Decay of Correlations, Central Limit Theorems and Approximation by Brownian Motion for Compact Lie Group Extensions *

Michael Field   Ian Melbourne †   Andrew Török ‡

Department of Mathematics
University of Houston
Houston, TX 77204-3476, USA

March 23, 2001
Revised: January 8, 2002

Abstract

Hölder continuous observations on hyperbolic basic sets satisfy strong statistical properties such as exponential decay of correlations, central limit theorems and invariance principles (approximation by Brownian motion).

Using an equivariant version of the Ruelle transfer operator studied by Parry & Pollicott, we obtain similar results for equivariant observations on compact group extensions of hyperbolic basic sets.

*The research of MF, IM and AT was supported in part by NSF Grant DMS-0071735.

†Permanent address: Department of Mathematics and Statistics, University of Surrey, Guildford, Surrey, GU2 7XH, UK

‡and Institute of Mathematics of the Romanian Academy, P.O. Box 1–764, RO-70700 Bucharest, Romania

1 Introduction

Let Λ be a topologically mixing hyperbolic basic set equipped with an equilibrium state (Gibbs measure). It is by now classical (see for example Bowen [5], Ratner [29], Ruelle [32], Parry & Pollicott [26]) that Hölder continuous observations $\phi : \Lambda \to \mathbb{R}$ satisfy strong statistical properties such as exponential decay of correlations and the central limit theorem (CLT). Furthermore, an almost sure invariance principle (ASIP), Denker & Philipp [10], states that the partial sums of ϕ approximate Brownian motion on the line.

Recently, interest has focused on (skew product) extensions $\Lambda \times G$ of a hyperbolic basic set Λ by a compact connected Lie group G. Such extensions are automatically partially hyperbolic and are typically ergodic and even stably ergodic and mixing [15, 27].

Counterexamples of Dolgopyat [11] show that even when $\Lambda \times G$ is stably ergodic, the decay of correlations of Hölder observations may be arbitrarily slow. The counterexamples are toral extensions of subshifts of finite type. On the other hand, if G is semisimple, or if Λ is Anosov, then Dolgopyat obtains rapid (though not exponential) decay of correlations when $\Lambda \times G$ is stably ergodic. The CLT then follows from Liverani [21].

Motivated by issues arising for certain noncompact (primarily Euclidean) group extensions, Nicol *et al.* [23] considered a class of G-*equivariant* n-dimensional observations $\phi : \Lambda \times G \to \mathbb{R}^n$. Under the same hypotheses as in [11], Nicol *et al.* [23] obtained an n-dimensional CLT with convergence to a normal distribution $N(0, \Sigma)$ where Σ is an $n \times n$ covariance matrix. Furthermore, Σ is typically nonsingular.

In this paper, we show that by restricting to equivariant observations from the outset, it is possible to bypass [11, 21]. Instead, we use an equivariant version of the Ruelle transfer operator studied by Parry & Pollicott [25, 26]. This leads both to cleaner proofs and stronger results. For example, the counterexamples of Dolgopyat [11] do not apply to equivariant observations, and we obtain *exponential* decay of correlations (not merely rapid decay) for equivariant Hölder observations on weak mixing extensions of arbitrary hyperbolic basic sets. Our results apply to principal extensions as well as to (skew product) extensions by compact connected Lie groups.

We also prove a version of the ASIP, where for almost every initial condition the partial sums of ϕ approximate n-dimensional Brownian motion. Immediate consequences of the ASIP are the law of the iterated logarithm (LIL), the CLT and the weak invariance principle (WIP).

The remainder of the paper is organized as follows. In Section 2, we state our main results for extensions of hyperbolic basic sets. In Section 3, we describe a general framework for proving exponential decay of correlations, the CLT, the WIP, and the upper half of the LIL for classes of observations for noninvertible dynamical systems. In Section 4, we prove exponential decay of correlations for skew and principal extensions of hyperbolic basic sets. In Section 5, we quickly review the definition and properties of the covariance matrix. In Section 6, we prove the ASIP for extensions of hyperbolic basic sets.

2 Statement of main results

Let M be a C^∞ compact manifold and suppose that $\Lambda \subset M$ is a topologically mixing hyperbolic basic set for the diffeomorphism $f : M \to M$. In the remainder of the paper, we always assume f is smooth – that is, of class at least C^1. Let d denote a choice of smooth Riemannian metric on M.

Recall that $\rho : \Lambda \to \mathbb{R}$ is *Hölder continuous* with exponent α, i.e., $\rho \in C^\alpha$, if $\alpha \in (0,1)$ and for all $x, y \in \Lambda$, we have $|f(x) - f(y)| \leq Cd(x,y)^\alpha$, for some constant $C \geq 0$. In the sequel, when we refer to Hölder continuous maps defined on subsets of manifolds, we always assume Hölder continuity is with respect to some fixed exponent α.

Let $\rho : \Lambda \to \mathbb{R}$ be Hölder continuous and μ denote the associated equilibrium state (Gibbs measure). In the sequel, we regard μ as a measure on M supported on Λ. Necessarily, $f|(\Lambda, \mu)$ is μ-mixing (and more, see [5]). Let G be a compact connected Lie group with Haar measure ν and define the product measure $m = \mu \times \nu$ on $M \times G$. In the sequel, unless otherwise stated, all integrals are over $M \times G$ (equivalently, $\Lambda \times G$) with respect to the measure m.

Let $f_h : M \times G \to M \times G$ be the compact group extension defined by $h : M \to G$. That is, $f_h(x,g) = (fx, gh(x))$, $(x,g) \in M \times G$. We make the standing assumptions that (1) all group extensions are determined by Hölder continuous skewing maps h, and that (2) $f_h : M \times G \to M \times G$ is *weak mixing* (with respect to the measure m). Note that it follows from Rudolph's theorem [31] that if f_h is weak mixing then it is mixing. However, as we only make use of properties of weak mixing in the sequel, we emphasize weak mixing rather than mixing.

Remark 2.1 Field and Parry [15] prove that weak mixing holds for an open and dense set of smooth G-extensions $f_h : M \times G \to M \times G$ provided either Λ is a hyperbolic attractor or G is semisimple (see also [14]). Even if none of these conditions holds, a residual set of G-extensions are weak mixing (this is an elementary consequence of Parry and Pollicott [27, Proposition 8]).

Suppose that $\mathbb{R}^n$ is an orthogonal representation of G. An observation $\phi : M \times G \to \mathbb{R}^n$ is *G-equivariant* if $\phi(x, ag) = a\phi(x,g)$ for all $a \in G$. Equivalently, $\phi(x,g) = gv(x)$ where $v : M \to \mathbb{R}^n$. Equivariant observations arise naturally in applications to problems with noncompact symmetry group, see [1, 23].

In the following, if b, c are vectors in $\mathbb{R}^n$, then bc^T denotes an $n \times n$ matrix in $\mathrm{Mat}_n(\mathbb{R})$. We always assume observations are Hölder continuous.

Theorem 2.2 (Exponential decay of correlations) *Let $f_h : M \times G \to M \times G$ be a weak mixing G-extension of a hyperbolic basic set.*

Then there is a constant $\tau \in (0,1)$ such that for all Hölder continuous G-equivariant observations $\phi, \psi : M \times G \to \mathbb{R}^n$ there is a constant $C > 0$ such that

$$|\textstyle\int \phi \circ f_h^j \psi^T - \int \phi \int \psi^T| \le C\tau^j, \quad \textit{for all } j \ge 1.$$

Analogous results hold for principal G-extensions.

Remark 2.3 The restriction to equivariant observations is crucial, since Dolgopyat [11] has counterexamples for general observations when Λ is a subshift of finite type and G is a torus — even when $\Lambda \times G$ is *stably* weak mixing. In fact, Dolgopyat obtains superpolynomial (but not exponential) decay of correlations for general observations provided G is semisimple or M is Anosov.

Let $\phi : M \times G \to \mathbb{R}^n$ be a G-equivariant observation, and define

$$\phi_N = \textstyle\sum_{j=0}^{N-1} \phi \circ f_h^j.$$

If ϕ has mean zero ($\int \phi = 0$), then it is an elementary consequence of Theorem 2.2 that the limit $\lim_{N\to\infty} \frac{1}{N} \int \phi_N \phi_N^T$ exists (see Section 3(b) below).

Definition 2.4 Let $\phi : M \times G \to \mathbb{R}^n$ be a G-equivariant observation with mean zero. Define its *covariance matrix* $\Sigma = \Sigma_\phi = \lim_{N\to\infty} \frac{1}{N} \int \phi_N \phi_N^T$.

Note that Σ is a G-equivariant symmetric positive semidefinite linear operator on $\mathbb{R}^n$. In fact, there are no further restrictions on Σ, and generically (for a C^0-open and C^∞-dense set of C^∞ observations ϕ) Σ is nonsingular (see Nicol *et al.* [23] and also Remark 5.4 below).

A stochastic process $W : [0,\infty) \times \Omega \to \mathbb{R}^n$ is called an n-dimensional Brownian motion if (i) $W(0) = 0$ almost surely, (ii) there is an $n \times n$ covariance matrix Σ such that $W(t)$ has distribution $N(0, t\Sigma)$ for each $t \ge 0$, and (iii) for each $0 \le t_1 < t_2 < \cdots < t_k$, the increments $W(t_1)$, $W(t_2) - W(t_1), \ldots, W(t_k) - W(t_{k-1})$ are independent random variables. When $\Sigma = I_n$, $W(t)$ is called a standard n-dimensional Brownian motion. It is a basic property of Brownian motion that $W \in C([0,\infty), \mathbb{R}^n)$ almost surely. That is, W is a *random element* with values in $C([0,\infty), \mathbb{R}^n)$. (This generalizes the notion of a random variable with values in $\mathbb{R}$.) We can now state our second main result.

Theorem 2.5 (Almost sure invariance principle (ASIP)) *Let $f_h : M \times G \to M \times G$ be a weak mixing G-extension of a hyperbolic basic set. Let $\phi : M \times G \to \mathbb{R}^n$ be a G-equivariant observation with mean zero and nonsingular covariance Σ.*

Then for each $c \in \mathbb{R}^n$, there is a one-dimensional Brownian motion W with variance $c^T\Sigma c$ and a sequence of random variables $S_c(N)$ such that the sequences $\{c^T\phi_N,\ N \geq 1\}$ and $\{S_c(N),\ N \geq 1\}$ are equal in distribution and such that for each $\delta > 0$,

$$S_c([t]) = W(t) + O(t^{1/4+\delta}) \quad \textit{as } t \to \infty,$$

almost surely.

Analogous results hold for principal G-extensions.

Remark 2.6 The error term $O(t^{1/4+\delta})$ is slightly better than the error term $(O(t^{1/2-\delta})$ that is more usual in statements of the ASIP [10, 28].

There are a number of standard consequences of the ASIP (see for example [28]). A few of these are listed below.

