the almost sure central limit theorem. In §5, we prove a large deviations estimate for observables belonging to BV and an upper bound for the Kantorovich distance using a direct approach. The Appendix is devoted to the proof of a formula used to establish the main result.

## 2. *Sequential dynamical systems of the unit interval*

In this section, we recall some background on sequential dynamical systems. We follow closely the lines of Conze and Raugi [**20**], and we refer to their paper for more details and results.

### 2.1. *Generalities on sequential dynamical systems.*

Let $(X, \mathcal{F}, m)$ be a probability space. A sequential dynamical system on $(X, m)$ is a sequence $(T_n)_{n\geq 1}$ of non-singular transformations† on $(X, \mathcal{F}, m)$.

For $n \leq k$, we denote by $T_n^k$ the composition $T_n^k = T_k \circ \cdots \circ T_n$. For a non-singular map $T$, we denote by $P_T$ the transfer operator; it acts on $L^1(m)$ and satisfies for all $f \in L^1(m)$ and $g \in L^\infty(m)$,

$$\int P_T f \cdot g \, dm = \int f \cdot g \circ T \, dm.$$

We clearly have $P_{T_n^k} = P_{T_k} \circ \cdots \circ P_{T_n}$. For notational convenience, we will drop the letter $T$ when the sequence of transformations is understood, and denote by $P_n^k$ the operator $P_{T_n^k}$.

† Recall that a non-singular transformation on a probability space $(X, \mathcal{F}, m)$ is a measurable map $T : X \to X$ such that $T_\star m \ll m$.

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

For a given sequential dynamical system $(T_n)$, we form the decreasing sequence of $\sigma$-algebras

$$\mathcal{F}_n = (T_1^n)^{-1}(\mathcal{F})$$

and define the asymptotic $\sigma$-algebra by

$$\mathcal{F}_\infty = \bigcap_{n\geq 1} \mathcal{F}_n.$$

Let $f \in L^1(m)$. Since $P_1^n f(x) = 0$ almost surely on the set $\{P_1^n \mathbb{1} = 0\}$, we define the quotient $(P_1^n f / P_1^n \mathbb{1})$ as 0 on $\{P_1^n \mathbb{1} = 0\}$. Then we have the relation

$$\mathbb{E}_m(f|\mathcal{F}_n) = \left(\frac{P_1^n f}{P_1^n \mathbb{1}}\right) \circ T_1^n.$$

By Doob's convergence theorem for martingales, the sequence of conditional expectations $\mathbb{E}_m(f|\mathcal{A}_n)$ converges almost surely and in $L^1(m)$ to $\mathbb{E}_m(f|\mathcal{F}_\infty)$. We say that the sequential dynamical system $(T_n)$ is exact when its associated asymptotic $\sigma$-algebra $\mathcal{F}_\infty$ is trivial modulo $m$. Equivalently, the system is exact if $\lim_n \|P_1^n f\|_{L_m^1} = 0$ for all $f \in L^1(m)$ with $\int f\, dm = 0$.

2.2. *A functional analytic framework.* In [**20**], Conze and Raugi have extended the spectral theory of the iterates of a single operator to the case of concatenations of different operators, having in mind applications to sequential dynamical systems. We recall briefly the setting and the main results.

Let $(\mathcal{B}, \|\cdot\|)$ be a Banach space and $\mathcal{V}$ be a subspace of $\mathcal{B}$ equipped with a norm $|\cdot|_v$ such that $\|\cdot\| \leq |\cdot|_v$. Let $\mathcal{P}$ be a set of contractions of $(\mathcal{B}, \|\cdot\|)$† leaving $\mathcal{V}$ invariant and satisfying:

(H1) the unit ball of $(\mathcal{V}, |\cdot|_v)$ is relatively compact in $(\mathcal{B}, \|\cdot\|)$;

(H2) there is a countable family in $\mathcal{V}$ which is dense in $(\mathcal{B}, \|\cdot\|)$;

(H3) there are an integer $r \geq 1$ and constants $0 < \rho_r < 1$, $M_0$, $C_r > 0$ such that

$$\text{for all } P \in \mathcal{P}, \quad |Pf|_v \leq M_0 |f|_v \quad \text{for all } f \in \mathcal{V}$$

and, for all $r$-tuples $P_1, \ldots, P_r$ of operators in $\mathcal{P}$,

$$\text{for all } f \in \mathcal{V}, \quad |P_r \cdots P_1 f|_v \leq \rho_r |f|_v + C_r \|f\|.$$

They considered mainly the case of one-dimensional systems, where $(X, \mathcal{F}, m)$ is the unit interval $[0, 1]$ endowed with its Borel $\sigma$-algebra and the Lebesgue measure, and where the maps $T$ are piecewise $C^2$ and uniformly expanding. In this setting, the class $\mathcal{P}$ will consist of the transfer operators of the class of maps considered, and the natural choice for the Banach spaces is to take $\mathcal{B} = L^1(m)$ and $\mathcal{V} = \mathrm{BV}$ the space of functions of bounded variations. The previous assumptions imply that there exists $M > 0$ such that $|P_n \cdots P_1 f|_v \leq M |f|_v$ for all $n \geq 1$, all $f \in \mathcal{V}$ and all choices of operators $P_1, \ldots, P_n$ in $\mathcal{P}$.

† That is, a set of linear operators $P : \mathcal{B} \to \mathcal{B}$ satisfying $\|Pf\| \leq \|f\|$ for all $f \in \mathcal{B}$.

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

Let $\mathcal{V}_0$ be a subspace of $\mathcal{V}$ which is invariant by all operators in $\mathcal{P}$. It will consist in concrete applications of the functions in $\mathcal{V}$ with zero average with respect to $m$. We say that a sequence of operators $(P_n)$ in $\mathcal{P}$ is exact if, for all $f \in \mathcal{V}_0$,

$$\lim_n \|P_n \cdots P_1 f\| = 0.$$

A single operator $P \in \mathcal{P}$ is exact in $\mathcal{V}_0$ if the sequence given by $P_n = P$ is exact in $\mathcal{V}_0$. When $P_n$ is the transfer operator of a transformation $T_n$ and $\mathcal{V}_0$ is the set of functions in $\mathcal{V}$ with zero average, then exactness of the sequence $(P_n)$ in $\mathcal{V}_0$ implies exactness of the sequential dynamical system $(T_n)$ as defined previously, as easily deduced from the facts that $\mathcal{V}$ is dense in $(\mathcal{B}, \|\cdot\|)$ and that all operators in $\mathcal{P}$ are contractions on $\mathcal{B}$.

2.3. *Exponential loss of memory.* It is well known since the work of Ionescu-Tulcea and Marinescu [**26**] that each single operator in $\mathcal{P}$ is quasi-compact on $\mathcal{V}$ and enjoys a nice spectral decomposition; see [**20**, Proposition 2.8] for a self-contained proof and [**7**, **11**, **25**] among others for further properties. As a direct consequence, it follows that if an operator $P \in \mathcal{P}$ is exact in $\mathcal{V}_0$, then the norm of $P^n$ seen as an operator on $\mathcal{V}_0$ decays exponentially fast: there exist $K > 0$ and $\theta < 1$ such that $|P^n f|_v \le K\theta^n |f|_v$ for all $n \ge 1$ and all $f \in \mathcal{V}_0$.

For compositions of different operators, the situation is slightly more difficult, and we need more assumptions in order to get exponential decay. First we state the decorrelation property.

*(Dec)*: a subset $\mathcal{P}_0 \subset \mathcal{P}$ satisfies the decorrelation property in $\mathcal{V}_0$ if there exist $\theta < 1$ and $K > 0$ such that, for all integers $l \ge 1$ and all $l$-tuples of operators $P_1, \ldots, P_l$ in $\mathcal{P}_0$,

$$\text{for all } f \in \mathcal{V}_0, \quad |P_l \cdots P_1 f|_v \le K\theta^l |f|_v.$$

Conze and Raugi [**20**] have given a condition ensuring that *(Dec)* is verified. Rather than stating their condition, we give two corollaries from their paper.

The first one, which is of a local nature, states that any exact operator $P$ admits a convenient neighborhood for which *(Dec)* holds. More precisely, for two operators $P, P' \in \mathcal{P}$, define

$$d(P, P') = \sup_{\{f \in \mathcal{V} : |f|_v \le 1\}} \|Pf - P'f\|$$

and, for $\delta > 0$, denote $B(P, \delta) = \{P' \in \mathcal{P} : d(P, P') < \delta\}$.

Then [**20**, Proposition 2.10] asserts that for all $P \in \mathcal{P}$ exact in $\mathcal{V}_0$, there exists a $\delta_0 > 0$ such that $\mathcal{P}_0 = \mathcal{P} \cap B(P, \delta_0)$ satisfies *(Dec)* in $\mathcal{V}_0$.

The other corollary, which does not need any sort of closeness, requires a compactness condition: a subset $\mathcal{P}_0 \subset \mathcal{P}$ satisfies the compactness condition *(C)* if for any sequence $(P_n)$ in $\mathcal{P}_0$, there exist a subsequence $(P_{n_j})$ and an operator $P \in \mathcal{P}_0$ such that

$$\text{for all } f \in \mathcal{B}, \quad \lim_j \|P_{n_j} f - Pf\| = 0.$$

In [**20**, Proposition 2.11], it is said that if $\mathcal{P}_0$ satisfies the compactness condition *(C)* and is such that all sequences in $\mathcal{P}_0$ are exact in $\mathcal{V}_0$, then it satisfies the decorrelation property *(Dec)* in $\mathcal{V}_0$.

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

2.4. *Application to one-dimensional systems.* We now describe how this theory applies to concrete situations, with piecewise expanding maps on $X = [0, 1]$. More precisely, we consider maps $T : X \to X$ uniformly expanding, i.e. $\lambda(T) := \inf_x |T'(x)| > 1$, and such that there exists a finite partition $\mathcal{A}_T$ of $X$ consisting of intervals with disjoint interiors, such that the map $T$ can be extended to a $C^2$ map on a neighborhood of each element of the partition. It is well known that the transfer operator $P_T : L^1(m) \to L^1(m)$ of $T$ satisfies

$$P_T f(x) = \sum_{Ty=x} \frac{f(y)}{|T'(y)|}$$

for all $f \in L^1(m)$ and is quasi-compact on the space BV of functions of bounded variation on $X$; see Baladi [**7**] or Boyarsky and Góra [**11**]. Recall that for $f \in L^1(m)$, we define

$$\mathrm{V} f = \inf\{\mathrm{var}\ \bar{f} : f = \bar{f} m\text{-almost everwehere}\},$$

where

$$\mathrm{var}\ \bar{f} = \sup\left\{\sum_{j=0}^{t-1} |\bar{f}(x_{j+1}) - \bar{f}(x_j)| : 0 = x_0 < \cdots < x_t = 1\right\}.$$

The space BV is equipped with the norm $\|\cdot\|_{\mathrm{BV}} = \mathrm{V}(\cdot) + \|\cdot\|_{L^1_m}$ and is a Banach space whose unit ball is compact in $L^1(m)$.

