Commun. Math. Phys. 316, 843–889 (2012)
Digital Object Identifier (DOI) 10.1007/s00220-012-1596-7

Communications in
**Mathematical
Physics**

Optimal Concentration Inequalities for Dynamical Systems

Jean-René Chazottes[1], **Sébastien Gouëzel**[2]

[1] CPHT, CNRS UMR 7644, Ecole Polytechnique, 91128 Palaiseau Cedex, France.
E-mail: chazottes@cpht.polytechnique.fr
[2] IRMAR, CNRS UMR 6625, Université de Rennes 1, 35042 Rennes, France.
E-mail: sebastien.gouezel@univ-rennes1.fr

Received: 3 November 2011 / Accepted: 4 May 2012
Published online: 30 October 2012 –

Abstract: For dynamical systems modeled by a Young tower with exponential tails, we prove an exponential concentration inequality for all separately Lipschitz observables of n variables. When tails are polynomial, we prove polynomial concentration inequalities. Those inequalities are optimal. We give some applications of such inequalities to specific systems and specific observables.

Contents

1. Introduction

Let X be a metric space. A function K on X^n is separately Lipschitz if, for all i, there exists a constant $\mathrm{Lip}_i(K)$ with

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

for all points $x_1, \dots, x_n, x_i'$ in X.

Consider a stationary process $(Z_0, Z_1, \dots)$ taking values in X. We say that this process satisfies an exponential concentration inequality if there exists a constant C such that, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$, one has

$$\mathbb{E}(e^{K(Z_0,\dots,Z_{n-1})-\mathbb{E}(K(Z_0,\dots,Z_{n-1}))}) \leq e^{C\sum_{j=0}^{n-1} \operatorname{Lip}_j(K)^2}. \tag{1.1}$$

One should stress that this inequality is valid for all n (i.e., the constant C does not depend on the number of variables one is considering). An important consequence of such an inequality is a control on the deviation probabilities: for all $t > 0$,

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

This inequality follows from the inequality $\mathbb{P}(Y > t) \leq e^{-\lambda t}\mathbb{E}(e^{\lambda Y})$ $(\lambda > 0)$ with $Y = K(Z_0, \dots, Z_{n-1}) - \mathbb{E}(K(Z_0, \dots, Z_{n-1}))$, then we use inequality (1.1) and optimize over λ by taking $\lambda = t/(2C\sum_{j=0}^{n-1}\operatorname{Lip}_j(K)^2)$.

In some cases, it is not reasonable to hope for such an exponential inequality. One says that $(Z_0, Z_1, \dots)$ satisfies a polynomial concentration inequality with moment $Q \geq 2$ if there exists a constant C such that, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$, one has

$$\mathbb{E}\Big(|K(Z_0, \dots, Z_{n-1}) - \mathbb{E}(K(Z_0, \dots, Z_{n-1}))|^Q\Big) \leq C\left(\sum_{j=0}^{n-1}\operatorname{Lip}_j(K)^2\right)^{Q/2}. \tag{1.2}$$

An important consequence of such an inequality is a control on the deviation probabilities: for all $t > 0$,

$$\mathbb{P}(|K(Z_0, \dots, Z_{n-1}) - \mathbb{E}(K(Z_0, \dots, Z_{n-1}))| > t) \leq Ct^{-Q}\left(\sum_{j=0}^{n-1}\operatorname{Lip}_j(K)^2\right)^{Q/2}. \tag{1.3}$$

The inequality (1.3) readily follows from (1.2) and the Markov inequality. However, it is weaker in general. We will say that $(Z_0, Z_1, \dots)$ satisfies a weak L^Q concentration inequality if (1.3) holds for any separately Lipschitz function K.

For instance, if $Z_0, Z_1, \dots$ is an i.i.d. process, then it satisfies an exponential concentration inequality if Z_i is bounded [Led01, p. 68], a polynomial concentration inequality with moment $Q \geq 2$ if $Z_i \in L^Q$ [BBLM05], and a weak L^Q concentration inequality if $\mathbb{P}(|Z_i| > t) \leq Ct^{-Q}$ (while we could not locate a proper reference in the literature, this follows easily from classical martingale techniques and a weak L^Q Rosenthal-Burkholder inequality – see Theorem 6.3 below).

Our main goal in this article is to study processes coming from dynamical systems: we consider a map T on a metric space X, and an invariant probability measure μ. Under suitable assumptions, we wish to show that the process $(x, Tx, \dots)$ (where x is distributed following μ) satisfies concentration inequalities. Equivalently, we are interested in the concentration properties of the measure μ_n on X^n given by $\mathrm{d}\mu_n(x_0, \dots, x_{n-1}) = \mathrm{d}\mu(x_0)\delta_{x_1=Tx_0}\cdots\delta_{x_{n-1}=Tx_{n-2}}$. This is not a product measure but, if the map T is sufficiently mixing, one may expect that $T^k(x)$ is more or less independent of x if k is large, making the process $(x, Tx, \dots)$ look like an independent process to some extent.

Such questions have already been considered in the literature. In particular, [CMS02] proves that a (non-necessarily Markov) piecewise uniformly expanding map of the interval satisfies an exponential concentration inequality. Polynomial concentration inequalities (with moment 2, also called Devroye inequalities) have been proved in less expanding situations (exponential Young towers – including Hénon maps – in [CCS05a], intermittent map with parameter close enough to 0 in [CCRV09]). Our goal is to prove optimal concentration inequalities for the same kind of systems. In particular, we will prove that Young towers with exponential tails satisfy an exponential concentration inequality, and that in Young towers with polynomial tails one can get polynomial concentration with a moment directly related to the tails of the return time on the basis of the tower.

Concentration inequalities are a tool to bound systematically the fluctuations of 'complicated' observables of the form $K(x, Tx, \dots, T^{n-1}x)$. For instance, the function K can have a complicated analytic expression or can be implicitly defined (e.g. as an optimization problem). If we are able to get a good estimate of the Lipschitz constants, we can apply the concentration inequality we have at our disposal. Various examples of observables have been studied in [CMS02,CCS05b,CCRV09]. Since we establish here optimal concentration inequalities, this improves automatically the bounds previously available for these observables. We shall state explicitly some of the new results which can be obtained.

Outline of the article. The proofs we will use for different classes of systems are all based on classical martingale arguments. It is enlightening to explain them in the simplest possible situation, subshifts of finite type endowed with a Gibbs measure. We will do so in Sect. 2. The following 4 sections are devoted to proofs of concentration inequalities in various kinds of dynamical systems with a combinatorial nature, namely Young towers with exponential tails in Sect. 3, with polynomial tails in Sect. 4 (the invertible case is explained in Sect. 5), and with weak polynomial tails in Sect. 6. Several applications to concrete dynamical systems and to specific observables are described in Sect. 7. Finally, an appendix is devoted to the proof of a particularly technical lemma.

In this paper, the letter C denotes a constant that can vary from line to line (or even on a single line).

2. Subshifts of Finite Type

In this section, we describe a strategy to prove concentration inequalities. It is very classical, uses martingales, and was for instance implemented for dynamical systems in [CMS02] and for weakly dependent processes in [Rio00]. Our proofs for more complicated systems will also rely on this strategy. However, it is enlightening to explain it in the most simple situation, subshifts of finite type.

2.1. Unilateral subshifts of finite type. Let $X \subset \Sigma^{\mathbb{N}}$ be the state space of a topologically mixing one-sided subshift of finite type, with an invariant Gibbs measure μ, and the combinatorial distance $d(x, y) = \beta^{s(x,y)}$, where $\beta < 1$ is some fixed number and $s(x, y)$ is the separation time of x and y, i.e., the minimum number n such that $T^n x$ and $T^n y$ do not belong to the same element of the Markov partition. Writing $x = (x_0 x_1 \dots)$ and $y = (y_0 y_1 \dots)$, then $s(x, y) = \inf\{n : x_n \neq y_n\}$.

Theorem 2.1. *The system (X, T, μ) satisfies an exponential concentration inequality.*

Fix a separately Lipschitz function $K(x_0, \dots, x_{n-1})$. We consider it as a function on $X^{\mathbb{N}}$ depending only on the first n coordinates (therefore, we will write $\mathrm{Lip}_i(K) = 0$ for $i \geq n$). We endow $X^{\mathbb{N}}$ with the measure μ_∞ limit of the μ_N when $N \to \infty$. On $X^{\mathbb{N}}$, let $\mathcal{F}_p$ be the σ-algebra of events depending only on the coordinates $(x_j)_{j \geq p}$ (this is a decreasing sequence of σ-fields). We want to write the function K as a sum of reverse martingale differences with respect to this sequence. Therefore, let $K_p = \mathbb{E}(K|\mathcal{F}_p)$ and $D_p = K_p - K_{p+1}$. The function D_p is $\mathcal{F}_p$-measurable and $\mathbb{E}(D_p|\mathcal{F}_{p+1}) = 0$. Moreover, $K - \mathbb{E}(K) = \sum_{p \geq 0} D_p$.

The main point of the proof is to get a good bound on D_p:

Lemma 2.2. *There exist $C > 0$ and $\rho < 1$ such that, for any p, one has*

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

We then use the Hoeffding-Azuma inequality (see e.g. [MS86, p. 33] or [Led01, p. 68]), saying that for such a sum of martingale increments,

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

The Cauchy-Schwarz inequality gives

$$\begin{aligned} \left(\sum_{j=0}^{p} \rho^{p-j} \, \mathrm{Lip}_j(K) \right)^2 &\leq \left(\sum_{j=0}^{p} \rho^{p-j} \, \mathrm{Lip}_j(K)^2 \right) \cdot \left(\sum_{j=0}^{p} \rho^{p-j} \right) \\ &\leq C \sum_{j=0}^{p} \rho^{p-j} \, \mathrm{Lip}_j(K)^2. \end{aligned}$$

Summing over p, we get $\sum_{p=0}^{P-1} \sup |D_p|^2 \leq C \sum_j \mathrm{Lip}_j(K)^2$. Using the Hoeffding-Azuma inequality at a fixed index P, and then letting P tend to infinity, we get $\mathbb{E}(e^{\sum D_p}) \leq e^{C \sum \mathrm{Lip}_j(K)^2}$, which is the desired exponential concentration inequality since $\sum D_p = K - \mathbb{E}(K)$.

It remains to prove Lemma 2.2. Let g denote the inverse of the jacobian of T, and $g^{(k)}$ the inverse of the jacobian of T^k. Let $\mathcal{L}$ denote the transfer operator associated to the map T, defined by duality by $\int u \cdot v \circ T \, \mathrm{d}\mu = \int \mathcal{L}u \cdot v \, \mathrm{d}\mu$. It can be written as $\mathcal{L}u(x) = \sum_{Ty=x} g(y)u(y)$. In the same way, $\mathcal{L}^k u(x) = \sum_{T^k y = x} g^{(k)}(y)u(y)$. One can define a Markov chain by jumping from a point x to one of its preimages y with the probability $g(y)$, then $\mathcal{L}$ is simply the Markov operator corresponding to this Markov chain. In particular,

$$\begin{aligned} K_p(x_p, x_{p+1}, \dots) &= \mathbb{E}(K|\mathcal{F}_p)(x_p, x_{p+1}, \dots) = \mathbb{E}(K(X_0, \dots, X_{p-1}, x_p, \dots)|X_p = x_p) \\ &= \sum_{T^p(y) = x_p} g^{(p)}(y) K(y, \dots, T^{p-1}y, x_p, \dots). \end{aligned}$$

To prove that D_p is bounded, i.e., K_p is close to K_{p+1}, one should show that this quantity does not depend too much on x_p. The preimages of x_p under T^p equidistribute in the space, therefore one should be able to show that K_p is close to an integral quantity. This is done in the following lemma.

Lemma 2.3. *We have*

$$\left|K_p(x_p,\dots) - \int K(y,\dots,T^{p-1}y,x_p,\dots)\,\mathrm{d}\mu(y)\right| \le C\sum_{j=0}^{p-1}\mathrm{Lip}_j(K)\rho^{p-1-j},$$

where $C > 0$ *and* $\rho < 1$ *only depend on* (X,T).

This lemma implies in particular that $K_p(x_p,x_{p+1},\dots) - K_p(x'_p,x_{p+1},\dots)$ is bounded by $C\sum_{j=0}^{p}\mathrm{Lip}_j(K)\rho^{p-j}$. Averaging over the preimages x'_p of x_{p+1}, we get the same bound for $D_p(x_p,x_{p+1},\dots)$, proving Lemma 2.2.

Proof of Lemma 2.3. The equidistribution of the Markov chain starting from x_p is formulated most conveniently in terms of the transfer operators, which act on functions of one variable. Therefore, we will eliminate the variables $x_0,\dots,x_{p-1}$ one after the other. Let us fix a point x_* in X, we decompose K_p as

$$\begin{aligned}K_p(x_p,\dots) = \sum_{i=0}^{p-1}\sum_{T^p(y)=x_p} g^{(p)}(y)(K(y,\dots,T^iy,x_*,\dots,x_*,x_p,\dots)\\ - K(y,\dots,T^{i-1}y,x_*,\dots,x_*,x_p,\dots))\\ +K(x_*,\dots,x_*,x_p,\dots).\end{aligned}$$

For fixed i, we may group together those points $y \in T^{-p}(x_p)$ that have the same image under T^i, splitting the sum $\sum_{T^p(y)=x_p}$ as $\sum_{T^{p-i}(z)=x_p}\sum_{T^i(y)=z}$. Since the jacobian is multiplicative, one has $g^{(p)}(y) = g^{(i)}(y)g^{(p-i)}(z)$. Let us define a function

$$\begin{aligned}f_i(z) &= \sum_{T^iy=z} g^{(i)}(y)(K(y,\dots,T^iy,x_*,\dots,x_*,x_p,\dots)\\ &\qquad - K(y,\dots,T^{i-1}y,x_*,\dots,x_*,x_p,\dots))\\ &= \sum_{T^iy=z} g^{(i)}(y)H(y,\dots,T^iy).\end{aligned} \tag{2.1}$$

Denoting by $\mathcal{L}$ the transfer operator (which satisfies $\mathcal{L}^k f(x) = \sum_{T^k(z)=x} g^{(k)}(z)f(z)$), we obtain

$$K_p(x_p,\dots) = \sum_{i=0}^{p-1}\mathcal{L}^{p-i}f_i(x_p) + K(x_*,\dots,x_*,x_p,\dots).$$

The function H is bounded by $\mathrm{Lip}_i(K)$, hence $|f_i| \le C\,\mathrm{Lip}_i(K)$ (since $\sum_{T^iy=z} g^{(i)}(y) = 1$ by invariance of the measure). To estimate the Lipschitz norm of f_i, we write

$$\begin{aligned}f_i(z) - f_i(z') &= \sum(g^{(i)}(y) - g^{(i)}(y'))H(y,\dots,T^iy)\\ &\quad+\sum g^{(i)}(y')(H(y,\dots,T^iy) - H(y',\dots,T^iy')),\end{aligned} \tag{2.2}$$

where z and z' are two points in the same partition element, and their respective preimages y, y' are paired according to the cylinder of length i they belong to. A distortion control gives $|g^{(i)}(y) - g^{(i)}(y')| \le Cg^{(i)}(y)d(z,z')$, hence the first sum is bounded by $C\,\mathrm{Lip}_i(K)d(z,z')$. For the second sum, substituting successively each T^jy with T^jy', we have

$$
\begin{aligned}
|H(y, \dots, T^i y) - H(y', \dots, T^i y')| &\le 2 \sum_{j=0}^{i} \mathrm{Lip}_j(K) d(T^j y, T^j y') \\
&\le 2 \sum_{j=0}^{i} \mathrm{Lip}_j(K) \beta^{i-j} d(z, z').
\end{aligned}
$$

Summing over the different preimages of z, we deduce that the Lipschitz norm of f_i is bounded by $C \sum_{j=0}^{i} \mathrm{Lip}_j(K)\beta^{i-j}$.

Let $\mathcal{C}$ be the space of Lipschitz functions on X, with its canonical norm $\|f\|_{\mathcal{C}} = \sup|f| + \mathrm{Lip}(f)$. The operator $\mathcal{L}$ has a spectral gap on $\mathcal{C}$: there exist $C > 0$ and $\rho < 1$ such that $\left\|\mathcal{L}^k f - \int f \,\mathrm{d}\mu\right\|_{\mathcal{C}} \le C\rho^k \|f\|_{\mathcal{C}}$. We get $\left\|\mathcal{L}^{p-i} f_i - \int f_i \,\mathrm{d}\mu\right\|_{\mathcal{C}} \le C\rho^{p-i} \sum_{j=0}^{i} \mathrm{Lip}_j(K)\beta^{i-j}$. This bound in $\mathcal{C}$ implies in particular a bound for the supremum. Increasing ρ if necessary, we can assume $\rho \ge \beta$. Summing those bounds, one obtains

$$
\begin{aligned}
&\left| K_p(x_p, \dots) - \sum_{i=0}^{p-1} \int f_i \,\mathrm{d}\mu - K(x_*, \dots, x_*, x_p, \dots) \right| \\
&\quad \le C \sum_{i=0}^{p-1} \rho^{p-i} \sum_{j=0}^{i} \mathrm{Lip}_j(K)\rho^{i-j} \le C \sum_{j=0}^{p-1} \mathrm{Lip}_j(K)\rho^{p-j}(p-j) \\
&\quad \le C' \sum_{j=0}^{p-1} \mathrm{Lip}_j(K)(\rho')^{p-j},
\end{aligned}
$$

for any $\rho' \in (\rho, 1)$.

Finally, when one computes the sum of the integrals of f_i, there are again cancelations, leaving only $\int K(y, \dots, T^{p-1}y, x_p, \dots) \,\mathrm{d}\mu(y)$. □

2.2. Bilateral subshifts of finite type. We consider now $X_{\mathbb{Z}} \subset \Sigma^{\mathbb{Z}}$ the state space of a topologically mixing bilateral subshift of finite type, together with an invariant Gibbs measure $\mu_{\mathbb{Z}}$. For two points $x = (\dots x_{-1}x_0x_1 \dots)$ and $y = (\dots y_{-1}y_0y_1 \dots)$ in $X_{\mathbb{Z}}$, let $s_{\mathbb{Z}}$ be their bilateral separation time, i.e., $\inf\{|n| : x_n \neq y_n\}$, and define a distance $d_{\mathbb{Z}}(x, y) = \beta^{s_{\mathbb{Z}}(x,y)}$ for some $\beta < 1$. We denote a function on $X_{\mathbb{Z}}^n$ by $K_{\mathbb{Z}}(x_0, \dots, x_{n-1})$, to emphasize the dependence both on the past and the future.

Theorem 2.4. *The system $(X_{\mathbb{Z}}, T, \mu_{\mathbb{Z}})$ satisfies an exponential concentration inequality.*

This is stronger than Theorem 2.1, which proves this statement for functions $K_{\mathbb{Z}}(x_0, \dots, x_{n-1})$ depending only on the future $(x_i)_0^\infty$ of each variable. We will deduce Theorem 2.4 from this statement by an approximation argument, by sending everything far away in the future.

Proof. Let us first assume that $X_{\mathbb{Z}}$ is the full shift. We fix a function $K_{\mathbb{Z}}(x_0, \dots, x_{n-1})$ depending both on the past and future of the variables. For $N \in \mathbb{N}$, we define $K_N(x_0, \dots, x_{n+N-1}) = K_{\mathbb{Z}}(x_N, \dots, x_{n+N-1})$. Thanks to the invariance of the measure, it is equivalent to prove concentration inequalities for $K_{\mathbb{Z}}$ or K_N.

Let us now define a function $\Phi_N : X_{\mathbb{Z}}^{n+N} \to X_{\mathbb{Z}}^{n+N}$ depending only on the future of the variables, and let us write $\tilde{K}_N = K_N \circ \Phi_N$. Since this function only depends on the future, Theorem 2.1 applies to it.

We set $\Phi_N(x_0, \dots, x_{n+N-1}) = (y_0, \dots, y_{n+N-1})$, where the y_i are defined inductively as follows. First, let us choose an arbitrary past $(p)_{-\infty}^{-1}$, and let $y_0 = ((p)_{-\infty}^{-1}, (x_0)_0^{\infty})$: it only depends on the future of x_0. If $y_0, \dots, y_{i-1}$ are already defined, we let $y_i = ((y_{i-1})_{-\infty}^{0}, (x_i)_0^{\infty})$. In other words,

$$y_i = ((p)_{-\infty}^{-1}, (x_0)_0, (x_1)_0, \dots, (x_{i-1})_0, (x_i)_0^{\infty}), \tag{2.3}$$

with an origin laid on $(x_i)_0$. This defines the function Φ_N, only depending on the future of the points.