Corollary 2.7 (Central limit theorem (CLT)) *Suppose that f_h and ϕ are as in Theorem 2.5. Then $\frac{1}{\sqrt{N}}\phi_N$ converges in distribution to an n-dimensional normal distribution with mean zero and covariance Σ as $N \to \infty$.*

Proof Let W be an n-dimensional Brownian motion with covariance Σ. Let $c \in \mathbb{R}^n$ and note that c^TW is a one-dimensional Brownian motion with variance $c^T\Sigma c$. It follows from the ASIP that $\frac{1}{\sqrt{N}}c^T\phi_N$ converges in distribution to a one-dimensional normal distribution with mean zero and variance $c^T\Sigma c$. By the Cramer-Wold technique (see for example [4, Theorem 29.4]), this implies the required result. ∎

Set $W_N(0) = 0$, and

$$W_N(t) = \tfrac{1}{\sqrt{N}}\phi_{Nt} = \frac{1}{\sqrt{N}}\sum_{j=0}^{Nt-1} \phi \circ f_h^j, \quad t = 1/N, 2/N, \ldots$$

Linearly interpolating on each interval $[(k-1)/N, k/N]$, $k \geq 1$, we obtain a sequence of random elements $W_N \in C([0,\infty), \mathbb{R}^n)$.

Corollary 2.8 (Weak invariance principle (WIP)) *Suppose that f_h, ϕ and W are as in Theorem 2.5. Then W_N converges weakly in $C([0,\infty), \mathbb{R}^n)$ to W as $N \to \infty$. (In other words, the measures induced by W_N on $C([0,\infty), \mathbb{R}^n)$ converge weakly to an n-dimensional Wiener measure.)*

Proof The ASIP implies that $c^T W_N$ converges weakly to $c^T W$ for each $c \in \mathbb{R}^n$ and again this is enough to imply the n-dimensional result. ∎

Remark 2.9 Let $\chi : C([0,\infty),\mathbb{R}^n) \to \mathbb{R}$ be continuous. A consequence [3] of the WIP is that the sequence of random variables $\chi(W_N)$ converges weakly to $\chi(W)$. Taking χ to be evaluation at $t = 1$, we see that the WIP implies the CLT. (The WIP is often called the functional CLT.)

As usual, a version of the LIL, together with its functional form which we do not state explicitly, follows directly from the ASIP:

Corollary 2.10 (Law of the iterated logarithm (LIL)) *Suppose that f_h and ϕ are as in Theorem 2.5. For each $c \in \mathbb{R}^n$,*

$$\limsup_{N\to\infty} \frac{c^T \phi_N}{\sqrt{2N \log\log N}} = \sigma_c,$$

almost surely, where $\sigma_c^2 = c^T \Sigma c$. ∎

Remark 2.11 We have proved n-dimensional versions of the CLT and WIP, but our versions of the ASIP and LIL are restricted to one-dimensional projections $c^T \phi_N$. We conjecture that the n-dimensional versions of the ASIP and (functional) LIL are valid under the same hypotheses, but the proof will require a different approach from the one in this paper (this is already the case for n-dimensional independent random variables, see [9]).

3 General framework

Suppose that (Y, m) is a probability space and that $T : Y \to Y$ is a measure-preserving endomorphism. In this section, all integrals are over Y with respect to the measure m. Let $U : L^2(Y,\mathbb{R}^n) \to L^2(Y,\mathbb{R}^n)$ be the induced isometry ($U\phi = \phi \circ T$), with adjoint operator U^*.

Let F be a Banach space embedded in $L^2(Y,\mathbb{R}^n)$, with norm $\| \, \|$ scaled so that $|\phi|_2 \le \|\phi\|$ for all $\phi \in F$. We assume (without loss) that the constant functions lie in F. Define the closed subspace $F_0 = \{\phi \in F : \int \phi = 0\}$. We suppose that U and U^* restrict to operators on F and that the following property holds:

$$\text{There are constants } C > 0 \text{ and } \rho \in (0,1) \text{ such that} \quad (3.1)$$
$$\|(U^*)^j \phi\| \le C\rho^j \|\phi\| \text{ for all } j \ge 1 \text{ and } \phi \in F_0.$$

Remark 3.1 All the results in this section go through with ρ^j replaced by any positive function $a(j)$ such that $\sum_{j=1}^{\infty} j\,a(j) < \infty$.

The remainder of the section is divided in several subsections. Exponential decay of correlations is derived in Subsection (a) and is used to define the covariance matrix in Subsection (b). Subsection (c) consists of a standard Gordin-type argument which yields information on the nondegeneracy of the covariance matrix, and at the same time leads to a multiplicative sequence of random variables. The CLT, WIP and upper LIL are consequences of multiplicativity.

(a) Exponential Decay of Correlations

Proposition 3.2 *There exist constants $K > 0$ and $\rho \in (0,1)$ such that*

$$|\textstyle\int (U^j\phi)\psi^T - \int\phi\int\psi^T| \le K\rho^j|\phi|_2\|\psi\|,$$

for all $j \ge 1$ and all $\phi \in L^2(Y,\mathbb{R}^n)$, $\psi \in F(Y,\mathbb{R}^n)$.

Proof Note that

$$\textstyle\int\big((U^j\phi)\psi^T - \int\phi\int\psi^T\big) = \int U^j\phi(\psi - \int\psi)^T = \int\phi\,(U^*)^j(\psi-\int\psi)^T$$

so that

$$|\textstyle\int U^j\phi\,\psi^T - \int\phi\int\psi^T| \le |\phi|_2|(U^*)^j(\psi-\int\psi)|_2.$$

But

$$|(U^*)^j(\psi - \textstyle\int\psi)|_2 \le \|(U^*)^j(\psi-\int\psi)\| \le C\rho^j(\|\psi\| + |\int\psi|) \le 2C\rho^j\|\psi\|,$$

so the result follows with $K = 2C$. ∎

(b) Covariance matrices

For notational convenience, given $\phi,\psi \in L^2(Y,\mathbb{R}^n)$, we write $E(\phi,\psi) = \int\phi\psi^T \in \mathrm{Mat}_n(\mathbb{R})$. Also, we set $\phi_N = \sum_{j=0}^{N-1} U^j\phi$.

We prove below that the covariance matrix (see Definition 2.4) is well defined, so the limit $\Sigma_\phi = \lim_{N\to\infty}\frac{1}{N}\int\phi_N\phi_N^T$ exists. We conclude that the *covariance matrix* Σ_ϕ is a symmetric positive semidefinite $n\times n$ matrix.

Proposition 3.3 *If $\phi \in F_0$, then the series*

$$\Sigma_\phi = E(\phi,\phi) + \sum_{j=1}^{\infty} E(U^j\phi,\phi) + \sum_{j=1}^{\infty} E(\phi,U^j\phi) \tag{3.2}$$

converges absolutely and

$$E(\phi_N,\phi_N) = N\Sigma_\phi + O(1), \quad \textit{as } N \to \infty.$$

Proof The case $n = 1$ is standard (see for example [8]) and the proof for general n is similar. We give the details for completeness.

Absolute convergence of the series for Σ_ϕ is an immediate consequence of decay of correlations. Next compute that

$$E(\phi_N,\phi_N) = \sum_{i,j=0}^{N-1} E(U^i\phi,U^j\phi) = \Big(\sum_{i=j} + \sum_{i>j} + \sum_{i<j}\Big) E(U^i\phi,U^j\phi).$$

Since $E(U^i\phi,U^i\phi) = E(\phi,\phi)$, the first term reduces to $NE(\phi,\phi)$. The second term can be written as

$$\begin{aligned}\sum_{i>j} E(U^i\phi,U^j\phi) &= \sum_{i>j} E(U^{i-j}\phi,\phi) = \sum_{r=1}^{N-1}(N-r)E(U^r\phi,\phi) \\ &= N\sum_{r=1}^{\infty} E(U^r\phi,\phi) - \sum_{r=1}^{N-1} rE(U^r\phi,\phi) - N\sum_{r=N}^{\infty} E(U^r\phi,\phi)\end{aligned}$$

It follows from decay of correlations that

$$\Big|\sum_{r=1}^{N-1} rE(U^r\phi,\phi)\Big| \le K|\phi|_2\|\phi\|\sum_{r=1}^{\infty} r\rho^r < \infty.$$

Similarly,

$$\Big|N\sum_{r=N}^{\infty} E(U^r\phi,\phi)\Big| \le N\sum_{r=N}^{\infty} K\rho^r|\phi|_2\|\phi\| = N\rho^N K|\phi|_2\|\phi\|/(1-\rho) \to 0,$$

as $N \to \infty$. Hence $\sum_{i>j} E(U^i\phi,U^j\phi) = N\sum_{r=1}^{\infty} E(U^r\phi,\phi) + O(1)$ as $N \to \infty$. Similarly, $\sum_{i<j} E(U^i\phi,U^j\phi) = N\sum_{r=1}^{\infty} E(\phi,U^r\phi) + O(1)$ as $N \to \infty$. We conclude that $E(\phi_N,\phi_N) = NE(\phi,\phi) + N\sum_{r=1}^{\infty} E(U^r\phi,\phi) + N\sum_{r=1}^{\infty} E(\phi,U^r\phi) + O(1) = N\Sigma_\phi + O(1)$ as required. ∎

(c) Multiplicative sequences and martingales

We recall a standard argument of Gordin [17].

Lemma 3.4 *If $\phi \in F_0$, then ϕ is cohomologous to an element $\psi \in F_0$ such that $U^*\psi = 0$. That is, there exists $\psi, \chi \in F_0$ such that $U^*\psi = 0$ and*

$$\phi = \psi + U\chi - \chi.$$

Moreover, $\Sigma_\phi = \Sigma_\psi = E(\psi, \psi)$.

Proof The sequence of partial sums $\sum_{j=1}^N (U^*)^j \phi$ is a Cauchy sequence in F_0 and hence converges to $\chi = \sum_{j=1}^\infty (U^*)^j \phi \in F_0$. Note that $\chi - U^*\chi = U^*\phi$. Now define $\psi = \phi + \chi - U\chi \in F_0$ and note that $U^*\psi = U^*\phi + U^*\chi - \chi = 0$.

If $j \geq 1$, then $E(U^j\psi, \psi) = \int U^j\psi\, \psi^T = \int \psi\big((U^*)^j\psi\big)^T = 0$, and similarly $E(\psi, U^j\psi) = 0$. Hence, by definition (equation (3.2)) of Σ_ψ, we have that $\Sigma_\psi = E(\psi, \psi)$.

