

INVARIANCE PRINCIPLES FOR ITERATED MAPS THAT CONTRACT ON AVERAGE

C. P. WALKDEN

ABSTRACT. We consider iterated function schemes that contract on average. Using a transfer operator approach, we prove a version of the almost sure invariance principle. This allows the system to be modelled by a Brownian motion, up to some error term. It follows that many classical statistical properties hold for such systems, such as the weak invariance principle and the law of the iterated logarithm.

1. INTRODUCTION AND STATEMENT OF RESULTS

1.1. Background. The study of the limiting behaviour of the sum of a sequence of observations or random variables is a key problem in dynamical systems and probability theory. For example, the ergodic theorem (alternatively, the strong law of large numbers) describes the average behaviour of such sums, and the central limit theorem describes the deviations of these sums from the average. One can then look for extensions of these results.

One such extension is the celebrated almost sure invariance principle [S1, S2]. This says the following: let X_j be a sequence of i.i.d. random variables with finite $(2+\delta)$ -moments and let $S_n = X_0 + \cdots + X_{n-1}$. Then there exists a Brownian motion B and a probability space Ω on which B and S_n can be redefined such that $S_n = B(n) + o(n^{1/2})$. This immediately allows one to deduce many classical statistical properties (such as the strong law of large numbers and various refinements of the central limit theorem) for S_n , given that they are known to hold for B . A general method for proving the almost sure invariance principle for sums of weakly dependent random variables can be found in [PS]; here the error term is of the form $O(n^{1/2-\delta})$ for some $\delta > 0$.

In the context of a dynamical system $T : X \rightarrow X$, it is natural to consider sums of the form $S_n = \sum_{j=0}^{n-1} fT^j$. The almost sure invariance principle in the case when T is a uniformly hyperbolic diffeomorphism or flow is now well understood [DK], again with an error term of the form $O(n^{1/2-\delta})$ for some $\delta > 0$.

More recently, the almost sure invariance principle has been re-examined for a variety of partially hyperbolic maps [FMT, MT] and non-uniformly hyperbolic maps with indifferent fixed points [PoS] using the spectral properties of a transfer operator. The existence of strong spectral properties (in particular a spectral gap) allows one to deduce an improved error term of the form $O(n^{1/4+\varepsilon})$ for any $\varepsilon > 0$.

Received by the editors March 19, 2003 and, in revised form, November 25, 2004.

2000 *Mathematics Subject Classification.* Primary 60F17; Secondary 37H99, 37A50.

©2006 American Mathematical Society
 Reverts to public domain 28 years from publication

The purpose of this note is to prove an almost sure invariance principle with an error term of the form $O(n^{1/4+\varepsilon})$ for any $\varepsilon > 0$ for a sequence of sums of observations arising from an iterated function scheme where the maps satisfy a notion of “contraction on average”, a form of non-uniform hyperbolicity.

1.2. Statement of results. Let (X, d) be a metric space. Consider a family of Lipschitz maps $T_j : X \rightarrow X$, $j = 1, \dots, M$. We are interested in studying the statistical properties of the iterated function scheme formed by applying the maps T_j chosen at random according to a Markov transition probability.

Let $p_j : X \rightarrow [0, 1]$ be continuous maps such that $\sum_{j=1}^M p_j(x) = 1$ for all $x \in X$. Define a Markov transition probability by

$$p(x, A) = \sum_{j=1}^M p_j(x) \chi_A(T_j x)$$

for each Borel subset $A \subset X$. (Here χ_A denotes the characteristic function of A .)

We say that the system *contracts on average* if there exists $r \in (0, 1)$ such that for all $x_1, x_2 \in X$ we have

$$(1) \quad \prod_{j=1}^M d(T_j x_1, T_j x_2)^{p_j(x_1)} \leq r d(x_1, x_2).$$

Examples of such systems include certain affine systems $T_j : \mathbb{R}^d \rightarrow \mathbb{R}^d : x \mapsto A_j x + b_j$ where $A_j \in \text{Gl}(d, \mathbb{R})$ and $b_j \in \mathbb{R}^d$; see §2.4. Other examples and applications are discussed in [DF].

We also assume that the $p_j > 0$ are continuous and satisfy a Dini condition (cf. [E1]).

With these assumptions, it is known [BDEG] that there exists a unique attractive stationary Borel probability measure ν on X , i.e. for all Borel sets A

$$(2) \quad \int p(x, A) d\nu(x) = \nu(A).$$

Let $\Sigma = \{i = (i_0, i_2, \dots) \mid 1 \leq i_j \leq M\}$ denote the one-sided full-shift on M symbols. For each $x \in X$ we define a probability measure μ_x on Σ by defining μ_x on cylinder sets by

$$\mu_x[i_0, i_2, \dots, i_{n-1}] = p_{i_0}(x) p_{i_1}(T_{i_0} x) \cdots p_{i_{n-1}}(T_{i_{n-2}} \cdots T_{i_0} x).$$

For each $x \in X$ and $i \in \Sigma$ we define

$$Z_k(x, i) = T_{i_{k-1}} \cdots T_{i_0}(x).$$

Then $Z_k(x, i)$ is an X -valued Markov chain with respect to μ_x , with initial state x and transition probability p .

We can relate μ_x with ν as follows. Define $\pi_x(i) = \lim_{k \rightarrow \infty} T_{i_0} T_{i_1} \cdots T_{i_k}(x)$ for μ_x -a.e. $i \in \Sigma$. Then for all $x \in X$ we have $\pi_x^* \mu_x = \nu$. See, for example, [E2].

