Gaussian Concentration bound for potentials satisfying Walters condition with subexponential continuity rates

J.-R Chazottes, J Moles, E Ugalde

Gaussian Concentration bound for potentials satisfying Walters condition with subexponential continuity rates

J.-R. Chazottes [*1], J. Moles [†2], and E. Ugalde [‡2]

[1]Centre de Physique Théorique, Ecole Polytechnique, CNRS, 91128 Palaiseau, France.
Email: chazottes@cpht.polytechnique.fr
[2]Instituto de Física, Universidad Autónoma de San Luis Potosí, S.L.P., 78290 México
Emails: molesjdn@ifisica.uaslp.mx, ugalde@ifisica.uaslp.mx

Dated: September 10, 2019

Abstract

We consider the full shift $T : \Omega \to \Omega$ where $\Omega = A^{\mathbb{N}}$, A being a finite alphabet. For a class of potentials which contains in particular potentials ϕ with variation decreasing like $O(n^{-\alpha})$ for some $\alpha > 2$, we prove that their corresponding equilibrium state μ_ϕ satisfies a Gaussian concentration bound. Namely, we prove that there exists a constant $C > 0$ such that, for all n and for all separately Lipschitz functions $K(x_0, \ldots, x_{n-1})$, the exponential moment of $K(x, \ldots, T^{n-1}x) - \int K(y, \ldots, T^{n-1}y)\, \mathrm{d}\mu_\phi(y)$ is bounded by $\exp\left(C \sum_{i=0}^{n-1} \mathrm{Lip}_i(K)^2\right)$. The crucial point is that C is independent of n and K. We then derive various consequences of this inequality. For instance, we obtain bounds on the fluctuations of the empirical frequency of blocks, the speed of convergence of the empirical measure, and speed of Markov approximation of μ_ϕ. We also derive an almost-sure central limit theorem.

Keywords: concentration inequalities, empirical measure, Kantorovich distance, Wasserstein distance, d-bar distance, relative entropy, Markov approximation, almost-sure central limit theorem.

[*]JRC benefited from an ECOS Nord project for his stay at San Luis Potosí.
[†]JM benefited from an ECOS Nord project for his stay at Ecole Polytechnique.
[‡]EU acknowledges Ecole Polytechnique for financial support (one-month stay).

Contents

1 Introduction

We consider the full shift $T : \Omega \to \Omega$ where $\Omega = A^{\mathbb{N}}$, A being a finite alphabet. Given an ergodic measure μ on Ω and a continuous observable $f : \Omega \to \mathbb{R}$, we know by Birkhoff's ergodic theorem that $n^{-1}S_n f(x)$ converges, for μ-almost every x, to $\int f \mathrm{d}\mu$. (We use the standard notation $S_n f = f + f \circ T + \cdots + f \circ T^{n-1}$.) To refine this result, we need more assumptions on μ and f. For instance, if $\mu = \mu_\phi$ is the equilibrium state for a Lipschitz potential $\phi : \Omega \to \mathbb{R}$ and f is also a Lipschitz function, then the following central limit theorem holds:

$$\lim_{n\to\infty} \mu_\phi \left(x : \frac{S_n f(x) - n \int f \mathrm{d}\mu_\phi}{\sqrt{n}} \leq u \right) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{u} \mathrm{e}^{-\frac{\xi^2}{2\sigma^2}} \mathrm{d}\xi \qquad (1.1)$$

for all $u \in \mathbb{R}$, where $\sigma^2 = \sigma_f^2$ is the variance of the process $\{f(T^n x)\}_{n\geq 0}$ where x is distributed according to μ_ϕ. [1] This result says in essence that the fluctuations of $S_n f(x) - n \int f \mathrm{d}\mu_\phi$ are with high probability of order $\sqrt{n}$, when $n \to \infty$. Fluctuations of order n, referred to as 'large deviations', are unlikely

[1] $\sigma_f^2 = \int f^2 \mathrm{d}\mu_\phi - \left(\int f \mathrm{d}\mu_\phi \right)^2 + 2 \sum_{\ell \geq 1} \left(\int f \cdot f \circ T^\ell \mathrm{d}\mu_\phi - \left(\int f \mathrm{d}\mu_\phi \right)^2 \right)$

to appear. Indeed, for instance one has

$$\mu_\phi\left(x : \frac{1}{n} S_n f(x) \geq \int f \mathrm{d}\mu_\phi + u\right) \asymp \mathrm{e}^{-nI(u+\int f \mathrm{d}\mu_\phi)} \tag{1.2}$$

where $u \geq 0$ and $I(u) \geq 0$ is the so-called 'rate function' which is (strictly) convex, such that $I(\int f \mathrm{d}\mu_\phi) = 0$, and equal to $+\infty$ outside a certain finite interval $(\underline{u}_f, \bar{u}_f)$.[2] Of course, both the central limit theorem and the large deviation asymptotics have been obtained for more general potentials, and for more general 'chaotic' dynamical systems. For a fairly recent review on probabilistic properties of nonuniformly hyperbolic dynamical systems modeled by Young towers, we refer to [4].

In this paper, we are interested in concentration inequalities which describe the fluctuations of observables of the form $K(x, Tx, \dots, T^{n-1}x)$ around their average. The only restriction on K is that it has to be separately Lipschitz. By this we mean that, for all $i = 0, \dots, n-1$, there exists a constant $\mathrm{Lip}_i(K)$ with

$$|K(x_0, \dots, x_i, \dots, x_{n-1}) - K(x_0, \dots, x_i', \dots, x_{n-1})| \leq \mathrm{Lip}_i(K)\, d(x_i, x_i').$$

for all points $x_0, \dots, x_i, \dots, x_{n-1}, x_i'$ in Ω, where d is the usual distance on Ω (see (2.1)). So K can be nonlinear and implicitly defined. Of course, such a class contains partial sums of Lipschitz functions, namely functions of the form $K(x_0, \dots, x_{n-1}) = f(x_0) + \cdots + f(x_{n-1})$ for which $\mathrm{Lip}_i(K) = \mathrm{Lip}(f)$ for all i. Beside considering very general observables, the other essential characteristics of concentration inequalities is that they are valid for all n, contrarily to the above two results which are valid only in the limit $n \to \infty$. More precisely, we shall prove the following 'Gaussian concentration bound'. There exists a constant C such that, for all n and for all separately Lipschitz functions $K(x_0, \dots, x_{n-1})$, we have

$$\begin{aligned} &\int \exp\left(K\left(x, Tx, \dots, T^{n-1}x\right)\right) \mathrm{d}\mu_\phi(x) \\ &\leq \exp\left(\int K\left(x, Tx, \dots, T^{n-1}x\right) \mathrm{d}\mu_\phi(x)\right) \exp\left(C \sum_{j=0}^{n-1} \mathrm{Lip}_j(K)^2\right). \end{aligned} \tag{1.3}$$

The crucial point is that C is independent of n and K. By a standard argument (see below), the previous inequality implies that for all $u > 0$

$$\begin{aligned} &\mu_\phi\left(x : K\left(x, Tx, \dots, T^{n-1}x\right) - \int K\left(y, Ty, \dots, T^{n-1}y\right) \mathrm{d}\mu_\phi(y) \geq u\right) \\ &\leq \exp\left(-\frac{u^2}{4C \sum_{i=0}^{n-1} \mathrm{Lip}_i(K)^2}\right). \end{aligned} \tag{1.4}$$

[2] For two positive sequences $(a_n), (b_n)$, $a_n \asymp b_n$ means that $\lim_n (1/n) \log a_n = \lim_n (1/n) \log b_n$.

The Gaussian concentration bound (1.3) is known for Lipschitz potentials [7]. We shall prove that it remains true for a large subclass of potentials ϕ satisfying Walters condition. For instance, the bound holds for a potential whose variation is $O(n^{-\alpha})$ for some $\alpha > 2$. The proof of our result relies on two main ingredients. First, we start with a classical decomposition of $K - \int K$ as a telescopic sum of martingale differences. Second, we have to do a second telescoping to use Ruelle's Perrons-Frobenius operator. But we do not have a spectral gap anymore as in [7] (in the case of Lipschitz potentials). Instead, we use a result of V. Maume-Deschamps [18] based on Birkhoff cones.

We apply the Gaussian concentration bound and its consequences, like (1.4), to various observables. On the one hand, we obtain concentration bounds for previously studied observables. We get the same bounds but they are no more limited to equilibrium states with Lipschitz potentials. On the other hand, we consider observables not considered before. Even when $K(x, \dots, T^{n-1}x) = S_n f(x)$, we get a non-trivial bound. We then obtain a control on the fluctuations of the empirical frequency of blocks $a^0, \dots, a^{k-1}$ around $\mu([a^0, \dots, a^{k-1}])$, uniformly in $a^0, \dots, a^{k-1} \in A^k$. We then consider an estimator of the entropy μ_ϕ based on hitting times. The next application is about the speed of convergence of the empirical measure $(1/n)\sum_{i=0}^{n-1} \delta_{T^i x}$ towards μ_ϕ in Wasserstein distance. Then we obtain an upper bound for the $\bar{d}$-distance between any shift-invariant probability measure and μ_ϕ. This distance is bounded by the square root of their relative entropy, times a constant. A consequence of this inequality is a bound for the speed of convergence of the Markov approximation of μ_ϕ in $\bar{d}$-distance. Then we quantify the 'shadowing' of an orbit by another one which has to start in a subset of Ω with μ_ϕ-measure $1/3$, say. Finally, we prove an almost-sure version of the central limit theorem. This application shows in particular that concentration inequalities can also be used to obtain limit theorems.

