Moment bounds and concentration inequalities for slowly mixing dynamical systems[*]

Sébastien Gouëzel[†]   Ian Melbourne[‡]

Abstract

We obtain optimal moment bounds for Birkhoff sums, and optimal concentration inequalities, for a large class of slowly mixing dynamical systems, including those that admit anomalous diffusion in the form of a stable law or a central limit theorem with nonstandard scaling $(n \log n)^{1/2}$.

Keywords: moment bounds ; concentration ; dynamical systems ; martingales ; intermittent maps.
AMS MSC 2010: 37A25 ; 37A50 ; 60F15.
Submitted to EJP on April 2, 2014, final version accepted on September 29, 2014.
Supersedes arXiv:1404.0645v2.
Supersedes HAL:hal-00971289.

1 Statement of results

Consider a dynamical system T on a space X, preserving an ergodic probability measure μ. If x is distributed according to μ, the process $x, Tx, T^2x, \ldots$ on $X^{\mathbb{N}}$ is stationary, with distribution $\mu \otimes \delta_{Tx} \otimes \delta_{T^2x} \otimes \cdots$ (equivalently, one considers a Markov chain on X, with stationary measure μ, for which the transitions from x to Tx are deterministic). In particular, if f is a real-valued function on X, the real process $f(x), f(Tx), \ldots$ is also stationary. We would like to understand to what extent these processes behave like independent or weakly dependent processes: Although they are purely deterministic once the starting point x is fixed, one expects a random-like behaviour if the map T is sufficiently chaotic and the observable f is regular enough. In such a situation, the Birkhoff sums $S_n f = \sum_{i=0}^{n-1} f \circ T^i$ of Hölder continuous functions with zero average typically satisfy the central limit theorem, and grow like $\sqrt{n}$. On the other hand, the moments $\int |S_n f|^p \, d\mu$ may grow faster than $n^{p/2}$: it is possible that some subsets of X with small measure give a dominating contribution to those moments. Estimating the precise growth rate is important from the point of view of large deviations. It turns

[†]IRMAR, CNRS UMR 6625, Université de Rennes 1, 35042 Rennes, France.
 E-mail: sebastien.gouezel@univ-rennes1.fr
[‡]Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK.
 E-mail: I.Melbourne@warwick.ac.uk

out that this precise growth rate depends on finer characteristics of the system, and displays a transition at some critical exponent p^* directly related to the lack of uniform hyperbolicity of the system.

The situation for uniformly expanding/hyperbolic (Axiom A) systems is easily described: all moments $\int |S_n f|^p \, \mathrm{d}\mu$ grow like $n^{p/2}$ and moreover $\int |n^{-1/2} S_n f|^p \, \mathrm{d}\mu$ converges to the p'th moment of the limiting Gaussian in the central limit theorem. [23] showed that convergence of all moments holds also for nonuniformly expanding/hyperbolic diffeomorphisms modelled by Young towers with exponential tails [30]. However, it follows from [21, 23] that the situation is quite different for systems modelled by Young towers with polynomial tails [31].

In this paper, we give optimal bounds for all moments of Birkhoff sums (by optimal, we mean that we have upper and lower bounds of the same order of magnitude), in the situation of Young towers. Many real systems are quotients of such Young towers, hence our bounds apply to such systems, including notably intermittent maps of the interval [19, 25]. See for instance [21] for a discussion of such applications. Our techniques also give a generalization of moment inequalities, to concentration inequalities (see [10] for a discussion of numerous applications of such bounds). By the methods in [10, 23], all results described here pass over to the situation of invertible systems and flows; for brevity we present the results only for noninvertible discrete time dynamics.

We formulate our results in the abstract setting of Young towers. To illustrate this setting, let us start with a more concrete example, intermittent maps, i.e., maps of the interval which are uniformly expanding away from an indifferent fixed point (see [31] for more details). For $\gamma \in (0, 1)$, consider for instance the corresponding Liverani-Saussol-Vaienti map $T_\gamma : [0, 1] \to [0, 1]$ given by

$$
T_\gamma(x) = \begin{cases} x(1 + 2^\gamma x^\gamma) & \text{if } x < 1/2; \\ 2x - 1 & \text{if } x \geqslant 1/2. \end{cases}
$$

The first return map to the subinterval $Y = [1/2, 1]$ is uniformly expanding and Markov. Define a new space $X = \{(x, i) \ : \ x \in [1/2, 1], i < \varphi(x)\}$ where $\varphi : Y \to \mathbb{N}^*$ is the first return time to Y, i.e., $\varphi(x) = \inf\{i > 0 \ : \ T_\gamma^i(x) \in Y\}$. On this new space, we define a dynamics by $T(x, i) = (x, i+1)$ if $i + 1 < \varphi(x)$, and $T(x, \varphi(x) - 1) = (T_\gamma^{\varphi(x)}(x), 0)$. We think of X as a tower, where the dynamics T is trivial when one climbs up while it has a large expansion when one comes back to the bottom of the tower. The point of this construction is that the combinatorics of T are simpler than those of the original map T_γ, while the essential features of T and T_γ are the same. More precisely, the two maps are semiconjugate: the projection $\pi : X \to [0, 1]$ given by $\pi(x, i) = T_\gamma^i(x)$ satisfies $T_\gamma \circ \pi = \pi \circ T$. Hence, results for the decay of correlations, or growth of moments, or concentration, for T readily imply corresponding results for T_γ. This situation is not specific to the maps T_γ: many concrete maps can be modelled by Young towers in the same way (although the Young tower is usually not as explicit as in this particular example).

Let us give a more formal definition. A Young tower is a space X endowed with a partition $\bigcup_\alpha \bigcup_{0 \leqslant i < h_\alpha} \Delta_{\alpha, i}$ (where α belongs to some countable set, and h_α are positive integers), a probability measure μ and a map T preserving μ. The dynamics T maps bijectively $\Delta_{\alpha, i}$ to $\Delta_{\alpha, i+1}$ for $i + 1 < h_\alpha$, and $\Delta_{\alpha, h_\alpha - 1}$ to $\Delta_0 = \bigcup \Delta_{\alpha, 0}$: the dynamics goes up while not at the top of the tower, and then comes back surjectively to the basis. The distance on X is defined by $d(x, y) = \rho^{s(x, y)}$ where $\rho < 1$ is fixed and $s(x, y)$, the separation time, is the number of returns to the basis before the iterates of the points x and y are not in the same element of the partition. Finally, we require a technical

distortion condition: Denoting by $g(x)$ the inverse of the jacobian of T for the measure μ, we assume that $|\log g(x) - \log g(y)| \leqslant Cd(Tx, Ty)$ for all x, y in the same partition element.

With the distance d, the map T is an isometry while going up the tower, and expands by a factor $\rho^{-1} > 1$ when going back to the basis: it is non-uniformly expanding, the time to wait before seeing the expansion being large on points in $\Delta_{\alpha,0}$ with h_α large. In particular, denoting by $\varphi(x)$ the return time to the basis, the quantities

$$\mathrm{tail}_n = \mu\{x \in \Delta_0 \ : \ \varphi(x) \geqslant n\} = \mu\left(\bigcup_{h_\alpha \geqslant n} \Delta_{\alpha,0}\right),$$

called the *tails of the return time*, dictate the statistical properties of the transformation T. By Kac's Formula, tail_n is summable since μ is finite by assumption. Various kinds of behaviour can happen for tail_n. For instance, in the case of intermittent maps of parameter $\gamma \in (0, 1)$, one has $\mathrm{tail}_n \sim C/n^{1/\gamma}$. In general, if $\mathrm{tail}_n = O(n^{-q})$ for some $q > 1$, then Lipschitz functions mix at a speed $O(n^{-(q-1)})$ by [31], and this speed is optimal, see [26] and [15]. If $q > 2$, then $n^{-1/2}S_n f$ converges in distribution to a Gaussian, and the variance is nonzero provided f is not a coboundary. (More generally, for convergence to a Gaussian it suffices that the return time function φ is square-integrable, i.e., $n\,\mathrm{tail}_n$ is summable.) When $q \in (1, 2]$, more precise information is required on tail_n, leading to the following result.

Theorem 1.1 ([14]). *Consider a Young tower with* $\mathrm{tail}_n \sim Cn^{-q}$ *for some* $q > 1$. *There is a sequence* a_n, *and a nonempty set* $\mathcal{U}$ *in the space of Lipschitz functions* $f : X \to \mathbb{R}$ *with mean zero, such that the following holds. For each* $f \in \mathcal{U}$, *there exists a nondegenerate law* Z *such that* $a_n^{-1}S_n f \to_d Z$. *Moreover,* a_n *and* Z *are given as follows:*

$q > 2$:	$a_n = n^{1/2}$,	Z *is Gaussian.*
$q = 2$:	$a_n = (n\log n)^{1/2}$,	Z *is Gaussian.*
$q \in (1, 2)$:	$a_n = n^{1/q}$,	Z *is a stable law of index* q.

The set $\mathcal{U}$ is rather big: it contains for instance all the functions that converge to a nonzero constant along points whose height in the tower tends to infinity.

Lower bounds for the growth of moments are well-known (see [21]) and can be summarized in the following proposition. We write $\|u\|$ for the Lipschitz norm of a function u, given by

$$\|u\| = \sup_x |u(x)| + \sup_{x,y} \frac{|u(x) - u(y)|}{d(x,y)},$$

where the supremum in the second term is restricted to those x and y that belong to the same partition element. Note that, changing the parameter ρ in the definition of the distance, Hölder functions for the old distance become Lipschitz functions for the new one. Hence, all results that are stated in this paper for Lipschitz functions also apply to Hölder functions.

Proposition 1.2. *Consider a Young tower with* $\mathrm{tail}_n \sim Cn^{-q}$ *for some* $q > 1$. *Then, for all* $p \in [1, \infty)$, *there exists* $c > 0$ *such that for all* $n \geqslant 1$

$$\sup_{\|f\| \leqslant 1, \int f \, \mathrm{d}\mu = 0} \int |S_n f|^p \, \mathrm{d}\mu \geqslant \begin{cases} c \max(n^{p/2}, n^{p-q+1}) & \text{if } q > 2, \\ c \max((n\log n)^{p/2}, n^{p-q+1}) & \text{if } q = 2, \\ c \max(n^{p/q}, n^{p-q+1}) & \text{if } q < 2. \end{cases}$$

The phase transition in these lower bounds happens at $p^* = 2q - 2$ for $q \geqslant 2$, and at $p^* = q$ for $q \leqslant 2$. Before this threshold, the first lower bound (that corresponds to an

average behavior over the whole space) is more important, while the second one (that corresponds to the Birkhoff sum being large on a small part of the space) is dominating afterwards.

Proof. For the lower bound n^{p-q+1}, we take f that is equal to 1 on $\bigcup_{h_\alpha \geqslant n} \bigcup_{i < h_\alpha} \Delta_{\alpha,i}$, and equal to another constant on the complement of this set, to make sure that $\int f \, \mathrm{d}\mu = 0$. Then $S_n f = n$ on $\bigcup_{h_\alpha \geqslant 2n} \bigcup_{i < h_\alpha/2} \Delta_{\alpha,i}$, hence

$$\int |S_n f|^p \, \mathrm{d}\mu \geqslant n^p \mu \left(\bigcup_{h_\alpha \geqslant 2n} \bigcup_{i < h_\alpha/2} \Delta_{\alpha,i} \right) = n^p \sum_{h \geqslant 2n} \frac{h}{2} \mu(\varphi = h).$$

Using a discrete integration by parts and the assumption $\mu(\varphi \geqslant n) \sim Cn^{-q}$, one checks that this is $\geqslant cn^{p-q+1}$.

For the other bound, we fix a mean zero Lipschitz function f in the set $\mathcal{U}$ constructed in Theorem 1.1. This theorem shows the existence of a_n and Z nondegenerate such that $a_n^{-1} S_n f \to_d Z$. Hence $a_n^{-p} \int |S_n f|^p \, \mathrm{d}\mu$ is bounded from below and we get the lower bound ca_n^p in all three cases. □

In the case $q < 2$ and $p = q$, the lower bound in the proposition is $\int |S_n f|^q \, \mathrm{d}\mu \geqslant cn$. It is not sharp: for $f \in \mathcal{U}$, $S_n f / n^{1/q}$ converges to a stable law Z of index q, whose q-th moment is infinite, hence $\int |S_n f / n^{1/q}|^q \, \mathrm{d}\mu$ tends to infinity. To get a better lower bound, one should study the speed of convergence of $S_n f / n^{1/q}$ to Z. We can do this under stronger assumptions on the tails (this is not surprising since it is well known that the speed of convergence to stable laws is related to regularity assumptions on the tails of the random variables):

Proposition 1.3. *Consider a Young tower with* $\mathrm{tail}_n = Cn^{-q} + O(n^{-q-\varepsilon})$ *for some* $q \in (1,2)$ *and some* $\varepsilon > 0$*. Then there exists* $c > 0$ *such that for all* $n > 0$

$$\sup_{\|f\| \leqslant 1, \int f \, \mathrm{d}\mu = 0} \int |S_n f|^q \, \mathrm{d}\mu \geqslant cn \log n.$$

This lower bound is considerably more complicated to establish than the ones in Proposition 1.2. Since the arguments are rather different from the rest of the paper (essentially, they reduce to a proof of a Berry-Esseen like bound for $S_n f / n^{1/q}$), we defer the proof of the proposition to Appendix A. The assumptions of this proposition are for instance satisfied for the classical Pomeau-Manneville intermittent maps [19, 25]. (See for example [22, Proposition 11.12].)

For $q = 2$, the bound $c \max((n \log n)^{p/2}, n^{p-q+1})$ is known to be optimal for all p, see Remarks 1.6 and 1.7 below. Also, for $q > 2$, the bound $c \max(n^{p/2}, n^{p-q+1})$ is known to be optimal for all $p \neq 2q - 2$. The remaining cases are much more subtle, and are solved for the first time in this paper. We note that for $q > 2$ and $p = 2q - 2$, [10] obtains an additional upper bound for the weak moment of $S_n f$, which implies for $p > 2q - 2$ the upper bound Cn^{p-q+1}, in accordance with the lower bound. Moreover, the very precise methods of [10] seemed to indicate that the upper bound for the weak moment at $p = 2q - 2$ was optimal, and that the discrepancy with the lower bound was due to a suboptimality of the (naive) lower bound. We prove below that this is not the case.

Theorem 1.4. *Consider a Young tower with* $\mathrm{tail}_n = O(n^{-q})$ *for some* $q > 1$*. Then, for all* $p \in [1, \infty)$*, there exists* $C > 0$ *such that for any Lipschitz function* f *with* $\|f\| \leqslant 1$ *and*

$\int f \, \mathrm{d}\mu = 0$, *for all* $n \geqslant 0$,

$$\int |S_n f|^p \, \mathrm{d}\mu \leqslant \begin{cases} C \max(n^{p/2}, n^{p-q+1}) & \text{if } q > 2, \\ C \max((n \log n)^{p/2}, n^{p-q+1}) & \text{if } q = 2, \\ C \max(n^{p/q}, n^{p-q+1}) & \text{if } q < 2 \text{ and } p \neq q. \end{cases}$$

If $q < 2$, we have for all $t > 0$

$$\mu\{x \,:\, |S_n f(x)| \geqslant t\} \leqslant C t^{-q} n \tag{1.1}$$

and therefore

$$\int |S_n f|^q \, \mathrm{d}\mu \leqslant C n \log n.$$

Our upper bounds all match the lower bounds given in Propositions 1.2 and 1.3, and are therefore optimal.