Let $f : X \rightarrow \mathbb{R}$ be a bounded and continuous function on X . We are interested in the distribution of the sequence of observations

$$f^n(x, i) = \sum_{k=0}^{n-1} f(Z_k(x, i)).$$

It is known [E1] that f^n satisfies a pointwise ergodic theorem (or strong law of large numbers): for all $x \in X$ and μ_x -a.e. $i \in \Sigma$,

$$(3) \quad \frac{1}{n} f^n(x, i) \rightarrow \nu(f).$$

A central limit theorem is also known [Pe]. Let $f : X \rightarrow \mathbb{R}$ be a bounded Lipschitz function and fix $x \in X$. Then

$$(4) \quad \frac{1}{\sqrt{n}} f^n(x, \cdot) \rightarrow_d \mathcal{N}(\nu(f), \sigma^2(f)),$$

provided that the variance $\sigma^2(f) \neq 0$. Here $\mathcal{N}(\nu(f), \sigma^2(f))$ denotes the normal distribution with mean $\nu(f)$ and variance $\sigma^2(f)$ and \rightarrow_d denotes convergence in distribution. The variance is given by

$$\sigma^2(f) = \lim_{n \rightarrow \infty} \frac{1}{n} \int (f^n(x, i) - n\nu(f))^2 d\mu_x(i).$$

This quantity is independent of the choice of origin x . Error terms and estimates on the rate of convergence in (4) (Berry-Esseen bounds) can be found in [Po].

The purpose of this note is to prove more refined statistical properties of f^n . In particular, we deduce the following results. (In the statements below, $C([0, 1], \mathbb{R})$ denotes the space of continuous functions on $[0, 1]$, and $D([0, 1], \mathbb{R})$ denotes the space of right-continuous functions on $[0, 1]$ that have left limits.)

Proposition 1.1 (Weak invariance principle). *Let $f : X \rightarrow \mathbb{R}$ be a bounded Lipschitz function. Suppose that $\nu(f) = 0$ and $\sigma^2(f) > 0$. Fix $x \in X$. For each $n > 0$ define the random function $\zeta_{n,x} : \Sigma \rightarrow D([0, 1], \mathbb{R})$ by defining*

$$\zeta_{n,x}(i)(t) = \frac{f^{[nt]}(x, i)}{\sigma(f)\sqrt{n}}.$$

Then for every $x \in X$ and μ_x -a.e. $i \in \Sigma$, the sequence of measures $\zeta_{n,x}^ \mu_x$ converges weakly as $n \rightarrow \infty$ to the standard Wiener measure on $C([0, 1], \mathbb{R})$.*

The weak invariance principle is also known as the functional central limit theorem. It immediately implies the central limit theorem (set $t = 1$).

Proposition 1.2 (Functional law of the iterated logarithm). *Let $f : X \rightarrow \mathbb{R}$ be a bounded Lipschitz function. Suppose that $\nu(f) = 0$ and $\sigma^2(f) > 0$. Fix $x \in X$. For each $n > 0$ define the random function $\xi_{n,x} : \Sigma \rightarrow D([0, 1], \mathbb{R})$ by defining*

$$\xi_{n,x}(i)(t) = \frac{1}{\sigma(f)\sqrt{2n \log \log n}} f^{[nt]}(x, i).$$

Then for each $x \in X$, $\{\xi_{n,x} \mid n \in \mathbb{N}\}$ has uniformly compact closure in $D([0, 1], \mathbb{R})$. Moreover, the closure is equal to the set

$$\left\{ \xi : [0, 1] \rightarrow \mathbb{R} \mid \xi \text{ is absolutely continuous, } \xi(0) = 0, \int_0^1 \xi'(s)^2 ds \leq 1 \right\}.$$

The FLIL immediately implies the law of the iterated logarithm (set $t = 1$).

Corollary 1.3 (Law of the iterated logarithm). *Let $f : X \rightarrow \mathbb{R}$ be a bounded Lipschitz function. Suppose that $\nu(f) = 0$ and $\sigma^2(f) > 0$. Then for all $x \in X$ and μ_x -a.e. $i \in \Sigma$,*

$$\limsup_{n \rightarrow \infty} \frac{f^n(x, i)}{\sigma(f)\sqrt{2n \log \log n}} = 1.$$

The above results are well-known corollaries of an almost sure invariance principle (Theorem 1.4 below). Other standard corollaries of the almost sure invariance principle are discussed in more detail in [HH].

Theorem 1.4 (Almost sure invariance principle). *Suppose that $f : X \rightarrow \mathbb{R}$ is a bounded Lipschitz function and is such that $\nu(f) = 0$ and $\sigma^2 = \sigma^2(f) > 0$. Fix $x \in X$. Then there exists a probability space (Ω, \mathcal{F}, P) and a one-dimensional Brownian motion $W : \Omega \rightarrow C(\mathbb{R}^+, \mathbb{R})$ such that the random variable $\omega \mapsto W(\omega)(t)$ has variance $\sigma^2 t$, and sequences of random variables $\phi_{n,x} : \Sigma \rightarrow \mathbb{R}$, $\psi_{n,x} : \Omega \rightarrow \mathbb{R}$ with the following properties:*

(i) *for all $\varepsilon > 0$, we have*

$$f^n(x, \cdot) = \phi_{n,x}(\cdot) + O(n^{\frac{1}{4}+\varepsilon}) \mu_x\text{-a.e.};$$

(ii) *the sequences $\{\phi_{n,x}(\cdot)\}_n$ and $\{\psi_{n,x}(\cdot)\}_n$ are equal in distribution;*

(iii) *for all $\varepsilon > 0$, we have*

$$(5) \quad \psi_{n,x}(\cdot) = W(\cdot)(n) + O(n^{\frac{1}{4}+\varepsilon}) P\text{-a.e.}$$

Remark. Theorem 1.4 holds for a more general class of observable f . See §4.