Let us study the Lipschitz constants of $\tilde{K}_N = K_N \circ \Phi_N$. If we fix x_j for $j \neq i$ and vary x_i, then we change y_j for $j \geq i$, at its coordinate with index $-(j-i)$. Therefore,

$$\mathrm{Lip}_i(\tilde{K}_N) \leq \sum_{j \geq i} \mathrm{Lip}_j(K_N)\beta^{j-i}.$$

With Cauchy-Schwarz inequality, we get $\sum \mathrm{Lip}_i(\tilde{K}_N)^2 \leq C \sum \mathrm{Lip}_i(K_N)^2 = C \sum \mathrm{Lip}_i(K_{\mathbb{Z}})^2$, for some constant C. Applying Theorem 2.1 to $\tilde{K}_N$ and changing variables by $x' = T^N x$, we obtain

$$\int e^{\tilde{K}_N(T^{-N}x', \dots, T^{-1}x', x', \dots, T^{n-1}x')} \, \mathrm{d}\mu_{\mathbb{Z}}(x')$$
$$\leq e^{\int \tilde{K}_N(T^{-N}x', \dots, T^{-1}x', x', \dots, T^{n-1}x') \, \mathrm{d}\mu_{\mathbb{Z}}(x')} e^{C \sum_{i=0}^{n-1} \mathrm{Lip}_i(K_{\mathbb{Z}})^2}.$$

By construction, the function $\tilde{K}_N(T^{-N}x', \dots, T^{-1}x', x', \dots, T^{n-1}x')$ converges to $K_{\mathbb{Z}}(x', \dots, T^{n-1}x')$ when N tends to infinity. Hence, the previous equation gives the desired exponential concentration.

When $X_{\mathbb{Z}}$ is not the full shift, there is an additional difficulty: one can not define Φ_N as above, since a point defined in (2.3) might use forbidden transitions. We should therefore modify the definition of Φ_N as follows. For any symbol a of the alphabet, we fix a legal past $p(a)$ of a. We define $\Phi_N(x_0, \dots, x_{N+n-1}) = (y_0, \dots, y_{N+n-1})$ by $y_0 = (p((x_0)_0), (x_0)_0^{\infty})$ (this point is admissible). Then, if the transition from $(x_{i-1})_0$ to $(x_i)_0$ is permitted, we let $y_i = ((y_{i-1})_{-\infty}^{0}, (x_i)_0^{\infty})$, and otherwise we let $y_i = (p((x_i)_0), (x_i)_0^{\infty})$. Therefore, the points y_i only use permitted transitions. The rest of the argument goes through without modification. □

3. Uniform Young Towers with Exponential Tails

There are two different definitions of Young towers, given respectively in [You98] and [You99]. The difference is on the definition of the separation time: in the first definition, one considers that the dynamics is expanding at every iteration, while in the second definition one considers that the dynamics is expanding only when one returns to the basis of the tower. Therefore, there is less expansion with the second definition than with the first one, making it more difficult to handle. We will say that Young towers in the first sense are uniform, while Young towers in the second sense are non-uniform. In this section, we work with the (easier) first definition, which turns out to be the most interesting when dealing with exponential tails. Here is the formal definition of a uniform Young tower: it is a space Δ satisfying the following properties:

(1) This space is partitioned into subsets $\Delta_{\alpha,\ell}$ (for $\alpha \in \mathbb{N}$ and $\ell \in [0, \phi(\alpha) - 1]$, where ϕ is an integer-valued return time function). The dynamics sends bijectively $\Delta_{\alpha,\ell}$ on $\Delta_{\alpha,\ell+1}$ if $\ell < \phi(\alpha) - 1$, and $\Delta_{\alpha,\phi(\alpha)-1}$ on $\Delta_0 := \bigcup_\alpha \Delta_{\alpha,0}$.
(2) The distance is given by $d(x, y) = \beta^{s(x,y)}$, where $\beta < 1$ and $s(x, y)$ is the separation time for the whole dynamics, i.e., the first n such that $T^n x$ and $T^n y$ are not in the same element of the partition.
(3) There is an invariant probability measure μ such that the inverse g of its jacobian satisfies $|g(x)/g(y) - 1| \leq Cd(Tx, Ty)$ for any x and y in the same element of the partition.
(4) We have $\gcd(\phi(\alpha) : \alpha \in \mathbb{N}) = 1$ (i.e., the tower is aperiodic).

When the return time function ϕ has exponential tails, i.e., there exists $c_0 > 0$ with $\int_{\Delta_0} e^{c_0\phi} \,\mathrm{d}\mu < \infty$, we say that the tower has exponential tails. We will write $h(x) = \ell$ if $x \in \Delta_{\alpha,\ell}$: this is the height of the point in the tower. For $x \in \Delta$, we will also denote by πx its projection in the basis, i.e., the unique point $y \in \Delta_0$ such that $T^{h(x)}(y) = x$.

Theorem 3.1. *Let (Δ, T, μ) be a uniform Young tower with exponential tails. It satisfies an exponential concentration inequality: there exists $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$,*

$$\int e^{K(x,Tx,\dots,T^{n-1}x)} \,\mathrm{d}\mu(x) \leq e^{\int K(x,\dots,T^{n-1}x)\,\mathrm{d}\mu(x)} e^{C\sum_{i=0}^{n-1} \mathrm{Lip}_i(K)^2}. \tag{3.1}$$

Let us first remark that, for any $\epsilon_0 > 0$, it is sufficient to prove the theorem for functions K such that $\mathrm{Lip}_i(K) \leq \epsilon_0$ for all i. Assume indeed that this is the case, and let us prove the general case. Let $K(x_0, \dots, x_{n-1})$ be a separately Lipschitz function. Let us fix an arbitrary point x_* in Δ. To any $(x_0, \dots, x_{n-1})$, we associate $(y_0, \dots, y_{n-1})$ by $y_i = x_i$ if $\mathrm{Lip}_i(K) \leq \epsilon_0$ and $y_i = x_*$ otherwise. The function $\tilde{K}(x_0, \dots, x_{n-1}) = K(y_0, \dots, y_{n-1})$ satisfies $\mathrm{Lip}_i(\tilde{K}) \leq \epsilon_0$ for all i. Moreover,

$$|K - \tilde{K}| \leq \sum_i \mathrm{Lip}_i(K) \mathbb{1}(\mathrm{Lip}_i(K) > \epsilon_0) \leq \sum_i \mathrm{Lip}_i(K)^2/\epsilon_0.$$

Therefore, the inequality (3.1) for $\tilde{K}$ readily implies the same inequality for K, with a different constant $C' = C + 2/\epsilon_0$.

Let us now fix a suitable ϵ_0 (the precise conditions will be given in the proof of Lemma 3.3), and let us consider a function K with $\mathrm{Lip}_i(K) \leq \epsilon_0$ for all i. To prove the exponential concentration inequality, we follow the strategy of Sect. 2. Let $K_p(x_p, \dots) = \mathbb{E}(K|\mathcal{F}_p)(x_p, \dots)$; the first step is to prove an analogue of Lemma 2.3. Since the transfer operator has a spectral gap on a suitable space of functions, as shown by Young in [You98], we can easily mimic the proof of this lemma.

Lemma 3.2. *For all $x_p \in \Delta_0$,*

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

where $C > 0$ and $\rho < 1$ only depend on Δ.

The main difference with the subshift case is that this bound is only valid for $h(x_p) = 0$. It is of course false if $h(x_p)$ is large, since there is no averaging mechanism in this case.

Proof. As in the proof of Lemma 2.3, we write

$$K_p(x_p,\dots) = \sum_{i=0}^{p-1} \mathcal{L}^{p-i} f_i(x_p) + K(x_*,\dots,x_*,x_p,\dots),$$

where the function f_i is bounded by $\mathrm{Lip}_i(K)$, and the Lipschitz norm of f_i on any partition element is at most $C\sum_{j=0}^{i}\mathrm{Lip}_j(K)\rho^{i-j}$ for some $\rho<1$.

Let $\mathcal{C}$ be the space of functions on Δ such that $|f(x)| \leq Ce^{\epsilon h(x)}$ and $|f(x)-f(y)| \leq Cd(x,y)e^{\epsilon h(x)}$ for all x, y in the same partition element. Young proves in [You98] that, if ϵ is small enough, then $\mathcal{L}$ has a spectral gap on $\mathcal{C}$: there exist $C>0$ and $\rho<1$ such that $\left\|\mathcal{L}^k f - \int f\,\mathrm{d}\mu\right\|_{\mathcal{C}} \leq C\rho^k \|f\|_{\mathcal{C}}$.

We obtain $\left\|\mathcal{L}^{p-i} f_i - \int f_i\,\mathrm{d}\mu\right\|_{\mathcal{C}} \leq C\rho^{p-i}\sum_{j=0}^{i}\mathrm{Lip}_j(K)\rho^{i-j}$. This bound in $\mathcal{C}$ gives in particular a bound on the supremum for points at height 0, and in particular at the point x_p. Summing those bounds over i, we get the desired result exactly as in the proof of Lemma 2.3. □

The next step of the proof is the following lemma. It is here that the Lipschitz constants $\mathrm{Lip}_j(K)$ should all be bounded by ϵ_0. As before, let $K_p = \mathbb{E}(K|\mathcal{F}_p)$, and $D_p = K_p - K_{p+1}$.

Lemma 3.3. *There exist $\epsilon_0 > 0$, $C_1 > 0$ and $\rho < 1$ such that any function $K(x_0,\dots,x_{n-1})$ with $\mathrm{Lip}_j(K) \leq \epsilon_0$ for all j satisfies, for any p,*

$$\mathbb{E}(e^{D_p}|\mathcal{F}_{p+1})(x_{p+1},\dots) \leq e^{C_1\sum_{j=0}^{p}\mathrm{Lip}_j(K)^2\rho^{p-j}}.$$

Proof. If the height of x_{p+1} is positive, then this point has a unique preimage y, and $D_p(y,x_{p+1},\dots) = 0$. Therefore, $\mathbb{E}(e^{D_p}|\mathcal{F}_{p+1})(x_{p+1},\dots) = 1$ and the estimate is trivial.

Assume now that $h(x_{p+1}) = 0$. Let us denote by $\{z_\alpha\}$ the preimages of x_{p+1} under T (with $z_\alpha \in \Delta_{\alpha,\phi(\alpha)-1}$). Let $A(z) = D_p(z,x_{p+1},\dots)$, we have $\mathbb{E}(e^{D_p}|\mathcal{F}_{p+1})(x_{p+1},\dots) = \sum g(z_\alpha)e^{A(z_\alpha)}$.

Fix a point $z = z_\alpha$, with height $h \geq 0$. If $h \leq p$, consider the projection πz of z in the basis of the tower. Since $K_p(z,\dots) = K_{p-h}(\pi z,\dots,z,\dots)$, Lemma 3.2 shows that $K_p(z,\dots)$ is equal to $\int K(y,\dots,T^{p-h}y,\pi z,\dots)\,\mathrm{d}\mu(y)$ up to $C\sum_{j=0}^{p-h-1}\mathrm{Lip}_j(K)\rho^{p-h-1-j}$. Up to an additional error $\sum_{j=p-h}^{p}\mathrm{Lip}_j(K)$, this is equal to $\int K(y,\dots,T^p y,x_{p+1},\dots)\,\mathrm{d}\mu(y)$. Applying again Lemma 3.2 (but to the point x_{p+1}), we obtain

$$\begin{aligned}|A(z)| &= |K_p(z,x_{p+1},\dots) - K_{p+1}(x_{p+1},\dots)|\\ &\leq C\sum_{j<p-h}\mathrm{Lip}_j(K)\rho^{p-h-j} + \sum_{j=p-h}^{p}\mathrm{Lip}_j(K).\end{aligned}$$

This estimate is also trivially true if $h > p$ (by convention, one sets $\mathrm{Lip}_j(K) = 0$ for $j<0$). In particular, since $\sup\mathrm{Lip}_j(K) \leq \epsilon_0$, we always get $|A(z)| \leq C_0(h+1)\epsilon_0$ for some $C_0 > 0$ (independent of the value of ϵ_0). Using the inequality $(x_1+\dots+x_k)^2 \leq k\sum x_i^2$, we get

$$
\begin{aligned}
|A(z)|^2 &\le C\left(\sum_{j<p-h}\operatorname{Lip}_j(K)\rho^{p-h-j}\right)^2 + C(h+1)\sum_{j=p-h}^{p}\operatorname{Lip}_j(K)^2 \\
&\le C\sum_{j<p-h}\operatorname{Lip}_j(K)^2\rho^{p-h-j} + C(h+1)\sum_{j=p-h}^{p}\operatorname{Lip}_j(K)^2, \qquad (3.2)
\end{aligned}
$$

where we used Cauchy-Schwarz inequality in the last inequality.

The function A satisfies a neat bound on points z_α with small height, but it is unbounded on points with large height. Therefore, the Hoeffding-Azuma inequality does not apply (contrary to the subshift of finite type case). While there are certainly exponential inequalities in the literature that can handle this situation, it is simpler to reprove everything since we are not interested in good constants.

We have $|e^A - 1 - A| \le A^2 e^{|A|}$ for any real number A. Therefore,

$$
\left|\sum_\alpha g(z_\alpha)(e^{A(z_\alpha)} - 1 - A(z_\alpha))\right| \le \sum g(z_\alpha)A(z_\alpha)^2 e^{|A(z_\alpha)|}.
$$

In the right-hand side, $g(z_\alpha) \le C\mu(\Delta_{\alpha,0})$ by bounded distortion, and $|A(z_\alpha)| \le C_0\epsilon_0(1+\phi(\alpha))$ as we explained above. Together with (3.2), we get

$$
\begin{aligned}
&\sum g(z_\alpha)A(z_\alpha)^2 e^{|A(z_\alpha)|} \\
&\quad\le C\sum_{h\ge 0}\mu(\phi = h)e^{C_0\epsilon_0 h}\left(\sum_{j<p-h}\operatorname{Lip}_j(K)^2\rho^{p-h-j} + (h+1)\sum_{j=p-h}^{p}\operatorname{Lip}_j(K)^2\right).
\end{aligned}
$$

Since the tower has exponential tails, we have $\mu(\phi = h) \le \rho_0^h$ for some $\rho_0 < 1$. If ϵ_0 is small enough, we get $\mu(\phi = h)e^{C_0\epsilon_0 h} \le \rho_1^h$ for some $\rho_1 < 1$. Therefore, in the previous bound, the coefficient of $\operatorname{Lip}_j(K)^2$ is at most

$$
\sum_{h<p-j}\rho_1^h\rho^{p-h-j} + \sum_{h\ge p-j}(h+1)\rho_1^h \le (p-j)\rho_2^{p-j} + \rho_2^{p-j},
$$

for some $\rho_2 < 1$. This is bounded by $C\rho^{p-j}$ for some $\rho < 1$. Hence, we have proved that

$$
\left|\sum_\alpha g(z_\alpha)(e^{A(z_\alpha)} - 1 - A(z_\alpha))\right| \le C\sum_{j\le p}\rho^{p-j}\operatorname{Lip}_j(K)^2.
$$

Since $\sum g(z_\alpha) = 1$ and $\sum g(z_\alpha)A(z_\alpha) = 0$, the left hand side is equal to $\left|\sum g(z_\alpha)e^{A(z_\alpha)} - 1\right|$. Finally,

$$
\begin{aligned}
|\mathbb{E}(e^{D_p}|\mathcal{F}_{p+1})(x_{p+1},\dots)| &= \left|\sum g(z_\alpha)e^{A(z_\alpha)}\right| \le 1 + C\sum_{j\le p}\rho^{p-j}\operatorname{Lip}_j(K)^2 \\
&\le e^{C\sum_{j\le p}\rho^{p-j}\operatorname{Lip}_j(K)^2}.
\end{aligned}
$$

This concludes the proof. □

Proof of Theorem 3.1. Consider a function K with $\mathrm{Lip}_j(K) \leq \epsilon_0$ for all j. Using inductively Lemma 3.3, we get for any P,

$$\mathbb{E}\left(e^{\sum_{p=0}^{P-1} D_p} | \mathcal{F}_P\right) \leq e^{C_1 \sum_{p=0}^{P-1} \sum_{j=0}^{p} \mathrm{Lip}_j(K)^2 \rho^{p-j}} \leq e^{C \sum \mathrm{Lip}_j(K)^2}.$$

Since $\sum_{p=0}^{P-1} D_p$ converges to $K - \mathbb{E}(K)$ when P tends to infinity, we obtain $\mathbb{E}(e^{K-\mathbb{E}(K)}) \leq e^{C \sum \mathrm{Lip}_j(K)^2}$. This proves the exponential concentration inequality in this case. The general case follows, as we explained after the statement of the theorem. □

The exponential concentration inequalities for uniform Young towers with exponential tails easily extends to invertible situations, as follows. Consider $T_{\mathbb{Z}} : \Delta_{\mathbb{Z}} \to \Delta_{\mathbb{Z}}$ the natural extension of such a Young tower, with bilateral separation time $s_{\mathbb{Z}}$, and distance $d_{\mathbb{Z}}(x, y) = \beta^{s_{\mathbb{Z}}(x,y)}$ for some $\beta < 1$.

Theorem 3.4. *The transformation $T_{\mathbb{Z}}$ satisfies an exponential concentration inequality.*

The proof is exactly the same as the proof of Theorem 2.4, exploiting the result for the non-invertible transformation.

4. Non-uniform Young Towers with Polynomial Tails

In this section, we consider Young towers in the sense of [You99], i.e., non-uniform Young towers. The combinatorial definition is the same as in Sect. 3, the difference is on the definition of the separation time (and therefore of the distance) as follows. Let Δ_0 be the basis of the tower, let $T_0 : \Delta_0 \to \Delta_0$ be the induced map on Δ_0 (i.e., $T_0(x) = T^{\phi(x)}(x)$, where $\phi(x)$ is the return time of x to Δ_0). For $x, y \in \Delta_0$, let $s(x, y)$ be the smallest integer n such that $T_0^n(x)$ and $T_0^n(y)$ are not in the same partition element. This separation time is extended to Δ as follows. For $x, y \in \Delta$, let $s(x, y) = s(\pi x, \pi y)$ if x and y are in the same partition element, and $s(x, y) = 0$ otherwise. In other words, $s(x, y)$ is the number of returns to the basis before the trajectories of x and y separate. Finally, the new distance is $d(x, y) = \beta^{s(x,y)}$ for some $\beta < 1$.

Intuitively, we are now considering maps that are expanding only when one returns to the basis, and can be isometries between successive returns, while the maps of Sect. 3 are always expanding. The setting is not uniformly expanding any more, rather non-uniformly expanding. For instance, intermittent maps can be modeled using non-uniform Young towers.

If the tails are not exponential any more, one can not hope to get exponential concentration inequalities. If the tails have a moment of order $q \geq 2$, then the moments of order $2q - 2$ of Birkhoff sums are controlled, and this is optimal [MN08, Thm. 3.1]. Our goal in this section is to generalize this result to a concentration inequality (with the same optimal moment).

Theorem 4.1. *Let $T : \Delta \to \Delta$ be a non-uniform Young tower. Assume that, for some $q \geq 2$, $\sum \phi(\alpha)^q \mu(\Delta_{\alpha,0}) < \infty$. Then T satisfies a polynomial concentration inequality with moment $2q - 2$, i.e., there exists a constant $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $K(x_0, \ldots, x_{n-1})$,*

$$\int \left| K(x, \dots, T^{n-1}x) - \int K(y, \dots, T^{n-1}y) \,\mathrm{d}\mu(y) \right|^{2q-2} \mathrm{d}\mu(x)$$
$$\leq C \left(\sum_j \mathrm{Lip}_j(K)^2 \right)^{q-1}.$$

The proof is considerably more difficult than the arguments in the previous section (and also than the arguments of [MN08] since the main inequality these arguments rely on, due to Rio, is of no help in our situation). The general strategy is the same as in the previous sections: decompose $K - \mathbb{E}(K)$ as $\sum D_p$, where D_p is a martingale difference sequence, obtain good estimates on D_p, and then apply a martingale inequality (in our case, the Rosenthal-Burkholder inequality) to obtain a bound on $K - \mathbb{E}(K)$. The difficulty comes from the non-uniform expansion of the map: instead of a uniformly decaying geometric series as in the previous sections, our estimates will be non-uniform, quantified by the number of visits to the basis in a definite amount of time.

The rest of this section is devoted to the proof of Theorem 4.1. In particular, we will always assume that Δ is a non-uniform Young tower satisfying $\sum \phi(\alpha)^q \mu(\Delta_{\alpha,0}) < \infty$ for some $q \geq 2$.

Remark 4.2. The arguments below also give an exponential concentration inequality in non-uniform Young towers with exponential tails, thereby strengthening Theorem 3.1. Since most interesting Young towers with exponential tails are uniform, we will not give further details in this direction.

4.1. Notations. As usual, the letter C denotes a constant that may change from one occurrence to the next. Let us also introduce a similar notation for sequences. For $Q \geq 0$, we will write $c_n^{(Q)}$ for a sequence of nonnegative numbers such that $\sum n^Q c_n^{(Q)} < \infty$, and we will allow this sequence to change from one line to the other (or even on the same line). We will also write $d_n^{(Q)}$ for a generic nonincreasing sequence with $\sum n^Q d_n^{(Q)} < \infty$.