Note that, in the proofs, if is sufficient to understand what happens at the critical exponent $p^* = 2q - 2$ for $q > 2$: a control on the L^{2q-2}-norm for $q > 2$ readily implies the control for any $p \in [1, \infty)$ thanks to the trivial inequalities

$$\|u\|_{L^p}^p \leqslant \|u\|_{L^\infty}^{p-r} \|u\|_{L^r}^r \text{ if } p > r, \quad \|u\|_{L^p} \leqslant \|u\|_{L^r} \text{ if } p < r. \tag{1.2}$$

In the same way, for $q < 2$, the control (1.1) on the weak q-th moment implies the corresponding L^p controls for any $p \in [1, \infty)$ thanks to the equality

$$\int |u|^p \, \mathrm{d}\mu = p \int_{s=0}^{\|u\|_{L^\infty}} s^{p-1} \mu\{|u| > s\} \, \mathrm{d}s. \tag{1.3}$$

This formula would also apply in the $q > 2$ case (combined with the control of $\mu\{|u| > s\}$ coming from the estimate at the exponent p^* and the Markov inequality), but it gives worse constants than (1.2) in this case.

On the other hand, for $q = 2$, the bound $\sqrt{n \log n}$ for the second moment does not give the desired upper bound for $p > 2$ (using the formulas (1.2) or (1.3), one only gets the upper bound $\int |S_n f|^p \, \mathrm{d}\mu \leqslant C n^{p-1} \log n$, with an extra $\log n$).

As an immediate consequence of the bounds on moments at the critical exponent, we obtain convergence of moments for all lower exponents.

Corollary 1.5. *Consider a Young tower with $\mathrm{tail}_n \sim C n^{-q}$ for some $q > 1$. Suppose that f, a_n and Z are as in Theorem 1.1. Then $\int |a_n^{-1} S_n f|^p \, \mathrm{d}\mu \to \mathbb{E}(|Z|^p)$ for all $p < p^*$ where $p^* = 2q - 2$ for $q \geq 2$ and $p^* = q$ for $q \in (1, 2)$.*

In particular, there exist nonzero constants $C = C_{p,q}$ such that

- *if $q > 2$, then $\int |S_n f|^p \, \mathrm{d}\mu \sim C n^{p/2}$ for all $p < 2q - 2$.*
- *if $q = 2$, then $\int |S_n f|^p \, \mathrm{d}\mu \sim C(n \log n)^{p/2}$ for all $p < 2$.*
- *if $q \in (1, 2)$, then $\int |S_n f|^p \, \mathrm{d}\mu \sim C n^{p/q}$ for all $p < q$.*

Proof. As in [23], this is an immediate consequence of Theorem 1.1, together with the fact that $\int |a_n^{-1/2} S_n f|^{p'} \, \mathrm{d}\mu$ is bounded for any $p' \in (p, p^*)$ as guaranteed by Theorem 1.4 (for $q \geqslant 2$, one can even take $p' = p^*$). $\qquad\square$

Remark 1.6. *Previously, no results were available on convergence of moments for $q < 2$. The case $q > 2$ in Corollary 1.5 recovers a result of [23] and the result for $q = 2$ was obtained by [5] in the context of dispersing billiards with cusps. [5] consider also the critical exponent $p = 2$ for dispersing billiards with cusps, and prove for this example that the limiting second moment is twice the moment of the limiting Gaussian: $\int |(n \log n)^{-1/2} S_n f|^2 \, \mathrm{d}\mu \to 2 \mathbb{E}(|Z|^2)$. This particular behaviour is due to the very specific geometric structure of the billiard.*

Remark 1.7. *Certain aspects of Theorem 1.4 and Corollary 1.5 do not require the full strength of the assumption that there is an underlying Young tower structure. We can consider the more general situation where f is a mean zero observable lying in L^∞ such that $\left|\int f\, g \circ T^n \,\mathrm{d}\mu\right| \leq C\|g\|_{L^\infty} n^{-(q-1)}$ for all $g \in L^\infty$, $n \geq 1$. (Such a condition is satisfied for f Lipschitz when X is a Young tower with $\mathrm{tail}_n = O(n^{-q})$.)*

In the case $q = 2$, this weaker condition is sufficient to recover all the moment estimates (and hence the convergence of moments for $p < 2$) described above. By [20, Lemma 2.1], $\int |S_n f|^p \,\mathrm{d}\mu \ll n^{p-1}$ for $p > 2$ and $\int |S_n f|^2 \,\mathrm{d}\mu \ll n \log n$.

In the case $q > 2$, $\int |S_n f|^p \,\mathrm{d}\mu \ll n^{p-q+1}$ for $p > 2q - 2$ by [20, Lemma 2.1], and $\int |S_n f|^p \,\mathrm{d}\mu \leq C n^{p/2}$ for $p < 2q - 2$ by [23], Again it follows that all moments converge for $p < 2q - 2$.

After we completed this article, we learned that, using techniques that are completely different from the ones we develop, Dedecker and Merlevède [12] also obtain the controls on moments given in Theorem 1.4, essentially under an assumption of the form $\left|\int f\, g \circ T^n \,\mathrm{d}\mu\right| \leq C\|g\|_{L^\infty} n^{-(q-1)}$. Their arguments (initially developed to control the behavior of the empirical measure) rely on general probabilistic inequalities for sums of random variables, and can apparently not give the concentration inequalities of Theorem 1.9 below.

Remark 1.8. *Proposition 1.2 and Theorem 1.4 clarify certain results in the Physics literature. As in [23], our results go over to flows, and apply in particular to infinite horizon planar periodic Lorentz gases. These can be viewed as suspension flows over Young towers with $\mathrm{tail}_n \sim Cn^{-2}$ so we are in the case $q = 2$. In particular, if $r(t)$ denotes position at time t, then $(t \log t)^{-1/2} r(t) \to_d Z$ where Z is a nondegenerate Gaussian [28]. [2] consider growth rate of moments for $r(t)$, but neglecting logarithmic factors. Defining $\gamma_p = \lim_{t \to \infty} \log \int |r(t)|^p \,\mathrm{d}\mu / \log t$, they argue heuristically that $\gamma_p = \max\{p/2, p-1\}$ in accordance with our main results. [11] conducted numerical simulations to verify the growth rates of the moments, including logarithmic factors, but based on the belief that $\int |r(t)|^p \,\mathrm{d}\mu$ scales like $(n \log n)^{p/2}$ for all p, whereas we have shown that this is correct only for $p \leq 2$.*

Two other examples of billiards that are modelled by Young towers with $\mathrm{tail}_n \sim Cn^{-2}$ are Bunimovich stadia (discrete and continuous time) [6] and billiards with cusps (discrete time) [4, 5]. Again, our results apply to these situations with $q = 2$.

The above optimal upper bounds for moments, dealing with Birkhoff sums, can be extended to concentration estimates, for any (possibly non-linear) function of the point and its iterates. More precisely, consider a function $K(x_0, x_1, \dots)$ (depending on finitely or infinitely many coordinates) which is separately Lipschitz: for all i, there exists a constant $\mathrm{Lip}_i(K)$ such that, for all $x_0, x_1, \dots$ and x_i',

$$|K(x_0, x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots) - K(x_0, x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots)| \leqslant \mathrm{Lip}_i(K) d(x_i, x_i').$$

If K does not depend on some variable x_i, we set by convention $\mathrm{Lip}_i(K) = 0$.

The function K is defined on the space $\tilde{X} = X^{\mathbb{N}}$. This space carries a natural probability measure, describing the deterministic dynamics once the starting point is chosen at random according to μ, i.e., $\tilde{\mu} := \mu \otimes \delta_{Tx} \otimes \delta_{T^2 x} \otimes \cdots$. Let

$$\mathbb{E}(K) = \int_X K(x, Tx, \dots) \,\mathrm{d}\mu(x) = \int_{\tilde{X}} K \,\mathrm{d}\tilde{\mu}.$$

This is the average of K with respect to the natural measure of the system. We are interested in the deviation of $K(x, Tx, \dots)$ from its average $\mathbb{E}(K)$. For instance, if $K(x_0, \dots, x_{n-1}) = \sum f(x_i)$, then $K(x, Tx, \dots)$ is simply the Birkhoff sum $S_n f(x)$. It

is separately Lipschitz if f is Lipschitz, with Lipschitz constants $\mathrm{Lip}_i(K) = \mathrm{Lip}(f)$ for $0 \leqslant i \leqslant n-1$, and $\mathrm{Lip}_i(K) = 0$ otherwise.

Theorem 1.9. *Consider a Young tower with* $\mathrm{tail}_n = O(n^{-q})$ *for some* $q > 1$. *Then, for all* $p \in [1, \infty)$, *there exists* $C > 0$ *such that, for all separately Lipschitz function* K,

- *if* $q > 2$, *the quantity* $\int |K(x, Tx, \dots) - \mathbb{E}K|^p \, \mathrm{d}\mu$ *is bounded by*

$$\begin{cases} C \left(\sum \mathrm{Lip}_i(K)^2 \right)^{p/2} & \text{if } p \leqslant 2q - 2, \\ C \left(\sum \mathrm{Lip}_i(K)^2 \right)^{q-1} \left(\sum \mathrm{Lip}_i(K) \right)^{p-(2q-2)} & \text{if } p \geqslant 2q - 2. \end{cases}$$

- *if* $q = 2$, *the quantity* $\int |K(x, Tx, \dots) - \mathbb{E}K|^p \, \mathrm{d}\mu$ *is bounded by*

$$\begin{cases} C \left(\sum \mathrm{Lip}_i(K)^2 \right)^{p/2} \left[1 + \log \left(\sum \mathrm{Lip}_i(K) \right) - \log \left(\sum \mathrm{Lip}_i(K)^2 \right)^{1/2} \right]^{p/2} & \text{if } p \leqslant 2, \\ C \left(\sum \mathrm{Lip}_i(K)^2 \right) \left(\sum \mathrm{Lip}_i(K) \right)^{p-2} & \text{if } p > 2. \end{cases}$$

- *if* $q < 2$, *then for all* $t > 0$

$$\mu\{x \; : \; |K(x, Tx, \dots) - \mathbb{E}K| \geqslant t\} \leqslant Ct^{-q} \sum_i \mathrm{Lip}_i(K)^q \tag{1.4}$$

and therefore $\int |K(x, Tx, \dots) - \mathbb{E}K|^p \, \mathrm{d}\mu$ *is bounded by*

$$\begin{cases} C \left(\sum \mathrm{Lip}_i(K)^q \right)^{p/q} & \text{if } p < q, \\ C \left(\sum \mathrm{Lip}_i(K)^q \right) \left[1 + \log \left(\sum \mathrm{Lip}_i(K) \right) - \log \left(\sum \mathrm{Lip}_i(K)^q \right)^{1/q} \right] & \text{if } p = q, \\ C \left(\sum \mathrm{Lip}_i(K)^q \right) \left(\sum \mathrm{Lip}_i(K) \right)^{p-q} & \text{if } p > q. \end{cases}$$

Note that $|K - \mathbb{E}(K)|$ is trivially bounded by $\sum_i \mathrm{Lip}_i(K)$. Hence, when $q > 2$, it is sufficient to prove the estimates for $p = 2q - 2$, as the other ones follow using (1.2). In the same way, for $q < 2$, it suffices to prove the weak moment bound (1.4), thanks to (1.3). On the other hand, for $q = 2$, the inequality for $p = 2$ is not sufficient to obtain the result for $p > 2$.

There are logarithmic terms in some of the above bounds when $q \leqslant 2$. This is not surprising, since such terms are already present in the simpler situation of Birkhoff sums, in Theorem 1.4. The precise form of these logarithmic terms may seem surprising at first sight, but it is in fact natural since such a bound has to be homogeneous: The logarithmic term should be invariant if one replaces K with λK, and therefore each $\mathrm{Lip}_i(K)$ with $\lambda \mathrm{Lip}_i(K)$. This would not be the case for the simpler bound $\log(\sum \mathrm{Lip}_i(K))$. When $\mathrm{Lip}_i(K)$ does not depend on i, the bound $\log \left(\sum \mathrm{Lip}_i(K) \right) - \log \left(\sum \mathrm{Lip}_i(K)^q \right)^{1/q}$ reduces to $(1 - 1/q) \log n$, a constant multiple of $\log n$ as we may expect.

Compared to moment controls, concentration results for arbitrary functions K have a lot more applications, especially when K is non-linear. We refer the reader to [10, Section 7] for a description of such applications.

Theorem 1.9 implies Theorem 1.4 (just take $K(x_0, \dots, x_{n-1}) = \sum f(x_i)$). However, the proof of Theorem 1.4 is considerably simpler, and motivates some techniques used in the proof of Theorem 1.9. Hence, we prove both theorems separately below. While some cases of Theorem 1.4 are already known (especially the case $q = 2$, see Remark 1.7), we nevertheless give again a full proof of these cases, for completeness and with the concentration case in mind.

The proofs of our results rely on two main tools: a dynamical one (very precise asymptotics of renewal sequences of operators) and a probabilistic one (inequalities

for martingales, of Burkholder-Rosenthal and von Bahr-Esseen type). In addition, for the concentration inequalities, we require analytic tools such as maximal inequalities and interpolation results, since the Lipschitz constants $\mathrm{Lip}_a(K)$ may vary considerably with a, which makes more usual inequalities too crude. All these tools are presented in Section 2. Theorem 1.4 is proved in Section 3, and Theorem 1.9 is proved in Section 4.

2 Preliminaries

2.1 Renewal sequences of operators

In this paragraph, we summarize the results on renewal sequences of operators that we need later on. They are proved in [26, 15, 16].

Consider a Young tower $T : X \to X$. The associated transfer operator $\mathcal{L}$, adjoint to the composition by T, is given by

$$\mathcal{L}u(x) = \sum_{Ty=x} g(y)u(y).$$

Denoting by $g^{(n)}(x) = g(x)\cdots g(T^{n-1}x)$ the inverse of the jacobian of T^n, one has $\mathcal{L}^n u(x) = \sum_{T^n y=x} g^{(n)}(y)u(y)$. Iterating the inequality $|\log g(x) - \log g(y)| \leqslant Cd(Tx, Ty)$ and using the uniform expansion when a trajectory returns to the basis, one has the following bounded distortion property: there exists $C > 0$ such that, for all n, for all points x and y in the same cylinder of length n (i.e., for $i < n$, the points $T^i x$ and $T^i y$ are in the same partition element),

$$|\log g^{(n)}(x) - \log g^{(n)}(y)| \leqslant Cd(T^n x, T^n y).$$

Among the trajectories of T, the only non-trivial behavior is related to the successive returns to the basis. Define a first return transfer operator at time n by $R_n u(x) = \sum_{T^n y=x} g^{(n)}(y)u(y)$ where $x \in \Delta_0$ and the sum is over those preimages y of x that belong to Δ_0 but $T^i y \notin \Delta_0$ for $1 \leqslant i \leqslant n-1$. Since R_n only involves preimages y with $\varphi(y) = n$, its operator norm $\|R_n\|$ with respect to the Lipschitz norm satisfies $\|R_n\| \leqslant C\mu(\varphi = n)$. In particular, R_n is easy to understand.