The couple $(\mathcal{B}, \|\cdot\|) = (L^1(m), \|\cdot\|_{L^1_m})$ and $(\mathcal{V}, |\cdot|_v) = (\mathrm{BV}, \|\cdot\|_{\mathrm{BV}})$ satisfies (H1) and (H2). To ensure (H3), we first need to recall the so-called Lasota–Yorke inequality [**30**]: for any piecewise expanding map $T$ and any $f \in \mathrm{BV}$, we have

$$\mathrm{V}(P_T f) \le \frac{2}{\lambda(T)} \mathrm{V} f + C(T) \|f\|_{L^1_m}$$

with $C(T) = \sup_x (|T''|/|T'|^2) + 2 \sup_{I \in \mathcal{A}_T} ((\sup_{x \in I} |T'(x)|^{-1})/m(I))$.

Then any class $\mathcal{C}$ of piecewise expanding maps $T$ for which $\sup_{T \in \mathcal{C}} C(T) < \infty$ and for which there exists a $r \ge 1$ such that

$$\inf_{T_1, \ldots, T_r \in \mathcal{C}} \lambda(T_1^r) > 2 \quad \text{and} \quad \sup_{T_1, \ldots, T_r \in \mathcal{C}} C(T_1^r) < \infty$$

will satisfy (H3) with $\mathcal{P}$ the class of transfer operators of all maps in $\mathcal{C}$. From now on, we will say that:

($D_r$) a class $\mathcal{C}$ satisfies this condition if the corresponding set of transfer operators satisfies (H3) with this particular $r$.

The class of maps described above hence satisfies ($D_r$).

It is shown in [**20**] that the class of $\beta$-transformations, i.e. maps of the form $x \mapsto \beta x \bmod 1$, with $\beta \ge 1 + \alpha$, $\alpha > 0$, satisfies ($D_r$) for some $r$ depending only on $\alpha$; see Theorem 3.4(c) therein. They also proved that this class of transformations is exact (Theorem 3.6) and that $d(P_1, P_2) \le C|\beta_1 - \beta_2|$ for some universal constant $C$, where $\beta_i > 1$ is arbitrary and $P_i$ is the transfer operator of the $\beta_i$-transformation, for $i = 1, 2$ (Lemma 3.9).

As a direct consequence, the class of $\beta$-transformations, with $\beta$ in between $\beta_{\min} > 1$ and $\beta_{\max} < \infty$, satisfies the compactness condition *(C)* and hence verifies the decorrelation property *(Dec)*.

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

2.5. *Limit theorems.* We state the two main limit theorems from [**20**]. Let $\mathcal{C}$ be a class of transformations that satisfies $(D_r)$ and *(Dec)*. Then, for any sequence of maps $(T_n)_{n\geq 1}$ in $\mathcal{C}$, we have the following strong law of large numbers (Theorem 3.7): for any $f \in \mathrm{BV}$,

$$\lim_{n\to\infty} \frac{S_n}{n} = 0, \quad m\text{-almost surely},$$

where $S_n = \sum_{k=0}^{n-1}[f \circ T_1^k - \int f \circ T_1^k \, dm]$.

In order to have the corresponding central limit theorem, one more assumption is needed:

(*Min*): the sequence of maps $(T_n)$ satisfies this condition if there exists a $\delta > 0$ such that $P_1^n \mathbb{1} \geq \delta$ for all $n$.

Note that this condition concerns only a particular sequence, and not the whole class $\mathcal{C}$.

Moreover, let $(T_n)$ be a sequence of maps which verifies *(Min)* and belongs to $\mathcal{C}$. Then (Theorem 5.1), for any $f \in \mathrm{BV}$, either $\|S_n\|_2$ is bounded and in this case $S_n$ is bounded almost surely, or $\|S_n\|_2$ is unbounded and in this case $\|S_n\|_2$ goes to $+\infty$ and we have

$$\frac{S_n}{\|S_n\|_2} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

2.6. *The minoration condition.* In this section, we discuss the condition *(Min)*. In the case of $\beta$-transformations, Conze and Raugi have shown that for all $\beta > 1$, there exists a neighborhood $[\beta - a, \beta + a]$ of $\beta$ such that if all maps $T_n$ are $\beta_n$-transformations, with $\beta_n \in [\beta - a, \beta + a]$, then *(Min)* holds true.

We now give a condition that automatically ensures the validity of *(Min)*. If $(T_n)$ is a sequential dynamical system on $[0, 1]$ made of piecewise monotonic maps, we denote by $\mathcal{A}_n$ the partition of monotonicity of $T_n$, and by $\mathcal{A}_n^m$ the partition of $T_n^m$, $n \leq m$.

*Definition 1.* The sequential dynamical system $(T_n)$ is covering if

$$\text{for all } n, \quad \text{there exists } N(n) : \text{for all } m, \text{ for all } I \in \mathcal{A}_{m+1}^{m+n}, \quad T_{m+1}^{m+N(n)}(I) = [0, 1].$$

Let $\mathcal{C}$ be a class of surjective piecewise expanding maps of $[0, 1]$ that satisfies $(D_r)$ for some $r \geq 1$ and for which there exist $\lambda > 1$ and $C > 0$ such that $\inf |T'| \geq \lambda$ and $\sup |T'| \leq C$ for all $T \in \mathcal{C}$.

PROPOSITION 2. *Let $(T_n)$ be a sequential dynamical system belonging to a class $\mathcal{C}$ as before, and suppose that $(T_n)$ is covering. Then the condition (Min) is satisfied.*

*Proof.* For $a > 0$, define

$$\mathcal{E}_a = \left\{ f \in \mathrm{BV} : f \geq 0,\ f \neq 0,\ \|f\|_{\mathrm{BV}} \leq a \int f \right\}.$$

Recall from Liverani [**32**, Lemma 3.2] that, given a partition $\mathcal{A}$ of $[0, 1]$ made of intervals of length less than $1/2a$, for any $f \in \mathcal{E}_a$, there exists $I \in \mathcal{A}$ such that $f(x) \geq \frac{1}{2} \int f$ for all $x \in I$. Since $\mathcal{C}$ satisfies $(D_r)$, for any $P_1, \ldots, P_r$ transfer operators associated to maps in $\mathcal{C}$ and any $f \in \mathcal{E}_a$, we have

$$\|P_r \ldots P_1 f\|_{\mathrm{BV}} \leq \rho_r \|f\|_{\mathrm{BV}} + C_r \|f\|_{L_m^1} \leq (a\rho_r + C_r) \int f.$$

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

Hence, if we choose $a \geq (C_r/(1-\rho_r))$, we have $(P_r \cdots P_1)(\mathcal{E}_a) \subset \mathcal{E}_a$ for any choice of $P_1, \ldots, P_r$. In order to also have $\mathbb{1} \in \mathcal{E}_a$, we will actually choose $a = \max\{1, (C_r/(1-\rho_r))\}$.

Let now $(T_n)$ be a sequence of maps in $\mathcal{C}$ which is covering. For any $n \leq m$, the diameter of $\mathcal{A}_n^m$ is less than $\lambda^{-(m-n+1)}$, so we can choose $n_0 \geq 1$ such that the diameter of $\mathcal{A}_{n+1}^{n+n_0}$ is less than $1/2a$ for all $n$.

Let $m \geq 0$. We have $P_1^{m+N(n_0)}\mathbb{1} = P_{m+1}^{m+N(n_0)} P_1^m \mathbb{1}$. Write $m = p_m r + q_m$ with $0 \leq q_m < r$. We have

$$P_1^{q_m}\mathbb{1}(x) = \sum_{T_1^{q_m} y = x} \frac{1}{|(T_1^{q_m})'(y)|} \geq C^{-q_m} \geq C^{-r}, \tag{1}$$

since all maps are surjective and have a derivative uniformly bounded by $C$. As a consequence, we have $P_1^{m+N(n_0)}\mathbb{1} \geq C^{-r} P_{m+1}^{m+N(n_0)} g_m$ with $g_m = P_{q_m+1}^m \mathbb{1}$. Since $P_{q_m+1}^m \mathbb{1}$ is a concatenation of $p_m$ blocks of $r$ operators applied to a function in $\mathcal{E}_a$, we obtain that $g_m$ belongs to $\mathcal{E}_a$. There thus exists an interval $I_m \in \mathcal{A}_{m+1}^{m+n_0}$ on which $g_m \geq \frac{1}{2}$. This implies that

$$\begin{aligned} P_1^{m+N(n_0)}\mathbb{1}(x) &\geq \frac{C^{-r}}{2} P_{m+1}^{m+N(n_0)} \mathbb{1}_{I_m}(x) \\ &= \frac{C^{-r}}{2} \sum_{T_{m+1}^{m+N(n_0)} y = x} \frac{\mathbb{1}_{I_m}(y)}{|(T_{m+1}^{m+N(n_0)})'(y)|} \geq \frac{C^{-r}}{2} C^{-N(n_0)}, \end{aligned}$$

since, for all $x \in [0, 1]$, there exists a $y \in I_m$ such that $T_{m+1}^{m+N(n_0)} y = x$, by the covering assumption. Using the same argument as in (1), $P_1^k \mathbb{1} \geq C^{-N(n_0)}$ for all $k < N(n_0)$ and this shows that *(Min)* is satisfied with $\delta = \min\{C^{-N(n_0)}, (C^{-(r+N(n_0))}/2)\}$. □

## 3. *A concentration inequality*

Let $(X, d)$ be a metric space. A function $K : X^n \to \mathbb{R}$ is said to be separately Lipschitz if, for all $i = 0, \ldots, n-1$, there exists a constant $\mathrm{Lip}_i(K)$ such that

$$\begin{aligned} &|K(x_0, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{n-1}) - K(x_0, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_{n-1})| \\ &\quad \leq \mathrm{Lip}_i(K) d(x_i, x_i') \end{aligned}$$

for all points $x_0, \ldots, x_{n-1}, x_i'$ in $X$.