If u_n and v_n are sequences, their convolution $u \star v$ is given by $(u \star v)_n = \sum_{k=0}^n u_k v_{n-k}$. One easily checks that, for $Q, Q' \geq 0$,

$$(c^{(Q)} \star c^{(Q')})_n \leq c_n^{(\min(Q,Q'))}. \tag{4.1}$$

Following the above convention, this statement should be understood as follows: if two sequences u and v satisfy, respectively, $\sum n^Q u_n < \infty$ and $\sum n^{Q'} v_n < \infty$, then $w = u \star v$ satisfies $\sum n^{\min(Q,Q')} w_n < \infty$. Indeed, letting $Q'' = \min(Q, Q')$,

$$\sum n^{Q''} w_n = \sum_{k,\ell} (k+\ell)^{Q''} u_k v_\ell \leq \sum_{k,\ell} (k+1)^{Q''} (\ell+1)^{Q''} u_k v_\ell$$
$$\leq \left(\sum (k+1)^Q u_k \right) \cdot \left(\sum (\ell+1)^{Q'} v_\ell \right) < \infty.$$

We also have for $Q \geq 1$,

$$\sum_{k=n}^{\infty} c_k^{(Q)} \leq d_n^{(Q-1)}. \tag{4.2}$$

Indeed,

$$\sum n^{Q-1} \sum_{k=n}^{\infty} c_k^{(Q)} = \sum_k \left(\sum_{n=0}^{k} n^{Q-1} \right) c_k^{(Q)} \leq \sum_k C k^Q c_k^{(Q)} < \infty,$$

and the sequence $\sum_{k=n}^{\infty} c_k^{(Q)}$ is nonincreasing.

4.2. Renewal sequences of operators, estimates on the returns to the basis. An important tool for our study will be renewal sequences of operators, as developed by Sarig and Gouëzel [Sar02,Gou04b,Gou04c], that we will now quickly describe.

Consider a function f; we wish to understand $\mathcal{L}^n f(x) = \sum_{T^n y = x} g^{(n)}(y) f(y)$ for $x \in \Delta_0$. For a preimage y of x under T^n, we can consider its first entrance into Δ_0, and then its successive returns to Δ_0. We obtain a decomposition

$$1_{\Delta_0} \mathcal{L}^n = \sum_{k+b=n} T_k B_b, \tag{4.3}$$

where T_k takes the successive returns to Δ_0 (during time k) into account, and B_b deals with the part of the trajectory outside Δ_0. Formally, for $x \in \Delta_0, T_k f(x) = \sum g^{(k)}(y) f(y)$, where the sum is restricted to those y such that $T^k y = x$ and $y \in \Delta_0$. The operator B_b, in turn, is given on Δ_0 by $B_b f(x) = \sum g^{(b)}(y) f(y)$, where the sum is restricted to those y with $T^b y = x$ and $y, \dots, T^{b-1} y \notin \Delta_0$.

The operators B_b are essentially trivial to understand, their behavior being controlled by the tails of the return time function ϕ. On the other hand, the operators T_k embody most of the dynamics of the transformation. To understand them, we introduce yet another operator R_j considering only the first return to Δ_0 at time j, i.e., $R_j f(x) = \sum g^{(j)}(y) f(y)$, where the sum is restricted to those y such that $T^j y = x$ and $y \in \Delta_0, Ty, \dots, T^{j-1} y \notin \Delta_0$. Splitting a trajectory into its successive excursions outside of Δ_0, one obtains

$$T_k = \sum_{\ell \geq 1} \sum_{j_1 + \dots + j_\ell = k} R_{j_1} \cdots R_{j_\ell}.$$

Formally, this equation can be written as

$$\sum T_k z^k = (I - \sum R_j z^j)^{-1}. \tag{4.4}$$

In fact, the series defined in this equation are holomorphic for $|z| < 1$ (as operators acting on the space $\mathcal{C}$ of Lipschitz functions on Δ_0) and this equality is a true equality between holomorphic functions. Moreover, the spectral radius of $\sum R_j z^j$ is at most 1 for $|z| \leq 1$.

A powerful way to use the previous equality is Banach algebra techniques. Simple examples of Banach algebras are given by Banach spaces $\mathcal{B}$ of sequences c_n such that, if $(c_n)_{n\in\mathbb{N}} \in \mathcal{B}$ and $(c'_n)_{n\in\mathbb{N}} \in \mathcal{B}$, then their convolution $c \star c'$ still belongs to $\mathcal{B}$. For instance, this is the case of sequences with a moment of order $Q \geq 0$ (by (4.1)), or of sequences satisfying $c_n = O(1/n^Q)$ for some $Q > 1$. Given such a Banach algebra of sequences $\mathcal{B}$, one can consider the Banach algebra $\tilde{\mathcal{B}}$ of sequences of operators $(M_n)_{n\in\mathbb{N}}$ (acting on some fixed Banach space $\mathcal{C}$) such that the sequence $(\|M_n\|)_{n\in\mathbb{N}}$ belongs to $\mathcal{B}$. One easily checks that $\tilde{\mathcal{B}}$ is again a Banach algebra (for the convolution product).

When the Banach algebra of sequences $\mathcal{B}$ satisfies a technical condition (its characters should all be given by evaluation of the power series $\sum c_n z^n$ at a point z of the unit disk), which is satisfied in all examples we mentioned above, then one can use the Wiener Lemma to obtain the following property: if a sequence of operators $(M_n)_{n\in\mathbb{N}}$ belongs to $\tilde{\mathcal{B}}$ and $\sum M_n z^n$ is invertible as an operator on $\mathcal{C}$ for any z in the closed unit disk, then $(M_n)_{n\in\mathbb{N}}$ is invertible in $\tilde{\mathcal{B}}$. In particular, the power series $\sum M'_n z^n = (\sum M_n z^n)^{-1}$ satisfies $(\|M'_n\|)_{n\in\mathbb{N}} \in \mathcal{B}$.

Using Banach algebra arguments and the renewal equation (4.4), the following proposition is proved in [Gou04c, Prop. 2.2.19].

Proposition 4.3. *Consider a Banach algebra of sequences $\mathcal{B}$ satisfying several technical conditions. If the sequence $(\sum_{k>n} \mu(\phi = k))_{n\in\mathbb{N}}$ belongs to $\mathcal{B}$, then this is also the case of the sequence $(\|T_{n+1} - T_n\|)_{n\in\mathbb{N}}$. Moreover, T_n converges to $\Pi : f \mapsto (\int_{\Delta_0} f) 1_{\Delta_0}$.*

The technical conditions on the Banach algebra (all the characters of $\mathcal{B}$ should be given by the evaluation at a point of the closed unit disk, and the symmetrized version of $\mathcal{B}$ should contain the Fourier coefficients of partitions of unity of the circle) will not be important for us; let us only mention that they are satisfied for the Banach algebras of series with moments of order $Q \geq 0$.

The contraction properties of the dynamics T are dictated by the number of returns to the basis. Their asymptotics are estimated in the next lemma.

Lemma 4.4. *For $x \in \Delta$, let $\psi_n(x) = \mathrm{Card}\{0 \leq k \leq n-1 \,:\, T^k x \in \Delta_0\}$ be the number of visits to the basis of x before time n, and let $\Psi_n(x) = \rho^{\psi_n(x)}$, where $\rho < 1$. If the return time on Δ_0 has a moment of order $q \geq 1$ (i.e., $\mu(\phi = n) \leq c_n^{(q)}$), we have*

$$\int_{T^{-n}\Delta_0} \Psi_n \,\mathrm{d}\mu(x) \leq c_n^{(q-1)}.$$

This bound is optimal: on $\Delta_{\alpha,\phi(\alpha)-n}$ (for α with $\phi(\alpha) > n$), we have $\Psi_n = 1$. Therefore, the integral in the lemma is bounded from below by $\mu(\bigcup_{\phi(\alpha)>n} \Delta_{\alpha,0}) \sim \sum_{n+1}^{\infty} c_k^{(q)} \sim c_n^{(q-1)}$.

Proof. Let us define an operator U_n by the series $\sum U_n z^n = \sum_{k=0}^{\infty} (\rho \sum R_n z^n)^k = (I - \rho \sum R_n z^n)^{-1}$. Then $U_n f(x) = \sum g^{(n)}(y) \Psi_n(y) f(y)$, where the sum is restricted to those $y \in \Delta_0$ with $T^n y = x$. Integrating and changing variables, we obtain

$$\int_{\Delta_0} U_n 1(x) \,\mathrm{d}\mu(x) = \int_{\Delta_0 \cap T^{-n}(\Delta_0)} \Psi_n(y) \,\mathrm{d}\mu(y).$$

Since the spectral radius of $\sum R_n z^n$ is at most 1 for $|z| \leq 1$, it follows that $I - \rho \sum R_n z^n$ is invertible on $\mathcal{C}$ (since $\rho < 1$). Moreover, the sequence $\|R_n\|$ satisfies $\|R_n\| \leq C\mu(\phi = n) \leq c_n^{(q)}$. It follows from Wiener's Lemma that $\sum U_n z^n = (I - \rho \sum R_n z^n)^{-1}$ belongs to the same Banach algebra of operators, i.e., $\|U_n\| \leq c_n^{(q)}$. We obtain

$$\int_{\Delta_0 \cap T^{-n}\Delta_0} \Psi_n(y) \,\mathrm{d}\mu(y) \leq c_n^{(q)}.$$

To study the integral of Ψ_n on $T^{-n}\Delta_0$, denote by Λ_b the set of points in Δ that enter Δ_0 exactly at time b. On Λ_b, we have $\Psi_n(y) = \Psi_{n-b}(T^b y)$. A distortion control gives

$$\int_{\Lambda_b \cap T^{-n}\Delta_0} \Psi_n \leq C\mu(\Lambda_b) \int_{\Delta_0 \cap T^{-(n-b)}\Delta_0} \Psi_{n-b} \leq C\mu(\Lambda_b) c_{n-b}^{(q)}.$$

Moreover, for $b > 0$, $\Lambda_b = \bigcup_{\phi(\alpha)\geq b} \Delta_{\alpha,\phi(\alpha)-b}$, hence $\mu(\Lambda_b) \leq \sum_{\ell\geq b} c_\ell^{(q)} \leq c_b^{(q-1)}$. We obtain

$$\int_{T^{-n}\Delta_0} \Psi_n(y)\,\mathrm{d}\mu(y) = \sum_{b=0}^{n} \int_{\Lambda_b\cap T^{-n}\Delta_0} \Psi_n(y)\,\mathrm{d}\mu(y) \leq C \sum_{b=0}^{n} c_b^{(q-1)} c_{n-b}^{(q)}.$$

By (4.1), this is bounded by $c_n^{(q-1)}$. □

4.3. Bounding D_p. To follow the same strategy as in the previous sections, we need to show that K_p is close to an integral, as in Lemma 2.3. To do so, as in the proof of this lemma, we define a function f_i as in (2.1), and control its iterates under the transfer operator. The first step is to control its Lipschitz constant.

Lemma 4.5. *For z and z' with zero height, $|f_i(z)| \leq C\,\mathrm{Lip}_i(K)$ and*

$$|f_i(z) - f_i(z')| \leq C d(z,z') \sum_{j=0}^{i} \mathrm{Lip}_j(K) c_{i-j}^{(q-1)}.$$

Proof. The inequality $|f_i(z)| \leq C\,\mathrm{Lip}_i(K)$ is trivial. To control the Lipschitz constant, as in (2.2), we decompose

$$\begin{aligned} f_i(z) - f_i(z') = &\sum (g^{(i)}(y) - g^{(i)}(y'))H(y,\dots,T^i y) \\ &+ \sum g^{(i)}(y')(H(y,\dots,T^i y) - H(y',\dots,T^i y')). \end{aligned}$$

Using distortion controls, we bound the first sum by $C\,\mathrm{Lip}_i(K) d(z,z')$. For the second sum, we replace successively each $T^j y$ with $T^j y'$, writing it as

$$\begin{aligned} \sum_{T^i y'=z'} \sum_{j=0}^{i} g^{(i)}(y')(&H(y,\dots,T^{j-1}y, T^j y, T^{j+1}y',\dots,T^i y') \\ &- H(y,\dots,T^{j-1}y, T^j y', T^{j+1}y',\dots,T^i y')). \end{aligned}$$

Since the distance between $T^j y$ and $T^j y'$ is bounded by $\Psi_{i-j}(T^j y')d(z,z')$, we obtain a bound

$$\begin{aligned} &\sum_{T^i y'=z'} \sum_{j=0}^{i} g^{(i)}(y')\Psi_{i-j}(T^j y')\,\mathrm{Lip}_j(K) d(z,z') \\ &\leq d(z,z') \sum_{j=0}^{i} \sum_{T^{i-j}(y'_j)=z'} g^{(i-j)}(y'_j)\Psi_{i-j}(y'_j)\,\mathrm{Lip}_j(K) \\ &\leq C d(z,z') \sum_{j=0}^{i} \mathrm{Lip}_j(K) \int_{T^{-(i-j)}\Delta_0} \Psi_{i-j}, \end{aligned}$$

by bounded distortion. With Lemma 4.4, this gives the result. □

To follow the strategy of proof of Lemma 2.3, we need to understand the iterates of f_i under the transfer operator. This is done in the next lemma.

Lemma 4.6. *For any $r \geq 0$ and any $z \in \Delta_0$, we have*

$$\left| \mathcal{L}^r f_i(z) - \int_\Delta f_i \right| \leq \sum_{j=0}^{i} \mathrm{Lip}_j(K) \left(\sum_{k=0}^{r} c_k^{(q-2)} c_{i-j+r-k}^{(q-1)} \right).$$

Proof. We will use the decomposition $1_{\Delta_0}\mathcal{L}^r = \sum_{k+b=r} T_k B_b$ given by (4.3) to understand $\mathcal{L}^r f_i$.

Let us first describe the asymptotics of T_k. Let $\mathcal{C}$ denote the space of Lipschitz functions on the basis Δ_0 of the tower. We define an operator Π on $\mathcal{C}$ by $\Pi f = (\int_{\Delta_0} f) 1_{\Delta_0}$. The operators T_n converge to Π. Since $\|T_n - T_{n+1}\| \leq c_n^{(q-1)}$ by Proposition 4.3, we have

$$\|T_k - \Pi\| \leq \sum_{n=k}^{\infty} \|T_n - T_{n+1}\| \leq \sum_{n=k}^{\infty} c_n^{(q-1)} \leq c_k^{(q-2)}, \tag{4.5}$$

by (4.2).

We will now estimate $\|B_b f_i\|_{\mathcal{C}}$ using Lemma 4.5. For $z \in \Delta_0$, we have

$$B_b f_i(z) = \sum_{\phi(\alpha) \geq b} g^{(b)}(z_\alpha) f_i(z_\alpha),$$

where z_α is the unique preimage of z under T^b in $\Delta_{\alpha, \phi(\alpha)-b}$. We have

$$|B_b f_i|_\infty \leq |f_i|_\infty \cdot C \sum_{\phi(\alpha) \geq b} \mu(\Delta_{\alpha,0}) \leq C |f_i|_\infty c_b^{(q-1)} \leq C \,\mathrm{Lip}_i(K) c_b^{(q-1)}. \tag{4.6}$$

Let us now estimate $B_b f_i(z) - B_b f_i(z')$ for z and z' in the same partition element. If we form the difference $g^{(b)}(z_\alpha) - g^{(b)}(z'_\alpha)$, the resulting term is bounded by $C d(z, z') \,\mathrm{Lip}_i(K) c_b^{(q-1)}$ (using distortion controls and the same computation as in (4.6)). On the other hand, denoting by $h_\alpha = \phi(\alpha) - b$ the height of z_α, we have

$$|f_i(z_\alpha) - f_i(z'_\alpha)| \leq C \left(\sum_{j=0}^{i-h_\alpha} \mathrm{Lip}_j(K) c_{i-j-h_\alpha}^{(q-1)} + \sum_{j=i-h_\alpha+1}^{i} \mathrm{Lip}_j(K) \right) d(z, z').$$

This follows from Lemma 4.5 applied to the function f_{i-h_α} and the points πz_α and $\pi z'_\alpha$. Summing over α, we obtain a bound for the Lipschitz constant of $B_b f_i$ of the form

$$\sum_{\phi(\alpha) \geq b} g^{(b)}(z_\alpha) \left[\sum_{j=0}^{i-h_\alpha} \mathrm{Lip}_j(K) c_{i-j-h_\alpha}^{(q-1)} + \sum_{j=i-h_\alpha+1}^{i} \mathrm{Lip}_j(K) \right].$$

By bounded distortion, $g^{(b)}(z_\alpha) \leq C\mu(\Delta_{\alpha,0})$. Taking the union over α and writing $\ell = \phi(\alpha)$, we get that the coefficient of $\mathrm{Lip}_j(K)$ in this sum is bounded by

$$C \sum_{\ell=b}^{b+i-j} \mu(\phi = \ell) c_{i-j-(\ell-b)}^{(q-1)} + C \sum_{\ell=b+i-j+1}^{\infty} \mu(\phi = \ell).$$

The second term is bounded by $c_{i-j+b}^{(q-1)}$ by (4.2), while the first term is bounded by

$$\sum_{\ell=0}^{i-j+b} c_\ell^{(q)} c_{i-j+b-\ell}^{(q-1)} \leq c_{i-j+b}^{(q-1)}$$

by (4.1). We have shown that

$$\|B_b f_i\|_{\mathcal{C}} \leq \sum_{j=0}^{i} \mathrm{Lip}_j(K) c_{i-j+b}^{(q-1)}.$$

(The contribution of (4.6) is compatible with this bound.)

Let us now study $\mathcal{L}^r f_i$ on Δ_0. We write $T_k = \Pi + E_k$ with $\|E_k\| \leq c_k^{(q-2)}$, by (4.5). Hence,

$$\mathcal{L}^r f_i = \sum_{k+b=r} T_k B_b f_i = \sum_{k+b=r} \Pi B_b f_i + \sum_{k+b=r} E_k B_b f_i. \tag{4.7}$$

The first term is a constant function equal to $\sum_{b=0}^{r} \int_{\Delta_0} B_b f_i$. Denoting by Λ_b the set of points that enter Δ_0 exactly at time b, we have $\int_{\Delta_0} B_b f_i = \int_{\Lambda_b} f_i$. As a consequence

$$\begin{aligned}\left|\sum_{b=0}^{r} \int_{\Delta_0} B_b f_i - \int f_i\right| = \left|-\int_{\bigcup_{b>r} \Lambda_b} f_i\right| &\leq |f_i|_\infty \sum_{b>r} \mu(\Lambda_b) \\ &\leq \mathrm{Lip}_i(K) \sum_{b>r} c_b^{(q-1)} \leq \mathrm{Lip}_i(K) c_r^{(q-2)},\end{aligned}$$

by (4.2). This bound is compatible with the statement of the lemma. The second term of (4.7) is bounded (in $\mathcal{C}$ norm, thus in sup norm) by

$$\sum_{k+b=r} c_k^{(q-2)} \|B_b f_i\|_{\mathcal{C}} \leq \sum_{j=0}^{i} \mathrm{Lip}_j(K) \cdot \sum_{k+b=r} c_k^{(q-2)} c_{i-j+b}^{(q-1)}.$$

This proves the lemma. □

We can now obtain the following lemma, which is the analogue in our setting of Lemma 2.3.

Lemma 4.7. *For all* $x_p \in \Delta_0$*,*

$$\left|K_p(x_p, \ldots) - \int K(y, \ldots, T^{p-1}y, x_p, \ldots)\, \mathrm{d}\mu(y)\right| \leq \sum_{j=0}^{p-1} \mathrm{Lip}_j(K) c_{p-j}^{(q-2)}.$$

Proof. Just like in the proof of Lemma 2.3,

$$\left|K_p(x_p, \ldots) - \int K(y, \ldots, T^{p-1}y, x_p, \ldots)\right| \leq \sum_{i=0}^{p-1} \left|\mathcal{L}^{p-i} f_i(x_p) - \int f_i\right|.$$

By Lemma 4.6, this quantity is bounded by

$$C\sum_{i=0}^{p-1}\sum_{j=0}^{i}\mathrm{Lip}_j(K)\left(\sum_{k=0}^{p-i}c_k^{(q-2)}c_{i-j+p-i-k}^{(q-1)}\right).$$

The coefficient of $\mathrm{Lip}_j(K)$ in this sum is

$$\sum_{k=0}^{p-j}c_k^{(q-2)}(p-k-j)c_{p-k-j}^{(q-1)}\leq\sum_{k=0}^{p-j}c_k^{(q-2)}c_{p-k-j}^{(q-2)}\leq c_{p-j}^{(q-2)}$$

by (4.1). This proves the lemma. □

The previous lemma makes it possible to control the moments of $D_p=K_p-K_{p+1}$:

Lemma 4.8. *For all $\kappa\leq 2q$,*

$$\mathbb{E}(|D_p|^\kappa|\mathcal{F}_{p+1})(x_{p+1},\dots)\leq C\sum_{j=0}^{p}\mathrm{Lip}_j(K)^\kappa c_{p-j}^{(q-2)}$$
$$+C\sum_{h\geq 0}c_h^{(q-\kappa/2)}\left(\sum_{j=p-h+1}^{p}\mathrm{Lip}_j(K)^2\right)^{\kappa/2}.$$

Proof. We follow closely the strategy of the proof of Lemma 3.3. If the height of x_{p+1} is positive, the estimate is trivial. Otherwise, let $\{z_\alpha\}$ denote the preimages of x_{p+1} under T, with respective height $h_\alpha=\phi(\alpha)-1$. Let $A(z)=D_p(z,x_{p+1},\dots)$; we have $\mathbb{E}(|D_p|^\kappa|\mathcal{F}_{p+1})(x_{p+1},\dots)=\sum g(z_\alpha)|A(z_\alpha)|^\kappa$.