Define a partial transfer operator $T_n = 1_{\Delta_0}\mathcal{L}^n 1_{\Delta_0}$. It can be written as $T_n u(x) = \sum_{T^n y=x} g^{(n)}(y)u(y)$, where x and y all have to belong to Δ_0. Decomposing a trajectory from Δ_0 to Δ_0 into successive excursions, one gets

$$T_n = \sum_{k=1}^{n} \sum_{\ell_1+\cdots+\ell_k=n} R_{\ell_1} \cdots R_{\ell_k}.$$

Formally, this is equivalent to the equality $\sum T_n z^n = (I - \sum R_k z^k)^{-1}$. This makes it possible to understand T_n. Denote by Π the projection on constant functions on Δ_0, given by $\Pi u(x) = \int_{\Delta_0} u \, d\mu/\mu(\Delta_0)$.

The following proposition is [16, Proposition 2.2.19 and Remark 2.4.8] in the specific case of polynomial growth rate (this proposition also holds for more exotic asymptotics such as $O(n^{-q}\log n)$ – it follows that most results of our paper could be extended to such speeds).

Proposition 2.1. *Assume that $\mu(\varphi \geqslant n) = O(n^{-q})$ for some $q > 1$. Then $\|T_{n+1} - T_n\| = O(n^{-q})$ and $\|T_n - \Pi T_n \Pi\| = O(n^{-q})$.*

In particular, $\|T_{n+1} - T_n\|$ is summable, hence T_n converges. Its limit is $\mu(\Delta_0)\Pi$.

Consider now a general function u and a point $x \in \Delta_0$, we wish to describe $\mathcal{L}^n u(x) = \sum_{T^n y=x} g^{(n)}(y)u(y)$. Splitting the trajectory of y into a first part until the first entrance

in Δ_0, of length $b \geqslant 0$, and then a second part starting from Δ_0 at time b and coming back to Δ_0 at time n, we obtain a decomposition

$$1_{\Delta_0} \mathcal{L}^n = \sum_{\ell + b = n} T_\ell B_b. \tag{2.1}$$

The operator B_b is given by $B_b u(x) = \sum_{T^b y = x} g^{(b)}(y) u(y)$, the sum being restricted to those preimages whose first entrance in Δ_0 is at time b (the projection in the basis of those points necessarily has $\varphi > b$). By bounded distortion, one gets

$$\|B_b\| \leqslant C\mu(\varphi > b). \tag{2.2}$$

2.2 Weak L^p spaces

If a function u belongs to L^p on a probability space, then $\mathbb{P}(|u| > s) \leqslant s^{-p} \mathbb{E}(|u|^p)$ by Markov's inequality. On the other hand, this condition $\mathbb{P}(|u| > s) = O(s^{-p})$ is not sufficient to belong to L^p. For instance, a stable law of index $p \in (1, 2)$ satisfies $\mathbb{P}(|Z| > s) \sim cs^{-p}$, it readily follows that it does not belong to L^p.

We say that a random variable u belongs to weak L^p if $\mathbb{P}(|u| > s) = O(s^{-p})$. We write

$$\|u\|_{L^{p,w}}^p = \sup_s s^p \mathbb{P}(|u| > s).$$

This is the analogue of the L^p norm in this context. It satisfies $\|u\|_{L^{p,w}} \leqslant \|u\|_{L^p}$. In general, $\|\cdot\|_{L^{p,w}}$ is not a norm (i.e., it does not satisfy the triangular inequality), however it is equivalent to a norm when $p > 1$ (see for instance [27, Paragraph V.3]). The weak L^p space is a particular instance of *Lorentz spaces*, corresponding to the space $L^{p,\infty}$ in the standard notation.

Apart from its natural appearance when considering stable laws, a major role of the weak L^p space comes from interpolation theory. The following is a particular case of the Marcinkiewicz interpolation theorem, see for instance [27, Theorem V.2.4].

Theorem 2.2. *If a linear operator is bounded from $L^1(\mu)$ to $L^{1,w}(\nu)$ and from $L^\infty(\mu)$ to $L^\infty(\nu)$, then it is bounded from $L^p(\mu)$ to $L^p(\nu)$ for any $1 < p < \infty$.*

This result can for instance be used to prove the boundedness of the Hardy-Littlewood maximal function on any L^p space, $1 < p \leqslant \infty$, since boundedness from L^1 to $L^{1,w}$ and from L^∞ to itself hold. We recall the statement in the case of $\mathbb{Z}$, since we need it later on. See for instance [27, Theorem II.3.7].

Theorem 2.3. *To a sequence $(u_n)_{n \in \mathbb{Z}}$, associate the sequence*

$$Mu(n) = \sup_{h \geqslant 0} \frac{1}{2h+1} \sum_{i=n-h}^{n+h} |u_i|.$$

For all $p \in (1, +\infty]$, there exists a constant C such that $\|Mu\|_{\ell^p} \leqslant C\|u\|_{\ell^p}$ for any sequence $u \in \ell^p$.

2.3 Martingale inequalities

Given a decreasing sequence of σ-algebras $\mathcal{F}_0 \supset \mathcal{F}_1 \supset \dots$ on a probability space, a sequence of reverse martingale differences with respect to this filtration is a sequence of random variables D_k such that $\mathbb{E}(D_k \mid \mathcal{F}_{k+1}) = 0$. This is a kind of one-sided independence condition. Moment inequalities, similar to classical inequalities for independent random variables, hold in this setting.

We will use the following Burkholder-Rosenthal inequality:

Theorem 2.4. *For any $Q \geqslant 2$, there exists a constant C such that any sequence of reverse martingale differences satisfies*

$$\mathbb{E}\left|\sum D_k\right|^Q \leqslant C\mathbb{E}\left(\sum \mathbb{E}(D_k^2 \mid \mathcal{F}_{k+1})\right)^{Q/2} + C\mathbb{E}(\max|D_k|^Q).$$

As a consequence,

$$\mathbb{E}\left|\sum D_k\right|^Q \leqslant C\left(\sum \|\mathbb{E}(D_k^2 \mid \mathcal{F}_{k+1})\|_{L^\infty}\right)^{Q/2} + C\sum \|\mathbb{E}(|D_k|^Q \mid \mathcal{F}_{k+1})\|_{L^\infty}. \qquad (2.3)$$

The first statement is due to Burkholder [8, Theorem 21.1]. The second (much weaker) statement readily follows, and is sufficient for our purposes. One interest of the second formulation is that the two terms look the same: in the applications we have in mind, we will control simultaneously $\|\mathbb{E}(D_k^2 \mid \mathcal{F}_{k+1})\|_{L^\infty}$ and $\|\mathbb{E}(|D_k|^Q \mid \mathcal{F}_{k+1})\|_{L^\infty}$.

For $Q \in (1,2)$, the (easier) analogue of the above theorem is the inequality of von Bahr and Esseen [29] stating that

$$\mathbb{E}\left|\sum D_k\right|^Q \leqslant C\sum \mathbb{E}|D_k|^Q.$$

However, we will rather need a version of this inequality involving weak L^Q norms (since the main part of Theorem 1.4 in the case $q < 2$ is the inequality (1.1), controlling the weak L^q norm of $S_n f$). Such an inequality holds:

Theorem 2.5. *For any $Q \in (1,2)$, there exists a constant C such that any sequence of reverse martingale differences D_k satisfies*

$$\left\|\sum D_k\right\|_{L^{Q,w}}^Q \leqslant C\sum_k \|D_k\|_{L^{Q,w}}^Q. \qquad (2.4)$$

Proof. This is a consequence of existing results in the literature, as we now explain. First, the $L^{Q,w}$-seminorm is not a norm, which can be a problem for the proof of inequalities involving an arbitrary number of terms. However, it is equivalent to a true norm, the Lorentz norm $L^{Q,\infty}$ (see [27, Paragraph V.3])), so this is not an issue.

[7, Theorem 7(1) on Page 39] proves that the space $L^{Q,w} = L^{Q,\infty}$ satisfies the von Bahr-Esseen property of index Q, i.e., the inequality (2.4) holds whenever the D_k are independent centered random variables.

Consider now a sequence of reverse martingale differences D_k. Let $(\tilde{D}_k)$ be independent random variables, such that $\tilde{D}_k$ is distributed as D_k. [3, Theorem 6.1] shows that

$$\left\|\sum D_k\right\|_{L^{Q,w}} \leqslant C\left\|\sum \tilde{D}_k\right\|_{L^{Q,w}}.$$

As the random variables $\tilde{D}_k$ are independent, they satisfy (2.4) by [7]. The same inequality follows for D_k. $\qquad \square$

2.4 Miscellaneous

We use repeatedly the following classical lemma, which is readily proved by a discrete integration by parts.

Lemma 2.6. *Let c_h be a sequence of nonnegative real numbers with $\sum_{h>n} c_h = O(n^{-q})$ for some $q > 1$. Then, for all $\alpha < q$, one has $\sum_{h>n} h^\alpha c_h = O(n^{\alpha-q})$. Moreover, for all $\alpha > q$, one has $\sum_{h<n} h^\alpha c_h = O(n^{\alpha-q})$. Finally, $\sum_{h<n} h^q c_h = O(\log n)$.*

We also use the following fact: If c_n is a summable sequence of nonnegative real numbers and $p \geqslant 1$,

$$\left|\sum c_n u_n\right|^p \leqslant \left(\sum c_n\right)^{p-1} \sum c_n |u_n|^p. \qquad (2.5)$$

Indeed, this follows from the convexity of $x \mapsto x^p$ for $\sum c_n = 1$, and the general case follows.

3 Moment bounds

Our goal in this section is to prove Theorem 1.4. We therefore fix a Young tower with $\text{tail}_n = O(n^{-q})$ for some $q > 1$.

The convolution of two sequences $(c_n)_{n \geqslant 0}$ and $(d_n)_{n \geqslant 0}$ is the sequence $c \star d$ given by

$$(c \star d)_n = \sum_{i=0}^{n} c_i d_{n-i}.$$

We write $c_n^{(q)}$ for a generic sequence of the form $C/(n+1)^q$, for a generic C that can change from one occurrence to the next, even on the same line, but only finitely many times in the whole article. We use repeatedly the fact that the convolution of two such sequences is bounded by a sequence of the same form. This fact reads

$$(c^{(q)} \star c^{(q)})_n = \sum_{i=0}^{n} c_i^{(q)} c_{n-i}^{(q)} \leqslant c_n^{(q)}. \tag{3.1}$$

(Note that the sequence $c_n^{(q)}$ on the right is not the same as the sequences on the left, in accordance with the above convention.)

We wish to understand the moments of Birkhoff sums $S_n f$. Since martingale inequalities are very powerful, we will reduce to such martingales in the most naive way. Let $\mathcal{F}_k = T^{-k}(\mathcal{F}_0)$ (where $\mathcal{F}_0$ is the Borel σ-algebra), a function is $\mathcal{F}_k$-measurable if and only if it can be written as $u \circ T^k$ for some function u. We have

$$S_n f = \sum_{k=0}^{n-1} (\mathbb{E}(S_n f \mid \mathcal{F}_k) - \mathbb{E}(S_n f \mid \mathcal{F}_{k+1})) + \mathbb{E}(S_n f \mid \mathcal{F}_n) = \sum_{k=0}^{n-1} A_k \circ T^k + \mathbb{E}(S_n f \mid \mathcal{F}_n),$$

for some functions A_k that we now describe. Note that this is a decomposition as a sum of reverse martingale differences, hence the moments of $S_n f$ will essentially be controlled by those of A_k.

Let $\mathcal{L}$ be the transfer operator, it satisfies $\mathbb{E}(u \mid \mathcal{F}_1) = (\mathcal{L}u) \circ T$. Hence, for $k < n$,

$$\mathbb{E}(S_n f \mid \mathcal{F}_k) - \mathbb{E}(S_n f \mid \mathcal{F}_{k+1}) = \left(\sum_{i=0}^{k} \mathcal{L}^i f \right) \circ T^k - \left(\sum_{i=1}^{k+1} \mathcal{L}^i f \right) \circ T^{k+1},$$

giving

$$A_k = \sum_{i=0}^{k} \mathcal{L}^i f - \left(\sum_{i=1}^{k+1} \mathcal{L}^i f \right) \circ T. \tag{3.2}$$

Let us define a function $F_k = \sum_{i=0}^{k} \mathcal{L}^i f$, this is the main function to understand.

Lemma 3.1. *If x is at height h and $Tx \in \Delta_0$, then*

$$F_k(x) - F_{k+1}(Tx) = O(1 + h \wedge k).$$

Proof. Clearly $\|F_k - F_{k+1} \circ T\|_{L^\infty} \leqslant 2(1+k)\|f\|_{L^\infty}$. We have to prove that $F_k(x) - F_{k+1}(Tx) = O(1+h)$.

First, we estimate $1_{\Delta_0} F_k$. We use the formalism of renewal transfer operators introduced in Paragraph 2.1. As in (2.1), we write $1_{\Delta_0} \mathcal{L}^n = \sum_{\ell+b=n} T_\ell B_b$, where T_ℓ counts the returns to the basis at time ℓ, and B_b is an average over preimages at time b that did not return to the basis in between. Write Π for the projection on constant functions on Δ_0. Proposition 2.1 shows that the operator $E_\ell = T_\ell - \Pi T_\ell \Pi$ satisfies $\|E_\ell\| \leqslant c_\ell^{(q)}$. We get

$$1_{\Delta_0} F_k = \sum_{i=0}^{k} 1_{\Delta_0} \mathcal{L}^i f = \sum_{\ell+b \leq k} T_\ell B_b f = \sum_{\ell+b \leq k} \Pi T_\ell \Pi B_b f + \sum_{\ell+b \leq k} E_\ell B_b f.$$

Since $\|B_b f\| \leqslant c_b^{(q)}$ by (2.2) and $\|E_\ell\| \leqslant c_\ell^{(q)}$, the second sum is uniformly $O(1)$. For the first sum, the function $\Pi T_\ell \Pi B_b f$ is constant by definition, and can be written as $t_\ell u_b(f)$ for $t_\ell = \int_{\Delta_0} T_\ell 1 \, \mathrm{d}\mu/\mu(\Delta_0)$ and $u_b(f) = \int_{\Delta_0} B_b f \, \mathrm{d}\mu/\mu(\Delta_0)$. We have obtained

$$1_{\Delta_0} F_k = \sum_{\ell+b \leq k} t_\ell u_b(f) + O(1),$$

where t_ℓ is uniformly bounded, and $u_b(f)$ is summable (with sum at most $\int |f| \, \mathrm{d}\mu/\mu(\Delta_0)$).

