Limit Theorems and Markov Approximations for Chaotic Dynamical Systems

N.I. Chernov

September 14, 2006

Department of Mathematics
University of Alabama at Birmingham
Birmingham, AL 35294, USA.
e-mail: chernov@vorteb.math.uab.edu Fax: (205)-934-9025

Abstract

We prove the central limit theorem and weak invariance principle for abstract dynamical systems based on bounds on their mixing coefficients. We also develop techniques of Markov approximations for dynamical systems. We apply our results to expanding interval maps, Axiom A diffeomorphisms, chaotic billiards and hyperbolic attractors.

Mathematical Subject Classification: 28D05, 58F11, 58F15, 60F05, 60F17

Running title: Limit theorems and Markov approximations

1 Introduction

The subject of this paper is the probabilistic aspects of the theory of chaotic dynamical systems. By a dynamical system we mean a measure preserving transformation $T: M \to M$ of a measurable space M (also called the phase space) with a probability measure μ. Flows are not discussed here. We consider mixing dynamical systems under various assumptions on mixing coefficients.

Our purpose is to establish bounds on correlations and central limit theorems for various classes of functions on the phase space. For any two observables $F, G \in L_2(M)$ the correlation function $C_{F,G}(n)$ is defined to be

$$C_{F,G}(n) = \langle F(T^n x)G(x)\rangle - \langle F(x)\rangle \cdot \langle G(x)\rangle, \quad (1)$$

where $\langle\cdot\rangle$ stands for the integration with respect to the invariant measure μ.

Let $S_N(x) = F(x) + F(Tx) + \cdots + F(T^{N-1}x)$. We say that F obeys the central limit theorem (CLT) if for any $z \in \mathbb{R}$ we have

$$\lim_{N\to\infty} \mu\left\{x : \frac{S_N - \langle S_N\rangle}{(\operatorname{Var} S_N)^{1/2}} \le z\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt, \tag{2}$$

i.e. the "centered" and "normed" partial sum S_N converges in distribution to the standard normal law (certainly, $\operatorname{Var} S_N = \langle S_N^2\rangle - \langle S_N\rangle^2$). In addition, the CLT says that $\operatorname{Var} S_N = \sigma_F^2 N + o(N)$, where

$$\sigma_F^2 = C_{F,F}(0) + 2\sum_{n=1}^{\infty} C_{F,F}(n) \tag{3}$$

A weak invariance principle (WIP) often accompanies the central limit theorem in the theory of random processes and in that of chaotic dynamical systems. For every $x \in M$ consider a function $W_N(t)$ for $t \in [0,1]$, such that

$$W_N(k/N) = \frac{S_k}{\sigma_F\sqrt{N}} \tag{4}$$

for $k = 0, 1, \ldots, N$, and $W_N(t)$ is a linear function between k/N and $(k+1)/N$ for each $0 \le k \le N-1$ (the graph of $W_N(t)$ is a polygonal line). Then $\{W_N\}$ for any N induces a measure on the space of continuous functions on $[0,1]$. The invariance principle then says that this measure converges weakly, as $N \to \infty$, to the Wiener measure. In other words, the partial sum process $\{S_N\}$, after a proper rescaling in space and time, converges in distribution to the Wiener process W. The weak invariance principle is also called the functional central limit theorem (FCLT).

The rate of the decay of correlations plays a crucial role in statistical physics, e.g., in the studies of relaxation to equilibrium and transport coefficients for dynamical variables. The central limit theorem plays a fundamental role in studies of diffusion and in statistical analysis on dynamical systems. The weak invariance principle provides an approximation to deterministic dynamical systems by a Brownian motion on large scales in space and time. In statistical physics such a time-space rescaling often means the transition from the microscopic time scale to the macroscopic one. As a result, a completely deterministic model will behave, at macroscopic times, as a Brownian motion. The most prominent example of that kind is the Lorentz gas, where a particle moves in space and bounces off fixed convex scatterers periodically arranged in space, see [7, 10, 12] for more detail.

All these properties – bounds on correlations, the CLT and WIP – are referred to as statistical properties of dynamical systems. We do not discuss further refinements of the CLT here, see, e.g., M. Denker's survey [14].

At present, strong statistical properties have been established for Anosov and Axiom A diffeomorphisms on compact manifolds with Hölder continuous functions by R. Bowen [5] and D. Ruelle [39]. After that, similar properties have been obtained for chaotic maps of the interval and so called functions of bounded p-variation by F. Hofbauer and

G. Keller [19], M. Rychlik [40] and others [50, 49, 26]. In all those systems the correlations decay exponentially fast. Recently, statistical properties have been obtained for chaotic billiards by L. Bunimovich, Ya. Sinai and N. Chernov [10, 12] and hyperbolic attractors by V. Afraimovich, N. Chernov and E. Sataev [1], again for Hölder continuous functions on the phase space. The correlations in the latter systems have been only bounded by a stretched exponential function, and the techniques of the proofs were rather specific and intricate. Further development of those techniques are needed to cover many more physically interesting systems, e.g., gases of hard spheres. We work in this direction here. We obtain rather strong bounds on correlations and limit theorems for abstract dynamical systems based on flexible and easy-to-check mixing assumptions. We also show how our abstract results work for concrete dynamical systems with both high and low mixing rates.

Some authors have raised a question how large the classes of functions satisfying the CLT are, see, e.g., a survey by Denker [14]. Sometimes those classes are much larger than those of Hölder continuous or bounded variation functions. In response to this question, we find very large classes of observables with rapid decay of correlations which also satisfy limit theorems.

A universal method of proving statistical properties is to find partitions of the phase space with sufficiently high mixing rates and then condition the given functions on atoms of those partitions. We also employ this strategy assuming the existence of partitions with mixing coefficients decaying at specific rates. We do not discuss how to construct such partitions in applications or how to obtain necessary bounds on mixing coefficients (the latter problem is, however, partially solved in Section 5).

We assume that the phase space is equipped with a metric and μ is a nonatomic measure. For any finite or countable measurable partition $\mathcal{A} = \{A_1, A_2, \ldots\}$ of the phase space M we denote $\operatorname{diam} \mathcal{A} = \sup_i \{\operatorname{diam} A_i\}$. We put $\mathcal{A}_n = T^{-n}\mathcal{A} = \{T^{-n}A_i\}$ for $n \geq 0$ and $\mathcal{A}_{n,k} = \mathcal{A}_n \vee \cdots \vee \mathcal{A}_k$ for $k \geq n \geq 0$. A measure of dependence between two partitions $\mathcal{A} = \{A_i\}$ and $\mathcal{B} = \{B_j\}$ of the phase space M is defined to be

$$\beta(\mathcal{A}, \mathcal{B}) = \sum_{i,j} |\mu(A_i \cap B_j) - \mu(A_i)\mu(B_j)|$$

Based on this measure, we put[1]

$$\beta(n) = \beta(\mathcal{A}_0, \mathcal{A}_n) \tag{5}$$

for any $n \geq 0$ and

$$\beta_N(n) = \max_{0 \leq k \leq N-n-1} \beta(\mathcal{A}_{0,k}, \mathcal{A}_{k+n,N-1}), \tag{6}$$

for any $N \geq n \geq 0$. We call $\beta(n)$ and $\beta_N(n)$ the mixing coefficients of the partition $\mathcal{A}$. Their rates of decay, as $n \to \infty$, are crucial for our proofs of statistical properties.

For any function $F \in L_2(M)$ and a partition $\mathcal{A}$ of the phase space M we denote $\bar{F}_{\mathcal{A}} = \langle F|\mathcal{A}\rangle$ the conditional expectation of F with respect to the σ-algebra generated

[1]note that $\beta(n)$ here does not have its usual meaning, $\sup_k \beta(\mathcal{A}_{0,k}, \mathcal{A}_{k+n,\infty})$.

by $\mathcal{A}$, and $\Delta_{\mathcal{A}} F = F - \bar{F}_{\mathcal{A}}$. Given an $F \in L_2(M)$, we define two functions, $\mathcal{H}_F(d)$ and $\mathcal{L}_F(d)$ defined for $d > 0$, that characterize F:

$$\mathcal{H}_F(d) = \sup_{\operatorname{diam} \mathcal{A} \leq d} ||\Delta_{\mathcal{A}} F||_2$$

and

$$\mathcal{L}_F(d) = \sup_B \int_B (F - \langle F \rangle)^2 \, d\mu(x)$$

where the supremum is taken over all subsets $B \subset M$ such that $\mu(B) \leq d$. For any $F \in L_2(M)$ both $\mathcal{H}_F(d)$ and $\mathcal{L}_F(d)$ approach zero as $d \to 0$, and the rate of decrease of those functions will be also crucial in our proofs of statistical properties. Certain properties of functions $\mathcal{H}$ and $\mathcal{L}$ are discussed in Section 2.

Theorem 1.1 (Bound on Correlations) *For any functions $F, G \in L_2(M)$, any $n \geq 1$ and any partition $\mathcal{A}$ we have*

$$\begin{aligned} |C_{F,G}(n)| &\leq 2|\mathcal{L}_F(\beta(n)) \cdot \mathcal{L}_G(\beta(n))|^{1/2} \\ &+ ||F||_2 \cdot ||\Delta_{\mathcal{A}} G||_2 + ||G||_2 \cdot ||\Delta_{\mathcal{A}} F||_2 + ||\Delta_{\mathcal{A}} F||_2 \cdot ||\Delta_{\mathcal{A}} G||_2 \end{aligned}$$

In particular, if $d = \operatorname{diam} \mathcal{A}$, then

$$\begin{aligned} |C_{F,G}(n)| &\leq 2|\mathcal{L}_F(\beta(n)) \cdot \mathcal{L}_G(\beta(n))|^{1/2} \\ &+ ||F||_2 \cdot \mathcal{H}_G(d) + ||G||_2 \cdot \mathcal{H}_F(d) + \mathcal{H}_F(d) \cdot \mathcal{H}_G(d) \end{aligned}$$

If F and G are essentially bounded, then the first term in both bounds does not exceed $2||F||_\infty ||G||_\infty \beta(n)$. If both F and G belong in $L_{2+\delta}(M)$ for some $\delta > 0$, then the first term in these bounds does not exceed $2||F||_{2+\delta} ||G||_{2+\delta} \beta(n)^{\delta/(2+\delta)}$.

We prove the CLT under various hypotheses. Our objective is to cover several types of dynamical systems: for the ones with high mixing rates one can considerably enlarge the classes of phase functions, slower mixing rates require more stringent assumptions on the functions. In all our CLT's we assume that the first moment of the autocorrelation function is finite:

$$\sum_{n=1}^{\infty} n |C_{F,F}(n)| < \infty \tag{7}$$

As usual, we assume that σ_F defined in (3) is strictly positive. A criterion for degeneracy ($\sigma_F = 0$) is given in Section 3.

Theorem 1.2 (Central Limit Theorem: General Format) *Let $F \in L_2(M)$ have an autocorrelation function $C_{F,F}(n)$ with a finite first moment (7) and $\sigma_F > 0$. Assume that for any $N \geq 1$ there is a partition $\mathcal{A} = \mathcal{A}^{(N)}$ of the phase space M such that*

$$||\Delta_{\mathcal{A}} F||_2 = o(N^{-1/2}) \tag{8}$$

Furthermore, assume that there is an integer valued function $n = n(N)$ such that $n \to \infty$ as $N \to \infty$ and $n = o(N)$ satisfying two conditions:

$$\beta_N(n) = o(n/N) \tag{9}$$

and

$$\mathcal{L}_F(n/N) = o(1/n) \tag{10}$$

Then the central limit theorem (2) holds.

A sufficient condition for (8) is

$$\mathcal{H}_F(\operatorname{diam} \mathcal{A}^{(N)}) = o(N^{-1/2}) \tag{11}$$

In the next theorem we specify two particular cases where Theorem 1.2 works:

Theorem 1.3 (CLT: Two Special Formats) *Let $F \in L_2(M)$ have an autocorrelation function $C_{F,F}(n)$ with a finite first moment (7) and $\sigma_F > 0$. Assume that for any $N \geq 1$ there is a partition $\mathcal{A} = \mathcal{A}^{(N)}$ of the phase space M such that $||\Delta_{\mathcal{A}} F||_2 = o(N^{-1/2})$. Furthermore, assume that either of the following two conditions holds:*
(i) *the function F is essentially bounded and there are $\varepsilon_1, \varepsilon_2 > 0, \varepsilon_1 > 2\varepsilon_2$, such that*

$$\beta_N(n) = O(n^{-1} \ln^{-\varepsilon_1} n) \tag{12}$$

where $n = [N^{1/2} \ln^{-\varepsilon_2} N]$;
(ii) *the function F belongs in $L_{2+\delta}(M)$ with some $\delta > 0$, and there are $0 < s < s_0 := (2 + 2/\delta)^{-1}$ and $t > s/(s_0^2 - ss_0)$ such that*

$$\beta_N(n) = O(n^{-\frac{2+\delta}{\delta} - t}) \tag{13}$$

where $n = [N^{\delta/(2+2\delta)-s}]$;
Then the central limit theorem (2) holds.

We prove WIP under slightly more stringent conditions than those of Theorem 1.3. It makes, however, no difference in all our applications in Sections 6 and 7.

Theorem 1.4 (Weak Invariance Principle) *Let $F \in L_2(M)$ have an autocorrelation function $C_{F,F}(n)$ with a finite first moment and $\sigma_F > 0$. Assume that for any $N \geq 1$ there is a partition $\mathcal{A} = \mathcal{A}^{(N)}$ of the phase space M such that*

$$||\Delta_{\mathcal{A}} F||_2 = o(N^{-1}) \tag{14}$$

Furthermore, assume that either of the following two conditions holds:
(i) *the function F is essentially bounded and there are $\varepsilon_1, \varepsilon_2 > 0, \varepsilon_1 > 2\varepsilon_2$ such that*

$$\beta_N(n) = O(n^{-1} \ln^{-\varepsilon_1} n) \tag{15}$$

where $n = [N^{1/2} \ln^{-\varepsilon_2} N]$*;*
(ii) *the function* F *belongs in* $L_{2+\delta}(M)$ *with some* $\delta > 0$*, and there are* $0 < s < s_0 = (2 + 2/\delta)^{-1}$ *and* $t > s/(s_0^2 - ss_0)$ *such that*

$$\beta_N(n) = O(n^{-\frac{2+\delta}{\delta} - t}) \tag{16}$$

for every $n \in [n_0, N]$*, where* $n_0 = [N^{\delta/(2+2\delta)-s}]$*;*
Then the weak invariance principle holds.

Next, we specify two classes of dynamical systems with high mixing rates and obtain two corollaries to our theorems.

Definition 1.5 *We say that a dynamical system* (M, T, μ) *has exponential mixing rates if for any* $a \in (0, 1]$ *and any* $N \geq 1$ *there is a partition* $\mathcal{A} = \mathcal{A}^{(N,a)}$ *of the space* M *such that diam*$(\mathcal{A}) \leq c_1 \lambda_1^{N^a}$ *and* $\beta_N([N^a]) \leq c_2 \lambda_2^{N^a}$ *with some* $c_1, c_2 > 0$ *and* $\lambda_1, \lambda_2 \in (0, 1)$ *(the last four constants may depend on* a *but not on* N*).*

Definition 1.6 *We say that a dynamical system* (M, T, μ) *has stretched exponential mixing rates if there is a constant* $\gamma \in (0, 1)$ *such that for any* $a \in (0, 1]$ *and any* $N \geq 1$ *there is a partition* $\mathcal{A} = \mathcal{A}^{(N,a)}$ *of the space* M *such that diam*$(\mathcal{A}) \leq c_1 \lambda_1^{N^{a\gamma}}$ *and* $\beta_N([N^a]) \leq c_2 \lambda_2^{N^{a\gamma}}$ *with some* $c_1, c_2 > 0$ *and* $\lambda_1, \lambda_2 \in (0, 1)$ *(again, these four constants may depend on* a *but not on* N*).*

In Sections 6 and 7 we show that Axiom A diffeomorphisms and expanding interval maps have exponential mixing rates. We also show that chaotic billiards and hyperbolic attractors have stretched exponential mixing rates (the latter does not, however, rule out possible exponential mixing rates for those systems!).

Corollary 1.7 *Let a dynamical system* (M, T, μ) *have exponential mixing rates. Assume that a function* $F \in L_2(M)$ *satisfies one of the two following conditions:*
(i) *it is essentially bounded and* $\mathcal{H}_F(d) = O(1/|\ln d|^{2+\varepsilon})$ *for some* $\varepsilon > 0$*;*
(ii) *it belongs in* $L_{2+\delta}(M)$ *for some* $\delta > 0$ *and* $\mathcal{H}_F(d) = O(1/|\ln d|^{2+2/\delta+\varepsilon})$ *for some* $\varepsilon > 0$*;*
Then both CLT and WIP hold provided $\sigma_F > 0$*.*

Corollary 1.8 *Let a dynamical system* (M, T, μ) *have stretched exponential mixing rates. Assume that a function* $F \in L_2(M)$ *satisfies one of the two following conditions:*
(i) *it is essentially bounded and* $\mathcal{H}_F(d) = O(1/|\ln d|^{2/\gamma+\varepsilon})$ *for some* $\varepsilon > 0$*;*
(ii) *it belongs in* $L_{2+\delta}(M)$ *for some* $\delta > 0$ *and* $\mathcal{H}_F(d) = O(1/|\ln d|^{(2+2/\delta)/\gamma+\varepsilon})$ *for some* $\varepsilon > 0$*;*
Then both CLT and WIP hold provided $\sigma_F > 0$*.*

We now recall how the partitions $\mathcal{A} = \mathcal{A}^{(N)}$ are constructed in the existing works. For Axiom A diffeomorphisms, see also Section 7, the partition $\mathcal{A}^{(\mathcal{N})}$ can obtained by a refinement of a finite generating Markov partition $\mathcal{A}_0$ according to the rule

$$\mathcal{A}^{(N)} = \vee_{i=-L}^{L} T^{-i} \mathcal{A}_0 \tag{17}$$

where $L = L(N)$ is some function of N. For expanding interval maps, see Section 6, one can take the natural partition $\mathcal{A}_0$ into intervals where the map T is monotone and put

$$\mathcal{A}^{(N)} = \vee_{i=0}^{L} T^{-i} \mathcal{A}_0 \tag{18}$$

where $L = L(N)$ is again some function in N. In both cases the mixing coefficients for $\mathcal{A}^{(N)}$ decay exponentially fast in N. Moreover, the existence of a common generator $\mathcal{A}_0$ for all $\mathcal{A}^{(N)}$ allows one to reduce the CLT and WIP to the existing limit theorems for strongly mixing stationary processes, see refs. [19, 14, 23], thus bypassing our theorems.