Fix a point $z=z_\alpha$ with height $h\geq 0$. If $h\leq p$, consider the projection πz of z in the basis of the tower. Using Lemma 4.7 (at time $p-h$ for the point πz, and at time $p+1$ for the point x_{p+1}), we get

$$|A(z)|\leq\sum_{j\leq p-h}\mathrm{Lip}_j(K)c_{p-h-j}^{(q-2)}+\sum_{j=p-h+1}^{p}\mathrm{Lip}_j(K).\tag{4.8}$$

This estimate also holds (trivially) if $h>p$.

To estimate $|A(z)|^\kappa$, we first use the inequality $(x+y)^\kappa\leq Cx^\kappa+Cy^\kappa$ to separate the two sums. Then, in the first sum, since $c_{p-h-j}^{(q-2)}$ is summable, we may use the Hölder inequality to get $\left(\sum_{j\leq p-h}\mathrm{Lip}_j(K)c_{p-h-j}^{(q-2)}\right)^\kappa\leq C\sum_{j\leq p-h}\mathrm{Lip}_j(K)^\kappa c_{p-h-j}^{(q-2)}$. For the second sum, we write $\left(\sum_{j=p-h+1}^{p}\mathrm{Lip}_j(K)\right)^2\leq h\sum_{j=p-h+1}^{p}\mathrm{Lip}_j(K)^2$, and we obtain

$$|A(z)|^\kappa\leq\sum_{j\leq p-h}\mathrm{Lip}_j(K)^\kappa c_{p-h-j}^{(q-2)}+Ch^{\kappa/2}\left(\sum_{j=p-h+1}^{p}\mathrm{Lip}_j(K)^2\right)^{\kappa/2}.$$

Summing over α, we get that $\sum g(z_\alpha)|A(z_\alpha)|^\kappa$ is at most

$$C \sum_{h=0}^{\infty} \mu(\phi = h) \left(\sum_{j \leq p-h} \mathrm{Lip}_j(K)^\kappa c_{p-h-j}^{(q-2)} + h^{\kappa/2} \left(\sum_{j=p-h+1}^{p} \mathrm{Lip}_j(K)^2 \right)^{\kappa/2} \right).$$

In the first sum, the coefficient of $\mathrm{Lip}_j(K)^\kappa$ is at most

$$\sum_{h=0}^{p-j} c_h^{(q)} c_{p-h-j}^{(q-2)} \leq c_{p-j}^{(q-2)}$$

by (4.1). In the second sum, $\mu(\phi = h)h^{\kappa/2} \leq c_h^{(q-\kappa/2)}$, yielding the statement of the lemma. □

4.4. Proof of Theorem 4.1. We will use the following Rosenthal-Burkholder martingale inequality [Bur73, Thm. 21.1 and Ineq. (21.5)]. Let $\mathcal{F}_p$ be a decreasing sequence of σ-algebras, and let D_p be a sequence of reverse martingale difference with respect to $\mathcal{F}_p$ (i.e., D_p is $\mathcal{F}_p$-measurable and $\mathbb{E}(D_p|\mathcal{F}_{p+1}) = 0$). For all $Q \geq 2$,

$$\left\| \sum D_p \right\|_{L^Q}^Q \leq C\mathbb{E} \left(\left[\sum_p \mathbb{E}(D_p^2|\mathcal{F}_{p+1}) \right]^{Q/2} \right) + C \sum_p \mathbb{E}(|D_p|^Q).$$

We apply this inequality to $\mathcal{F}_p$ the σ-algebra of sets depending only on $x_p, x_{p+1}, \ldots$, to $D_p = K_p - K_{p+1}$ and to $Q = 2q - 2$. By Lemma 4.8 with $\kappa = 2$, we have

$$\mathbb{E}(D_p^2|\mathcal{F}_{p+1})(x_{p+1}, \ldots) \leq C \sum_{j=0}^{p} \mathrm{Lip}_j(K)^2 c_{p-j}^{(q-2)} + C \sum_{h \geq 0} c_h^{(q-1)} \sum_{j=p-h+1}^{p} \mathrm{Lip}_j(K)^2. \tag{4.9}$$

The coefficient of $\mathrm{Lip}_j(K)^2$ in this estimate is bounded by $c_{p-j}^{(q-2)} + \sum_{h \geq p-j+1} c_h^{(q-1)} \leq c_{p-j}^{(q-2)}$. Hence, the first term in the Rosenthal-Burkholder inequality is bounded by

$$C \left(\sum_p \sum_{j=0}^{p} \mathrm{Lip}_j(K)^2 c_{p-j}^{(q-2)} \right)^{q-1} \leq C \left(\sum_j \mathrm{Lip}_j(K)^2 \right)^{q-1}.$$

For the second term, we should bound $\sum_p \mathbb{E}(|D_p|^{2q-2})$. We sum the estimates of Lemma 4.8 (with $\kappa = 2q - 2$), to get

$$\sum_p \mathbb{E}(|D_p|^{2q-2}) \leq C \sum_j \sum_{p \geq j} \mathrm{Lip}_j(K)^{2q-2} c_{p-j}^{(q-2)} + C \sum_{h \geq 0} c_h^{(1)} \sum_p \left(\sum_{j=p-h+1}^{p} \mathrm{Lip}_j(K)^2 \right)^{q-1}. \tag{4.10}$$

In the first sum, the coefficient of $\mathrm{Lip}_j(K)^{2q-2}$ is $\sum_k c_k^{(q-2)} \le C$, therefore this sum is bounded by $C \sum_j \mathrm{Lip}_j(K)^{2q-2} \le C \left(\sum \mathrm{Lip}_j(K)^2\right)^{q-1}$.

The second sum is more delicate. Let us fix h and $p_0 \in [0, h)$, and let us consider the contribution of those p in $p_0 + \mathbb{Z}h$. The intervals $[p-h+1, p]$ are disjoint. The inequality $\sum x_i^{q-1} \le (\sum x_i)^{q-1}$ yields

$$\sum_{p \equiv p_0\ [h]} \left(\sum_{j=p-h+1}^{p} \mathrm{Lip}_j(K)^2 \right)^{q-1} \le \left(\sum_{p \equiv p_0\ [h]} \sum_{j=p-h+1}^{p} \mathrm{Lip}_j(K)^2 \right)^{q-1} \le \left(\sum_j \mathrm{Lip}_j(K)^2 \right)^{q-1}.$$

Summing over the h possible values of p_0, we get that the second sum of (4.10) is bounded by

$$C \sum_{h \ge 0} c_h^{(1)} h \left(\sum_j \mathrm{Lip}_j(K)^2 \right)^{q-1} \le C \left(\sum_j \mathrm{Lip}_j(K)^2 \right)^{q-1},$$

since $\sum h c_h^{(1)} < \infty$ by definition.

We have proved that $\left\| \sum D_p \right\|_{L^{2q-2}}^{2q-2} \le C \left(\sum_j \mathrm{Lip}_j(K)^2 \right)^{q-1}$. Since $\sum D_p = K - \mathbb{E}(K)$, this proves Theorem 4.1. □

5. Invertible Non-uniform Young Towers

Let $T : X \to X$ be a non-uniform Young tower, with invariant measure μ. Its natural extension $T_{\mathbb{Z}} : X_{\mathbb{Z}} \to X_{\mathbb{Z}}$ preserves a probability measure $\mu_{\mathbb{Z}}$. There is a natural distance on $X_{\mathbb{Z}}$, defined as follows. First, the positive separation time $s(x, y)$ is defined as for T. One can also define a negative separation time $s_-(x, y)$ in the same way, but towards the past: one iterates towards the past until the points are in different elements of the Markov partition, and one counts the number of visits to Δ_0 in between. The distance $d_{\mathbb{Z}}$ is then defined by $d_{\mathbb{Z}}(x, y) = \beta^{\min(s(x,y), s_-(x,y))}$. Geometrically, this distance is interpreted as follows: when one returns to the basis, there is uniform contraction along stable manifolds (corresponding to the past), and uniform expansion along unstable manifolds. Two points are close in the unstable direction if they remain close in the future for a long time (distance $\beta^{s(x,y)}$), while they are close in the stable direction if they have a long common past (distance $\beta^{s_-(x,y)}$).

Theorem 5.1. *Let $(T_{\mathbb{Z}}, X_{\mathbb{Z}}, \mu_{\mathbb{Z}})$ be the natural extension of a Young tower in which the return time function ϕ has a moment of order q. This system satisfies a concentration inequality with moment $2q-2$, i.e., there exists a constant $C > 0$ such that, for any $n \in \mathbb{N}$, for any function $K_{\mathbb{Z}}(x_0, \dots, x_{n-1})$ which is separately Lipschitz for the distance $d_{\mathbb{Z}}$,*

$$\int \left| K_{\mathbb{Z}}(x, \ldots, T^{n-1}x) - \int K_{\mathbb{Z}}(y, \ldots, T^{n-1}y)\, \mathrm{d}\mu_{\mathbb{Z}}(y) \right|^{2q-2} \mathrm{d}\mu_{\mathbb{Z}}(x)$$
$$\leq C \left(\sum_j \mathrm{Lip}_j(K_{\mathbb{Z}})^2 \right)^{q-1}.$$

This implies Theorem 4.1 (if one considers a function $K_{\mathbb{Z}}$ depending only on the future of the points), but the converse is not true: since the contraction is not uniform, we are not able to reduce this theorem to Theorem 4.1, contrary to what we have done for subshifts of finite type or uniform Young towers.

For the proof, we will work with the non-invertible system X, or rather with an augmented space $X_* = X \cup \{x_*\}$, where x_* is a new point (at distance 1 of any point of X, with zero measure).

Let us start with a function $K_{\mathbb{Z}}$ on $X_{\mathbb{Z}}$, depending on the past and the future of points. We define a new function K on X_*^n as follows. We let $K(x_0, \ldots, x_{n-1}) = K_{\mathbb{Z}}(y_0, \ldots, y_{n-1})$ where the y_i are defined inductively. For each element a of the partition, let us fix an admissible past $p(a)$. Let us also fix a point $y_* \in X_{\mathbb{Z}}$. Let $y_0 = (p((x_0)_0), x_0)$ (unless $x_0 = x_*$, in which case let $y_0 = y_*$). If y_{i-1} is defined, let us define y_i. If $x_i = x_*$, we take $y_i = y_*$. If the transition from $(x_{i-1})_0$ to $(x_i)_0$ is not permitted, let $y_i = (p((x_i)_0), x_i)$. Otherwise, let $y_i = ((y_{i-1})_{-\infty}^0, x_i)$.

We claim that this function K satisfies an inequality

$$\int_{X_*} \left| K(x, \ldots, T^{n-1}x) - \int K(y, \ldots, T^{n-1}y)\, \mathrm{d}\mu(y) \right|^{2q-2} \mathrm{d}\mu(x)$$
$$\leq C \left(\sum_{j=0}^{n-1} \mathrm{Lip}_j(K_{\mathbb{Z}})^2 \right)^{q-1}. \tag{5.1}$$

This implies Theorem 5.1 by using the same argument as in Subsect. 2.2: let $K_N(y_0, \ldots, y_{n+N-1}) = K_{\mathbb{Z}}(y_N, \ldots, y_{N+n-1})$, and let $\tilde{K}_N$ be the function obtained from K_N by applying the above procedure. After a change of variables, we get from (5.1),

$$\int_{X_{\mathbb{Z}}} \left| \tilde{K}_N(T^{-N}x, \ldots, x, Tx, \ldots, T^{n-1}x) - \mathbb{E}(\tilde{K}_N) \right|^{2q-2} \mathrm{d}\mu_{\mathbb{Z}}(x)$$
$$\leq C \left(\sum_{j=0}^{n-1} \mathrm{Lip}_j(K_{\mathbb{Z}})^2 \right)^{q-1}.$$

When N tends to ∞, $\tilde{K}_N(T^{-N}x, \ldots, x, Tx, \ldots, T^{n-1}x)$ converges to $K_{\mathbb{Z}}(x, \ldots, T^{N-1}x)$. Hence, we obtain the desired concentration inequality by letting N tend to infinity in the previous equation.

To prove (5.1), we follow the same strategy as in the previous section. Note that we can not directly apply Theorem 4.1 since the Lipschitz constants of K are not easily bounded in terms of those of $K_{\mathbb{Z}}$, due to the non-uniform expansion. Therefore, we have to reimplement the strategy from scratch.

Let us first start with a crucial remark. When one controls the Lipschitz constants of K in terms of those of $K_{\mathbb{Z}}$, a point x_* blocks the propagation of modifications, in the following sense: consider a difference $K(x_0, \ldots, x_{n-1}) - K(x'_0, \ldots, x'_{n-1})$ where x_i and

x_i' coincide at all indices but j. By construction of K, this is equal to $K_{\mathbb{Z}}(y_0, \dots, y_{n-1}) - K_{\mathbb{Z}}(y_0', \dots, y_{n-1}')$ for some points $y_i, y_i' \in X_{\mathbb{Z}}$. The definition shows that $y_i = y_i'$ for $i < j$. On the other hand, y_i and y_i' might be different for all $i \geq j$, not only for $i = j$. However, if there is an index $k > j$ such that $x_k = x_k' = x_*$, then $y_i = y_i'$ for $i \geq k$: this follows directly from the construction. Therefore, $K(x_0, \dots, x_{n-1}) - K(x_0', \dots, x_{n-1}')$ will be estimated only in terms of $\mathrm{Lip}_i(K_{\mathbb{Z}})$ for $j \leq i < k$.

To follow the same strategy as in the previous sections, we need to show that K_p is close to an integral, as in Lemma 2.3. To do so, as in the proof of this lemma, we define a function f_i as in (2.1), and control its iterates under the transfer operator. We decompose $K_p(x_p, \dots) = \sum_{i=0}^{p-1} \mathcal{L}^{p-i} f_i(x_p) + K(x_*, \dots, x_*, x_p, \dots)$, where

$$
\begin{aligned}
f_i(z) = \sum_{T^i y = z} g^{(i)}(y)(K(y, \dots, T^i y, x_*, \dots, x_*, x_p, \dots) \\
- K(y, \dots, T^{i-1} y, x_*, \dots, x_*, x_p, \dots)).
\end{aligned}
$$

When $i < p-1$, there is a point x_* in the definition of f_i, blocking the propagation of modifications as we explained above. Therefore, we may follow the proofs of Lemmas 4.5 and 4.6 in this setting, to obtain the following:

Lemma 5.2. *If $i < p-1$, we have for any $r \geq 0$ and any $z \in \Delta_0$,*

$$
\left| \mathcal{L}^r f_i(z) - \int_{\Delta} f_i \right| \leq \sum_{j=0}^{i} \mathrm{Lip}_j(K_{\mathbb{Z}}) \left(\sum_{k=0}^{r} c_k^{(q-2)} c_{i-j+r-k}^{(q-1)} \right).
$$

On the other hand, there is no such blocking effect for f_{p-1}, yielding a worse estimate. Indeed, in f_{p-1}, one considers averages of terms of the form $K(y, \dots, T^{p-1} y, x_p, \dots) - K(y, \dots, T^{p-2} y, x_*, x_p, \dots)$. Considering the definition of K in terms of $K_{\mathbb{Z}}$, this difference reads $K_{\mathbb{Z}}(y_0', \dots, y_{n-1}') - K_{\mathbb{Z}}(y_0'', \dots, y_{n-1}'')$ where the points y_j', y_j'' belong to $X_{\mathbb{Z}}$, coincide for $j < p-1$ and may differ for $j \geq p-1$. For $j > p-1$, the points y_j' and y_j'' have the same future, and the same past up to the index $j-p$. Therefore, $d_{\mathbb{Z}}(y_j', y_j'') \leq \beta^{\mathrm{Card}\{k \in [p,j] : x_k \in \Delta_0\}}$. Averaging over the points y with $T^{p-1}(y) = z$, we get

$$
|f_{p-1}(z)| \leq \sum_{j=p-1}^{n-1} \mathrm{Lip}_j(K_{\mathbb{Z}}) \beta^{\mathrm{Card}\{k \in [p,j] : x_k \in \Delta_0\}}.
$$

The functions $\mathcal{L} f_{p-1}$ and $\mathcal{L} f_{p-1} - \int f_{p-1}$ also satisfy the same bound.

Still following the strategy of proof of Sect. 4, we deduce from those estimates an analogue of Lemma 4.7, with an additional error term coming from f_{p-1}: for all $x_p \in \Delta_0$,

$$
\begin{aligned}
&\left| K_p(x_p, \dots) - \int K(y, \dots, T^{p-1} y, x_p, \dots) \, \mathrm{d}\mu(y) \right| \\
&\leq C \sum_{j=0}^{p-1} \mathrm{Lip}_j(K_{\mathbb{Z}}) c_{p-j}^{(q-2)} + C \sum_{j=p}^{n-1} \mathrm{Lip}_j(K_{\mathbb{Z}}) \beta^{\mathrm{Card}\{k \in [p,j] : x_k \in \Delta_0\}}.
\end{aligned}
$$

In turn, this yields an analogue of Lemma 4.8, still with an additional error term: for all $\kappa \le 2q$, and for all $x_{p+1} \in \Delta_0$,

$$\mathbb{E}(|D_p|^\kappa | \mathcal{F}_{p+1})(x_{p+1}, \dots) \le C \left(\sum_{j \ge p+1} \mathrm{Lip}_j(K_{\mathbb{Z}}) \beta^{\mathrm{Card}\{k \in [p+1, j] \,:\, x_k \in \Delta_0\}} \right)^\kappa$$
$$+ C \sum_{j=0}^{p} \mathrm{Lip}_j(K_{\mathbb{Z}})^\kappa c_{p-j}^{(q-2)} + C \sum_{h \ge 0} c_h^{(q-\kappa/2)} \left(\sum_{j=p-h+1}^{p} \mathrm{Lip}_j(K_{\mathbb{Z}})^2 \right)^{\kappa/2}. \quad (5.2)$$

On the other hand, $\mathbb{E}(|D_p|^\kappa | \mathcal{F}_{p+1})(x_{p+1}, \dots) = 0$ if $h(x_{p+1}) > 0$.

We can now conclude the proof of (5.1), following the strategy we used to prove Theorem 4.1 in Subsect. 4.4. By the Rosenthal-Burkholder inequality, we have

$$\mathbb{E}|K - \mathbb{E}K|^{2q-2} = \mathbb{E}\left|\sum D_p\right|^{2q-2}$$
$$\le C\mathbb{E}\left(\left[\sum_p \mathbb{E}(D_p^2 | \mathcal{F}_{p+1})\right]^{q-1}\right) + C \sum \mathbb{E}(|D_p|^{2q-2}).$$

The conditional expectations are estimated thanks to (5.2). The terms that were already present in the proof of Theorem 4.1 are handled exactly in the same way. Therefore, we only need to deal with the additional term. Let us define a function $\Phi_j(x) = \beta^{\mathrm{Card}\{k \in [1, j] \,:\, T^k(x) \in \Delta_0\}}$ for $x \in \Delta_0$, and $\Phi_j(x) = 0$ elsewhere (it is closely related to the function Ψ_j of Lemma 4.4, with the difference that it is supported in Δ_0). The additional term in the Rosenthal-Burkholder inequality is bounded by

$$C \int \left[\sum_{p \ge 0} \left(\sum_{j \ge p+1} \mathrm{Lip}_j(K_{\mathbb{Z}}) \Phi_{j-p-1}(T^{p+1}x) \right)^2 \right]^{q-1} \mathrm{d}\mu(x)$$
$$+ C \sum_{p \ge 0} \int \left(\sum_{j \ge p+1} \mathrm{Lip}_j(K_{\mathbb{Z}}) \Phi_{j-p-1}(T^{p+1}x) \right)^{2q-2} \mathrm{d}\mu(x).$$

The inequality $\sum x_i^{q-1} \le (\sum x_i)^{q-1}$ shows that the second term is bounded by the first one. Therefore, to conclude the proof, it is sufficient to prove the following inequality:

$$\int \left[\sum_{p \ge 0} \left(\sum_{j \ge p+1} \mathrm{Lip}_j(K_{\mathbb{Z}}) \Phi_{j-p-1}(T^{p+1}x) \right)^2 \right]^{q-1} \mathrm{d}\mu(x) \le C \left(\sum \mathrm{Lip}_j(K_{\mathbb{Z}})^2 \right)^{q-1}. \quad (5.3)$$

This estimate is formulated solely in terms of the non-invertible system. Its proof is technical and complicated. Therefore, we defer it to Theorem A.1 in Appendix A. Modulo this result, this concludes the proof of (5.1), and of Theorem 5.1.