Consider now an arbitrary x, at height $h < k$, and with $Tx \in \Delta_0$. Then $F_k(x) = F_{k-h}(\pi x) + O(h)$ where πx is the projection of x in the basis of the tower, i.e., the unique preimage of x under T^h. We get

$$F_k(x) - F_{k+1}(Tx) = \sum_{\ell+b \leq k-h} t_\ell u_b(f) - \sum_{\ell+b \leq k+1} t_\ell u_b(f) + O(1+h)$$
$$= \sum_{k-h < \ell+b \leqslant k+1} t_\ell u_b(f) + O(1+h).$$

For each b, there are at most $h+1$ values of ℓ for which $k-h < \ell+b \leqslant k+1$. Since t_ℓ is bounded, we obtain

$$|F_k(x) - F_{k+1}(Tx)| \leqslant (h+1) \sum_b |u_b(f)| + O(1+h) = O(1+h). \qquad \square$$

3.1 The case $q > 2$

In this paragraph, we prove Theorem 1.4 in the case $q > 2$. It suffices to prove the desired estimate for $p = 2q - 2$, since the other estimates follow using (1.2).

We start from the decomposition

$$S_n f = \sum A_k \circ T^k + \mathbb{E}(S_n f \mid \mathcal{F}_n). \tag{3.3}$$

First, we control the last term, which is easier. Write $Q = 2q - 2$, we have

$$\|\mathbb{E}(S_n f \mid \mathcal{F}_n)\|_{L^Q} = \left\| \left(\sum_{k=1}^n \mathcal{L}^k f \right) \circ T^n \right\|_{L^Q} = \left\| \sum_{k=1}^n \mathcal{L}^k f \right\|_{L^Q} \leqslant \sum_{k=1}^n \|\mathcal{L}^k f\|_{L^Q}.$$

One can use transfer operators techniques, or argue directly as in [21]: since the speed of decay of correlations against bounded functions is $O(1/n^{q-1})$ by [31], we have

$$\int |\mathcal{L}^k f|^Q \, \mathrm{d}\mu \leqslant \|f\|_\infty^{Q-1} \int \mathrm{sgn}(\mathcal{L}^k f) \cdot \mathcal{L}^k f \, \mathrm{d}\mu \leqslant C/k^{q-1}.$$

Hence, $\|\mathcal{L}^k f\|_{L^Q} \leqslant C/k^{(q-1)/Q} = k^{-1/2}$, giving $\|\mathbb{E}(S_n f \mid \mathcal{F}_n)\|_{L^Q} \leqslant Cn^{1/2}$.

Then, we turn to the first sum $\sum A_k \circ T^k = \sum D_k$ in (3.3). It is a sum of reverse martingale differences, hence we may apply Burkholder-Rosenthal inequality in the form of (2.3):

$$\mathbb{E}\left| \sum D_k \right|^Q \leqslant C \left(\sum \|\mathbb{E}(D_k^2 \mid \mathcal{F}_{k+1})\|_{L^\infty} \right)^{Q/2} + C \sum \|\mathbb{E}(|D_k|^Q \mid \mathcal{F}_{k+1})\|_{L^\infty}. \tag{3.4}$$

For $r \in \{2, Q\}$, we have $\mathbb{E}(|D_k|^r \mid \mathcal{F}_{k+1})(y) = \mathcal{L}(|A_k|^r)(T^{k+1}y)$. This implies $\sum \|\mathbb{E}(|D_k|^r \mid \mathcal{F}_{k+1})\|_{L^\infty} = \|\mathcal{L}(|A_k|^r)\|_{L^\infty}$.

Consider a point $x \in X$. If it does not belong to Δ_0, it has a unique preimage z, and moreover $A_k(z) = 0$. Hence, $\mathcal{L}(|A_k|^r)(x) = 0$. Suppose now $x \in \Delta_0$. Let z_α denote its

preimages (with respective heights $h_\alpha - 1$). Lemma 3.1 gives $A_k(z_\alpha) = O(1 + h_\alpha \wedge k)$. Hence,

$$\mathcal{L}(|A_k|^r)(x) = \sum_\alpha g(z_\alpha)|A_k(z_\alpha)|^r \leqslant C + \sum_\alpha g(z_\alpha)(h_\alpha \wedge k)^r \leqslant C \int_{\Delta_0} (\varphi \wedge k)^r \, \mathrm{d}\mu.$$

We have proved that

$$\|\mathbb{E}(|D_k|^r \mid \mathcal{F}_{k+1})\|_{L^\infty} \leqslant C \int_{\Delta_0} (\varphi \wedge k)^r \, \mathrm{d}\mu. \tag{3.5}$$

We use this inequality to estimate the two sums on the right hand side of (3.4). For $r = 2$, the above integral is uniformly bounded since φ has a moment of order 2. Hence, the first sum in (3.4) is bounded by $n^{Q/2} = n^{q-1}$. For $r = Q = 2q - 2$, the above integral is bounded by k^{q-2} thanks to Lemma 2.6. Summing over k, it follows that the second sum in (3.4) is bounded by n^{q-1}, as desired. $\qquad \square$

3.2 The case $q < 2$

In this paragraph, we prove Theorem 1.4 in the case $q \in (1, 2)$. Again, it suffices to prove the estimate (1.1) regarding the weak q-moment, i.e., $\|S_n f\|_{L^{q,w}} \leqslant C n^{1/q}$, since the other estimates follow using (1.3).

We start again from the decomposition $S_n f = \sum A_k \circ T^k + \mathbb{E}(S_n f \mid \mathcal{F}_n)$. We rely on the von Bahr-Esseen result for weak moments given in Theorem 2.5, for $Q = q$.

First, we control the last term, as above: we have $\|\mathbb{E}(S_n f \mid \mathcal{F}_n)\|_{L^q} \leqslant \sum_{k=1}^n \|\mathcal{L}^k f\|_{L^q}$. Moreover, we have as above $\|\mathcal{L}^k f\|_{L^q} \leqslant C/k^{(q-1)/q}$. Summing over k,

$$\|\mathbb{E}(S_n f \mid \mathcal{F}_n)\|_{L^q} \leqslant C \sum_{k=1}^n k^{1/q-1} \leqslant C n^{1/q}.$$

As the weak L^q-norm is dominated by the strong L^q-norm, this is the desired control.

Now, we turn to the contribution of A_k. We want to estimate $\|A_k\|_{L^{q,w}}$. If $Tx \notin \Delta_0$, then $A_k(x) = 0$. If $Tx \in \Delta_0$, then $|A_k(x)| \leqslant C(1 + k \wedge h) \leqslant C(1 + h)$ by Lemma 3.1. Hence, for s larger than a fixed constant,

$$\mu\{x \ : \ |A_k(x)| \geqslant s\} \leqslant C\mu\{y \in \Delta_0 \ : \ \varphi(y) \geqslant C^{-1}s\} = O(s^{-q}).$$

This shows that $\|A_k\|_{L^{q,w}}$ is uniformly bounded. Summing over k and using Theorem 2.5, we get $\|\sum A_k \circ T^k\|_{L^{q,w}}^q \leqslant Cn$, as desired. $\qquad \square$

3.3 The case $q = 2$

In this paragraph, we prove Theorem 1.4 in the case $q = 2$. Contrary to the previous cases, it is not sufficient to prove the result at the critical exponent $p = 2$, one should also control all $p > 2$. The arguments in the proof of the case $q > 2$ (notably Burkholder's inequality (3.4) combined with (3.5)) give, for a general $p \geqslant 2$,

$$\|S_n f\|_{L^p}^p \leqslant C \left(\sum_{k=0}^{n-1} \int_{\Delta_0} (\varphi \wedge k)^2 \right)^{p/2} \mathrm{d}\mu + C \sum_{k=0}^{n-1} \int_{\Delta_0} (\varphi \wedge k)^p \, \mathrm{d}\mu + \left(\sum_{k=1}^n \|\mathcal{L}^k f\|_{L^p} \right)^p. \tag{3.6}$$

First, we have $\int |\mathcal{L}^k f|^p \, \mathrm{d}\mu \leqslant C/k$ since the speed of decay of correlations is $1/k$. Hence, $\|\mathcal{L}^k f\|_{L^p} \leqslant k^{-1/p}$ and the last term in (3.6) is bounded by n^{p-1}.

Let us now deal with $p = 2$. Lemma 2.6 gives $\int_{\Delta_0} (\varphi \wedge k)^2 \, \mathrm{d}\mu \leqslant \log k$ since we are precisely at the critical exponent for which there is an additional logarithmic factor. Summing over k and using (3.6), we obtain $\|S_n f\|_{L^2}^2 \leqslant n \log n$ as desired.

Consider then $p > 2$. Again, $\int_{\Delta_0} (\varphi \wedge k)^2 \, \mathrm{d}\mu \leqslant \log k$, hence the first sum in (3.6) gives a contribution $C(n \log n)^{p/2}$, which is bounded by Cn^{p-1} as $p/2 < p - 1$. For the second sum in (3.6), Lemma 2.6 gives $\int (\varphi \wedge k)^p \, \mathrm{d}\mu \leqslant Ck^{p-2}$. Summing over k, we get a bound n^{p-1}. □

4 Concentration bounds

In this section, we prove Theorem 1.9 about concentration inequalities in Young towers with $\mathrm{tail}_n = O(n^{-q})$ for some $q > 1$. As before, we write $c_n^{(q)}$ for a generic sequence that is $O(n^{-q})$.

Consider a general function $K(x_0, x_1, \dots)$ which is separately Lipschitz in each variable, with corresponding constants $\mathrm{Lip}_i(K)$. Fix any reference point x_* in the space.

To study the magnitude of $K(x, Tx, \dots)$, the idea is to decompose it as a sum of reverse martingale differences. We consider K as a function defined on the space $\tilde{X} = X^{\mathbb{N}}$, endowed with the probability measure $\tilde{\mu} = \mu \otimes \delta_{Tx} \otimes \delta_{T^2 x} \otimes \cdots$. Let $\mathcal{F}_k$ be the σ-algebra generated by indices starting with k (i.e., a function $f(x_0, x_1, \dots)$ on $\tilde{X}$ is $\mathcal{F}_k$-measurable if it does not depend on $x_0, \dots, x_{k-1}$). Let

$$K_k(x_k, \dots) = \mathbb{E}(K \mid \mathcal{F}_k)(x_k, \dots) = \sum_{T^k x = x_k} g^{(k)}(x) K(x, \dots, T^{k-1}x, x_k, \dots). \tag{4.1}$$

This function plays the role of the function F_k (defined after (3.2)) for Birkhoff sums, and is the main object to understand.

As in the proof of Lemma 3.1, we want to express $K_k(x_k, \dots)$, for $x_k \in \Delta_0$, using the transfer operator restricted to the basis, i.e., T_n. Define for $i \leqslant k$ a function w_i on the basis by

$$w_i(x) = \sum_{T^i y = x} g^{(i)}(y) \Big[K(y, Ty, \dots, T^{j(y)-1}y, T^{j(y)}y, \dots, T^{i-1}y, x, \underbrace{x_*, \dots, x_*}_{k-i-1 \text{ terms}}, x_k, \dots)$$
$$- K(y, Ty, \dots, T^{j(y)-1}y, T^{j(y)}y, \underbrace{x_*, \dots, x_*}_{k-j(y)-1 \text{ terms}}, x_k, \dots) \Big],$$

where for each y we define $j(y)$ as the last time in $[0, i-1]$ for which $T^{j(y)}(y) \in \Delta_0$. If there is no such time, then $j(y) = -1$. The idea is that, for each preimage y of x under T^i, we replace its last excursion outside of Δ_0 by the trivial sequence $x_*, \dots, x_*$.

A simple telescoping argument then gives:

$$K_k(x_k, \dots) = \sum_{i=0}^{k} \sum_{x \in \Delta_0, T^{k-i} x = x_k} g^{(k-i)}(x) w_i(x) + K(x_*, \dots, x_*, x_k, \dots).$$

Indeed, in the expression (4.1), if one starts replacing successively each excursion outside of Δ_0, one ends up adding sums of the functions $w_i(x)$, and the remaining term (where all excursions have been replaced) is $\sum_{T^k x = x_k} g^{(k)}(x) K(x_*, \dots, x_*, x_k, \dots)$, which reduces to $K(x_*, \dots, x_*, x_k, \dots)$ since $\sum g^{(k)}(x) = 1$ as the measure is invariant.

The above expression also reads

$$K_k(x_k, \dots) = \sum_{i=0}^{k} T_{k-i} w_i(x_k) + K(x_*, \dots, x_*, x_k, \dots). \tag{4.2}$$

We will be able to use it since we know a lot about the operators T_n (their properties, expressed in Proposition 2.1, were already at the heart of the proof of Lemma 3.1), but we first need to understand w_i more properly.

Lemma 4.1. *We have* $\|w_i\| \leqslant \sum_{a+b=i} \mathrm{Lip}_a(K) c_b^{(q)}$.

Proof. First, we control the supremum of w_i. Write $w_i(x) = \sum_y g^{(i)}(y) H(y)$, then $|H(y)| \leqslant \sum_{j(y)+1}^{i-1} \mathrm{Lip}_\ell(K)$. The sum of $g^{(i)}(y)$ over those points with $j(y) < \ell$ is given by $\sum_{T^{i-\ell} z = x} g^{(i-\ell)}(z)$, where the sum is restricted to those points z that do not come back to the basis before time $i - \ell$. By bounded distortion, this is comparable to $\mu\{\varphi > i - \ell\} \leqslant c_{i-\ell}^{(q)}$. We get

$$\|w_i\|_\infty \leqslant \sum_\ell \mathrm{Lip}_\ell(K) c_{i-\ell}^{(q)}. \tag{4.3}$$

We estimate now the Lipschitz constant of w_i. Write for $x, x' \in \Delta_0$

$$\begin{aligned} w_i(x) - w_i(x') &= \sum g^{(i)}(y) H(y) - g^{(i)}(y') H(y') \\ &= \sum g^{(i)}(y)(H(y) - H(y')) + \sum (g^{(i)}(y) - g^{(i)}(y')) H(y'), \end{aligned}$$

where we have paired together the preimages y and y' of x and x' under T^i that belong to the same cylinder of length i. For the second sum, bounded distortion gives $|g^{(i)}(y) - g^{(i)}(y')| \leqslant C d(x, x') g^{(i)}(y')$, hence the Lipschitz norm of this sum is at most $C\|w_i\|_\infty$, which has already been controlled in (4.3). For the first sum, we have

$$|H(y) - H(y')| \leqslant 2 \sum_{\ell=0}^{i-1} \mathrm{Lip}_\ell(K) d(T^\ell y, T^\ell y') \leqslant 2 \sum_{\ell=0}^{i-1} \mathrm{Lip}_\ell(K) \Psi_{i-\ell}(T^\ell y) d(x, x'),$$

where $\Psi_a(z) = \rho^{\mathrm{Card}\{0 \leqslant t < a, T^t z \in \Delta_0\}}$: this function measures the expansion of the map T^a applied to z, since each return to the basis gives an expansion factor of $\rho^{-1} > 1$ by definition of the distance. Using bounded distortion, we get

$$\left| \sum g^{(i)}(y)(H(y) - H(y')) \right| \leqslant C \sum_{\ell=0}^{i-1} \mathrm{Lip}_\ell(K) d(x, x') \int_{T^{-(i-\ell)} \Delta_0} \Psi_{i-\ell} \, \mathrm{d}\mu.$$

By [10, Lemma 4.4], the sequence $\int_{T^{-n} \Delta_0} \Psi_n \, \mathrm{d}\mu$ is $\leqslant c_n^{(q)}$. The desired bound for the Lipschitz constant of w_i follows. $\qquad\square$

Then, we turn to the analogue of Lemma 3.1.