Let $(Z_n)_{n \geq 0}$ be a discrete-time random process with values in $X$. We say that this process satisfies an exponential concentration inequality if there exists $C > 0$ such that for all $n \geq 1$ and all functions $K : X^n \to \mathbb{R}$ separately Lipschitz, one has

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

This implies a large deviations estimate: for all $t > 0$,

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

which follows by optimizing over $\lambda > 0$ the inequality $\mathbb{P}(Y > t) \leq e^{-\lambda t} \mathbb{E}(e^{\lambda Y})$ with $Y = K(Z_0, \ldots, Z_{n-1}) - \mathbb{E}(K(Z_0, \ldots, Z_{n-1}))$ (see e.g. [**16**]).

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

We will consider processes generated by sequential dynamical systems on the unit interval, that is, processes of the form $Z_n = T_1^n$, where $(T_n)_{n\geq 1}$ is a sequential dynamical system on $[0, 1]$. These processes are defined on the probability space $([0, 1], \mathcal{F}, m)$, where $m$ is the Lebesgue measure defined on the Borel $\sigma$-algebra, which is endowed with the usual distance $d(x, y) = |x - y|$.

The maps $T_n$ will belong to a class $\mathcal{C}$ that satisfies the following properties.

*Assumptions on the class $\mathcal{C}$.*

The class $\mathcal{C}$ is a class of piecewise expanding maps on $[0, 1]$ such that:

(1) $\mathcal{C}$ satisfies $(D_r)$ for some $r \geq 1$ and *(Dec)*;
(2) there exist $\lambda > 1$ and $M \geq 1$ such that $\inf |T'| \geq \lambda$ and $\sup |T''| \leq M$ for all $T \in \mathcal{C}$;
(3) there exists $N \geq 1$ such that $\sharp \mathcal{A}_T \leq N$ for all $T \in \mathcal{C}$.

The main result is the following theorem.

THEOREM 3. *Let $(T_n)_{n\geq 1}$ be a sequence in $\mathcal{C}$ verifying the condition (Min). Then the process $Z_n = T_1^n$ satisfies an exponential concentration inequality.*

In order to prove the theorem, we will use McDiarmid's bounded difference method [**34**, **35**], already used in [**16**, **19**], which we will adapt to the non-stationary case.

We first extend the function $K$ as a function $K : X^{\mathbb{N}} \to \mathbb{R}$ that depends only on the $n$ first coordinates. Here $X^{\mathbb{N}}$ is endowed with the product $\sigma$-algebra $\tilde{\mathcal{F}}$ and the probability measure $\tilde{m}$, which is the image of $m$ by the map $\Phi : X \to X^{\mathbb{N}}$ defined by $\Phi(x) = (T_1^n x)_{n\geq 0}$, with the convention $T_1^0 = \mathrm{Id}$. With this notation, $\mathbb{E}_m(e^{K(Z_0,\dots,Z_{n-1})-\mathbb{E}_m(K(Z_0,\dots,Z_{n-1}))}) = \mathbb{E}_{\tilde{m}}(e^{K-\mathbb{E}_{\tilde{m}}(K)})$. Let $\tilde{\mathcal{F}}_p$ be the $\sigma$-algebra on $X^{\mathbb{N}}$ of events depending only on the coordinates $(x_k)_{k\geq p}$, and define $K_p = \mathbb{E}_{\tilde{m}}(K|\tilde{\mathcal{F}}_p)$. Since the sequential dynamical system $(T_n)$ is exact, as a consequence of the assumption *(Dec)*, the $\sigma$-algebra $\mathcal{F}_\infty$ is trivial mod $m$. Consequently, the $\sigma$-algebra $\tilde{\mathcal{F}}_\infty := \bigcap_{p\geq 0} \tilde{\mathcal{F}}_p$ is trivial mod $\tilde{m}$, since it is easy to check that $\Phi^{-1}(\tilde{\mathcal{F}}_\infty) \subset \mathcal{F}_\infty$. Hence, by Doob's convergence theorem for martingales, $K_p$ goes $\tilde{m}$-almost surely to $\mathbb{E}_{\tilde{m}}(K)$, and in all $L^q$, with $1 \leq q < \infty$. We then have $K - \mathbb{E}_{\tilde{m}}(K) = \sum_{p\geq 0} D_p$, where $D_p = K_p - K_{p+1}$. Using the Azuma–Hoeffding inequality, see [**4**], we deduce the existence of a universal constant $C > 0$ such that, for all $P \geq 0$,

$$\mathbb{E}_{\tilde{m}}(e^{\sum_{p=0}^{P} D_p}) \leq e^{C \sum_{p=0}^{P} \sup |D_p|^2}.$$

It remains to bound $D_p$.

PROPOSITION 4. *There exist $\rho < 1$ and $C > 0$ depending only on $(T_n)_{n\geq 1}$ such that, for all $p$ and all $K$, one has*

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

This proposition, together with the Schwarz inequality, implies the desired concentration inequality, in the same manner as in [**16**]: indeed, we have

$$\left(\sum_{j=0}^{p} \rho^{p-j} \operatorname{Lip}_j(K)\right)^2 \leq \left(\sum_{j=0}^{p} \rho^{p-j} \operatorname{Lip}_j^2(K)\right)\left(\sum_{j=0}^{p} \rho^{p-j}\right) \leq C \sum_{j=0}^{p} \rho^{p-j} \operatorname{Lip}_j^2(K).$$

After summation over $p$, we get $\sum_{p=0}^{P} \sup |D_p|^2 \leq C \sum_j \operatorname{Lip}_j^2(K)$.

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

Proposition 4 will follow immediately from the Lipschitz condition on $K$ and the following lemma.

LEMMA 5. *There exist $\rho < 1$ and $C > 0$ depending only on $(T_n)_{n\geq 1}$ such that, for all $K$, all $p$ and $\tilde{m}$-almost every $\underline{x} = (x_0, x_1, \ldots) \in X^{\mathbb{N}}$, one has*

$$\left| K_p(\underline{x}) - \int_X K(y, T_1 y, \ldots, T_1^{p-1} y, x_p, \ldots)\, dm(y) \right| \leq C \sum_{j=0}^{p-1} \mathrm{Lip}_j(K) \rho^{p-j}.$$

The rest of this section is devoted to the proof of this lemma.

Firstly, we remark that

$$K_p(\underline{x}) = \frac{1}{P_1^p \mathbb{1}(x_p)} \sum_{T_1^p y = x_p} \frac{K(y, T_1 y, \ldots, T_1^{p-1} y, x_p, \ldots)}{|(T_1^p)'(y)|}.$$

The proof of this identity can be found in the Appendix.

We fix a $x_\star \in X$ and we decompose $K_p$ as

$$K_p(\underline{x}) = K(x_\star, \ldots, x_\star, x_p, \ldots) + \frac{1}{P_1^p \mathbb{1}(x_p)} \sum_{i=0}^{p-1} \sum_{T_1^p y = x_p} \frac{H_i(y)}{|(T_1^p)'(y)|},$$

where $H_i(y) = K(y, \ldots, T_1^i y, x_\star, \ldots, x_\star, x_p, \ldots) - K(y, \ldots, T_1^{i-1} y, x_\star, \ldots, x_\star, x_p, \ldots)$.

Using the chain rule, we obtain $K_p(\underline{x}) = K(x_\star, \ldots, x_\star, x_p, \ldots) + (1/(P_1^p \mathbb{1}(x_p))) \sum_{i=0}^{p-1} P_{i+1}^p f_i(x_p)$, with

$$f_i(y) = \sum_{T_1^i z = y} \frac{H_i(z)}{|(T_1^i)'(z)|} = P_1^i H_i(y).$$

We remark that $\int f_i\, dm = \int H_i\, dm$, whence

$$\sum_{i=0}^{p-1} \int f_i\, dm = \int K(y, \ldots, T_1^{p-1} y, x_p, \ldots)\, dm(y) - K(x_\star, \ldots, x_\star, x_p, \ldots).$$

Thus,

$$\begin{aligned} &K_p(\underline{x}) - \int K(y, T_1 y, \ldots, T_1^{p-1} y, x_p, \ldots)\, dm(y) \\ &\quad = \sum_{i=0}^{p-1} \left( \frac{P_{i+1}^p \mathbb{1}(x_p)}{P_1^p \mathbb{1}(x_p)} - 1 \right) \int f_i\, dm \\ &\qquad + \frac{1}{P_1^p \mathbb{1}(x_p)} \sum_{i=0}^{p-1} \left( P_{i+1}^p f_i(x_p) - \left( \int f_i\, dm \right) P_{i+1}^p \mathbb{1}(x_p) \right). \end{aligned}$$

For the first term, we have

$$\left| \frac{P_{i+1}^p \mathbb{1}(x_p)}{P_1^p \mathbb{1}(x_p)} - 1 \right| = \frac{|P_{i+1}^p \mathbb{1}(x_p) - P_1^p \mathbb{1}(x_p)|}{P_1^p \mathbb{1}(x_p)} \leq C \| P_{i+1}^p (P_1^i \mathbb{1} - \mathbb{1}) \|_{\mathrm{BV}},$$

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

where we used *(Min)* and the fact that the BV norm dominates the supremum norm. Moreover, from *(Dec)*, we get $\|P_{i+1}^p(P_1^i\mathbb{1}-\mathbb{1})\|_{\mathrm{BV}} \le K\theta^{p-i-1}\|P_1^i\mathbb{1}-\mathbb{1}\|_{\mathrm{BV}} \le C\theta^{p-i-1}$, since $(P_1^i\mathbb{1})_i$ is bounded in BV by (H3). Finally, since $|\int f_i\,dm| \le \mathrm{Lip}_i(K)$, we obtain

$$\left|\sum_{i=0}^{p-1}\left(\frac{P_{i+1}^p\mathbb{1}(x_p)}{P_1^p\mathbb{1}(x_p)}-1\right)\int f_i\,dm\right| \le C\sum_{i=0}^{p-1}\theta^{p-i-1}\,\mathrm{Lip}_i(K). \tag{2}$$

For the second term, the hypothesis *(Dec)* allows us to estimate

$$\left\|P_{i+1}^p f_i-\left(\int f_i\,dm\right)P_{i+1}^p\mathbb{1}\right\|_{\mathrm{BV}} \le K\theta^{p-i-1}\left\|f_i-\int f_i\,dm\right\|_{\mathrm{BV}} \le C\theta^{p-i-1}\|f_i\|_{\mathrm{BV}}. \tag{3}$$

Using *(Min)*, (3) and $\|\cdot\|_{\sup}\le\|\cdot\|_{\mathrm{BV}}$, we have

$$\left|\frac{1}{P_1^p\mathbb{1}(x_p)}\sum_{i=0}^{p-1}\left(P_{i+1}^p f_i(x_p)-\left(\int f_i\,dm\right)P_{i+1}^p\mathbb{1}(x_p)\right)\right| \le C\sum_{i=0}^{p-1}\theta^{p-i-1}\|f_i\|_{\mathrm{BV}}. \tag{4}$$

Thus, it remains to estimate $\|f_i\|_{\mathrm{BV}}$. One can observe that $\|f_i\|_{\mathrm{BV}} \le \|f_i\|_{\sup}+\mathrm{V}(f_i)$ and that, for any $y$, we have

$$|f_i(y)| \le P_1^i\mathbb{1}(y)\|H_i\|_{\sup} \le P_1^i\mathbb{1}(y)\,\mathrm{Lip}_i(K) \le C\,\mathrm{Lip}_i(K), \tag{5}$$

since $(P_1^i\mathbb{1})_i$ is bounded in BV. The crucial point hence lies in the estimate of the variation of $f_i$.