It remains to prove that $\Sigma_\phi = \Sigma_\psi$. Equivalently, $c^T\Sigma_\phi c = c^T\Sigma_\psi c$ for all $c \in \mathbb{R}^n$. But $c^T\phi_N = c^T\psi_N + c^T(U^N\chi - \chi)$, so taking L^2-norms,

$$|c^T\phi_N|_2 \leq |c^T\psi_N|_2 + 2|c^T\chi|_2.$$

Since $|c^T\phi_N|_2^2 = \int c^T\phi_N\phi_N^T c = c^T E(\phi_N, \phi_N)c$, we deduce that $\frac{1}{N}c^T E(\phi_N, \phi_N)c \leq \frac{1}{N}c^T E(\psi_N, \psi_N)c + o(1)$. Hence $c^T\Sigma_\phi c \leq c^T\Sigma_\psi c$ and the reverse inequality follows by the same argument. ∎

Corollary 3.5 *Let $\phi \in F_0$. Then $\Sigma_\phi = 0$ if and only if $\phi = U\chi - \chi$ for some $\chi \in F_0$ (that is, ϕ is a coboundary in F_0).*

Proof If ϕ is a coboundary, then $\phi_N = U^N\chi - \chi$ and $E(\phi_N, \phi_N) = O(1)$, so that $\Sigma_\phi = 0$. Conversely, suppose that $\Sigma_\phi = 0$ and write $\phi = \psi + U\chi - \chi$ as in the lemma. Then $0 = \Sigma_\phi = E(\psi, \psi) = \int \psi\psi^T$ so that $\psi = 0$ and $\phi = U\chi - \chi$. ∎

Remark 3.6 This argument, which directly produces a coboundary in F_0, answers positively a question raised by Bowen [5, p. 39]. The usual approach is to first construct an L^2 coboundary and to then prove regularity, see for example [26, Proposition 4.12]. (The L^2 coboundary criterion for degeneracy holds very generally, see [26, Proposition 4.12] and for a simpler argument [8, Lemma 2.4].)

Definition 3.7 A sequence of random variables $\{X_j\}$ is *multiplicative* if

$$\int X_{j_1}X_{j_2}\cdots X_{j_r} = 0 \quad \text{for all } j_1 > j_2 > \cdots > j_r \geq 0,\ r \geq 1. \tag{3.3}$$

Proposition 3.8 *Suppose that $U^*\phi = 0$ and let $c \in \mathbb{R}^n$. Define $X_j = c^T U^j \phi$. Then $\{X_j\}$ is a multiplicative sequence.*

Proof The isometries U on $L^2(Y, \mathbb{R}^n)$ and $L^2(Y, \mathbb{R})$ are related by $U(c^T\psi) = c^T U\psi$ and hence the corresponding adjoints U^* satisfy $U^*(c^T\psi) = c^T U^*\psi$. In particular, $U^*X_0 = c^T U^*\phi = 0$. Suppose that $j_1 \geq \cdots \geq j_{r-1} > j_r$. Let $k_i = j_i - j_r$ and compute that

$$\begin{aligned}\textstyle\int X_{j_1} X_{j_2} \cdots X_{j_r} &= \textstyle\int X_{k_1} \cdots X_{k_{r-1}} X_0 = \int U\{X_{k_1-1} \cdots X_{k_{r-1}-1}\} X_0 \\ &= \textstyle\int X_{k_1-1} \cdots X_{k_{r-1}-1} U^* X_0 = 0\end{aligned}$$

as required. ∎

Corollary 3.9 *Suppose that $T : Y \to Y$ is ergodic and $\phi \in F_0$. Then $\frac{1}{\sqrt{N}}\phi_N$ converges in distribution to an n-dimensional normal distribution with mean zero and covariance matrix Σ_ϕ.*

Proof By Lemma 3.4, $\phi_N = \psi_N + O(1)$ where $U^*\psi = 0$, and $\Sigma_\phi = \Sigma_\psi$. Hence, without loss of generality we may suppose that $U^*\phi = 0$. In particular, $\Sigma_\phi = \int \phi\phi^T$.

By the Cramer-Wold technique it is sufficient to prove the CLT for the sequence $c^T U^j \phi$ for each vector $c \in \mathbb{R}^n$. It follows from Proposition 3.8 that this is a multiplicative sequence, and hence we can apply results of McLeish [22] or Fukuyama [16]. ∎

Remark 3.10 Similarly, the WIP holds for multiplicative sequences by [16, 22]. Furthermore, if $F \subset L^\infty$, it follows from Takahashi [36, Theorem 2] that the upper half of the law of the iterated logarithm holds for ϕ_N. That is, for each $c \in \mathbb{R}^n$,

$$\limsup_{N\to\infty} \frac{c^T\phi_N}{\sqrt{2N \log\log N}} \leq \sigma_c,$$

almost surely, where $\sigma_c^2 = c^T \Sigma_\phi c$.

Remark 3.11 We have not assumed that Σ_ϕ is nonsingular in Corollary 3.9 or Remark 3.10.

Remark 3.12 In this section, we have used probabilistic limit theorems for multiplicative sequences rather than for martingales. However, the martingale approach proves useful in Section 6, so we recall this formulation for later reference.

Suppose that $T : Y \to Y$ is ergodic and let $\mathcal{M}$ be the σ-algebra on Y. Of course, T and ϕ are $\mathcal{M}$-measurable and $T^{-1}\mathcal{M} \subset \mathcal{M}$. Suppose that $U^*\phi = 0$. The conditional expectation $E(\cdot|T^{-1}\mathcal{M})$ is equal to the projection UU^*, from which it follows that $E(\phi|T^{-1}\mathcal{M})=0$.

It follows immediately that $T^{-j}\mathcal{M}$ is a decreasing sequence of σ-algebras, and that for all $j\geq 0$, $\phi\circ T^j$ is $T^{-j}\mathcal{M}$-measurable and $E(\phi\circ T^j|T^{-(j+1)}\mathcal{M})=0$.

The CLT and WIP can now be recovered by a standard argument, which is implicit in Gordin [17]. For completeness, we sketch the argument — although the approach using multiplicative sequences is more elementary.

First embed $T:Y\to Y$ in its natural extension $\widehat{T}:\widehat{Y}\to\widehat{Y}$ (see [30]). The observation $\phi:Y\to\mathbb{R}^n$ lifts to an observation $\widehat{\phi}:\widehat{Y}\to\mathbb{R}^n$, and the joint distributions of $\widehat{\phi}_N$ are identical to those of ϕ_N for $N\geq 1$. Let $\widehat{\mathcal{M}}$ be the σ-algebra $\mathcal{M}$ lifted to $\widehat{Y}$. Then $\widehat{T}^j\widehat{\mathcal{M}}$, $j\geq 1$, is an *increasing* filtration of σ-algebras and moreover for all $j\geq 1$, $\widehat{\phi}\circ\widehat{T}^{-j}$ is $\widehat{T}^j\widehat{\mathcal{M}}$-measurable, and $E(\widehat{\phi}\circ\widehat{T}^{-j}|\widehat{T}^{j-1}\widehat{\mathcal{M}})=0$. In other words, $\{\sum_{j=1}^N\widehat{\phi}\circ\widehat{T}^{-j},\,N\geq 1\}$ is a stationary martingale.

If $T:Y\to Y$ is ergodic, then $\widehat{T}:\widehat{Y}\to\widehat{Y}$ is ergodic, so that $\{\sum_{j=1}^N\widehat{\phi}\circ\widehat{T}^{-j},\,N\geq 1\}$ is an ergodic stationary martingale. It now follows from Billingsley [2, 3] that the CLT and WIP hold for $\sum_{j=1}^N\widehat{\phi}\circ\widehat{T}^{-j}$. But $\sum_{j=0}^{N-1}\widehat{\phi}\circ\widehat{T}^j=\sum_{j=1}^N\widehat{\phi}\circ\widehat{T}^{-j}\circ\widehat{T}^N$ so that the partial sums $\sum_{j=0}^{N-1}\phi\circ T^j$, $\sum_{j=0}^{N-1}\widehat{\phi}\circ\widehat{T}^j$, and $\sum_{j=1}^N\widehat{\phi}\circ\widehat{T}^{-j}$ have the same distribution. It follows easily that the CLT and WIP hold for ϕ as $N\to\infty$.

4 Exponential decay of correlations

In this section we prove exponential decay of correlations for equivariant observations. For one-sided shifts this is a direct consequence of the fact that condition (3.1) of Section 3 holds in our setting (Proposition 4.5). For two-sided shifts (Theorem 4.10) we use the one-sided results and an adaptation of the proof for scalar-valued observations.

In Subsections (a) and (b), we consider G-extensions of one-sided subshifts and two-sided subshifts respectively. In Subsection (c) we consider G-extensions of general hyperbolic basic sets. In Subsection (d) we show that these results extend to a larger class which includes principal G-extensions over hyperbolic basic sets as well as examples of partially hyperbolic sets where the G-action is not free.

(a) Extensions of one-sided subshifts

The Ruelle operator Let $\sigma:X^+\to X^+$ denote an aperiodic subshift of finite type. The Hölder spaces $F_\theta(X^+,\mathbb{R})$, $0<\theta<1$, are defined in the usual way, see the appendix. Given $f\in F_\theta(X^+,\mathbb{R})$, the *Ruelle operator* $L_f:C(X^+,\mathbb{R})\to C(X^+,\mathbb{R})$ is defined by

$$(L_f w)(x) = \sum_{\sigma y = x} e^{f(y)} w(y).$$

Note that $L_f((v \circ \sigma)w) = v L_f w$. We recall some basic facts about the Ruelle operator and refer to the books by Ruelle [32], Bowen [5], or Parry and Pollicott [26] for details.

First of all, f can be 'normalized' so that $L_f 1 = 1$. We shall always assume that f is normalized. It follows immediately that $L_f U = I$ where $Uw = w \circ \sigma$. There is a unique σ-invariant equilibrium state μ such that $L_f^* \mu = \mu$. Note that L_f acts on $L^2 = L^2(X^+, \mu)$ and that U is an isometry on L^2. Moreover, $L_f = U^*$.

It is clear that the spectrum of L_f lies inside the unit disk and that 1 is an eigenvalue (since $L_f 1 = 1$). In fact, 1 is a simple eigenvalue and there are no further eigenvalues on the unit circle. Restricting to Hölder spaces, it can be shown that $L_f : F_\theta(X^+, \mathbb{R}) \to F_\theta(X^+, \mathbb{R})$ has essential spectral radius θ. Outside this radius, the spectrum consists of isolated eigenvalues of finite multiplicity. Hence, the spectrum of L_f, excluding the simple eigenvalue 1, has radius $\rho_0 \in (\theta, 1)$.

The subspace of $F_\theta(X^+, \mathbb{R})$ consisting of functions w with $\int_{X^+} w \, d\mu = 0$ is preserved by L_f, and on this subspace, L_f has spectral radius ρ_0. Hence, for any $\rho \in (\rho_0, 1)$ there is a constant $C > 0$ such that $\|L_f^j w\|_\theta \le C\rho^j \|w\|_\theta$ for all $j \ge 1$ and $w \in F_\theta(X^+, \mathbb{R})$ with $\int_{X^+} w \, d\mu = 0$.

Similarly, $L_f : F_\theta(X^+, \mathbb{R}^n) \to F_\theta(X^+, \mathbb{R}^n)$ has an eigenvalue 1 of multiplicity n (the eigenfunctions are the constant functions) and the reminder of the spectrum lies in the disk of radius ρ_0. Again, $\|L_f^j w\|_\theta \le C\rho^j \|w\|_\theta$, for all $j \ge 1$ and $w \in F_\theta(X^+, \mathbb{R}^n)$ with $\int_{X^+} w \, d\mu = 0$.

Therefore, it follows from [5, 32, 26] that the Banach space $F = F_\theta(X^+, \mathbb{R}^n)$ satisfies the property (3.1) required in Section 3 and that exponential decay of correlations holds for Hölder observations.