This note is organised as follows. In §2 we remark on some generalisations and examples of the set-up described above. The Skorokhod Embedding Theorem, a key tool in our analysis, is stated in §3. In §4 we define a transfer operator acting on a suitable family of functions spaces and state its spectral properties. The sequence of observations f^n is reduced to a martingale in §5. Following some moment estimates in §6, we prove Theorem 1.4(i), (ii). The error term in Theorem 1.4(iii) is estimated in §7.

2. EXAMPLES, REMARKS AND GENERALISATIONS

2.1. Generalisation to countably many transformations. In §1 we stated our results for iterated function schemes consisting of finitely many transformations T_j . Our results continue to hold in the more general case of countably many transformations provided that one makes the following additional assumptions:

(i)

$$\sup_{x,y,z \in X, y \neq z} \sum_{j=0}^{\infty} \frac{d(T_j y, T_j z)}{d(y, z)} p_j(x) < \infty,$$

(ii) *for some (hence any) choice of x_0*

$$\sup_{x,y \in X} \sum_{j=0}^{\infty} \frac{d(T_j y, x_0)}{1 + d(y, x_0)} p_j(x) < \infty,$$

(iii) *for some (hence any) choice of x_0*

$$\sup_{x \in X} \sum_{j=0}^{\infty} \left(\frac{d(T_j x, x_0)}{1 + d(x, x_0)} \left(\sup_{y,z \in X, y \neq z} \frac{|p_j(y) - p_j(z)|}{d(y, z)} \right) \right).$$

The above conditions are required to ensure that a suitably defined transfer operator (see §4.2) has the appropriate spectral properties [Pe]. Note that the above conditions are trivially satisfied in the case of finitely many T_j s.

2.2. Generalisation to non-negative probabilities. In §1 we assumed that the probabilities p_j are strictly positive. In addition to the generalisations given in §2.1, we can weaken this to $p_j(x) \geq 0$, provided that we assume the following: for all $x, y \in X$ there exists $i, j \in \Sigma$ such that

$$\lim_{n \rightarrow \infty} d(Z_n(x, i), Z_n(y, j))(1 + d(Z_n(y, j), x_0)) = 0$$

with $p_{i_n}(Z_n(x, i)) > 0$ and $p_{j_n}(Z_n(y, j)) > 0$ for all $n \geq 1$. Again, these conditions are required to ensure that a certain transfer operator has the appropriate spectral properties [Pe].

2.3. Recoding. It may happen that a given system fails to satisfy (1) but does satisfy this condition after it has been recoded. For example, take $X = \mathbb{R}^2$ and $T_1(x, y) = (5x/4, y/4)$ and $T_2(x, y) = (x/4, 5y/4)$, both chosen with probability $1/2$. This system does not satisfy (1). However, the recoded system $T_{i,j} = T_i \circ T_j$, $i, j \in \{1, 2\}$, each chosen with probability $1/4$, does satisfy (1). Hence Theorem 1.4 holds for the recoded system. It is then simple to see that the conclusions of Theorem 1.4 hold for the original system. Note that neither T_1 nor T_2 are contractive.

2.4. Example: Affine maps. Specific examples of systems satisfying our hypotheses can be given by using affine maps; see [Pe, DF]. Let $A_j \in \text{Gl}(d, \mathbb{R})$ and $b_j \in \mathbb{R}^d$. Define

$$T_j : \mathbb{R}^d \rightarrow \mathbb{R}^d : x \mapsto A_j x + b_j.$$

In this setting, the contraction on average condition (1) becomes

$$\sup_{x \in \mathbb{R}^d} \sum_j \|A_j\| p_j(x) < 1,$$

where $\|A\| = \sup_{x \neq 0} \|Ax\|/\|x\|$ denotes the matrix norm of A .

3. THE SKOROKHOD EMBEDDING THEOREM

Let (Ω, \mathcal{F}, P) be a probability space.

Definition. A stochastic process $W : \Omega \rightarrow C(\mathbb{R}^+, \mathbb{R})$ is called a *Brownian motion* (with mean zero and variance $\sigma^2 > 0$) if:

- (i) for P -a.e. $\omega \in \Omega$, we have $W(\omega)(0) = 0$;
- (ii) there exists $\sigma^2 > 0$ such that for each $t > 0$, the random variable

$$\omega \mapsto W(\omega)(t) : \Omega \rightarrow \mathbb{R}$$

is normally distributed with mean zero and variance $\sigma^2 t$;

- (iii) for all $t_0 < t_1 < \dots < t_n$ the random variables

$$\omega \mapsto W(\omega)(t_i) - W(\omega)(t_{i-1})$$

are independent.

Suppose that \mathcal{F}_n is an increasing sequence of sub- σ -algebras of \mathcal{F} .

Definition. A sequence of random variables S_n is called a *martingale* if, for each n , S_n is \mathcal{F}_n -measurable and $\mathbb{E}(S_n | \mathcal{F}_{n-1}) = S_{n-1}$ a.s. We call $X_n = S_n - S_{n-1}$ a *martingale difference*. Setting $X_0 = S_0$ we write $S_n = \sum_{j=0}^n X_j$.