However, the proofs of the CLT and WIP for hyperbolic systems with singularities, such as billiards and attractors, are based on partitions $\mathcal{A}^{(N)}$ constructed for every N independently, so that those partitions are by no means related to each other. This makes direct applications of the theory of strongly mixing random processes no longer possible. Thus, the CLT and WIP have to be proven by separate arguments. Such arguments, in an abstract manner, are developed here. It is very likely that systems with even more dilute chaotic properties (such as nonuniformly hyperbolic systems, e.g., gases of hard balls [43]) will require further work in this direction.

The structure of the paper is the following. In Section 2 we discuss the classes of phase functions to which our results apply and prove Theorem 1.1. In Section 3 we prove all our CLT's (Theorems 1.2 and 1.3 and their variations, including a CLT for nonmixing transformations). In the next section we prove the WIP (Theorem 1.4). In Section 5 we develop techniques for bounding mixing coefficients for partitions with strong Markov properties. To this end we define and explore Markov approximations. In Section 6 we apply our theorems to expanding interval maps (both uniform and nonuniform ones). Lastly, in Section 7 we discuss hyperbolic maps: Anosov and Axiom A diffeomorphisms, chaotic billiards and attractors.

2 Mixing and decay of correlations

A dynamical system (M, T, μ) is said to be mixing if for any two measurable subsets $A, B \subset M$ one has $\lim \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$, or, equivalently, for any $F, G \in L_2(M)$ one has $\lim C_{F,G}(n) = 0$.

The rates of the decay of correlations characterize the degree of 'chaoticity' of the dynamical system. It is, however, known long ago (see, e.g., [13]) that even for extremely chaotic dynamical systems there are functions in $L_2(M)$ for which the correlations decay arbitrarily slowly. At present, it is even known that slow decay of correlations is not just a 'bad luck', but a typical feature of functions in $L_2(M)$:

Proposition 2.1 *Let $T : M \to M$ be an invertible aperiodic ergodic dynamical system. Let K_n be an arbitrary sequence such that $K_n \to \infty$ as $n \to \infty$. Then for any F in a second category (i.e., G_δ) subset of $L_2(M)$ one has*

$$\limsup_{n\to\infty} K_n \frac{C_{F,F}(0) + \cdots + C_{F,F}(n-1)}{n} = \infty$$

This proposition follows from a recent category theorem by D. Volný [47]. The latter says that if an invertible dynamical system is aperiodic and ergodic and $K_n \to \infty$ as $n \to \infty$, then for each F in a G_δ subset of the Hilbert space $\{F \in L_2(M),\ \langle F \rangle = 0\}$ one has

$$\limsup_{n\to\infty} K_n ||n^{-1} \sum_{i=0}^{n-1} F(T^i x)||_2 = \infty$$

From this result Proposition 2.1 follows due to the equation (21) from Section 3.

Volný's category theorem shows that for any ergodic and aperiodic dynamical system the CLT fails for a G_δ subset of functions in $L_2(M)$. One can also observe that the CLT is closely related to the rate of convergence in Birkhoff's ergodic theorem, so that a slow convergence in that theorem rules out the CLT. The convergence in the ergodic theorem is known to be arbitrarily slow for typical functions in $L_p(M)$ for every $p \geq 1$ (including L_∞) and even for typical continuous functions in case M is an interval (see [24]). It is very likely that the correlations decay arbitrarily slowly for typical functions in $L_p(M)$ for every $p \geq 2$, in L_∞, and even for typical continuous functions, whatever chaotic a dynamical system.

We conclude that a fast decay of autocorrelations and the CLT are, in a sense, a good luck for phase functions, and one has to restrict oneself to certain narrow classes of "nice" functions in order to bound the correlations and prove the CLT.

Favorite types of functions for which the correlations have been effectively bounded and the CLT has been proven for smooth dynamical systems on manifolds are smooth and 'nearly smooth' functions, such as Hölder continuous ones and those of bounded p-variation (cf. Section 6). Various generalizations of Hölder continuous functions were introduced in [44, 38, 1, 10, 30, 12], but none of them went beyond bounded and almost everywhere continuous functions. On the other hand, for certain classes of maps with uniform mixing rates (e.g., for Axiom A diffeomorphisms and expanding maps of the interval), strong bounds on correlations and limit theorems can be carried out for strikingly large classes of functions [14], see our Sections 6 and 7 for more detail.

The classes of functions for which our theorems work are much larger than those of Hölder continuous or bounded p-variation functions or all the generalizations in [44, 38, 1, 10, 30, 12]. Our theorems do not, however, cover the classes of functions that have been covered in [14] for systems with uniform mixing rates. We intentionally do not assume any uniform mixing rates here, because that would prevent us from studuing even relatively 'nice' systems like chaotic billiards or attractors, let alone hyperbolic systems with nonuniform expansion rates such as gases of hard spheres.

We now discuss the properties of the functions $\mathcal{H}$ and $\mathcal{L}$ defined is Section 1.

First, if the observable F is continuous and satisfies the condition $|F(x) - F(y)| \leq h(\mathrm{dist}(x, y))$ for some function $h(d), d > 0$, then $\mathcal{H}_F(d) \leq h(d)$. In particular, if F is a Hölder continuous function with exponent a, then $\mathcal{H}_F(d) \leq \mathrm{const} \cdot d^a$. Ya. Sinai [44] and

M. Ratner [38] have studied a class of functions F such that

$$|F(x) - F(y)| < \text{const} \cdot \rho^{|\ln \text{dist}(x,y)|^{\kappa}} \tag{19}$$

with some $\rho, \kappa \in (0,1)$. For such functions $\mathcal{H}_F(d) \leq \text{const} \cdot \rho^{|\ln d|^{\kappa}}$. In recent works [10, 12, 1] a class of piecewise Hölder continuous functions on compact manifolds M has been introduced. Those are defined to be Hölder continuous with an exponent a on a finite number of open subsets in M with piecewise smooth boundaries that generate a (mod 0) partition of M. Assume also that the measure μ does not grow too fast near the boundaries of those open sets, i.e. $\mu(\varepsilon-$neighborhood of those boundaries) <const$\cdot\varepsilon^{a'}$ with some $a' > 0$. Then $\mathcal{H}_F(d) \leq \text{const} \cdot d^{a_1}$ with $a_1 = \min\{a, a'/2\}$. In Section 6 we show that for all the functions of bounded p-variation on $[0,1]$ one has $\mathcal{H}_F(d) \leq \text{const} \cdot d^a$ with $a = \min\{1/2, 1/p\}$.

Therefore, the function $\mathcal{H}_F(d)$ decreases as $d \to 0$ as a power function in d for all the popular classes of phase functions, except for Sinai-Ratner functions (19) for which the speed of decrease is still very high. As Corollaries 1.7 and 1.8 show, even a much slower asymptotics, such as $O(1/|\ln d|^p)$ with a $p > 2$, is enough for both CLT and WIP.

Furthermore, no bounds on the function $\mathcal{H}_F(d)$ will imply a.e. continuity or even boundedness of the phase function F. Even if $\mathcal{H}_F(d) \leq \text{const} \cdot d$ (the strongest possible bound for nonconstant functions), the function F may be everywhere discontinuous and even essentially unbounded on every open subset of M.

We now turn to the function $\mathcal{L}_F(d)$. It is monotone increasing, continuous and convex in $d \geq 0$ for any $F \in L_2(M)$. It satisfies $\mathcal{L}_F(d) \geq \text{const} \cdot d$ for all $d \in [0,1]$ unless the function F is a constant a.e., in which case $\mathcal{L}_F(d) \equiv 0$. One has $\mathcal{L}_F(d) \leq \text{const} \cdot d$ iff the function F is essentially bounded. One has $\mathcal{L}_F(d) \leq \text{const} \cdot d^{\delta/(2+\delta)}$ for some $\delta > 0$ iff $F \in L_{2+\delta}(M)$.

It is useful to note that the function $\mathcal{L}_F(d)$ can be defined solely in terms of the distribution of the observable F. That is, if F and F_1 have the same distribution, then $\mathcal{L}_F(d) \equiv \mathcal{L}_{F_1}(d)$. In particular, $\mathcal{L}_F(d) \equiv \mathcal{L}_{F \circ T^n}(d)$ for all $n \geq 1$. It is also clear that for any partition $\mathcal{A}$ we have $\mathcal{L}_{\bar{F}_{\mathcal{A}}}(d) \leq \mathcal{L}_F(d)$.

Proof of Theorem 1.1. Since $C_{F+a,G+b}(n) = C_{F,G}(n)$ for any constants a and b and any $n \geq 1$, we can (and will) assume that $\langle F \rangle = \langle G \rangle = 0$.

We first observe that given an $n \geq 0$ and a partition $\mathcal{A}$, we have

$$\begin{aligned} C_{F,G}(n) &= \langle \bar{F}(T^n x) \bar{G}(x) \rangle \\ &+ \langle \Delta F(T^n x) \bar{G}(x) \rangle + \langle \bar{F}(T^n x) \Delta G(x) \rangle + \langle \Delta F(T^n x) \Delta G(x) \rangle \end{aligned} \tag{20}$$

(for brevity, we omit the subscript $\mathcal{A}$). We apply Schwarz' inequality to the last three terms in the RHS of (20) and immediately obtain the last three terms of the bounds claimed in Theorem 1.1. Notice that these terms depend on the partition $\mathcal{A}$ and the functions F and G but have nothing to do with the map T. These terms measure the error of the approximation of the functions F and G by their mean values on the atoms of the partition $\mathcal{A}$.

The first term in the RHS of (20) is $C_{\bar F,\bar G}(n)$, the correlation of two discrete functions $\bar F(T^n x)$ and $\bar G(x)$, the latter is $\mathcal{A}$ measurable, and the former is $\mathcal{A}_n = T^{-n}\mathcal{A}$ measurable. This term can be rewritten as

$$\begin{aligned} C_{\bar F,\bar G}(n) &= \sum_{i,j} \bar F_j \bar G_i \Big(\mu(T^{-n}A_j \cap A_i) - \mu(A_j)\mu(A_i)\Big) \\ &\le \sum_{i,j} |\bar F_j| \cdot |\bar G_i| \cdot |\mu(T^{-n}A_j \cap A_i) - \mu(A_j)\mu(A_i)|. \end{aligned}$$

where $\bar F_i = \bar F(x)$ and $\bar G_i = \bar G(x)$ for $x \in A_i \in \mathcal{A}$. The quantity $C_{\bar F,\bar G}(n)$ depends mainly on how fast the map T mixes up the atoms of the partition $\mathcal{A}$, the functions F and G play minor roles.

We will denote $\Sigma^+_{i,j}$ (and $\Sigma^-_{i,j}$) the summation over such i and j that the difference $\mu(T^{-n}A_j \cap A_i) - \mu(A_j)\mu(A_i)$ is positive (respectively, negative). We then make an observation that

$$\Sigma^+_{i,j} \Big[\mu(T^{-n}A_j \cap A_i) - \mu(A_j)\mu(A_i)\Big] \le \beta(n),$$

and so

$$\begin{aligned} &\Sigma^+_{i,j} |\bar F_j| \cdot |\bar G_i| \cdot |\mu(T^{-n}A_j \cap A_i) - \mu(A_j)\mu(A_i)| \\ &\quad \le \sup_{B:\,\mu(B)=\beta(n)} \int_B |\bar F(T^n x)\bar G(x)|\, d\mu(x). \end{aligned}$$

Given a set $B \subset M$ such that $\mu(B) = \beta(n)$, applying Schwarz' inequality yields

$$\begin{aligned} &\int_B |\bar F(T^n x)\bar G(x)|\, d\mu(x) \\ &\le \Big(\int_B \bar F^2(T^n x)\, d\mu(x)\Big)^{1/2} \cdot \Big(\int_B \bar G^2(x)\, d\mu(x)\Big)^{1/2} \\ &\le \mathcal{L}_F(\beta(n)) \cdot \mathcal{L}_G(\beta(n)) \end{aligned}$$

In a similar fashion we observe that

$$\Sigma^-_{i,j} \Big[\mu(A_j)\mu(A_i) - \mu(T^{-n}A_j \cap A_i)\Big] \le \beta(n),$$

and then estimate

$$\begin{aligned} &\Sigma^-_{i,j} |\bar F_j| \cdot |\bar G_i| \cdot |\mu(A_j)\mu(A_i) - \mu(T^{-n}A_j \cap A_i)| \\ &\le \Big(\Sigma^-_{i,j} \bar F_j^2 [\mu(A_j)\mu(A_i) - \mu(T^{-n}A_j \cap A_i)]\Big)^{1/2} \cdot \Big(\Sigma^-_{i,j} \bar G_i^2 [\mu(A_j)\mu(A_i) - \mu(T^{-n}A_j \cap A_i)]\Big)^{1/2} \\ &\le \left(\sup_{B:\mu(B)=\beta(n)} \int_B \bar F^2(x) d\mu(x)\right)^{1/2} \cdot \left(\sup_{B:\mu(B)=\beta(n)} \int_B \bar G^2(x) d\mu(x)\right)^{1/2} \\ &\le \mathcal{L}_F(\beta(n)) \cdot \mathcal{L}_G(\beta(n)) \end{aligned}$$

Theorem 1.1 is proven.

3 Central limit theorem

We start with a few helpful observations. First,

$$\operatorname{Var} S_N = N \cdot C_{F,F}(0) + 2 \sum_{n=1}^{N-1} (N-n) C_{F,F}(n) \tag{21}$$

Our assumption on the first moment of the autocorrelation function (7) implies that $\operatorname{Var} S_N = N\sigma_F^2 + O(1)$. More specifically,

$$\lim_{N\to\infty} (\operatorname{Var} S_N - N\sigma_F^2) = -2 \sum_{n=1}^{\infty} n C_{F,F}(n) = \text{const} < \infty \tag{22}$$

There are two distinct cases now. The first one is $\sigma_F^2 = 0$. Then, due to (22), the sequence $\{\operatorname{Var} S_N\}$ is bounded and has a finite limit. Due to a general result by V. Leonov [29], see also [21, Theorem 18.2.2], there is another function $G \in L_2(M)$ such that

$$F(x) = G(Tx) - G(x) \quad \text{a.e..} \tag{23}$$

In this case the function F is called a coboundary, and one has $S_N(x) = G(T^N x) - G(x)$ and $\lim_{N\to\infty} \operatorname{Var} S_N = 2\operatorname{Var} G$ (the latter was shown in [29]). The central limit theorem (2) certainly fails in this case.

We will discuss only the main case: $\sigma_F^2 > 0$. The equation (22) allows one to replace $\operatorname{Var} S_N$ by $N\sigma_F^2$ in the CLT (2), which then looks more like the classical central limit theorem in probability theory.

We first prove a general CLT based on a Lindeberg-type condition. A similar theorem has been obtained first in applications [7, 10]. We prove this theorem in a very general setup.

Theorem 3.1 *Let $F \in L_2(M)$ have an autocorrelation function $C_{F,F}(n)$ with a finite first moment and $\sigma_F > 0$. Assume that for any $N \geq 1$ there is a partition $\mathcal{A} = \mathcal{A}^{(N)}$ of the phase space M such that $||\Delta_{\mathcal{A}} F||_2 = o(N^{-1/2})$. Furthermore, assume that there are two integer valued functions $p = p(N)$ and $q = q(N)$, both monotone increasing to infinity as $N \to \infty$, such that $p = o(N)$ and $q = o(p)$, and*

$$\lim_{N\to\infty} N p^{-1} \beta_N(q) = 0 \tag{24}$$

and for any $\varepsilon > 0$

$$\lim_{N\to\infty} \langle p^{-1} S_p^2 \cdot \chi_{\{x:\, S_p^2 > \varepsilon^2 N\}} \rangle = 0 \tag{25}$$

Then the CLT (2) holds.

Here χ_B stands for the characteristic function (indicator) of the subset $B \subset M$.

Proof. Our proof is based on Bernstein's technique of approximation to sums of weakly dependent random variables by those of independent ones. This method is also known as “the big small block technique”. We partition the ‘time’ interval [0,N-1] into an alternating sequence of ‘big’ blocks of length $p = p(N)$ and ‘small’ blocks of length $q = q(N)$. The number of big blocks is $k = [N/(p+q)] \sim N/p$. The last remaining block is of length $N - kp - (k-1)q < p + q$.