2 Setting and preliminary results

Let $\Omega = A^{\mathbb{N}}$ where A is a finite set. We denote by $x = x^0 x^1 \dots$ the elements of Ω (hence $x^i \in A$), and by T the shift map: $(Tx)^k = x^{k+1}$, $k \in \mathbb{N}$. (We use upper indices instead of lower indices because we will need to consider bunches of points in Ω, *e.g.*, $x_0, x_1, \dots, x_p$, $x_i \in \Omega$.) We use the classical distance

$$d_\theta(x, y) = \theta^{\inf\{k : x^k \neq y^k\}} \tag{2.1}$$

where $\theta \in (0, 1)$ is some fixed number. Probability measures are defined on the Borel sigma-algebra of Ω which is generated by cylinder sets. Let $\phi : \Omega \to \mathbb{R}$ be a continuous potential, which means that

$$\mathrm{var}_n(\phi) := \sup\{|\phi(x) - \phi(y)| : x^i = y^i, 0 \leq i \leq n-1\} \xrightarrow{n\to\infty} 0.$$

The sequence $(\mathrm{var}_n(\phi))_{n\geq 1}$ is the modulus of continuity of ϕ and it is called the 'variation' of ϕ in our context. By the way, we denote by $\mathscr{C}(\Omega)$ the Banach space of real-valued continuous functions on Ω equipped with the supremum norm $\|\cdot\|_\infty$. We put further restrictions on ϕ, namely that it must satisfy the Walters condition [22]. For x, y in Ω let

$$W(\phi, x, y) = \sup_{n\in\mathbb{N}} \sup_{a\in A^n} |S_n\phi(ax) - S_n\phi(ay)| \ .$$

We assume that $W(\phi, x, y)$ exists and that there exists $W(\phi) > 0$ such that

$$\sup_{x,y\,\in\,\Omega} W(\phi, x, y) \leq W(\phi)\,. \tag{2.2}$$

Now for $p \in \mathbb{N}$ let

$$W_p(\phi) := \sup\{W(\phi, x, y) : x^i = y^i, 0 \leq i \leq p-1\}\,.$$

Definition 2.1. *ϕ is said to satisfy Walters' condition if $(W_p(\phi))_{p\in\mathbb{N}}$ is a strictly positive sequence and decreases to 0 as $p \to \infty$.*

We now make several remarks on Walters' condition. First, observe that locally constant potentials do not satisfy this condition because $W_p(\phi) = 0$ for all p larger than some p_0. But one can in fact work with any strictly positive sequence $(\widetilde{W}_p(\phi))_{p\in\mathbb{N}}$ decreasing to zero such that $W_p(\phi) \leq \widetilde{W}_p(\phi)$ for all p, *e.g.*, $\max(W_p(\phi), \eta^p)$ for some fixed $\eta \in (0, 1)$. Second, one easily checks that

$$\mathrm{var}_{p+1}(\phi) \leq W_p(\phi) \leq \sum_{k=p+1}^{\infty} \mathrm{var}_k(\phi)\,,\ p \in \mathbb{N}. \tag{2.3}$$

Hence the set of potentials satisfying Walters' condition contains the set of potentials with summable variation. In particular, $(W_p(\phi))_p$ is bounded above by a geometric sequence if and only if $(\mathrm{var}_p(\phi))_p$ is also bounded above by a geometric sequence. This corresponds to the case of Lipschitz or Hölder potentials (with respect to d_θ).

Now define Ruelle's Perron-Frobenius operator $P_\phi : \mathscr{C}(\Omega) \to \mathscr{C}(\Omega)$ as

$$P_\phi f(x) = \sum_{Ty=x} f(y)\, \mathrm{e}^{\phi(y)}\ .$$

The next step is to define a function space preserved by P_ϕ and on which it has good spectral properties. We take the space of Lipschitz functions with respect to a new distance d_ϕ built out of ϕ as follows.

Definition 2.2 (The distance d_ϕ)**.** *For $x, y \in \Omega$ let*

$$d_\phi(x, y) = W_p(\phi) \quad \text{if} \quad d_\theta(x, y) = \theta^p$$

and $d_\phi(x, x) = 0$.

Now define

$$\mathcal{L}_\phi = \{f \in \mathscr{C}(\Omega) : \exists M > 0 \text{ such that } \mathrm{var}_n(f) \leq MW_n(\phi), n = 1, 2, \ldots\}$$

and

$$\mathrm{Lip}_\phi(f) = \sup\left\{\frac{|f(x) - f(y)|}{d_\phi(x, y)} : x \neq y\right\} = \sup\left\{\frac{\mathrm{var}_n(f)}{W_n(\phi)} : n \in \mathbb{N}\right\}.$$

One can then define a norm on $\mathcal{L}_\phi$, making it a Banach space, by setting

$$\|f\|_{\mathcal{L}_\phi} = \|f\|_\infty + \mathrm{Lip}_\phi(f).$$

Remark 2.1. *The usual Banach space of Lipschitz functions is defined as follows. Let*

$$\mathcal{L}_\theta = \{f \in \mathscr{C}(\Omega) : \exists M > 0 \text{ such that } \mathrm{var}_n(f) \leq M\theta^n, n = 1, 2, \ldots\}$$

and

$$\mathrm{Lip}_\theta(f) = \sup\left\{\frac{|f(x) - f(y)|}{d_\theta(x, y)} : x \neq y\right\} = \sup\left\{\frac{\mathrm{var}_n(f)}{\theta^n} : n \in \mathbb{N}\right\}.$$

The canonical norm making $\mathcal{L}_\theta$ a Banach space is $\|f\|_{\mathcal{L}_\theta} = \|f\|_\infty + \mathrm{Lip}_\theta(f)$.

In view of (2.3), if we have $W_n(\phi) = O(\theta^n)$, then $\mathcal{L}_\theta = \mathcal{L}_\phi$. If we now have, for instance, $W_n(\phi) = O(n^{-q})$ for some $q > 0$, then we get a bigger space which contains in particular all functions f such that $\mathrm{var}_n(f) = O(n^{-r})$ with $r \geq q$.

The following result is instrumental to this article. In brief, it tells us that a potential ϕ satisfying Walters' condition has a unique equilibrium state, which will be denoted by μ_ϕ, and gives a speed of convergence for the properly normalized iterates of the associated Ruelle's Perron-Frobenius operator. The first part of the theorem is due to Walters, while the second one is due to Maume-Deschamps and can be found in her PhD thesis [18, Chapter I.2]. Unfortunately, her result was not published even though it is much sharper than the result in [16].

Theorem 2.1 ([22], [18])**.** *Let $\phi : \Omega \to \mathbb{R}$ satisfying Walters' condition as above. Then the following holds.*

A. There exists a unique triplet $(h_\phi, \lambda_\phi, \nu_\phi)$ such that $h_\phi \in \mathcal{L}_\phi$ and is strictly positive, $\|\log h_\phi\|_\infty < \infty$, $\lambda_\phi > 0$, ν_ϕ a fully supported probability measure such that $\int h_\phi \, \mathrm{d}\nu_\phi = 1$. Moreover, $P_\phi h_\phi = \lambda_\phi h_\phi$ and $P_\phi^* \nu_\phi = \lambda_\phi \nu_\phi$, and ϕ has a unique equilibrium state $\mu_\phi = h_\phi \nu_\phi$ which is mixing. [3]

[3] This means that for any pair of cylinders B, B' $\lim_{n\to\infty} \mu_\phi(B \cap T^{-n}B') = \mu_\phi(B)\mu_\phi(B')$. In particular μ_ϕ is ergodic.

B. There exists a positive sequence $(\epsilon_n)_{n\in\mathbb{N}}$ converging to zero, such that, for any $f \in \mathcal{L}_\phi$,

$$\left\| \frac{P_\phi^n f}{\lambda_\phi^n} - h_\phi \int f \mathrm{d}\nu_\phi \right\|_\infty \leq C_{(2.1)}\, \epsilon_n \|f\|_{\mathcal{L}_\phi},\ \forall n \in \mathbb{N}\,. \tag{2.4}$$

Morover, one has the following behaviors:

1. *If $W_n(\phi) = O(\eta^n)$ for some $\eta \in (0,1)$, then there exists $\eta' \in (0,1)$ such that $\epsilon_n = O(\eta'^n)$.*
2. *If $W_n(\phi) = O(n^{-\alpha})$ for some $\alpha > 0$, then $\epsilon_n = O(n^{-\alpha})$.*
3. *If $W_n(\phi) = O(\theta^{(\log n)^\alpha})$ for some $\theta \in (0,1)$ and $\alpha > 1$, then, for any $\epsilon > 0$, $\epsilon_n = O(\theta^{(\log n)^{\alpha-\epsilon}})$.*
4. *If $W_n(\phi) = O(\mathrm{e}^{-cn^\alpha})$ for some $c > 0$ and $\alpha \in (0,1)$, then there exists $c' > 0$ such that $\epsilon_n = O\big(\mathrm{e}^{-c' n^{\frac{\alpha}{\alpha+1}}}\big)$.*

The fact that μ_ϕ is an equilibrium state means that it maximizes the functional $\mu \mapsto h(\nu) + \int \phi \,\mathrm{d}\nu$ over the set of shift-invariant probability measures on Ω, where $h(\nu)$ is the entropy of ν, and the maximum is equal to the topological pressure $P(\phi)$ of ϕ (see *e.g.* [15]), and we have $P(\phi) = \log \lambda_\phi$.

Let us give examples of potentials. First consider $A = \{-1, 1\}$ and $p > 1$, and define

$$\phi(x) = -\sum_{n\geq 2} \frac{x^0 x^{n-1}}{n^p}\,.$$

One can check that $W_n(\phi) = O(n^{-p+2})$. This is the analog of the so-called long-range Ising model on $\mathbb{N}$. Let us now take $A = \{0, 1\}$ and let $[0^k 1] = \{x \in \Omega : x^i = 0, 0 \leq i \leq k-1, \text{and}\, x_k = 1\}$. Let (v_n) be a monotone decreasing sequence of real numbers converging to 0 and define

$$\phi(x) = \begin{cases} v_k & \text{if } x \in [0^k 1] \\ 0 & \text{if } x = (0, 0, \ldots) \\ 0 & \text{otherwise}\,. \end{cases}$$

One can check that $\mathrm{var}_n(\phi) = v_n$. This example is taken from [19].

Remark 2.2. *Let us briefly explain how we can interpret an equilibrium state for a non Lipschitz potential as an absolutely continuous invariant measure of a piecewise expanding map of the unit interval with a Markov partition. It is well-known that a uniformly expanding map S of the unit interval with a finite Markov partition which is piecewise $C^{1+\eta}$, for some $\eta > 0$, can be coded by a subshift of finite type (Ω, T) over a finite alphabet. Then, $-\log|S'|$ induces a potential ϕ on Ω which is Lipschitz (with respect to d_θ). The pullback of μ_ϕ is then the unique absolutely continuous invariant probability measure for*

S. In [10], the authors showed that, given ϕ which is not Lipschitz, one can construct a uniformly expanding map of the unit interval with a finite Markov partition which is piecewise C^1, but not piecewise $C^{1+\eta}$ for any $\eta > 0$, and such that the pullback of μ_ϕ is the Lebesgue measure.

3 Main result and applications

3.1 Gaussian concentration bound

We can now state our main theorem whose proof is deferred to Section 4. We start by the definition of separately d_θ-Lipschitz functions.