The equivariant Ruelle operator Let G be a compact Lie group with Haar measure ν. Suppose that G acts (orthogonally) on $\mathbb{R}^n$ and identify G with a subset of $\mathbf{O}(n)$. For $f \in F_\theta(X^+, \mathbb{R})$ normalized as above, and $h \in F_\theta(X^+, G)$, define the *equivariant Ruelle operator* $L_{f,h} : F_\theta(X^+, \mathbb{R}^n) \to F_\theta(X^+, \mathbb{R}^n)$ by

$$L_{f,h} w = L_f(h^{-1} \cdot w).$$

The operator $L_{f,h}$ was introduced and studied extensively by Parry and Pollicott [25], see also [26]. Again, $L_{f,h}$ has spectral radius at most 1 and essential spectral radius θ.

We define the skew product $\sigma_h : X^+ \times G \to X^+ \times G$ by $\sigma_h(x, g) = (\sigma x, gh(x))$. The product measure $m = \mu \times \nu$ is invariant under σ_h. Recall that σ_h is weak mixing if the only measurable eigenfunctions are the constants (that is, the equation

$\phi \circ \sigma_h = \alpha\phi$ ae where $\phi : X^+ \times G \to \mathbb{C}$ is measurable and $\alpha \in \mathbb{C}$, has only the 'trivial' solutions $\alpha = 1$ and ϕ constant). It is a consequence of a Livšic-type regularity theorem [24] that measurable solutions ϕ have Hölder continuous versions (and hence σ_h is weak mixing if and only if it is topologically weak mixing).

In the remainder of this subsection, we suppose that $f \in F_\theta(X^+, \mathbb{R})$ and $h \in F_\theta(X^+, G)$ are such that σ_h is weak mixing.

Corollary 4.1 *Suppose that* $\mathrm{Fix}\, G = \{0\}$. *Then* $L_{f,h} : F_\theta(X^+, \mathbb{R}^n) \to F_\theta(X^+, \mathbb{R}^n)$ *has spectral radius less than one.*

Proof (cf. [26, Theorem 8.3]) Since $L_{f,h}$ has essential spectral radius $\theta < 1$, it is sufficient to rule out the existence of eigenvalues $\alpha \in \mathbb{C}$ with $|\alpha| = 1$. Such an eigenvalue leads to the equation $L_{f,h}w = \alpha w$ and by a standard convexity argument [26, p. 130], $h(x)w(\sigma x) = \alpha^{-1}w(x)$. Equivalently, $\phi \circ \sigma_h = \alpha^{-1}\phi$ where $\phi(x, g) = gw(x)$. By definition of weak mixing, ϕ is constant. Since $\mathrm{Fix}\, G = \{0\}$ we deduce that $w = 0$. ∎

Write $\mathbb{R}^n = \mathrm{Fix}\, G \oplus (\mathrm{Fix}\, G)^\perp$. Then $F_\theta(X^+, \mathbb{R}^n) = F_\theta(X^+, \mathrm{Fix}\, G) \oplus F_\theta(X^+, (\mathrm{Fix}\, G)^\perp)$, and $L_{f,h}$ preserves the splitting. Moreover, $L_{f,h} = L_f$ on the first summand, and Corollary 4.1 applies to the second summand. Hence there is the eigenvalue 1 of multiplicity equal to the dimension of $\mathrm{Fix}\, G$ and, provided σ_h is weak mixing, the remainder of the spectrum of $L_{f,h}$ lies within a disk of radius $\rho_0 < 1$.

Corollary 4.2 *Let* $\rho \in (\rho_0, 1)$. *Then there is a constant* C *(that depends only on* θ, ρ, f *and* h*) such that*

$$\|L^j_{f,h}w - \textstyle\int_{X^+ \times G} g \cdot w\, dm\|_\theta \le C\|w\|_\theta\, \rho^j,$$

for all $j \ge 1$ *and all* $w \in F_\theta(X^+, \mathbb{R}^n)$.

Proof For the $\mathrm{Fix}\, G$ components of w, we have $L^j_{f,h}w = L^j_f w \to \int_{X^+} w$ exponentially quickly, as in the nonequivariant case. For the $(\mathrm{Fix}\, G)^\perp$ components, $\int_{X^+ \times G} g \cdot w\, dm = 0$ and we apply Corollary 4.1. ∎

Remark 4.3 Viewed as an operator on $L^2(X^+, \mathbb{R}^n)$, $L_{f,h}$ is the adjoint of the isometry $\widehat{U}$ defined by $\widehat{U}v = hv \circ \sigma$.

The one sided shift $\sigma_h : X^+ \times G \to X^+ \times G$ induces an isometry $U : L^2_G(X^+ \times G, \mathbb{R}^n) \to L^2_G(X^+ \times G, \mathbb{R}^n)$ defined by $U\phi = \phi \circ \sigma_h$. We have a natural bijective isometry $L^2_G(X^+ \times G, \mathbb{R}^n) \cong L^2(X^+, \mathbb{R}^n)$, defined by mapping $\phi \in L^2_G(X^+ \times G, \mathbb{R}^n)$ to $\widehat{\phi} \in L^2(X^+, \mathbb{R})$, where $\phi(x, g) = g\widehat{\phi}(x)$, for all $(x, g) \in X^+ \times G$. Under this natural isomorphism, the isometry U of $L^2_G(X^+ \times G, \mathbb{R}^n)$ determines the isometry $\widehat{U}$

of $L^2(X^+,\mathbb{R}^n)$ defined above. The next result is an immediate consequence of these observations.

Proposition 4.4 *The adjoint $U^\star$ of $U : L^2_G(X^+\times G,\mathbb{R}^n) \to L^2_G(X^+\times G,\mathbb{R}^n)$ is given by $\widehat{U^*} = \widehat{U}^\star$. (That is, given $\phi \in L^2_G(X^+ \times G,\mathbb{R}^n)$, $\widehat{U^*\phi} = L_{f,h}\widehat{\phi}$.)*

Let $F^G_\theta(X^+ \times G,\mathbb{R}^n)$ denote the space of equivariant observations $\phi(x,g) = gv(x)$ where $v \in F_\theta(X^+,\mathbb{R}^n)$, and define $\|\phi\|_\theta = \|v\|_\theta$. Just as above, we have the natural $\| \,\|_\theta$ preserving isomorphism $F^G_\theta(X^+ \times G,\mathbb{R}^n) \cong F_\theta(X^+,\mathbb{R}^n)$. The operators U, U^* on $L^2_G(X^+ \times G,\mathbb{R}^n)$ restrict to operators on $F^G_\theta(X^+ \times G,\mathbb{R}^n)$.

Proposition 4.5 *Let $C > 0$ and $\rho \in (0,1)$ be as in Corollary 4.2. Suppose that $\phi \in F^G_\theta(X^+ \times G,\mathbb{R}^n)$ and $\int \phi = 0$. Then*

$$\|(U^*)^j\phi\|_\theta \le C\rho^j\|\phi\|_\theta, \quad \textit{for all } j \ge 1,$$

Proof It follows from Corollary 4.2 that $\|L^j_{f,h}\widehat{\phi}\|_\theta \le C\rho^j\|\widehat{\phi}\|_\theta$. Hence

$$\|(U^*)^j\phi\|_\theta = \|L^j_{f,h}\widehat{\phi}\|_\theta \le C\rho^j\|\widehat{\phi}\|_\theta = C\rho^j\|\phi\|_\theta.$$

∎

It follows that $F = F^G_\theta(X^+ \times G,\mathbb{R}^n)$ satisfies the requirement (3.1) of Section 3. In particular, we have established exponential decay of correlations for extensions of one-sided subshifts:

Theorem 4.6 *Let $C > 0$ and $\rho \in (0,1)$ be as in Corollary 4.2. Then*

$$\left|\int\phi \circ \sigma^j_h\, \psi^T - \int\phi\int\psi^T\right| \le C|\phi|_2\,\|\psi\|_\theta\,\rho^j,$$

for all $j \ge 1$, and for all $\phi \in L^2_G(X^+ \times G,\mathbb{R}^n)$, $\psi \in F^G_\theta(X^+ \times G,\mathbb{R}^n)$.

(b) Extensions of two-sided subshifts

Let $\sigma : X \to X$ be an aperiodic two-sided subshift of finite type. The Hölder spaces $F_\theta(X,\mathbb{R}^n)$, $F_\theta(X,G)$, $\theta \in (0,1)$ are defined as in the one-sided case, see Appendix A. Let $f \in F_\theta(X,\mathbb{R})$ be a Gibbs potential. It is known [26] that f is cohomologous to a potential $\widetilde{f} \in F_{\theta^{1/2}}(X^+,\mathbb{R})$. We denote the corresponding equilibrium states on X, X^+ by $\mu, \widetilde{\mu}$ respectively. If $\phi : X^+ \to \mathbb{R}$ is $\widetilde{\mu}$-measurable, then the natural map induced by ϕ on X is μ-measurable and the integrals over X, X^+ coincide (if they exist).

We consider skew products $\sigma_h : X \times G \to X \times G$ induced by $h \in F_\theta(X, G)$, and write $\sigma_h^j(x, g) = (\sigma^j x, g h_j(x))$ for $j \in \mathbb{Z}$. In particular, $h_j(x) = h(x)h(\sigma x) \cdots h(\sigma^{j-1}x)$ when $j \geq 1$.

Suppose that G acts orthogonally on $\mathbb{R}^n$. Again, there is a natural one-to-one correspondence between equivariant observations $\phi : X \times G \to \mathbb{R}^n$ and functions $\widehat{\phi} : X \to \mathbb{R}^n$. We define the spaces of Hölder equivariant observations $F_\theta^G(X \times G, \mathbb{R}^n) \cong F_\theta(X, \mathbb{R}^n)$ just as we did in the one-sided case.

The following result enables us to apply results for one-sided subshifts.

Lemma 4.7 *Suppose that $\widehat{\phi} \in F_\theta(X, \mathbb{R}^n)$, $h \in F_\theta(X, G)$, where $0 < \theta < 1$. Then there exist functions $\widetilde{h}, \widehat{M} \in F_{\theta^{1/2}}(X, G)$ and $\widehat{\psi}, \chi \in F_{\theta^{1/4}}(X, \mathbb{R}^n)$ such that $\widehat{\psi}$ and $\widetilde{h}$ depend only on future variables and*

$$h = \widehat{M}\widetilde{h}(\widehat{M} \circ \sigma)^{-1}, \qquad \widehat{M}^{-1}\widehat{\phi} = \widehat{\psi} + \chi - \widetilde{h}\chi \circ \sigma.$$

Moreover, $\sigma_{\widetilde{h}}$ is weak mixing if and only σ_h is weak mixing.

Proof The construction of $\widetilde{h}$ and $\widehat{M}$ can be found in Parry and Pollicott [26, Appendix II]. The construction of $\widehat{\psi}$ and χ is a simple adaptation of the standard proof where $\widetilde{h}$ is absent and is proved in Appendix A. ∎

Let $\phi, \psi : X \times G \to \mathbb{R}^n$ be the equivariant observations corresponding to $\widehat{\phi}, \widehat{\psi}$ respectively. We define $M : X \times G \to G$ by $M(x, g) = g\widehat{M}(x)g^{-1}$. For $N \geq 1$, define $\psi_N = \sum_{j=0}^{N-1} \psi \circ \sigma_{\widetilde{h}}^j$.