We denote by $\Delta_r, 1 \le r \le k$, the big blocks and set

$$S_p^{(r)}(x) = \sum_{i \in \Delta_r} F(T^i x),$$

$$S'_N = \sum_{r=1}^{k} S_p^{(r)}, \quad \text{and} \quad S''_N = S_N - S'_N$$

The residual sum S''_N contains no more than $w = kq + p$ summands $F(T^n x)$. Hence,

$$\operatorname{Var} S''_N \le w \left(|C_{F,F}(0)| + 2 \sum_{n=1}^{\infty} |C_{F,F}(n)| \right) \le \text{const} \cdot w$$

Using Chebyshev’s inequality gives

$$\mu \left\{ x : \frac{|S''_N|}{\sigma_F \sqrt{N}} > \varepsilon \right\} \le \frac{\operatorname{Var} S''_N}{\varepsilon^2 \sigma_F^2 N} \le \text{const} \cdot \left(\frac{q}{p} + \frac{p}{N} \right) \to 0 \tag{26}$$

as $N \to \infty$. So, one can replace S_N by S'_N in (2), cf. Lemma 18.4.1 in [21] and also our Lemma 3.4 below. We will also replace $\operatorname{Var} S_N$ by $k \operatorname{Var} S_p^{(1)}$ based on the following observation, cf. (21):

$$|\operatorname{Var} S_N - k \operatorname{Var} S_p^{(1)}| \le w |C_{F,F}(0)|$$
$$+2 \sum_{n=1}^{p-1} (w + nk) |C_{F,F}(n)| + 2 \sum_{n=p}^{N-1} (N-n) |C_{F,F}(n)|,$$

and so $|\operatorname{Var} S_N - k \operatorname{Var} S_p^{(1)}| = o(N) = o(\operatorname{Var} S_N)$ as $N \to \infty$. As a result, (2) is equivalent to

$$\lim_{N \to \infty} \mu \left\{ x : \frac{S'_N}{(k \operatorname{Var} S_p^{(1)})^{1/2}} \le z \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} \, dt \tag{27}$$

The next step in the proof is the approximation of the variable

$$U_N = \frac{S'_N}{(k \operatorname{Var} S_p^{(1)})^{1/2}} = \frac{S_p^{(1)} + \cdots + S_p^{(k)}}{(k \operatorname{Var} S_p^{(1)})^{1/2}} \tag{28}$$

by the sum

$$U'_N = u_N^{(1)} + \cdots + u_N^{(k)} \tag{29}$$

of k independent random variables $u_N^{(i)}, 1 \le i \le k$, each of which has the same distribution as that of

$$w_N = \frac{S_p^{(1)}}{(k \operatorname{Var} S_p^{(1)})^{1/2}} \tag{30}$$

By approximation we mean that the limit distributions of U_N and U'_N as $N \to \infty$ coincide provided one of them exists. In virtue of the continuity theorem of probability theory it suffices to show that

$$\sup_{\lambda \in \Lambda} |\varphi_{U_N}(\lambda) - \varphi_{U'_N}(\lambda)| \to 0 \tag{31}$$

for any compact subset $\Lambda \in \mathbb{R}$, where $\varphi_U(\lambda)$ stands for the characteristic function of a random variable U.

Lemma 3.2 *Let $U, U' \in L_2(M)$ and $\langle U \rangle = \langle U' \rangle$. Then*

$$|\varphi_U(\lambda) - \varphi_{U'}(\lambda)| \leq 4\lambda^{2/3} ||U - U'||_2^{2/3}$$

Proof. Applying Schwarz' inequality gives

$$|\varphi_U(\lambda) - \varphi_{U'}(\lambda)| = |\langle e^{i\lambda U} \cdot (e^{i\lambda(U'-U)} - 1)\rangle|$$

$$\leq \langle |e^{i\lambda(U'-U)} - 1|^2 \rangle^{1/2} = 2\langle \sin^2(\lambda(U' - U)/2)\rangle^{1/2}$$

Therefore, for any $\varepsilon > 0$

$$\begin{aligned} |\varphi_U(\lambda) - \varphi_{U'}(\lambda)| &\leq \varepsilon + 2\mu \left\{ x : |U' - U| \geq \varepsilon |\lambda|^{-1} \right\} \\ &\leq 2\varepsilon + 2\varepsilon^{-2}\lambda^2 \mathrm{Var}\,(U' - U) \end{aligned}$$

where we have used Chebyshev's inequality. Minimizing the RHS of the last bound with respect to ε gives the lemma.

Corollary 3.3 *Let $U' = (1 + \delta)U$ with a constant δ and $\langle U \rangle = 0$. Then*

$$|\varphi_U(\lambda) - \varphi_{U'}(\lambda)| \leq 4\lambda^{2/3} |\delta|^{2/3} ||U||_2^{2/3}$$

Corollary 3.4 *Let*

$$U = \rho \sum_{i=1}^{n} F(T^{r_i} x) \quad \text{and} \quad U' = \rho \sum_{i=1}^{n} \bar{F}_{\mathcal{A}}(T^{r_i} x)$$

for some $n \geq 1$ and $0 \leq r_1 < r_2 < \cdots < r_n \leq N$, with a constant factor $\rho > 0$. Then

$$|\varphi_U(\lambda) - \varphi_{U'}(\lambda)| \leq 4\lambda^{2/3} \rho^{2/3} n^{2/3} ||\Delta_{\mathcal{A}} F||_2^{2/3}$$

We now ‘discretize’ the functions U_N and U'_N. For any $r, 1 \le r \le k$, denote

$$\bar{S}_p^{(r)}(x) = \sum_{i \in \Delta_r} \bar{F}_{\mathcal{A}}(T^i x)$$

(we omit the subscript $\mathcal{A}$ for the sake of brevity). Then replacing all the $S_p^{(i)}$ by $\bar{S}_p^{(i)}$ in (28)-(30) gives new, discrete functions $\bar{U}_N$ and $\bar{U}'_N$.

Observe that $||U_N||_2 \le$const$< \infty$, $||U'_N||_2 \le$const$< \infty$ and

$$|\mathrm{Var}\, S_p^{(r)} - \mathrm{Var}\, \bar{S}_p^{(r)}| < \mathrm{const} \cdot p^{3/2} ||\Delta_{\mathcal{A}} F||_2 \tag{32}$$

for every $r, 1 \le r \le k$. Recall that $\sqrt{N}||\Delta_{\mathcal{A}}F||_2 \to 0$ as $N \to \infty$. Then Corollaries 3.3 and 3.4 imply that

$$|\varphi_{U_N}(\lambda) - \varphi_{\bar{U}_N}(\lambda)| \le \mathrm{const} \cdot \lambda^{2/3} (\sqrt{N}||\Delta_{\mathcal{A}}F||_2)^{2/3} \to 0$$

as $N \to \infty$, and

$$|\varphi_{U'_N}(\lambda) - \varphi_{\bar{U}'_N}(\lambda)| \le \mathrm{const} \cdot \lambda^{2/3} (\sqrt{N}||\Delta_{\mathcal{A}}F||_2)^{2/3} \to 0$$

as $N \to \infty$, where the convergence is uniform in λ on any compact subset $\Lambda \subset \mathbb{R}$.

It remains to show that

$$\sup_{\lambda \in \Lambda} |\varphi_{\bar{U}_N}(\lambda) - \varphi_{\bar{U}'_N}(\lambda)| \to 0 \tag{33}$$

for any compact $\Lambda \subset \mathbb{R}$, and we will complete the proof of (31).

Lemma 3.5 *Let $0 \le n < n + q < N$. Let $\Phi_1(x)$ and $\Phi_2(x)$ be two complex valued functions on M such that Φ_1 is $\mathcal{A}_{0,n}$ measurable, Φ_2 is $\mathcal{A}_{n+q,N-1}$ measurable, and $|\Phi_k| \le M_k$ for $k = 1, 2$. Then*

$$|\langle \Phi_1 \Phi_2 \rangle - \langle \Phi_1 \rangle \langle \Phi_2 \rangle| \le M_1 M_2 \beta_N(q) \tag{34}$$

Applying (34) to characteristic functions involved in (33) for each small block of length q gives the bound

$$|\varphi_{\bar{U}_N}(\lambda) - \varphi_{\bar{U}'_N}(\lambda)| \le (k-1)\beta_N(q) \le N p^{-1} \beta_N(q)$$

for any real λ. Making use of the assumption (24) completes the proof of (33) and that of (31).

The last step consists in proving the CLT for the sum (29) of independent identically distributed random variables. It is known in probability theory (see, e.g., the proof of Theorem 18.4.1 in [21]) that the condition (25) is sufficient for that CLT, i.e. U'_N converge, as $N \to \infty$, in distribution to the standard normal law. Theorem 3.1 is proven.

The condition (25) in Theorem 3.1 is an analogue of Lindeberg's condition in probability theory and deserves a separate discussion. Note that

$$\lim_{N\to\infty} \langle p^{-1}S_p^2\rangle = \sigma_F^2 < \infty,$$

and

$$\mu\{x:\ S_p^2 > \varepsilon^2 N\} \leq \frac{\langle S_p^2\rangle}{\varepsilon^2 N} \leq \frac{2\sigma_F^2}{\varepsilon^2}\cdot\frac{p}{N} \to 0 \tag{35}$$

as $N\to\infty$. These simple observations make (25) look trivial. However, (25) is the most subtle and controversial among all the hypotheses of the theorem. It is really hard to check in applications. It imposes many unpleasant constraints in the theory of strongly mixing stationary processes, see, e.g., the refs. [20, 21, 34, 14].

We prove Lindeberg's condition (25) by the following trick:

$$\begin{aligned}\langle p^{-1}S_p^2\cdot\chi_{\{x:\,S_p^2>\varepsilon^2N\}}\rangle &\leq p\mathcal{L}_F(\mu\{x:\ S_p^2>\varepsilon^2N\})\\ &\leq p\mathcal{L}_F\left(\frac{2\sigma_F^2}{\varepsilon^2}\cdot\frac{p}{N}\right)\leq\frac{2\sigma_F^2}{\varepsilon^2}\cdot p\mathcal{L}_F(p/N)\end{aligned}\tag{36}$$

(recall that $\mathcal{L}_F(d)$ is a convex function in d). Therefore, a sufficient condition for (25) is

$$\lim_{N\to\infty} p\mathcal{L}_F(p/N) = 0 \tag{37}$$

It certainly works only if $p = o(\sqrt{N})$. We can afford functions $p(N)$ like this, as opposite to the theory of strongly mixing stationary processes, where $p(N)$ must grow faster than $\sqrt{N}$, see [20, 34]. Such a fast growth is necessary in that theory to ensure a linear in N asymptotics of $\operatorname{Var} S_N$. We guarantee this asymptotics by a different assumption, that of the finiteness of the first moment of the autocorrelation function (7).

We now derive Theorem 1.2 from Theorem 3.1. Let $q(N) = n(N)$. In order to define the function $p(N)$ we consider the sequence $a_N = q\mathcal{L}_F(q/N)$, which approaches zero, as $N\to\infty$, due to (10). Let $p(N) = [a_N^{-1/3}q(N)]$. Then

$$p\mathcal{L}_F(p/N) \leq a_N^{-2/3}q\mathcal{L}_F(q/N)\to 0$$

as $N\to\infty$, and we get (37). Thus, Theorem 3.1 applies and we obtain Theorem 1.2.

Theorem 1.3 immediately follows from Theorem 1.2 by a direct inspection. We leave the details to the reader.

Warning. Even though Theorem 1.3 requires solely power law mixing rates, one may attempt to relax those rates even further. We have certain reasons to believe, however, that the mixing rates in (12) and (13) are very close to the ones necessary for the CLT. Our belief is based on the following two counterexamples due to R. Bradley [6] in the theory of absolutely regular processes.

Let $\{X_i\}$ be a stationary random process and $\mathcal{F}_{m,n}$ denote σ-algebra generated by $X_m, \ldots, X_n$ for $-\infty \le m \le n \le \infty$. For any two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ one defines

$$\beta_*(\mathcal{A}, \mathcal{B}) = \sup \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |\mu(A_i \cap B_j) - \mu(A_i)\mu(B_j)|$$

where supremum is taken over all pairs of partitions $\{A_1, \ldots, A_I\}$ and $\{B_1, \ldots, B_J\}$ such that $A_i \in \mathcal{A}$ and $B_j \in \mathcal{B}$. Then one defines $\beta_*(n) = \beta_*(\mathcal{F}_{-\infty,0}, \mathcal{F}_{n,\infty})$. A process $\{X_i\}$ is said to be absolutely regular if $\beta_*(n) \to 0$ as $n \to \infty$. Clearly, the function $\beta_*(n)$ is an analogue of our mixing coefficient $\beta_N(n)$.

Bradley [6] has found two absolutely regular stationary processes that do not obey the CLT. His first example satisfies

$$|X_i| \le C < \infty \quad \text{and} \quad \beta_*(n) = O(n^{-1} \ln^3 n)$$

and the other one satisfies

$$\langle |X_i|^{2+\delta} \rangle < \infty \ \text{ and } \ \beta_*(n) = O\left((n^{-1} \ln^3 n)^{(2+\delta)/\delta} \right)$$

It is very likely that there exist dynamical systems whose mixing coefficients, $\beta_N(n)$, have the above asymptotics, which do not obey the CLT.

On the other hand, our central limit theorems may produce a wrong impression that the decay of correlations (or, equivalently, the mixing property) is a necessary condition for the CLT (2). As R. Burton and M. Denker have recently proven [11], for any ergodic and aperiodic dynamical system (M, T, μ), mixing or not, there is a dense subset of functions in $L_2(M)$ that obey the CLT. In particular, they proved this statement for irrational rotations of the circle, based on Fourier representations. Of course, the functions satisfying the CLT for irrational circle rotations, are very peculiar, and generic smooth functions on the circle do not obey the CLT.

There is, however, a wide class of nonmixing transformations for which generic smooth functions do obey the CLT. This class, as we will see in Sections 6 and 7, appears in many important applications, so that we need to devote a few lines to it.

Let $M = M_1 \cup \cdots \cup M_r$ be a partition into disjoint subsets, and $T : M_i \to M_{i+1}$ for all $i = 1, \ldots, r-1$ and $T(M_r) = M_1$. In that case we call T a cyclic permutation. Denote T_i the restriction of T^r on M_i and μ_i the conditional μ-measure on M_i. For any function F on M denote F_i its restriction on M_i.

Proposition 3.6 *Let T be a cyclic permutation as defined above. Let the dynamical system (M_i, T_i, μ_i) and the function F_i satisfy the hypotheses of Theorem 3.1 for every $i = 1, \ldots, r$. Then the function F obeys the CLT (2), unless F is a coboundary.*

To prove this proposition one can apply Theorem 3.1 to the function $F(x) + F(Tx) + \cdots + F(T^{r-1}x)$ on M_1. The argument is straightforward and we omit it.

Note that the autocorrelation function $C_{F,F}(n)$ in the case of the last proposition does not necessarily decay at all. It decays if and only if $\langle F_i \rangle = 0$ for every $i = 1, \dots, r$. Otherwise the correlation function converges to a periodic nonzero function. Its average over the period is zero, and so the correlations cancel out in long partial sums S_N, as $N \to \infty$, giving way to the CLT.

We conclude this section with a vector CLT. Let $\mathbf{F} = \{F_1, \dots, F_r\}$ be a vector function on M. Assume that each component F_i satisfies the hypotheses of Theorem 3.1 (in particular, $\langle F_i \rangle = 0$), and let

$$\sum_{n=1}^{\infty} n |C_{F_i,F_j}(n)| < \infty$$

for any pair F_i, F_j. Consider partial sums $\mathbf{S}_n = \{S_{1,n}, \dots, S_{r,n}\} = \mathbf{F} + \mathbf{F} \circ T + \cdots + \mathbf{F} \circ T^{n-1}$. Define an $r \times r$ matrix, $\mathbf{V} = \{v_{ij}\}$, with components

$$\begin{aligned} v_{ij} &= \lim_{n\to\infty} \frac{1}{n} \langle S_{i,n} S_{j,n} \rangle \\ &= C_{F_i F_j}(0) + \sum_{n=1}^{\infty} C_{F_i F_j}(n) + \sum_{n=1}^{\infty} C_{F_j F_i}(n) \end{aligned}$$

In particular, $v_{ii} = \sigma_{F_i}^2$. The matrix $\mathbf{V}$ is symmetric and nonnegative definite.

Theorem 3.7 (Vector CLT) *Under the above assumptions, if the matrix* $\mathbf{V}$ *is nondegenerate (i.e., if* $\det \mathbf{V} \neq 0$*), the vector function* $\mathbf{S}_n/\sqrt{n}$ *converges in distribution to a normal law with zero mean and covariance matrix* $\mathbf{V}$*.*

The following lemma provides a useful criterion for the degeneracy of the matrix $\mathbf{V}$:

Lemma 3.8 *The matrix* $\mathbf{V}$ *is degenerate (*$\det \mathbf{V} = 0$*) iff there is a linear combination* $H = a_1 F_1 + \cdots + a_r F_r$ *with some* $a_1^2 + \cdots + a_r^2 \neq 0$*, which is a coboundary function:* $H(x) = G(Tx) - G(x)$ *a.e. for some* $G \in L_2(M)$*.*

To prove this lemma, we denote $\mathbf{A} = \{a_1, \dots, a_r\}$ and observe that

$$\sigma_H^2 = \sum_{i,j} a_i a_j v_{ij} = \mathbf{A}\mathbf{V}\mathbf{A}^* \tag{38}$$

(here $*$ means the transposition), so that $\det \mathbf{V} = 0$ is equivalent to $\sigma_H^2 = 0$ for some $\mathbf{A} \neq 0$. Then we recall (23) and obtain the lemma.

Theorem 3.7 can be proven in two ways. First, one can just repeat the proof of Theorem 3.1 almost word by word. Alternatively, we can reduce Theorem 3.7 to Theorem 3.1 by a simple trick. For any linear combination $H = \mathbf{A}\mathbf{F}^* = a_1 F_1 + \cdots + a_r F_r$ we have $\sigma_H \neq 0$ (due to Lemma 3.8), and so by Theorem 3.1 the function

$$U_n := \mathbf{A}\mathbf{S}_n^*/\sqrt{n} = [H(x) + H(Tx) + \cdots + H(T^{n-1}x)]/\sqrt{n}$$

converges in distribution to a normal law with zero mean and variance σ_H^2. By the continuity theorem the characteristic function

$$\varphi_{U_n}(t) = \varphi_{\mathbf{S}_n/\sqrt{n}}(a_1 t, \ldots, a_r t)$$

converges, as $n \to \infty$, to $\exp(-\sigma_H^2 t^2/2) = \exp(-\mathbf{AVA}^* t^2/2)$ for any real t. Therefore, the characteristic function $\varphi_{\mathbf{S}_n/\sqrt{n}}(t_1, \ldots, t_r)$ converges pointwise to $\exp(-\mathbf{TVT}^*/2)$, where $\mathbf{T} = \{t_1, \ldots, t_r\}$. This limit function is the characteristic function of a normal distribution with zero mean and covariance matrix $\mathbf{V}$, so that Theorem 3.7 again follows by the continuity theorem.

4 Weak invariance principle

The classical strategy to prove a weak invariance principle consists of two steps. First, one shows that finite dimensional distributions of W_N, see (4), converge to those of the Wiener process W. For one-dimensional distributions this convergence is equivalent to the central limit theorem. Thus, the above convergence can be termed the multidimensional central limit theorem. Its exact statement follows.