Definition 3.1. *A function $K : \Omega^n \to \mathbb{R}$ is said to be separately d_θ-Lipschitz if, for all i, there exists a constant $\mathrm{Lip}_{\theta,i}(K)$ with*

$$|K(x_0, \ldots, x_i, \ldots, x_{n-1}) - K(x_0, \ldots, x_i', \ldots, x_{n-1})| \leq \mathrm{Lip}_{\theta,i}(K)\, d_\theta(x_i, x_i').$$

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

Theorem 3.1. *Suppose that ϕ satisfies one of the following conditions:*

1. *$W_n(\phi) = O(\theta^n)$ (that is, ϕ is d_θ-Lipschitz);*
2. *$W_n(\phi) = O(n^{-\alpha})$ for some $\alpha > 1$;*
3. *$W_n(\phi) = O(\theta^{(\log n)^\alpha})$ for some $\theta \in (0,1)$ and $\alpha > 1$;*
4. *$W_n(\phi) = O(\mathrm{e}^{-cn^\alpha})$ for some $c > 0$ and $\alpha \in (0,1)$.*

Then the process $(x, Tx, \ldots)$, with x distributed according to μ_ϕ, satisfies the following Gaussian concentration bound. There exists $C_{(3.1)} > 0$ such that for any $n \in \mathbb{N}$ and for any separately d_θ-Lipschitz function $K : \Omega^n \to \mathbb{R}$, we have

$$\begin{aligned} &\int \exp\left(K\left(x, Tx, \ldots, T^{n-1}x\right)\right) \mathrm{d}\mu_\phi(x) \\ &\leq \exp\left(\int K\left(x, Tx, \ldots, T^{n-1}x\right) \mathrm{d}\mu_\phi(x)\right) \exp\left(C_{(3.1)} \sum_{j=0}^{n-1} \mathrm{Lip}_{\theta,j}(K)^2\right). \end{aligned} \tag{3.1}$$

Three remarks are in order. First, we conjecture that this theorem is valid under the condition $\sum_n \mathrm{var}_n(\phi) < \infty$. Second, it would be useful to have an explicit formula for $C_{(3.1)}$ in (3.1). Unfortunately, this constant is proportional to $C_{(2.1)}$ (see Theorem 2.1) which is cumbersome since it involves the eigendata of P_ϕ . Third, for the sake of simplicity, we considered the full shift $A^{\mathbb{N}}$. In fact, our results remain true if $\Omega \subset A^{\mathbb{N}}$ is a topologically mixing one-sided subshift of finite type. Moreover, one can extend Theorem 3.1 to bilateral subshifts of finite type by a trick used in [7].

We now give some corollaries of our main theorem that we will be used in the section on applications. First, by (2.3) we immediately obtain the following corollary.

Corollary 3.2. *If there exists $\alpha > 2$ such that*

$$\mathrm{var}_n(\phi) = O\left(\frac{1}{n^\alpha}\right)$$

then we have the Gaussian concentration bound (3.1).

Next, we get the following concentration inequalities from (3.1).

Corollary 3.3. *For all $u > 0$, we have*

$$\mu_\phi\left(x \in \Omega : K\left(x, Tx, \ldots, T^{n-1}x\right) - \int K\left(y, Ty, \ldots, T^{n-1}y\right) \mathrm{d}\mu_\phi(y) \geq u\right)$$
$$\leq \exp\left(-\frac{u^2}{4C_{(3.1)} \sum_{i=0}^{n-1} \mathrm{Lip}_{\theta,i}(K)^2}\right) \tag{3.2}$$

and

$$\mu_\phi\left(x \in \Omega : \left|K\left(x, Tx, \ldots, T^{n-1}x\right) - \int K\left(y, Ty, \ldots, T^{n-1}y\right) \mathrm{d}\mu_\phi(y)\right| \geq u\right)$$
$$\leq 2\exp\left(-\frac{u^2}{4C_{(3.1)} \sum_{i=0}^{n-1} \mathrm{Lip}_{\theta,i}(K)^2}\right). \tag{3.3}$$

Proof. Inequality (3.2) follows by a well-known trick referred to as Chernoff's bounding method [2]. Let us give the proof for completeness. Let $u > 0$. For any random variable Y, Markov's inequality tells us that $\mathbb{P}(Y \geq u) \leq \mathrm{e}^{-\xi u}\, \mathbb{E}\left(\mathrm{e}^{\xi Y}\right)$ for all $\xi > 0$. Now let

$$Y = K\left(x, Tx, \ldots, T^{n-1}x\right) - \int K\left(y, Ty, \ldots, T^{n-1}y\right) \mathrm{d}\mu_\phi(y)\,.$$

Using (3.1) and optimizing over ξ, we get (3.2). Inequality (3.3) follows by applying (3.2) to $-K$ and then summing up the two bounds. □

The last corollary we want to state is about the variance of any separately d_θ-Lipschitz function.

Corollary 3.4. *We have*

$$\int \left(K\left(x, Tx, \ldots, T^{n-1}x\right) - \int K\left(y, Ty, \ldots, T^{n-1}y\right) \mathrm{d}\mu_\phi(y)\right)^2 \mathrm{d}\mu_\phi(x)$$
$$\leq 2C_{(3.1)} \sum_{i=0}^{n-1} \mathrm{Lip}_{\theta,i}(K)^2\,. \tag{3.4}$$

Proof. To alleviate notations, we simply write K instead of $K\left(x, Tx, \dots, T^{n-1}x\right)$, $\int K$ instead of $\int K\left(y, Ty, \dots, T^{n-1}y\right) \mathrm{d}\mu_\phi(y)$, and so on and so forth. Applying (3.1) to ξK where ξ is any real number different from 0, we get

$$\int \exp\left(\xi\Big(K - \int K\Big)\right) \le \exp\left(C_{(3.1)}\xi^2 \sum_{j=0}^{n-1} \mathrm{Lip}_{\theta,j}(K)^2\right).$$

Now by Taylor expansion we get

$$1 + \frac{\xi^2}{2}\int \Big(K - \int K\Big)^2 + o(\xi^2) \le 1 + C_{(3.1)}\xi^2 \sum_{j=0}^{n-1} \mathrm{Lip}_{\theta,j}(K)^2 + o(\xi^2).$$

Dividing by ξ^2 on both sides and then taking the limit $\xi \to 0$, we obtain the desired inequality. □

Although we were not able to prove the Gaussian concentration bound for separately d_ϕ-Lipschitz functions, for many applications separately d_θ-Lipschitz functions are more natural. Furthermore there is a notable class of separately d_ϕ-Lipschitz functions, namely Birkhoff sums of the potential itself, for which our theorem holds. Indeed, when $\phi \in \mathcal{L}_\phi$, the function $K(x, \dots, T^{n-1}x) = S_n\phi(x)$ is obviously separately d_ϕ-Lipschitz and $\mathrm{Lip}_{\phi,j}(K) = \mathrm{Lip}_\phi(\phi)$ for all j. We have the following result.

Theorem 3.5. *Under the hypotheses of Theorem 3.1, there exists $C_{(3.5)} > 0$ such that, for any $\psi \in \mathcal{L}_\phi$, for all $u > 0$, and for all $n \in \mathbb{N}$, we have*

$$\mu_\phi\left(x \in \Omega : \frac{1}{n}S_n\psi(x) - \int \psi \,\mathrm{d}\mu_\phi \ge u\right) \le \exp\left(-\frac{nu^2}{4C_{(3.5)}\mathrm{Lip}_\phi(\psi)^2}\right). \tag{3.5}$$

The proof is left to the reader. The main (simple) modification lies in the proof of Lemma 4.3 in which considering a Birkhoff sum of a d_ϕ-Lipschitz function works fine, whereas we are stuck for a general separately d_ϕ-Lipschitz function.

We will apply this result with $\psi = -\phi$ to derive concentration bounds for hitting times. Note that under the assumptions of this theorem, $\{\psi(T^n x)\}_{n\ge 0}$ satisfies the central limit theorem [18, Chapter 2].

3.2 Related works

The novelty here is to prove a Gaussian concentration bound for potentials with a variation decaying subexponentially. For ϕ is Lipschitz, Theorem 3.1 was proved in [7]. The main goal of [7] was then to deal with nonuniformly

hyperbolic systems modeled by a Young tower. For a tower with a return-time to the base with exponential tails, the authors of [7] proved a Gaussian concentration bound. For polynomial tails, they proved moment concentration bounds. For $C^{1+\eta}$ maps of the unit interval with an indifferent fixed point, which are thus nonuniformly expanding, we are in the latter situation. In view of Remark 2.2 above, we deal here with maps whose derivative is not Hölder continuous, but which are still uniformly expanding.

Let us also mention the paper [14] in which the authors prove a Gaussian concentration bound for ϕ of summable variation (whereas we need a bit more than summable). Their proof is based on coupling. However, they consider functions K on A^n, not on $\left(A^{\mathbb{N}}\right)^n = \Omega^n$ as in this paper. For such functions, the analogue of $\mathrm{Lip}_{\theta,i}(K)$ is $\delta_i(K) = \sup\{|K(a^0,\dots,a^i,\dots,a^{n-1}) - K(b^0,\dots,b^i,\dots,b^{n-1})| : a^j = b^j, \forall j \neq i\}$. It is clear that a Gaussian concentration bound for functions $K : \left(A^{\mathbb{N}}\right)^n \to \mathbb{R}$ implies a Gaussian concentration bound for functions $K : A^n \to \mathbb{R}$, but the converse is not true.

3.3 Applications

We now give several applications of the Gaussian concentration bound (3.1) and its corollaries. Throughout this section, μ_ϕ is the equilibrium state for a potential ϕ satisfying one of the conditions *1-4* in Theorem 3.1.

3.3.1 Birkhoff sums

Let $f : \Omega \to \mathbb{R}$ be a d_θ-Lipschitz function and define

$$K(x_0,\dots,x_{n-1}) = f(x_0) + \cdots + f(x_{n-1})$$

whence $K(x, Tx, \dots, T^{n-1}x) = f(x) + f(Tx) + \cdots + f(T^{n-1}x) := S_n f(x)$ is the Birkhoff sum of f. Clearly, $\mathrm{Lip}_{\theta,i}(K) = \mathrm{Lip}_\theta(f)$ for all $i = 0,\dots,n-1$. Applying Corollary 3.3 we immediately get

$$\mu_\phi\left(x : \left|\frac{S_n f(x)}{n} - \int f \mathrm{d}\mu_\phi\right| \geq u\right) \leq 2\exp\left(-c_{(3.6)} n u^2\right) \tag{3.6}$$

for all $n \geq 1$ and $u \in \mathbb{R}_+$, where

$$c_{(3.6)} = \frac{1}{4C_{(3.1)}\mathrm{Lip}_\theta(f)^2}\,.$$

This bound can be compared with the large deviation asymptotics (1.2). We see that it has the right behavior in n. Replacing u by $u/\sqrt{n}$ in (3.6) we get

$$\mu_\phi\left(x : \left|S_n f(x) - n\int f \mathrm{d}\mu_\phi\right| \geq u\sqrt{n}\right) \leq 2\exp\left(-c_{(3.6)} u^2\right)$$

for all n and $u > 0$. This can be compared with the central limit theorem (1.1). We can see that the previous bound is consistent with that theorem. Note that the central limit is about convergence in law, whereas here we obtain a (non-asymptotic) bound from which one cannot deduce a convergence in law.