The following standard result [HH, p. 269] allows us to model a martingale by a Brownian motion.

Proposition 3.1 (Skorokhod Embedding Theorem). *Let $\{\phi_n = \sum_{j=0}^{n-1} X_j, \mathcal{B}_n\}$ be a square integrable, zero mean martingale on a probability space (X, \mathcal{B}, μ) . Then there exists a probability space (Ω, \mathcal{F}, P) , a standard Brownian motion $W : \Omega \rightarrow C(\mathbb{R}^+, \mathbb{R})$ and a sequence of non-negative random variables τ_0, τ_1, \dots with the following properties: Let $T_n = \sum_{j=0}^{n-1} \tau_j$, $\psi_n = W(T_n) : \Omega \rightarrow \mathbb{R}$, and denote by \mathcal{G}_n the σ -algebra generated by $\psi_0, \dots, \psi_{n-1}, W(t), 0 \leq t \leq T_n$. Then:*

- (i) *the sequences $\{\phi_n\}_n$ and $\{\psi_n\}_n$ are equal in distribution;*
- (ii) *T_n is \mathcal{G}_n -measurable;*
- (iii) *$\mathbb{E}(\tau_n | \mathcal{G}_{n-1}) = \mathbb{E}((\psi_n - \psi_{n-1})^2 | \mathcal{G}_{n-1})$ a.s.;*
- (iv) *for each $r > 1$ there exists a constant $C_r < \infty$ such that $\mathbb{E}(\tau_n^r | \mathcal{G}_{n-1}) \leq C_r \mathbb{E}((\psi_n - \psi_{n-1})^{2r} | \mathcal{G}_{n-1})$ a.s.*

4. FUNCTION SPACES AND TRANSFER OPERATORS

4.1. Function spaces. We will be interested in the following family of function spaces. Fix a choice of origin $x_0 \in X$ (the exact choice of x_0 is immaterial). Let $\alpha, \beta \in (0, 1)$ and define $a(t) = 1 + t^\alpha$, $b(t) = t^\beta$. Define

$$C_{a,b}(X, \mathbb{R}) = \{f \in C(X, \mathbb{R}) \mid \|f\|_{a,b} < \infty\},$$

where $\|f\|_{a,b} = |f|_a + |f|_{a,b}$ and

$$\begin{aligned} |f|_a &= \sup_{x \in X} \frac{|f(x)|}{a(d(x, x_0))}, \\ |f|_{a,b} &= \sup_{x, y \in X, x \neq y} \frac{|f(x) - f(y)|}{a(d(x, x_0))b(d(x, y))}. \end{aligned}$$

The following is easily proved from the definitions.

Lemma 4.1. *Let $f_j \in C_{a,b}(X, \mathbb{R})$ for $j = 1, 2, \dots, \ell$. Then $\prod_{j=1}^\ell f_j \in C_{a^\ell, b+b^\ell}(X, \mathbb{R})$. Moreover,*

$$(6) \quad \|f_1 f_2\|_{a^2, b+b^2} \leq \|f_1\|_{a,b} \|f_2\|_{a,b}.$$

We shall use the following simple observation in what follows.

Lemma 4.2. *Let a and b be as above. Let $f : X \rightarrow \mathbb{R}$ be a bounded Lipschitz function. Then $f \in C_{a,b}(X, \mathbb{R})$.*

Proof. Clearly $|f|_a = \sup_x |f(x)|/a(d(x, x_0)) \leq |f|_\infty$, where $|f|_\infty = \sup_{x \in X} |f(x)|$. Similarly,

$$\begin{aligned} &|f|_{a,b} \\ &\leq \sup_{x, y \in X, d(x, y) \leq 1} \frac{|f(x) - f(y)|}{a(d(x, x_0))b(d(x, y))} + \sup_{x, y \in X, d(x, y) \geq 1} \frac{|f(x) - f(y)|}{a(d(x, x_0))b(d(x, y))} \\ &\leq \sup_{x, y \in X, d(x, y) \leq 1} \frac{|f|_{\text{Lip}} d(x, y)}{a(d(x, x_0))b(d(x, y))} + \sup_{x, y \in X, d(x, y) \geq 1} \frac{|f(x)| + |f(y)|}{a(d(x, x_0))b(d(x, y))} \\ &\leq |f|_{\text{Lip}} + 2|f|_\infty < \infty. \end{aligned}$$

□

4.2. Transfer operators. Define the operator $P : C(X, \mathbb{R}) \rightarrow C(X, \mathbb{R})$ by

$$Pf(x) = \sum_{j=1}^M p_j(x) f(T_j x).$$

It follows that

$$\begin{aligned} P^n f(x) &= \sum_{i_{n-1}, \dots, i_0} p_{i_0}(x) p_{i_1}(T_{i_0} x) \cdots p_{i_{n-1}}(T_{i_{n-2}} \cdots T_{i_0} x) f(T_{i_{n-1}} \cdots T_{i_0} x) \\ (7) \quad &= \int_{\Sigma} f(Z_n(x, i)) d\mu_x(i). \end{aligned}$$

More generally, let \mathcal{F}_k denote the sub- σ -algebra of Σ generated by the cylinders of length k . Then

$$(8) \quad Pf(Z_{n-1}(x, \cdot)) = \mathbb{E}_x(f(Z_n(x, \cdot)) \mid \mathcal{F}_{n-1}),$$

where \mathbb{E}_x denotes the conditional expectation on $L^1(\Sigma, \mu_x)$. As x is fixed, we will often write fZ_k for $f(Z_k(x, \cdot))$.