Corollary 4.8 $\phi_N = M\psi_N + O(1)$.

Proof It suffices to show that $\widehat{\phi}_N = \widehat{M}\widehat{\psi}_N + O(1)$. Since $h_j = \widehat{M}\widetilde{h}_j(\widehat{M} \circ \sigma^j)^{-1}$, we have

$$\begin{aligned}
\widehat{U}^j\widehat{\phi} &= h_j\widehat{\phi} \circ \sigma^j = \widehat{M}\widetilde{h}_j(\widehat{M} \circ \sigma^j)^{-1}\widehat{\phi} \circ \sigma^j = \widehat{M}\widetilde{h}_j(\widehat{M}^{-1}\widehat{\phi}) \circ \sigma^j \\
&= \widehat{M}\widetilde{h}_j(\widehat{\phi} \circ \sigma^j + \chi \circ \sigma^j - \widetilde{h} \circ \sigma^j \chi \circ \sigma^{j+1}) \\
&= \widehat{M}\widehat{U}^j\widehat{\psi} + \widehat{M}(\widetilde{h}_j\chi \circ \sigma^j - \widetilde{h}_{j+1}\chi \circ \sigma^{j+1}).
\end{aligned}$$

The result follows. ∎

Lemma 4.9 *Suppose that $h \in F_\theta(X^+, G) \hookrightarrow F_\theta(X, G)$ and that σ_h is weak mixing. Let $\rho \in (\rho_0, 1)$ where ρ_0 is as in Corollary 4.2. Then there is a constant $C > 0$ such that*

$$\left|\int \phi \circ \sigma_h^j \, \psi^T - \int \phi \int \psi^T\right| \leq C\|\phi\|_\theta \, \|\psi\|_\theta (\rho^{1/3})^j, \tag{4.1}$$

for all $j \geq 1$, and for all $\phi, \psi \in F_\theta^G(X \times G, \mathbb{R}^n)$.

Proof We follow the proof of [26, Proposition 2.4], taking into account the extra complications due to the presence of h. We may restrict to the case $\mathrm{Fix}\, G = 0$ (since when G acts trivially we are in the nonequivariant case which is proved in [26, Proposition 2.4]). In particular, ϕ and ψ automatically have mean zero. Since $\int_{X\times G} \phi\circ \sigma_h^j\, \psi^T dm = \int_X h_j v\circ\sigma^j\, w^T d\mu$, we reduce to proving that

$$|\textstyle\int_X h_j v\circ\sigma^j\, w^T d\mu| \le C\|v\|_\theta\, \|w\|_\theta (\rho^{1/3})^j.$$

Following [26], for each $k\ge 1$, we choose v_k depending only on coordinates x_i, $|i|\le k$, such that $|v_k|_\infty \le |v|_\infty$ and $|v-v_k|_\infty \le |v|_\theta\, \theta^k$. Similarly, choose w_k for w. (For example, choose $\varphi_k : X\to X$ depending only on coordinates x_i, $|i|\le k$, and such that $(\varphi_k(x))_i = x_i$ for $|i|\le k$. Define $v_k(x) = v(\varphi_k(x))$ and note that $|v-v_k|_\infty \le \mathrm{var}_k(v)$.) Write

$$\begin{aligned}\textstyle\int_X h_j v\circ\sigma^j\, w^T d\mu =& \textstyle\int_X h_j (v-v_k)\circ\sigma^j\, w^T d\mu + \int_X h_j v_k\circ\sigma^j\, (w-w_k)^T d\mu \\ &+ \textstyle\int_X h_j v_k\circ\sigma^j\, w_k^T d\mu.\end{aligned}$$

The first two terms are estimated just as in [26]. The first term is dominated by $|h_j(v-v_k)\circ\sigma^j|_\infty |w|_\infty \le |v|_\theta |w|_\infty \theta^k$. Similarly, the second term is dominated by $|v|_\infty |w|_\theta \theta^k$. Hence

$$|\textstyle\int_X h_j v\circ\sigma^j\, w^T d\mu| \le (|v|_\theta\, |w|_\infty + |v|_\infty\, |w|_\theta)\theta^k + |I_{jk}|, \tag{4.2}$$

where $I_{jk} = \int_X h_j v_k\circ\sigma^j\cdot w_k^T d\mu$.

Since μ is σ-invariant,

$$I_{jk} = \textstyle\int_X h_j\circ\sigma^k v_k\circ\sigma^k\circ\sigma^j\cdot w_k^T\circ\sigma^k d\mu.$$

Note that $h_{j+k} = h_k h_j\circ\sigma^k = h_j h_k\circ\sigma^j$ so that $h_j\circ\sigma^k = h_k^{-1} h_j h_k\circ\sigma^j$. It follows that

$$I_{jk} = \textstyle\int_X h_j (h_k v_k\circ\sigma^k)\circ\sigma^j\cdot (h_k w_k)^T\circ\sigma^k d\mu.$$

By definition of v_k, w_k and by our assumption on h, the integrand depends only on future coordinates. Therefore, it follows from Theorem 4.6 that we can choose a constant $C_1\ge 1$ such that

$$|I_{jk}| \le C_1\, |h_k v_k\circ\sigma^k|_\infty\, \|h_k w_k\circ\sigma^k\|_\theta\, \rho^j.$$

But $|h_k v_k\circ\sigma^k|_\infty = |v_k|_\infty \le |v|_\infty$ and we claim that

$$\|h_k w_k\circ\sigma^k\|_\theta \le C_2 |w|_\infty \theta^{-2k},$$

where $C_2 \geq 1$ depends only on h and θ. Hence

$$|I_{jk}| \leq C|v|_\infty |w|_\infty \theta^{-2k}\rho^j, \tag{4.3}$$

where $C = C_1C_2$. To verify the claim, compute that

$$\|h_k w_k \circ \sigma^k\|_\theta = |h_k w_k \circ \sigma^k|_\infty + |h_k w_k \circ \sigma^k|_\theta \leq |w_k|_\infty + |w_k \circ \sigma^k|_\theta + |h_k|_\theta |w_k|_\infty,$$

and use the inequalities $|w_k|_\infty \leq |w|_\infty$, $|w_k \circ \sigma^k|_\theta \leq \theta^{-k}|w_k|_\theta \leq \theta^{-k}2\theta^{-k+1}|w|_\infty$, and $|h_k|_\theta \leq |h|_\theta + \cdots + |h \circ \sigma^{k-1}|_\theta \leq |h|_\theta(1 + \cdots + \theta^{-k+1}) \leq |h|_\theta \theta^{-k+1}/(1-\theta)$. This establishes the claim.

Combining (4.2) and (4.3), we have

$$|\textstyle\int_X h_j v \circ \sigma^j\, w^T d\mu| \leq C(|v|_\theta\, |w|_\infty + |v|_\infty\, |w|_\theta)\theta^k + C|v|_\infty\, |w|_\infty \theta^{-2k}\rho^j.$$

Now, let $r > 1$ be such that $\rho^r = \theta$ and set $k = j/3r$ to obtain the required result. ∎

Theorem 4.10 *Suppose that $h \in F_\theta(X,G)$ and that σ_h is weak mixing. There are constants $C > 0$ and $\tau \in (0,1)$ such that*

$$\left|\textstyle\int \phi \circ \sigma_h^j\, \psi^T - \int \phi \int \psi^T\right| \leq C\, \|\phi\|_{\theta^{1/2}} \|\psi\|_{\theta^{1/2}}\, \tau^j,$$

for all $j \geq 1$, and for all $\phi, \psi \in F_\theta^G(X \times G, \mathbb{R}^n)$.

Proof Again, we may suppose that $\operatorname{Fix} G = 0$ so that it suffices to estimate $\int \phi \circ \sigma_h^j\, \psi^T = \int \widehat{U}^j \widehat{\phi}\widehat{\psi}^T$.

By Lemma 4.7, we may write $h = \widehat{M}\widetilde{h}(\widehat{M} \circ \sigma)^{-1}$ where $\widehat{M}, \widetilde{h} \in F_{\theta^{1/2}}(X,G)$ and $\widetilde{h}$ depends only on future coordinates. Since $h_j = \widehat{M}\widetilde{h}_j(\widehat{M} \circ \sigma^j)^{-1}$, we have

$$h_j \widehat{\phi} \circ \sigma^j \widehat{\psi}^T = \widehat{M}\widetilde{h}_j(\widehat{M}^{-1}\widehat{\phi}) \circ \sigma^j \widehat{\psi}^T = \widehat{M}\widetilde{h}_j \widehat{\Phi} \circ \sigma^j \widehat{\Psi}^T \widehat{M}^T = \widetilde{h}_j \widehat{\Phi} \circ \sigma^j \widehat{\Psi}^T,$$

where $\widehat{\Phi} = \widehat{M}^{-1}\widehat{\phi}$, $\widehat{\Psi} = \widehat{M}^{-1}\widehat{\psi}$ and $\Phi, \Psi \in F^G_{\theta^{1/2}}(X \times G, \mathbb{R}^n)$. Hence

$$\textstyle\int \phi \circ \sigma_h^j\, \psi^T = \int \widehat{U}^j \widehat{\phi}\widehat{\psi}^T = \int \widetilde{h}_j \widehat{\Phi} \circ \sigma^j \widehat{\Psi}^T = \int \Phi \circ \sigma_{\widetilde{h}}^j\, \Psi^T.$$

Note that $\Phi, \Psi \in F^G_{\theta^{1/2}}(X \times G, \mathbb{R}^n)$.

Since $\widetilde{h}$ depends only on future coordinates, it follows from Lemma 4.9 that there are constants $D > 0$, $\tau \in (0,1)$ such that

$$|\textstyle\int \phi \circ \sigma_h^j\, \psi^T| \leq D\|\Phi\|_{\theta^{1/2}} \|\Psi\|_{\theta^{1/2}} \tau^j = D\|\widehat{\Phi}\|_{\theta^{1/2}} \|\widehat{\Psi}\|_{\theta^{1/2}} \tau^j.$$

Finally, note that there is a constant $K > 1$ depending only on M and θ such that

$$|\widehat{\Phi}|_{\theta^{1/2}} \leq K|\phi|_{\theta^{1/2}}, \quad |\widehat{\Psi}|_{\theta^{1/2}} \leq K|\psi|_{\theta^{1/2}}.$$

The result follows with $C = DK^2$. ∎

Remark 4.11 Actually, Theorem 4.10 holds for all $\phi, \psi \in F^G_{\theta^{1/2}}(X \times G, \mathbb{R}^n)$. (Of course, h must lie in $F_\theta(X,G)$.)

(c) Extensions of Axiom A diffeomorphisms

Suppose that $\Lambda \subset M$ is a hyperbolic basic set for a diffeomorphism $f : M \to M$, where M is a compact manifold. Let $\rho : \Lambda \to \mathbb{R}$ be Hölder continuous and μ_Λ denote the corresponding equilibrium state on Λ. Bowen [5] proved that there is a two-sided subshift of finite type $\sigma : X \to X$ and a Hölder semiconjugacy $\pi : X \to \Lambda$ such that $\pi\sigma = f\pi$. Further, π is an isomorphism with respect to μ_Λ and the equilibrium state μ_X on X defined by the potential $\rho\pi$.