3.3.2 Empirical frequency of blocks

Take $f(x) = \mathbb{1}_{[a^{0,k-1}]}(x)$ where

$$[a^{0,k-1}] = \{x \in \Omega : x^i = a^i, i = 0, \dots, k-1\}$$

is a given k-cylinder. Let

$$\mathfrak{f}_n(x, a^{0,k-1}) = \frac{\sum_{k=0}^{n-k} \mathbb{1}_{[a^{0,k-1}]}(T^k x)}{n-k+1}\,.$$

This is the 'empirical frequency' of the block $a^{0,k-1} \in A^k$ in the orbit of x up to time $n-k$. By Birkhoff's ergodic theorem, we know that, for each $a^{0,k-1}$, $\mathfrak{f}_n(x, a^{0,k-1})$ goes to $\mu_\phi([a^{0,k-1}])$ for μ_ϕ-almost all x. The next theorem quantifies this asymptotic statement. Notice that we can control the fluctuations of $\mathfrak{f}_n(x, a^{0,k-1})$ around $\mu_\phi([a^{0,k-1}])$ uniformly in $a^{0,k-1}$.

Theorem 3.6. *For all $n \in \mathbb{N}$, for all $1 \le k \le n$ and for all $u > 0$ we have*

$$\mu_\phi\left(x : \max_{a^{0,k-1}} \left|\mathfrak{f}_n(x, a^{0,k-1}) - \mu_\phi([a^{0,k-1}])\right| \ge \frac{(u + c\sqrt{k})\,\theta^{-k}}{\sqrt{n-k+1}}\right) \le \mathrm{e}^{-\frac{u^2}{4C_{(3.1)}}}$$

where $c = 2\sqrt{2C_{(3.1)} \log |A|}$. Moreover, if $k = k(n) = \zeta \log n$ for some $\zeta > 0$, then

$$\mu_\phi\left(x : \max_{a^{0,k(n)-1}} \left|\mathfrak{f}_n(x, a^{0,k(n)-1}) - \mu_\phi([a^{0,k(n)-1}])\right| \ge \frac{(u + c'\sqrt{\log n})n^{\zeta|\log\theta|}}{\sqrt{n-k(n)+1}}\right)$$
$$\le \mathrm{e}^{-\frac{u^2}{4C_{(3.1)}}}$$

where $c' = 2\sqrt{2\zeta C_{(3.1)} \log |A|}$.

Proof. Define the function $K : \Omega^{n-k+1} \to \mathbb{R}$ by

$$K(x_0, \dots, x_{n-k}) = \max_{a^{0,k-1}} Z(a^{0,k-1}; x_0, \dots, x_{n-k})$$

where

$$Z(a^{0,k-1}; x_0, \dots, x_{n-k}) = \left|\frac{\sum_{j=0}^{n-k} \mathbb{1}_{[a^{0,k-1}]}(x_j)}{n-k+1} - \mu_\phi([a^{0,k-1}])\right|\,.$$

It is left to the reader to check that $\mathrm{Lip}_{\theta,j}(K) = \frac{\mathrm{Lip}_\theta(f)}{n-k+1} = \frac{1}{\theta^k(n-k+1)}$, so we get immediately from 3.6

$$\mu_\phi\left(x \in \Omega : K(x, Tx, \dots, T^{n-k}x) \ge u + \int K\left(y, Ty, \dots, T^{n-k}y\right) \mathrm{d}\mu_\phi(y)\right)$$
$$\le \exp\left(-\frac{\theta^{2k}}{4C_{(3.1)}}\,(n-k+1)u^2\right)$$

for all $n \geq 1$ and $u > 0$. To complete the proof, we need a good upper bound for $\int K\left(y, Ty, \ldots, T^{n-k-1}y\right) \mathrm{d}\mu_\phi(y)$. Actually, this can be done by using again the Gaussian concentration bound. Using (3.1) and Jensen's inequality we get for any $\xi > 0$

$$\begin{aligned}
&\exp\left(\xi \int K\left(x, Tx, \ldots, T^{n-k}x\right) \mathrm{d}\mu_\phi(x)\right) \\
&\leq \int \exp\left(\xi \max_{a^{0,k-1}} Z(a^{0,k-1}; x, Tx, \ldots, T^{n-k}x)\right) \mathrm{d}\mu_\phi(x) \\
&\leq \sum_{a^{0,k-1} \in A^k} \int \exp\left(\xi Z(a^{0,k-1}; x, Tx, \ldots, T^{n-k}x)\right) \mathrm{d}\mu_\phi(x) \\
&\leq 2|A|^k \exp\left(\frac{C_{(3.1)}\theta^{-2k}\xi^2}{n-k+1}\right) .
\end{aligned}$$

The third inequality is obtained by using the trivial inequality

$$\mathrm{e}^{\max_{i=1}^p a_i} \leq \sum_{i=1}^p \mathrm{e}^{a_i} .$$

Taking logarithms on both sides and then dividing by ξ, we have

$$\int K\left(x, Tx, \ldots, T^{n-k}x\right) \mathrm{d}\mu_\phi(x) \leq \frac{\log 2 + k \log|A|}{\xi} + \frac{C_{(3.1)}\theta^{-2k}\xi}{n-k+1} .$$

There is a unique $\xi > 0$ minimizing the right-hand side, hence

$$\int K\left(x, Tx, \ldots, T^{n-k}x\right) \mathrm{d}\mu_\phi(x) \leq 2\theta^{-k}\sqrt{\frac{C_{(3.1)}(k+1)\log|A|}{n-k+1}}$$

where we used that $\log 2 \leq \log|A|$. Hence we get the desired estimate. □

Note that $\log|A|$ is the topological entropy of the full shift with alphabet A.

3.3.3 Hitting times and entropy

For $x, y \in \Omega$, let

$$T_{x^{0,n-1}}(y) = \inf\{j \geq 1 : y^{j,j+n-1} = x^{0,n-1}\} .$$

This is the first time that the n first symbols of x appear in y. We assume that ϕ satisfies

$$\mathrm{var}_n(\phi) = O\left(\frac{1}{n^\alpha}\right) \quad \text{for some } \alpha > 2 . \tag{3.7}$$

One can prove (see [9]) that

$$\lim_{n\to\infty} \frac{1}{n} \log T_{x^{0,n-1}}(y) = h(\mu_\phi)\,, \quad \text{for } \mu_\phi \otimes \mu_\phi\text{-almost every } (x,y)\,.$$

Roughly, this means that, if we pick x and y independently, each one according to μ_ϕ, then the time it takes to see the first n symbols of x appearing in y for the first time is $\approx \mathrm{e}^{nh(\mu_\phi)}$.

Theorem 3.7. *If ϕ satisfies* (3.7)*, then there exist strictly positive constants c_1, c_2 and u_0 such that, for all n and for all $u > u_0$,*

$$(\mu_\phi \otimes \mu_\phi)\Big\{(x,y) : \frac{1}{n} \log T_{x^{0,n-1}}(y) \geq h(\mu_\phi) + u\Big\} \leq c_1\, \mathrm{e}^{-c_2 n u^2}$$

and

$$(\mu_\phi \otimes \mu_\phi)\Big\{(x,y) : \frac{1}{n} \log T_{x^{0,n-1}}(y) \leq h(\mu_\phi) - u\Big\} \leq c_1\, \mathrm{e}^{-c_2 n u}\ .$$

These bounds were obtained in [8] when ϕ is Lipschitz. Observe that the probability of being above $h(\mu_\phi)$ is bounded above by $c_1\, \mathrm{e}^{-c_2 n u^2}$, whereas the probability of being below $h(\mu_\phi)$ is bounded above by $c_1\, \mathrm{e}^{-c_2 n u}$. The proof of this theorem being very similar to that given in [8], we omit the details and only sketch it. We cannot directly deal with $T_{x^{0,n-1}}(y)$ but we have $\log T_{x^{0,n-1}}(y) = \log\big(T_{x^{0,n-1}}(y)\mu_\phi([x^{0,n-1}])\big) - \log \mu_\phi([x^{0,n-1}])$. Then we use Theorem 3.5 for $\psi = -\phi$, assuming (without loss of generality) that $P(\phi) = 0$, that is, $h(\mu_\phi) = -\int \phi \,\mathrm{d}\mu_\phi$, because we can control uniformly in x the approximation $-\log \mu_\phi([x^{0,n-1}]) \approx S_n(-\phi)(x)$. To control the other term, we use that the law of $T_{x^{0,n-1}}(y)\mu_\phi([x^{0,n-1}])$ is well approximated by an exponential law.

Another estimator of $h(\mu_\phi)$ is the so-called plug-in estimator. We could also obtain concentration bounds for it in the spirit of [8].

3.3.4 Speed of convergence of the empirical measure

Instead of looking at the frequency of a block a_1^k we can consider a global object, namely the empirical measure

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

For μ_ϕ-almost every x, we know that

$$\frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x} \xrightarrow{n\to\infty} \mu_\phi$$

where the convergence is in the weak topology on the space of probability measures $\mathscr{M}(\Omega)$ on Ω. This is a consequence of Birkhoff's ergodic theorem. The natural question is: how fast does this convergence takes place? We can answer this question by using the Kantorovich distance d_K which metrizes weak topology on $\mathscr{M}(\Omega)$:

$$d_K(\nu_1,\nu_2) = \sup\left\{\int f\mathrm{d}\nu_1 - \int f\mathrm{d}\nu_2 : f:\Omega\to\mathbb{R} \text{ such that } \mathrm{Lip}_\theta(f)=1\right\}.$$

We have the following result.

Theorem 3.8. *For all $u>0$ and all $n\geq 1$ we have*

$$\mu_\phi\left(x : \left|d_K(\mathscr{E}_n(x),\mu_\phi) - \int d_K(\mathscr{E}_n(y),\mu_\phi)\,\mathrm{d}\mu_\phi(y)\right| \geq u\right) \leq 2\ \mathrm{e}^{-c_{(3.8)}nu^2} \quad (3.8)$$

where $c_{(3.8)} = (4C_{(3.1)})^{-1}$.