The following result, ensuring that on suitable function spaces P has a spectral gap, will be a key tool in our analysis.

Proposition 4.3 ([Pe]). *Let $a(t) = 1 + t^\alpha$, $b(t) = t^\beta$ where $\alpha, \beta \in (0, 1/2)$ and $\beta < \alpha$. Then:*

- (i) *the operator P maps $C_{a,b}(X, \mathbb{R}) \rightarrow C_{a,b}(X, \mathbb{R})$;*
- (ii) *the operator P , when restricted to $C_{a,b}(X, \mathbb{R})$, has 1 as a simple maximal eigenvalue with an associated eigenprojection ν ;*
- (iii) *the remainder of the spectrum lies inside a disc of radius $\rho = \rho(a, b) \leq r^\alpha < 1$ (where r is as in (1)).*

Remarks. 1. Thus we can write $P = \nu + Q$, where $Q : C_{a,b}(X, \mathbb{R}) \rightarrow C_{a,b}(X, \mathbb{R})$ has spectral radius at most ρ . Moreover, we have that $\nu Q = Q\nu = 0$ so that $P^n = \nu + Q^n$.

2. The eigenprojection ν corresponds to the stationary probability measure defined in (2).

Suppose that we are in the simpler case of a place-independent probability, that is, $p_i(x) = p_i$ is independent of x for each i . Then we can instead work on the simpler function space

$$C^\alpha(X, \mathbb{R}) = \{f : X \rightarrow \mathbb{R} \mid \|f\|_\alpha = |f| + |f|_{(\alpha)} < \infty\},$$

where

$$\begin{aligned} |f| &= \sup_{x \in X} \frac{|f(x)|}{1 + d(x, x_0)}, \\ |f|_{(\alpha)} &= \sup_{x, y \in X, x \neq y} \frac{|f(x) - f(y)|}{d(x, y)^\alpha}. \end{aligned}$$

This space is invariant under the transfer operator P . It is shown in [Pe] that if α is sufficiently small, then P is quasi-compact and the conclusions of Proposition 4.3 hold.

5. REDUCTION TO A MARTINGALE

Let $f : X \rightarrow \mathbb{R}$ be a bounded Lipschitz function. We assume that $\nu(f) = 0$. In [Pe] it is shown that $a, b : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ can be chosen as in §4.1 in such a way that P acts on the different spaces $C_{a,b}(X, \mathbb{R})$, $C_{a^2, b+b^2}(X, \mathbb{R})$ and $C_{a^4, b+b^4}(X, \mathbb{R})$ and has a spectral gap.

Let

$$w = \sum_{k=0}^{\infty} P^k f = \sum_{k=0}^{\infty} Q^k f \in C_{a,b}(X, \mathbb{R}).$$

Then

$$w - Pw = \sum_{k=0}^{\infty} P^k f - \sum_{k=1}^{\infty} P^k f = f.$$

Recalling that $f^n(x, i) = \sum_{k=0}^{n-1} fZ_k(x, i)$ we see that

$$\begin{aligned} f^n(x, i) &= \sum_{k=0}^{n-1} wZ_k(x, i) - PwZ_k(x, i) \\ &= \sum_{k=1}^{n-1} (wZ_k(x, i) - PwZ_{k-1}(x, i)) + w(x) - PwZ_{n-1}(x, i) \\ &= \sum_{k=1}^{n-1} u_k(x, i) + w(x) - PwZ_{n-1}(x, i), \end{aligned}$$

where we have set $u_k(x, i) = wZ_k(x, i) - PwZ_{k-1}(x, i)$. Let

$$\phi_{n,x}(i) = \sum_{k=1}^{n-1} u_k(x, i)$$

so that

$$f^n(x, i) = \phi_{n,x}(i) + w(x) - PwZ_{n-1}(x, i).$$

Proposition 5.1. *Let $\varepsilon > 0$. Then for each $x \in X$*

$$f^n(x, i) = \phi_{n,x}(i) + O(n^{\frac{1}{4}+\varepsilon})$$

for μ_x -a.e. $i \in \Sigma$.

Proof. For each $n > 0$ let

$$\Sigma_{n,x} = \{i \in \Sigma \mid PwZ_{n-1}(x, i) > n^{\frac{1}{4}+\varepsilon}\}.$$

To prove the proposition it is sufficient to prove that μ_x -a.e. $i \in \Sigma$ is in at most finitely many $\Sigma_{n,x}$. By the Borel-Cantelli lemma it is sufficient to prove that $\sum_n \mu_x(\Sigma_{n,x}) < \infty$. As $Pw \in C_{a,b}(X, \mathbb{R})$ we have that $(Pw)^4 \in C_{a^4, b+b^4}(X, \mathbb{R})$,

a space on which P acts and has a spectral gap. Now

$$\begin{aligned}
\mu_x(\Sigma_{n,x}) &= \int_{\Sigma_{n,x}} 1 d\mu_x \\
&\leq \int_{\Sigma_{n,x}} \left(\frac{PwZ_{n-1}}{n^{\frac{1}{4}+\varepsilon}} \right)^4 d\mu_x \\
&\leq \frac{1}{n^{1+4\varepsilon}} \int_{\Sigma_{n,x}} (Pw)^4 Z_{n-1} d\mu_x \\
&= \frac{1}{n^{1+4\varepsilon}} P^{n-1}((Pw)^4)(x) \\
&\leq \frac{1}{n^{1+4\varepsilon}} (\nu((Pw)^4) + Q^{n-1}((Pw)^4))(x) \\
&\leq \text{Const} \frac{1}{n^{1+4\varepsilon}},
\end{aligned}$$

where the constant is independent of n . As this series is summable, the result follows. \square

The following is the key observation in our analysis.