6. Weak Polynomial Concentration Inequalities

The results of Sect. 4 are not completely satisfactory for the significant example of intermittent maps. Indeed, for Pomeau-Manneville maps of index $\alpha \in (0,1)$ (with $T(x) = x + cx^{1+\alpha}(1+o(1))$ for small x, see (7.4) below), the return time function to the rightmost interval satisfies a bound $\mu\{\phi = n\} \sim C/n^{1/\alpha+1}$. Therefore, the corresponding Young tower has a moment of order q for any $q < 1/\alpha$ (which yields a concentration inequality of order Q for any $Q < 2/\alpha - 2$ when $\alpha < 1/2$), but it does not have a moment of order $1/\alpha$. Indeed, it only has a *weak* moment of order $1/\alpha$, meaning that $\mu\{\phi > t\} \leq Ct^{-1/\alpha}$. An optimal concentration statement for such a map would therefore be formulated in terms of weak moments. This is our goal in this section.

Theorem 6.1. *Let $T : \Delta \to \Delta$ be a non-uniform Young tower. Assume that, for some $q > 2$, the return time ϕ to the basis of the tower has a weak moment of order q, i.e., there exists a constant $C > 0$ such that $\mu\{x \in \Delta_0 : \phi(x) > t\} \leq Ct^{-q}$ for all $t > 0$. Then T satisfies a weak polynomial concentration inequality with moment $2q-2$, i.e., there exists a constant $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$, and any $t > 0$,*

$$\mu\left\{x : \left|K(x, \dots, T^{n-1}x) - \int K(y, \dots, T^{n-1}y)\,\mathrm{d}\mu(y)\right| > t\right\}$$
$$\leq Ct^{-(2q-2)}\left(\sum_j \mathrm{Lip}_j(K)^2\right)^{q-1}.$$

Let us introduce a convenient notation. When Z is a real-valued random variable and $Q \geq 1$, we write $\|Z\|_{L^{Q,w}} = \sup t P(|Z| > t)^{1/Q}$, so that $\mathbb{P}(|Z| > t) \leq t^{-Q}\|Z\|_{L^{Q,w}}^Q$. This is the weak L^Q (semi)norm of Z. With this notation, the statement of the theorem becomes $\|K - \mathbb{E}(K)\|_{L^{2q-2,w}}^{2q-2} \leq C\left(\sum_j \mathrm{Lip}_j(K)^2\right)^{q-1}$, in close analogy with the statement of Theorem 4.1. Note that $\|Z\|_{L^{Q,w}}$ is not a true norm: the triangle inequality fails, and is replaced by $\|Z + Z'\|_{L^{Q,w}} \leq C(\|Z\|_{L^{Q,w}} + \|Z'\|_{L^{Q,w}})$. On the other hand,

$$\left\|\max(|Z|, |Z'|)\right\|_{L^{Q,w}}^Q \leq \|Z\|_{L^{Q,w}}^Q + \|Z'\|_{L^{Q,w}}^Q.$$

Since a sequence with a weak moment of order $q > 2$ has a strong moment of order 2, we may use intermediate results of the proof of Theorem 4.1 (and especially Lemma 4.7) to prove Theorem 6.1. The proofs diverge at the level of Lemma 4.8: the version we will need in the weak moments case is the following.

Lemma 6.2. *Assume that ϕ has a weak moment of order $q > 2$. For all $t > 0$,*

$$\mathbb{P}(|D_p| > t|\mathcal{F}_{p+1})(x_{p+1}, \dots) \leq Ct^{-(2q-2)}\sum_{j=0}^{p}\mathrm{Lip}_j(K)^{2q-2}c_{p-j}^{(0)}$$
$$+Ct^{-(2q-2)}\left(\sum \mathrm{Lip}_j(K)^2\right)^{q-2}\sup_{h>0}\left(h^{-1}\sum_{j=p-h+1}^{p}\mathrm{Lip}_j(K)\right)^2.$$

Proof. If $h(x_{p+1}) > 0$, then x_{p+1} has a unique preimage x_p, and $D_p(x_p, x_{p+1}, \dots) = 0$. Therefore, there is nothing to prove. Assume now that $h(x_{p+1}) = 0$, and let $\{z_\alpha\}$ denote the preimages of x_{p+1} under T. Writing $A(z) = D_p(z, x_{p+1}, \dots)$, we have

$$\mathbb{P}(|D_p| > t|\mathcal{F}_{p+1})(x_{p+1}, \dots) = \sum_{|A(z_\alpha)|>t} g(z_\alpha).$$

Since ϕ has a weak moment of order $q > 2$, it has a strong moment of order 2. Therefore, (4.8) gives

$$|A(z)| \le \sum_{j\le p-h} \operatorname{Lip}_j(K) c^{(0)}_{p-h-j} + \sum_{j=p-h+1}^{p} \operatorname{Lip}_j(K) =: A_1(z) + A_2(z).$$

If $|A(z)| > t$, then $A_1(z) > t/2$ or $A_2(z) > t/2$. Therefore, $\mathbb{P}(|D_p| > t|\mathcal{F}_{p+1})$ is bounded by

$$\sum_{A_1(z_\alpha)>t/2} g(z_\alpha) + \sum_{A_2(z_\alpha)>t/2} g(z_\alpha). \tag{6.1}$$

For the first sum,

$$\begin{aligned}\sum_{A_1(z_\alpha)>t/2} g(z_\alpha) &\le C \sum g(z_\alpha)(A_1(z_\alpha)/t)^{2q-2}\\ &\le C \sum_{h\ge 0} \mu(\phi = h) t^{-(2q-2)} \left(\sum_{j\le p-h} \operatorname{Lip}_j(K) c^{(0)}_{p-h-j} \right)^{2q-2}\\ &\le C t^{-(2q-2)} \sum_{h\ge 0} \mu(\phi = h) \sum_{j\le p-h} \operatorname{Lip}_j(K)^{2q-2} c^{(0)}_{p-h-j}.\end{aligned}$$

The coefficient of $\operatorname{Lip}_j(K)^{2q-2}$ in this expression is $\sum_{h=0}^{p-j} c^{(2)}_h c^{(0)}_{p-h-j} \le c^{(0)}_{p-j}$. Therefore, this is bounded by $C t^{-(2q-2)} \sum_{j\le p} \operatorname{Lip}_j(K)^{2q-2} c^{(0)}_{p-j}$.

The second sum of (6.1) is bounded by $C \sum \mu(\phi = \ell)$, where the sum is restricted to those ℓ with $\sum_{p-\ell+1}^{p} \operatorname{Lip}_j(K) > t/2$. Let h be the smallest such ℓ, the sum is bounded by

$$\mu(\phi \ge h) \le C h^{-q} \le C h^{-q} \left(\sum_{p-h+1}^{p} \operatorname{Lip}_j(K)/t \right)^{2q-2}.$$

To bound the last sum, we use the inequality $(\sum_{p-h+1}^{p} x_j)^2 \le h \sum x_j^2$, to obtain

$$\begin{aligned}h^{-q} \left(\sum_{p-h+1}^{p} \operatorname{Lip}_j(K) \right)^{2q-2} &= h^{-q} \left(\sum_{p-h+1}^{p} \operatorname{Lip}_j(K) \right)^{2} \cdot \left(\sum_{p-h+1}^{p} \operatorname{Lip}_j(K) \right)^{2q-4}\\ &\le h^{-q} \left(\sum_{p-h+1}^{p} \operatorname{Lip}_j(K) \right)^{2} \cdot \left(h \sum_{p-h+1}^{p} \operatorname{Lip}_j(K)^2 \right)^{q-2}\\ &\le h^{-2} \left(\sum_{p-h+1}^{p} \operatorname{Lip}_j(K) \right)^{2} \cdot \left(\sum_{j\in\mathbb{Z}} \operatorname{Lip}_j(K)^2 \right)^{q-2}.\end{aligned}$$

This concludes the proof. □

To proceed, we need an analogue of the Rosenthal-Burkholder inequality for weak moments. Although it is not written explicitly in Burkholder's article [Bur73], it follows easily from the techniques developed there, giving the following statement.

Theorem 6.3. *Let (D_p) be a sequence of reverse martingale differences with respect to a decreasing filtration $\mathcal{F}_p$ (i.e., D_p is $\mathcal{F}_p$-measurable and $\mathbb{E}(D_p|\mathcal{F}_{p+1}) = 0$). For all $Q \geq 2$,*

$$\left\| \sum D_p \right\|_{L^{Q,w}}^Q \leq C \left\| \sum \mathbb{E}(D_p^2|\mathcal{F}_{p+1}) \right\|_{L^{Q/2,w}}^{Q/2} + C \left\| \sup |D_p| \right\|_{L^{Q,w}}^Q .$$

In particular,

$$\left\| \sum D_p \right\|_{L^{Q,w}}^Q \leq C \left\| \sum \mathbb{E}(D_p^2|\mathcal{F}_{p+1}) \right\|_{L^{Q/2,w}}^{Q/2} + C \sum \left\| D_p \right\|_{L^{Q,w}}^Q .$$

Proof. By a truncation argument, it suffices to prove the result for bounded random variables, and $p \in [0, P]$. Define three random variables

$$X = \sup_{0 \leq p \leq P} \left| \sum_{k=p}^{P} D_k \right|, \quad Y = \left(\sum \mathbb{E}(D_p^2|\mathcal{F}_{p+1}) \right)^{1/2}, \quad Z = \max_{0 \leq p \leq P} |D_p|.$$

The inequality (21.2) in [Bur73] gives, for any $0 < \delta < \beta - 1$,

$$\mathbb{P}(X > \beta t, \max(Y, Z) \leq \delta t) \leq \epsilon \mathbb{P}(X > t),$$

where $\epsilon = \delta^2/(\beta - \delta - 1)^2$. In particular,

$$\begin{aligned}(\beta t)^Q \mathbb{P}(X > \beta t) &\leq (\beta t)^Q \mathbb{P}(\max(Y, Z) > \delta t) + (\beta t)^Q \epsilon \mathbb{P}(X > t) \\ &\leq \beta^Q \delta^{-Q} \left\| \max(Y, Z) \right\|_{L^{Q,w}}^Q + \beta^Q \epsilon \left\| X \right\|_{L^{Q,w}}^Q .\end{aligned}$$

Taking the supremum over t, we obtain

$$\|X\|_{L^{Q,w}}^Q \leq \beta^Q \delta^{-Q} \left\| \max(Y, Z) \right\|_{L^{Q,w}}^Q + \beta^Q \epsilon \left\| X \right\|_{L^{Q,w}}^Q .$$

If $\beta > 1$ is fixed, and δ is chosen small enough so that $\beta^Q \epsilon < 1$, this yields $\|X\|_{L^{Q,w}}^Q \leq C \left\| \max(Y, Z) \right\|_{L^{Q,w}}^Q$. Since $\left| \sum_0^P D_p \right| \leq X$ and $\|Y\|_{L^{Q,w}}^Q = \left\| Y^2 \right\|_{L^{Q/2,w}}^{Q/2}$, this proves the theorem. □

Proof of Theorem 6.1. We have $K - \mathbb{E}(K) = \sum D_p$, hence

$$\|K - \mathbb{E}(K)\|_{L^{2q-2,w}}^{2q-2} \leq C \left\| \sum \mathbb{E}(D_p^2|\mathcal{F}_{p+1}) \right\|_{L^{q-1,w}}^{q-1} + C \sum \left\| D_p \right\|_{L^{2q-2,w}}^{2q-2} .$$

For the first term, we use the inequality $\|\cdot\|_{L^{Q,w}} \leq \|\cdot\|_{L^Q}$. Therefore, this term is bounded by

$$C \mathbb{E} \left(\left[\sum_p \mathbb{E}(D_p^2|\mathcal{F}_{p+1}) \right]^{q-1} \right).$$

Since ϕ has a weak moment of order q, it has a strong moment of order 2. Therefore, (4.9) gives $\mathbb{E}(D_p^2|\mathcal{F}_{p+1}) \le \sum_{j\le p} c_{p-j}^{(0)} \operatorname{Lip}_j(K)^2$. Hence, the first term in the Rosenthal-Burkholder inequality is bounded by

$$C\left(\sum_p \sum_{j=0}^{p} \operatorname{Lip}_j(K)^2 c_{p-j}^{(0)}\right)^{q-1} \le C\left(\sum_j \operatorname{Lip}_j(K)^2\right)^{q-1}.$$

Let us now turn to $\|D_p\|_{L^{2q-2,w}}$. Integrating the estimates of Lemma 6.2, we get

$$\|D_p\|_{L^{2q-2,w}}^{2q-2} \le C\sum_{j\le p} \operatorname{Lip}_j(K)^{2q-2} c_{p-j}^{(0)} + C\left(\sum \operatorname{Lip}_j(K)^2\right)^{q-2} \sup_{h>0}\left(h^{-1}\sum_{j=p-h+1}^{p} \operatorname{Lip}_j(K)\right)^2. \quad (6.2)$$

We should sum those estimates over p. For the first sum, we obtain

$$\sum_j \operatorname{Lip}_j(K)^{2q-2}\sum_{p\ge j} c_{p-j}^{(0)} \le C\sum_j \operatorname{Lip}_j(K)^{2q-2} \le C\left(\sum_j \operatorname{Lip}_j(K)^2\right)^{q-1}.$$

For the second sum, let us define a function f on $\mathbb{Z}$ by $f(j) = \operatorname{Lip}_j(K)$. This function belongs to $\ell^2(\mathbb{Z})$. The corresponding maximal function $Mf(p) = \sup_{h>0}\frac{1}{2h+1}\sum_{j=p-h}^{p+h} f(j)$ also belongs to $\ell^2(\mathbb{Z})$ and satisfies $\|Mf\|_{\ell^2} \le C\|f\|_{\ell^2}$, by Hardy-Littlewood maximal inequality. In particular,

$$\sum_p \sup_{h>0}\left(h^{-1}\sum_{p-h+1}^{p} \operatorname{Lip}_j(K)\right)^2 \le C\sum_j \operatorname{Lip}_j(K)^2.$$

Therefore, the contribution of the second term in (6.2) is bounded by $C\left(\sum \operatorname{Lip}_j(K)^2\right)^{q-1}$. This concludes the proof of Theorem 6.1. □

Remark 6.4. In view of Theorems 5.1 and 6.1, it would seem natural to try to prove a weak polynomial concentration inequality in invertible systems with weak moment controls on the return time. We have not been able to prove such a statement.

7. Applications

In this section, we first give examples of dynamical systems satisfying an exponential concentration inequality or only a polynomial concentration inequality. We also give examples of systems satisfying a weak polynomial concentration inequality. Second, we present several applications of these inequalities to specific observables. We shall not attempt to be exhaustive. Previous results are found in [CMS02,CCS05b,CCRV09]. For instance, we strengthen the bounds obtained in [CCS05b] since for dynamical systems modeled by a uniform Young tower with exponential tails, we can now use an exponential concentration inequality instead of a polynomial concentration inequality with moment 2 as in [CCS05b]. For systems modeled by a non-uniform Young tower, only a polynomial concentration inequality with moment 2 was known for intermittent maps of the interval (under some restrictions on the parameter). We now have at our disposal an optimal polynomial concentration inequality for these maps, and more generally, for dynamical systems modeled by non-uniform Young towers with polynomial tails.

7.1. Examples of dynamical systems. There are well-known dynamical systems (X, T) which can be modeled by a uniform Young tower with exponential tails [You98]. Examples of invertible dynamical systems fitting this framework are for instance Axiom A attractors, Hénon attractors for Benedicks-Carleson parameters [BY00], piecewise hyperbolic maps like the Lozi attractor, some billiards with convex scatterers, etc. Such systems admit an SRB measure μ and there is an invertible uniform Young tower $(\Delta_{\mathbb{Z}}, \hat{T}_{\mathbb{Z}}, \hat{\mu}_{\mathbb{Z}})$ and a projection map $\pi : \Delta_{\mathbb{Z}} \to X$ such that $T \circ \pi = \pi \circ \hat{T}_{\mathbb{Z}}$ and $\mu = \hat{\mu}_{\mathbb{Z}} \circ \pi^{-1}$. In the non-invertible case, there is a non-invertible Young tower $(\Delta, \hat{T}, \hat{\mu})$ and a corresponding projection map. A non-invertible example is the quadratic family for Benedicks-Carleson parameters. In both cases, it can also be ensured that the projection map is contracting, i.e., $d(\pi x, \pi y) \leq \hat{d}_\beta(x, y)$ for every x, y in the same partition element. Here, $\hat{d}_\beta$ denotes the (unilateral or bilateral) symbolic distance in the tower given by $\hat{d}_\beta(x, y) = \beta^{s(x,y)}$ for some $\beta < 1$. In particular, if f is a bounded Lipschitz function on X, it lifts to a function $f \circ \pi$ which is Lipschitz in the tower. More generally, if f is Hölder continuous, then its lift is Lipschitz for $\hat{d}_\beta$ if β is close enough to 1. Therefore, all the results we proved in the previous sections for Lipschitz observables K have a counterpart about Hölder ones; we will not give further details in this direction and restrict to the Lipschitz situation for ease of exposition. We will also assume for simplicity that X is bounded.

Theorem 7.1. *Let (X, T) be a dynamical system modeled by a uniform Young tower with exponential tails and let μ be its SRB measure. There exists $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $K(x_0, \ldots, x_{n-1})$,*

$$\int e^{K(x,Tx,\ldots,T^{n-1}x)} \,\mathrm{d}\mu(x) \leq e^{\int K(x,\ldots,T^{n-1}x)\,\mathrm{d}\mu(x)} e^{C\sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2}. \tag{7.1}$$

This theorem is an obvious consequence of Theorem 3.4 in the invertible case and of Theorem 3.1 in the non-invertible case. Inequality (7.1) was previously known only for uniformly piecewise expanding maps of the interval and subshifts of finite type equipped with a Gibbs measure [CMS02]. Under the assumptions of the previous theorem, only a polynomial concentration with moment 2 had been proven [CCS05a].

An immediate consequence of (7.1) is the following inequality for upper deviations: for all $t > 0$ and for all $n \in \mathbb{N}$,

$$\mu\left\{x \in X : K(x, Tx, \ldots, T^{n-1}x) - \int K(y, \ldots, T^{n-1}y)\,\mathrm{d}\mu(y) > t\right\}$$
$$\leq e^{-\frac{t^2}{4C\sum_{j=0}^{n-1}\mathrm{Lip}_j(K)^2}}. \tag{7.2}$$

The same bound holds for lower deviations by applying (7.2) to $-K$.

Let us now consider dynamical systems modeled by a non-uniform Young tower with polynomial tails. In the invertible case, there is an invertible non-uniform Young tower $(\Delta_{\mathbb{Z}}, \hat{T}_{\mathbb{Z}}, \hat{\mu}_{\mathbb{Z}})$ and a projection map $\pi : \Delta_{\mathbb{Z}} \to X$, and the SRB measure is $\mu = \hat{\mu}_{\mathbb{Z}} \circ \pi^{-1}$, provided that $\sum \phi(\alpha)\hat{\mu}_{\mathbb{Z}}(\Delta_{\alpha,0}) < \infty$. If $\sum \phi(\alpha)^q \hat{\mu}_{\mathbb{Z}}(\Delta_{\alpha,0}) < \infty$, we shall simply say that the tower has L^q tails. Similarly, if $\sum_{\phi(\alpha)>n} \hat{\mu}_{\mathbb{Z}}(\Delta_{\alpha,0}) \leq Cn^{-q}$, we shall say that the tower has weak L^q tails. We can of course rephrase what we have just said in the non-invertible case.

Theorem 7.2. *Let (X, T) be a dynamical system modeled by a non-uniform Young tower with L^q tails, for some $q \geq 2$. Then T satisfies a polynomial concentration inequality with moment $2q-2$, i.e., there exists a constant $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$,*

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

Using Markov's inequality we get at once that, for any $t > 0$ and for any $n \in \mathbb{N}$,

$$\mu\left\{ x \in X : \left| K(x, Tx, \dots, T^{n-1}x) - \int K(y, \dots, T^{n-1}y)\,\mathrm{d}\mu(y) \right| > t \right\} \leq C\, \frac{\left(\sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2 \right)^{q-1}}{t^{2q-2}}. \tag{7.3}$$

If the tails are only in weak L^q, Theorem 6.1 shows that (7.3) still holds.