Proof. Let

$$K(x_0,\dots,x_{n-1}) = \sup\left\{\frac{1}{n}\sum_{j=0}^{n-1} f(x_j) - \int f\mathrm{d}\mu_\phi : f:\Omega\to\mathbb{R} \text{ with } \mathrm{Lip}_\theta(f)=1\right\}.$$

Of course, $K(x,Tx,\dots,T^{n-1}x) = d_K(\mathscr{E}_n(x),\mu_\phi)$. It is left to the reader to check that

$$\mathrm{Lip}_{\theta,i}(K) \leq \frac{1}{n}\,,\ i=0,\dots,n-1\,.$$

The result follows at once by applying inequality (3.3). □

It is natural to ask for a good upper bound for $\int d_K(\mathscr{E}_n(y),\mu_\phi)\mathrm{d}\mu_\phi(y)$ because this would give a control on the fluctuations of $d_K(\mathscr{E}_n(x),\mu_\phi)$ around 0. Getting such a bound turns out to be difficult. In [5, Section 8] it is proved that

$$\int d_K(\mathscr{E}_n(y),\mu_\phi)\,\mathrm{d}\mu_\phi(y) \preceq \frac{1}{n^{\frac{1}{2(1+\log|A|)}}}\,.$$

For two positive sequences $(a_n),(b_n)$, $a_n\preceq b_n$ means that $\limsup_n \frac{\log a_n}{\log b_n}\leq 1$. One could in principle get a non-asymptotic but messy bound.

3.3.5 Relative entropy, $\bar{d}$-distance and speed of Markov approximation

Given $n\in\mathbb{N}$ and $x^{0,n-1},y^{0,n-1}\in A^n$ the (non normalized) Hamming distance between x and y is

$$\bar{d}_n(x^{0,n-1},y^{0,n-1}) = \sum_{i=0}^{n-1}\mathbb{1}_{\{x^i\neq y^i\}}\,. \quad (3.9)$$

Now, given two shift-invariant probability measures μ, ν on Ω, denote by μ_n and ν_n their projections on A^n, and define their $\bar{d}_n$-distance by

$$\bar{d}_n(\mu_n, \nu_n) = \inf \sum_{x^{0,n-1} \in A^n} \sum_{y^{0,n-1} \in A^n} \bar{d}_n(x, y)\, \mathbb{P}_n(x^{0,n-1}, y^{0,n-1})$$

where the infimum is taken over all the joint shift-invariant probability distributions $\mathbb{P}_n$ on $A^n \times A^n$ such that $\sum_{y^{0,n-1} \in A^n} \mathbb{P}_n(x^{0,n-1}, y^{0,n-1}) = \mu_n(x^{0,n-1})$ and $\sum_{x^{0,n-1} \in A^n} \mathbb{P}_n(x^{0,n-1}, y^{0,n-1}) = \nu_n(y^{0,n-1})$. By [20, Theorem I.9.6, p. 92], the limit following exists:

$$\bar{d}(\mu, \nu) := \lim_{n \to \infty} \frac{1}{n}\, \bar{d}_n(\mu_n, \nu_n) \tag{3.10}$$

and defines a distance on the set of shift-invariant probability measures. It induces a finer topology than the weak topology and, in particular, the $\bar{d}$-limit of ergodic measures is ergodic, and the entropy is $\bar{d}$-continuous on the class of ergodic measures.[4]

Next, given $n \in \mathbb{N}$ and a shift-invariant probability measure ν on Ω, define the n-block relative entropy of ν with respect to μ_ϕ by

$$H_n(\nu|\mu_\phi) = \sum_{x^{0,n-1} \in A^n} \nu_n(x^{0,n-1}) \log \frac{\nu_n(x^{0,n-1})}{\mu_{\phi,n}(x^{0,n-1})}\,.$$

One can easily prove that the following limit exists and defines the relative entropy of ν with respect to μ_ϕ:

$$\lim_{n \to \infty} \frac{1}{n}\, H_n(\nu_n|\mu_{\phi,n}) =: h(\nu|\mu_\phi) = P(\phi) - \int \phi \,\mathrm{d}\nu - h(\nu) \tag{3.11}$$

where $P(\phi)$ is the topological pressure of ϕ:

$$P(\phi) = \lim_{n \to \infty} \frac{1}{n} \log \sum_{a^{0,n-1} \in A^n} \mathrm{e}^{\sup\left\{S_n\phi(x) : x \in [a^{0,n-1}]\right\}}\,.$$

This limit exists for any continuous ϕ. (To prove (3.11), we use that there exists a positive sequence $(\varepsilon_n)_n$ going to 0 such that, for any $a^{0,n-1} \in A^n$ and any $x \in [a^{0,n-1}]$, $\mu_\phi([a^{0,n-1}])/\exp(-nP(\phi) + S_n\phi(x))$ is bounded below by $\exp(-n\varepsilon_n)$ and above by $\exp(-n\varepsilon_n)$.) By the variational principle, $h(\nu|\mu_\phi) \geq 0$ with equality if and only if $\nu = \mu_\phi$ (recall that μ_ϕ is the unique equilibrium state of ϕ). We refer to [21] for details. We can now formulate the first theorem of this section.

Theorem 3.9. *For every shift-invariant probability measure ν on Ω and for all $n \in \mathbb{N}$, we have*

$$\bar{d}_n(\nu_n, \mu_{\phi,n}) \leq c_{(3.9)} \sqrt{n H_n(\nu_n|\mu_{\phi,n})} \tag{3.12}$$

[4] These two properties are false in the weak topology.

where $c_{(3.9)} = \sqrt{2C_{(3.1)}}$. *In particular*

$$\bar{d}(\nu, \mu_\phi) \leq c_{(3.9)} \sqrt{h(\nu|\mu_\phi)}\,. \tag{3.13}$$

Proof. For a function $f : A^n \to \mathbb{R}$, define for each $i = 0, \ldots, n-1$

$$\delta_i(f) = \sup\{|f(a^{0,n-1}) - f(b^{0,n-1})| : a^j = b^j,\, \forall j \neq i\}\,.$$

We obviously have that for all $a^{0,n-1}, b^{0,n-1} \in A^n$

$$|f(a^{0,n-1}) - f(b^{0,n-1})| \leq \sum_{j=0}^{n-1} \mathbb{1}_{\{a^j \neq b^j\}} \delta_j(f)\,.$$

A function $f : A^n \to \mathbb{R}$ such that $\delta_j(f) = 1$, $i = 0, \ldots, n-1$ is 1-Lipschitz for the Hamming distance (3.9). We now consider the set of functions

$$\mathcal{H}(n, \phi) = \left\{ f : A^n \to \mathbb{R} : f \text{ 1-Lipschitz for } \bar{d}_n \,,\, \int_{A^n} f \,\mathrm{d}\mu_{\phi,n} = 0 \right\}\,.$$

We can identify a function $f \in \mathcal{H}(n, \phi)$ with a function $\tilde{f} : \Omega^n \to \mathbb{R}$ in a natural way: $\tilde{f}(x_0, \ldots, x_{n-1}) = f(\pi(x_0), \ldots, \pi(x_{n-1}))$ where $\pi : \Omega \to A$ is defined by $\pi(x) = x^0$. We obviously have $\int \tilde{f} \,\mathrm{d}\mu_\phi = 0$ and it is easy to check that $\mathrm{Lip}_j(\tilde{f}) = \delta_j(f) = 1$, $j = 0, \ldots, n-1$. Therefore we can apply the Gaussian concentration bound (3.1) to get

$$\int_{A^n} \mathrm{e}^{\xi f} \,\mathrm{d}\mu_{\phi,n} \leq \mathrm{e}^{C_{(3.1)} n \xi^2}\,, \text{ for all } f \in \mathcal{H}(n, \phi) \text{ and for all } \xi \in \mathbb{R}\,. \tag{3.14}$$

We now apply an abstract result [1, Theorem 3.1] which says that (3.14) is equivalent to

$$\bar{d}(\nu_n, \mu_{\phi,n}) \leq \sqrt{2C_{(3.1)} n H_n(\nu_n|\mu_{\phi,n})} \quad \text{for all probability measures } \nu_n \text{ on } A^n\,.$$

Hence (3.12) is proved. To get (3.13), divide by n on both sides and take the limit $n \to \infty$ and use (3.10) and (3.11). □

We now give an application of inequality (3.13). Let

$$\phi_1(x) = \log \mu_\phi(x^0) \quad \text{and} \quad \phi_n(x) = \log \mu_\phi(x^{n-1}|x^{0,n-2}), n \geq 2.$$

The equilibrium state for ϕ_n is a $(n-1)$-step Markov measure. One can prove that in the weak topology $(\mu_{\phi_n})_n$ converges to μ_ϕ, but one cannot get any speed of convergence. We get the following upper bound on the speed of convergence of $(\mu_{\phi_n})_n$ to μ_ϕ in the finer $\bar{d}$ topology.

Corollary 3.10. *Assume, without loss of generality, that ϕ is normalized in the sense that*

$$\sum_{a\in A} \mathrm{e}^{\phi(ax)} = 1,\ \forall x \in \Omega.$$

Then there exists $n_\phi \geq 1$ such that, for all $n \geq n_\phi$, we have

$$\bar{d}(\mu_{\phi_n}, \mu_\phi) \leq \rho_\phi \operatorname{var}_n(\phi) \tag{3.15}$$

where

$$\rho_\phi = \sqrt{2|A|\, C_{(3.1)}}\, (\mathrm{e}-1)\, \mathrm{e}^{\frac{3}{2}\|\phi\|_\infty}\ .$$

More details on how to normalize a potential are given in Subsection 4.1.