Proposition 5.2. *The sequence $\phi_{n,x}(\cdot)$ is a square-integrable, zero-mean martingale on Σ with respect to the filtration $\{\mathcal{F}_n\}$, where \mathcal{F}_n denotes the σ -algebra of cylinders of length n .*

Proof. To see that $\phi_{n,x}$ is a martingale, simply observe that

$$\begin{aligned}
u_k(x, i) &= wZ_k(x, i) - PwZ_{k-1}(x, i) \\
(9) \quad &= wZ_k(x, i) - \mathbb{E}_x(wZ_k \mid \mathcal{F}_{k-1})(x, i)
\end{aligned}$$

by (8). Hence

$$\begin{aligned}
\mathbb{E}_x(u_k(x, i) \mid \mathcal{F}_{k-1}) &= \mathbb{E}_x(wZ_k - \mathbb{E}_x(wZ_k \mid \mathcal{F}_{k-1}) \mid \mathcal{F}_{k-1}) \\
&= \mathbb{E}_x(wZ_k \mid \mathcal{F}_{k-1}) - \mathbb{E}_x(wZ_k \mid \mathcal{F}_{k-1}) \\
&= 0
\end{aligned}$$

so that u_k is a martingale difference operator and $\phi_{n,x}$ is a martingale. Moreover, integrating (9) with respect to μ_x shows that $\int u_k(x, i) d\mu_x(i) = 0$. Hence $\phi_{n,x}$ has zero mean. That $\phi_{n,x}$ is square integrable follows from Proposition 6.2 below. \square

6. MOMENT ESTIMATES

Here we collect several estimates that will be useful below. We begin with the following observation.

Lemma 6.1. *Let $f, g \in L^1(X, \nu)$. Then*

$$\int fZ_{n+m}(x, i)gZ_n(x, i) d\mu_x(i) = (P^n(P^m f \cdot g))(x).$$

Proof. Simply observe that

$$\begin{aligned}
 & \int f Z_{n+m}(x, i) g Z_n(x, i) d\mu_x(i) \\
 &= \sum_{i_0, \dots, i_{n-1}} p_{i_0}(x) \cdots p_{i_{n-1}}(Z_{n-1}(x, i)) \\
 & \quad \times \left(\sum_{i_n, \dots, i_{n+m-1}} p_{i_n}(Z_n(x, i)) \cdots p_{i_{n+m-1}}(Z_{n+m-1}(x, i)) f(Z_{n+m}(x, i)) \right) \\
 & \quad \times g(Z_n(x, i)) \\
 &= \sum_{i_0, \dots, i_{n-1}} p_{i_0}(x) \cdots p_{i_{n-1}}(Z_{n-1}(x, i)) (P^m f) Z_n(x, i) g Z_n(x, i) \\
 &= (P^n(P^m f)g)(x).
 \end{aligned}$$

□

We need the following estimate.

Proposition 6.2. *The variances $\sigma^2(\phi_{n,x})$ of $\phi_{n,x}$ with respect to the measure μ_x satisfy*

$$\sigma^2(\phi_{n,x}) = n\sigma^2(f) + O(1).$$

Proof. Observe that $w^2, (Pw)^2 \in C_{a^2, b+b^2}(X, \mathbb{R})$, a space on which P acts and has a spectral gap. Recall that $\phi_{n,x}$ has zero mean. Also recall that the u_k are orthogonal as they are martingale differences. Hence

$$\sigma^2(\phi_{n,x}) = \sum_{k=1}^n \int u_k^2(x, i) d\mu_x(i).$$

Now

$$\begin{aligned}
 & \int u_k^2(x, i) d\mu_x(i) \\
 &= \int w^2 Z_k(x, i) - 2w Z_k(x, i) Pw Z_{k-1}(x, i) + (Pw)^2 Z_{k-1}(x, i) d\mu_x(i).
 \end{aligned}$$

By (7) and Lemma 6.1, we have that

$$(10) \quad \int u_k^2(x, i) d\mu_x(i) = P^k(w^2)(x) - P^{k-1}((Pw)^2)(x)$$

so that

$$\begin{aligned}
 \sigma^2(\phi_{n,x}) &= \sum_{k=1}^n (P^k(w^2) - P^{k-1}((Pw)^2))(x) \\
 &= n(\nu(w^2) - \nu((Pw)^2)) + \sum_{k=1}^n Q^k(w^2)(x) - Q^{k-1}((Pw)^2)(x).
 \end{aligned}$$

Proposition 4.3 implies

$$\left| \sum_{k=1}^n Q^k(w^2)(x) - Q^{k-1}((Pw)^2)(x) \right| \leq \text{Const} \sum_{k=1}^{\infty} \rho^k < \infty.$$

A calculation, using the fact that $w = \sum_{k=0}^{\infty} P^k f$, shows that $\nu(w^2) - \nu((Pw)^2) = \sigma^2(f)$. □

Remark. One can easily show that the degenerate case $\sigma^2(f) = 0$ holds if and only if there exists $u : X \rightarrow \mathbb{R}$ such that, for each j , $fT_j = uT_j - u$ ν -a.e. [Pe].