Lemma 4.2. *If x_k is at height h and $x_{k+1} \in \Delta_0$, then*

$$\begin{aligned} |K_k(x_k, & x_{k+1}, x_{k+2}, \dots) - K_{k+1}(x_{k+1}, x_{k+2}, \dots)| \\ &\leqslant \sum_{a=0}^{k-h} \mathrm{Lip}_a(K) \min\left((h+1) c_{k-h-a}^{(q)}, c_{k-h-a}^{(q-1)}\right) + \sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K). \end{aligned} \tag{4.4}$$

When $h > k$, the first sum vanishes, and the second one reduces to $\sum_{a=0}^{k} \mathrm{Lip}_a(K)$ since $\mathrm{Lip}_a(K) = 0$ for $a < 0$.

If all the $\mathrm{Lip}_a(K)$ are of order 1 (which is the case for instance with Birkhoff sums), it is easy to check that the expression in the lemma reduces to $O(1 + h \wedge k)$ as in Lemma 3.1.

Proof. The case $h > k$ is easy (it suffices to substitute each variable in the expression of $K_k(x_k, \dots)$ with the corresponding variable in $K_{k+1}(x_{k+1}, \dots)$). Let us deal with the more interesting case $h \leqslant k$.

We first prove the inequality

$$
\begin{aligned}
&|K_k(x_k, x_{k+1}, x_{k+2}, \dots) - K_{k+1}(x_{k+1}, x_{k+2}, \dots)| \\
&\leqslant \sum_{a=0}^{k-h} \mathrm{Lip}_a(K) \left[\sum_{b=0}^{k-h-a} c_b^{(q)} \left(\sum_{j=k-h-a-b}^{k-a-b} c_j^{(q)} \right) + \sum_{b=k-h-a+1}^{k-a+1} c_b^{(q)} \right] + \sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K).
\end{aligned}
\tag{4.5}
$$

We replace successively all the variables with index in $(k-h, k]$ in the expressions of $K_k(x_k, \dots)$ and $K_{k+1}(x_{k+1}, \dots)$ with x_*, introducing an error at most $\sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K)$ that corresponds to the last term in (4.5). Letting

$$
\tilde{K}(x_0, \dots, x_{k-h}) = K(x_0, \dots, x_{k-h}, x_*, \dots, x_*, x_{k+1}, \dots) - K(x_*, \dots, x_*, x_{k+1}, \dots),
$$

we may then work with $\tilde{K}$ instead of K. It satisfies $\mathrm{Lip}_a(\tilde{K}) \leqslant \mathrm{Lip}_a(K)$ for $a \leqslant k-h$, and $\mathrm{Lip}_a(\tilde{K}) = 0$ for $a > k-h$. Let w_i be the corresponding functions for $\tilde{K}$, and let $x = \pi x_k$ be the projection of x_k in the basis of the tower. We get from (4.2)

$$
\tilde{K}_k(x_k) = \tilde{K}_{k-h}(x) = \sum_{i=0}^{k-h} T_{k-h-i} w_i(x), \quad \tilde{K}_{k+1}(x_{k+1}) = \sum_{i=0}^{k+1} T_{k+1-i} w_i(x_{k+1}).
$$

We write $T_\ell = \Pi T_\ell \Pi + E_\ell$, where Π is the projection on constant functions, and $\|E_\ell\| \leqslant c_\ell^{(q)}$ by Proposition 2.1. We have

$$
\left| \sum_{i=0}^{k-h} E_{k-h-i} w_i(x) \right| \leqslant \sum_{\ell+i=k-h} \|E_\ell\| \|w_i\| \leqslant \sum_{\ell+i=k-h} c_\ell^{(q)} \sum_{a+b=i} \mathrm{Lip}_a(\tilde{K}) c_b^{(q)}
$$

$$
= \sum_{a+j=k-h} \mathrm{Lip}_a(\tilde{K}) (c^{(q)} \star c^{(q)})_j.
$$

By (3.1), this is bounded by $\sum_{a+j=k-h} \mathrm{Lip}_a(K) c_j^{(q)}$, which is bounded by (4.5) (to see this, in (4.5), take $b=0$ in the first sum over b, and then $j = k-h-a$ in the next sum). In the same way, we have

$$
\left| \sum_{i=0}^{k+1} E_{k+1-i} w_i(x_{k+1}) \right| \leqslant \sum_{a+j=k+1} \mathrm{Lip}_a(\tilde{K}) c_j^{(q)},
$$

which is again bounded by (4.5) (up to a shift of one in the indices, take $b=0$ in the first sum of (4.5) and $j = k-a$ in the second sum).

We turn to the main terms, coming from $\Pi T_\ell \Pi$. We have $\Pi T_\ell \Pi w_i = t_\ell u_i$, for some scalar sequences t_ℓ and u_i. Moreover, u_i is bounded by $\|w_i\|$, and $|t_\ell - t_{\ell+1}| \leqslant c_\ell^{(q)}$ by Proposition 2.1. The resulting term is

$$
\left| \sum_{i=0}^{k-h} t_{k-h-i} u_i - \sum_{i=0}^{k+1} t_{k+1-i} u_i \right| \leqslant \sum_{i=0}^{k-h} |u_i| \sum_{j=k-h-i}^{k-i} |t_{j+1} - t_j| + \sum_{i=k-h+1}^{k+1} |u_i|.
$$

Bounding $|u_i|$ by $\sum_{a+b=i} \mathrm{Lip}_a(\tilde{K}) c_b^{(q)}$ and $|t_{j+1} - t_j|$ with $c_j^{(q)}$, we readily check that all those terms are bounded by (4.5).

This concludes the proof of (4.5). To conclude, we should show that the coefficient of $\mathrm{Lip}_a(K)$ in this expression is bounded by $\min\left((h+1) c_{k-h-a}^{(q)}, c_{k-h-a}^{(q-1)} \right)$. We have

$\sum_{i=\ell}^{\infty} c_i^{(q)} \leqslant c_\ell^{(q-1)}$. In particular,

$$\sum_{b=0}^{k-h-a} c_b^{(q)} \left(\sum_{j=k-h-a-b}^{k-a-b} c_j^{(q)} \right) \leqslant \sum_{b=0}^{k-h-a} c_b^{(q)} c_{k-h-a-b}^{(q-1)} = (c^{(q)} \star c^{(q-1)})_{k-h-a}$$

$$\leqslant (c^{(q-1)} \star c^{(q-1)})_{k-h-a} \leqslant c_{k-h-a}^{(q-1)},$$

as the sequences that are $O(1/n^{q-1})$ are stable under convolution. This proves the upper bound $c_{k-h-a}^{(q-1)}$. For the other one, note that $\sum_{j=k-h-a-b}^{k-a-b} c_j^{(q)} \leqslant (h+1)c_j^{(q)}$. From this point on, one can continue the computation as above, getting in the end the bound $(h+1)c_{k-h-a}^{(q)}$. □

Remark 4.3. *The article [10] already proved concentration estimates in Young towers, but only for $q > 2$. In this case, the estimates were not as good as those in Theorem 1.9. Moreover, all the estimates started diverging when $q \leqslant 2$. There are three main differences in the current approach that make it possible to improve upon [10]:*

- *The decomposition (4.2) of K_k, where one replaces one excursion at a time in the definition of w_i, is more efficient than the corresponding decomposition of [10] where one only replaces one variable at a time (this creates some useless redundancy in the estimates, which is not a problem when $q > 2$ but causes divergence of the estimates when $q \leqslant 2$).*

- *The main difference between the current paper and [10] is that, in Lemma 4.2, we compare directly K_k to K_{k+1}. On the contrary, in [10], K_k and K_{k+1} are compared to explicit integral quantities (see for instance Lemma 2.3 there). This is more intuitive and natural, since it expresses the mixing properties of the system. However, when $q \leqslant 2$, the convergence towards these integrals is rather slow, making again the estimates diverge. In the proof of Lemma 4.2, we do not claim that K_k is close to any explicit or meaningful quantity, only that it is close to $\sum t_{k-i} u_i$. This is sufficient to prove that K_k is close to K_{k+1} since t_n is close to t_{n+1} by Proposition 2.1. Both are also close to $\lim t_i$ if n is large enough, and this is essentially what is used in [10], but this gives a worse estimate.*

- *In the case $q > 2$, the main new ingredient compared to the techniques of [10] is Lemma 4.4 below.*

We can now deduce concentration bounds in the different situations we considered for moment bounds.

4.1 The case $q > 2$

In this paragraph, we prove Theorem 1.9 in the case $q > 2$. As we explained after the statement of this theorem, it suffices to prove the result for $p = 2q - 2$.

In this situation, we use (4.4) in the form

$$|K_k(x_k, x_{k+1}, \dots) - K_{k+1}(x_{k+1}, \dots)| \leqslant \sum_{a=0}^{k-h} \text{Lip}_a(K) c_{k-h-a}^{(q-1)} + \sum_{a=k-h+1}^{k} \text{Lip}_a(K), \quad (4.6)$$

i.e., we always use the same term $c_{k-h-a}^{(q-1)}$ in the minimum in (4.4).

Let us start the proof of the theorem. The quantity $K - \mathbb{E}(K)$ can be decomposed as $\sum_{k \geqslant 0}(K_k - K_{k+1})$. Since this is a sum of reverse martingale differences, we may use

Burkholder-Rosenthal inequality in the form of (2.3), to obtain a bound

$$\int |K(x, Tx, \dots) - \mathbb{E}(K)|^{2q-2} \, d\mu(x)$$
$$\leqslant C \left(\sum \|\mathbb{E}(D_k^2 \mid \mathcal{F}_{k+1})\|_\infty \right)^{q-1} + C \sum \|\mathbb{E}(|D_k|^{2q-2} \mid \mathcal{F}_{k+1})\|_\infty,$$

where $D_k = K_k - K_{k+1}$. Hence, for $r \in \{2, 2q-2\}$, we should estimate $\|\mathbb{E}(|D_k|^r \mid \mathcal{F}_{k+1})\|_\infty$. If x_{k+1} is not in the basis of the tower, then $\mathbb{E}(|D_k|^r \mid \mathcal{F}_{k+1})(x_{k+1}, \dots) = 0$ and there is nothing to do. Assume now that x_{k+1} is in the basis. Let z_α denote its preimages, with respective heights h_α. We have

$$\mathbb{E}(|D_k|^r \mid \mathcal{F}_{k+1})(x_{k+1}, \dots) = \sum g(z_\alpha) |K_k(z_\alpha, x_{k+1}, \dots) - K_{k+1}(x_{k+1}, \dots)|^r.$$

With (4.6), we get

$$\|\mathbb{E}(|D_k|^r \mid \mathcal{F}_{k+1})\|_\infty \leqslant \sum_{h \geqslant 0} \mu(\varphi = h+1) \left(\sum_{a=0}^{k-h} \mathrm{Lip}_a(K) c_{k-h-a}^{(q-1)} + \sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K) \right)^r$$

Using the inequality $(X + Y)^r \leqslant C(X^r + Y^r)$ to separate the two sums, we get two different terms. We should then sum over k, and get a bound in terms of $\sum \mathrm{Lip}_a(K)^2$.

First, we deal with the first sum $\sum_{a=0}^{k-h} \mathrm{Lip}_a(K) c_{k-h-a}^{(q-1)}$. Since the sequence $c_n^{(q-1)}$ is summable, we have by (2.5)

$$\left(\sum_{a=0}^{k-h} \mathrm{Lip}_a(K) c_{k-h-a}^{(q-1)} \right)^r \leqslant C \sum_{a=0}^{k-h} \mathrm{Lip}_a(K)^r c_{k-h-a}^{(q-1)}.$$

Summing over k, we get a term

$$\sum_k \sum_{h=0}^{k} \sum_{a=0}^{k-h} \mu(\varphi = h+1) \mathrm{Lip}_a(K)^r c_{k-h-a}^{(q-1)}.$$

Writing $k = \ell + h$, this becomes

$$\sum_{h \geqslant 0} \sum_{\ell \geqslant 0} \sum_{a=0}^{\ell} \mu(\varphi = h+1) \mathrm{Lip}_a(K)^r c_{\ell-a}^{(q-1)}.$$

The sum over h factorizes out. Then, for each a, the sum over ℓ gives a finite contribution since $c^{(q-1)}$ is summable. We are left with $\sum_a \mathrm{Lip}_a(K)^r$, which is bounded by $\left(\sum_a \mathrm{Lip}_a(K)^2 \right)^{r/2}$ as desired.

Then, we deal with the second sum $\sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K)$. Summing over k, the corresponding term is

$$\sum_k \sum_h \mu(\varphi = h+1) \left(\sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K) \right)^r.$$

We need to treat separately the cases $r = 2$ and $r = 2q - 2$. For $r = 2$, we simply use Cauchy-Schwarz inequality:

$$\sum_k \sum_h \mu(\varphi = h+1) \left(\sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K) \right)^2 \leqslant \sum_k \sum_h \mu(\varphi = h+1) h \sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K)^2$$
$$= \sum_a \mathrm{Lip}_a(K)^2 \sum_h \mu(\varphi = h+1) h \sum_{k=a}^{a+h-1} 1.$$

We can factorize out $\sum_h \mu(\varphi = h+1)h^2$, which is finite since $q > 2$, by Lemma 2.6. We are left with $C \sum_a \operatorname{Lip}_a(K)^2$ as desired.

For the case $r = 2q - 2$, we should prove an inequality

$$\sum_k \sum_h \mu(\varphi = h+1) \left(\sum_{a=k-h+1}^{k} \operatorname{Lip}_a(K) \right)^{2q-2} \leqslant C \left(\sum_a \operatorname{Lip}_a(K)^2 \right)^{q-1}.$$

It turns out that this inequality is more difficult than the previous ones. It is given in Lemma 4.4 below. With this lemma, the proof is complete. $\qquad\square$

Lemma 4.4. *Let $q > 2$. Consider a sequence $a_n \geqslant 0$ with $\sum_{n \geqslant N} a_n = O(N^{-q})$. There exists a constant C such that, for any sequence $(u_n) \in \ell^2(\mathbb{Z})$,*

$$\sum_{n \in \mathbb{Z}} \sum_{h \geqslant 0} a_h \left| \sum_{i=n-h}^{n+h} u_i \right|^{2q-2} \leqslant C \left(\sum_{n \in \mathbb{Z}} u_n^2 \right)^{q-1}.$$

Although the statement of the lemma is completely elementary, this result is not trivial, even for $a_n = 1/n^{q+1}$ (as is maybe indicated by the fact that it fails for $q = 2$). In particular, we have not been able to find a direct proof: We need to resort to maximal inequalities and interpolation.

Proof. We associate to a sequence u_n the sequence

$$v(n,h) = \frac{\sum_{i=n-h}^{n+h} u_i}{(2h+1)^{1/2}}.$$

We consider v as a function on the space $\mathbb{Z} \times \mathbb{N}$ endowed with the measure $\nu(n,h) = (h+1)^{q-1} a_h$.