Proof. Using (3.11) and the variational principle we get

$$h(\mu_{\phi_n}|\mu_\phi) = -\int \phi\, \mathrm{d}\mu_{\phi_n} - h(\mu_{\phi_n}) = \int (\phi_n - \phi)\, \mathrm{d}\mu_{\phi_n}. \tag{3.16}$$

Indeed, since ϕ and ϕ_n are normalized, we have in particular that $P(\phi) = P(\phi_n) = 0$, and by the variational principle $h(\mu_{\phi_n}) = -\int \phi_n \mathrm{d}\mu_{\phi_n}$. Now

$$\begin{aligned}\int (\phi_n - \phi)\, \mathrm{d}\mu_{\phi_n} &= \int \log\left(\frac{\mathrm{e}^{\phi_n}}{\mathrm{e}^{\phi}}\right)\, \mathrm{d}\mu_{\phi_n} = \int \log\left(1 + \frac{\mathrm{e}^{\phi_n} - \mathrm{e}^{\phi}}{\mathrm{e}^{\phi}}\right) \mathrm{d}\mu_{\phi_n} \\ &\leq \int \frac{\mathrm{e}^{\phi_n} - \mathrm{e}^{\phi}}{\mathrm{e}^{\phi}}\, \mathrm{d}\mu_{\phi_n}\end{aligned} \tag{3.17}$$

where we used the inequality $\log(1+u) \leq u$ for all $u > -1$. Now using the shift-invariance of μ_{ϕ_n} and replacing e^{ϕ_n} by $\mathrm{e}^{\phi_n} - \mathrm{e}^{\phi} + \mathrm{e}^{\phi}$ we get

$$\begin{aligned}&\int \frac{\mathrm{e}^{\phi_n} - \mathrm{e}^{\phi}}{\mathrm{e}^{\phi}}\, \mathrm{d}\mu_{\phi_n} = \int \mathrm{d}\mu_{\phi_n}(x) \sum_{a\in A} \mathrm{e}^{\phi_n(ax)}\ \frac{\mathrm{e}^{\phi_n(ax)} - \mathrm{e}^{\phi(ax)}}{\mathrm{e}^{\phi(ax)}} \\ &= \int \mathrm{d}\mu_{\phi_n}(x) \sum_{a\in A} \frac{\left(\mathrm{e}^{\phi_n(ax)} - \mathrm{e}^{\phi(ax)}\right)^2}{\mathrm{e}^{\phi(ax)}} + \int \mathrm{d}\mu_{\phi_n}(x) \sum_{a\in A} \left(\mathrm{e}^{\phi_n(ax)} - \mathrm{e}^{\phi(ax)}\right) \\ &\leq |A|\ \mathrm{e}^{\|\phi\|_\infty} \left(\|\, \mathrm{e}^{\phi_n} - \mathrm{e}^{\phi}\, \|_\infty\right)^2\end{aligned} \tag{3.18}$$

where we used that $\sum_{a\in A}(\mathrm{e}^{\phi_n(ax)} - \mathrm{e}^{\phi(ax)}) = 0$. Combining (3.13), (3.16), (3.17) and (3.18) we thus obtain

$$\bar{d}(\mu_{\phi_n}, \mu_\phi) \leq \sqrt{2|A|\, C_{(3.1)}\, \mathrm{e}^{\|\phi\|_\infty}}\, \|\, \mathrm{e}^{\phi_n} - \mathrm{e}^{\phi}\, \|_\infty.$$

It remains to estimate $\|\, \mathrm{e}^{\phi_n} - \mathrm{e}^{\phi}\, \|_\infty$ in terms of $\operatorname{var}_n(\phi)$. We have

$$\|\, \mathrm{e}^{\phi_n} - \mathrm{e}^{\phi}\, \|_\infty = \left\|\, \mathrm{e}^{\phi_n} - \mathrm{e}^{\phi}\, \right\|_\infty \leq \mathrm{e}^{\|\phi\|_\infty} \left\|\, \mathrm{e}^{\phi-\phi_n} - 1\right\|_\infty \leq (\mathrm{e}-1)\, \mathrm{e}^{\|\phi\|_\infty}\, \|\phi - \phi_n\|_\infty$$

provided that $\|\phi-\phi_n\|_\infty < 1$, where we used the inequality $|\,\mathrm{e}^u - 1| \le (\mathrm{e}-1)|u|$ valid for $|u| < 1$. Finally, since $\|\phi - \phi_n\|_\infty \le \mathrm{var}_n(\phi)$, we define n_ϕ to be the smallest integer sucht $\mathrm{var}_n(\phi) < 1$ and we can take

$$\rho_\phi = \sqrt{2|A|\, C_{(3.1)}}\,(\mathrm{e}-1)\,\mathrm{e}^{\frac{3}{2}\|\phi\|_\infty}\,.$$

We thus proved (3.15). □

Let us mention the paper [13] in which the authors obtain the same bound for the speed of convergence of Markov approximation, up to the constant. Their approach is a direct estimation of $\bar{d}(\mu_{\phi_n}, \mu_\phi)$ by using a coupling method. The point here is to obtain the same speed of convergence as an easy corollary of inequality (3.13). Let us remark that from (4.8) we get a worse result since we end up with a bound proportional to $\sqrt{\mathrm{var}_n(\phi)}$. The trick which leads to the correct bound was told us by Daniel Takahashi.

3.3.6 Shadowing of orbits

Let A be a Borel subset of Ω such that $\mu_\phi(A) > 0$ and define for all $n \in \mathbb{N}$

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

A basic example of set A is a cylinder set $[a^{0,k-1}]$. The quantity $\mathcal{S}_A(x,n)$, which lies between 0 and 1, measures how we can trace, in the best possible way, the orbit of some initial condition not in A by an orbit starting in A.

Theorem 3.11. *For any Borel subset $A \subset \Omega$ such that $\mu_\phi(A) > 0$, for any $n \in \mathbb{N}$ and for any $u > 0$*

$$\mu_\phi\left\{x \in \Omega : \mathcal{S}_A(x,n) \ge \frac{u_A + u}{\sqrt{n}}\right\} \le \mathrm{e}^{-\frac{u^2}{4C_{(3.1)}}}$$

where

$$u_A = 2\sqrt{-\,C_{(3.1)} \ln \mu_\phi(A)}\,.$$

We give a shorter and simpler proof than in [11].

Proof. Let $K(x_0,\dots,x_{n-1}) = \frac{1}{n}\inf_{y\in A}\sum_{j=0}^{n-1} d_\theta(x_j, T^j y)$. One can easily check that

$$\mathrm{Lip}_{\theta,i}(K) = \frac{1}{n},\ \forall i = 0,\dots,n-1.$$

It follows from (3.2) that

$$\mu_\phi\left\{\mathcal{S}_A(x,n) \ge \int \mathcal{S}_A(y,n)\,\mathrm{d}\mu_\phi(y) + \frac{u}{\sqrt{n}}\right\} \le \mathrm{e}^{-\frac{u^2}{4C_{(3.1)}}}$$

for all $n \geq 1$ and for all $u > 0$. We now need an upper bound for $\int \mathcal{S}_A(y,n)\mathrm{d}\mu_\phi(y)$. We simply observe that by (3.1) and the definition of $\mathcal{S}_A(\cdot,n)$

$$\begin{aligned}\mu_\phi(A) &= \int \mathrm{e}^{-\xi\mathcal{S}_A(x,n)}\,\mathbb{1}_A(x)\,\mathrm{d}\mu_\phi(x) \leq \int \mathrm{e}^{-\xi\mathcal{S}_A(x,n)}\,\mathrm{d}\mu_\phi(x)\\ &\leq \mathrm{e}^{-\xi\int \mathcal{S}_A(y,n)\,\mathrm{d}\mu_\phi(y)}\,\mathrm{e}^{\frac{C_{(3.1)}\xi^2}{n}}\end{aligned}$$

for all $\xi > 0$. Hence

$$\int \mathcal{S}_A(y,n)\,\mathrm{d}\mu_\phi(y) \leq \frac{C_{(3.1)}\xi}{n} + \frac{1}{\xi}\ln(\mu_\phi(A)^{-1})$$

Optimizing this bound over $\xi > 0$ gives

$$\int \mathcal{S}_A(y,n)\,\mathrm{d}\mu_\phi(y) \leq 2\sqrt{\frac{C_{(3.1)}\ln(\mu_\phi(A)^{-1})}{n}}.$$

The theorem follows at once. □

3.3.7 Almost-sure central limit theorem

It was proved in [18, Chapter 2] that (Ω, T, μ_ϕ) satisfies the central limit theorem for the class of d_θ-Lipschitz functions $f : \Omega \to \mathbb{R}$ such $\int f\mathrm{d}\mu_\phi = 0$, that is, for any such f the process $\{f \circ T^n\}_{n\geq 0}$ satisfies

$$\mu_\phi\Big(x : \frac{S_k f(x)}{\sqrt{k}} \leq u\Big) = \int \mathbb{1}_{\left\{\frac{S_k f(x)}{\sqrt{k}} \leq t\right\}}\mathrm{d}\mu_\phi(x) \xrightarrow{k\to\infty} G_{0,\sigma^2(f)}((-\infty,t]) \quad (3.19)$$

where

$$\sigma^2(f) = \int f^2\,\mathrm{d}\mu_\phi + 2\sum_{i\geq 1}\int f\cdot f\circ T^i\,\mathrm{d}\mu_\phi \in [0,+\infty)\,.$$

If $\sigma^2(f) > 0$, G_{0,σ^2} denotes the law of a Gaussian random variable with mean 0 and variance $\sigma^2(f)$, that is,

$$\mathrm{d}G_{0,\sigma^2(f)}(u) = \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{u^2}{2\sigma^2(f)}}\,\mathrm{d}u,\ u \in \mathbb{R}\,.$$

When $\sigma^2(f) = 0$ we set $G_{0,0} = \delta_0$, the Dirac mass at zero.

Remark 3.1. *In fact, a more general statement was proved in [18, Chapter I.2]. Namely,* (3.19) *holds when ϕ is such that $\sum_k \epsilon_k < +\infty$ and $f \in \mathcal{L}_\phi$.*

Now, for each $N \geq 1$ and $x \in \Omega$, define the probability measure

$$\mathcal{A}_N(x) = \frac{1}{L_N}\sum_{n=1}^{N}\frac{1}{n}\,\delta_{\frac{S_n f(x)}{\sqrt{n}}} \quad (3.20)$$

where $L_N = \sum_{n=1}^N \frac{1}{n}$ and where, as usual, δ_u is the Dirac mass at point $u \in \mathbb{R}$. Of course, $L_N = \log N + \mathcal{O}(1)$. Notice that $\mathcal{A}_N$ is a random probability measure. Finally, the Wasserstein distance between two probability measures ν, ν' on the Borel sigma-algbra $\mathscr{B}(\mathbb{R})$ is

$$W_1(\nu, \nu') = \inf_{\pi \in \Pi(\nu, \nu')} \int d_\theta(x, x') \, \mathrm{d}\pi(x, x')$$

where the infimum is taken over all probability measures such that

$$\int \pi(B, x') \, \mathrm{d}x' = \nu(B) \quad \text{and} \quad \int \pi(x, B) \, \mathrm{d}x = \nu'(B)$$

for any Borel subset of $\mathbb{R}$. By the Kantorovich-Rubinstein duality theorem, $W_1(\nu, \nu')$ is equal to the Kantorovich distance which is the supremum of $\int \ell \, \mathrm{d}\nu - \int \ell \, \mathrm{d}\nu'$ over the set of 1-Lipschitz functions $\ell : \mathbb{R} \to \mathbb{R}$. We refer to [12] for background and proofs.

Now we can formulate the almost-sure central limit theorem.