Theorem 4.1 (Multidimensional CLT) *Under the assumptions of Theorem 3.1 for any $k \geq 1$ and $0 < t_1 < t_2 < \cdots < t_k \leq 1$ the joint distribution of the functions $W_N(t_1), W_N(t_2), \ldots, W_N(t_k)$ converges, as $N \to \infty$, to that of $W(t_1), W(t_2), \ldots, W(t_k)$.*

Proof. This theorem is a direct generalization of Theorem 3.1. Indeed, it is enough to prove that, given $0 < t_1 < \cdots < t_k \leq 1$, the joint distribution of the variables

$$\frac{S_{[Nt_1]}}{\sigma_F\sqrt{Nt_1}}, \; \frac{S_{[Nt_2]} - S_{[Nt_1]}}{\sigma_F\sqrt{N(t_2 - t_1)}}, \ldots, \frac{S_{[Nt_k]} - S_{[Nt_{k-1}]}}{\sigma_F\sqrt{N(t_k - t_{k-1})}}, \tag{39}$$

converges to the k-dimensional normal distribution with zero mean and unit covariance matrix. Individually, the functions (39) converge in distribution to the standard normal law by Theorem 3.1. Then we remove a small block of length q from each sum $S_{[Nt_{i+1}]-[Nt_i]}, 1 \leq i \leq k-1$, starting at its left end, i.e. we remove terms $F \circ T^{[Nt_i]}, \ldots, F \circ T^{[Nt_i]+q-1}$. By applying the arguments of the proof of Theorem 3.1 it is easy to show that this removal does not harm the limit distribution of the vector (39). Furthermore, the joint characteristic function of (39) can be approximated by the product of individual characteristic functions of the involved variables. This completes the proof of Theorem 4.1.

The second part of the proof of the WIP is far more difficult: one has to verify the tightness of W_N. The weak invariance principle follows from the multidimensional CLT and the tightness of $\{W_N\}$, see the book by P. Billingsley [4] and discussions in surveys [34, 36].

To prove the tightness, certain criteria are available, see [4, 34]. In particular, the following condition is sufficient for the tightness of W_N, see [4, Theorem 8.4]. If for any $\varepsilon > 0$ there is an $A_\varepsilon > 1$ and an $N_\varepsilon \geq 1$ such that

$$\mu\{\max_{0\leq i\leq N} |S_i| \geq A_\varepsilon \sigma_F \sqrt{N}\} \leq \frac{\varepsilon}{A_\varepsilon^2} \tag{40}$$

for all $N \geq N_\varepsilon$, then $\{W_N\}$ is tight.

We will also need the following simple lemma:

Lemma 4.2 *Under the assumptions of Theorem 3.1 the sequence $\{S_N^2/N\}$ is uniformly integrable.*

Proof. This lemma is an easy consequence of the central limit theorem, cf. [21, Theorem 18.4.2]. Indeed, let $F_N(z) = \mu\{x : S_N(x)/\sqrt{N} \leq z\}$. Then, by the CLT, for any $A > 0$

$$\int_{|z|\leq A} z^2\, dF_N(z) \to \sigma_F^2 \int_{|z|\leq A} z^2\, d\Phi(z) \quad \text{as } N \to \infty$$

where $\Phi(z)$ is the distribution function of the standard normal law. In addition,

$$\int_{-\infty}^{\infty} z^2\, dF_N(z) = \mathrm{Var}\left(\frac{S_N}{\sqrt{N}}\right) \to \sigma_F^2 \quad \text{as } N \to \infty$$

These prove Lemma 4.2.

Proof of Theorem 1.4 (i). We follow a short and elegant proof of a similar theorem in the theory of strongly mixing random processes by H. Oodaira and K. Yoshihara [33]. All we have to do is verify the sufficient condition (40).

Let $k = [N/n]$. First, we 'discretize' the sums S_i for all $i \leq N$ with respect to the partition $\mathcal{A}$. We denote

$$\bar{S}_i = \sum_{j=0}^{i-1} \bar{F}_{\mathcal{A}}(T^j x) \quad \text{and} \quad \Delta S_i = \sum_{j=0}^{i-1} \Delta_{\mathcal{A}} F(T^j x)$$

and so

$$\mu\{\max_{i\leq N} |S_i| \geq A\sigma_F\sqrt{N}\} \leq \mu\{\max_{i\leq N} |\bar{S}_i| \geq 2^{-1}A\sigma_F\sqrt{N}\} + \mu\{\max_{i\leq N} |\Delta S_i| \geq 2^{-1}A\sigma_F\sqrt{N}\}$$

The last term in the RHS does not exceed

$$\begin{aligned} &\mu\{|\Delta_{\mathcal{A}}F| + \cdots + |(\Delta_{\mathcal{A}}F)\circ T^{N-1}| \geq 2^{-1}A\sigma_F\sqrt{N}\} \\ &\leq N\mu\{|\Delta_{\mathcal{A}}F| \geq 2^{-1}A\sigma_F/\sqrt{N}\} \leq \frac{4N^2||\Delta_{\mathcal{A}}F||_2^2}{A^2\sigma_F^2} \end{aligned} \tag{41}$$

This last quantity approaches zero as $N \to \infty$ due to (14).

It remains to show that

$$\mu\{\max_{i\le N}|\bar S_i| \ge 2^{-1}A_\varepsilon\sigma_F\sqrt{N}\} \le \frac{5}{6}\cdot\frac{\varepsilon}{A_\varepsilon^2} \tag{42}$$

for some $A_\varepsilon > 1$ and all sufficiently large N.

Given $\varepsilon > 0$, choose an $A = A_\varepsilon$ (> 1) so that

$$\mu\{|S_i| > 12^{-1}A\sigma_F\sqrt{i}\} \le \frac{\varepsilon}{6A^2}$$

for all $i \ge 1$, which is possible due to Lemma 4.2. Then for any $N \ge 1$ and $i \in [1, N]$

$$\begin{aligned} &\mu\{|\bar S_i| > 6^{-1}A\sigma_F\sqrt{i}\} \\ &\le \mu\{|S_i| > 12^{-1}A\sigma_F\sqrt{i}\} + \mu\{|\Delta S_i| > 12^{-1}A\sigma_F\sqrt{i}\} \\ &\le \frac{\varepsilon}{6A^2} + \frac{144 i^2 ||\Delta_{\mathcal{A}}F||_2^2}{A^2\sigma_F^2} \le \frac{\varepsilon}{3A^2} \end{aligned} \tag{43}$$

for all sufficiently large N and all $i \le N$, based again on the bound (41).

Since $F(x)$ is essentially bounded,

$$|\bar F(x)| + \cdots + |\bar F(T^{2n}x)| \le 6^{-1}A\sigma_F\sqrt{N} \quad \text{a.e.} \tag{44}$$

for all sufficiently large N.

We now prove (42). For every $i \in [1, N]$ we consider a subset $E_i = \{x \in M : \max_{j<i}|\bar S_j(x)| < 2^{-1}A\sigma_F\sqrt{N} \le |\bar S_i(x)|\}$. Obviously, $\{E_i\}$ are disjoint and their total measure enters (42). Now,

$$\begin{aligned} &\mu\{\max_{i\le N}|\bar S_i| \ge 2^{-1}A\sigma_F\sqrt{N}\} \\ &\le \mu\{|\bar S_N| \ge 6^{-1}A\sigma_F\sqrt{N}\} + \mu\left(\cup_{i=1}^{N-1}[E_i \cap \{|\bar S_N - \bar S_i| \ge 3^{-1}A\sigma_F\sqrt{N}\}]\right) \\ &\le \mu\{|\bar S_N| \ge 6^{-1}A\sigma_F\sqrt{N}\} + \sum_{r=0}^{k-2}\mu\left(\cup_{i=1}^{n}[E_{rn+i} \cap \{|\bar S_N - \bar S_{rn+i}| \ge 3^{-1}A\sigma_F\sqrt{N}\}]\right) \\ &\quad + \sum_{i=(k-1)n+1}^{N-1}\mu\{|\bar S_N - \bar S_i| \ge 3^{-1}A\sigma_F\sqrt{N}\} \end{aligned}$$

(from now on we assume that N is large enough and make use of (44))

$$\begin{aligned} &\le \mu\{|\bar S_N| \ge 6^{-1}A\sigma_F\sqrt{N}\} + \sum_{r=0}^{k-2}\mu\left(\cup_{i=1}^{n}[E_{rn+i} \cap \{|\bar S_N - \bar S_{(r+2)n}| \ge 6^{-1}A\sigma_F\sqrt{N}\}]\right) \\ &= \mu\{|\bar S_N| \ge 6^{-1}A\sigma_F\sqrt{N}\} + \sum_{r=0}^{k-2}\mu\left([\cup_{i=1}^{n}E_{rn+i}] \cap \{|\bar S_N - \bar S_{(r+2)n}| \ge 6^{-1}A\sigma_F\sqrt{N}\}\right) \end{aligned}$$

Since $\cup_{i=1}^n E_{rn+i}$ is $\mathcal{A}_{0,(r+1)n}$ measurable and $\{|\bar S_N - \bar S_{(r+2)n}| \ge 3^{-1} A\sigma_F \sqrt{N}\}$ is $\mathcal{A}_{(r+2)n,N-1}$ measurable, we apply Lemma 3.5 and (15) and get

$$\sum_{r=0}^{k-2} \mu\left([\cup_{i=1}^n E_{rn+i}] \cap \{|\bar S_N - \bar S_{(r+2)n}| \ge 6^{-1} A\sigma_F \sqrt{N}\}\right)$$

$$\le \sum_{r=0}^{k-2} \mu\left(\cup_{i=1}^n E_{rn+i}\right) \mu\left\{|\bar S_N - \bar S_{(r+2)n}| \ge 6^{-1} A\sigma_F \sqrt{N}\right\} + k\beta_N(n)$$

$$\le \sum_{r=0}^{k-2} \mu\left(\cup_{i=1}^n E_{rn+i}\right) \mu\left\{|\bar S_{N-(r+2)n}| \ge 6^{-1} A\sigma_F \sqrt{N-(r+2)n}\right\} + k\beta_N(n)$$

$$\le \frac{\varepsilon}{3A^2} + k\beta_N(n) = \frac{\varepsilon}{3A^2} + o(1),$$

where we have used (43). The clause (i) of Theorem 1.4 is now proven.

Proof of Theorem 1.4 (ii). As in the previous proof, we only have to verify (40). We start with a discretization of S_N with respect to the partition $\mathcal{A}$. This repeats word by word that from the previous proof, and we again reduce (40) to (42).

The rest of the proof follows that of the corresponding theorem in the theory of strongly mixing random processes [33, Theorem 2]. We truncate the discrete function $\bar F_{\mathcal{A}}(x)$ by the rule

$$\hat F(x) = \begin{cases} \bar F_{\mathcal{A}}(x) & \text{if } |\bar F_{\mathcal{A}}(x)| \le N^{\frac{1}{2(1+\delta)}} \\ 0 & \text{otherwise} \end{cases}$$

and denote $F^*(x) = \bar F_{\mathcal{A}}(x) - \hat F(x)$. We then put $\hat S_n(x) = \sum_{i=0}^{n-1} \hat F(T^i x)$ and $S_n^*(x) = \sum_{i=0}^{n-1} F^*(T^i x)$ for each $n \ge 1$.

Lemma 4.3 $N^{-1}\mathrm{Var}\,\hat S_N \to \sigma_F^2$ *as* $N \to \infty$.

Proof. First recall that $N^{-1}\mathrm{Var}\,S_N \to \sigma_F^2$ as $N \to \infty$, and (32) can be rewritten as

$$|\mathrm{Var}\,\bar S_N - \mathrm{Var}\,S_N| \le \mathrm{const} \cdot N^{3/2} ||\Delta_{\mathcal{A}} F||_2$$

Therefore $N^{-1}\mathrm{Var}\,\bar S_N \to \sigma_F^2$ as $N \to \infty$. Then, by the expansion (21) it is enough to prove that

$$\lim_{N\to\infty} \sum_{n=0}^{N-1} |C_{\bar F,\bar F}(n) - C_{\hat F,\hat F}(n)| = 0$$

Observe that

$$C_{\bar F,\bar F}(n) - C_{\hat F,\hat F}(n) = C_{\hat F,F^*}(n) + C_{F^*,\hat F}(n) + C_{F^*,F^*}(n)$$

We make use of Theorem 1.1 and the assumption (16), and obtain

$$\sum_{n=n_0}^{N-1} |C_{\bar F,\bar F}(n) - C_{\hat F,\hat F}(n)| \le \mathrm{const} \cdot ||F||_{2+\delta} ||F^*||_{2+\delta} \to 0$$

as $N \to \infty$. Besides,

$$\sum_{n=0}^{n_0-1} |C_{\bar{F},\bar{F}}(n) - C_{\hat{F},\hat{F}}(n)| \leq \text{const} \cdot n_0 \cdot \left(||\hat{F}||_\infty \langle |F^*| \rangle + \langle |F^*|^2 \rangle \right) \to 0$$

as $N \to \infty$, cf. also (47) below. Lemma 4.3 is proven.

Next, we have

$$\begin{aligned} &\mu\{\max_{i \leq N} |\bar{S}_i| \geq 2^{-1} A\sigma_F \sqrt{N}\} \\ &\leq \mu\{\max_{i \leq N} |\hat{S}_i - i\langle \hat{F} \rangle| \geq 4^{-1} A\sigma_F \sqrt{N}\} + \mu\{\max_{i \leq N} |S_i^* - i\langle F^* \rangle| \geq 4^{-1} A\sigma_F \sqrt{N}\} \qquad (45) \\ &\leq \mu\{\max_{i \leq N} |\hat{S}_i - i\langle \hat{F} \rangle| \geq 4^{-1} A\sigma_F \sqrt{N}\} + \mu \left\{ \sum_{i=0}^{N-1} |F^*(T^i x) - \langle F^* \rangle| \geq 4^{-1} A\sigma_F \sqrt{N} \right\} \end{aligned}$$

We put $k = [N/n_0]$. In virtue of (16) $k\beta_N(n_0) \to 0$ as $N \to \infty$. Since the function $|\hat{F}(x) - \langle \hat{F} \rangle|$ is bounded ($\leq 2N^{1/2(1+\delta)}$), we have

$$\sum_{i=0}^{2n_0-1} |\hat{F}(T^i x) - \langle \hat{F} \rangle| \leq 4n_0 N^{\frac{1}{2(1+\delta)}} < 50^{-1} A\sigma_F \sqrt{N} \quad \text{a.e.}$$

for all sufficiently large A. Based on these observations and Lemma 4.3 and by repeating the arguments of the proof of (42) we obtain that for any $\varepsilon > 0$ there is an $A_\varepsilon > 1$ such that

$$\mu\{\max_{i \leq N} |\hat{S}_i - i\langle \hat{F} \rangle| \geq 4^{-1} A\sigma_F \sqrt{N}\} \leq \frac{5\varepsilon}{12A^2}$$

for all $A \geq A_\varepsilon$.

It remains to bound the last term in the RHS of (45). Due to Chebyshev's inequality we have

$$\mu \left\{ \sum_{i=0}^{N-1} |F^*(T^i x) - \langle F^* \rangle| \geq 4^{-1} A\sigma_F \sqrt{N} \right\} \leq \frac{16}{A^2 \sigma_F^2 N} \left\langle \left(\sum_{i=0}^{N-1} |F^*(T^i x) - \langle F^* \rangle| \right)^2 \right\rangle$$

Next,

$$\begin{aligned} &N^{-1} \left\langle \left(\sum_{i=0}^{N-1} |F^*(T^i x) - \langle F^* \rangle| \right)^2 \right\rangle \\ &\leq \langle |F^*(x) - \langle F^* \rangle|^2 \rangle + 2 \sum_{i=1}^{N-1} \langle |F^*(x) - \langle F^* \rangle| \cdot |F^*(T^i x) - \langle F^* \rangle| \rangle \qquad (46) \\ &\leq 2n_0 ||F^*||_2^2 + 2 \sum_{i=n_0}^{N-1} \langle |F^*(x) - \langle F^* \rangle| \cdot |F^*(T^i x) - \langle F^* \rangle| \rangle \end{aligned}$$

Observe that

$$\langle |F^*| \rangle \leq ||F^*||_{2+\delta} ||F||_{2+\delta}^{1+\delta} N^{-1/2} \quad \text{and} \quad \langle |F^*|^2 \rangle \leq ||F^*||_{2+\delta}^2 ||F||_{2+\delta}^{\delta} N^{-\delta/2(1+\delta)} \qquad (47)$$

Therefore, the first term in the RHS of the last inequality in (46) vanishes as $N \to \infty$. The second term involves the correlations between two functions, one is $\mathcal{A}$ measurable, the other is $\mathcal{A}_i$ measurable for an $i \geq n_0$. Thus, we can apply Theorem 1.1 and get

$$\langle |F^*(x) - \langle F^* \rangle| \cdot |F^*(T^i x) - \langle F^* \rangle| \rangle$$

$$\leq \langle |F^*(x) - \langle F^* \rangle| \rangle \cdot \langle |F^*(T^i x) - \langle F^* \rangle| \rangle + 2||F^* - \langle F^* \rangle||_{2+\delta}^2 \beta_N^{\frac{\delta}{2+\delta}}(i)$$

$$\leq 4\langle |F^*| \rangle^2 + 8||F^*||_{2+\delta}^2 \beta_N^{\frac{\delta}{2+\delta}}(i)$$

$$\leq 4N^{-1}||F||_{2+\delta}^{2(1+\delta)}||F^*||_{2+\delta}^2 + \text{const} \cdot ||F^*||_{2+\delta}^2 i^{-1-t\delta/(2+\delta)}$$

Summing over $i \in [n_0, N-1]$ shows that the second term in the RHS of (46) vanishes as $N \to \infty$, too. Therefore, for all sufficiently large N we have

$$\mu\left\{\sum_{i=0}^{N-1} |F^*(T^i x) - \langle F^* \rangle| \geq 4^{-1} A \sigma_F \sqrt{N}\right\} \leq \frac{5\varepsilon}{12A^2}$$

and the bound (42) is established. The clause (ii) of Theorem 1.4 is now proven.