The fundamental example is an expanding map of the interval with an indifferent fixed point [You99]. For the sake of definiteness, we consider for $\alpha \in (0, 1)$ the so-called "intermittent" map $T : [0, 1] \to [0, 1]$ defined by

$$T(x) = \begin{cases} x(1 + 2^\alpha x^\alpha) & \text{if } 0 \leq x \leq 1/2, \\ 2x - 1 & \text{if } 1/2 < x \leq 1. \end{cases} \tag{7.4}$$

There is a unique absolutely continuous invariant probability measure $\mathrm{d}\mu(x) = h(x)\,\mathrm{d}x$ such that $h(x) \sim x^{-\alpha}$ as $x \to 0$. This map is modeled by a non-uniform Young tower $(\Delta, \hat{\mu})$ such that $\hat{\mu}\{\phi = n\} \sim C/n^{\frac{1}{\alpha}+1}$. The return time has a weak moment of order $1/\alpha$. Thus, for $\alpha \in (0, 1/2)$, the previous results yield:

Corollary 7.3. *Let T be the map (7.4) and μ be its absolutely continuous invariant probability measure. There exists a constant $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $K(x_0, \dots, x_{n-1})$,*

$$\mu\left\{ x \in X : \left| K(x, Tx, \dots, T^{n-1}x) - \int K(y, \dots, T^{n-1}y)\,\mathrm{d}\mu(y) \right| > t \right\} \leq C\, \frac{\left(\sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2 \right)^{1/\alpha-1}}{t^{\frac{2}{\alpha}-2}}.$$

This estimate readily gives bounds for the moments of order $q \neq 2/\alpha - 2$. Indeed, if Z is a random variable satisfying $\mathbb{P}(|Z| > t) \leq (A/t)^Q$, then using the formula $\mathbb{E}(|Z|^q) = \int qt^{q-1}\mathbb{P}(|Z| > t)\,\mathrm{d}t$ and the tail estimates, one gets

$$\mathbb{E}(|Z|^q) \leq \frac{Q}{Q-q} A^q \quad \text{for } q < Q,$$

and if Z is bounded

$$\mathbb{E}(|Z|^q) \leq \frac{q}{q-Q} A^Q \, \|Z\|_{L^\infty}^{q-Q} \quad \text{for } q > Q.$$

For $q < 2/\alpha - 2$, this generalizes to arbitrary separately Lipschitz functions of n variables the moment bounds obtained for ergodic sums of Lipschitz functions in [MN08] (while the moment bounds for $q > 2/\alpha - 2$ are apparently new, even for ergodic sums). On the other hand, we improve the result in [CCRV09] in two respects: first, we obtain a polynomial concentration inequality with moment 2 for any $\alpha \in (0, 1/2)$ instead of $(0, 4 - \sqrt{15})$; second, we also obtain a polynomial concentration inequality with a moment whose order is larger than 2 and depends on $\alpha \in (0, 1/2)$.

Remark 7.4. There is a difference between Theorems 4.1 (about strong moments) and 6.1 (about weak moments): in the former, the range of parameters is $q \geq 2$, while we require $q > 2$ in the latter. It turns out that Theorem 6.1 is *false* for $q = 2$, as testified by the map (7.4) with $\alpha = 1/2$. For such a map, if f is a Hölder function with $\int f \,\mathrm{d}\mu = 0$ and $f(0) \neq 0$, then $S_n f/\sqrt{n \log n}$ converges in distribution to a gaussian [Gou04a, p. 88]. If Theorem 6.1 were true for $q = 2$, we would have $\mu\{|S_n f| > t\} \leq Ct^{-2}n$, hence $\mu\{|S_n f/\sqrt{n \log n}| > t\} \leq Ct^{-2}(n \log n)^{-1} n \to 0$, implying that $S_n f/\sqrt{n \log n}$ tends in probability to 0 and giving a contradiction.

There are also invertible examples exhibiting an intermittent behavior, notably coming from billiards. Indeed, apart from the stadium billiard (with a weak moment of order 2 and therefore not covered by our results), Chernov and Zhang studied in [CZ05a, CZ05b] several classes of billiards for which the decay of correlations behaves like $O((\log n)^C / n^{1/\alpha - 1})$, for some parameter α that can be chosen freely in $(0, 1/2]$ and some $C > 0$. This decay rate is obtained by modeling those billiards by nonuniform invertible Young towers with well controlled tails. Therefore, we can apply Theorem 7.2 to those maps, yielding polynomial concentration inequalities for any exponent $p < 2/\alpha - 2$, just like in the above one-dimensional non-invertible situation.

7.2. Empirical covariance. For a Lipschitz observable f such that $\int f \,\mathrm{d}\mu = 0$, the auto-covariance of the process $\{f \circ T^k\}$ is defined as usual by

$$C_f(\ell) = \int f \cdot f \circ T^\ell \,\mathrm{d}\mu. \tag{7.5}$$

An obvious estimator for $C_f(\ell)$ is

$$\widehat{C}_f(n, \ell, x) = \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) f(T^{j+\ell} x).$$

We could as well consider the covariance between $\{f \circ T^k\}$ and $\{g \circ T^k\}$, for a pair of Lipschitz observables f, g. For each $\ell \geq 0$, the ergodic theorem tells us that $\widehat{C}_f(n, \ell, x) \to C_f(\ell)$ μ-almost surely, as $n \to \infty$. Considering the function of $n + \ell$ variables $K(x_0, \ldots, x_{n+\ell-1}) = \frac{1}{n}\sum_{j=0}^{n-1} f(x_j) f(x_{j+\ell})$, we obtain immediately (noting that $\int \widehat{C}_f(n, \ell, x)\, \mathrm{d}\mu(x) = C_f(\ell)$) the following theorems.

Theorem 7.5. *Let (X, T) be a dynamical system modeled by a uniform Young tower with exponential tails and μ its SRB measure. Let $f : X \to \mathbb{R}$ be a Lipschitz function with $\int f\, \mathrm{d}\mu = 0$. There exists a constant $c > 0$ such that, for any $n, \ell \in \mathbb{N}$ and for any $t > 0$,*

$$\mu\left\{x \in X : \left|\widehat{C}_f(n, \ell, x) - C_f(\ell)\right| > t\right\} \leq 2e^{-c\frac{n^2t^2}{n+\ell}}.$$

Theorem 7.6. *Let (X, T) be a dynamical system modeled by a non-uniform Young tower with weak L^q tails, for some $q \geq 2$, and μ its SRB measure. Let $f : X \to \mathbb{R}$ be a Lipschitz function with $\int f\, \mathrm{d}\mu = 0$. There exists a constant $c > 0$ such that, for any $n, \ell \in \mathbb{N}$ and for any $t > 0$,*

$$\mu\left\{x \in X : \left|\widehat{C}_f(n, \ell, x) - C_f(\ell)\right| > t\right\} \leq c\left(\frac{n+\ell}{n^2}\right)^{q-1} \frac{1}{t^{2q-2}}.$$

7.3. Empirical measure. Given $x \in X$ in an ergodic compact dynamical system (X, T, μ), let

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

be the associated empirical measure. By Birkhoff's ergodic theorem, $\mathcal{E}_n(x)$ vaguely converges to μ, for μ-almost every x. Our aim is to quantify the 'speed' at which this convergence takes place. We use the Kantorovich distance (compatible with vague convergence): for two probability measures μ_1, μ_2 on X, let

$$\mathrm{dist}_K(\mu_1, \mu_2) = \sup\left\{\int g\, \mathrm{d}\mu_1 - \int g\, \mathrm{d}\mu_2 : g : X \to \mathbb{R} \text{ is 1-Lipschitz}\right\}.$$

Set

$$\mathcal{D}_n(x) = \mathrm{dist}_K(\mathcal{E}_n(x), \mu).$$

We have the following general bounds.

Theorem 7.7. *Let (X, T) be a dynamical system modeled by a uniform Young tower with exponential tails and μ its SRB measure. Let $f : X \to \mathbb{R}$ be a Lipschitz function. There exists a constant $C > 0$ such that, for any $n \in \mathbb{N}$ and for any $t > 0$,*

$$\mu\left\{x \in X : \left|\mathcal{D}_n(x) - \int \mathcal{D}_n(y)\, \mathrm{d}\mu(y)\right| > \frac{t}{\sqrt{n}}\right\} \leq 2e^{-Ct^2}.$$

Theorem 7.8. *Let (X, T) be a dynamical system modeled by a non-uniform Young tower with weak L^q tails, for some $q \geq 2$, and μ its SRB measure. Let $f : X \to \mathbb{R}$ be a Lipschitz function. There exists a constant $C > 0$ such that, for all $n \in \mathbb{N}$ and all $t > 0$,*

$$\mu\left\{x \in X : \left|\mathcal{D}_n(x) - \int \mathcal{D}_n(y)\, \mathrm{d}\mu(y)\right| > \frac{t}{\sqrt{n}}\right\} \leq \frac{C}{t^{2q-2}}.$$

These bounds follow at once by applying either (7.2) or (7.3) to the function

$$K(x_0, \ldots, x_{n-1}) = \sup\left\{\frac{1}{n}\sum_{j=0}^{n-1} g(x_j) - \int g\, \mathrm{d}\mu : g : X \to \mathbb{R} \text{ is } 1-\text{Lipschitz}\right\}$$

whose Lipschitz constants are uniformly bounded by $1/n$. The natural next step is to seek for an upper bound for $\int \mathcal{D}_n(y)\, \mathrm{d}\mu(y)$. We are not able to obtain an *a priori* sufficiently good estimate unless we restrict to one-dimensional systems.

Corollary 7.9. *Let (X, T) be a one-dimensional dynamical system satisfying the assumptions of Theorem 7.7. There exist some constants $B, C > 0$ such that, for any $n \in \mathbb{N}$ and for any $t > 0$,*

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

Corollary 7.10. *Let (X, T) be a one-dimensional dynamical system satisfying the assumptions of Theorem 7.8. There exist some constants $B, C > 0$ such that, for any $n \in \mathbb{N}$ and for any $t > 0$,*

$$\mu\left\{x \in X : \mathcal{D}_n(x) > \frac{t}{n^{1/2}} + \frac{B}{n^{1/4}}\right\} \leq \frac{C}{t^{2q-2}}.$$

These two corollaries follow immediately if we can prove that there exists $B > 0$ such that, for any $n \in \mathbb{N}$,

$$\int \mathcal{D}_n\, \mathrm{d}\mu \leq \frac{B}{n^{1/4}}.$$

The proof is found in [CCS05b, Thm. 5.2]. The point is that in dimension one, there is a special representation of Kantorovich distance in terms of the distribution functions. The estimate then follows easily using the fact that the auto-covariance of Lipschitz observables is summable under the above assumptions.

For the map (7.4), we can use Corollary 7.3 to get the bound

$$\mu\left\{x \in X : \mathcal{D}_n(x) > \frac{t}{n^{1/2}} + \frac{B}{n^{1/4}}\right\} \leq \frac{C}{t^{\frac{2}{\alpha}-2}},$$

for any $n \in \mathbb{N}$ and for any $t > 0$.

Remark 7.11. What explains the power $1/4$ of n is the fact that at some stage, one has to approximate a characteristic function of a set by a Lipschitz function. If one can control the auto-covariance of functions with bounded variation, one gets

$$\int \mathcal{D}_n\, \mathrm{d}\mu \leq \frac{B}{\sqrt{n}}.$$

This is the case for uniformly piecewise expanding maps of the interval [CMS02]. This is also the case for the quadratic map with Benedicks-Carleson parameters [You92].

Since we proved that this system satisfies an exponential concentration inequality, we get

$$\mu\left\{x \in X : \mathcal{D}_n(x) > \frac{t}{\sqrt{n}}\right\} \le e^{-Ct^2},$$

for any $n \in \mathbb{N}$ and for any t greater than some $t_0 > 0$.

7.4. Kernel density estimation. The estimation from an orbit of the density h of the invariant measure of a one-dimensional dynamical system (X, T) is based on the estimator

$$h_n(s; x) = \frac{1}{na_n}\sum_{j=0}^{n-1} \psi\Big(\frac{s - T^j x}{a_n}\Big),$$

where a_n is a sequence of positive numbers going to 0 but such that na_n goes to ∞, and ψ is a 'kernel', that is, a non-negative Lipschitz function with compact support. We suppose that it is fixed in the sequel.

As proved in [CCS05a, App. C], the density of the invariant measure for a one-dimensional system modeled by a uniform Young tower with exponential tails has the following property: there exist some constants $B > 0$ and $\tau > 0$ such that

$$\int \big|h(s) - h(s-t)\big|\,\mathrm{d}s \le B|t|^\tau, \quad \forall t \in \mathbb{R}. \tag{7.6}$$

We have the following result about the L^1 convergence of empirical densities.

Theorem 7.12. *Let (X, T) be a one-dimensional dynamical system modeled by a uniform Young tower with exponential tails and μ its SRB measure. There exist $c_1, c_2 > 0$ such that, for any $t > c_1(a_n^\tau + 1/(\sqrt{n}a_n^2))$ and for any $n \in \mathbb{N}$,*

$$\mu\left\{x \in X : \int \big|h_n(s; x) - h(s)\big|\,\mathrm{d}s > t\right\} \le e^{-c_2 n a_n^2 t^2}.$$

The proof is similar to the proof of Theorem 5.2 in [CCS05a] except that we use an exponential concentration inequality instead of a polynomial concentration inequality with moment 2; hence we obtain a much stronger bound. (See also [CMS02, Thm. III.2] for uniformly piecewise expanding maps of the interval.) The property (7.6) is used to obtain an upper bound for $\int \big|h_n(s; x) - h(s)\big|\,\mathrm{d}s\,\mathrm{d}\mu$.

We do not know if the property (7.6) holds for the density of the invariant measure of all one-dimensional systems modeled by a non-uniform Young tower with polynomial tails. But for the special case of the intermittent map (7.4), it is easy to check that (7.6) is true with $\tau = 1 - \alpha$. Therefore, applying Corollary 7.3 we get the following result.

Theorem 7.13. *Let T be the map (7.4) and μ its absolutely continuous invariant probability measure. There exist $c_1, c_2 > 0$ such that for any $t > c_1(a_n^{1-\alpha} + 1/(\sqrt{n}a_n^2))$ and for any $n \in \mathbb{N}$,*

$$\mu\left\{x \in X : \int \big|h_n(s; x) - h(s)\big|\,\mathrm{d}s > t\right\} \le \frac{c_2}{n^{\frac{1}{\alpha}-1} a_n^{\frac{2}{\alpha}-2} t^{\frac{2}{\alpha}-2}}.$$

7.5. Tracing orbit properties. Let A be a measurable subset of X such that $\mu(A) > 0$ and define for all $n \in \mathbb{N}$,

$$\mathcal{S}_A(x,n) = \frac{1}{n} \inf_{y\in A} \sum_{j=0}^{n-1} d(T^j x, T^j y),$$

where d is the distance on X. This quantity, between 0 and 1, measures how well we can trace the orbit of some initial condition not in A by an orbit from an element of A.

Theorem 7.14. *Let (X, T) be a dynamical system modeled by a uniform Young tower with exponential tails and μ its SRB measure. There exist constants $c_1, c_2 > 0$ such that, for any measurable subset $A \subset X$ with $\mu(A) > 0$, for any $n \in \mathbb{N}$ and for any $t > 0$,*

$$\mu\left\{x \in X \,:\, \mathcal{S}_A(x,n) > c_1 \frac{\sqrt{\log n}}{\mu(A)\sqrt{n}} + \frac{t}{\sqrt{n}}\right\} \le e^{-c_2 t^2}.$$

Again, the proof is the same as [CMS02, Thm. IV.1] because it relies only on the exponential concentration inequality.

Theorem 7.15. *Let (X, T) be a dynamical system modeled by a non-uniform Young tower with weak L^q tails, for some $q \ge 2$, and μ its SRB measure. There exist constants $c_1, c_2 > 0$ such that, for any measurable subset $A \subset X$ with $\mu(A) > 0$, for any $n \in \mathbb{N}$ and for any $t > 0$,*

$$\mu\left\{x \in X \,:\, \mathcal{S}_A(x,n) > \frac{1}{n^{(q-1)/(2q-1)}}\left(t + \frac{c_1}{\mu(A)}\right)\right\} \le \frac{c_2}{n^{(q-1)/(2q-1)} t^{2q-2}}.$$

The proof follows the lines of that of [CMS02, Thm. IV.1] except that one uses the weak polynomial concentration inequality instead of the exponential concentration inequality as in the previous theorem.

For the intermittent maps (7.4), we can use Corollary 7.3. We get that there exist constants $c_1, c_2 > 0$ such that for any subset $A \subset [0, 1]$ with $\mu(A) > 0$, for any $n \in \mathbb{N}$ and for any $t > 0$,

$$\mu\left\{x \in [0,1] : \mathcal{S}_A(x,n) > \frac{1}{n^{(1/\alpha-1)/(2/\alpha-1)}}\left(t + \frac{c_1}{\mu(A)}\right)\right\} \le \frac{c_2}{n^{(\frac{1}{\alpha}-1)/(\frac{2}{\alpha}-1)} t^{\frac{2}{\alpha}-2}}.$$

We now formulate similar results for the number of mismatches at a given precision. Let A be a measurable subset of X such that $\mu(A) > 0$ and $\epsilon > 0$. For all $n \in \mathbb{N}$ define

$$\mathcal{M}_A(x,n,\epsilon) = \frac{1}{n} \inf_{y\in A} \mathrm{Card}\{0 \le j \le n-1 \,:\, d(T^j x, T^j y) > \epsilon\}.$$

We have the following result.

Theorem 7.16. *Let (X, T) be a dynamical system modeled by a Young tower with exponential tails and μ its SRB measure. There exist constants $c_1, c_2 > 0$ such that, if $A \subset X$ is such that $\mu(A) > 0$, then for any $0 < \epsilon < 1/2$, for any $n \in \mathbb{N}$ and for any $t > 0$,*

$$\mu\left\{x \in X : \mathcal{M}_A(x,n,\epsilon) > c_1 \epsilon^{-1} \frac{\sqrt{\log n}}{\mu(A)\sqrt{n}} + \frac{t\epsilon^{-1}}{\sqrt{n}}\right\} \le e^{-c_2 t^2}.$$

Theorem 7.17. *Let (X, T) be a dynamical system modeled by a non-uniform Young tower with weak L^q tails, for some $q \geq 2$, and μ its SRB measure. There exist constants $c_1, c_2 > 0$ such that, if $A \subset X$ is such that $\mu(A) > 0$, then for any $0 < \epsilon < 1/2$, for any $n \in \mathbb{N}$ and for any $t > 0$,*

$$\mu\left\{x \in X \,:\, \mathcal{M}_A(x, n, \epsilon) > \frac{1}{\epsilon^{(q-1)/(q-1/2)} n^{(q-1)/(2q-1)}} \left(t + \frac{c_1}{\mu(A)}\right)\right\} \leq \frac{c_2}{\epsilon^{(q-1)/(q-1/2)} n^{(q-1)/(2q-1)} t^{2q-2}}.$$

Once more, the proofs are almost the same as [CMS02, Thm. IV.2].

7.6. Integrated periodogram. Let (X, T, μ) be a dynamical system and $f : X \to \mathbb{R}$ be a Lipschitz function such that $\int f \,\mathrm{d}\mu = 0$. Define the empirical integrated periodogram function of the process $\{f \circ T^k\}_{k\geq 0}$ by

$$J_n(x, \omega) = \int_0^\omega \frac{1}{n} \Big| \sum_{j=0}^{n-1} e^{-\mathbf{i}js} f(T^j x) \Big|^2 \mathrm{d}s, \quad \omega \in [0, 2\pi].$$

Let

$$J(\omega) = C_f(0)\omega + 2 \sum_{\ell=1}^{\infty} \frac{\sin(\omega\ell)}{\ell} C_f(\ell),$$

where $C_f(\ell)$ is defined in (7.5).

Theorem 7.18. *Let (X, T) be a dynamical system modeled by a uniform Young tower with exponential tails and μ its SRB measure. Let $f : X \to \mathbb{R}$ be a Lipschitz function such that $\int f \,\mathrm{d}\mu = 0$. There exist some positive constants c_1, c_2 such that for any $n \in \mathbb{N}$ and for any $t > 0$,*

$$\mu\left\{x \in X \,:\, \sup_{\omega\in[0,2\pi]} \big|J_n(x, \omega) - J(\omega)\big| > t + \frac{c_1(1 + \log n)^{3/2}}{\sqrt{n}}\right\} \leq e^{-c_2 n t^2/(1+\log n)^2}.$$

The observable $\sup_{\omega\in[0,2\pi]} \big|J_n(x, \omega) - J(\omega)\big|$ was studied in [CCS05b] in the same setting but using the polynomial concentration inequality with moment 2. We get here a stronger result since we now have the exponential concentration inequality at hand.

Proof. Let

$$K(x_0, \ldots, x_{n-1}) = \sup_{\omega\in[0,2\pi]} \left| \int_0^\omega \frac{1}{n} \Big| \sum_{j=0}^{n-1} e^{-\mathbf{i}js} f(x_j) \Big|^2 \mathrm{d}s - J(\omega) \right|. \tag{7.7}$$

The reader can verify that

$$\sup_{0\leq\ell\leq n-1} \mathrm{Lip}_\ell(K) \leq \frac{c(1 + \log n)}{n} \tag{7.8}$$

for some constant $c > 0$. Let

$$Q_n(x) = \sup_{\omega\in[0,2\pi]} \left|J_n(x,\omega) - J(\omega)\right|. \tag{7.9}$$

The major task is to estimate from above $\int Q_n \, \mathrm{d}\mu$. We partly proceed as in [CCS05b, p. 2345]: We discretize ω, that is, given any integer $N \in \mathbb{N}$, we define the finite sequence of numbers (ω_p) by $\omega_p = 2\pi p/N$, $p = 0, \dots, N$. We then define

$$\overline{Q}_n(x) := \sup_{0\le p\le N} \left|J_n(x,\omega_p) - J(\omega_p)\right|.$$

One can then show that there exists some $C > 0$ such that

$$Q_n(x) \le \overline{Q}_n(x) + \frac{C}{N} \tag{7.10}$$

for all $x \in X$ and for all integers $n, N \in \mathbb{N}$.