By Cauchy-Schwarz inequality, the function v is bounded in L^∞ by $\|u\|_{\ell^2}$. Let us now consider its weak L^2-norm. Let $Mu(n) = \sup_h \left(\sum_{i=n-h}^{n+h} |u_i| \right) / (2h+1)$ be the maximal function associated to u. It is bounded in ℓ^2 by $C\|u\|_{\ell^2}$, by Theorem 2.3. Since $v(n,h) \leqslant (2h+1)^{1/2} Mu(n)$, we have for all $s \geqslant 0$

$$\nu\{(n,h) \;:\; |v(n,h)| \geqslant s\} \leqslant \nu\{(n,h) \;:\; (2h+1)^{1/2} \geqslant s/Mu(n)\}$$
$$\leqslant \sum_n \sum_{h \geqslant ((s/Mu(n))^2 - 1)/2} (h+1)^{q-1} a_h.$$

By Lemma 2.6, we have $\sum_{h \geqslant t}(h+1)^{q-1} a_h \leqslant C/(t+1)$. Hence,

$$\nu\{(n,h) \;:\; |v(n,h)| \geqslant s\} \leqslant C \sum_n Mu(n)^2/s^2 \leqslant Cs^{-2}\|Mu\|_{\ell^2}^2 \leqslant Cs^{-2}\|u\|_{\ell^2}^2.$$

This shows that v is bounded in L^∞ and in weak L^2 by $C\|u\|_{\ell^2}$. One could deduce boundedness in any L^p for $2 < p < \infty$ by using classical interpolation arguments, but it is simpler to use the formula (1.3): we get

$$\int |v|^p \, d\nu \leqslant C \int_{s=0}^{\|v\|_\infty} s^{p-1} Cs^{-2}\|u\|_{\ell^2}^2 \, ds \leqslant C\|u\|_{\ell^2}^2 \|v\|_\infty^{p-2} \leqslant C\|u\|_{\ell^2}^p. \qquad (4.7)$$

Taking $p = 2q - 2$, we get

$$C \left(\sum u_n^2 \right)^{q-1} \geqslant \sum_n \sum_h \nu(n,h)|v(n,h)|^{2q-2} = \sum_n \sum_h (h+1)^{q-1} a_h \left| \frac{\sum_{i=n-h}^{n+h} u_i}{(2h+1)^{1/2}} \right|^{2q-2}.$$

The powers of h cancel on the right, and we are left with the statement of the lemma. $\qquad\square$

4.2 The case $q < 2$

In this case, it is sufficient to prove the weak moment estimate (1.4), since it implies all the other ones thanks to (1.3). Let us for instance explain how to get the most complicated moment estimate, for $p = q$. Write $A = \sum \operatorname{Lip}_i(K)^q$, so that $\mu\{|K - \mathbb{E}K| \geqslant s\} \leqslant CAs^{-q}$, and $B = \sum \operatorname{Lip}_i(K) \geqslant \|K - \mathbb{E}K\|_{L^\infty}$. Then

$$
\begin{aligned}
\int |K - \mathbb{E}K|^q \, \mathrm{d}\mu &= q \int s^{q-1} \mu\{|K - \mathbb{E}K| \geqslant s\} \, \mathrm{d}s \leqslant q \int_{s=0}^{B} s^{q-1} \min(1, CAs^{-q}) \, \mathrm{d}s \\
&\leqslant q \int_{s=0}^{A^{1/q}} s^{q-1} \, \mathrm{d}s + q \int_{s=A^{1/q}}^{B} s^{q-1} CAs^{-q} \, \mathrm{d}s \\
&= A + qCA(\log B - \log A^{1/q}).
\end{aligned}
\tag{4.8}
$$

This is the desired moment estimate.

Let us now start the proof of (1.4). Thanks to Proposition 2.5, the decomposition $K - \mathbb{E}K = \sum D_k$ (with $D_k = K_k - K_{k+1}$) gives

$$
\|K - \mathbb{E}K\|_{L^{q,w}}^q \leqslant C \sum \|D_k\|_{L^{q,w}}^q.
$$

We have $D_k(x) = 0$ if $Tx \notin \Delta_0$, and otherwise Lemma 4.2 gives the bound

$$
|D_k(x)| \leqslant \sum_{a=0}^{k-h} \operatorname{Lip}_a(K) \min\left((h+1)c_{k-h-a}^{(q)}, c_{k-h-a}^{(q-1)}\right) + \sum_{a=k-h+1}^{k} \operatorname{Lip}_a(K),
$$

where $h = h(x)$. We should bound the weak L^q norm of both terms on the right to conclude. Let us denote them by $U_k(x)$ and $V_k(x)$.

We start with V_k. Fix $s \geqslant 0$. Let $h_0(k)$ be minimal such that $\sum_{a=k-h_0+1}^{k} \operatorname{Lip}_a(K) \geqslant s$. Then

$$
\mu\left\{x \in T^{-1}\Delta_0 : \sum_{a=k-h+1}^{k} \operatorname{Lip}_a(K) \geqslant s\right\} = \mu\{x \in T^{-1}\Delta_0 : h(x) \geqslant h_0\}.
$$

This measure is exactly $\mu\{y \in \Delta_0 : \varphi(y) > h_0\} = O(h_0^{-q})$. Hence,

$$
\begin{aligned}
\mu\left\{x : \sum_{a=k-h+1}^{k} \operatorname{Lip}_a(K) \geqslant s\right\} &\leqslant C(1+h_0)^{-q} \leqslant C(1+h_0)^{-q} \left(\sum_{a=k-h_0+1}^{k} \operatorname{Lip}_a(K)/s\right)^q \\
&\leqslant Cs^{-q}M(k)^q,
\end{aligned}
$$

where $M(k)$ is the maximal function associated to $\operatorname{Lip}_i(K)$, i.e.,

$$
M(k) = \sup_{h \geqslant 0} \frac{1}{2h+1} \sum_{i=k-h}^{k+h} \operatorname{Lip}_i(K).
\tag{4.9}
$$

We have proved that $\|V_k\|_{L^{q,w}}^q \leqslant CM(k)^q$. Summing over k, we obtain

$$
\sum \|V_k\|_{L^{q,w}}^q \leqslant C \sum_k M(k)^q \leqslant C \sum \operatorname{Lip}_i(K)^q,
$$

since M is bounded in $\ell^q(\mathbb{Z})$ by $C\|\operatorname{Lip}_i(K)\|_{\ell^q(\mathbb{Z})}$, by Theorem 2.3. This is the desired upper bound.

We turn to U_k. We have

$$
\begin{aligned}
\sum_k \|U_k\|_{L^{q,w}}^q &\leqslant \sum_k \|U_k\|_{L^q}^q \\
&\leqslant \sum_k \sum_h \mu(\varphi = h+1) \left(\sum_{a=0}^{k-h} \operatorname{Lip}_a(K) \min\left((h+1)c_{k-h-a}^{(q)}, c_{k-h-a}^{(q-1)}\right)\right)^q.
\end{aligned}
$$

The next lemma shows that this is bounded by $C \sum \mathrm{Lip}_i(K)^q$ (set $\varepsilon = q - 1$, $n = k - h$, $i = a$ and $u_i = \mathrm{Lip}_i(K)$ to reduce to this statement). This concludes the proof. $\square$

Lemma 4.5. *Let $q > 1$ and $\varepsilon > 0$. Consider a sequence $a_n \geqslant 0$ with $\sum_{n \geqslant N} a_n = O(N^{-q})$. There exists a constant C such that, for any sequence $(u_n) \in \ell^q(\mathbb{Z})$,*

$$\sum_{n \in \mathbb{Z}} \sum_{h \geqslant 0} a_h \left| \sum_{i \in \mathbb{Z}} u_i \cdot \min \left(\frac{h+1}{1 + |n-i|^{1+\varepsilon}}, \frac{1}{1 + |n-i|^{\varepsilon}} \right) \right|^q \leqslant C \sum_{n \in \mathbb{Z}} |u_n|^q.$$

Proof. We proceed as in the proof of Lemma 4.4. Define a sequence

$$v(n, h) = \frac{1}{(1+h)^{1-\varepsilon}} \sum_{i \in \mathbb{Z}} u_i \cdot \min \left(\frac{h+1}{1 + |n-i|^{1+\varepsilon}}, \frac{1}{1 + |n-i|^{\varepsilon}} \right).$$

We consider it as a function on the space $\mathbb{Z} \times \mathbb{N}$ with the measure $\nu(n, h) = a_h (1+h)^{q(1-\varepsilon)}$. We have for any $n \in \mathbb{Z}$

$$\sum_{i \in \mathbb{Z}} \min \left(\frac{h+1}{1 + |n-i|^{1+\varepsilon}}, \frac{1}{1 + |n-i|^{\varepsilon}} \right)$$
$$\leqslant \sum_{|m| \leqslant h} \frac{1}{1 + |m|^{\varepsilon}} + (h+1) \sum_{|m| > h} \frac{1}{1 + |m|^{1+\varepsilon}} \leqslant C(1+h)^{1-\varepsilon}.$$

This shows that the operator $A : u \mapsto v$ is bounded from $\ell^{\infty}(\mathbb{Z}, \mu)$ (where μ is the counting measure) to $\ell^{\infty}(\mathbb{Z} \times \mathbb{N}, \nu)$. Moreover, writing $m = n - i$,

$$\sum_{n, h} \nu(n, h) v(n, h)$$
$$= \sum_{n \in \mathbb{Z}} \sum_{h \geqslant 0} a_h (1+h)^{q(1-\varepsilon)} \frac{1}{(1+h)^{1-\varepsilon}} \sum_{i \in \mathbb{Z}} u_i \cdot \min \left(\frac{h+1}{1 + |n-i|^{1+\varepsilon}}, \frac{1}{1 + |n-i|^{\varepsilon}} \right)$$
$$= \sum_{i \in \mathbb{Z}} u_i \cdot \sum_{h \geqslant 0} a_h (1+h)^{(q-1)(1-\varepsilon)} \sum_{m \in \mathbb{Z}} \min \left(\frac{h+1}{1 + |m|^{1+\varepsilon}}, \frac{1}{1 + |m|^{\varepsilon}} \right).$$

As we have seen above, the last sum over m is $O((1+h)^{1-\varepsilon})$. Hence, the sum over h and m reduces to $\sum_h a_h (1+h)^{q(1-\varepsilon)}$, which is finite by Lemma 2.6. This shows that $\|v\|_{\ell^1(\nu)} \leqslant C \|u\|_{\ell^1(\mu)}$.

The operator $A : u \mapsto v$ is bounded from $\ell^r(\mu)$ to $\ell^r(\nu)$ for $r = 1$ and ∞. By interpolation (see Theorem 2.2), it is also bounded from $\ell^q(\mu)$ to $\ell^q(\nu)$. This is the desired inequality. $\square$

4.3 The case $q = 2$

In this paragraph, we prove Theorem 1.9 in the case $q = 2$. As we explained after the statement of this theorem, it suffices to prove the result for $p \geqslant 2$. We follow essentially the same steps as in the $q > 2$ case. We start with Burkholder-Rosenthal inequality (2.3)

$$\int |K(x, Tx, \dots) - \mathbb{E}(K)|^p \, \mathrm{d}\mu(x)$$
$$\leqslant C \left(\sum \|\mathbb{E}(D_k^2 \mid \mathcal{F}_{k+1})\|_{\infty} \right)^{p/2} + C \sum \|\mathbb{E}(|D_k|^p \mid \mathcal{F}_{k+1})\|_{\infty}.$$

Moreover, for $r \in \{2, p\}$, we have

$$
\|\mathbb{E}(|D_k|^r \mid \mathcal{F}_{k+1})\|_\infty \leqslant C \sum_{h \geqslant 0} \mu(\varphi = h+1) \left(\sum_{a=0}^{k-h} \mathrm{Lip}_a(K) \min\left((h+1)c_{k-h-a}^{(q)}, c_{k-h-a}^{(q-1)}\right) \right)^r
$$

$$
+ C \sum_{h \geqslant 0} \mu(\varphi = h+1) \left(\sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K) \right)^r. \tag{4.10}
$$

Let us first consider the contribution of the first line when we sum over k. For $r = 2$, Lemma 4.5 shows that the resulting term is bounded by $C \sum \mathrm{Lip}_a(K)^2$. Its contribution to Burkholder-Rosenthal inequality is therefore at most

$$
C \left(\sum \mathrm{Lip}_i(K)^2 \right)^{p/2} = C \left(\sum \mathrm{Lip}_i(K)^2 \right) \cdot \left(\sum \mathrm{Lip}_i(K)^2 \right)^{p/2-1}
$$

$$
\leqslant C \left(\sum \mathrm{Lip}_i(K)^2 \right) \left(\sum \mathrm{Lip}_i(K) \right)^{p-2},
$$

since $\sum x_i^2 \leqslant (\sum x_i)^2$. This bound is compatible with the statement of the theorem. For $r = p$, we write

$$
\left(\sum_{a=0}^{k-h} \mathrm{Lip}_a(K) \min\left((h+1)c_{k-h-a}^{(q)}, c_{k-h-a}^{(q-1)}\right) \right)^p
$$

$$
\leqslant \left(\sum_{a=0}^{k-h} \mathrm{Lip}_a(K) \min\left((h+1)c_{k-h-a}^{(q)}, c_{k-h-a}^{(q-1)}\right) \right)^2 \cdot \left(\sum_{a \in \mathbb{Z}} \mathrm{Lip}_a(K) \right)^{p-2}.
$$

Using again Lemma 4.5, it follows that the contribution of this term to Burkholder-Rosenthal inequality is at most $\left(\sum \mathrm{Lip}_i(K)^2 \right) \left(\sum \mathrm{Lip}_i(K) \right)^{p-2}$ as desired.

Let us now turn to the second line of (4.10). We define a sequence

$$
v(k, h) = \sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K)
$$

on the space $\mathbb{Z} \times \mathbb{N}$ with the measure $\nu(k, h) = \mu(\varphi = h+1)$. It satisfies $\|v\|_{\ell^\infty} \leqslant \sum_{a \in \mathbb{Z}} \mathrm{Lip}_a(K)$. Let us control its weak L^2 norm. Let $s \geqslant 0$. For fixed k, let $h_0(k)$ be the smallest h such that $\sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K) \geqslant s$. Then

$$
\nu\{(k, h) \,:\, v(k, h) \geqslant s\} = \sum_k \sum_{h \geqslant h_0(k)} \mu(\varphi = h+1) \leqslant C \sum_k (1 + h_0(k))^{-2}
$$

$$
\leqslant C \sum_k (1 + h_0(k))^{-2} \left(\sum_{a=k-h_0(k)+1}^{k} \mathrm{Lip}_a(K)/s \right)^2 \leqslant Cs^{-2} \sum_k M(k)^2,
$$

where $M(k)$ is the maximal function associated to $\mathrm{Lip}_a(K)$, defined in (4.9). By Theorem 2.3, it satisfies $\sum_k M(k)^2 \leqslant C \sum \mathrm{Lip}_a(K)^2$. Hence, we have proved that the weak L^2 norm of v is bounded by $C \left(\sum \mathrm{Lip}_a(K)^2 \right)^{1/2}$.

Using the bounds on the weak L^2 norm of v and on its L^∞ norm, one deduces a bound on its strong L^2 norm as in (4.8), and on its strong L^p norm for $p > 2$ as in (4.7).