The following extension of our WIP can be obtained in the same fashion as Proposition 3.6:

Proposition 4.4 *Let T be a cyclic permutation as defined in Section 3. Let the dynamical system (M_i, T_i, μ_i) and the function F_i satisfy the clause (i) or the clause(ii) of Theorem 1.4 for every $i = 1, \dots, r$. Then the function F satisfies WIP as well, unless it is a coboundary.*

5 Markov approximation

In the present paper we do not discuss how to establish high mixing rates required by our theorems. However, we show how the hypotheses on mixing coefficients in our theorems can be considerably relaxed, with the help of a novel general technique involving Markov approximations generated by finite or countable partitions of the phase space.

The idea of Markov approximations defined here goes back to that of coarse-graining of the phase space, popular among physicists, see a discussion in the survey [32]. It is also close to Ulam's construction [46]. Markov approximations for dynamical systems are motivated mainly by Markov partitions first obtained for toral automorphisms by R. Adler and B. Weiss [3], then for Anosov systems by Ya. Sinai [42] and for Axiom A diffeomorphisms by R. Bowen [5], cf. also Section 7. Historically, however, Markov partitions were defined by very fine topological properties. Those are of little help for Markov approximations in the measure-theoretic sense, which we need here. Other partitions with good Markov properties in the measure-theoretic sense alone (like Markov sieves, cf. Section 7) can work even better than original Markov partitions and are often much easier to construct. This is why we will define Markov approximations based on an *arbitrary* partition of the phase space.

Let $\mathcal{A} = \{A_i\}$ be a finite or countable partition of the phase space M into subsets of positive measure. We define a probabilistic stationary Markov chain with transition matrix

$$\pi_{ij} = \mu(T^{-1}A_j/A_i) = \mu(T^{-1}A_j \cap A_i)/\mu(A_i) \tag{48}$$

and stationary distribution

$$p_i = \mu(A_i). \tag{49}$$

We say that this Markov chain approximates the dynamics within n iterates of T if $(n+1)$-dimensional distributions of the Markov chain

$$p_{i_0 i_1 \cdots i_n} = p_{i_0}\pi_{i_0 i_1}\cdots\pi_{i_{n-1}i_n} \tag{50}$$

are close to those of the dynamical system

$$\begin{aligned}\mu(T^{-n}A_{i_n} \cap T^{-(n-1)}A_{i_{n-1}} \cap \cdots \cap A_{i_0}) &= \mu(T^{-n}A_{i_n}/T^{-(n-1)}A_{i_{n-1}} \cap \cdots \cap A_{i_0}) \\ \times\mu(T^{-(n-1)}A_{i_{n-1}}/T^{-(n-2)}A_{i_{n-2}} \cap \cdots \cap A_{i_0}) &\cdots \mu(T^{-1}A_{i_1}/A_{i_0})\mu(A_{i_0})\end{aligned} \tag{51}$$

The expansion (51) holds whenever $\mu(T^{-(n-1)}A_{i_{n-1}} \cap \cdots \cap A_{i_0}) > 0$.

It is clear that (50) and (51) are close provided

$$\mu(T^{-n}A_{i_n}/T^{-(n-1)}A_{i_{n-1}} \cap \cdots \cap A_{i_0}) \approx \mu(T^{-1}A_{i_n}/A_{i_{n-1}}) = \pi_{i_{n-1}i_n} \tag{52}$$

whenever $\mu(T^{-1}A_{i_l} \cap A_{i_{l-1}}) > 0$ for all $l = 1, 2, \ldots, n$. We specify below what the sign "$\approx$" in (52) means. One can think of (52) as a 'short memory' condition for the dynamical system (M, T, μ) within the first n iterates.

We now introduce a measure of closeness of the 'long-memory' and 'short-memory' conditional probabilities involved in (52) by

$$\nu_N = \sup_{n\le N} \sum_{i_0,\ldots,i_n} |\mu(T^{-n}A_{i_n}/T^{-(n-1)}A_{i_{n-1}} \cap \cdots \cap A_{i_0}) - \mu(T^{-1}A_{i_n}/A_{i_{n-1}})|$$

$$\times\mu(T^{-(n-1)}A_{i_{n-1}} \cap \cdots \cap A_{i_0}) \tag{53}$$

In this and other formulas the summation is taken over such indexes that all the conditional measures are defined.

Certainly, (53) is an 'integral' measure of closeness of long- and short-memory conditional probabilities rather than their individual closeness indicated by (52). Recall that given two probability distributions $P = \{p_i\}$ and $Q = \{q_i\}$ on the same index set $\{i\}$, the distance in variation between P and Q is defined to be

$$\mathrm{Var}(P, Q) = \frac{1}{2}\sum_i |p_i - q_i| \tag{54}$$

Then (53) is twice the mean distance in variation between the long- and short-memory conditional distributions on $\{T^{-n}A_i\}, i \geq 1$.

Based on (53), one can estimate how close the finite dimensional distributions (50) and (51) are in the variational metric:

$$\sum_{i_0,\dots,i_n} |\mu(T^{-n}A_{i_n} \cap \dots \cap A_{i_0}) - p_{i_0 i_1 \cdots i_n}| \leq (n-1)\nu_N \tag{55}$$

for $n \leq N$. The proof of (55) goes by induction in n.

Certain bounds on the closeness of finite dimensional conditional distributions with long and short memories follow from (53). First,

$$\begin{aligned}\sum_{i_0,\dots,i_{n+k}} &|\mu(T^{-(n+k)}A_{i_{n+k}} \cap \dots \cap T^{-n}A_{i_n}/T^{-(n-1)}A_{i_{n-1}} \cap \dots \cap A_{i_0}) \\ &-\mu(T^{-(n+k)}A_{i_{n+k}} \cap \dots \cap T^{-n}A_{i_n}/T^{-(n-1)}A_{i_{n-1}})| \\ &\times\mu(T^{-(n-1)}A_{i_{n-1}} \cap \dots \cap A_{i_0}) \leq (2k+1)\nu_N,\end{aligned} \tag{56}$$

for any $n, k \geq 1$ such that $n + k \leq N$. This can be derived from (53) by induction in k. Second,

$$\begin{aligned}\sum_{i_0,\dots,i_m,i_n,\dots,i_{n+k}} &|\mu(T^{-(n+k)}A_{i_{n+k}} \cap \dots \cap T^{-n}A_{i_n}/T^{-m}A_{i_m} \cap \dots \cap A_{i_0}) \\ -\mu(T^{-(n+k)}A_{i_{n+k}} \cap \dots \cap T^{-n}A_{i_n}&/T^{-m}A_{i_m})| \cdot \mu(T^{-m}A_{i_m} \cap \dots \cap A_{i_0}) \\ &\leq (2k+2n-2m+1)\nu_N\end{aligned} \tag{57}$$

for any $n > m$ and $k \geq 0$ such that $n+k \leq N$. This readily follows from (56). Note that there is "a gap in time", between m and n in (57). In a similar manner another useful bound follows:

$$\begin{aligned}\sum_{i_0,\dots,i_m,i_n,\dots,i_{n+k}} &|\mu(T^{-(n+k)}A_{i_{n+k}} \cap \dots \cap T^{-n}A_{i_n}/T^{-m}A_{i_m} \cap \dots \cap A_{i_0}) \\ &-\mu(T^{-(n+k)}A_{i_{n+k}} \cap \dots \cap T^{-n}A_{i_n})| \cdot \mu(T^{-m}A_{i_m} \cap \dots \cap A_{i_0}) \\ \leq \sum_{i_0,\dots,i_m,i_n} &|\mu(T^{-n}A_{i_n}/T^{-m}A_{i_m} \cap \dots \cap A_{i_0}) - \mu(T^{-n}A_{i_n})| \cdot \mu(T^{-m}A_{i_m} \cap \dots \cap A_{i_0}) \\ &+(2k+2n-2m)\nu_N \\ \leq \sum_{i_m,i_n} &|\mu(T^{-n}A_{i_n} \cap T^{-m}A_{i_m}) - \mu(T^{-n}A_{i_n})\mu(T^{-m}A_{i_m})| + (2k+4n-4m)\nu_N\end{aligned} \tag{58}$$

We will utilize this last bound later.

The Markov approximations introduced here can be also interpreted in terms of symbolic dynamics. For every point $x \in M$ we define a one-sided symbolic sequence $\omega(x) = \{\omega_i(x)\}_{i=0}^{\infty}$ by $T^i x \in A_{\omega_i(x)}$ for all $i = 0, 1, \dots$. Denote Ω_M the set of symbolic sequences $\omega(x)$ for all $x \in M$. The transformation T of M is conjugate to the left shift σ_L on Ω_M, defined by $\sigma_L(\omega_0, \omega_1, \ldots) = (\omega_1, \omega_2, \ldots)$. The invariant measure μ on M induces a measure μ_Ω on Ω_M, which is invariant under the shift σ_L. The dynamical system $(\Omega_M, \sigma_L, \mu_\Omega)$ is commonly referred to as the symbolic representation of (M, T, μ) generated by the partition $\{A_i\}$. One usually demands that different points $x, y \in M$ have different symbolic sequences $\omega(x) \neq \omega(y)$, in which case the partition $\{A_i\}$ is called a generator. If the transformation T is invertible, one usually assigns double-sided sequences $\omega(x) = \{\omega_i(x)\}_{i=-\infty}^{\infty}$ to points $x \in M$ by the rule $T^i x \in A_{\omega_i(x)}, -\infty < i < \infty$.

The equations (48) and (49) define another, Markov, measure $\mu^{(M)}$ on the set of all one-sided symbolic sequences $\mathbb{N}^{\mathbb{Z}_+} \supset \Omega_M$. (In the case of invertible T we get a Markov measure $\mu^{(M)}$ on the set of double-sided symbolic sequences $\mathbb{N}^{\mathbb{Z}} \supset \Omega_M$). The measure $\mu^{(M)}$ and the invariant measure μ_Ω always have common one- and two-dimensional distributions on the cylinders $\{\omega_0\}$ and $\{\omega_0\omega_1\}$. The condition (52) requires that the multidimensional distributions of $\mu^{(M)}$ and μ_Ω are close on cylinders $\{\omega_0, \ldots, \omega_n\}$ of length $n + 1$.

Let us emphasize that good Markov approximations can be defined not only for chaotic dynamical systems. For example, let $M = S^1$ and T be a rotation through a rational angle of $360m/n$ degrees, with $m, n \in \mathbb{N}$. Then a partition of S^1 into equal arcs of $360/n$ degrees generates a perfect Markov approximation ($\nu_N \equiv 0$ for all $N \geq 1$). If the fraction m/n is irreducible, then the associated Markov chain is ergodic, but not mixing. Needless to say, the system is extremely regular (all the points are periodic).

In other words, the quality of the Markov approximation (the smallness of ν_N in (53)) is an exclusive property of the partition $\mathcal{A}$, and does not say anything about the ergodic or statistical properties of the dynamical system (M, T, μ). One needs additional information about the approximating Markov chain in order to derive statistical properties of the dynamical system. Next we specify what kind of information is sufficient. Recall that our purpose is to bound mixing coefficients $\beta_N(n)$ and $\beta(n)$ for the partition $\mathcal{A}$.

The following lemma readily follows from the estimate (58):

Lemma 5.1 *$\beta_N(n) \leq \beta(n) + 2(N + n)\nu_N$ for any $n < N$.*

This lemma shows that, if the Markov approximation generated by the partition $\mathcal{A}$ is good enough (so that ν_N is sufficiently small), then bounds on a more intricate coefficient $\beta_N(n)$ can be reduced to bounds on $\beta(n)$.

In order to bound the function $\beta(n)$, a Markov approximation can be used again:

$$\beta(n) = \sum_{i,j} |\mu(T^{-n}A_j \cap A_i) - \mu(A_i)\mu(A_j)|$$

$$\leq \sum_{i,i_1,\ldots,i_{n-1},j} |\mu(T^{-n}A_j \cap T^{-(n-1)}A_{i_{n-1}} \cap \cdots \cap T^{-1}A_{i_1} \cap A_i) - p_{ii_1\ldots i_{n-1}j}|$$

$$+ \sum_{i,j} |p_i \pi_{ij}^{(n)} - p_i p_j| \leq (n-1)\nu_N + \sum_{i,j} |\pi_{ij}^{(n)} - p_j| p_i$$

for any $n < N$, based on (55). Here $\pi_{ij}^{(n)}$ stands for the n-step transition probability from A_i to A_j. Denote $\hat{\beta}(n) = \sum |\pi_{ij}^{(n)} - p_j| p_i$.

Corollary 5.2 *$\beta_N(n) \le (2N + 3n)\nu_N + \hat{\beta}(n)$ for any $n < N$.*

The value of $\hat{\beta}(n)$ measures the convergence of the transition probabilities $\pi_{ij}^{(n)}$, as $n \to \infty$, to the equilibrium distribution $\{p_j\}$, and so it is $\hat{\beta}(n)$ who is responsible for mixing rates in the dynamical system.

If the dynamical system (T, M, μ) is ergodic and mixing, then any approximating Markov chain is irreducible and aperiodic. If the latter is finite, then the difference $\pi_{ij}^{(n)} - p_j$ converges to zero exponentially fast in n for any pair i, j. However, in applications the partition $\mathcal{A}$ itself depends on n, through the way of refinements (17) and (18), or by direct construction. So, the above observation is not very helpful.

There are certain sufficient conditions that provide effective bounds on $\hat{\beta}(n)$ and work both for finite and countable Markov chains. Those are motivated by classical Doeblin's condition in the theory of Markov chains [16] and by Dobrushin's coefficient of ergodicity for countable Markov chains [15].

For any pair of states i, j and $k \ge 1$ we denote

$$V_{ij}(k) = \frac{1}{2} \sum_l |\pi_{il}^{(k)} - \pi_{jl}^{(k)}| \tag{59}$$

the distance in variation between two conditional distributions, $\{\pi_{il}^{(k)}\}$ and $\{\pi_{jl}^{(k)}\}$. For any $k \ge 1$ and a real $s > 0$ we put

$$q_1(k, s) = \sum_{i,j:\ V_{ij}(k) > 1-s} p_i p_j \tag{60}$$

We also denote

$$q_2(k, s) = \inf \Big\{ q > 0 : \sum_{i:\ \sum_{j:\ V_{ij}(k) > 1-s} p_j < q} p_i > 1 - q \Big\} \tag{61}$$

The classical Doeblin condition in a modified form, due to L. Bunimovich and Ya. Sinai [7, 10], says that $V_{ij}(k) < 1 - s$ for some $s > 0$ and $k \ge 1$ for all the pairs i, j (so that both quantities $q_1(k, s)$ and $q_2(k, s)$ vanish). Our functions $q_1(k, s)$ and $q_2(k, s)$ measure the 'total probability' of pairs i, j that violate Doeblin's condition.

Remark. $\frac{1}{2} q_1(k, s) \le q_2(k, s) \le \sqrt{q_1(k, s)}$.

The following proposition provides a bound on the quantity $\hat{\beta}(n)$ in terms of k, s and $q_2(k, s)$:

Proposition 5.3 *For any $n > k \ge 1$ and $s > 0$ we have*

$$\hat{\beta}(n) \le 2(1 - s/2)^{[n/k]} + 2[n/k] \cdot q_2(k, s)$$

Proof. Since the function $\hat{\beta}(n)$ is monotone decreasing in n, it is enough to prove the proposition for all $n = kr$, where $r \geq 1$ is an integer. The proof goes by induction in r.

We will denote Σ_j^+ and Σ_j^- the summations over such j that the difference $\pi_{ij}^{(kr+k)} - p_j$ is positive and negative, respectively. In the same manner, we will denote Σ_m^+ and Σ_m^- the summations over such m that the difference $\pi_{im}^{(kr)} - p_m$ is positive and negative, respectively. Then we have

$$\hat{\beta}(rk+k) = 2\Sigma_i p_i \Sigma_j^+ (\pi_{ij}^{(kr+k)} - p_j) =$$

$$2\Sigma_i p_i \Sigma_j^+ \left(\Sigma_m (\pi_{im}^{(rk)} - p_m)\pi_{mj}^{(k)}\right) \leq 2\Sigma_i p_i \Sigma_j^+ \left(\Sigma_m^+ (\pi_{im}^{(rk)} - p_m)\pi_{mj}^{(k)}\right)$$

$$\leq 2\Sigma_i p_i \left(\Sigma_m^+ (\pi_{im}^{(rk)} - p_m)\Sigma_j^+ \pi_{mj}^{(k)}\right) \leq \left(1 - \frac{s}{2}\right)\hat{\beta}(rk) + 2 \times \sum_{m:\, \Sigma_j^+ \pi_{mj}^{(k)} > 1-s/2} p_m$$

In a similar fashion we have

$$\hat{\beta}(rk+k) \leq 2\Sigma_i p_i \left(\Sigma_m^- (p_m - \pi_{im}^{(rk)})\Sigma_j^- \pi_{mj}^{(k)}\right)$$

$$\leq \left(1 - \frac{s}{2}\right)\hat{\beta}(rk) + 2 \times \sum_{m:\, \Sigma_j^- \pi_{mj}^{(k)} > 1-s/2} p_m$$

Observe that for any m_1, m_2

$$\Sigma_j^+ \pi_{m_1 j}^{(k)} + \Sigma_j^- \pi_{m_2 j}^{(k)} = 1 + \Sigma_j^- (\pi_{m_2 j}^{(k)} - \pi_{m_1 j}^{(k)}) \leq 1 + V_{m_1 m_2}(k)$$

Therefore, if $\Sigma_j^+ \pi_{m_1 j}^{(k)} > 1 - s/2$ and $\Sigma_j^- \pi_{m_2 j}^{(k)} > 1 - s/2$ for some m_1 and m_2, then $V_{m_1 m_2}(k) > 1 - s$. Due to (61) we conclude that either

$$\sum_{m:\, \Sigma_j^+ \pi_{mj}^{(k)} > 1-s/2} p_m \leq q_2(k, s)$$

or

$$\sum_{m:\, \Sigma_j^- \pi_{mj}^{(k)} > 1-s/2} p_m \leq q_2(k, s)$$

Thus, the proof of Proposition 5.3 is completed.