To do so, we first establish a distortion control. The proof is standard, but we reproduce it here for completeness. If $T$ is a piecewise $C^2$ map of the interval, with partition of monotonicity $\mathcal{A}_T$, we define its distortion $\mathrm{Dist}(T)$ as the least constant $C$ such that $|T'(x)-T'(y)| \le C|T'(x)||Tx-Ty|$ for all $x,\ y\in I$ and $I\in\mathcal{A}_T$. Note that $(1+C)^{-1} \le (|T'(x)|/|T'(y)|) \le (1+C)$ for all $x,\ y\in I$.

LEMMA 6. *There exists $C>0$ such that $\mathrm{Dist}(T_1^n)\le C$ for all $n$.*

*Proof.* As previously, we denote by $\mathcal{A}_n$ the partition of monotonicity of $T_n$, and by $\mathcal{A}_n^m$ the partition of $T_n^m$, $n\le m$. We have, for any $x,\ y\in I\in\mathcal{A}_1^n$,

$$\left|\log\frac{|(T_1^n)'(x)|}{|(T_1^n)'(y)|}\right| \le \sum_{j=1}^{n}|\log|T_j'(T_1^{j-1}x)|-\log|T_j'(T_1^{j-1}y)|| \le C\sum_{j=1}^{n}|T_1^{j-1}x-T_1^{j-1}y|,$$

where we have used the fact that the Lipschitz constant of $\log|T_j'|$ is bounded independently of $j$, since $|T_j''|\le M$. As all maps are uniformly expanding by a factor of at least $\lambda$, and $x,\ y$ belong to the same partition element, we have $|T_1^{j-1}x-T_1^{j-1}y| \le \lambda^{-(n-j-1)}|T_1^n x-T_1^n y|$. Since $\sum_j\lambda^{-(n-j-1)}$ is bounded, this concludes the proof. □

We will need the following technical lemma, which is adapted from [**19**, Lemma II.4].

LEMMA 7. *There exists a constant $C>0$ such that, for all $n\ge 1$,*

$$\sum_{I\in\mathcal{A}_1^n}\sup_I\frac{1}{|(T_1^n)'|}\le C.$$

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

*Proof.* For any $I \in \mathcal{A}_1^n$, there exists a least integer $p = p_I \le n-1$ such that $T_1^p(I) \cap \partial\mathcal{A}_{p+1} \neq \emptyset$. Denote by $\mathcal{A}_1^{n,p}$ the set of all $I \in \mathcal{A}_1^n$ for which $p_I = p$. Consider $I \in \mathcal{A}_1^{n,p}$. There exists $a \in \partial I$ such that $b = T_1^p a \in \partial\mathcal{A}_{p+1}$. From Lemma 6, we deduce that there exists $C > 0$ such that, for any $x \in I$,

$$|(T_1^n)'(x)| \ge C|(T_1^n)'(a)| = C|(T_{p+1}^n)'(T_1^p a)||(T_1^p)'(a)| \ge C\lambda^{n-p}|(T_1^p)'(a)|.$$

Hence, one has $\sup_I (1/|(T_1^n)'|) \le C^{-1}\lambda^{-(n-p)}(1/|(T_1^p)'(a)|)$. Since a pre-image by $T_1^p$ of an element $b \in \mathcal{A}_{p+1}$ can only belong to at most two different $I \in \mathcal{A}_1^n$, it follows that

$$\begin{aligned}\sum_{I\in\mathcal{A}_1^n} \sup_I \frac{1}{|(T_1^n)'|} &\le 2C^{-1}\sum_{p=0}^{n-1}\lambda^{-(n-p)} \sum_{b\in\partial\mathcal{A}_{p+1}} \sum_{T_1^p a=b} \frac{1}{|(T_1^p)'(a)|}\\ &= 2C^{-1}\sum_{p=0}^{n-1}\lambda^{-(n-p)} \sum_{b\in\partial\mathcal{A}_{p+1}} P_1^p \mathbb{1}(b)\\ &\le 2C^{-1}\sum_{p=0}^{n-1}\lambda^{-(n-p)} \|P_1^p\mathbb{1}\|_{\mathrm{BV}} \sharp\partial\mathcal{A}_{p+1}.\end{aligned}$$

This quantity is bounded independently of $n$, since $(P_1^p)_p$ is bounded in BV, and the number of elements in $\mathcal{A}_{p+1}$ is bounded independently of $p$, by assumption. □

As a direct corollary of the two previous lemmas, we have the following result.

COROLLARY 8. *There exists $C > 0$ such that, for all $n \ge 1$,*

$$\sum_{I\in\mathcal{A}_1^n} \mathrm{V}_I\left(\frac{1}{|(T_1^n)'|}\right) \le C,$$

*where $\mathrm{V}_I(f)$ denotes the total variation of $f$ over the subinterval $I \subset [0, 1]$.*

*Proof.* Let $x_0 < \cdots < x_j$ be a sequence of elements of $I$. Then

$$\begin{aligned}\sum_{i=0}^{j-1}\left|\frac{1}{|(T_1^n)'(x_{i+1})|} - \frac{1}{|(T_1^n)'(x_i)|}\right| &= \sum_{i=0}^{j-1}\frac{||(T_1^n)'(x_{i+1})| - |(T_1^n)'(x_i)||}{|(T_1^n)'(x_i)||(T_1^n)'(x_{i+1})|}\\ &\le \sum_{i=0}^{j-1}\frac{\mathrm{Dist}(T_1^n)|T_1^n x_{i+1} - T_1^n x_i|}{|(T_1^n)'(x_i)|} \le C\sup_I\frac{1}{|(T_1^n)'|},\end{aligned}$$

where, for the last inequality, we have used Lemma 6 and the fact that $(T_1^n x_i)_i$ is a monotone sequence in [0, 1]. A direct application of Lemma 7 proves the corollary. □

We are now able to estimate $\mathrm{V}(f_i)$. For $I \in \mathcal{A}_1^i$, we denote by $S_{i,I}$ the inverse branch of the restriction of $T_1^i$ to $I$. Then we can write

$$f_i = P_1^i H_i = \sum_{I\in\mathcal{A}_1^i}\left(\frac{H_i}{|(T_1^i)'|}\right)\circ S_{i,I}\,\mathbb{1}_{T_1^i(I)}.$$

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

Using standard properties of the total variation, it follows that

$$\mathrm{V}(f_i) \le \sum_{I\in\mathcal{A}_1^i} \mathrm{V}_I\left(\frac{H_i}{|(T_1^i)'|}\right) + 2\sum_{a\in\partial\mathcal{A}_1^i} \frac{|H_i(a)|}{|(T_1^i)'(a)|} \le \mathrm{I}_i + \mathrm{II}_i + \mathrm{III}_i,$$

where

$$\mathrm{I}_i = \sum_{I\in\mathcal{A}_1^i} \mathrm{V}_I\left(\frac{1}{|(T_1^i)'|}\right) \sup_I |H_i|,$$

$$\mathrm{II}_i = \sum_{I\in\mathcal{A}_1^i} \sup_I \frac{1}{|(T_1^i)'|} \mathrm{V}_I(H_i),$$

$$\mathrm{III}_i = 2\sum_{a\in\partial\mathcal{A}_1^i} \frac{|H_i(a)|}{|(T_1^i)'(a)|}.$$

Using the Lipschitz condition on $K$, one gets $\mathrm{I}_i \le C\,\mathrm{Lip}_i(K)\sum_{I\in\mathcal{A}_1^i} \mathrm{V}_I(1/|(T_1^i)'|)$, which gives $\mathrm{I}_i \le C\,\mathrm{Lip}_i(K)$ by Corollary 8.

Let us now estimate $\mathrm{II}_i$. Let $y_0 < \cdots < y_l$ be a sequence of points in $I$. In order to estimate $\sum_{j=0}^{l-1} |H_i(y_{j+1}) - H_i(y_j)|$, we split $H_i$ into two terms in an obvious way, and we deal with the first one, the second being completely similar. We have

$$\begin{aligned}\sum_{j=0}^{l-1}\sum_{k=0}^{i} &|K(y_{j+1},\dots,T_1^k y_{j+1}, T_1^{k+1}y_j,\dots,T_1^i y_j,\dots)\\ &\qquad - K(y_{j+1},\dots,T_1^{k-1}y_{j+1}, T_1^k y_j,\dots,T_1^i y_j,\dots)|\\ &\le \sum_{j=0}^{l-1}\sum_{k=0}^{i} \mathrm{Lip}_k(K)|T_1^k y_{j+1} - T_1^k y_j| \le \sum_{k=0}^{i} \mathrm{Lip}_k(K) m(T_1^k(I)).\end{aligned}$$

Since $I$ belongs to $\mathcal{A}_1^i$, the interval $T_1^k(I)$ is included in an interval of monotonicity of $T_{k+1}^i$ and hence its length is less than $\lambda^{-(i-k)}$. Therefore, one has $\mathrm{V}_I(H_i) \le 2\sum_{k=0}^{i} \lambda^{-(i-k)}\,\mathrm{Lip}_k(K)$. An application of Lemma 7 then yields $\mathrm{II}_i \le C\sum_{k=0}^{i} \lambda^{-(i-k)}\,\mathrm{Lip}_k(K)$.

Using again Lemma 7, we can bound the third term by $\mathrm{III}_i \le C\,\mathrm{Lip}_i(K)$.

Putting together all the estimates, we find that

$$\mathrm{V}(f_i) \le C\sum_{k=0}^{i} \lambda^{-(i-k)}\,\mathrm{Lip}_k(K). \tag{6}$$

Finally, (5) and (6) give the same estimate for $\|f_i\|_{\mathrm{BV}}$.