Remark. The above calculation immediately implies that $\phi_{n,x}(\cdot) \in L^2(\Sigma, \mu_x)$. By Proposition 3.1, we can find an abstract probability space (Ω, \mathcal{F}, P) , a Brownian motion $W : \Omega \rightarrow C(\mathbb{R}^+, \mathbb{R})$ and a sequence of stopping times T_n such that if we define $\psi_{n,x} = W(T_n)$, then the sequences $\{\phi_{n,x}\}$ and $\{\psi_{n,x}\}$ are equal in distribution. This immediately implies part (ii) of Theorem 1.4.

We will also need the following fourth moment estimate. Here and throughout, u_k denotes $u_k(x, \cdot)$.

Lemma 6.3. *There exists $K > 0$ such that $\mathbb{E}_x(u_k^4) \leq K < \infty$ for all $k \geq 1$.*

Proof. Observe that

$$\begin{aligned} \mathbb{E}_x(u_k^4) &= \mathbb{E}_x(w^4 Z_k(x, i) + 4w^3 Z_k(x, i)(Pw)Z_{k-1}(x, i) \\ &\quad + 6w^2 Z_k(x, i)(Pw)^2 Z_{k-1}(x, i) \\ &\quad + 4w Z_k(x, i)(Pw)^3 Z_{k-1}(x, i) + (Pw)^4 Z_{k-1}(x, i)) \\ &= P^{k-1}(P(w^4) + 4P(w^3)Pw + 6P(w^2)(Pw)^2 + 5(Pw)^4) \\ &\leq \nu(P(w^4) + 4P(w^3)Pw + 6P(w^2)(Pw)^2 + 5(Pw)^4) + \text{Const } \rho^k. \end{aligned}$$

□

7. ESTIMATION OF THE ERROR TERM

To estimate the error term and to prove part (iii) of Theorem 1.4, we follow an approach outlined in [PS]. Let $\sigma^2 = \sigma^2(f) = \nu(w^2) - \nu((Pw)^2)$. Let W be a standard Brownian motion satisfying the conclusions of Proposition 3.1. We estimate

$$\begin{aligned} T_n - \sigma^2 n &= \sum_{k=0}^{n-1} \tau_k - \mathbb{E}(\tau_k \mid \mathcal{G}_{k-1}) \\ &\quad + \sum_{k=0}^{n-1} \mathbb{E}(\tau_k \mid \mathcal{G}_{k-1}) - \mathbb{E}((\psi_k - \psi_{k-1})^2 \mid \mathcal{G}_{k-1}) \\ &\quad + \sum_{k=0}^{n-1} \mathbb{E}((\psi_k - \psi_{k-1})^2 \mid \mathcal{G}_{k-1}) - (\psi_k - \psi_{k-1})^2 \\ &\quad + \sum_{k=0}^{n-1} (\psi_k - \psi_{k-1})^2 - \mathbb{E}_x(u_k^2) \\ &\quad + \sum_{k=0}^{n-1} \mathbb{E}_x(u_k^2) - \sigma^2 n \\ &= \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)} + \text{(V)}. \end{aligned}$$

By Proposition 3.1(iii), we have that (II) = 0. By Proposition 6.2, (V) = $O(1)$.

We will then need the following lemma, which is a special case of a version of the Kronecker lemma for martingales; see [F, p. 243].

Lemma 7.1. *Let $S_n = X_0 + \cdots + X_n$ be a zero mean martingale. Suppose that for some K , we have that $\mathbb{E}(X_n^2) \leq K < \infty$ for all $n \geq 0$. Then for any $\varepsilon > 0$ we have that*

$$\frac{S_n}{n^{\frac{1}{2}+\varepsilon}} \rightarrow 0 \text{ a.s.}$$

We claim that both (I) and (III) are $O(n^{\frac{1}{2}+\varepsilon})$ a.s.

Lemma 7.2. *Let $x \in X$ and $\varepsilon > 0$. Then*

$$(I) = \sum_{k=0}^{n-1} \tau_k - \mathbb{E}(\tau_k \mid \mathcal{G}_{k-1}) = O(n^{\frac{1}{2}+\varepsilon}) \text{ a.s.}$$

Proof. Let $X_k = \tau_k - \mathbb{E}(\tau_k \mid \mathcal{G}_{k-1})$. Proposition 3.1(ii) implies that $\tau_k - \mathbb{E}(\tau_k \mid \mathcal{G}_{k-1})$ is \mathcal{G}_k -measurable and that

$$\mathbb{E}(\tau_k - \mathbb{E}(\tau_k \mid \mathcal{G}_{k-1}) \mid \mathcal{G}_{k-1}) = \mathbb{E}(\tau_k \mid \mathcal{G}_{k-1}) - \mathbb{E}(\tau_k \mid \mathcal{G}_{k-1}) = 0.$$

Hence the summand in (I) is a martingale difference.