In applications, it is more convenient to bound the quantity

$$V_{ij}^{\mu}(k) := \frac{1}{2} \sum_l |\mu(T^{-k} A_l / A_i) - \mu(T^{-k} A_l / A_j)| \tag{62}$$

rather than $V_{ij}(k)$, since $V_{ij}^{\mu}(k)$ is defined in terms of the invariant measure μ. In other words, we assume the smallness of

$$q_1^{\mu}(k, s) := \sum_{i,j:\, V_{ij}^{\mu}(k) > 1-s} p_i p_j \tag{63}$$

Lemma 5.4 *For any* $r \in (0, s)$ *we have*

$$q_1(k,r) \leq q_1^{\mu}(k,s) + (s-r)^{-1} k\nu_N$$

Proof. Denote Σ^*_{ij} the summation over such i, j that $V^{\mu}_{ij}(k) < 1-s$ and $V_{ij}(k) > 1-r$. Then it is a straightforward calculation based on (55) that

$$(s-r)\Sigma^*_{ij} p_i p_j \leq \Sigma^*_{ij}\left(V_{ij}(k) - V^{\mu}_{ij}(k)\right) p_i p_j \leq k\nu_N$$

The lemma is proven.

Corollary 5.5 *For any* $r \in (0, s)$ *we have*

$$\hat{\beta}(n) \leq 2(1-r/2)^{[n/k]} + 2[n/k] \cdot \left(q_1^{\mu}(k,s) + (s-r)^{-1} k\nu_N\right)^{1/2}$$

Based on Corollary 5.5 one can exploit the following strategy for bounding the function $\hat{\beta}(n)$: to bound the quantity $V^{\mu}_{ij}(k)$ away from unity for an "overwhelming majority" of the pairs i, j, and also bound the measure of the remaining pairs i, j in (63). That strategy has been successfully employed in [1] for two-dimensional hyperbolic attractors, see Section 7 for more detail.

We now outline another approach to bounding the function $\hat{\beta}(n)$. For any $k \geq 1$ and $t > 0$ denote

$$q_3(k,t) = \sum_{i,j:\, \pi^{(k)}_{ij} < tp_j} p_i p_j \tag{64}$$

If the quantity $q_3(k,t)$ is small enough, one can say that the conditional distribution $\pi^{(k)}_{ij}$ recovers 'a fraction' t of the equilibrium one.

Lemma 5.6 *For any* $s \in (0, t)$ *we have*

$$q_1(k,s) \leq \frac{4t}{t-s} \cdot q_3(k,t)$$

The key point in the proof is the observation that if

$$\sum_{l:\, \pi^{(k)}_{il} < tp_l} p_l < \frac{1}{2}\left(1 - \frac{s}{t}\right) \quad \text{and} \quad \sum_{l:\, \pi^{(k)}_{jl} < tp_l} p_l < \frac{1}{2}\left(1 - \frac{s}{t}\right)$$

then $V_{ij}(k) < 1-s$. We leave the details to the reader.

Corollary 5.7 *For any* $n > k \geq 1$ *and* $t > s > 0$ *we have*

$$\hat{\beta}(n) \leq 2(1-s/2)^{[n/k]} + 4[n/k] \cdot \sqrt{t(t-s)^{-1} q_3(k,t)}$$

Again, in applications it is more natural to seek a bound for the quantity

$$q_3^\mu(k,u) := \sum_{i,j:\, \mu(T^{-k}A_j/A_i)<u\mu(A_j)} \mu(A_i)\mu(A_j) \tag{65}$$

rather than for $q_3(k,t)$.

Lemma 5.8 *For any $t \in (0,u)$ we have*

$$q_3(k,t) < q_3^\mu(k,u) + (u-t)^{-1}k\nu_N$$

Proof. Denote Σ_{ij}^* the summation over such i,j that $\mu(T^{-k}A_j/A_i) \geq u\mu(A_j)$ and $\pi_{ij}(k) < tp_j$. Then it is a straightforward calculation based on (55) that

$$(u-t)\Sigma_{ij}^* p_i p_j \leq \Sigma_{ij}^* \left(\mu(T^{-k}A_j/A_i) - \pi_{ij}^{(k)}\right) p_i \leq k\nu_N$$

The lemma is proven.

Corollary 5.9 *For any $n > k \geq 1$ and $u > t > s > 0$ we have*

$$\hat{\beta}(n) \leq 2(1-s/2)^{[n/k]} + 4[n/k]\cdot\left(t(t-s)^{-1}(q_3^\mu(k,u) + (u-t)^{-1}k\nu_N)\right)^{1/2}$$

This corollary provides an alternative strategy for bounding $\hat{\beta}(n)$, based on proving the inequality $\mu(T^{-k}A_j/A_i) \geq u\mu(A_j)$ for some constant $u > 0$ and an "overwhelming majority" of pairs i,j. Such a strategy has been successfully implemented for chaotic billiards in [10, 12], see Section 7 for more detail. The implementation of either of the above two strategies for any new class of dynamical systems is beyond the scope of this article.

6 Applications to expanding interval maps

Here we apply our results to expanding maps of an interval.

6.1 Uniformly expanding maps. Let $M = [0,1]$ and $\mathcal{B} = \{I_1, \ldots, I_n, \ldots\}$ be a finite or countable partition of $[0,1]$ into disjoint open subintervals such that the Lebesgue measure of $M \setminus (\cup I_i)$ is zero. Let $T : [0,1] \to [0,1]$ be a map which is C^1-smooth and monotonic on each interval $I_i, i \geq 1$. Two assumptions are imposed on T:

(i) $\inf\{|T'x|,\, x \in \cup I_i\} = \Lambda > 1$, i.e. T is a (uniformly) expanding map;

(ii) either T is C^2 smooth on each $I_i, i \geq 1$, or the function $g(x) = 1/|T'(x)|$ is of bounded variation on $[0,1]$.

Ergodic and statistical properties of such interval maps (in the case of finite partitions $\mathcal{B}_0$) have been studied by Hofbauer and Keller [19]. Rychlik [40] has extended that theory to countable partitions $\mathcal{B}_0$. Their theory was later extended to interval maps with nonpositive Schwarzian by Ziemian [50] and to certain quadratic maps of the interval by Young [49], Keller and Nowicki [26], but we do not go that far.

It has been shown in [19, 40] that the map T satisfying (i) and (ii) has an absolutely continuous invariant measure μ whose density $h(x)$ is of bounded variation on $[0,1]$.

Generally, T need not be ergodic. The interval $[0,1]$ can be decomposed into a finite number of ergodic components [19, 40]. On every ergodic component T need not be mixing. In turn, every component can be decomposed into a finite number of "mixing subcomponents", which are permuted cyclically by T, and on each of which an appropriate iterate of T is an exact endomorphism. Such a decomposition of $[0,1]$ is obtained and described [19, 40] by spectral properties of the adjoint operator in $L_1(M)$, and so it is commonly referred to as the spectral decomposition.

From now on we assume that T is weakly mixing. It has been shown in [19, 40] that T is then mixing, a K-system and Bernoulli. Furthermore, for functions of bounded p-variation on $[0,1]$ with $p \geq 1$ (see below) the correlations decay exponentially fast and the central limit theorem along with its invariance principle holds. A function $F(x)$ on $[0,1]$ is said to be of bounded p-variation if

$$\sup \sum_{i=1}^{n} |F(x_i) - F(x_{i-1})|^p < \infty$$

where the supremum is taken over all finite subsets $a_0 < a_1 < \cdots < a_n$ of $[0,1]$. Any Hölder continuous function on $[0,1]$ with exponent $1/p$ is of bounded p-variation.

Lemma 6.1 *Any function $F(x)$ on $[0,1]$ of bounded p-variation, $p \geq 1$, is bounded and satisfies $\mathcal{H}_F(d) \leq \text{const} \cdot d^a$ with $a = \min\{1/2, 1/p\}$.*

Proof. Boundedness of F is obvious. If the lemma holds for an $F(x)$, it will hold for any $cF(x) + d$ with $c, d \in \mathbb{R}$. Therefore, we can assume that $0 < F(x) < 1$ for all $x \in [0,1]$. In computing the function $\mathcal{H}_F(d)$ the integration can be done with respect to the Lebesgue measure m rather than the invariant measure μ, because the density $h(x)$ of the latter is bounded. Denote $d = \operatorname{diam} \mathcal{A}$ and for any subset $B \subset [0,1]$ denote $\operatorname{osc}(F, B) = \sup_B F(x) - \inf_B F(x)$. Then we have

$$\langle (\Delta_{\mathcal{A}} F)^2 \rangle \leq \sum_i \operatorname{osc}^2(F, A_i) m(A_i)$$

$$\leq 4d \sum_{j=0}^{[(4d)^{-1}]} \operatorname{osc}^2(F, [4jd, (4j+4)d]) + 4d \sum_{j=0}^{[(4d)^{-1}]} \operatorname{osc}^2(F, [(4j+2)d, (4j+6)d])$$

If $p \leq 2$, we get a bound $\langle (\Delta_{\mathcal{A}} F)^2 \rangle \leq \text{const} \cdot d$. If $p > 2$, we apply Hölder's inequality and get $\langle (\Delta_{\mathcal{A}} F)^2 \rangle \leq \text{const} \cdot d^{2/p}$. Lemma 6.1 is proven.

Consider partitions $\mathcal{B}_{k,n} = T^{-k}\mathcal{B}_0 \vee \cdots \vee T^{-n}\mathcal{B}_0$ for all $n \geq k \geq 0$, and denote

$$\beta_*^{(\mathcal{B})}(n) = \sup_{k,l \geq 0} \beta(\mathcal{B}_{0,k}, \mathcal{B}_{k+n,k+n+l})$$

Then we have two key estimates from [19] and [40]:

$$\operatorname{diam} \mathcal{B}_{0,n} \leq \Lambda^{-n}$$

(see Lemma 2 in [19] and Lemma 3 in [40]) and

$$\beta_*^{(\mathcal{B})}(n) \leq K r^n$$

for certain $K > 0$ and $r \in (0, 1)$ (see Theorem 4 in [19] and Theorem 5 in [40]).

These estimates show that the expanding interval maps are, in our terminology, dynamical systems with exponential mixing rates. Corollary 1.7 immediately applies to these systems. It covers phase functions of bounded p-variation and far larger classes of functions as well. For instance, it covers continuous functions with modulus of continuity $O(1/|\ln d|^{2+\varepsilon})$, $\varepsilon > 0$. It also covers piecewise smooth functions with a finite number of cusps of type $(x - a)^{-b}$ with $b < 1/2$.

The character of mixing in these systems is, however, stronger than the one assumed in our Corollary 1.7. We refer the reader to [14, 23] for more detail. This strong mixing allowed Keller [25] to apply Gordin's results in the theory of weakly dependent stationary processes and to prove the CLT for any function $F \in L_2([0, 1])$ such that there are functions F_n of bounded variation satisfying

$$\sum_{n \geq 1} \langle |F - F_n|^2 \rangle^{1/2} < \infty \quad \text{and}$$

$$\lim_{n \to \infty} \operatorname{var}(F_n)(1 + h)^{-n} = 0 \quad \text{for all } h > 0$$

This class is larger than the one covered by Corollary 1.7. It includes, in particular, piecewise smooth functions with a finite number of cusps of type $(x-a)^{-1/2}|\ln(x-a)|^{-p}$ with $p > 3/2$.

Remark. The CLT and WIP hold for ergodic, but not necessarily mixing expanding transformations T. This follows from the spectral decomposition and our Propositions 3.6 and 4.4.

In the case of nonergodic expanding interval maps H. Ishitani [22, 23] has proven that the limit distribution of $S_N/||S_N||_2$, as $N \to \infty$, is a mixture of normal distributions.

All the results obtained in [19, 40, 25] and here will work if, instead of a one-step expansive condition (i), one assumes that

$$\inf |(T^m)' x| = \Lambda > 1 \tag{66}$$

for some finite iterate T^m of the map T. For example, the famous Gauss transformation $Tx = \{1/x\}$ on $(0, 1)$, where $\{\cdot\}$ stands for the fractional part of a real number, has an absolutely continuous invariant measure with density $h(x) = (\ln 2)^{-1}(1 + x)^{-1}$ and satisfies (66) for $m = 2$ but not for $m = 1$.

6.2 Nonuniformly expanding maps. The interval maps described by (66) are still uniformly expanding 'eventually'. Further relaxation of the assumption (i) inevitably leads to maps $T(x)$ such that

$$\inf |(T^m)' x| = 1 \tag{67}$$

for all $m \geq 1$ (such maps are said to be almost expanding). In particular, the map T or one of its iterates, $T^r, r \geq 1$, may have a so called indifferent fixed point, $T^r x_0 = x_0, |(T^r)' x_0| = 1$. In order to ensure the existence of a finite absolutely continuous invariant measure for such a map, one has to discard the C^2 smoothness of T at indifferent fixed points. Some examples of such maps have been studied in the literature [28, 37, 48], and they have proven to enjoy relatively rich statistical properties.

We consider one example of this type, introduced by P. Gaspard and X.-J. Wang [18, 48]. It does not pretend to be a typical one, nor has it any physical applications, it will only illustrate our theorems.

Let T be a piecewise linear map on $[0,1]$ defined by

$$T(x) = \begin{cases} \frac{\xi_{k-2}-\xi_{k-1}}{\xi_{k-1}-\xi_k}(x-\xi_k)+\xi_{k-1}, & \text{if} \quad \xi_k \leq x < \xi_{k-1} \\ \frac{x-\xi_0}{1-\xi_0} & \text{if} \quad \xi_0 \leq x \leq 1 \end{cases}$$

with $\xi_k = \xi_0/(1+k)^\alpha, k = 1, 2, \dots$ and $\xi_0 < 1$. This map is linear on each segment $\Delta_k = [\xi_k, \xi_{k-1}], k \geq 1$ and on $\Delta_0 = [\xi_0, 1]$. The point $x = 0$ is fixed, and the one-sided derivative of T at $x = 0$ is equal to one. The function $T'(x)$ is only Hölder continuous at $x = 0$ with an exponent $1/\alpha$. For any $\alpha > 1$ the map T has a finite invariant measure μ with a density that is constant on each Δ_k and takes a value

$$\rho(x) = \rho_k = \frac{(1-\xi_0)\rho_0}{1-(k/(k+1))^\alpha}$$

for $x \in \Delta_k, k \geq 1$, where ρ_0 is the value of the density on Δ_0. The density $\rho(x)$ approaches infinity as $x \to 0$. The following asymptotic formulas help to study the map T:

$$|\Delta_k| \sim \frac{\alpha\xi_0}{k^{\alpha+1}}, \quad \rho_k \sim \frac{(1-\xi_0)\rho_0 k}{\alpha}, \quad \mu(\Delta_k) \sim \frac{\xi_0(1-\xi_0)\rho_0}{k^\alpha}$$

as $k \to \infty$. The map T is ergodic and mixing.

The partition $\mathcal{B} = \{\Delta_k\}, k \geq 0$, of the unit interval is a Markov partition for the map T, because $T(\Delta_k) = \Delta_{k-1}$ for every $k \geq 1$ and $T(\Delta_0) = [0,1]$. Since T is linear on each atom of $\mathcal{B}$, the measure μ is Markov as well. In particular, the Markov approximation generated by $\mathcal{B}$ (as defined in Section 5) is perfect, i.e., $\nu_N = 0$ in (53) for all $N \geq 1$.

Statistical properties of the map T have been studied by Wang [48] and Lambert et al. in [28]. It has been proven that for all $\alpha > 3$ the correlations for any Hölder continuous functions F, G on $[0,1]$ decay by a power law. Also, for Hölder continuous functions on $[0,1]$ a central limit theorem has been established provided $\alpha > 30$. This is one of a few known examples where a CLT is proven for a system with an algebraic decay of correlations.

Let $n \geq 1$, $n_1 = [n^\nu]$ and $n_2 = [n^z]$ with some $0 < z < \nu < 1$ specified below (we use here the notations of [28] for the reader's easy reference). Let $\tilde{\mathcal{B}} = \{\Delta_0, \Delta_1, \dots, \Delta_{n_2}, \Delta_{-1}\}$ be a finite partition of $[0,1]$, where $\Delta_{-1} = \cup_{i>n_2}\Delta_i$. Denote $\Delta_{-1}(n) = \cup_{k=0}^{n-1} T^{-k}\Delta_{-1}$. It has been shown in [28, Lemma 3.1] that

$$\mu(\Delta_{-1}(n)) \le \frac{C_1}{n_2^{\alpha-1}} \cdot \max\left\{1, \frac{n}{n_2}\right\} \tag{68}$$

Now consider a partition $\mathcal{A} = \vee_{i=0}^{n_1-1} T^{-i}\tilde{\mathcal{B}}$. It has been shown in [28] (see Eq. (3.7) there) that

$$\max\{\operatorname{diam} A \,:\, A \in \mathcal{A},\, A \not\subset \Delta_{-1}(n_1)\} \le \text{const} \cdot \left(1 + \frac{\alpha+1}{n_2}\right)^{-n_1} \tag{69}$$

We now estimate the mixing coefficient $\beta(n)$ of the partition $\mathcal{A}$ based on [28, Lemma 3.8]:

$$\beta(n) \le 4\left[1 - \frac{1}{n_2}\left(1 - \frac{C_2 n}{n_2^{\alpha}}\right)^{n_2+1}\right]^{\frac{n-n_1}{n_2+1}} + \frac{C_1 n}{n_2^{\alpha}} \tag{70}$$

The last term in the RHS comes from the bound (68) on the measure of the 'residual' set $\Delta_{-1}(n)$. As in [28], we assume that $\alpha > 3$ and $z \in ((\alpha-1)^{-1}, 1/2)$, in which case the first term in the RHS of (70) decays as a stretched exponential function in n, and the second term is then the leading one. Thus, (70) reduces to

$$\beta(n) \le \frac{C_3 n}{n^{\alpha z}} \tag{71}$$

where the constant $C_3 > 0$ depends on the choice of z.