Coming back to $K_p(\underline{x})$, we then have by (2) and (4)

$$\begin{aligned}&\left|K_p(\underline{x}) - \int K(y, T_1 y,\dots,T_1^{p-1}y, x_p,\dots)\,dm(y)\right|\\ &\qquad \le C\sum_{i=0}^{p-1}\theta^{p-i-1}\sum_{k=0}^{i}\lambda^{-(i-k)}\,\mathrm{Lip}_k(K),\end{aligned}$$

which is less than $C\sum_{k=0}^{p-1}\rho^{p-k}\,\mathrm{Lip}_k(K)$, as shown by a simple computation. This concludes the proof of our main theorem.

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

## 4. *Applications*

Let $(T_n)$ be a sequential dynamical system on $[0, 1]$ that satisfies all the conditions of Theorem 3. We describe some applications of the concentration inequality, following ideas from [**14**, **19**].

### 4.1. *Large deviations for ergodic sums.*

Considering observables $K$ of the form

$$K(x_0, \ldots, x_{n-1}) = \sum_{k=0}^{n-1} f(x_k)$$

with $f$ Lipschitz, an immediate application of Theorem 3 gives us the following result.

PROPOSITION 9. *There exists $C > 0$ such that, for all $f : [0, 1] \to \mathbb{R}$ Lipschitz, all $n \geq 1$ and all $t > 0$,*

$$m\left(\frac{1}{n}\sum_{k=0}^{n-1}\left[f \circ T_1^k - \int f \circ T_1^k \, dm\right] > t\right) \leq e^{-(Cnt^2/\operatorname{Lip}(f)^2)}.$$

In §5, we prove a similar statement for observables $f$ in BV.

### 4.2. *Empirical measure.*

For $x \in [0, 1]$, define for $n \geq 1$ the empirical measure

$$\mathcal{E}_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T_1^k x}.$$

By the strong law of large numbers, we know that for almost every $x$, this measure approximates in the weak topology, as $n \to \infty$, the measure $m_n$ defined by

$$m_n = \frac{1}{n}\sum_{k=0}^{n-1} (T_1^k)_\star m.$$

It is natural to try to quantify this phenomenon. For this, we need first to introduce a notion of distance on the set of probability measures. We will consider the Kantorovich distance, which, for probability measures $\mu_1$ and $\mu_2$ on $[0, 1]$, can be defined as

$$\kappa(\mu_1, \mu_2) = \int_0^1 |F_{\mu_1}(t) - F_{\mu_2}(t)| \, dt,$$

where $F_{\mu_i}(t) = \mu_i([0, t])$ is the distribution function of $\mu_i$. Thus, we have the following result.

PROPOSITION 10. *There exist $t_0 > 0$ and $C > 0$ such that, for all $t > t_0$ and $n \geq 1$,*

$$m\left(\kappa(\mathcal{E}_n, m_n) > \frac{t}{\sqrt{n}}\right) \leq e^{-Ct^2}.$$

The following proof is an application of our main theorem; in §5, we prove this proposition using a direct approach.

*Proof.* For $n \geq 1$, define

$$K_n(x_0, \ldots, x_{n-1}) = \int_0^1 \left|\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{[0,t]}(x_k) - F_{m_n}(t)\right| dt.$$

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

We clearly have $\kappa(\mathcal{E}_n(x), m_n) = K_n(x, \dots, T_1^{n-1}x)$ and $\mathrm{Lip}_j(K_n) \le 1/n$ for any $0 \le j \le n-1$. By the concentration inequality, and its large deviation counterpart, we derive

$$m\left(\kappa(\mathcal{E}_n, m_n) - \mathbb{E}_m(\kappa(\mathcal{E}_n, m_n)) > \frac{t}{\sqrt{n}}\right) \le e^{-Ct^2}.$$

To conclude, it is then sufficient to prove that $\mathbb{E}_m(\kappa(\mathcal{E}_n, m_n))$ is of order $1/\sqrt{n}$.

Denote by $\chi_t$ the characteristic function $\mathbb{1}_{[0,t]}$. Using the Schwarz inequality, we have

$$\begin{aligned}\mathbb{E}_m(\kappa(\mathcal{E}_n, m_n)) &= \int\left(\int_0^1 \left|\frac{1}{n}\sum_{k=0}^{n-1}\chi_t(T_1^k x) - F_{m_n}(t)\right| dt\right) dm(x)\\ &\le \left[\int_0^1\left(\int\left|\frac{1}{n}\sum_{k=0}^{n-1}(\chi_t\circ T_1^k - F_{(T_1^k)_\star m}(t))\right|^2 dm\right) dt\right]^{1/2}.\end{aligned}$$

Expanding the square, we have

$$\begin{aligned}&\int\left|\frac{1}{n}\sum_{k=0}^{n-1}(\chi_t\circ T_1^k - F_{(T_1^k)_\star m}(t))\right|^2 dm\\ &= \frac{1}{n^2}\sum_{k,l=0}^{n-1}\int(\chi_t\circ T_1^k - F_{(T_1^k)_\star m}(t))(\chi_t\circ T_1^l - F_{(T_1^l)_\star m}(t))\, dm.\end{aligned}$$

If $k \le l$, by the properties of transfer operators, and since $F_{(T_1^k)_\star m}(t) = \int \chi_t\circ T_1^k\, dm$,

$$\begin{aligned}&\int\left(\chi_t\circ T_1^k - F_{(T_1^k)_\star m}(t)\right)\left(\chi_t\circ T_1^l - F_{(T_1^l)_\star m}(t)\right) dm\\ &= \int\left(\chi_t - \int\chi_t\circ T_1^l\right)P_{k+1}^l\left(\left(\chi_t - \int\chi_t\circ T_1^k\right)P_1^k\mathbb{1}\right) dm.\end{aligned}$$

Since $(\chi_t - \int\chi_t\circ T_1^k)P_1^k\mathbb{1}$ has 0 integral, $\{\chi_t\}_t$ is a bounded family in BV, and $(P_1^k\mathbb{1})_k$ is also bounded in BV, we can use property *(Dec)* to get

$$\left|\int\left(\chi_t - \int\chi_t\circ T_1^l\right)P_{k+1}^l\left(\left(\chi_t - \int\chi_t\circ T_1^k\right)P_1^k\mathbb{1}\right) dm\right| \le C\theta^{l-k}. \tag{7}$$

It follows that

$$\mathbb{E}_m(\kappa(\mathcal{E}_n, m_n)) \le \frac{C}{n}\left(\sum_{k,l=0}^{n-1}\theta^{-|k-l|}\right)^{1/2} = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \qquad \square$$

4.3. *Efficiency of shadowing.* Let $A$ be a measurable subset of $[0, 1]$ with positive measure. How well can we approximate the trajectory of a point $x \in [0, 1]$ by a trajectory starting in $A$? The following result provides an estimation of the average quality of this shadowing.

PROPOSITION 11. *There exist constants $C_1, C_2 > 0$ such that, for all $A$ with $m(A) > 0$ and all $n \ge 1$, the sequence $(Z_n)$ of functions defined by*

$$Z_n(x) = \inf_{y\in A}\frac{1}{n}\sum_{k=0}^{n-1}|T_1^k x - T_1^k y|$$

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

*satisfies for any $t > 0$,*

$$m\left(Z_n > C_1 \frac{\sqrt{|\log m(A)|}}{\sqrt{n}} + \frac{t}{\sqrt{n}}\right) \le e^{-C_2 t^2}.$$

*Proof.* We apply the concentration inequality to the functions

$$K_n(x_0, \dots, x_{n-1}) = \frac{1}{n} \inf_{y \in A} \sum_{k=0}^{n-1} |T_1^k y - x_k|.$$

They satisfy $\mathrm{Lip}_k(K_n) \le 1/n$, so we get

$$m\left(Z_n - \mathbb{E}_m(Z_n) > \frac{t}{\sqrt{n}}\right) \le e^{-Ct^2}.$$

We now estimate $\mathbb{E}_m(Z_n)$. One has

$$m(A) \le m\left(\mathbb{E}_m(Z_n) - Z_n > \frac{1}{\sqrt{n}} \frac{\sqrt{n}|\mathbb{E}_m(Z_n)|}{2}\right),$$

since $Z_n = 0$ on $A$. By the concentration inequality, we deduce that $m(A) \le e^{-Cn(\mathbb{E}_m(Z_n))^2/4}$, whence

$$\mathbb{E}_m(Z_n) \le C_1 \frac{\sqrt{|\log m(A)|}}{\sqrt{n}}. \qquad \square$$

4.4. *Almost sure central limit theorem.* Let $(Z_n)$ be a random process that satisfies the central limit theorem, i.e. there exist $A_n \in \mathbb{R}$ and $B_n > 0$ such that $S_n = \sum_{k=0}^{n-1} Z_k$ verifies

$$\frac{S_n - A_n}{B_n} \to W$$

in the weak topology, where $W$ is distributed as a standard normal law $\mathcal{N}(0, 1)$.

By definition, this means that for every $t \in \mathbb{R}$,

$$\mathbb{E}(\mathbb{1}_{\{((S_n - A_n)/B_n) \le t\}}) \to \mathbb{P}(W \le t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-s^2/2}\, ds.$$

We can then ask about the pointwise behavior of $\mathbb{1}_{\{((S_n - A_n)/B_n) \le t\}}$. If this quantity cannot in general converge almost surely, it appears that after a correct averaging, one can recover the convergence. It turns out that the correct averaging is a logarithmic one; more precisely, we say that $(Z_n)$ satisfies an almost sure central limit theorem if almost surely

$$\frac{1}{H_n} \sum_{k=1}^{n} \frac{1}{k} \delta_{((S_k - A_k)/B_k)} \to W$$

in the weak topology, where $H_n = \sum_{k=1}^{n} 1/k$ (one can observe that $H_n \sim \log n$).

We refer the interested reader to [**9**, **27**] for particularly nice reviews of this subject, and to [**15**] for a description of applications in dynamical systems.

Let now $(T_n)$ be a sequential dynamical system on $[0, 1]$ which satisfies the same assumptions as before. Let $f : [0, 1] \to \mathbb{R}$ be a Lipschitz observable, and define

$$S_n = \sum_{k=0}^{n-1} \left[ f \circ T_1^k - \int f \circ T_1^k \, dm \right].$$

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

If $\|S_n\|_2$ is unbounded, we have by Conze and Raugi that $S_n/\|S_n\|_2$ goes in distribution to $W \sim \mathcal{N}(0, 1)$.