These bounds read:

$$\sum_k \sum_h \mu(\varphi = h + 1) \left(\sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K) \right)^2$$
$$\leqslant C \left(\sum_{a \in \mathbb{Z}} \mathrm{Lip}_a(K)^2 \right) \left[1 + \log \left(\sum_{a \in \mathbb{Z}} \mathrm{Lip}_a(K) \right) - \log \left(\sum_{a \in \mathbb{Z}} \mathrm{Lip}_a(K)^2 \right)^{1/2} \right]$$

and for $p > 2$

$$\sum_k \sum_h \mu(\varphi = h + 1) \left(\sum_{a=k-h+1}^{k} \mathrm{Lip}_a(K) \right)^p \leqslant C \left(\sum_{a \in \mathbb{Z}} \mathrm{Lip}_a(K)^2 \right) \left(\sum_{a \in \mathbb{Z}} \mathrm{Lip}_a(K) \right)^{p-2}.$$

For $p = 2$, we deduce directly that the contribution of the second line of (4.10) to Burkholder-Rosenthal inequality is bounded as in the statement of the theorem.

For $p > 2$, we also obtain that the contribution of this line, for $r = p$, is bounded as desired. It remains to check the contribution of this line with $r = 2$. Writing $u_a = \mathrm{Lip}_a(K)$, we should prove that

$$\left(\sum_{a \in \mathbb{Z}} u_a^2 \right)^{p/2} \left[1 + \log \left(\sum_{a \in \mathbb{Z}} u_a \right) - \log \left(\sum_{a \in \mathbb{Z}} u_a^2 \right)^{1/2} \right]^{p/2} \leqslant C \left(\sum_{a \in \mathbb{Z}} u_a^2 \right) \left(\sum_{a \in \mathbb{Z}} u_a \right)^{p-2}.$$

Since this equation is homogeneous, it suffices to prove it when $\sum u_a^2 = 1$. In this case, writing $x = \sum u_a \geqslant 1$, it reduces to the inequality $(1 + \log x)^{p/2} \leqslant C x^{p-2}$, which is trivial on $[1, \infty)$. $\qquad \square$

A Speed of convergence to stable laws

In this appendix, our goal is to prove Proposition 1.3. To do so, we estimate the speed of convergence of the Birkhoff sums to the stable law, first on the basis Δ_0 of the tower using the Nagaev-Guivarc'h spectral method. Then, we induce back those estimates to the whole tower. Those ideas are classical: the first step comes from [1], the second step from [24] (see [17] for a general explanation of the method). However, since we want quantitative estimates, we need to go beyond the results of these articles.

The standing assumptions are those of Proposition 1.3: (Δ, T) is a Young tower with $\mathrm{tail}_n = Cn^{-q} + O(n^{-q-\varepsilon})$, for some $q \in (1, 2)$ and some $\varepsilon > 0$. Without loss of generality, we can assume $q + \varepsilon < 2$.

Let $Y = \Delta_0$ be the basis of the Young tower. We denote by $T_Y : Y \to Y$ the induced map on the basis, by $\mu_Y = \mu_{|Y} / \mu(Y)$ the induced probability measure, by S_n^Y the Birkhoff sums for T_Y, and by $\varphi : Y \to \mathbb{N}$ the first return time to Y.

We define a function f on the tower, by $f = 1 - 1_Y / \mu(Y)$, so that $\int f \, \mathrm{d}\mu = 0$. The induced function on the basis of the tower is by definition

$$f_Y(x) = \sum_{k=0}^{\varphi(x)-1} f(T^k x) = \varphi(x) - 1_Y / \mu(Y).$$

Denote by L the transfer operator associated to T_Y, and define a family of perturbed transfer operators $L_t(u) = L(e^{\mathrm{i}t f_Y} u)$. Their interest is that

$$\int_Y e^{\mathrm{i}t S_n^Y f_Y} \, \mathrm{d}\mu_Y = \int_Y L_t^n 1 \, \mathrm{d}\mu_Y. \tag{A.1}$$

Hence, spectral properties of L_t make it possible to understand the characteristic function of $S_n^Y f_Y$, and therefore its closeness to the limiting stable law.

Lemma A.1. *The family of operators $t \mapsto L_t$ is C^1.*

Proof. We omit the standard argument which shows in fact that the family is C^q, see for instance [1, Theorem 2.4]. □

The unperturbed operator $L = L_0$ has a simple eigenvalue at 1, and the rest of its spectrum is contained in a disk of strictly smaller radius. This spectral description persists for small t, see [18]. Denote by λ_t the leading eigenvalue of L_t, by Π_t the corresponding (one-dimensional) spectral projection, and by $Q_t = L_t - \lambda_t \Pi_t$ the part of L_t corresponding to the rest of the spectrum. All those quantities depend in a C^1 way on t, by the previous proposition. Moreover, for small t, we have

$$L_t^n = \lambda_t^n \Pi_t + Q_t^n, \quad \text{with } \|Q_t^n\| \leqslant Cr^n, \tag{A.2}$$

for some fixed $r < 1$. The main contribution in this equation comes from the perturbed eigenvalue λ_t.

Lemma A.2. *We have for small $t > 0$*

$$\lambda_t = 1 + ct^q + O(t^{q+\varepsilon}),$$

where c is a complex number with $\Re c < 0$.

Proof. Let ξ_t denote the C^1 family of eigenfunctions of L_t for the eigenvalue λ_t, normalized so that $\int_Y \xi_t \, d\mu_Y = 1$. In particular, $\xi_0 \equiv 1$.

Now

$$
\begin{aligned}
\lambda_t &= \int_Y \lambda_t \xi_t \, d\mu_Y = \int_Y L_t \xi_t \, d\mu_Y = \int_Y L_t 1 \, d\mu_Y + \int_Y (L_t - L)(\xi_t - 1) \, d\mu_Y \\
&= \int_Y e^{\mathbf{i}t f_Y} \, d\mu_Y + O(t^2) \\
&= 1 + \int_Y (e^{\mathbf{i}t f_Y} - 1 - \mathbf{i}t f_Y) \, d\mu_Y + O(t^2).
\end{aligned}
$$

Let $G(s) = \mu_Y(f_Y < s)$ denote the distribution function of f_Y. It vanishes for $s < -M$, for $M = 1 - 1/\mu(Y)$. The asymptotics of the tails of the return time yield $1 - G(s) = c_1 s^{-q} + A(s)$ where $A(s) = O(s^{-q-\varepsilon})$. Hence

$$
\begin{aligned}
\int_Y (e^{\mathbf{i}t f_Y} - 1 - \mathbf{i}t f_Y) \, d\mu_Y &= \int_{-M}^\infty (e^{\mathbf{i}ts} - 1 - \mathbf{i}ts) \, dG(s) \\
&= -\int_{-M}^\infty (e^{\mathbf{i}ts} - 1 - \mathbf{i}ts) \, d(1 - G(s)) \\
&= (e^{-\mathbf{i}tM} - 1 + \mathbf{i}tM) + \mathbf{i}t \int_{-M}^\infty (e^{\mathbf{i}ts} - 1)(1 - G(s)) \, ds \\
&= \mathbf{i}t \int_1^\infty (e^{\mathbf{i}ts} - 1)(1 - G(s)) \, ds + O(t^2) \\
&= \mathbf{i} \int_t^\infty (e^{\mathbf{i}\sigma} - 1)(1 - G(\sigma/t)) \, d\sigma + O(t^2) \\
&= \mathbf{i}c_1 t^q \int_t^\infty (e^{\mathbf{i}\sigma} - 1)\sigma^{-q} \, d\sigma + \mathbf{i} \int_t^\infty (e^{\mathbf{i}\sigma} - 1)A(\sigma/t) \, d\sigma + O(t^2) \\
&= c_2 t^q + E_1 + E_2 + O(t^2),
\end{aligned}
$$

where

$$c_2 = \mathbf{i}c_1 \int_0^\infty (e^{\mathbf{i}\sigma} - 1)\sigma^{-q}\,\mathrm{d}\sigma, \quad E_1 = -\mathbf{i}c_1 t^q \int_0^t (e^{\mathbf{i}\sigma} - 1)\sigma^{-q}\,\mathrm{d}\sigma,$$

$$E_2 = \mathbf{i} \int_t^\infty (e^{\mathbf{i}\sigma} - 1)A(\sigma/t)\,\mathrm{d}\sigma.$$

Note that c_2 is well-defined since $q \in (1,2)$. Also, $|E_1| \leq c_1 t^q \int_0^t \sigma^{-(q-1)}\,\mathrm{d}\sigma = O(t^2)$. There is a constant $C > 0$ such that $|A(s)| \leq Cs^{-q-\varepsilon}$ for $s \geq 1$. Hence

$$|E_2| \leq Ct^{q+\varepsilon} \int_t^\infty |e^{\mathbf{i}\sigma} - 1|\sigma^{-q-\varepsilon}\,\mathrm{d}\sigma \leqslant Ct^{q+\varepsilon} \int_0^\infty \min(\sigma^{1-q-\varepsilon}, \sigma^{-q-\varepsilon})\,\mathrm{d}\sigma = O(t^{q+\varepsilon}),$$

where the integral is finite since $q + \varepsilon \in (1,2)$. $\qquad\square$

Let Z_Y be the real probability distribution whose characteristic function is given for $t > 0$ by

$$\mathbb{E}(e^{\mathbf{i}tZ_Y}) = e^{\mathbf{i}ct^q},$$

where c is given by Lemma A.2. It is a totally asymmetric stable law of index q. We can now estimate the speed of convergence of $S_n^Y f_Y$ to Z_Y:

Proposition A.3. *There exists $C > 0$ such that for any $n > 0$ and for any $s \in \mathbb{R}$,*

$$\left|\mu_Y\{x \,:\, S_n^Y f_Y(x)/n^{1/q} > s\} - \mathbb{P}(Z_Y > s)\right| \leqslant Cn^{-\varepsilon/q}.$$

In particular, we recover the (already known) convergence of $S_n^Y f_Y/n^{1/q}$ to Z_Y, the novelty being the control on the speed of convergence. Below, in Proposition A.4 and Theorem A.5, we also recover known stable limits, with additional controls on the speed of convergence.

Proof. The quantity to estimate is the L^∞-norm of the difference between the distribution functions of $S_n^Y f_Y/n^{1/q}$ and Z_Y. Berry-Esseen's lemma (see for instance [13, Lemma XVI.3.2]) ensures that, for any $M > 0$, this quantity is bounded by

$$C \int_0^M \frac{|\varphi_n(t) - \psi(t)|}{t}\,\mathrm{d}t + \frac{C}{M} \tag{A.3}$$

where C is a universal constant, and φ_n and ψ denote respectively the characteristic functions of $S_n^Y f_Y/n^{1/q}$ and Z_Y. We estimate this integral, taking $M = M_n = \alpha_0 n^{1/q}$ for some suitably small α_0.

First, for $t < 1/n$, we have

$$|\varphi_n(t) - 1| = |\mathbb{E}(e^{\mathbf{i}tS_n^Y f_Y/n^{1/q}} - 1)| \leqslant t \int |S_n^Y f_Y|\,\mathrm{d}\mu_Y/n^{1/q} \leqslant Ctn^{1-1/q}.$$

In the same way, $|\psi(t) - 1| \leqslant Ct^q$. Hence,

$$\int_0^{1/n} \frac{|\varphi_n(t) - \psi(t)|}{t} \leqslant Cn^{-1/q} + Cn^{-q} \leqslant Cn^{-1/q}.$$

Now, we turn to the interval $t \in [1/n, M_n]$. Combining the formula (A.1) for φ_n and the spectral expansion (A.2) of L_t^n, we get

$$\varphi_n(t) = \lambda(t/n^{1/q})^n u(t/n^{1/q}) + r_n(t/n^{1/q}),$$

where r_n is exponentially small, u is a C^1 function close to 0 and the asymptotic expansion of λ is given in Lemma A.2. The contribution of r_n to the integral (A.3) is exponentially small (this is why we had to discard the interval $[0, 1/n]$). We can write $\lambda(s) = e^{cs^q + B(s)}$ where $B(s) = O(s^{q+\varepsilon})$, by Lemma A.2. Hence,

$$\lambda(t/n^{1/q})^n = e^{n(ct^q/n + B(t/n^{1/q}))} = e^{ct^q} e^{nB(t/n^{1/q})} = \psi(t)e^{nB(t/n^{1/q})}.$$

The remaining part of the integral (A.3) can be written as

$$\int_{1/n}^{M_n} \frac{|\psi(t)e^{nB(t/n^{1/q})}u(t/n^{1/q}) - \psi(t)|}{t}\,\mathrm{d}t$$

$$\leqslant \int_0^{M_n} \left|\psi(t)e^{nB(t/n^{1/q})}\right| \frac{|u(t/n^{1/q}) - 1|}{t}\,\mathrm{d}t + \int_0^{M_n} |\psi(t)| \frac{|e^{nB(t/n^{1/q})} - 1|}{t}\,\mathrm{d}t =: I_1 + I_2.$$

In I_1, we have

$$\frac{n|B(t/n^{1/q})|}{t^q} \leqslant C(t/n^{1/q})^{q+\varepsilon-q} \leqslant C\alpha_0^\varepsilon.$$

Hence, $\left|\psi(t)e^{nB(t/n^{1/q})}\right| \leqslant e^{\Re(c)t^q} e^{C\alpha_0^\varepsilon t^q}$. If α_0 is small enough, this is bounded by e^{-at^q}, for some $a > 0$. Since the function u is C^1 with $u(0) = 1$, it follows that

$$I_1 \leqslant C \int_0^{M_n} e^{-at^q} n^{-1/q}\,\mathrm{d}t \leqslant Cn^{-1/q}.$$

Finally, in I_2, we use the inequality $|e^s - 1| \leqslant |s|e^{|s|}$, to get a bound

$$I_2 \leqslant \int_0^{M_n} \left|\psi(t)e^{n|B(t/n^{1/q})|}\right| \frac{n|B(t/n^{1/q})|}{t}\,\mathrm{d}t.$$

As above, the factor $\left|\psi(t)e^{n|B(t/n^{1/q})|}\right|$ is bounded by e^{-at^q}. Moreover, the second factor is bounded by $t^{q+\varepsilon-1}n^{-\varepsilon/q}$. This gives $I_2 \leqslant Cn^{-\varepsilon/q}$.

Finally, we obtain a bound for (A.3) of the form $Cn^{-1/q} + Cn^{-\varepsilon/q}$, which is bounded by $Cn^{-\varepsilon/q}$ as $\varepsilon < 2 - q < 1$. $\qquad\square$

We can then lift the above bound to the original Birkhoff sums $S_n f$. Let $Z = \mu(Y)^{1/q}Z_Y$, it is again a (completely asymmetric) stable law of index q.

Proposition A.4. *Let* $\delta = \min((q-1)/(1+2q^2), \varepsilon/q) > 0$. *There exists* $C > 0$ *such that for any* $n > 0$ *and for any* $s \in \mathbb{R}$,

$$\left|\mu_Y\{x \,:\, S_n f(x)/n^{1/q} > s\} - \mathbb{P}(Z > s)\right| \leqslant Cn^{-\delta}.$$

Proof. For $x \in Y$, the Birkhoff sums $S_n f(x)$ and $S_{n\mu(Y)}^Y f_Y(x)$ should be close (since a return to Y takes on average $1/\mu(Y)$ iterates of T, both sums involve roughly the same number of iterations of T), and we know that $S_{n\mu(Y)}^Y f_Y(x)/(n\mu(Y))^{1/q}$ is close to Z_Y in distribution. (We write $n\mu(Y)$ instead of its integer part for notational simplicity.) The result follows if we can show that the different errors are suitably small.