Let G be a compact connected Lie group with Haar measure ν. Let $S : \Lambda \times G \to \Lambda \times G$ be the compact group extension defined by the Hölder map $h : \Lambda \to G$. Then $h\pi : X \to G$ is Hölder and defines a group extension $\sigma_{h\pi} : X \times G \to X \times G$. Note that $\Pi = \pi \times \mathrm{id} : X \times G \to \Lambda \times G$ defines a semiconjugacy between $\sigma_{h\pi}$ and S. Also Π is an isomorphism with respect to the measures $m_X = \mu_X \times \nu$ and $m_\Lambda = \mu_\Lambda \times \nu$. In particular, S is weak mixing if and only if $\sigma_{h\pi}$ is weak mixing.

It is therefore immediate that exponential decay of correlations for subshifts implies the same for hyperbolic basic sets, proving Theorem 2.2.

(d) Principal extensions

The previous results extend to principal extensions. Specifically, suppose that M is a compact G-manifold, $F : M \to M$ is a G-equivariant diffeomorphism and $\Omega \subset M$ is a G- and F-invariant subset of M. We assume that G acts *freely* on Ω. It follows by standard results on smooth G-actions that G acts freely on an open G-invariant neighborhood U of Ω (see Bredon [6]). If we choose an open G-invariant neighborhood V of Ω, $V \subset U$, such that $F(V) \subset U$, then F induces a smooth embedding $f : V/G \subset U/G$ and $\Lambda = \Omega/G$ is then a compact f-invariant subset of U/G. We assume that Λ is a hyperbolic basic set for f. Equivalently, and more naturally, we may assume that Ω is a *partially hyperbolic* subset of M with center foliation given by the G-orbits (see [13]).

Choose a Hölder continuous potential ρ on Λ and let μ_Λ denote the corresponding equilibrium state on Λ. (Equivalently, we could have chosen a Hölder continuous G-invariant potential on Ω.) As usual, we assume that f is topologically mixing and therefore is μ_Λ-mixing. Let m denote the Haar extension [13] of μ_Λ to Ω. Extend m by zero to all of M. With this convention, F is measure preserving. We further assume that F is weak mixing. Just as in the case of skew extensions, it is straightforward to verify that weak mixing holds generically (we refer to [15, 13] for a precise description of allowable perturbations). Note that if the principal bundle $\Omega \to \Lambda$ admits no principal H-subbundles, for any compact Lie subgroup H of G, then F is always mixing (cf. [27, 7]).

Following the case of skew extensions of Axiom A diffeomorphisms, we choose a two-sided subshift of finite type $\sigma : X \to X$ and a Hölder semiconjugacy $\pi : X \to \Lambda$ such that $\pi\sigma = f\pi$. Let μ_X denote the equilibrium state on X defined by the Gibbs potential $\rho\pi$.

Let $p : \Omega \to \Lambda$ denote the orbit map. Recall that the pullback $p^\star X$ by p of X is defined to be $p^\star X = \{(x,\omega) \in X \times \Omega : \pi(x) = p(\omega)\}$. Clearly, G acts freely on $p^\star X$ and the natural projection $P : p^\star X \to X$ gives $p^\star X$ the structure of a G-principal Hölder bundle over X. Since X is totally disconnected, $p^\star X$ is a trivial bundle and so we may write $p^\star X = X \times G$. The map $H : p^\star X \to p^\star X$ induced by F and σ is G-equivariant and hence we can write $H = \sigma_h : X \times G \to X \times G$ for some Hölder continuous skewing map $h : X \to G$. Let $\Pi : X \times G \to \Omega$ denote the corresponding projection on Ω. Suppose $\phi : \Omega \to \mathbb{R}^n$ is a Hölder continuous equivariant observation on Ω. Then ϕ lifts to the Hölder continuous equivariant observation $\phi\Pi : X \times G \to \mathbb{R}^n$. Everything now goes through just as in the skew extension case and so it follows that Theorem 2.2 holds for principal extensions.

Remarks 4.12 (1) Note that the result on principal extensions holds even though equivariant observations on Ω do not correspond to observations on Ω/G. That is, $\mathbb{R}^n$-valued maps on Ω/G do not determine equivariant maps on Ω unless $p : \Omega \to \Omega/G$ is a trivial G-principal bundle.

(2) Theorem 2.2 continues to hold if we can construct equivariant Hölder surjections of twisted products $X \times_H G$ onto Ω. This is relevant for generalizations of our main results to more general partial hyperbolic G-invariant sets for which the action of G is no longer free (see [13]). For example, if $H \subset G$ is a finite group and Λ is an H-invariant hyperbolic set for an H-equivariant diffeomorphism, then Theorem 2.2 applies to the twisted product $\Lambda \times_H G$.

5 Covariance matrix for equivariant observations

In Section 3(b), we defined the covariance matrix $\Sigma = \Sigma_\phi$ within a quite general framework. In particular, we used decay of correlations to show that $\Sigma = \lim_{N\to\infty} \frac{1}{N} \int \phi_N \phi_N^T$.

In the specific context of equivariant observations on a compact group extension, the covariance matrices have additional structure. Moreover, we have good necessary and sufficient conditions for nondegeneracy, cf. [23]. We recall these results now within the context of G-extensions of one-sided subshifts.

As in Section 4(a), we fix $f \in F_\theta(X^+, \mathbb{R})$ and $h \in F_\theta(X^+, G)$, and suppose that σ_h is weak mixing.

Proposition 5.1 *Let $\phi \in F_\theta^G(X^+ \times G, \mathbb{R}^n)$ with $\int_{X^+\times G} \phi \, dm = 0$. Define the covariance $\Sigma = \Sigma_\phi$ as in Proposition 3.3. Then Σ commutes with the action of G*

on $\mathbb{R}^n$.

Proof By Proposition 3.3, $\Sigma = \lim_{N\to\infty} \frac{1}{N}\int \phi_N\phi_N^T dm$. Equivariance of ϕ means that $\phi(x, ag) = a\phi(x, g)$ for all $a \in G$. It follows from invariance of Haar measure that $\int \phi_N\phi_N^T dm = a(\int \phi_N\phi_N^T dm)a^T = a(\int \phi_N\phi_N^T dm)a^{-1}$ for all $a \in G$. Hence $\Sigma = a\Sigma a^{-1}$ for all $a \in G$. ∎

Recall that Σ is degenerate if $\det\Sigma = 0$.

Remark 5.2 If G acts irreducibly on $\mathbb{R}^n$, then the real version of Schur's lemma [20, Theorem 2, p. 119] states that the space of commuting linear maps on $\mathbb{R}^n$ is a real division ring and hence isomorphic to $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$. Positive semidefiniteness implies that Σ is a *real* scalar multiple of the identity. It follows that when G acts irreducibly, Σ is degenerate if and only if $\Sigma = 0$. Consequently, by Corollary 3.5, if G acts irreducibly, then Σ is degenerate if and only if ϕ is a Hölder coboundary.

Proposition 5.3 *The covariance matrix* Σ *is degenerate if and only if there is an irreducible subspace* $V \subset \mathbb{R}^n$, *with orthogonal projection* $\pi_V : \mathbb{R}^n \to V$, *such that* $\pi_V\phi$ *is a Hölder coboundary.*

Proof Degeneracy means that $\dim\ker\Sigma \geq 1$. Since $\ker\Sigma$ is G-invariant, there is a G-irreducible subspace $V \subset \ker\Sigma$. Restricting to V, we have $\Sigma|_V = 0$. But $\Sigma|_V$ is the covariance matrix corresponding to the sequence of random variables $\frac{1}{\sqrt{N}}(\pi_V\phi)_N$. Since V is G-invariant, $\pi_V\phi$ is a G-equivariant observation and we remain within the setting of Section 3. Now apply Corollary 3.5. ∎

Remark 5.4 Nicol *et al.* [23, Lemma 5.8] show that typically $\pi_V\phi$ is not a coboundary for any projection π_V. Indeed, it is sufficient to perturb ϕ at one point. Hence, we conclude that for an open and dense subset of equivariant observations $\phi \in F_\theta^G(X \times G, \mathbb{R}^n)$, the covariance matrix Σ is positive definite. Furthermore, Σ is a general symmetric positive definite G-equivariant $n \times n$ matrix.

6 Almost sure invariance principle

In Subsection (a) we prove the ASIP for extensions of two-sided shifts. This immediately leads to similar results for (principal) extensions of hyperbolic sets, as well as for extensions of one-sided shifts. These are mentioned in Subsection (b).

Throughout this section, the statement $\{A_N,\ N \geq 1\} =_d \{B_N,\ N \geq 1\}$ means that the two sequences are equal in distribution. This is equivalent to demanding equality of the finite distributions $\{A_j,\ j = 1,\dots,N\}$ and $\{B_j,\ j = 1,\dots,N\}$ for all $N \geq 1$ [19, Theorem 3.29].

(a) Extensions of two-sided subshifts

Suppose that $\sigma : X \to X$ is a two-sided subshift of finite type and that $h \in F_\theta(X,G)$ induces a weak mixing G-extension $\sigma_h : X \times G \to X \times G$. Suppose also that $\phi \in F_\theta^G(X \times G, \mathbb{R}^n)$ has mean zero and nonsingular covariance Σ.

Recall that if $\widehat{M} : X \to G$, then $M : X \times G \to G$ is defined by $M(x,g) = g\widehat{M}(x)g^{-1}$.

Lemma 6.1 *There exists $\widetilde{h} \in F_{\theta^{1/2}}(X,G)$ such that $\sigma_{\widetilde{h}} : X \times G \to X \times G$ is weak mixing, and there exists $\psi \in F^G_{\theta^{1/4}}(X \times G, \mathbb{R}^n)$ with mean zero, and $\widehat{M} : X \to G$ $\theta^{1/2}$-Hölder continuous, such that*

(i) $\{\psi_N = \sum_{j=0}^{N-1} \psi \circ \sigma_{\widetilde{h}}^j,\ N \geq 1\}$ is a stationary, ergodic, essentially bounded ($\mathbb{R}^n$-valued) martingale,

(ii) $\phi_N = M\psi_N + O(1)$, and $\int \psi\psi^T = \Sigma_\psi = \Sigma$.

Proof Since the class of G-extensions of two-sided subshifts is closed under time-reversal, it is sufficient to consider reverse time partial sums $\phi_{-N} = \sum_{j=0}^{N-1} \phi \circ \sigma_h^{-j}$.

Construct $\widetilde{h}$ and $\widehat{\psi}$ depending only on future coordinates as in Lemma 4.7. As in Corollary 4.8, we can write $\phi_{-N} = M\psi_{-N} + O(1)$. By Lemma 3.4, we can replace ψ by a new ψ depending only on future coordinates and satisfying $U^*\psi = 0$. (As usual, the adjoint U^* corresponds to the isometry $U : L^2(X^+ \times G) \to L^2(X^+ \times G)$ induced by $\sigma_{\widetilde{h}}$).