We shall also use the fact (see [CCS05b] for more details) that there exists some $C > 0$ such that, for all ω and for any $n \in \mathbb{N}$,

$$\left|J(\omega) - \int J_n(x,\omega)\, \mathrm{d}\mu(x)\right| \le \frac{C}{n}. \tag{7.11}$$

We now depart from [CCS05b] and use that for any real $\beta > 0$,

$$\int e^{\beta\overline{Q}_n}\, \mathrm{d}\mu \le \sum_{p=0}^{N} \int e^{\beta[J_n(x,\omega_p)-J(\omega_p)]}\, \mathrm{d}\mu(x) + \sum_{p=0}^{N} \int e^{\beta[J(\omega_p)-J_n(x,\omega_p)]}\, \mathrm{d}\mu(x). \tag{7.12}$$

We estimate each term in the first sum of the right-hand side of this inequality by using the exponential concentration inequality (7.1), (7.8) and (7.11):

$$\begin{aligned}
&\int e^{\beta[J_n(x,\omega_p)-J(\omega_p)]}\, \mathrm{d}\mu(x) \\
&= \int e^{\beta[J_n(x,\omega_p)-\int J_n(y,\omega_p)\,\mathrm{d}\mu(y)]}\, \mathrm{d}\mu(x) \cdot e^{\beta[\int J_n(y,\omega_p)\,\mathrm{d}\mu(y) - J(\omega_p)]} \\
&\le e^{C\beta^2(1+\log n)^2/n} \cdot e^{C\beta/n}.
\end{aligned}$$

We get the same bound for each term in the second sum of the right-hand side of (7.12), hence

$$\int e^{\beta\overline{Q}_n}\, \mathrm{d}\mu \le 2(N+1)e^{C\beta^2(1+\log n)^2/n} \cdot e^{C\beta/n}.$$

We now use Jensen's inequality, (7.10) and (7.9) to get

$$\begin{aligned}
&\int \sup_{\omega\in[0,2\pi]} \left|J_n(x,\omega) - J(\omega)\right| \mathrm{d}\mu(x) \\
&\le \inf_{N\in\mathbb{N}} \left\{ \frac{1}{\beta}\log[2(N+1)] + C\beta\frac{(1+\log n)^2}{n} + \frac{C}{n} + \frac{C}{N} \right\}.
\end{aligned}$$

It remains to optimize over $N \in \mathbb{N}$ and $\beta > 0$ to obtain

$$\int \sup_{\omega\in[0,2\pi]} \left|J_n(x,\omega) - J(\omega)\right| \mathrm{d}\mu(x) \leq \frac{c_1(1+\log n)^{3/2}}{\sqrt{n}}.$$

We conclude the proof by applying (7.2) to the function (7.7), taking into account (7.8) and the previous estimate. □

Appendix A. A Technical Lemma

Our goal in this section is to prove a technical result that was required to obtain polynomial concentration estimates in non-uniform invertible Young towers. Let us consider a non-invertible non-uniform Young tower in which the return time has a moment of order $q \geq 2$ (i.e., $\sum h^q \mu\{x \in \Delta_0 : \phi(x) = h\} < \infty$). We define a function Φ_n by $\Phi_n(x) = \beta^{\mathrm{Card}\{j\in[1,n] : T^j x \in \Delta_0\}}$ for $x \in \Delta_0$, and $\Phi_n = 0$ otherwise, where $\beta < 1$ is fixed.

The estimate we need in (5.3) is given in the following theorem.

Theorem A.1. *For all nonnegative real numbers L_k,*

$$\int \left(\sum_r \left(\sum_{k\geq r} L_k \Phi_{k-r} \circ T^r\right)^2\right)^{q-1} \leq C\left(\sum L_k^2\right)^{q-1}.$$

For the proof, let us expand the square on the left, the resulting function is bounded by $\sum_r \sum_{k\geq\ell\geq r} L_k L_\ell \Phi_{k-r} \circ T^r$, since $\Phi_{\ell-r} \circ T^r \leq 1$. Bounding $L_k L_\ell$ by $L_k^2 + L_\ell^2$, we get two terms that will be studied separately (but with very similar techniques). The theorem follows from the following lemmas.

Lemma A.2. *We have*

$$\int \left(\sum_r \sum_{k\geq r} L_k^2 (k-r+1)\Phi_{k-r} \circ T^r\right)^{q-1} \leq C\left(\sum L_k^2\right)^{q-1}.$$

Lemma A.3. *We have*

$$\int \left(\sum_r \sum_{k\geq r} \sum_{\ell=r}^{k-1} L_\ell^2 \Phi_{k-r} \circ T^r\right)^{q-1} \leq C\left(\sum L_k^2\right)^{q-1}.$$

We will prove a more general result, encompassing those two lemmas and better suited to induction. We will need the following notion.

Definition A.4. *A weight system is a set of numbers $u(r,k)$ for $r < k$ such that*

1. *either $u(r,k) = M_k$ for all $r < k$,*
2. *or $u(r,k) = (\sum_{j=r}^{k-1} M_j)/(k-r)$ for all $r < k$,*

where M_k is a summable sequence of nonnegative real numbers. In both cases, let $\Sigma = \sum M_k$ be the sum of the weight system.

Weight systems satisfy the following property.

Lemma A.5. *Let $u(r,k)$ be a weight system. For all $m>0$, we have $\sum_r u(r,r+m)\leq\Sigma$.*

Proof. If $u(r,k)=M_k$, then $\sum u(r,r+m)=\sum M_{r+m}\leq\sum M_r=\Sigma$. If $u(r,k)=(\sum_{j=r}^{k-1}M_j)/(k-r)$, then

$$\sum u(r,r+m)=m^{-1}\sum_r\sum_{j=0}^{m-1}M_{r+j}\leq m^{-1}\sum_{j=0}^{m-1}\Sigma=\Sigma.\square$$

We will also need the following fact.

Lemma A.6. *Let $u(r,k)$ be a weight system with sum Σ, and let $c_n^{(1)}$ be a sequence with a moment of order 1. There exists a weight system $v(r,k)$ with sum at most $C\Sigma$ such that, for all $s<k$, we have $\sum_{r<s}u(r,k)c_{s-r}^{(1)}\leq v(s,k)$.*

Proof. Let $w(s,k)=\sum_{r<s}u(r,k)c_{s-r}^{(1)}$. If $u(r,k)$ is of the first type (i.e., $u(r,k)=M_k$), then $w(s,k)=\sum_{r<s}M_k c_{s-r}^{(1)}\leq CM_k$, and one can take $v(s,k)=CM_k$. If $u(r,k)$ is of the second type (i.e., $u(r,k)=(\sum_{j=r}^{k-1}M_j)/(k-r)$), then

$$\begin{aligned}w(s,k)&=\sum_{r<s}u(r,k)c_{s-r}^{(1)}=\sum_{r<s}\frac{1}{k-r}\left(\sum_{j=r}^{k-1}M_j\right)c_{s-r}^{(1)}\\&\leq\frac{1}{k-s}\left(\sum_{j<s}M_j\sum_{r\leq j}c_{s-r}^{(1)}+\sum_{j=s}^{k-1}M_j\sum_{r<s}c_{s-r}^{(1)}\right)\\&\leq\frac{1}{k-s}\left(\sum_{j<s}M_jc_{s-j}^{(0)}+C\sum_{j=s}^{k-1}M_j\right).\end{aligned}$$

Let $M_s'=CM_s+\sum_{j<s}M_jc_{s-j}^{(0)}$, we get $w(s,k)\leq\frac{1}{k-s}(M_s'+C\sum_{j=s+1}^{k-1}M_j)$, which is bounded by $\frac{1}{k-s}\sum_{j=s}^{k-1}M_j'$. Moreover, $\sum M_j'\leq C\sum M_j$ since the sequence $c_n^{(0)}$ is summable. This shows that w is bounded by a weight system v with sum at most $C\Sigma$. □

The main lemma is the following:

Lemma A.7. *Consider a weight system $u(r,k)$, and real numbers $\gamma\geq 1$ and $Q\geq 1$ with $\gamma Q\leq q-1$. We have*

$$\int\left(\sum_{k>r}u(r,k)(k-r)^\gamma\Phi_{k-r}\circ T^r\right)^Q\leq C\Sigma^Q.$$

This result implies Lemmas A.2 and A.3, using it with $\gamma=1$, $Q=q-1$ and the weights L_k^2 for the former, $(\sum_{\ell=r}^{k-1}L_\ell^2)/(k-r)$ for the latter.

We will prove the lemma directly for $Q\in[1,2]$, while an induction will be required for $Q>2$. When u is a weight system, let us write $S(\gamma,u)=\sum_{k>r}u(r,k)(k-$

$r)^{\gamma}\Phi_{k-r}\circ T^r$. We will construct another weight system $v(r,k)$ (with sum at most $C\Sigma$) such that

$$\int |S(\gamma,u)|^Q \leq C\Sigma^Q + C\Sigma^{Q/2}\int |S(2\gamma,v)|^{Q/2}.$$

By induction, the last integral is bounded by $C\Sigma^{Q/2}$, and we obtain the desired result.

Let us explain the strategy of the proof. First, since $\int \Phi_n \leq c_n^{(q-1)}$ by Lemma A.8 below, we have

$$\begin{aligned}\mathbb{E}(S(\gamma,u)) &\leq \sum_{k>r}(k-r)^{\gamma}u(r,k)c_{k-r}^{(q-1)} = \sum_m m^{\gamma}c_m^{(q-1)}\left(\sum_r u(r,r+m)\right)\\ &\leq \sum_m m^{\gamma}c_m^{(q-1)}\Sigma,\end{aligned}$$

by Lemma A.5. As $\gamma \leq \gamma Q \leq q-1$, the sum in m is finite, and we get $\mathbb{E}(S(\gamma,u)) \leq C\Sigma$. Consequently, to prove the lemma, it suffices to bound $\int |S(\gamma,u)-\mathbb{E}(S(\gamma,u))|^Q$.

We decompose $S=S(\gamma,u)$ as $\mathbb{E}(S)+\sum_{s\geq 0}S_s\circ T^s$, where $S_s\circ T^s$ is a sequence of reverse martingale differences: writing $\mathcal{F}_0$ for the Borel σ-algebra and $\mathcal{F}_s=T^{-s}\mathcal{F}_0$, the function $S_s\circ T^s$ is $\mathcal{F}_s$-measurable and $\mathbb{E}(S_s\circ T^s|\mathcal{F}_{s+1})=0$, i.e., $\mathbb{E}(S_s|\mathcal{F}_1)=0$. For any function f, one has $E(f|\mathcal{F}_s)=(\mathcal{L}^s f)\circ T^s$, where $\mathcal{L}$ is the transfer operator. Therefore, S_s is given by $S_s(z)=\mathcal{L}^s S(z)-\mathcal{L}^{s+1}S(Tz)$.

For $Q\in[1,2]$, the von Bahr-Esseen inequality [vBE65] yields

$$\int |S-\mathbb{E}(S)|^Q \leq \sum_s \mathbb{E}(|S_s|^Q|),$$

while for $Q>2$ the Rosenthal-Burkholder inequality gives an additional term as follows:

$$\int |S-\mathbb{E}(S)|^Q \leq \mathbb{E}\left(\sum_s \mathbb{E}(S_s^2|\mathcal{F}_1)\circ T^s\right)^{Q/2} + \sum_s \mathbb{E}(|S_s|^Q).$$

We will split each function S_s into several parts that will be estimated separately. Plugging those bounds into the inequalities of von Bahr-Esseen (for $Q\in[1,2]$) and Rosenthal-Burkholder (for $Q>2$) will give the desired result.

More precisely, if $h(x)\neq 0$, we have $\mathbb{E}(|S_s||\mathcal{F}_1)=0$ at the (unique) preimage of x and there is nothing to estimate. On the other hand, if $h(x)=0$ and if z is a preimage of x under T, we have

$$\begin{aligned}S_s(z)=\mathcal{L}^s S(z)-\mathcal{L}^{s+1}S(x) = \sum_{k>r}(k-r)^{\gamma}u(r,k)(&\mathcal{L}^s(\Phi_{k-r}\circ T^r)(z)\\ &-\mathcal{L}^{s+1}(\Phi_{k-r}\circ T^r)(x)).\end{aligned}$$

When estimating $\mathbb{E}(S_s^2|\mathcal{F}_1)$ or $\mathbb{E}(|S_s|^Q|\mathcal{F}_1)$, there is a contribution coming from $\mathcal{L}^{s+1}S(x)$ (involving a sum over $k>r$), and a contribution coming from the sum over the preimages z of x of $\mathcal{L}^s S(z)$ (involving a sum over z and over $k>r$). We will treat separately those contributions depending on the positions of k and r with respect to s and to $s-h$ (where h is the height of the preimage z of x one is considering). Let πz

be the projection of z in the basis of the tower. If $h \le s$, we have $\mathcal{L}^s S(z) = \mathcal{L}^{s-h} S(\pi z)$. (This is the interesting case: if $h > s$, then all the following estimates become easier, we will not indicate the trivial modifications to be done in this case.)

We will study separately the following cases:

(1) $k > r \ge s+1$, contribution of $\mathcal{L}^{s-h} S(\pi z) - \mathcal{L}^{s+1} S(x)$;
(2) $k > s+1 > r$, contribution solely of $\mathcal{L}^{s+1} S(x)$;
(3) $k > s-h$, $\min(s+1, k) > r$, contribution solely of $\mathcal{L}^{s-h} S(\pi z)$;
(4) $s+1 \ge k > s-h$, $r < k$, contribution solely of $\mathcal{L}^{s+1} S(x)$;
(5) $s-h \ge k > r$, contribution of $\mathcal{L}^{s-h} S(\pi z) - \mathcal{L}^{s+1} S(x)$.

We will treat separately those five contributions, and see that all of them satisfy the desired bounds. We will need very precise estimates on the transfer operator, given in the following lemma. We recall that the notation $d_n^{(Q)}$ indicates a nonincreasing sequence with a moment of order Q.

Lemma A.8. *We have $\int \Phi_m \le c_m^{(q-1)}$. For $h(z) = 0$, we have $\mathcal{L}^n \Phi_m(z) \le c_n^{(q)} \Phi_{m-n}(z)$ if $n \le m$, and*

$$\left| \mathcal{L}^n(\mathcal{L}^m \Phi_m)(z) - \sum_{b \le n} e(b, m) \right| \le \sum_{b=0}^{n} d_{n-b}^{(q-2)} \sum_{i=0}^{m} c_{b+m-i}^{(q)} c_i^{(q)},$$

where the scalar $e(b, m)$ only depends on b and m and is bounded by $\sum_{i=0}^m c_{b+m-i}^{(q)} c_i^{(q)}$.

The function Φ_m involves m iterates of the transformation. While the transfer operator is eliminating some number $n \le m$ of those iterates, the improvement in the estimates depends on n, and $m-n$ iterates remain ready to be used (under the form of Φ_{m-n}). Once all the variables are eliminated, $\mathcal{L}^n(\mathcal{L}^m \Phi_m)$ converges to the integral of Φ_m (which is equal to $\sum_{b \ge 0} e(b, m)$), with a more complicated error term whose precise form will play an important role later on.

Proof. Let us first assume $n \le m$. In this case, $\mathcal{L}^n \Phi_m(z) = \Phi_{m-n}(z) \cdot U_n 1(z)$, where the operator U_n was introduced in the proof of Lemma 4.4. We proved there that $\|U_n\| \le c_n^{(q)}$, the desired estimate follows.

For any point x with height $i \in [0, m]$, we obtain $\mathcal{L}^m \Phi_m(x) = \mathcal{L}^{m-i} \Phi_m(\pi x) \le c_{m-i}^{(q)}$. On the other hand, if $h(x) = i > m$, we have $\mathcal{L}^m \Phi_m(x) = \Phi_m(T^{-m} x) = 0$, since Φ_m vanishes on points with positive height by definition. Let $\Gamma = \mathcal{L}^m \Phi_m$.

We obtain

$$\int \Phi_m = \int \Gamma \le \sum_{i=0}^{m} \mu\{h = i\} c_{m-i}^{(q)} \le \sum_{i=0}^{m} c_i^{(q-1)} c_{m-i}^{(q)} \le c_m^{(q-1)}.$$

Let us now study $\mathcal{L}^n(\mathcal{L}^m \Phi_m) = \mathcal{L}^n \Gamma$, using the previous information regarding Γ. We will use the operators T_k and B_b that were introduced in Subsect. 4.2, so that $\mathcal{L}^n \Gamma(z) = \sum_{k+b=n} T_k B_b \Gamma(z)$ for $h(z) = 0$. We explained there that $T_k = \Pi + E_k$, where $\Pi f = (\int f) 1_{\Delta_0}$, and $\|E_k\| \le d_k^{(q-2)}$. Hence,

$$\mathcal{L}^n \Gamma(z) = \Pi \cdot \sum_{b \le n} B_b \Gamma + \sum_{k+b=n} E_k B_b \Gamma(z).$$

We estimate first $\|B_b \Gamma\|$. We have $B_b \Gamma(x) = \sum g^{(b)}(y) \Gamma(y)$, where we sum over the points $y \in T^{-b}(x)$ not returning to Δ_0 before time b. If $h(y) = i$, the point πy has a return time to the basis equal to $b+i$. Therefore, $|B_b\Gamma(x)| \leq \sum_{i=0}^m c_{b+i}^{(q)} c_{m-i}^{(q)} = \sum_{i=0}^m c_{b+m-i}^{(q)} c_i^{(q)}$ (in view of the bound on Γ at height i). The Lipschitz norm of $B_b\Gamma$ is estimated in the same way. Thus,

$$\sum_{k+b=n} \|E_k B_b \Gamma\| \leq \sum_{k+b=n} d_k^{(q-2)} \sum_{i=0}^m c_{b+m-i}^{(q)} c_i^{(q)}.$$

Finally, the statement of the lemma is satisfied letting $e(b,m) = \int B_b\Gamma = \Pi(B_b\Gamma)$. This scalar is independent of n and bounded by $\sum_{i=0}^m c_{b+m-i}^{(q)} c_i^{(q)}$. □

We will use the following simple remark. For $\kappa \geq 2$ and $x, y \geq 0$, we have $(x+y)^\kappa \leq x^\kappa + Cy(x+y)^{\kappa-1}$ (by Taylor's formula). By induction, this implies

$$\left(\sum_{i=1}^n x_i\right)^\kappa \leq C \sum_{i=1}^n x_i \cdot \left(\sum_{j=1}^i x_j\right)^{\kappa-1}. \tag{A.1}$$

A.1. The case $k > r \geq s+1$. When $k > r \geq s+1$, we have $\mathcal{L}^{s+1}(\Phi_{k-r} \circ T^r)(x) = \Phi_{k-r} \circ T^{r-s-1}(x)$, while $\mathcal{L}^{s-h}(\Phi_{k-r} \circ T^r)(\pi z) = \Phi_{k-r} \circ T^{r-s+h}(\pi z)$. Since $T^{h+1}(\pi z) = x$, those terms coincide, and their contribution to $S_s(z)$ vanishes.

A.2. The case $k > s+1 > r$, contribution of $\mathcal{L}^{s+1} S(x)$. The contribution from $\Phi_{k-r} \circ T^r$ satisfies

$$\mathcal{L}^{s+1}(\Phi_{k-r} \circ T^r) = \mathcal{L}^{s+1-r} \Phi_{k-r} \leq c_{s+1-r}^{(q)} \Phi_{k-s-1}(x),$$

by Lemma A.8. Summing those contributions to $S_s(z)$ (for varying k and r) gives a term which is bounded by

$$S_s^{(2)} = \sum_{k>s+1>r} (k-r)^\gamma u(r,k) c_{s+1-r}^{(q)} \Phi_{k-s-1}(x).$$

Let us note that this term does not depend on z. Since $k-r = (k-s-1)+(s+1-r) \leq 2(k-s-1)(s+1-r)$ and since $(s+1-r)^\gamma c_{s+1-r}^{(q)} \leq c_{s+1-r}^{(q-\gamma)}$, we have

$$S_s^{(2)} \leq \sum_{k>s+1} \sum_{r \leq s} u(r,k) c_{s+1-r}^{(q-\gamma)} (k-s-1)^\gamma \Phi_{k-s-1}(x).$$

By Lemma A.6, there exists a new weight system v such that $\sum_{r\leq s} u(r,k) c_{s+1-r}^{(q-\gamma)} \leq v(s+1,k)$, yielding $S_s^{(2)} \leq \sum_{k>s+1} v(s+1,k)(k-s-1)^\gamma \Phi_{k-s-1}(x)$. Moreover, the sum of the weight v is at most $C\Sigma$.