Theorem 3.12. *Let $f : \Omega \to \mathbb{R}$ be a d_θ-Lipschitz function. Then, for μ_ϕ almost every $x \in \Omega$, we have*

$$W_1(\mathcal{A}_N(x), G_{0,\sigma^2(f)}) \xrightarrow[N \to +\infty]{} 0 \, .$$

We make several comments. Recall that the Wasserstein distance metrizes the weak topology on the set of probability measures ν on $\mathscr{B}(\mathbb{R})$. Moreover, if $(\nu_n)_{n \geq 1}$ is a sequence of probability measures on $\mathscr{B}(\mathbb{R})$ and ν a probability measure on $\mathscr{B}(\mathbb{R})$, then

$$\lim_{n \to \infty} W_1(\nu_n, \nu) = 0 \quad \Longleftrightarrow \quad \nu_n \xrightarrow{\text{law}} \nu \quad \text{and} \int |u| \, \mathrm{d}\nu_n(u) \xrightarrow{n \to +\infty} \int |u| \, \mathrm{d}\nu(u)$$

where "$\xrightarrow{\text{law}}$" means weak convergence of probability measures on $\mathscr{B}(\mathbb{R})$.
To compare with (3.19), observe that Theorem 3.12 implies that for μ_ϕ-almost every x, $\mathcal{A}_N(x) \xrightarrow{\text{law}} G_{0,\sigma^2(f)}$, which in turn implies that

$$\int \mathbb{1}_{\{u \leq t\}} \mathrm{d}\mathcal{A}_N(u) = \frac{1}{L_N} \sum_{n=1}^N \frac{1}{n} \mathbb{1}_{\{S_n f / \sqrt{n} \leq t\}} \xrightarrow[N \to +\infty]{} G_{0,\sigma^2(f)}((-\infty, t]).$$

Therefore, the expectation with respect to μ_ϕ in (3.19) is replaced by a pathwise logarithmic average in the almost-sure central limit theorem.

Proof. The proof follows from an abstract theorem proved in [6]. In words, that theorem says the following. Let $(X_n)_{n \geq 0}$ be a stochastic stationary process where the X_n's are random variables taking values in Ω. Assume that

if $f : \Omega \to \mathbb{R}$ is d-Lipschitz and such that $\mathbb{E}[f(X_0)] = 0$, then it satisfies the central limit theorem, that is, for all $u \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{\sum_{j=0}^{n-1} f(X_j)}{\sqrt{n}} \leq u\right) \xrightarrow{n\to\infty} G_{0,\sigma^2(f)}((-\infty, u])$$

where $\sigma^2(f) := \mathbb{E}[f^2(X_0)] + 2\sum_{\ell\geq 1} \mathbb{E}[f(X_0)f(X_\ell)]$ is assumed to be $\neq 0$. Moreover, assume that the process $(X_n)_{n\geq 0}$ satisfies the following variance inequality: There exists $C > 0$ such that for all separately d-Lipschitz functions $K : \Omega^n \to \mathbb{R}$ for some distance d on Ω,

$$\mathbb{E}\big[(K(X_0, \ldots, X_{n-1}) - \mathbb{E}[K(X_0, \ldots, X_{n-1})])^2\big] \leq C \sum_{i=0}^{n-1} \mathrm{Lip}_i(K)^2 \,.$$

Then, the conclusion is that, almost surely,

$$\frac{1}{L_N} \sum_{n=1}^{N} \frac{1}{n} \delta_{\frac{X_0+\cdots+X_{n-1}}{\sqrt{n}}}$$

converges in Wasserstein distance (or, equivalently, in Kantorovich distance) to $G_{0,\sigma^2(f)}((-\infty, u])$. We apply this abstract theorem to the process $(x, Tx, \ldots)$ where $x \in \Omega$ is distributed according to μ_ϕ with $\Omega = A^{\mathbb{N}}$ and $d = d_\theta$. Since we have (3.19) and (3.4), the theorem follows. □

Remark 3.2. *The previous result relies only upon the variance inequality* (3.4)*, which is much weaker than the Gaussian concentration bound of Theorem 3.1. On the one hand, the variance inequality* (3.4) *should be true for less regular potentials than the ones we consider here. On the other hand, the Gaussian concentration bound should provide a strengthening of Theorem 3.12, namely a speed of convergence.*

4 Proof of Theorem 3.1

We follow the proof given in [7] with the appropriate modifications to go beyond Lipschitz potentials.

4.1 Some preparatory results

It is convenient to normalize the potential ϕ or, equivalently, the operator P_ϕ in the following way. We use the notations of Theorem 2.1. Let

$$\widetilde{P}_\phi f = \lambda_\phi^{-1} h_\phi^{-1} P_\phi(f h_\phi).$$

Thus

$$\widetilde{P}_\phi 1 = 1 \quad \text{and} \quad \widetilde{P}_\phi^* \mu_\phi = \mu_\phi. \tag{4.1}$$

Let g denote the inverse of the Jacobian of T, and $g^{(k)}$ the inverse of the Jacobian of T^k, that is,

$$g = \frac{h_\phi}{\lambda_\phi h_\phi \circ T} \exp(\phi) \quad \text{and} \quad g^{(k)} = \frac{h_\phi}{\lambda_\phi^k h_\phi \circ T} \exp\Big(\sum_{i=0}^{k-1} \phi \circ T^i\Big)\,.$$

(Of course $g = g^{(1)}$.) Therefore we have

$$\widetilde{P}_\phi f(x) = \sum_{Ty=x} g(y) f(y) \quad \text{and} \quad \widetilde{P}_\phi^k f(x) = \sum_{T^k y = x} g^{(k)}(y) f(y)\,. \tag{4.2}$$

Estimate (2.4) now takes the form

$$\left\| \widetilde{P}_\phi^n f - \int f \mathrm{d}\mu_\phi \right\|_\infty \le C_{(2.1)} \, \|f\|_{\mathcal{L}_\phi} \, \epsilon_n, \ n \ge 1, \tag{4.3}$$

for any $f \in \mathcal{L}_\phi$. Finally, we will need the following distortion estimate. Let $x, y \in \Omega$ such that $x^i = y^i$ for $i = 0, \dots, n-1$ and $x', y' \in \Omega$ such that $T^k x' = x$ and $T^k y' = y$. Then it is easy to check (see [18, Chapter 2]) that, for any k,

$$\left| 1 - \frac{g^{(k)}(x')}{g^{(k)}(y')} \right| \le c_{(4.4)} \, d_\phi(x, y) \tag{4.4}$$

for some constant $c_{(4.4)} > 0$ depending only on ϕ.

We will use the following inequality relating the distances d_θ and d_ϕ.

Lemma 4.1. *Suppose that $W_n(\phi) = O(\theta^n)$, $n \ge 1$, or*

$$\lim_n \frac{W_n(\phi)}{W_{n+1}(\phi)} = 1\,. \tag{4.5}$$

Then there exists $c_{(4.1)} > 0$

$$\sup_n \frac{\theta^n}{W_n(\phi)} \le c_{(4.1)}$$

or, equivalently,

$$d_\theta(x, y) \le c_{(4.1)} d_\phi(x, y)$$

for all x, y.

Proof. The statement is trivial when $W_n(\phi) = O(\theta^n)$. If (4.5) holds, then there exists n_0 such that for all $n \ge n_0$

$$\frac{W_n(\phi)}{W_{n+1}(\phi)} \le \frac{1}{\theta},$$

hence $W_n(\phi) \ge \theta^{n-n_0} W_{n_0}(\phi)$. Then the desired inequalities follow easily from the definitions. □

4.2 Proof of Theorem 3.1

Fix a separately d_θ-Lipschitz function $K : \Omega^n \to \mathbb{R}$. It is convenient to think of it as a function on $\Omega^{\mathbb{N}}$ depending only on the first n coordinates, therefore $\mathrm{Lip}_{\theta,i}(K) = 0$ for $i \geq n$. We endow $\Omega^{\mathbb{N}}$ with the measure μ^∞ obtained as the limit when $k \to \infty$ of the measure μ_k^∞ on Ω^k given by $\mathrm{d}\mu_k^\infty(x_0, \dots, x_{k-1}) = \mathrm{d}\mu_\phi(x_0)\delta_{x_1=Tx_0} \cdots \delta_{x_{k-1}=Tx_{k-2}}$. On $\Omega^{\mathbb{N}}$, let $\mathcal{F}_p$ be the σ-algebra of events depending only on the coordinates $(x_j)_{j\geq p}$ (this is a decreasing sequence of σ-fields). We want to write the function K as a sum of reverse martingale differences with respect to this sequence. Therefore, let $K_p = \mathbb{E}(K|\mathcal{F}_p)$ and $D_p = K_p - K_{p+1}$. More precisely,

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

The function D_p is $\mathcal{F}_p$-measurable and $\mathbb{E}(D_p|\mathcal{F}_{p+1}) = 0$. Moreover

$$K - \mathbb{E}(K) = \sum_{p\geq 0} D_p. \tag{4.6}$$

We then apply Azuma-Hoeffding inequality (see e.g. [17, Page 68]) which says that

$$\mathbb{E}\left(\mathrm{e}^{\sum_{p=0}^{P-1} D_p}\right) \leq \mathrm{e}^{\frac{1}{2}\sum_{p=0}^{P-1} \|D_p\|_\infty^2}. \tag{4.7}$$

Therefore, the point is to obtain a good bound on D_p. This is the claim of the following lemma.

Lemma 4.2. *There exists $C_{(4.2)} > 0$, depending only on ϕ, such that for any $p \in \mathbb{N}$ one has*

$$\|D_p\|_\infty \leq C_{(4.2)} \sum_{i=0}^{p} \epsilon_{p-i} \sum_{j=0}^{i} \mathrm{Lip}_{\theta,j}(K)\, \theta^{i-j} + \mathrm{Lip}_{\theta,p}(K)\,.$$

Using this lemma and applying Young's inequality for convolutions [3, p.