Define the sequence

$$Y_j = \psi \circ \sigma_{\widetilde{h}}^{-j}, \qquad \mathcal{F}_j = \sigma_{\widetilde{h}}^j(\mathcal{M}^+ \times G),$$

where $\mathcal{M}^+$ is the σ-algebra on X^+ lifted up to X. Then $\{\mathcal{F}_j : j \in \mathbb{Z}\}$ is an increasing sequence of σ-algebras. We claim that $\{\psi_{-N} = \sum_{j=0}^{N-1} Y_j, \mathcal{F}_N, N \in \mathbb{Z}\}$ is a martingale. That is, Y_j is $\mathcal{F}_j$-measurable and $E(Y_j|\mathcal{F}_{j-1}) = 0$ for all j. By stationarity, it suffices to prove these for $j = 0$. Since ψ depends only on future coordinates, Y_0 is $\mathcal{F}_0$-measurable. Since in addition $\widetilde{h}$ depends only on future coordinates, $E(\psi|\sigma_{\widetilde{h}}^{-1}\mathcal{F}_0) = 0$ by Remark 3.12, proving the claim. ∎

Proposition 6.2 *Let $c \in \mathbb{R}^n$ and set $\mu_c^2 = c^T \int \phi\phi^T c$. Define $X_j = (1/\mu_c)c^T\phi \circ \sigma_h^j$, and define $s_N^2 = \sum_{j=0}^{N-1} X_j^2$. Then for any $\delta > 0$,*

$$s_N^2 = N + O(N^{1/2+\delta}) \quad \textit{almost surely.}$$

Proof Let $\phi' = \phi\phi^T - \int \phi\phi^T$. Then $\phi' : X^+ \times G \to \mathrm{Mat}_n(\mathbb{R})$ is Hölder continuous with mean zero and is G-equivariant with respect to the obvious representation of G on $\mathrm{Mat}_n(\mathbb{R})$. It follows from Lemma 6.1 that $\phi'_N = M'\psi'_N + O(1)$ where ψ'_N is a stationary ergodic Hölder continuous martingale. By Proposition B.2 (strong law of large numbers for martingales), $\phi'_N = O(N^{1/2+\delta})$ almost surely.

Now compute that almost surely

$$s_N^2 = \sum_{j=0}^{N-1} X_j^2 = (1/\mu_c^2)c^T(\phi\phi^T)_N\, c = (1/\mu_c^2)[c^T\phi'_N c + N\mu_c^2] = N + O(N^{1/2+\delta}),$$

as required. ∎

Proposition 6.3 *If $M : X \times G \to G$ is measurable, then $\{M\phi_N,\, N \geq 1\} =_d \{\phi_N,\, N \geq 1\}$.*

Proof Fix $N \geq 1$. It suffices to show that the finite sequences $\{M\phi_j,\, j = 1, \ldots, N\} =_d \{\phi_j,\, j = 1, \ldots, N\}$. Since the sequences are G-equivariant, their distributions are determined by G-invariant subsets $I \subset (\mathbb{R}^n)^N$. But since $M \in G$, it is immediate that for such subsets $(\phi_1, \ldots, \phi_N)$ and $(M\phi_1, \ldots, M\phi_N)$ have equal probability of lying in I. ∎

Theorem 6.4 *For any $c \in \mathbb{R}^n$, there is a one-dimensional Brownian motion W with variance $c^T\Sigma c$ and a sequence of random variables $S_c(N)$ (defined on the same space) such that $\{c^T\phi_N,\, N \geq 1\} =_d \{S_c(N),\, N \geq 1\}$ and for every $\delta > 0$,*

$$S_c([t]) = W(t) + O(t^{1/4+\delta}) \quad \textit{almost surely.}$$

Proof Set $\sigma_c^2 = c^T\Sigma c$. Rescaling variances to 1, it is equivalent to prove the result with $c^T\phi$ replaced by $(1/\sigma_c)c^T\phi$ and W replaced by $(1/\sigma_c)W$ which is a standard Brownian motion on the line.

Let ψ, $\widetilde{h}$ be as in Lemma 6.1. Define $X_j = (1/\sigma_c)c^T\psi \circ \sigma_{\widetilde{h}}^j$. Then $\{\sum_{j=0}^{N-1} X_j,\, N \geq 1\}$ is a (stationary ergodic) martingale.

Since X_j is stationary and continuous, hypothesis (a) of Theorem B.3 is trivially satisfied. Hypothesis (b) follows from Proposition 6.2. Combining Theorems B.1 and B.3, we have that $\{c^T\psi_N,\, N \geq 1\} =_d \{S'_c(N),\, N \geq 1\}$ and $S'_c([t]) = W(t) + O(t^{1/4+\delta})$ almost surely.

By Proposition 6.3, $\{M\psi_N\} =_d \{\psi_N\}$. Altogether, using Lemma 6.1, we have

$$c^T\phi_N = c^T M\psi_N + O(1) \tag{6.1}$$

$$\{c^T M\psi_N,\, N \geq 1\} =_d \{S'_c(N),\, N \geq 1\}$$

$$S'_c([t]) = W(t) + O(t^{1/4+\delta}) \text{ almost surely.} \tag{6.2}$$

As in [28, p. 23], the probability space on which $S_c'(N)$ and W are defined may be enlarged so as to support a copy in distribution of $\{c^T\phi_N, c^T M\psi_N,\ N \geq 1\}$. This can be done in such a way that equations (6.1) and (6.2) remain valid, while $c^T M\psi_N$ is identified with $S_c'(N)$. The result follows. ∎

(b) Extensions of hyperbolic sets and one-sided subshifts

Suppose that $\Lambda \subset M$ is a basic hyperbolic set, $\pi : X \to \Lambda$ its coding by a two sided-shift (as in Section 4(c)), and $\Pi = \pi \times \mathrm{id} : X \times G \to \Lambda \times G$. If $\phi : \Lambda \times G \to \mathbb{R}^n$, then define $\widehat{\phi} = \phi \circ \Pi : X \times G \to \mathbb{R}^n$. Since π is an isomorphism of probability spaces, $\{\phi_N\} =_d \{\widehat{\phi}_N\}$, and therefore the ASIP for ϕ follows from Theorem 6.4 applied to $\widehat{\phi}$.

As in Section 4(d), we can reduce the case of principal extensions to extensions of a basic hyperbolic set. Therefore, Theorem 6.4 holds also for principal extensions. This completes the proof of Theorem 2.5.

Finally, it is clear that, by taking the associated two-sided shift, the ASIP (together with all of its consequences) for extensions of two-sided subshifts implies the corresponding results for extensions of one-sided subshifts.

Remark 6.5 The CLT, WIP and upper LIL for one-sided subshifts can be proved directly within the abstract framework of Section 3. However, our proof of the ASIP (and the full LIL) for one-sided subshifts relies on the ASIP for two-sided subshifts and hence on the time-reversal step at the beginning of the proof of Lemma 6.1.

Similarly, our proof of the ASIP for two-sided subshifts relies on the fact that the class of such dynamical systems is closed under time-reversals. We note that the martingale approach of Conze and Le Borgne [8] (in the setting of geodesic flows on spaces of constant negative curvature) relies also on this time-reversal property [8, Remarques 2.7(1)].

Appendix A Hölder spaces

Suppose that $\sigma : X \to X$ is a two-sided subshift of finite type. If $f : X \to \mathbb{R}^n$ is continuous, we define for each $N \geq 1$, the N'th variation $\mathrm{var}_N f = \sup |f(x) - f(y)|$, where the supremum is over all $x, y \in X$ with $x_i = y_i$ for $|i| < N$. If $\theta \in (0,1)$, define $|f|_\theta = \sup_{N\geq 1} \mathrm{var}_N f/\theta^N$. Then $F_\theta(X, \mathbb{R}) = \{f \in C(X, \mathbb{R}^n) : |f|_\theta < \infty\}$ is a Banach space in the Hölder norm $\|f\|_\theta = |f|_\infty + |f|_\theta$.

Similarly, we define $F_\theta(X^+, \mathbb{R}^n)$ for the one-sided subshift $\sigma : X^+ \to X^+$ induced by the two-sided subshift σ. There is a natural embedding of $F_\theta(X^+, \mathbb{R}^n)$ inside $F_\theta(X, \mathbb{R}^n)$. Indeed, we can identify functions of $F_\theta(X^+, \mathbb{R}^n)$ with those functions in $F_\theta(X, \mathbb{R}^n)$ that depend only on future coordinates.

We also define the Hölder spaces $F_\theta(X, \mathbf{O}(n))$ as subspaces of $F_\theta(X, \mathrm{Mat}_n(\mathbb{R}))$.

Lemma A.1 *Let $\sigma : X \to X$ be a two-sided subshift. Suppose that $v \in F_\theta(X, \mathbb{R}^n)$ and $h \in F_\theta(X, \mathbf{O}(n))$, where $\theta \in (0, 1)$. Suppose further that h depends only on future coordinates.*

Then there exists $w \in F_{\theta^{1/2}}(X, \mathbb{R}^n)$ depending only on future coordinates and $\chi \in F_{\theta^{1/2}}(X, \mathbb{R}^n)$, such that

$$v = w + \chi - h \cdot \chi \circ \sigma.$$

Proof This result is a minor modification of [33, 5]. Our proof follows [26, Proposition 1.2]. Fix a map $\varphi : X \to X$ that depends only on future coordinates. Define

$$\chi(x) = \sum_{n=0}^{\infty} \big(h_n(x)v(\sigma^n x) - h_n(x)v(\sigma^n \varphi x)\big).$$

(This series converges, since h_n is orthogonal and $\mathrm{var}_n(v) \le |v|_\theta\, \theta^n$.)

Compute that $v = w + \chi - h \cdot \chi \circ \sigma$ where

$$w(x) = \sum_{n=0}^{\infty} \big(h_n(x)v(\sigma^n \varphi x) - h_{n+1}(x)v(\sigma^n \varphi \sigma x),$$

which clearly depends only on future coordinates.

It remains to show that χ (and hence w) lies in $F_{\theta^{1/2}}(X, \mathbb{R}^n)$. First note that

$$|\chi(x) - \chi(y)| \le A(x, y) + A(\varphi x, \varphi y) + B(x) + B(y),$$

where

$$A(x, y) = \sum_{n=0}^{N} |h_n(x)v(\sigma^n x) - h_n(y)v(\sigma^n y)|, \qquad B(x) = \sum_{n=N+1}^{\infty} |v(\sigma^n x) - v(\sigma^n \varphi x)|,$$

for all x, y.

We claim that there is a constant $K > 0$ such that (i) $B(x) \le K\theta^N$ for all $N \ge 1$ and $x \in X$, and (ii) $A(x, y) \le K\theta^N$ for all $N \ge 1$ and $x, y \in X$ with $x_i = y_i$ for $|i| \le 2N$. It then follows that $\mathrm{var}_{2N}(\chi) \le 4K\theta^N$ for all $N \ge 1$, proving the result.