Define a function H on Δ as follows: if x is at height i (i.e., it belongs to $\Delta_{\alpha,i}$ for some α), let $\pi x = T^{-i}x$ be its unique preimage in the basis, and let $H(x) = \sum_{j=0}^{i-1} f(T^j \pi x)$. Let $N(n, x)$ denote the number of returns to Y of a point $x \in Y$ before time n. We get $S_n f(x) = S_{N(n,x)}^Y f_Y(x) + H(T^n x)$. We expect $N(n, x)$ to be close to $n\mu(Y)$, hence we decompose further as

$$S_n f(x) = S_{n\mu(Y)}^Y f_Y(x) + (S_{N(n,x)}^Y f_Y(x) - S_{n\mu(Y)}^Y f_Y(x)) + H(T^n x)$$

$$= S_{n\mu(Y)}^Y f_Y(x) + E_n(x) + F_n(x).$$

Suppose that, for $u_n = n^{-\delta}$ for some $\delta \in (0, \varepsilon/q]$, we have

$$\mu_Y\{|E_n|/n^{1/q} > u_n\} \leqslant C u_n, \quad \mu_Y\{|F_n|/n^{1/q} > u_n\} \leqslant C u_n. \tag{A.4}$$

We deduce from the above equation that

$$\mu_Y\{S_n f/n^{1/q} > s\} \leqslant \mu_Y\{S^Y_{n\mu(Y)} f_Y/n^{1/q} > s - 2u_n\} + 2C u_n.$$

By Proposition A.3, this is bounded by

$$\mathbb{P}(\mu(Y)^{1/q} Z_Y > s - 2u_n) + C n^{-\varepsilon/q} + 2C u_n.$$

As Z_Y has a bounded density, the probability that $\mu(Y)^{1/q} Z_Y$ belongs to the interval $[s - 2u_n, s)$ is bounded by $C u_n$. Finally, we obtain

$$\mu_Y\{S_n f/n^{1/q} > s\} \leqslant \mathbb{P}(\mu(Y)^{1/q} Z_Y > s) + C n^{-\varepsilon/q} + C u_n.$$

The lower bound is similar, and we obtain the conclusion of the proposition.

It remains to prove (A.4). We first deal with the bound involving F_n. We have

$$\mu(F_n \geqslant u_n n^{1/q}) = \mu(H \circ T^n \geqslant u_n n^{1/q}) = \mu(H \geqslant u_n n^{1/q}).$$

The function H can only be $\geqslant A$ on the set of points with height at least A. The set of points with height i has measure $\text{tail}_{i+1} \sim C i^{-q}$, hence $\mu(H \geqslant A) \leqslant C A^{-q+1}$. We get

$$\mu(F_n \geqslant u_n n^{1/q}) \leqslant C(u_n n^{1/q})^{-(q-1)}.$$

This is bounded by $C u_n$ if $u_n = n^{-\delta}$ with $\delta \leqslant (q-1)/q^2$.

We turn to E_n. Let $M_n = n^r$, for some $r \in (1/q, 1)$. We have

$$\{E_n \geqslant u_n n^{1/q}\} \subset \{E_n(x) \geqslant u_n n^{1/q}, \ |N(n, x) - n\mu(Y)| < M_n\} \cup \{|N(n, x) - n\mu(Y)| \geqslant M_n\}.$$

In the first set, as $S^Y_{N(n,x)} f_Y(x)$ and $S^Y_{n\mu(Y)} f_Y(x)$ are separated by $u_n n^{1/q}$, one of them is distant from $S^Y_{n\mu(Y) - M_n} f_Y(x)$ by at least $u_n n^{1/q}/2$. Hence, the first set is included in

$$\left\{ \max_{0 \leqslant k \leqslant 2M_n} |S^Y_{n\mu(Y) - M_n + k} f_Y(x) - S^Y_{n\mu(Y) - M_n} f_Y(x)| \geqslant u_n n^{1/q}/2 \right\}.$$

By the invariance of the measure μ_Y under T_Y, the measure of this set is

$$\mu_Y \left\{ \max_{0 \leqslant k \leqslant 2M_n} |S^Y_k f_Y| \geqslant u_n n^{1/q}/2 \right\}.$$

The sequence $S^Y_i f_Y/i^{1/q}$ converges in distribution, but more is true: It follows from [9, Lemma 7.1 and proof of Theorem 2.10] that this sequence remains bounded in L^1, and that the weak L^1 norm of the corresponding maxima also remain bounded. Hence, the above equation is bounded by $C M_n^{1/q}/(u_n n^{1/q})$. This is bounded by $C u_n$ if $u_n = n^{-\delta}$ with $\delta \leqslant (1-r)/(2q)$.

Finally, if $|N(n, x) - n\mu(Y)| \geqslant M_n$, then either $N(n, x) \geqslant n\mu(Y) + M_n$, or $N(n, x) \leqslant n\mu(Y) - M_n$. In the first case, $S^Y_{n\mu(Y) + M_n} \varphi_Y \leqslant n$, i.e., $S^Y_{n\mu(Y) + M_n} f_Y \leqslant -M_n/\mu(Y)$. By Proposition A.3, this can only happen with probability $\mathbb{P}(Z_Y \leqslant -c M_n \mu(Y)/n^{1/q}) + C n^{-\varepsilon/q}$. The stable law Z_Y has tails of order q, i.e., $\mathbb{P}(|Z_Y| > s) \leqslant C s^{-q}$. Hence, this is bounded by $u_n = n^{-\delta}$ if $\delta \leqslant \min(q(r - 1/q), \varepsilon/q)$. The second case is handled similarly.

We have proved that, if δ is small enough, then (A.4) holds. More specifically, we can choose r so that $q(r - 1/q) = (1 - r)/2q$, i.e., $r = (1 + 2q)/(1 + 2q^2)$. The resulting constraints on δ are

$$\delta \leqslant \min((q-1)/q^2, (q-1)/(1 + 2q^2), \varepsilon/q).$$

The first constraint can be removed since it is implied by the second one. $\qquad \square$

We can now conclude the proof of Proposition 1.3. The probability distribution Z has heavy tails, since it is a stable law of index q: there exists $c > 0$ such that, for all $s \geqslant 1$, we have $\mathbb{P}(Z > s) \geqslant cs^{-q}$. It follows from Proposition A.4 that $\mu_Y\{S_nf/n^{1/q} > s\} \geqslant cs^{-q} - Cn^{-\delta}$. This is $\geqslant cs^{-q}/2$ if $Cn^{-\delta} \leqslant cs^{-q}/2$, which holds for $s \in [1, n^r]$ if $r < \delta/q$ and n is large enough. In this range, it follows that $\mu\{S_nf/n^{1/q} > s\} \geqslant c's^{-q}$.

Using (1.3) for the first equality, we have

$$\int |S_nf/n^{1/q}|^q \, \mathrm{d}\mu = q \int_{s=0}^{\infty} s^{q-1} \mu\left\{|S_nf/n^{1/q}| \geqslant s\right\} \mathrm{d}s \geqslant q \int_{s=1}^{n^r} s^{q-1} c' s^{-q} \, \mathrm{d}s = c'qr \log n.$$

This is the desired lower bound. $\qquad\square$

One can also deduce from Proposition A.4 a speed of convergence towards the stable law Z on the whole space (Δ, μ). Although this is not needed for Proposition 1.3, we include it for completeness:

Theorem A.5. *Let* $\delta = \min((q-1)/(1+2q^2), \varepsilon/q)$. *There exists* $C > 0$ *such that for any* $n > 0$ *and for any* $s \in \mathbb{R}$,

$$\left|\mu\{x \ : \ S_nf(x)/n^{1/q} > s\} - \mathbb{P}(Z > s)\right| \leqslant Cn^{-\delta}.$$

Proof. Consider a set $\Delta_{\alpha,i}$, together with its renormalized probability measure $\mu_{\alpha,i} = \mu_{|\Delta_{\alpha,i}}/\mu(\Delta_{\alpha,i})$. This measure is sent by $T_*^{h_\alpha - i}$ to a measure on Y, which is equivalent to μ_Y, with a density bounded from above and from below, and with uniformly bounded Lipschitz constant. Proposition A.3 still works for this measure, with uniform constants, since all we need to apply the spectral argument is that the density is Lipschitz. It follows that Proposition A.4 also works for these measures. Adding the additional error coming from the $h_\alpha - i$ first steps needed to reach Y (bounded by $(h_\alpha - i)/n^{1/q}$), we deduce: for $n > h_\alpha - i$,

$$\left|\mu_{\alpha,i}\{x \in \Delta_{\alpha,i} \ : \ S_nf(x)/n^{1/q} > s\} - \mathbb{P}(Z > s)\right| \leqslant C(n - (h_\alpha - i))^{-\delta} + C(h_\alpha - i)/n^{1/q}.$$

Let Λ_k denote the set of points in Δ that enter Y after exactly k steps. Multiplying the above inequality by $\mu(\Delta_{\alpha,i})$ and summing over (α, i), we obtain:

$$\left|\mu\{x \in \Delta \ : \ S_nf(x)/n^{1/q} > s\} - \mathbb{P}(Z > s)\right|$$
$$\leqslant C \sum_{k < n} \mu(\Lambda_k) \min((n-k)^{-\delta} + k/n^{1/q}, 1) + \sum_{k \geqslant n} \mu(\Lambda_k).$$

We have $\mu(\Lambda_k) = \mathrm{tail}_{k+1} \sim Ck^{-q}$. Splitting the above sum into $k \leqslant n^{1/q}$ and $k > n^{1/q}$, we get the bound

$$C \sum_{k \leqslant n^{1/q}} k^{-q}(n^{-\delta} + k/n^{1/q}) + C \sum_{k > n^{1/q}} k^{-q} \leqslant Cn^{-\delta} + Cn^{(2-q)/q}/n^{1/q} + Cn^{(1-q)/q}.$$

This is bounded by $Cn^{-\delta'}$ for $\delta' = \min(\delta, 1 - 1/q) = \delta$. $\qquad\square$

References

[1] Jon Aaronson and Manfred Denker, *Local limit theorems for partial sums of stationary sequences generated by Gibbs-Markov maps*, Stoch. Dyn. **1** (2001), 193–237. MR-1840194

[2] Douglas N. Armstead, Brian R. Hunt, and Edward Ott, *Anomalous diffusion in infinite horizon billiards*, Phys. Rev. E (3) **67** (2003), 021110, 7 pages. MR-1974617

[3] Sergey Astashkin, Fedor Sukochev, and Chin Pin Wong, *Disjointification of martingale differences and conditionally independent random variables with some applications*, Studia Math. **205** (2011), 171–200. MR-2824894

[4] Péter Bálint, Nikolai Chernov, and Dmitry Dolgopyat, *Limit theorems for dispersing billiards with cusps*, Comm. Math. Phys. **308** (2011), 479–510. MR-2851150

[5] ———, *Convergence of moments for dispersing billiards with cusps*, Preprint, 2013.

[6] Péter Bálint and Sébastien Gouëzel, *Limit theorems in the stadium billiard*, Comm. Math. Phys. **263** (2006), 461–512. MR-2207652

[7] Michael Sh. Braverman, *Independent random variables and rearrangement invariant spaces*, London Mathematical Society Lecture Note Series, vol. 194, Cambridge University Press, Cambridge, 1994. MR-1303591

[8] Donald L. Burkholder, *Distribution function inequalities for martingales*, Ann. Probability **1** (1973), 19–42. MR-0365692

[9] Jean-René Chazottes and Sébastien Gouëzel, *On almost-sure versions of classical limit theorems for dynamical systems*, Probab. Theory Related Fields **138** (2007), 195–234. MR-2288069

[10] ———, *Optimal concentration inequalities for dynamical systems*, Comm. Math. Phys. **316** (2012), 843–889. MR-2993935

[11] M. Courbage, M. Edelman, S. M. Saberi Fathi, and G. M. Zaslavsky, *Problem of transport in billiards with infinite horizon*, Phys. Rev. E (3) **77** (2008), 036203, 5 pages. MR-2495438

[12] Jérôme Dedecker and Florence Merlevède, *Moment bounds for dependent sequences in smooth banach spaces*, Preprint, 2014.

[13] William Feller, *An introduction to probability theory and its applications. Vol. II*, John Wiley & Sons Inc., New York, 1966. MR-0210154

[14] Sébastien Gouëzel, *Central limit theorem and stable laws for intermittent maps*, Probab. Theory Related Fields **128** (2004), 82–122. MR-2027296

[15] ———, *Sharp polynomial estimates for the decay of correlations*, Israel J. Math. **139** (2004), 29–65. MR-2041223

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

[17] ———, *Limit theorems in dynamical systems using the spectral method*, preprint, 2013.

[18] Tosio Kato, *Perturbation theory for linear operators*, Die Grundlehren der mathematischen Wissenschaften, Band 132, Springer-Verlag New York, Inc., New York, 1966. MR-0203473

[19] Carlangelo Liverani, Benoît Saussol, and Sandro Vaienti, *A probabilistic approach to intermittency*, Ergodic Theory Dynam. Systems **19** (1999), 671–685. MR-1695915

[20] Ian Melbourne, *Large and moderate deviations for slowly mixing dynamical systems*, Proc. Amer. Math. Soc. **137** (2009), 1735–1741. MR-2470832

[21] Ian Melbourne and Matthew Nicol, *Large deviations for nonuniformly hyperbolic systems*, Trans. Amer. Math. Soc. **360** (2008), 6661–6676. MR-2434305

[22] Ian Melbourne and Dalia Terhesiu, *Operator renewal theory and mixing rates for dynamical systems with infinite measure*, Invent. Math. **189** (2012), 61–110. MR-2929083

[23] Ian Melbourne and Andrei Török, *Convergence of moments for Axiom A and non-uniformly hyperbolic flows*, Ergodic Theory Dynam. Systems **32** (2012), 1091–1100. MR-2995657

[24] Ian Melbourne and Andrew Török, *Statistical limit theorems for suspension flows*, Israel J. Math. **144** (2004), 191–209. MR-2121540

[25] Yves Pomeau and Paul Manneville, *Intermittent transition to turbulence in dissipative dynamical systems*, Comm. Math. Phys. **74** (1980), 189–197. MR-576270

[26] Omri Sarig, *Subexponential decay of correlations*, Invent. Math. **150** (2002), 629–653. MR-1946554

[27] Elias M. Stein and Guido Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton University Press, Princeton, N.J., 1971, Princeton Mathematical Series, No. 32. MR-0304972

[28] Domokos Szász and Tamás Varjú, *Limit laws and recurrence for the planar Lorentz process with infinite horizon*, J. Stat. Phys. **129** (2007), 59–80. MR-2349520

[29] Bengt von Bahr and Carl-Gustav Esseen, *Inequalities for the rth absolute moment of a sum of random variables,* $1 \leqslant r \leqslant 2$, Ann. Math. Statist **36** (1965), 299–303. MR-0170407

[30] Lai-Sang Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. (2) **147** (1998), 585–650. MR-1637655

[31] _____ , *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188. MR-1750438