We now estimate the mixing coefficient $\beta_N(n)$ of the partition $\mathcal{A}$. This partition is not Markov, but the Markovness of $\mathcal{B}$ allows us to compute $\beta_N(n)$ in the same way as we have computed $\beta(n)$. We skip the calculation – the result is almost the same, we only have to replace $C_1 n$ by $C_1 N$ in the last term in the RHS of (70), and for any $\alpha > 3$ we get

$$\beta_N(n) \le \frac{C_3 N}{n^{\alpha z}} \tag{72}$$

Since the last term in (70) comes from the bound (68) on the measure of the 'residual' set $\Delta_{-1}(n)$, one can partition this set into arbitrarily small fragments, and the estimates (70)-(72) will survive. In other words, if $\Delta_{-1} = \cup_{j\ge 1}\Delta_{-1}^{(j)}$ for arbitrarily small disjoint subsets $\Delta_{-1}^{(j)}$, then one can redefine $\tilde{\mathcal{B}}$ to be $\{\Delta_0, \Delta_1, \ldots, \Delta_{n_2}, \Delta_{-1}^{(1)}, \Delta_{-1}^{(2)}, \ldots\}$ and again set $\mathcal{A} = \vee_{i=0}^{n_1-1} T^{-i}\tilde{\mathcal{B}}$, and still have the bounds (70)-(72). A refinement of Δ_{-1} is only necessary for a better control on the value of $\operatorname{diam}\mathcal{A}$, since the bound (69) does not cover the atoms of $\mathcal{A}$ within the residual set $\Delta_{-1}(n_1)$. After an appropriate refinement of Δ_{-1} we convert (69) into

$$\operatorname{diam}\mathcal{A} \le \text{const} \cdot \left(1 + \frac{\alpha+1}{n_2}\right)^{-n_1} \le \text{const} \cdot e^{-(\alpha+1)n^{\nu-z}} \tag{73}$$

We are now ready to apply our Theorems 1.3 and 1.4.

Theorem 6.2 *Let $F \in L_2([0,1])$ and $\mathcal{H}_F(d) \leq \mathrm{const}/|\ln d|^p$ with a sufficiently large $p > 0$. Assume that either of the following two conditions holds:*
(i) *F is essentially bounded and $\alpha > 6$;*
(ii) *$F \in L_{2+\delta}([0,1])$ with some $\delta > 0$ and $\alpha > (8 + 6\delta)/\delta$.*
Then both CLT and WIP hold provided $\sigma_F \neq 0$.

The proof consists of a simple inspection of necessary relations between the involved parameters, and we leave it to the reader. One only has to choose z in (72) sufficiently close to $1/2$.

The central limit theorem in the clause (i) recovers Theorem 2 from [28] and extends to larger classes of functions. In addition, it covers a larger region in the parameter space ($\alpha > 6$ instead of $\alpha > 30$) and partially answers a conjecture stated in [28] that the CLT holds for Hölder continous functions and any $\alpha > 2$. The invariance principle established here is a novel result. [2]

7 Applications to hyperbolic maps

Here we apply our theorems to hyperbolic maps: Anosov and Axiom A diffeomorphisms, chaotic billiards and hyperbolic attractors. Technically, these maps differ from the interval maps by their invertibility and hyperbolicity (instead of expandingness).

7.1 Anosov maps and Smale's Axiom A diffeomorphisms. These are two basic types of smooth invertible chaotic dynamical systems. An extensive study of those has been done in an excellent monograph by R. Bowen [5], and we will follow his notations here. For the reader's convenience, we provide necessary definitions.

Let $T : M \to M$ be a diffeomorphism of a compact C^∞ manifold M. A closed subset $\Lambda \subset M$ is said to be hyperbolic if $T\Lambda = \Lambda$ and for any $x \in \Lambda$ the tangent space $\mathcal{T}_x M$ is a direct sum $\mathcal{T}_x M = E_x^u \oplus E_x^s$, so that
(i) $DT(E_x^{u,s}) = E_{Tx}^{u,s}$ (invariance under DT);
(ii) there are constants $c > 0$ and $\lambda \in (0,1)$ such that

$$||DT^n v|| \leq c\lambda^n ||v|| \quad \text{for all} \quad v \in E_x^s,\ n \geq 0$$

$$||DT^{-n} v|| \leq c\lambda^n ||v|| \quad \text{for all} \quad v \in E_x^u,\ n \geq 0$$

(iii) E_x^u and E_x^s depend on x continuously.

A point $x \in M$ is said to be nonwandering if $U \cap (\cup_{n>0} T^n U) \neq \emptyset$ for any neighborhood U of x. The set $\Omega = \Omega(T)$ of all the nonwandering points is closed and T-invariant.

A diffeomorphism T is said to satisfy Axiom A [45] if $\Omega(T)$ is a hyperbolic set and periodic points are dense in $\Omega(T)$.

If the entire manifold M is a hyperbolic set, then T is called an Anosov diffeomorphism, see [5, 2]. In that case T always satisfies Axiom A.

[2] After this manuscript had been submitted to the journal, the author learned that the CLT for functions of bounded variations and any $\alpha > 3/2$ was proven in [31].

For any point $x \in \Lambda$ there is a local stable manifold $W^s(x) \subset M$ and a local unstable manifold $W^u(x) \subset M$ such that $\mathcal{T}_x W^{u,s}(x) = E_x^{u,s}$ and $\operatorname{diam} T^n W^s(x)$ $(\operatorname{diam} T^{-n} W^u(x))$ approaches zero exponentially fast in n as $n \to \infty$. Analytically, $W^{u,s}(x)$ are C^r disks provided T is a C^r diffeomorphism. The manifolds $W^{u,s}$ allow one to define local coordinates in $\Omega(T)$: there is a $\delta > 0$ such that if $\operatorname{dist}(x, y) < \delta$ for some $x, y \in \Omega(T)$, then the intersection $W^s(x) \cap W^u(y)$ consists of one point denoted by $[x, y]$, and $[x, y] \in \Omega(T)$. This describes the local structure of $\Omega(T)$.

The first (global) property of $\Omega(T)$ is a spectral decomposition, see [5, Theorem 3.5]: $\Omega(T) = \Omega_1 \cup \cdots \cup \Omega_r$, where $\{\Omega_i\}$ are disjoint closed T-invariant sets (called basic sets) such that $T|_{\Omega_i}$ is topologically transitive for every i. Furthermore, $\Omega_i = \Omega_{i,1} \cup \cdots \cup \Omega_{i,r_i}$, where $\Omega_{i,j}$ are disjoint closed subsets such that $T(\Omega_{i,j}) = \Omega_{i,j+1}$ (and $T(\Omega_{i,r_i}) = \Omega_{i,1}$) and $T^{r_i}|_{\Omega_{i,1}}$ is topologically mixing.

From now on we work on an arbitrary basic set Ω_s. A subset $R \subset \Omega_s$ is called a rectangle if $[x, y] \in R$ for all $x, y \in R$. A rectangle R is said to be proper if it is closed and $R = \overline{\operatorname{int} R}$ (where $\operatorname{int} R$ stands for the interior of R in Ω_s). For $x \in R$ we put $W^{u,s}(x, R) = W^{u,s}(x) \cap R$.

A Markov partition of Ω_s is, by definition, a finite covering $\mathcal{R} = \{R_1, \ldots, R_m\}$ of Ω_s by proper rectangles such that
(i) $\operatorname{int} R_i \cap \operatorname{int} R_j = \emptyset$ for $i \neq j$;
(ii) $TW^u(x, R_i) \supset W^u(Tx, R_j)$ and $TW^s(x, R_i) \subset W^s(Tx, R_j)$ if $x \in \operatorname{int} R_i$ and $Tx \in \operatorname{int} R_j$.

Bowen [5, Theorem 3.12] has proven that for any basic set Ω_s there are Markov partitions with rectangles of arbitrary small diameter.

Notice that no invariant measure for T has been defined or introduced so far. So, one has to be constructed. Bowen did that by using symbolic dynamics. Let $\mathcal{R} = \{R_1, \ldots, R_m\}$ be a Markov partition of Ω_s. Define a transition matrix $A = A(\mathcal{R})$ by

$$A_{ij} = \begin{cases} 1 & \text{if } \operatorname{int} R_i \cap T^{-1} \operatorname{int} R_j \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{74}$$

We then invoke the symbolic dynamics defined in Section 5, in a slightly modified form. Consider $\tilde{\Omega}_s \subset \Omega_s$, a T-invariant set of points such that $T^n x \in \operatorname{int} R_{\omega_n(x)}$ for all $n \in \mathbb{Z}$ (the set $\tilde{\Omega}_s$ is residual in topological sense). To any point $x \in \tilde{\Omega}_s$ we assign a symbolic sequence $\omega(x) = \{\omega_n(x)\}_{-\infty}^{\infty}$ as we did in Section 5. The closure of the space of all the sequences $\{\omega(x) : x \in \tilde{\Omega}_s\}$ is the following set:

$$\Sigma_A = \{\omega \in \{1, \ldots, m\}^{\mathbb{Z}} : A_{\omega_i \omega_{i+1}} = 1 \text{ for all } -\infty < i < \infty\}$$

Therefore, a symbolic representation of the map T on $\tilde{\Omega}_s$ is a subshift of finite type, also called topological Markov chain, (Σ_A, σ_L) with the transition matrix A and the left shift σ_L. The shift σ_L is topologically transitive, as is the map T on $\tilde{\Omega}_s$, hence the matrix A is irreducible. It is convenient to fix a metric on Σ_A such that $\operatorname{dist}(\omega', \omega'') = d^n$, where n is the maximal nonnegative integer such that $\omega'_i = \omega''_i$ for all $|i| < n$, and $d \in (0, 1)$ is a fixed parameter.

From now on we assume that the map T on Ω_s is topologically mixing. Then the subshift (Σ_A, σ_L) is topologically mixing, too [5, Theorem 3.19], i.e. A^M contains no zeroes for some $M > 0$. Due to Bowen [5, Theorem 1.4], for any Hölder continuous function φ^* on Σ_A there is a (unique) Gibbs σ_L-invariant measure $\mu^* = \mu^*_{\varphi^*}$ on Σ_A, see [5] for definitions and basic properties of Gibbs measures. Any Hölder continuous function φ on Ω_s lifted up to Σ_A yields a Hölder continuous function φ^* on Σ_A [5, Lemma 4.2], and so it produces a Gibbs measure μ^* on Σ_A. That Gibbs measure can be then projected down to Ω_s and will yield a T-invariant measure $\mu = \mu_\varphi$ on Ω_s, which is called an equilibrium state for φ [5]. Every equilibrium measure is ergodic, mixing and Bernoulli [5, Theorem 4.1].

Bowen has established an exponential bound on the decay of correlations [5, Theorem 1.26] and a central limit theorem [5, Theorem 1.27] for all Hölder continuous functions and Gibbs measures on Σ_A. These are automatically carried over to Hölder continuous functions and equilibrium states on Ω_s. Ruelle has also bounded correlations for C^1 functions on Axiom A attractors in a separate paper [39]. We now cite certain technical estimates from [5].

Consider partitions $\mathcal{R}_{k,n} = T^{-k}\mathcal{R} \vee \cdots \vee T^{-n}\mathcal{R}$ for all $n \geq k$, and denote

$$\beta^{(\mathcal{R})}_*(n) = \sup_{k,l\geq 0} \beta(\mathcal{R}_{0,k}, \mathcal{R}_{k+n,k+n+l})$$

Then we have two key estimates:

$$\operatorname{diam} \mathcal{R}_{-n,n} \leq \alpha^n \tag{75}$$

for a certain $\alpha \in (0,1)$ (see Lemma 4.2 in [5]), and

$$\beta^{(\mathcal{R})}_*(n) \leq A'\gamma^n \tag{76}$$

for certain $A' > 0$ and $\gamma \in (0,1)$ (see the proof of Theorem 1.25 in [5]).

These estimates show that Axiom A diffeomorphisms are, in our terminology, dynamical systems with exponential mixing rates. Corollary 1.7 immediately applies to these systems. It covers Hölder continuous functions as well as far larger classes of phase functions, for instance, those with modulus of continuity $O(1/|\ln d|^{2+\varepsilon})$, $\varepsilon > 0$.

M. Denker [14] has noticed that the label process $\{X_n\} : \Sigma \to \{1, \ldots, m\}$ defined by $X_n(\omega) = \omega_n, n \in \mathbb{Z}$, is not only β-mixing as specified by (76) but also ψ-mixing with exponentially decaying ψ-mixing coefficients. Those coefficients are defined to be $\psi(n) = \sup_{k,l\geq 0} \psi(\mathcal{R}_{0,k}, \mathcal{R}_{k+n,k+n+l})$, where for any two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ we have

$$\psi(\mathcal{A}, \mathcal{B}) = \sup_{A,B} \left| \frac{\mu(A \cap B)}{\mu(A)\mu(B)} - 1 \right|$$

the supremum being taken over all $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Bowen [5] has shown that $\psi(n) \leq \text{cost} \cdot \gamma^n$ for a $\gamma \in (0,1)$. Based on this observation, Denker [14] concluded that all the functions $F \in L_2(M)$ such that

$$\sum_{n\geq 1} ||F - \langle F|\mathcal{R}_{-n,n}\rangle||_2 < \infty \tag{77}$$

satisfy the CLT. In particular, all the functions $F \in L_2(M)$ with $\mathcal{H}_F(d) \leq \text{const}/|\ln d|^{1+\varepsilon}$, for any $\varepsilon > 0$ are covered, with no restrictions on $\mathcal{L}_F(d)$ whatsoever. This class of functions is far larger than the one covered by our Corollary 1.7.

Remark. The CLT and WIP hold in the case of ergodic, but not necessarily mixing transformation T on Ω_s. This follows from the spectral decomposition of Ω_s and our Propositions 3.6 and 4.4.

Remark. Smooth Anosov diffeomorphisms with absolutely continuous invariant measures are very popular in studies of chaotic dynamical systems. As Bowen has shown [5, Corollary 4.13], if T is a topologically mixing C^2 Anosov diffeomorphism of M with an absolutely continuous invariant measure μ, then $\mu = \mu_{\varphi^u}$ for the Hölder continuous function $\varphi^u(x) = -\log \lambda(x)$, where $\lambda(x)$ is the Jacobian of the linear map $DT : E^u_x \to E^u_{Tx}$. Thus, our Corollary 1.7 and Denker's result cover this class of systems as well.

7.2 Chaotic billiards. Apparently, the first class of chaotic dynamical systems with direct physical applications and with mathematically established statistical properties is that of billiards. However, all the papers on billiards involve so heavy and specific techniques that many researches stay away from this class of systems. We cannot go into any detailed discussion of billiards. We only sketch necessary definitions and then focus on the differences between Anosov diffeomorphisms and billiard maps.

Let Q be a compact closed domain on a plane or 2-torus with a piecewise smooth (of class C^3) boundary ∂Q. We call Q a billiard table. A point particle inside Q moves freely at unit speed and elastically bounces off the boundary ∂Q. Let

$$M = \{x = (q, v) : q \in \partial Q,\ v \text{ is a unit vector attached to } q \text{ pointing inside } Q\}$$

The transformation T on M sends each point $x = (q, v)$ along its velocity vector v to the point of the next reflection at ∂Q at which the velocity vector *after* the reflection is attached. The map T is called the billiard ball map. It preserves a finite measure μ absolutely continuous with respect to the natural uniform measure m on $M \subset \partial Q \times S^1$. The density $d\mu/dm = \cos\varphi$, where φ is the angle between the velocity vector v and the normal vector to ∂Q at q. More detailed definitions may be found in [9, 10, 17].

The billiard ball map T on M is invertible and smooth, except for some discontinuity curves. The map T is chaotic (has nonzero Lyapunov exponents a.e.) for certain classes of billiard tables: Sinai's billiards with strictly concave boundary (also called dispersing billiards) [43], generic tables with nonstrictly concave boundary (called semidispersing billiards) and certain special tables with convex components of the boundary (see [9, 10] for a more detailed discussion and references). We call such billiards chaotic.

Locally, the map T for any chaotic billiard is similar to Anosov diffeomorphisms. There is a difference, however, in their global structures: the map T has a finite or countable number of discontinuity lines in M. Local stable and unstable manifolds $W^{u,s}(x)$ exist for a.e. point $x \in M$, but unlike those for Anosov maps they may be arbitrary short, depending on x, and arbitrary short $W^{u,s}(x)$ are dense in M.

Markov partitions for chaotic billiards are defined, but in a weaker sense than for Axiom A maps. A Markov partition is a *countable* (mod 0) covering $\mathcal{R} = \{R_i\}$ of M by

rectangles (which are closed but not proper) such that
(i) $\mu(R_i \cap R_i) = 0$ for $i \neq j$;
(ii) $TW^u(x, R_i) \supset W^u(Tx, R_j)$ and $TW^s(x, R_i) \subset W^s(Tx, R_j)$ for every $x \in R_i \cap T^{-1}R_j$ provided $\mu(R_i \cap T^{-1}R_j) > 0$.

Markov partitions with rectangles of arbitrary small diameter exist [9, 27], but they are always countable. There are several reasons why those partitions are not good (at least now) to study statistical properties of billiards. First, they are constructed mostly by topological arguments and do not generate any good Markov approximations in the sense of Section 5. Second, even if they generated good approximations by Markov chains, those chains would not satisfy the Doeblin condition (cf. [16] and Section 5). Thus, in order to bound the mixing coefficients and correlations in that Markov chain, one would have to do a cut-off and lump an infinite number of states into one 'bad' state. Such a plan has been elaborated in [7], but the arguments there were complicated and not quite conclusive (see [8]).

Another strategy has been implemented in [10], wich did not require Markov partitions. A finite collection of rectangles in M with a weaker Markov property was constructed, and it generated a good approximation for T by a finite Markov chain. The measure of the residual set (the one not covered by the rectangles) was bounded separately. We called that collection a Markov sieve.