Now, $|v(\sigma^n x) - v(\sigma^n \varphi x)| \le \mathrm{var}_n(v) \le |v|_\theta\, \theta^n$, for all $n \ge 1$. Hence

$$B(x) \le \frac{|v|_\theta\, \theta}{(1 - \theta)}\, \theta^N,$$

verifying (i).

Next, $A(x,y) \le \sum_{n=0}^N |v(\sigma^n x) - v(\sigma^n y)| + |v|_\infty \sum_{n=0}^N |h_n(x) - h_n(y)|$. Since $x_i = y_i$ for $|i| \le 2N$, we have that $(\sigma^j x)_i = (\sigma^j y)_i$ for $|i| \le 2N - j$. It follows that $|v(\sigma^n x) - v(\sigma^n y)| \le \mathrm{var}_{2N-n}(v) \le |v|_\theta\, \theta^{2N-n}$. Also

$$\begin{aligned}|h_n(x) - h_n(y)| &\le |h(x) - h(y)| + |h(\sigma x) - h(\sigma y)| + \cdots + |h(\sigma^{n-1}x) - h(\sigma^{n-1}y)| \\ &\le |h|_\theta(\theta^{2N} + \theta^{2N-1} + \cdots + \theta^{2N-n+1}) \\ &\le |h|_\theta\, \theta^{2N-n+1}/(1-\theta).\end{aligned}$$

Hence

$$A(x,y) \le \Big(\frac{|v|_\theta}{1-\theta} + \frac{|v|_\infty |h|_\theta\, \theta}{(1-\theta)^2}\Big)\theta^N,$$

verifying (ii). ∎

Appendix B Skorokhod representation

We will require the following embedding theorem for martingales which generalizes the Skorokhod representation theorem [34] for sums of independent random variables.

Theorem B.1 ([18, Theorem A.1]) *Let $\{S_N = \sum_{j=0}^{N-1} X_j,\ N \ge 1\}$ be a square-integrable ($\mathbb{R}$-valued) martingale. Then there exists a probability space supporting a Brownian motion W and a sequence of nonnegative variables $\tau_0, \tau_1, \ldots$ with the following properties:*

If $T_N = \sum_{j=0}^{N-1} \tau_j$, $W(T_N) = \sum_{j=0}^{N-1} Y_j$ and $\mathcal{G}_N$ is the σ-field generated by $\{W(T_j) : j = 0, \ldots, N-1\}$ and $\{W(t) : 0 \le t \le T_N\}$, then

(i) $\{S_N : N \ge 1\} =_d \{W(T_N) : N \ge 1\}$,

(ii) T_N is $\mathcal{G}_N$-measurable for each $N \ge 1$,

(iii) There is a constant $C > 0$ such that $E(\tau_N^2) \le CE(Y_N^4)$ for each $N \ge 1$, and

(iv) $E(\tau_N|\mathcal{G}_{N-1}) = E(Y_N^2|\mathcal{G}_{N-1})$ almost surely for each $N \ge 1$. ∎

Theorem B.1 was extended by Strassen [35] to obtain an ASIP for certain martingales. The error term is typically of the form $O(t^{1/2-\alpha})$ for sufficiently small $\alpha > 0$. In our context, we obtain an improved error term of the form $O(t^{1/4+\delta})$ for arbitrarily small $\delta > 0$. This is due to the fact that the martingales that we consider in this paper have the property that the partial sums of $X_j^2 - 1$ are as amenable as the partial sums of X_j. (Controlling $X_j^2 - 1$ is usually the difficult part of the procedure, see [28, p. 11].)

We require the following strong law of large numbers for martingales (see for example [12, p. 238]).

Proposition B.2 *Suppose that* $\{\sum_{j=0}^{N-1} R_j,\ N \geq 1\}$ *is a square integrable martingale and that the sequence* $\|R_j\|_2$ *is bounded. Then for any* $\delta > 0$, $\sum_{j=0}^{N-1} R_j = O(N^{1/2+\delta})$ *almost surely.* ∎

Theorem B.3 *Suppose that* $\{S_N = \sum_{j=0}^{N-1} X_j,\ N \geq 1\}$ *is a martingale and that*

(a) There is a constant $M > 0$ *such that* $\int X_j^4 < M$ *for all* j*, and*

(b) For any $\delta > 0$, $\sum_{j=0}^{N-1} X_j^2 - N = O(N^{1/2+\delta})$ *almost surely.*

Let W *and* T_N *be as in Theorem B.1. Then, for any* $\delta > 0$,

$$W(T_N) = W(N) + O(N^{1/4+\delta}) \quad \text{almost surely.}$$

Proof (Following [28].) Let $\delta > 0$. We claim that $T_N - N = O(N^{1/2+\delta})$ almost surely. It follows from the claim that almost surely

$$W(T_N) = W(N + O(N^{1/2+\delta})) = W(N) + O(N^{1/4+\delta'}).$$

It remains to prove the claim. Using property (iv), write

$$T_N - N = \sum_{j=0}^{N-1}(\tau_j - E(\tau_j|\mathcal{G}_{j-1})) + \sum_{j=0}^{N-1}(E(Y_j^2|\mathcal{G}_{j-1}) - Y_j^2) + \sum_{j=0}^{N-1} Y_j^2 - N.$$

By the definition of $\mathcal{G}_N$ and property (ii), the first and second terms are martingales. Set $R_j = E(Y_j^2|\mathcal{G}_{j-1}) - Y_j^2$. Then

$$\|R_j\|_2^2 \leq \|E(Y_j^2|\mathcal{G}_{j-1})\|_2^2 + \|Y_j^2\|_2^2 \leq 2\|Y_j^2\|_2^2 = 2\|X_j^2\|_2^2 < 2M.$$

Hence Proposition B.2 takes care of the second term. The first term is treated similarly, using property (iii). By property (i), the third term is equal in distribution to $\sum_{j=0}^{N-1} X_j^2 - N$ and hence is almost surely $O(N^{1/2+\delta})$ by (b). ∎

Acknowledgments IM is grateful to Matthew Nicol for valuable discussions and suggestions, and to Mark Pollicott for pointing out errors in an earlier version of this paper. The authors wish to thank the referee for several helpful comments.

References

[1] P. Ashwin, I. Melbourne and M. Nicol. Hypermeander of spirals; local bifurcations and statistical properties. *Physica D* **156** (2001), 364–382.

[2] P. Billingsley. The Lindeberg-Lévy theorem for martingales. *Proc. Amer. Math. Soc.* **12** (1961), 788–792.

[3] P. Billingsley. *Convergence of Probability Measures.* Wiley, New York, 1968.

[4] P. Billingsley. *Probability and Measure.* Wiley, New York, 1986.

[5] R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms.* Lecture Notes in Math. **470**, Springer, Berlin, 1975.

[6] G. E. Bredon. *Introduction to compact transformation groups.* Pure and Applied Mathematics **46**, Academic Press, New York and London, 1972.

[7] M. I. Brin. Topological transitivity of one class of dynamic systems and flows of frames on manifolds of negative curvature. *Funkts. Anal. Prilozh* **9** (1975), 9–19.

[8] J.-P. Conze and S. Le Borgne. Méthode de martingales et flow géodésique sur une surface de courbure constante négative. *Ergod. Th. & Dynam. Sys.* **21** (2001), 421–441.

[9] M. Csörgö and P. Révész. A new method to prove Strassen type laws of invariance principle, I. *Z. Wahrscheinlichkeitstheorie verw. Gebiete:* **31** (1975), 255–259.

[10] M. Denker and W. Philipp. Approximation by Brownian motion for Gibbs measures and flows under a function. *Ergod. Th. & Dynam. Sys.* **4** (1984), 541–552.

[11] D. Dolgopyat. On mixing properties of compact group extensions of hyperbolic systems. To appear in *Israel J. Math.*

[12] W. Feller. *An Introduction to Probability Theory and its Applications, II.* Wiley, New York, 1966.

[13] M. J. Field and M. Nicol. Ergodic Theory of Equivariant Diffeomorphisms I & II. Preprint, 2000.

[14] M. J. Field and V. Niţică. Stable topological transitivity of skew and principal extensions. *Nonlinearity* **14** (2001), 1055–1069.

[15] M. J. Field and W. Parry. Stable ergodicity of skew extensions by compact Lie groups. *Topology* **38** (1999), 167–187.

[16] K. Fukuyama. Function central limit theorem and Strassen's law of the iterated logarithms for weakly multiplicative systems. *J. Math. Kyoto. Univ.* **30** (1990), 625–635.

[17] M. I. Gordin. The central limit theorem for stationary processes. *Soviet Math. Dokl.* **10** (1969), 1174–1176.

[18] P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application.* Probability and Mathematical Statistics, Academic Press, New York, 1980.

[19] O. Kallenberg, *Foundations of Modern Probability.* Springer series in Statistics, Probability and its Applications, Springer, 1997.

[20] A. A. Kirillov. *Elements of the Theory of Representations.* Grundlehren **220**, Springer, Berlin, 1976.

[21] C. Liverani. Central limit theorems for deterministic systems. (F. Ledrappier, J. Lewowics, and S. Newhouse, eds.), Pitman Research Notes in Math. **363**, Longman Group Ltd, Harlow, 1996, 56–75.

[22] D. L. McLeish. Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** (1974), 620–628.

[23] M. Nicol, I. Melbourne, and P. Ashwin. Euclidean extensions of dynamical systems. *Nonlinearity* **14** (2001), 275–300.

[24] W. Parry. Skew products of shifts with a compact Lie group. *J. London Math. Soc.* **56** (1997), 395–404.

[25] W. Parry and M. Pollicott. The Chebotarev theorem for Galois coverings of Axiom A flows. *Ergod. Th. & Dynam. Sys.* **6** (1986), 133–148.

[26] W. Parry and M. Pollicott. *Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics.* Astérique **187-188**, Société Mathématique de France, Montrouge, 1990.

[27] W. Parry and M. Pollicott. Stability of mixing for toral extensions of hyperbolic systems. *Proc. Steklov Inst.* **216** (1997), 354–363.

[28] W. Philipp and W. F. Stout. *Almost Sure Invariance Principles for Partial Sums of Weakly Dependent Random Variables.* Memoirs of the Amer. Math. Soc. **161**, Amer. Math. Soc., Providence, RI, 1975.

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

[30] V. A. Rohlin. Exact endomorphisms of a Lebesgue space. *Izv. Akad. Nauk SSSR Ser. Mat.* **25** (1961), 499–530.

[31] D. J. Rudolph. Classifying the isometric extensions of a Bernoulli shift. *Journal d'Analyse Math.* **34** (1978), 36–60.

[32] D. Ruelle. *Thermodynamic Formalism.* Encyclopedia of Math. and its Applications **5**, Addison Wesley, Massachusetts, 1978.

[33] Y. G. Sinai. Gibbs measures in ergodic theory. *Russ. Math. Surv.* **27** (1972), 21–70.

[34] A. V. Skorokhod. *Studies in the Theory of Random Processes.* Addison Wesley, Reading, Massachusetts, 1965.

[35] V. Strassen. Almost sure behavior of sums of independent random variables and martingales. *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, **2** (1967), 315–343.

[36] S. Takahashi. Notes on the law of the iterated logarithm. *Studia Sci. Math. Hung.* **7** (1972), 21–24.