Let $\kappa \geq 1$, we estimate $|S_s^{(2)}(z)|^\kappa$. We apply the inequality (A.1) to $x_k = v(s+1,k)$ $(k-s-1)^\gamma \Phi_{k-s-1}$, yielding

$$|S_s^{(2)}|^\kappa \leq \sum_{k>s+1} v(s+1,k)(k-s-1)^\gamma \Phi_{k-s-1} \cdot \left(\sum_{s+1<\ell\leq k} v(s+1,\ell)(\ell-s-1)^\gamma\right)^{\kappa-1}.$$

We claim that the last sum is bounded by $C(k-s-1)^\gamma \Sigma$. Indeed, if the weight v is of the first type (i.e., $v(r,\ell)=M_\ell$), then we bound $(\ell-s-1)^\gamma$ by $(k-s-1)^\gamma$, to obtain $(k-s-1)^\gamma \sum_{\ell=s+2}^k M_\ell \le C(k-s-1)^\gamma \Sigma$. On the other hand, if v is of the second type (i.e., $v(r,\ell)=(\sum_{j=r}^{\ell-1} M_j)/(\ell-r)$), then the sum is bounded by

$$\sum_{\ell=s+2}^{k}\sum_{j=s+1}^{\ell-1} M_j(\ell-s-1)^{\gamma-1} \le (k-s-1)^{\gamma-1}\sum_{j=s+1}^{k-1} M_j(k-j)$$
$$\le (k-s-1)^{\gamma}\sum_{j=s+1}^{k-1} M_j \le (k-s-1)^\gamma \Sigma.$$

We have proved that, for all $\kappa \ge 1$,

$$|S_s^{(2)}|^\kappa \le C\sum_{k>s+1} v(s+1,k)(k-s-1)^{\kappa\gamma}\Phi_{k-s-1}\Sigma^{\kappa-1}. \tag{A.2}$$

Let us now assume that $Q\in[1,2]$, and let us consider the contribution of $S_s^{(2)}$ to the von Bahr-Esseen inequality. It is given by

$$\sum_s \mathbb{E}(|S_s^{(2)}|^Q) = \sum_s \mathbb{E}(\mathbb{E}(|S_s^{(2)}|^Q|\mathcal{F}_1))$$
$$\le \sum_s C\sum_{k>s+1} v(s+1,k)(k-s-1)^{Q\gamma}\mathbb{E}(\Phi_{k-s-1})\Sigma^{Q-1},$$

by (A.2). Since $\mathbb{E}(\Phi_{k-s-1}) \le c_{k-s-1}^{(q-1)}$, this can be written (letting $k=s+1+m$) as $\Sigma^{Q-1}\sum_m m^{Q\gamma}c_m^{(q-1)}\sum_s v(s+1,s+1+m)$. For fixed m, the sum $\sum_s v(s+1,s+1+m)$ is bounded by $C\Sigma$ by Lemma A.5. As $Q\gamma \le q-1$, $m^{Q\gamma}c_m^{(q-1)}$ is summable, and we obtain a bound $C\Sigma^Q$ as desired.

Assume now $Q>2$. In this case, the second term in the Rosenthal-Burkholder inequality is bounded by $C\Sigma^Q$ as above. Using (A.2) (with $\kappa=2$), the first term is at most

$$C\int\left(\sum_s\sum_{k>s+1} v(s+1,k)(k-s-1)^{2\gamma}\Phi_{k-s-1}\circ T^{s+1}\cdot\Sigma\right)^{Q/2}$$
$$= C\Sigma^{Q/2}\int |S(2\gamma,v)|^{Q/2}.$$

Since $\gamma'=2\gamma$ and $Q'=Q/2$ satisfy $\gamma'Q'\le q-1$, we can argue by induction to show that this term is again bounded by Σ^Q.

A.3. The case $k>s-h$, $\min(s+1,k)>r$, contribution of $\mathcal{L}^{s-h}S(\pi z)$. We should study $S_s^{(3)}(z)=\mathcal{L}^{s-h}(\sum_{k>s-h}\sum_{r\le\min(s,k-1)} u(r,k)(k-r)^\gamma\Phi_{k-r}\circ T^r)(\pi z)$.

If $k>s-h$ and $r\in(s-h,s]$ with $r<k$, we have $\mathcal{L}^{s-h}(\Phi_{k-r}\circ T^r)(\pi z)=\Phi_{k-r}\circ T^{r-(s-h)}(\pi z)$. Since the point $T^{r-(s-h)}(\pi z)$ has positive height, the function Φ_{k-r} vanishes here. Therefore, we only have to consider the contribution of $k>s-h\ge r$.

This is exactly the same thing as in the previous subsection, but for the point πz instead of x. The inequality (A.2) gives, for all $\kappa \geq 1$,

$$|S_s^{(3)}(z)|^\kappa \leq C \sum_{k>s-h} v(s-h,k)(k-s+h)^{\kappa\gamma} \Phi_{k-s+h}(\pi z)\Sigma^{\kappa-1},$$

where v is a weight system with sum at most $C\Sigma$. For $k \in (s-h, s+1]$, we simply bound $\Phi_{k-s+h}(\pi z)$ by 1, while for $k > s+1$ we bound it by $\Phi_{k-s-1}(x)$, since $T^{h+1}(\pi z) = x$. Summing over the preimages z of x, we get

$$\mathbb{E}(|S_s^{(3)}|^\kappa|\mathcal{F}_1) \leq C\Sigma^{\kappa-1}\sum_{h\geq 0} c_h^{(q)}\Bigg(\sum_{k=s-h+1}^{s+1} v(s-h,k)(k-s+h)^{\kappa\gamma} + \sum_{k>s+1} v(s-h,k)(k-s+h)^{\kappa\gamma}\Phi_{k-s-1}(x)\Bigg).$$

In the first sum, we bound $k-s+h$ by $h+1$ and we use the inequality $(h+1)^{\kappa\gamma} c_h^{(q)} \leq c_h^{(q-\kappa\gamma)}$. In the second sum, we have $c_h^{(q)}(k-s+h)^{\kappa\gamma} \leq c_h^{(q-\kappa\gamma)}(k-s-1)^{\kappa\gamma}$ by the same argument. If $\kappa\gamma \leq q-1$, the quantity $\sum_{h\geq 0} c_h^{(q-\kappa\gamma)} v(s-h,k)$ is bounded by $w(s+1,k)$, where w is a weight system with sum at most $C\Sigma$, by Lemma A.6. We obtain

$$\mathbb{E}(|S_s^{(3)}|^\kappa|\mathcal{F}_1) \leq C\Sigma^{\kappa-1}\Bigg(\sum_{h\geq 0}\sum_{k=s-h+1}^{s+1} c_h^{(q-\kappa\gamma)} v(s-h,k) + \sum_{k>s+1} w(s+1,k)(k-s-1)^{\kappa\gamma}\Phi_{k-s-1}(x)\Bigg). \tag{A.3}$$

The second term is identical to the term appearing in the previous subsection, in (A.2). It follows in the same way that its contribution to the inequalities of von Bahr-Esseen (case $Q \in [1,2]$) and Rosenthal-Burkholder (case $Q > 2$) is bounded by $C\Sigma^Q$.

Let us consider the first term, first in von Bahr-Esseen inequality (case $Q \in [1,2]$). Thanks to (A.3) (with $\kappa = Q$), its contribution is given by

$$\begin{aligned}
&\sum_s C\Sigma^{Q-1}\sum_{h\geq 0}\sum_{k=s-h+1}^{s+1} c_h^{(q-Q\gamma)} v(s-h,k)\\
&= C\Sigma^{Q-1}\sum_{h\geq 0} c_h^{(q-Q\gamma)}\sum_{m=1}^{h+1}\sum_s v(s-h,s-h+m)\\
&\leq C\Sigma^{Q-1}\sum_{h\geq 0} c_h^{(q-Q\gamma)}\sum_{m=1}^{h+1}\Sigma = C\Sigma^Q\sum_{h\geq 0} c_h^{(q-Q\gamma-1)},
\end{aligned}$$

where we used Lemma A.5 for the inequality. Since $Q\gamma \leq q-1$, this is bounded by $C\Sigma^Q$.

When $Q > 2$, we use the Rosenthal-Burkholder inequality. As above, the last term in this inequality is bounded by $C\Sigma^Q$. Using (A.3) (with $\kappa = 2$), the first term is bounded by

$$\left(\sum_s C\Sigma \sum_{h\geq 0} \sum_{k=s-h+1}^{s+1} c_h^{(q-2\gamma)} v(s-h,k) \right)^{Q/2}.$$

The same computation as above shows that this is bounded by $(C\Sigma^2)^{Q/2}$.

A.4. The case $s+1 \geq k > s-h$, $r < k$, contribution of $\mathcal{L}^{s+1}S(x)$. The contribution coming from $\Phi_{k-r} \circ T^r$ satisfies

$$\mathcal{L}^{s+1}(\Phi_{k-r} \circ T^r) = \mathcal{L}^{s+1-k}\mathcal{L}^{k-r}\Phi_{k-r},$$

which is controlled by Lemma A.8. Summing over $k \in [s-h+1, s+1]$ and $r < k$, we obtain that the resulting contribution $S_s^{(4)}$ is bounded by

$$\sum_{k=s-h+1}^{s+1} \sum_{r<k} u(r,k)(k-r)^\gamma \left(\sum_{b\leq s+1-k} \sum_{i=0}^{k-r} c_{b+k-r-i}^{(q)} c_i^{(q)} + \sum_{b\leq s+1-k} d_{s+1-k-b}^{(q-2)} \sum_{i=0}^{k-r} c_{b+k-r-i}^{(q)} c_i^{(q)} \right).$$

Since $d_{s+1-k-b}^{(q-2)}$ is bounded, the second term is bounded by the first one. Since $k-r \leq (b+k-r-i)+i$, we have $k-r \leq (b+k-r-i+1)(i+1)$, yielding $(k-r)^\gamma c_{b+k-r-i}^{(q)} c_i^{(q)} \leq c_{b+k-r-i}^{(q-\gamma)} c_i^{(q-\gamma)}$. For $\kappa \geq 1$, we obtain (letting $m = k-r$)

$$\mathbb{E}(|S_s^{(4)}|^\kappa | \mathcal{F}_1) \leq \sum_{h\geq 0} c_h^{(q)} \left(\sum_{k=s-h+1}^{s+1} \sum_{b\leq s+1-k} \sum_{i\geq 0} c_i^{(q-\gamma)} \sum_{m\geq i} u(k-m,k) c_{b+m-i}^{(q-\gamma)} \right)^\kappa.$$

Summing over s and using the inequality $\sum x_i^\kappa \leq (\sum x_i)^\kappa$, we get

$$\sum_s \mathbb{E}(|S_s^{(4)}|^\kappa | \mathcal{F}_1) \circ T^s$$
$$\leq \sum_{h\geq 0} c_h^{(q)} \left(\sum_s \sum_{k=s-h+1}^{s+1} \sum_{b\leq s+1-k} \sum_{i\geq 0} c_i^{(q-\gamma)} \sum_{m\geq i} u(k-m,k) c_{b+m-i}^{(q-\gamma)} \right)^\kappa.$$

We reorganize the sums as follows. First, we write $s+1 = k+a$ for some $a \in [0,h]$, so that the first three sums are replaced by $\sum_{a=0}^h \sum_k \sum_{b\leq a}$. Then, we move the sum over k to the end: since $\sum_k u(k-m,k) \leq \Sigma$ for all m by Lemma A.5, we get a bound

$$\Sigma^\kappa \sum_{h\geq 0} c_h^{(q)} \left(\sum_{a=0}^h \sum_{b\leq a} \sum_{i\geq 0} c_i^{(q-\gamma)} \sum_{m\geq i} c_{b+m-i}^{(q-\gamma)} \right)^\kappa.$$

The sum over $m \geq i$ is bounded by $d_b^{(q-\gamma-1)}$. The (finite) quantity $\sum_{i\geq 0} c_i^{(q-\gamma)}$ can be factorized out, giving a multiplicative constant. Since the sum $\sum_{b\leq a} d_b^{(q-\gamma-1)}$ is uniformly bounded, we get an upper bound $\Sigma^\kappa \sum_{h\geq 0}(h+1)^\kappa c_h^{(q)} \leq C\Sigma^\kappa$, when $\kappa \leq q$.

This readily implies that the contributions of $S_s^{(4)}$ to the inequalities of von Bahr-Esseen (case $1 \leq Q \leq 2$) and Rosenthal-Burkholder (case $Q > 2$) are bounded by Σ^Q, as desired.

A.5. The case $s - h \geq k > r$. The contribution coming from $\Phi_{k-r} \circ T^r$ reads

$$\begin{aligned}&\mathcal{L}^{s-h}(\Phi_{k-r} \circ T^r)(\pi z) - \mathcal{L}^{s+1}(\Phi_{k-r} \circ T^r)(x)\\&\quad = \mathcal{L}^{s-h-k}\mathcal{L}^{k-r}\Phi_{k-r}(\pi z) - \mathcal{L}^{s+1-k}\mathcal{L}^{k-r}\Phi_{k-r}(x).\end{aligned}$$

To estimate those contributions, we use Lemma A.8. The main terms $e(b, k-r)$ simplify partially: only those corresponding to $s-h-k < b \leq s+1-k$ remain. As a consequence, the global contribution $S_s^{(5)}(z)$ is bounded by

$$\sum_{s-h\geq k>r}(k-r)^\gamma u(r,k)\left(\sum_{b=s-h-k+1}^{s+1-k}\sum_{i=0}^{k-r} c_{b+k-r-i}^{(q)} c_i^{(q)} + \sum_{b=0}^{s-h-k} d_{s-h-k-b}^{(q-2)} \sum_{i=0}^{k-r} c_{b+k-r-i}^{(q)} c_i^{(q)}\right).$$

Let us first note that $(k-r)^\gamma c_{b+k-r-i}^{(q)} c_i^{(q)} \leq c_{b+k-r-i}^{(q-\gamma)} c_i^{(q-\gamma)}$ as in the previous subsection. We will then handle separately the two pieces $S_s^{(5.1)}(z)$ and $S_s^{(5.2)}(z)$ of this expression.

Summing over h and then over s, and using the inequality $\sum x_i^\kappa \leq (\sum x_i)^\kappa$ as in the previous subsection, we get

$$\begin{aligned}&\sum_s \mathbb{E}(|S_s^{(5.1)}|^\kappa|\mathcal{F}_1) \circ T^s\\&\quad \leq \sum_{h\geq 0} c_h^{(q)}\left(\sum_s\sum_{k\leq s-h}\sum_{b=s-h-k+1}^{s+1-k}\sum_{i\geq 0} c_i^{(q-\gamma)}\sum_{m\geq i} u(k-m,k) c_{b+m-i}^{(q-\gamma)}\right)^\kappa.\end{aligned}$$

Let us reorganize the sums essentially as in the previous subsection. First, let $s+1-h = k+a$ for some $a \geq 1$, so that the first sums become $\sum_{a\geq 1}\sum_k\sum_{b=a}^{a+h}$. Then, we move the sum over k to the end, and we use the inequality $\sum_k u(k-m,k) \leq \Sigma$ for all m. This yields a bound

$$\Sigma^\kappa \sum_{h\geq 0} c_h^{(q)}\left(\sum_{a\geq 1}\sum_{b=a}^{a+h}\sum_{i\geq 0} c_i^{(q-\gamma)}\sum_{m\geq i} c_{b+m-i}^{(q-\gamma)}\right)^\kappa.$$

The last sum over m is bounded by $d_b^{(q-\gamma-1)}$, which is independent of i. Therefore, we may factorize out the sum over i, since $\sum_i c_i^{(q-\gamma)} < \infty$. Since $d_b^{(q-\gamma-1)}$ is nonincreasing, we have $\sum_{b=a}^{a+h} d_b^{(q-\gamma-1)} \leq (h+1)d_a^{(q-\gamma-1)}$. As $q-\gamma-1 \geq 0$, the sequence $d_a^{(q-\gamma-1)}$ is summable, giving yet another multiplicative constant. We obtain a bound $C\Sigma^\kappa \sum_{h\geq 0}(h+1)^\kappa c_h^{(q)} \leq C\Sigma^\kappa$ when $\kappa \leq q$.

Let us now study $S_s^{(5.2)}(z)$. We have

$$\sum_s \mathbb{E}(|S_s^{(5.2)}|^\kappa|\mathcal{F}_1) \circ T^s$$

$$\le \sum_{h\ge 0} c_h^{(q)} \left(\sum_s \sum_{k\le s-h} \sum_{b=0}^{s-h-k} d_{s-h-k-b}^{(q-2)} \sum_{i\ge 0} c_i^{(q-\gamma)} \sum_{m\ge i} u(k-m,k) c_{b+m-i}^{(q-\gamma)} \right)^\kappa .$$

We proceed exactly as above, with the difference that the sum over b goes from 0 to $a-1$. We get a bound

$$C\Sigma^\kappa \sum_{h\ge 0} c_h^{(q)} \left(\sum_{a\ge 1} \sum_{b=0}^{a-1} d_{a-1-b}^{(q-2)} \cdot d_b^{(q-\gamma-1)} \right)^\kappa .$$

Since $q-\gamma-1 \le q-2$, the convolution between $d_{a-1-b}^{(q-2)}$ and $d_b^{(q-\gamma-1)}$ is bounded by $c_{a-1}^{(q-\gamma-1)}$. As $\gamma+1 \le q$, the sum over a is finite, and we obtain a bound Σ^κ.

Gluing the two pieces together, we have shown that $\sum_s \mathbb{E}(|S_s^{(5)}|^\kappa|\mathcal{F}_1) \circ T^s \le C\Sigma^\kappa$ for all $\kappa \le q$. This readily implies that the contributions of $S_s^{(5)}$ to the inequalities of von Bahr-Esseen (case $1 \le Q \le 2$) and Rosenthal-Burkholder (case $Q > 2$) are bounded by Σ^Q, as desired.

References

[BBLM05] Boucheron, S., Bousquet, O., Lugosi, G., Massart, P.: Moment inequalities for functions of independent random variables. Ann. Probab. **33**, 514–560 (2005)

[Bur73] Burkholder, D.L.: Distribution function inequalities for martingales. Ann. Probab. **1**, 19–42 (1973)

[BY00] Benedicks, M., Young, L.-S.: *Markov extensions and decay of correlations for certain Hénon maps*. Géométrie complexe et systèmes dynamiques (Orsay, 1995), Astérisque (2000), xi, pp. 13–56

[CCRV09] Chazottes, J.-R., Collet, P., Redig, F., Verbitskiy, E.: A concentration inequality for interval maps with an indifferent fixed point. Ergo. Th. Dynam. Sys. **29**, 1097–1117 (2009)

[CCS05a] Chazottes, J.-R., Collet, P., Schmitt, B.: Devroye inequality for a class of non-uniformly hyperbolic dynamical systems. Nonlinearity **18**, 2323–2340 (2005)

[CCS05b] Chazottes, J.-R., Collet, P., Schmitt, B.: Statistical consequences of devroye inequality for processes. applications to a class of non-uniformly hyperbolic dynamical systems. Nonlinearity **18**, 2341–2364 (2005)

[CMS02] Collet, P., Martínez, S., Schmitt, B.: Exponential inequalities for dynamical measures of expanding maps of the interval. Probab. Th. Rel. Fields **123**, 301–322 (2002)

[CZ05a] Chernov, N., Zhang, H.-K.: Billiards with polynomial mixing rates. Nonlinearity **18**, 1527–1553 (2005)

[CZ05b] Chernov, N., Zhang, H.-K.: A family of chaotic billiards with variable mixing rates. Stoch. Dyn. **5**, 535–553 (2005)

[Gou04a] Gouëzel, S.: Central limit theorem and stable laws for intermittent maps. Probab. Th. Rel. Fields **128**, 82–122 (2004)

[Gou04b] Gouëzel, S.: Sharp polynomial estimates for the decay of correlations. Israel J. Math. **139**, 29–65 (2004)

[Gou04c] Gouëzel, S.: *Vitesse de décorrélation et théorèmes limites pour les applications non uniformément dilatantes*. Ph.D. thesis, Université Paris Sud, 2004

[Led01] Ledoux, M.: *The concentration of measure phenomenon*. Mathematical Surveys and Monographs, vol. **89**. Providence, RI: Amer. Math. Soc., 2001

[MN08] Melbourne, I., Nicol, M.: Large deviations for nonuniformly hyperbolic systems. Trans. Amer. Math. Soc. **360**, 6661–6676 (2008)

[MS86] Milman, V.D., Schechtman, G.: *Asymptotic theory of finite-dimensional normed spaces*. Lecture Notes in Mathematics, Vol. **1200**, Berlin: Springer-Verlag, 1986, (With an appendix by M. Gromov)

[Rio00] Rio, E.: Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. C. R. Acad. Sci. Paris Sér. I Math. **330**, 905–908 (2000)

[Sar02] Sarig, O.: Subexponential decay of correlations. Invent. Math. **150**, 629–653 (2002)

[vBE65] von Bahr, B., Esseen, C.-G.: Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. Ann. Math. Statist **36**, 299–303 (1965)

[You92] Young, L.-S.: Decay of correlations for certain quadratic maps. Comm. Math. Phys. **146**, 123–138 (1992)

[You98] Young, L.-S.: Statistical properties of dynamical systems with some hyperbolicity. Ann. of Math. (2) **147**, 585–650 (1998)

[You99] Young, L.-S.: Recurrence times and rates of mixing. Israel J. Math. **110**, 153–188 (1999)

Communicated by G. Gallavotti