In order to prove an almost sure version of this convergence, we will need an additional assumption on the growth of the variance of the ergodic sum, namely that there exists a $C > 0$ such that $\|S_n\|_2 \geq Cn^{1/2}$ for all $n$.

PROPOSITION 12. *Let $f : [0, 1] \to \mathbb{R}$ be a Lipschitz observable, and define $S_n$ as above. Assume that $\|S_n\|_2 \geq Cn^{1/2}$ for some $C > 0$. Then $S_n$ satisfies the almost sure central limit theorem:*

$$\frac{1}{H_n}\sum_{k=1}^{n}\frac{1}{k}\delta_{(S_k(x)/\|S_k\|_2)} \xrightarrow{\mathcal{L}} W$$

*for almost every $x \in [0, 1]$.*

*Proof.* We have to show that almost everywhere, $(1/H_n)\sum_{k=1}^{n}(1/k)g(S_k/\|S_k\|_2) \to \mathbb{E}(g(W))$ for all $g : \mathbb{R} \to \mathbb{R}$ continuous and bounded. By standard density and separability arguments (see e.g. [**27**, Theorem 2.4]), it is sufficient to prove that for all $g$ bounded and Lipschitz, this convergence occurs almost surely. We remark also that we can assume without loss of generality that $g(0) = 0$. Hence, there exists $C > 0$ such that $|g(x)| \leq C|x|$ for all $x$.

Define

$$K_n(x_0, \ldots, x_{n-1}) = \frac{1}{H_n}\sum_{k=1}^{n}\frac{1}{k}g\left(\frac{\sum_{j=0}^{k-1}[f(x_j) - \int f \circ T_1^j]}{\|S_k\|_2}\right).$$

We then have $K_n(x, \ldots, T_1^{n-1}x) = (1/H_n)\sum_{k=1}^{n}(1/k)g(S_k/\|S_k\|_2)$, and the assumption on $\|S_n\|_2$ gives

$$\mathrm{Lip}_j(K_n) \leq \frac{1}{H_n}\sum_{k>j}\frac{1}{k}\mathrm{Lip}(g)\frac{\mathrm{Lip}(f)}{\|S_k\|_2} = \mathcal{O}\left(\frac{1}{H_n}\sum_{k>j}k^{-3/2}\right) = \mathcal{O}\left(\frac{1}{j^{1/2}H_n}\right).$$

Hence,

$$\sum_{j=0}^{n-1}\mathrm{Lip}_j^2(K_n) = \mathcal{O}\left(\frac{1}{H_n^2}\sum_{j=0}^{n-1}j^{-1}\right) = \mathcal{O}\left(\frac{1}{\log n}\right).$$

Using the concentration inequality, we deduce that

$$m\left(\left|\frac{1}{H_n}\sum_{k=1}^{n}\frac{1}{k}g\left(\frac{S_k}{\|S_k\|_2}\right) - \mathbb{E}_m\left(\frac{1}{H_n}\sum_{k=1}^{n}\frac{1}{k}g\left(\frac{S_k}{\|S_k\|_2}\right)\right)\right| \geq t\right) \leq e^{-Ct^2\log n}.$$

Hence, if we define the subsequence $n_j = e^{j^\alpha}$ with $\alpha > 0$, we have that $(1/H_{n_j})\sum_{k=1}^{n_j}(1/k)g(S_k/\|S_k\|_2)$ goes to $\mathbb{E}(g(W))$ by the Borel–Cantelli lemma, since $\mathbb{E}_m((1/H_{n_j})\sum_{k=1}^{n_j}(1/k)g(S_k/\|S_k\|_2))$ goes to $\mathbb{E}(g(W))$ when $j \to \infty$ by the usual central limit theorem.

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

It remains to control the gaps. If $n_j < n \le n_{j+1}$, we have

$$\frac{1}{H_n}\sum_{k=1}^{n}\frac{1}{k}g\left(\frac{S_k}{\|S_k\|_2}\right)-\frac{1}{H_{n_j}}\sum_{k=1}^{n_j}\frac{1}{k}g\left(\frac{S_k}{\|S_k\|_2}\right)$$
$$=\frac{1}{H_n}\sum_{k=n_j+1}^{n}\frac{1}{k}g\left(\frac{S_k}{\|S_k\|_2}\right)+\left(\frac{H_{n_j}}{H_n}-1\right)\frac{1}{H_{n_j}}\sum_{k=1}^{n_j}\frac{1}{k}g\left(\frac{S_k}{\|S_k\|_2}\right).$$

Using the previous observations, the supremum over $n_j < n \le n_{j+1}$ of the second term tends to 0 when $j \to \infty$. For the first term, the assumption on $g$ gives us

$$\sup_{n_j<n\le n_{j+1}}\left|\frac{1}{H_n}\sum_{k=n_j+1}^{n}\frac{1}{k}g\left(\frac{S_k}{\|S_k\|_2}\right)\right|\le\frac{C}{H_{n_j}}\sum_{k=n_j+1}^{n_{j+1}}\frac{|S_k|}{k\|S_k\|_2}.$$

We now prove the almost sure convergence to zero of

$$G_j=\frac{1}{H_{n_j}}\sum_{k=n_j+1}^{n_{j+1}}\frac{|S_k|}{k\|S_k\|_2}.$$

Using the Schwarz inequality, we have

$$\mathbb{E}_m(G_j^2)\le\frac{1}{H_{n_j}^2}\sum_{k,l=n_j+1}^{n_{j+1}}\frac{\mathbb{E}_m(S_k^2)^{1/2}\mathbb{E}_m(S_l^2)^{1/2}}{k\|S_k\|_2 l\|S_l\|_2}=\frac{1}{H_{n_j}^2}\left(\sum_{k=n_j+1}^{n_{j+1}}k^{-1}\right)^2$$
$$=\frac{(\log n_{j+1}-\log n_j+\mathcal{O}(1))^2}{H_{n_j}^2}\le\frac{\mathcal{O}(1)}{j^2}.$$

The result follows by the Borel–Cantelli lemma, since $\mathbb{E}_m(G_j^2)$ is summable. □

## 5. *Large deviations and empirical measure: an alternative approach*

In the previous section, we gave some applications of Theorem 3. In this section, we will show that one can use a direct approach to get large deviations estimates for observables in BV and also an upper bound on the Kantorovich distance.

### 5.1. *Large deviations estimates for observables in BV.*

Proposition 9 gives exponential large deviations estimates for ergodic sums associated to Lipschitz observables, for sequential dynamical systems on $[0, 1]$ that satisfy all the assumptions required by Theorem 3. As the strong law of large numbers holds for observables in BV under slightly weaker assumptions for the dynamical system, it is natural to ask whether large deviations estimates remain valid in this more general situation. The following proposition answers positively this question.

PROPOSITION 13. *Let $(T_n)$ be a sequential dynamical system on $[0, 1]$ such that all maps $T_n$ belong to a class $\mathcal{C}$ that satisfies $(D_r)$ and (Dec), and for which the condition (Min) holds. Then, for any $f \in \mathrm{BV}$ and all $t > 0$, there exist $\tau = \tau(f) > 0$ and $C = C(f, t) > 0$ such that, for all $n \ge 1$,*

$$m\left(\frac{1}{n}\sum_{k=0}^{n-1}\left[f\circ T_1^k-\int f\circ T_1^k\,dm\right]>t\right)\le Ce^{-\tau nt^2}.$$

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

*Proof.* We follow the strategy employed in [**3**, Proposition 2.5], using the martingale approximation described in [**20**, §5]. Write $f_n = f - \int f \circ T_1^n \, dm$ and define the operators $Q_n$, for $n \geq 1$, by

$$Q_n g = \frac{P_n(g P_1^{n-1} \mathbb{1})}{P_1^n \mathbb{1}}.$$

Let $h_n$ be defined by the relation $h_{n+1} = Q_{n+1} f_n + Q_{n+1} h_n$ with $h_0 = 0$. We get

$$h_n = \frac{1}{P_1^n \mathbb{1}} \sum_{k=0}^{n-1} P_{k+1}^n (f_k P_1^k \mathbb{1}).$$

Since BV is a Banach algebra and $(P_1^k \mathbb{1})_k$ is bounded in BV, we get that $f_k P_1^k \mathbb{1}$ is a bounded sequence in BV of functions with integral 0. As a consequence of *(Dec)*, the BV norm of the term $P_{k+1}^n (f_k P_1^k \mathbb{1})$ decays exponentially fast with $n - k$, which implies, thanks to condition *(Min)*, that the sequence $(h_n)$ is bounded for the supremum norm.

Define $\varphi_n = f_n + h_n - h_{n+1} \circ T_{n+1}$ and $U_n = \varphi_n \circ T_1^n$. We have that $(U_n)$ is a sequence of reversed martingales for the filtration $(\mathcal{F}_n)$, and is bounded for the supremum norm. Denoting $S_n = \sum_{k=0}^{n-1} f_k \circ T_1^k = \sum_{k=0}^{n-1} [f \circ T_1^k - \int f \circ T_1^k \, dm]$, we have

$$S_n = \sum_{k=0}^{n-1} U_k + h_n \circ T_1^n.$$

Then $m(S_n > nt) = m(\sum_{k=0}^{n-1} U_k + h_n \circ T_1^n > nt) \leq m(\sum_{k=0}^{n-1} U_k > (nt/2))$, since $(h_n)$ is bounded for the supremum norm and thus $h_n \circ T_1^n < (nt/2)$ for all $n$ large enough. On the other hand, using the Azuma–Hoeffding inequality, there exists a constant $\tau = \tau(f)$ such that $m(\sum_{k=0}^{n-1} U_k > (nt/2)) \leq e^{-\tau n t^2}$, since $(U_n)$ is also bounded for the supremum norm. This proves the proposition. □

5.2. *Empirical measure and Kantorovich distance.* In Proposition 10, an upper bound for the Kantorovich distance was obtained using Theorem 3. Following Dedecker and Merlevède [**21**], we will prove this result for a larger class of sequential dynamical systems via an alternative approach.

Recall that for a sequential dynamical system $(T_n)$ on $[0, 1]$, we define the empirical measure by $\mathcal{E}_n(x) = (1/n) \sum_{k=0}^{n-1} \delta_{T_1^k x}$ and the measure $m_n$ by $m_n = (1/n) \sum_{k=0}^{n-1} (T_1^k)_\star m$.