316] twice we obtain

$$\begin{aligned}
&\sum_{p=0}^{P-1} \|D_p\|_\infty^2 \\
&\qquad \le 2C_{(4.2)}^2 \sum_{p=0}^{P-1} \left(\sum_{i=0}^{p} \epsilon_{p-i} \sum_{j=0}^{i} \mathrm{Lip}_{\theta,j}(K)\, \theta^{i-j} \right)^2 + 2 \sum_{p=0}^{P-1} \mathrm{Lip}_{\theta,p}(K)^2 \\
&\qquad \le 2C_{(4.2)}^2 \left(\sum_{k\ge 1} \epsilon_k \right)^2 \sum_{p=0}^{P-1} \left(\sum_{j=0}^{p} \mathrm{Lip}_{\theta,j}(K)\, \theta^{p-j} \right)^2 + 2 \sum_{p=0}^{P-1} \mathrm{Lip}_{\theta,p}(K)^2 \\
&\qquad \le 2 \left(C_{(4.2)}^2 (1-\theta)^{-2} \left(\sum_{k\ge 1} \epsilon_k \right)^2 + 1 \right) \sum_{p=0}^{P} \mathrm{Lip}_{\theta,p}(K)^2 .
\end{aligned}$$

Remark 4.1. *If $u = (u_n)_n$ and $v = (v_n)_n$ are sequences of reals, their convolution $u \star v$ is given by $(u \star v)_n = \sum_{k=0}^{n} u_k v_{n-k}$. Young's inequality tells us that if $u \in \ell^p(\mathbb{N})$, $u \in \ell^q(\mathbb{N})$ and $1 \le p, q, r \le \infty$ with $r^{-1} + 1 = p^{-1} + q^{-1}$, then*

$$\|u \star v\|_r \le \|u\|_p \|v\|_q \,.$$

We used it twice with $r = 2$, $p = 2$ and $q = 1$.

Notice that by assumption and by Theorem 2.1 we have $\sum_{k\ge 1} \epsilon_k < +\infty$. Therefore, using (4.7) at a fixed index P and then letting P tend to infinity, we get by the dominated convergence theorem

$$\mathbb{E}\left(\mathrm{e}^{\sum_{p\ge 0} D_p} \right) \le \mathrm{e}^{\frac{1}{2} \sum_{p\ge 0} \|D_p\|_\infty^2}$$

which is, in view of (4.6), exactly (3.1) with

$$C_{(3.1)} = 1 + C_{(4.2)}^2 (1-\theta)^{-2} \left(\sum_{k\ge 1} \epsilon_k \right)^2 .$$

Now we are going to prove Lemma 4.2 by proving that K_p is close to an integral quantity. This is the content of the following lemma which is the core of the proof.

Lemma 4.3. *There exists $C_{(4.3)} > 0$, depending only on ϕ, such that, for all $p \in \mathbb{N}$,*

$$\begin{aligned}
&\left| K_p(x_p, \dots) - \int K(y, \dots, T^{p-1}y, x_p, \dots)\, \mathrm{d}\mu_\phi(y) \right| \\
&\qquad\qquad \le C_{(4.3)} \sum_{i=0}^{p-1} \epsilon_{p-i} \sum_{j=0}^{i} \mathrm{Lip}_{\theta,j}(K)\, \theta^{i-j}
\end{aligned}$$

where

$$C_{(4.3)} = C_{(2.1)}(c_{(4.4)} + 2c_{(4.1)}).$$

Proof of Lemma 4.2. Applying Lemma 4.3 yields

$$|K_p(x_p, x_{p+1}, \dots) - K_p(x'_p, x_{p+1}, \dots)| \\ \leq 2C_{(4.3)} \sum_{i=0}^{p-1} \epsilon_{p-i} \sum_{j=0}^{i} \mathrm{Lip}_{\theta,j}(K)\, \theta^{i-j} + \mathrm{Lip}_{\theta,p}(K)\,.$$

Averaging $K_p(x'_p, x_{p+1}, \dots)$ over the preimages of x'_p we get exactly $K_{p+1}(x_{p+1}, \dots)$, hence the previous bound holds for $|D_p|$, proving the lemma. □

Proof of Lemma 4.3. Let us fix a point x_* in Ω and decompose K_p as

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

For fixed i, we can group together those points $y \in T^{-p}(x_p)$ which have the same image under T^i, splitting the sum $\sum_{T^p(y)=x_p}$ as $\sum_{T^{p-i}(z)=x_p} \sum_{T^i(y)=z}$. Since the jacobian is multiplicative, one has $g^{(p)}(y) = g^{(i)}(y) g^{(p-i)}(z)$. Let us define two functions f_i and H as follows:

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

Bearing in mind (4.2), we obtain

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

Now we want to prove that $f_i \in \mathcal{L}_\phi$ to use (4.3). First observe that for any $z \in \Omega$

$$|f_i(z)| \leq \sum_{T^i y = z} g^{(i)}(y)\, \mathrm{Lip}_{\theta,i}(K)\, d_\theta(x_*, T^i y) \leq \mathrm{Lip}_{\theta,i}(K)$$

since $d_\theta(x_*, T^i y) \leq 1$ and $\sum_{T^i y = z} g^{(i)}(y) = 1$. Hence

$$\|f_i\|_\infty \leq \mathrm{Lip}_{\theta,i}(K)\,.$$

We now estimate the d_ϕ-Lipschitz norm of f_i. We write

$$\begin{aligned} f_i(z) - f_i(z') = & \sum (g^{(i)}(y) - g^{(i)}(y')) H(y, \dots, T^i y) \\ & + \sum g^{(i)}(y') (H(y, \dots, T^i y) - H(y', \dots, T^i y')) \end{aligned} \tag{4.8}$$

where z and z' are two points in the same partition element, and their respective preimages y, y' are paired according to the cylinder of length i they belong to. Using the distorsion control (4.4) we have

$$|g^{(i)}(y) - g^{(i)}(y')| \leq c_{(4.4)}\, g^{(i)}(y)\, d_\phi(z, z')$$

hence the first sum in (4.8) is bounded in absolute value by

$$c_{(4.4)}\, \mathrm{Lip}_{\theta,i}(K)\, d_\phi(z, z')\,.$$

For the second sum, substituting successively each $T^j y$ with $T^j y'$, we have

$$\begin{aligned} |H(y, \dots, T^i y) - H(y', \dots, T^i y')| &\leq 2 \sum_{j=0}^{i} \mathrm{Lip}_{\theta,j}(K)\, d_\theta(T^j y, T^j y') \\ &\leq 2 \sum_{j=0}^{i} \mathrm{Lip}_{\theta,j}(K)\, \theta^{i-j} d_\theta(z, z') \\ &\leq 2 c_{(4.1)} \sum_{j=0}^{i} \mathrm{Lip}_{\theta,j}(K)\, \theta^{i-j} d_\phi(z, z') \end{aligned}$$

where we used Lemma 4.1 for the third inequality.
Summing over the different preimages of z, we deduce that

$$\|f_i\|_{\mathcal{L}_\phi} \leq (c_{(4.4)} + 2c_{(4.1)}) \sum_{j=0}^{i} \mathrm{Lip}_{\theta,j}(K)\, \theta^{i-j}.$$

Therefore we can apply (4.3) to get

$$\left\| \widetilde{P}_\phi^{p-i} f_i - \int f_i \, \mathrm{d}\mu_\phi \right\|_\infty \leq C_{(2.1)} \big(c_{(4.4)} + 2c_{(4.1)}\big)\, \epsilon_{p-i} \sum_{j=0}^{i} \mathrm{Lip}_{\theta,j}(K)\, \theta^{i-j}\,.$$

Summing those bounds, one obtains

$$\begin{aligned} & \Big| K_p(x_p, \dots) - \sum_{i=0}^{p-1} \int f_i \, \mathrm{d}\mu_\phi - K(x_*, \dots, x_*, x_p, \dots) \Big| \\ & \qquad \leq C_{(2.1)} \big(c_{(4.4)} + 2c_{(4.1)}\big) \sum_{i=0}^{p-1} \epsilon_{p-i} \sum_{j=0}^{i} \mathrm{Lip}_{\theta,j}(K)\, \theta^{i-j}\,. \end{aligned}$$

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

References

[1] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.

[2] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities.* Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

[3] P. Bullen. *Dictionary of inequalities.* Monographs and Research Notes in Mathematics. CRC Press, Boca Raton, FL, second edition, 2015.

[4] J.-R. Chazottes. Fluctuations of observables in dynamical systems: from limit theorems to concentration inequalities. In *Nonlinear dynamics new directions*, volume 11 of *Nonlinear Syst. Complex.*, pages 47–85. Springer, Cham, 2015.

[5] J.-R. Chazottes, P. Collet, and F. Redig. On concentration inequalities and their applications for Gibbs measures in lattice systems. *Journal of Statistical Physics*, 169(3):504–546, Nov 2017.

[6] J.-R. Chazottes, P. Collet, and B. Schmitt. Statistical consequences of the Devroye inequality for processes. Applications to a class of non-uniformly hyperbolic dynamical systems. *Nonlinearity*, 18(5):2341–2364, 2005.

[7] J.-R. Chazottes and S. Gouëzel. Optimal concentration inequalities for dynamical systems. *Comm. Math. Phys.*, 316(3):843–889, 2012.

[8] J.-R. Chazottes and C. Maldonado. Concentration bounds for entropy estimation of one-dimensional Gibbs measures. *Nonlinearity*, 24(8):2371–2381, 2011.

[9] J.-R. Chazottes and E. Ugalde. Entropy estimation and fluctuations of hitting and recurrence times for Gibbsian sources. *Discrete Contin. Dyn. Syst. Ser. B*, 5(3):565–586, 2005.

[10] P. Collet and A. Galves. Chains of infinite order, chains with memory of variable length, and maps of the interval. *J. Stat. Phys.*, 149(1):73–85, 2012.

[11] P. Collet, S. Martínez, and B. Schmitt. Exponential inequalities for dynamical measures of expanding maps of the interval. *Probab. Theory Related Fields*, 123(3):301–322, 2002.

[12] R. Dudley. *Real Analysis and Probability.* CRC Press, 2018.

[13] R. Fernández and A. Galves. Markov approximations of chains of infinite order. *Bull. Braz. Math. Soc., New Series*, 33(3): 295–306, 2002.

[14] S. Gallo and D. Y. Takahashi. Attractive regular stochastic chains: perfect simulation and phase transition. *Ergodic Theory Dynam. Systems*, 34(5):1567–1586, 2014.

[15] G. Keller. *Equilibrium states in ergodic theory*, volume 42 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1998.

[16] A. Kondah, V. Maume, and B. Schmitt. Vitesse de convergence vers l'état d'équilibre pour des dynamiques markoviennes non höldériennes. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(6):675–695, 1997.

[17] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[18] V. Maume-Deschamps. *Propriétés de mélange pour des systèmes dynamiques markoviens*. PhD thesis, Université de Bourgogne, 1998.

[19] M. Pollicott. Rates of mixing for potentials of summable variation. *Transactions of the American Mathematical Society*, 352, 02 2000.

[20] P. C. Shields. *The ergodic theory of discrete sample paths*, volume 13 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.

[21] P. Walters. Ruelle's operator theorem and g-measures. *Trans. Amer. Math. Soc.*, 214:375–387, 1975.

[22] P. Walters. Invariant measures and equilibrium states for some mappings which expand distances. *Trans. Amer. Math. Soc.*, 236:121–153, 1978.