Definition 7.1 *Given two integers, $N > m > 0$, a Markov sieve $\mathcal{R}_{N,m}$ is a collection of disjoint rectangles $R_1, \dots, R_I$ in M such that*
(i) $\max_i\{\mathrm{diam} R_i\} \leq c_1\alpha_1^m$;
(ii) *for the residual set $R_0 := M \setminus \cup_{i=1}^I R_i$ one has $\mu(R_0) \leq c_2\alpha_2^m$;*
(iii) *For any $l \leq N$ and $i_0, \dots i_l \in [1, I]$ one has*

$$\mu(T^{-l}R_{i_l}/T^{-(l-1)}R_{i_{l-1}} \cap \cdots \cap R_{i_0}) = \mu(T^{-1}R_{i_l}/R_{i_{l-1}})(1+\epsilon) \tag{78}$$

with some $|\epsilon| \leq c_3\alpha_3^m$;
(iv) *there are constants $g_0, g_1 > 0$ independent of N and m such that for every $k \geq [g_0 m]$ and for a majority of pair $i, j \in [1, I]$ one has*

$$\mu(T^{-k}R_j/R_i) \geq g_1\mu(R_j) \tag{79}$$

where the "majority of pairs" means that

$$\sum_{i:\sum_{j:\,(79)\text{ holds}} \mu(B_j) > 1-q} \mu(B_i) > 1 - q \tag{80}$$

with $q = c_4 N\alpha_4^m$. The constants $c_i > 0$ and $\alpha_i \in (0, 1)$ in these estimates do not depend on N or m.

The Markov sieves have been constructed for several classes of planar chaotic billiards in [10]. Those are generic Sinai's dispersing billiards (the Lorentz gases with and without horizon included), generic semidispersing billiards, Bunimovich's stadium and other examples of Bunimovich's focusing billiard tables. In [12] Markov sieves have been constructed for multidimensional Lorentz gas with finite horizon.

Theorem 7.2 *Let (M,T,μ) be a billiard ball map for which Markov sieves exist. Then the dynamical system (M,T,μ) has stretched exponential mixing rates with $\gamma = 1/2$.*

Proof. Pick an $a \in (0,1]$ and for any $N \geq 1$ put $n_1 = [N^a] - 1$ and $m = [n_1^{1/2}]$. We define a partition $\mathcal{A}^{(N,a)}$ to be one into the rectangles $R_1, \dots, R_I$ of the Markov sieve $\mathcal{R}_{N,m}$ and some sufficiently small fragments of the residual set R_0 constructed in an arbitrary manner. Then

$$\operatorname{diam} \mathcal{A}^{(N,a)} \leq c_1 \alpha_1^m \sim c_1 \alpha_1^{N^{a/2}}$$

Based on the properties (iii) and (ii) of the Markov sieves we bound the quantity (53) for the partition $\mathcal{A} = \mathcal{A}^{(N,a)}$:

$$\nu_N \leq c_3 \alpha_3^m + N c_2 \alpha_2^m$$

The property (iv) allows us to bound the quantity (65):

$$q_3^\mu(k, g_1) \leq 2(N c_4 \alpha_4^m + c_2 \alpha_2^m)$$

for any $k \geq [g_0 m]$. Therefore, Corollary 5.9 yields a bound

$$\hat{\beta}(n) \leq 2\,(1 - g_1/8)^{[n/g_0 m]} + 4[n/g_0 m] \cdot \left(4 N c_4 \alpha_4^m + 4 g_0 g_1^{-1} m \cdot (c_3 \alpha_3^m + N c_2 \alpha_2^m)\right)^{1/2}$$

for any $n > g_0 m$. We now assume that $n \in [n_1, N-1]$. Combining the above estimates with Corollary 5.2 gives

$$\beta_N(n) \leq N^2 c_5 \alpha_5^m \sim N^2 c_5 \alpha_5^{N^{a/2}}$$

with some $c_5 > 0$ and $\alpha_5 \in (0,1)$ for any $n \in [N^a - 1, N-1]$. Theorem 7.2 is proven.

Thus, Corollary 1.8 applies to chaotic billiards. It ensures the CLT and WIP for Hölder continuous functions and for much larger classes of phase functions as well. In particular, continuous functions with modulus of continuity $O(1/|\ln d|^{4+\varepsilon}), \varepsilon > 0$, are covered. Piecewise smooth functions with a finite number of cusps of type $||x-a||^{-p}, a \in M, p < 1$, also satisfy the CLT and WIP.

We should like to note that Theorem 1.1 cannot provide an exponential bound on the correlation function even for Hölder continuous or smooth observables in the case of billiards. This is not a drawback of the method we use. The bound we get is the best one that Markov sieves can provide. A possibility of improvement of this bound is, at present, an open problem.

7.3 Hyperbolic attractors. Attractors have become a very popular type of dynamical systems during the past two decades. Recently Ya. Pesin [35] has introduced a large class of generalized hyperbolic attractors. It covers, for example, popular attractors generated by Lorenz, Lozi and Belykh maps, see [1, 35, 41]. Statistical properties of 2-dimensional generalized hyperbolic attractors have been studies by Afraimovich, Chernov and Sataev [1]. We briefly outline necessary definitions from the ref. [1].

Let M be a smooth two-dimensional manifold, $U \subset M$ an open connected subset with compact closure, $\Gamma \subset U$ a closed subset. We assume that the set $S^+ = \Gamma \cup \partial U$ consists of a finite number of compact smooth curves. Let $T : U \setminus \Gamma \to U$ be a C^2-diffeomorphism from the open set $U \setminus \Gamma$ onto its image $T(U \setminus \Gamma)$. We assume that T is twice differentiable on $U \setminus \Gamma$ up to its boundary S^+. The set S^+ is the singularity set for the map T. Denote $U^+ = \{x \in U : T^n(x) \notin S^+, n = 0, 1, 2, \ldots\}$ and $D = \cap_{n \geq 0} T^n(U^+)$. The set D is invariant for both T and T^{-1}. Its closure $\Lambda = \bar{D}$ is called the attractor for T.

A hyperbolic structure, essentially similar to that of Axiom A diffeomorphisms, is defined for the map T. Technically, it is defined in terms of families of stable and unstable cones [35, 41] rather than stable and unstable invariant subspaces $E_x^{u,s}$. There are two other technical assumptions on the singularity set Γ. First, it is supposed to be transversal to the unstable cones. Second, for any $k \geq 1$ the number of smooth components of the singularity set for the map T^k that can intersect at any point in M does not exceed $c(1+\varepsilon)^k$ for some $c > 0$ and a sufficiently small $\varepsilon > 0$.

Generally, the map T contracts Riemannian volume in M, and then it cannot preserve an absolutely continuous measure. An invariant measure for T can be constructed by a weak Cesaro limit of any absolutely continuous measure evolving under the map T. Such limits are called u-Gibbs invariant measures, and also Sinai-Bowen-Ruelle (SBR) measures. For any u-Gibbs measure almost every point of the attractor Λ has stable and unstable manifolds, called also fibers. Conditional measures on unstable fibers are absolutely continuous unlike those on stable fibers.

Ya. Pesin [35] and E. Sataev [41] have studied topological and ergodic properties of generalized hyperbolic attractors with u-Gibbs measures. Their main result is a spectral decomposition. In our context, it is a decomposition of the attractor Λ into a finite number of ergodic components. That decomposition is very similar to the ones for expanding interval maps and Axiom A diffeomorphisms, and we omit the details.

We then take an arbitrary mixing subcomponent Λ_* of the attractor Λ, the conditional measure μ_* on Λ_* and study an appropriate iterate T_* of the map T that leaves Λ_* invariant. The dynamical system (Λ_*, T_*, μ_*) is mixing. Its structure is similar, in many respects, to that of chaotic billiards. Both systems are locally hyperbolic, have stable and unstable manifolds and singularities. However, unlike billiards, attractors are essentially nonsymmetric under time reversal. In particular, conditional measures on unstable fibers are absolutely continuous, but those on stable fibers are singular.

In ref. [1], Markov sieves have been constructed for the above system (Λ_*, T_*, μ_*). The definition of Markov sieves for hyperbolic attractors differs from that for billiards in the clause (iv). One has to replace the clause (iv) in the definition of Markov sieves by the following one:
(iv') there are constants $g_0, g_1 > 0$ independent of N and m such that for every $k \geq [g_0 m]$

and for any pair $(i,j) \in [1,I]$ one has

$$\frac{1}{2}\sum_{l=0}^{I} |\mu_*(T_*^{-k}R_l/R_i) - \mu_*(T_*^{-k}R_l/R_j)| \leq 1 - g_1 \tag{81}$$

The reason of such a difference is the above nonsymmetry of the dynamics in forward and backward evolution.

We are now ready to obtain a counterpart of Theorem 7.2 in the context of attractors.

Theorem 7.3 *Let* (Λ_*, T_*, μ_*) *be an induced mixing dynamical system on a mixing subcomponents of a generalized hyperbolic attractor, for which Markov sieves exist. Then this dynamical system has stretched exponential mixing rates with* $\gamma = 1/2$.

The proof of this theorem is nearly a replica of that of Theorem 7.2. The first half of it goes word by word. In the last half we utilize the property (iv') of the Markov sieves to bound the quantity (62) by $1-g_1$, so that $q_1^{\mu_*}(k, g_1) \leq 2c_2\alpha_2^m$ for any $k \geq [g_0 m]$. After that we employ Corollary 5.5 and complete the proof of Theorem 7.3.

Therefore, Corollary 1.8 applies to the dynamical system (Λ_*, T_*, μ_*). We then obtain the CLT and WIP for the same classes of phase functions on Λ_* as in the case of chaotic billiards.

Remark. The CLT and WIP hold on every ergodic component of the attractor Λ as well. This follows from Pesin-Sataev spectral decomposition of Λ and our Propositions 3.6 and 4.4.

Acknowledgements. This work has been done during my visit at Princeton university, and I gladly acknowledge the hospitality. I should like to thank Ya. Sinai, M. Blank, D. Kosygin and S. Troubetzkoy for helpful discussions. I am grateful to the referee for pointing out some imprecisions in the manuscript.

References

[1] Afraimovich, V.S., Chernov, N.I. and Sataev, E.A.: Statistical properties of 2-D generalized hyperbolic attractors. To appear in Chaos/Xaoc **4** (1994).

[2] Anosov, D.V.: Geodesic flows on closed Riemannian manifolds with negative curvature. Proc. Steklov Inst. Math. **90** (1967).

[3] Adler, R. and Weiss, B.: Entropy a complete metric invariant for automorphisms of the torus. Proc. Nat. Acad. Sci. USA **57**, 1573–1576 (1967)

[4] Billingsley, P.: Convergence of Probability Measures. New York: Wiley, 1968.

[5] Bowen, R.: Equilibrium states and the ergodic theory of Anosov diffeomorphisms. Lect. Notes Math., **470**. Berlin: Springer-Verlag 1975.

[6] Bradley, R.: On the central limit question under absolute regularity. Ann. Probab. **13**, 1314–1325 (1985)

[7] Bunimovich, L.A. and Sinai, Ya.G.: Statistical properties of Lorentz gas with periodic configuration of scatterers. Commun. Math. Phys. **78**, 479–497 (1981)

[8] Bunimovich, L.A. and Sinai, Ya.G.: Markov partitions for dispersed billiards (Erratum). Commun. Math. Phys. **107**, 357–358 (1986)

[9] Bunimovich, L.A., Sinai, Ya.G. and Chernov, N.I.: Markov partitions for two-dimensional billiards. Russ. Math. Surv. **45**, 105–152 (1990)

[10] Bunimovich, L.A., Sinai, Ya.G. and Chernov, N.I.: Statistical properties of two-dimensional hyperbolic billiards. Russ. Math. Surv. **46**, 47–106 (1991)

[11] Burton, R. and Denker, M.: On the central limit theorem for dynamical systems. Trans. Amer. Math. Soc. **302**, 715–726 (1987)

[12] Chernov, N.I.: Statistical properties of the periodic Lorentz gas. Multidimensional case. J. Stat. Phys. **74**, 11–53 (1994)

[13] Crawford, J.D. and Cary, J.R.: Decay of correlations in a chaotic measure-preserving transformation. Physica D **6**, 223–232 (1983)

[14] Denker, M.: The central limit theorem for dynamical systems. Dyn. Syst. Ergod. Th. Banach Center Publ., **23**, Warsaw: PWN–Polish Sci. Publ., 1989.

[15] Dobrushin, R.L.: Central limit theorem for nonstationary Markov chains I. Prob. Theory Probab. Its Appl. **1**, 65–79 (1956)

[16] Doob, J.L.: Stochastic Processes. New York: Wiley, 1990.

[17] Gallavotti, G. and Ornstein, D.: Billiards and Bernoulli schemes. Commun. Math. Phys. **38**, 83–101 (1974)

[18] Gaspard, P. and Wang, X.-J.: Sporadicity: between periodic and chaotic dynamical behaviors. Proc. Nat. Acad. Sci. USA **85**, 4591–4595 (1988)

[19] Hofbauer, F. and Keller, G.: Ergodic properties of invariant measures for piecewise monotonic transformations. Math. Z. **180**, 119–140 (1982)

[20] Ibragimov, I.A.: Some limit theorems for stationary processes. Theory Probab. Its Appl. **7**, 349–382 (1962)

[21] Ibragimov, I.A. and Linnik, Y.V.: Independent and stationary sequences of random variables. Groningen: Wolters-Noordhoff, 1971.

[22] Ishitani, H.: A central limit theorem of mixed type for a class of 1-dimensional transformations. Hiroshima Math. J. **16**, 161–188 (1986)

[23] Ishitani, H.: Perron Frobenius operator and central limit theorem. In: The Study of Dynamical Systems (Kyoto, 1989), 122–139, World Sci. Adv. Ser. Dyn. Syst. **7**, Teaneck, NJ: World Sci. Publ., 1989.

[24] del Junco, A. and Rosenblatt, J.: Counterexamples in ergodic theory and number theory. Math. Ann. **245**, 185–197 (1979)

[25] Keller, G.: Un théorème de la limite centrale pour une classe de transformations monotonnes par morceaux. C. R. Acad. Sci. Paris, Ser. A **291**, 155–158 (1980)

[26] Keller, G. and Nowicki, T.: Spectral theory, zeta functions and the distribution of periodic points for Collet-Eckmann maps. Commun. Math. Phys. **149**, 31–69 (1992)

[27] Krüger, T. and Troubetzkoy, S.: Markov partitions and shadowing for nonuniformly hyperbolic systems with singularities. Ergod. Th. and Dynam. Sys. **12**, 487–508 (1992)

[28] Lambert, A., Siboni, S. and Vaienti, S.: Statistical properties of a nonuniformly hyperbolic map of the interval. J. Stat. Phys. **72**, 1305–1330 (1993)

[29] Leonov, V.P.: On the dispersion of time-dependent means of a stationary stochastic process. Theory Probab. Its Appl. **6**, 87–93 (1961)

[30] Loviscach, J.: A large class of functions with good properties under Anosov maps. Preprint BiBoS, no. 586/93, 1993.

[31] Mori, M.: On the intermittency of a piecewise linear map. Tokyo J. Math. **16**, 411–428 (1993)

[32] Nicolis, G., Martinez, S. and Tirapegui, E.: Finite coarse-graining and Chapman-Kolmogorov equation in conservative dynamical systems. Chaos, Solutions and Fractals **1**, 25–37 (1991)

[33] Oodaira, H. and Yoshihara, K.: Functional central limit theorems for strictly stationary processes satisfying the strong mixing condition. Kodai Math. Sem. Rep. **24**, 259–269 (1972)

[34] Peligrad, M.: Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (A survey). In: Dependence in Probability and Statistics, Boston: Birkhäuser, 1986, 193–223.

[35] Pesin, Ya.B.: Dynamical systems with generalized hyperbolic attractors: hyperbolic, ergodic and topological properties. Ergod. Th. Dynam. Sys. **12**, 123–151 (1992)

[36] Philipp, W.: Invariance principles for independent and weakly dependent random variables. In: Dependence in Probability and Statistics, Boston: Birkhäuser, 1986, 225–268.

[37] Prellberg, T. and Slawny, J.: Maps of intervals with indifferent fixed points: thermodynamic formalism and phase transition. J. Stat. Phys. **66**, 503–514 (1992)

[38] Ratner, M.: The central limit theorem for geodesic flows on n-dimensional manifolds of negative curvature. Israel J. Math. **16**, 181–197 (1973)

[39] Ruelle, D.: A measure associated with Axiom-A attractors. Amer. J. Math. **98**, 619–654 (1976)

[40] Rychlik, M.: Bounded variation and invariant measures. Studia Math. **LXXVI**, 69–80 (1983)

[41] Sataev, E.A.: Invariant measures for hyperbolic maps with singularities. Russ. Math. Surv. **47**:1, 191–251 (1992)

[42] Sinai, Ya.G.: Markov partitions and C-diffeomorphisms. Funct. Anal. Its Appl. **2**, 61–82 (1968)

[43] Sinai, Ya.G.: Dynamical systems with elastic reflections. Ergodic properties of scattering billiards. Russ. Math. Surveys **25**, 137–189 (1970)

[44] Sinai, Ya.G.: Gibbs measures in ergodic theory, Russ. Math. Surveys **27**, 21–69 (1972)

[45] Smale, S.: Differentiable dynamical systems. Bull. Amer. Math. Soc. **73**, 747–817 (1967)

[46] Ulam, S.: Problems in Modern Mathematics, New York, Interscience Publishers, 1960.

[47] Volný, D.: On limit theorems and category for dynamical systems. Yokohama Math. J. **38**, 29–35 (1990)

[48] Wang, X.-J.: Statistical physics of temporal intermittency. Phys. Rev. A. **40**, 6647–6661 (1989)

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

[50] Ziemian, K.: Almost sure invariance principle for some maps of an interval. Ergod. Th. and Dyn. Syst. **5**, 625–640 (1985)

Abstract

We develop a unified approach to study statistical properties of chaotic dynamical systems, namely, to bound correlation functions and to prove the central limit theorem and its (weak) invariance principle. Our methods are based on partitions of the phase space of a dynamical system, for which certain bounds on mixing coefficients are assumed. We also use partitions of the phase space to approximate the dynamics by Markov chains and employ those to bound mixing coefficients in applications. We apply our results to basic types of smooth chaotic dynamical systems: expanding interval maps (both uniform and nonuniform), Axiom A diffeomorphisms, billiards and hyperbolic attractors. We recover the existing limit theorems, obtain some new, and often extend those theorems to larger classes of functions.