PROPOSITION 14. *Let $(T_n)$ be a sequential dynamical system on* $[0, 1]$ *such that all maps $T_n$ belong to a class $\mathcal{C}$ that satisfies $(D_r)$ and* (Dec)*, and for which the condition (Min) holds. Then there exists $C > 0$ such that, for all $t > 0$ and all $n \geq 1$,*

$$m\left(\kappa(\mathcal{E}_n, m_n) > \frac{t}{\sqrt{n}}\right) \leq 2e^{-Ct^2}.$$

*Proof.* Let

$$S_n(t) = \sum_{k=0}^{n-1} \mathbb{1}_{Z_k \leq t} - \mathbb{E}(\mathbb{1}_{Z_k \leq t}).$$

Let $\| \cdot \|_{L_m^p}$ denote the $L^p$-norm under $m$; one can observe that

$$\kappa(\mathcal{E}_n, m_n) = \frac{1}{n} \|S_n\|_{L_m^1} \leq \frac{1}{n} \|S_n\|_{L_m^2}.$$

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

Let $\mathcal{M}_n = \{\emptyset, [0, 1]\}$ be the trivial $\sigma$-algebra and let $\mathcal{M}_k = \sigma(T_1^k)$ for $0 \le k \le n-1$. Thus, using Yurinskii's idea [**42**], since $\mathbb{E}(S_n|\mathcal{M}_n) = \mathbb{E}(S_n) = 0$, we have

$$S_n = \sum_{i=0}^{n-1} \mathbb{E}(S_n|\mathcal{M}_i) - \mathbb{E}(S_n|\mathcal{M}_{i+1}) = \sum_{i=0}^{n-1} d_{i,n},$$

where

$$d_{i,n}(t) = \sum_{k=0}^{i} \mathbb{E}(\mathbb{1}_{Z_k \le t}|\mathcal{M}_i) - \mathbb{E}(\mathbb{1}_{Z_k \le t}|\mathcal{M}_{i+1}).$$

Thus, using [**39**, Theorem 3], an Azuma-type inequality in Hilbert spaces, we get

$$m\left(\kappa(\mathcal{E}_n, m_n) > \frac{t}{\sqrt{n}}\right) \le m\left(\frac{1}{\sqrt{n}}\|S_n\|_{2,m} > t\right) \le 2e^{-(nt^2/2b_n^2)},$$

where

$$b_n^2 = \sum_{i=0}^{n-1} \|\|d_{i,n}\|_{L_m^2}\|_\infty^2. \tag{8}$$

We thus only need to bound $b_n^2$. We observe that

$$\|d_{i,n}\|_{L_m^2} \le \sum_{k=0}^{i} \|\mathbb{E}(\mathbb{1}_{Z_k \le \cdot}|\mathcal{M}_i) - \mathbb{E}(\mathbb{1}_{Z_k \le \cdot})\|_{L_m^2} + \sum_{k=0}^{i} \|\mathbb{E}(\mathbb{1}_{Z_k \le \cdot}|\mathcal{M}_{i+1}) - \mathbb{E}(\mathbb{1}_{Z_k \le \cdot})\|_{L_m^2}. \tag{9}$$

Now one can use the properties of the transfer operator to show that

$$\mathbb{E}(\mathbb{1}_{Z_k \le t}|\mathcal{M}_i) - \mathbb{E}(\mathbb{1}_{Z_k \le t}) = \frac{P_{k+1}^i((\chi_t - \mathbb{E}(\chi_t(Z_k)))P_1^k(\mathbb{1}))}{P_1^i(\mathbb{1})} \circ T_1^i;$$

moreover, using *(Min)*,

$$\|\mathbb{E}(\mathbb{1}_{Z_k \le \cdot}|\mathcal{M}_i) - \mathbb{E}(\mathbb{1}_{Z_k \le \cdot})\|_{L_m^2} \le \frac{\sup_{t\in[0,1]} |P_{k+1}^i((\chi_t - \mathbb{E}(\chi_t(Z_k)))P_1^k(\mathbb{1}))|}{\delta}.$$

Finally, using *(Dec)* as in (7), we obtain

$$\|\mathbb{E}(\mathbb{1}_{Z_k \le \cdot}|\mathcal{M}_i) - \mathbb{E}(\mathbb{1}_{Z_k \le \cdot})\|_{L_m^2} \le \frac{C\theta^{i-k}}{\delta}.$$

Since $\sum_{k=0}^{i} \theta^{i-k}$ is bounded uniformly in $i$, from (8) and (9) we get

$$b_n^2 \le \sum_{i=0}^{n-1}\sum_{i=0}^{n-1} \|\|d_{i,n}\|_{L_m^2}\|_\infty^2 \le Cn,$$

which gives us

$$m\left(\kappa(\mathcal{E}_n, m_n) > \frac{t}{\sqrt{n}}\right) \le 2\exp(-Ct^2)$$

and concludes the proof. □

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

*Acknowledgements.* The authors would like to thank the anonymous referee for useful comments and in particular for bringing to our knowledge the paper [**21**] and for suggesting another proof for the upper bound on the Kantorovich distance. RA is also grateful to Sébastien Gouëzel, who suggested to him an improvement for Proposition 11. RA would like to thank Carlangelo Liverani for a discussion on the condition *(Min)* and for suggesting to him to use his paper [**32**] and Sandro Vaienti for several discussions and constant encouragement. RA was partially supported by Conseil Régional Provence-Alpes-Côte d'Azur, the ANR-Project *Perturbations*, by the PICS (Projet International de Coopération Scientifique), Propriétés statistiques des systèmes dynamiques deterministes et aléatoires, with the University of Houston, no. PICS05968, and by the European Advanced Grant Macroscopic Laws and Dynamical Systems (MALADY) (ERC AdG 246953). Most of this work was done when RA was affiliated to Aix Marseille Université, CNRS, CPT, UMR 7332, 13288 Marseille, France and Université de Toulon, CNRS, CPT, UMR 7332, 83957 La Garde, France. RA would also like to acknowledge the Newton Institute of Mathematics at Cambridge, where part of this work was done. JR was partially supported by CNPq and FAPESB. JR would also like to acknowledge the Université de Toulon, where part of this work was done.

## A. *Appendix. Proof of a formula on conditional expectations*

In this appendix, we prove the following relation:

$$K_p(\underline{x}) = \frac{1}{P_1^p \mathbb{1}(x_p)} \sum_{T_1^p y = x_p} \frac{K(y, T_1 y, \ldots, T_1^{p-1} y, x_p, \ldots)}{|(T_1^p)'(y)|}, \tag{A1}$$

which holds for $\tilde{m}$-almost every $\underline{x} \in X^{\mathbb{N}}$.

First, we recall some basic results on transfer operators that will be needed in the proof. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two probability spaces and let $T : X \to Y$ be a measurable non-singular map. Its associated transfer operator $P_T : L^1(\mu) \to L^1(\nu)$ is defined as follows: for $f \in L^1(\mu)$, $P_T(f)$ is the Radon–Nikodym derivative $(d(\lambda \circ T^{-1})/d\nu)$, where $d\lambda = f d\mu$.

It satisfies the following properties.

LEMMA A.1. *For all $f \in L^1(\mu)$ and $g \in L^1(\nu)$, we have:*

(1) $\int_X fg \circ T \, d\mu = \int_Y P_T(f) g \, d\nu$;

(2) $P_T(g \circ T) = g P_T \mathbb{1}$;

(3) $\mathbb{E}_\mu(f | T^{-1}\mathcal{B}) = (P_T(f)/P_T\mathbb{1}) \circ T$, *where $P_T(f)/P_T\mathbb{1}$ is defined as* 0 *on the set* $\{P_T\mathbb{1} = 0\}$. *In particular,* $(P_T(f_1)/P_T(f_2)) \circ T = (\mathbb{E}_\mu(f_1|T^{-1}\mathcal{B})/\mathbb{E}_\mu(f_2|T^{-1}\mathcal{B}))$ *for all* $f_1, f_2 \in L^1(\mu)$.

We now recall our setting. Let $(T_n)_{n\geq 1}$ be a sequence of piecewise expanding maps of the unit interval $(X, \mathcal{F}, m)$. We define $\Phi : X \to X^{\mathbb{N}}$ by $\Phi(x) = (T_1^n x)_{n\geq 0}$. The transfer operator of the map $T_1^p : X \to X$ is denoted by $P_1^p$. Also, $X^{\mathbb{N}}$ is endowed with the product $\sigma$-algebra $\tilde{\mathcal{F}}$ and the probability measure $\tilde{m}$, which is the image of $m$ by $\Phi$.

Let $\tilde{\mathcal{F}}_p$ be the $\sigma$-algebra on $X^{\mathbb{N}}$ of events depending only on the coordinates $(x_k)_{k\geq p}$. Clearly, $\tilde{\mathcal{F}}_p = \sigma^{-p}\tilde{\mathcal{F}}$, where $\sigma : X^{\mathbb{N}} \to X^{\mathbb{N}}$ is the shift map. Let $K : X^n \to \mathbb{R}$ be a

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

separately Lipschitz function, which we extend as a function $K : X^{\mathbb{N}} \to \mathbb{R}$ belonging to $L^1(\tilde{m})$. Finally, we set $K_p = \mathbb{E}_{\tilde{m}}(K|\tilde{F}_p)$.

LEMMA A.2. *We have* $K_p \circ \Phi = (P_1^p(K \circ \Phi)/P_1^p \mathbb{1}) \circ T_1^p$, *m-almost everywhere.*

*Proof.* By Lemma A.1(3), we have $K_p = \mathbb{E}_{\tilde{m}}(K|\sigma^{-p}\tilde{\mathcal{F}}) = (P_\sigma^p K/P_\sigma^p \mathbb{1}) \circ \sigma^p$. Introduce the map $\Phi_p : X \to X^{\mathbb{N}}$ defined by $\Phi_p(x) = (x, T_{p+1}x, T_{p+1}^{p+2}x, \ldots) = (T_{p+1}^{p+n}x)_{n\geq 0}$. This map satisfies $\sigma^p \circ \Phi = \Phi_p \circ T_1^p$. It is not non-singular and then its transfer operator is not well defined, but the previous relation implies that $P_{\Phi_p}(P_1^p f)$ is well defined for all $f \in L^1(m)$, is equal to $P_\sigma^p P_\Phi f$, and satisfies all the properties described in Lemma A.1.

We then have

$$K_p \circ \Phi = \left(\frac{P_\sigma^p K}{P_\sigma^p \mathbb{1}}\right) \circ \Phi_p \circ T_1^p = \left(\frac{P_\sigma^p P_\Phi(K \circ \Phi)}{P_\sigma^p P_\Phi \mathbb{1}}\right) \circ \Phi_p \circ T_1^p,$$

by Lemma A.1(2), since $P_\Phi \mathbb{1} = \mathbb{1}$.

Moreover, since $P_\sigma^p P_\Phi = P_{\Phi_p} P_1^p$, one can use Lemma A.1(3) to obtain

$$\begin{aligned} K_p \circ \Phi &= \left(\frac{P_{\Phi_p} P_1^p(K \circ \Phi)}{P_{\Phi_p} P_1^p \mathbb{1}}\right) \circ \Phi_p \circ T_1^p \\ &= \left(\frac{((P_{\Phi_p}\mathbb{1}) \circ \Phi_p)\mathbb{E}_m(P_1^p(K \circ \Phi)|(\Phi_p)^{-1}\tilde{\mathcal{F}})}{(((P_{\Phi_p}\mathbb{1}) \circ \Phi_p)\mathbb{E}_m(P_1^p \mathbb{1})|(\Phi_p)^{-1}\tilde{\mathcal{F}})}\right) \circ T_1^p. \end{aligned}$$

The lemma is proved as soon as we remark that $(\Phi_p)^{-1}\tilde{\mathcal{F}} = \mathcal{F}$. □

Finally, this lemma implies (A1). Indeed, if we define $\pi_p : X^{\mathbb{N}} \to X$ to be the map that sends $\underline{x}$ to its $p$th coordinate $x_p$, then $T_1^p = \pi_p \circ \Phi$, and the lemma asserts that $K_p \circ \Phi = (P_1^p(K \circ \Phi)/P_1^p \mathbb{1}) \circ \pi_p \circ \Phi$, $m$-almost everywhere.

This implies that $K_p = (P_1^p(K \circ \Phi)/P_1^p \mathbb{1}) \circ \pi_p$, $\tilde{m}$-almost everywhere, and proves the relation, since

$$\left(\frac{P_1^p(K \circ \Phi)}{P_1^p \mathbb{1}}\right) \circ \pi_p(\underline{x}) = \frac{1}{P_1^p \mathbb{1}(x_p)} \sum_{T_1^p y = x_p} \frac{K(y, T_1 y, \ldots, T_1^{p-1} y, x_p, \ldots)}{|(T_1^p)'(y)|}.$$

To see that, set $A = \{K_p \circ \Phi = (P_1^p(K \circ \Phi)/P_1^p \mathbb{1}) \circ \pi_p \circ \Phi\}$. We have $A \in \mathcal{F}$, and $m(A) = 1$ by Lemma A.2. Since $(X, \mathcal{F})$ and $(X^{\mathbb{N}}, \tilde{\mathcal{F}})$ are both Polish spaces endowed with their Borel $\sigma$-algebras, and the map $\Phi : X \to X^{\mathbb{N}}$ is injective and measurable, we obtain that $\Phi(A) \in \tilde{\mathcal{F}}$ by [**18**, Theorem 8.3.7]. Then $\tilde{m}(\Phi(A)) = m(\Phi^{-1}(\Phi(A))) \geq m(A) = 1$, which shows that $K_p = (P_1^p(K \circ \Phi)/P_1^p \mathbb{1}) \circ \pi_p$, $\tilde{m}$-almost everywhere, since the set on which this relation holds contains $\Phi(A)$.

## REFERENCES

[**1**] R. Aimino, H. Hu, M. Nicol, A. Török and S. Vaienti. Polynomial loss of memory for maps of the interval with a neutral fixed point. *Discrete Contin. Dynam. Syst. A* **35**(3) (2015), 793–806.

[**2**] R. Aimino, M. Nicol and S. Vaienti. Annealed and quenched limit theorems for random expanding dynamical systems. *Probab. Theory Related Fields*, to appear, doi:10.1007/s00440-014-0571-y. *Preprint*, 2013, arXiv:1310.4359.

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

[**3**] J. F. Alves, J. M. Freitas, S. Luzzato and S. Vaienti. From rates of mixing to recurrence times via large deviations. *Adv. Math.* **228** (2011), 1203–1236.
[**4**] K. Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J.* (2) **19** (1967), 357–367.
[**5**] V. I. Bakhtin. Random processes generated by a hyperbolic sequence of mappings, I. *Russ. Acad. Sci. Izv. Math.* **44** (1995), 247–279.
[**6**] V. I. Bakhtin. Random processes generated by a hyperbolic sequence of mappings, II. *Russ. Acad. Sci. Izv. Math.* **44** (1995), 617–627.
[**7**] V. Baladi. *Positive Transfer Operators and Decay of Correlations (Advanced Series in Nonlinear Dynamics,* 16*)*. World Scientific, River Edge, NJ, 2000.
[**8**] D. Berend and V. Bergelson. Ergodic and mixing sequences of transformations. *Ergod. Th. & Dynam. Sys.* **4** (1984), 353–366.
[**9**] I. Berkes. Results and problems related to the pointwise central limit theorem. *Asymptotic Methods in Probability and Statistics (Ottawa, ON, 1997)*. Ed. B. Szyszkowicz. North-Holland, Amsterdam, 1998, pp. 59–96.
[**10**] S. Boucheron, G. Lugosi and P. Massart. *Concentration Inequalities: a Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013.
[**11**] A. Boyarsky and P. Góra. *Laws of Chaos. Invariant Measures and Dynamical Systems in One Dimension (Probability and its Applications)*. Birkhaüser, Boston, 1997.
[**12**] J.-R. Chazottes, P. Collet, F. Redig and E. Verbitskiy. A concentration inequality for interval maps with an indifferent fixed point. *Ergod. Th. & Dynam. Sys.* **29** (2009), 1097–1117.
[**13**] J.-R. Chazottes, P. Collet and B. Schmitt. Devroye inequality for a class of non-uniformly hyperbolic dynamical systems. *Nonlinearity* **18** (2005), 2323–2340.
[**14**] 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.
[**15**] 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.
[**16**] J.-R. Chazottes and S. Gouëzel. Optimal concentration inequalities for dynamical systems. *Comm. Math. Phys.* **316** (2012), 843–889.
[**17**] G. Chen and Y. Shi. Chaos of time-varying discrete dynamical systems. *J. Difference Equ. Appl.* **15** (2009), 429–449.
[**18**] D. L. Cohn. *Measure Theory: Second Edition (Birkhaüser Advanced Texts)*. Birkhaüser/Springer, New York, 2013.
[**19**] P. Collet, S. Martinez and B. Schmitt. Exponential inequalities for dynamical measures of expanding maps of the interval. *Probab. Theory Related Fields* **123** (2002), 301–322.
[**20**] J.-P. Conze and A. Raugi. Limit theorems for sequential expanding dynamical systems on [0, 1]. *Ergodic Theory and Related Fields (Contemporary Mathematics, 430)*. American Mathematical Society, Providence, RI, 2007, pp. 89–121.
[**21**] J. Dedecker and F. Merlevède. The empirical distribution function for dependent variables: asymptotic and nonasymptotic results in $L^p$. *ESAIM Probab. Stat.* **11** (2007), 102–114.
[**22**] S. Gouëzel and I. Melbourne. Moment bounds and concentration inequalities for slowly mixing dynamical systems. *Electron. J. Probab.* **19**(93) (2014), 30.
[**23**] C. Gupta, W. Ott and A. Török. Memory loss for time-dependent piecewise-expanding systems in higher dimension. *Math. Res. Lett.* **20** (2013), 141–161.
[**24**] N. Haydn, M. Nicol, A. Török and S. Vaienti. Almost sure invariance principle for sequential and non-stationary dynamical systems. *Preprint*, 2014.
[**25**] H. Hennion and L. Hervé. *Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasicompactness (Lecture Notes in Mathematics, 1766)*. Springer, Berlin, 2001.
[**26**] C. T. Ionescu-Tulcea and G. Marinescu. Théorie ergodique pour des classes d'opérations non complètement continues. *Ann. Math.* **52** (1950), 140–147.
[**27**] F. Jonsson. Almost sure central limit theory. *Project Report*, Department of Mathematics, Uppsala University, 2007.
[**28**] C. Kawan. Metric entropy of nonautonomous dynamical systems. *Nonauton. Dyn. Syst.* **1** (2014), 1–71.
[**29**] S. Kolyada and L. Snoha. Topological entropy of nonautonomous dynamical systems. *Random Comput. Dyn.* **4** (1996), 205–223.
[**30**] A. Lasota and J.-A. Yorke. On the existence of invariant measures for piecewise monotonic transformations. *Trans. Amer. Math. Soc.* **186** (1973), 481–488.
[**31**] M. Ledoux. *The Concentration of Measure Phenomenon (Mathematical Surveys and Monographs, 89)*. American Mathematical Society, Providence, RI, 2001.
[**32**] C. Liverani. Decay of correlations for piecewise expanding maps. *J. Stat. Phys.* **78** (1995), 1111–1129.

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press

[**33**] C. Maldonado. Fluctuation bounds for chaos plus noise in dynamical systems. *J. Stat. Phys.* **148** (2012), 548–564.
[**34**] C. McDiarmid. On the method of bounded differences. *Surveys in Combinatorics (London Mathematical Society Lecture Note Series, 141)*. Cambridge University Press, Cambridge, 1989, pp. 148–188.
[**35**] C. McDiarmid. Concentration. *Probabilistic Methods for Algorithmic Discrete Mathematics (Algorithms and Combinatorics, 16)*. Springer, Berlin, 1998, pp. 195–248.
[**36**] A. Mohapatra and W. Ott. Memory loss for nonequilibrium open dynamical systems. *Discrete Contin. Dyn. Syst.* **34**(9) (2014), 3747–3759.
[**37**] P. Nándori, D. Szász and T. Varjú. A central limit theorem for time-dependent dynamical systems. *J. Stat. Phys.* **146** (2012), 1213–1220.
[**38**] W. Ott, M. Stenlund and L.-S. Young. Memory loss for time-dependent dynamical systems. *Math. Res. Lett.* **16** (2009), 463–475.
[**39**] I. Pinelis. An approach to inequalities for the distributions of infinite-dimensional martingales. *Probability in Banach Spaces (Proceedings of the Eighth International Conference, 8,* Brunswick, ME, 1991*) (Progress in Probability, 30)*. Birkhauser, Boston, MA, 1992, pp. 128–134.
[**40**] M. Stenlund. Non-stationary compositions of Anosov diffeomorphisms. *Nonlinearity* **24** (2011), 2991–3018.
[**41**] M. Stenlund, L.-S. Young and H. Zhang. Dispersing billiards with moving scatterers. *Comm. Math. Phys.* **322** (2013), 909–955.
[**42**] V. V. Yuriniskii. Exponential bounds for large deviations. *Theory Probab. Appl.* **19** (1974), 154–155.

https://doi.org/10.1017/etds.2015.19 Published online by Cambridge University Press