Note that

$$\begin{aligned} \mathbb{E}((\tau_k - \mathbb{E}(\tau_k \mid \mathcal{G}_{k-1}))^2) &\leq \mathbb{E}(\tau_k^2) \\ &\leq \mathbb{E}(\mathbb{E}(\tau_k^2 \mid \mathcal{G}_{k-1})) \\ &\leq C_2 \mathbb{E}(\mathbb{E}((\psi_k - \psi_{k-1})^4 \mid \mathcal{G}_{k-1})) \\ &= C_2 \mathbb{E}((\psi_k - \psi_{k-1})^4) \\ &= C_2 \mathbb{E}_x(u_k^4) \\ &\leq K < \infty \end{aligned}$$

(where we have used Proposition 3.1(i), (iii) and Lemma 6.3) for some constant K independent of k . \square

Lemma 7.3. *Let $x \in X$ and let $\varepsilon > 0$. Then*

$$(III) = \sum_{k=0}^{n-1} \mathbb{E}((\psi_k - \psi_{k-1})^2 \mid \mathcal{G}_{k-1}) - (\psi_k - \psi_{k-1})^2 = O(n^{\frac{1}{2}+\varepsilon}) \text{ a.s.}$$

Proof. Let $X_k = \mathbb{E}((\psi_k - \psi_{k-1})^2 \mid \mathcal{G}_{k-1}) - (\psi_k - \psi_{k-1})^2$. The definition of \mathcal{G}_k in Proposition 3.1 implies that $\mathbb{E}((\psi_k - \psi_{k-1})^2 \mid \mathcal{G}_{k-1}) - (\psi_k - \psi_{k-1})^2$ is \mathcal{G}_k -measurable. Moreover,

$$\mathbb{E}((\psi_k - \psi_{k-1})^2 - \mathbb{E}((\psi_k - \psi_{k-1})^2 \mid \mathcal{G}_{k-1}) \mid \mathcal{G}_{k-1}) = 0.$$

Hence the summand in (III) is a martingale difference.

Note that

$$\begin{aligned} \mathbb{E}((\mathbb{E}((\psi_k - \psi_{k-1})^2 \mid \mathcal{G}_{k-1}) - (\psi_k - \psi_{k-1})^2)^2) &\leq \mathbb{E}((\psi_k - \psi_{k-1})^4) \\ &= \mathbb{E}_x(u_k^4) \leq K \end{aligned}$$

for some constant K independent of k . By Lemma 7.1, (III) = $O(n^{\frac{1}{2}+\varepsilon})$ a.s. \square

Finally, it remains to consider (IV).

Lemma 7.4. *We have that*

$$(IV) = \sum_{k=0}^{n-1} ((\psi_k - \psi_{k-1})^2 - \mathbb{E}_x(u_k^2)) = O(n^{\frac{1}{2}+\varepsilon}) \mu_x\text{-a.e.}$$

The summand in (IV) is not in general a martingale difference, and the analysis above does not immediately generalise. We prove Lemma 7.4 in §8 below.

Combining the above we see that

$$T_n - \sigma^2 n = O(n^{\frac{1}{2}+\varepsilon})$$

μ_x -a.e. To prove part (iii) of Theorem 1.4 it suffices to note that for any $\varepsilon > 0$,

$$\begin{aligned} W(T_n) &= W(\sigma^2 n + O(n^{\frac{1}{2}+\varepsilon'})) \text{ a.s.} \\ &= W(\sigma^2 n) + O(n^{\frac{1}{4}+\varepsilon'}) \text{ a.s.} \end{aligned}$$

for some $0 < \varepsilon' < \varepsilon$.

8. PROOF OF LEMMA 7.4

In order to prove Lemma 7.4 we need the following special case of the Ga'al-Koksma inequality.

Lemma 8.1 ([PS, Appendix I]). *Let X_k be a zero-mean random variable such that $\mathbb{E}(X_k^2) \leq K < \infty$, where K is independent of k . Suppose in addition that for each $m, n \in \mathbb{N}$*

$$\mathbb{E} \left(\left(\sum_{k=m}^{m+n} X_k \right)^2 \right) = O(n),$$

where the implied constant is independent of m . Then for any $\varepsilon > 0$, we have that

$$\sum_{k=0}^{n-1} X_k = O(n^{\frac{1}{2}+\varepsilon}) \text{ a.s.}$$

We apply Lemma 8.1 to (IV) with $X_k = (\psi_k - \psi_{k-1})^2 - \mathbb{E}_x(u_k^2)$. By Lemma 6.3 and Proposition 3.1(i) we see that

$$(11) \quad \mathbb{E}(X_k^2) = \mathbb{E}_x((u_k^2 - \mathbb{E}_x(u_k^2))^2) = \mathbb{E}_x(u_k^4) - \mathbb{E}_x(u_k^2)^2 \leq K < \infty.$$

Now

$$(12) \quad \mathbb{E} \left(\left(\sum_{k=m}^{m+n} X_k \right)^2 \right) = \sum_{k=m}^{n+m} \mathbb{E}(X_k^2) + 2 \sum_{d=1}^n \sum_{k=m}^{m+n-d} \mathbb{E}(X_{k+d} X_k).$$

The first term is clearly $O(n)$ (independently of m) by (11).

We claim that

$$(13) \quad |\mathbb{E}(X_{k+d} X_k)| \leq \text{Const } \rho^d,$$

where the constant is independent of m, n, d, k . To see that (13) is sufficient to conclude that (12) is $O(n)$, simply note that

$$\sum_{d=1}^n \sum_{k=m}^{m+n-d} \rho^d = \sum_{d=1}^n (n-d) \rho^d \leq \sum_{d=1}^n n \rho^d \leq \frac{n}{1-\rho} = O(n).$$

It remains to prove (13). Using Proposition 3.1(i) we have

$$\mathbb{E}(X_{k+d} X_k) = \mathbb{E}_x(u_{k+d}^2 u_k^2) - \mathbb{E}_x(u_{k+d}^2) \mathbb{E}_x(u_k^2).$$

As $u_k^2 = w^2 Z_k - 2wZ_k PwZ_{k-1} + (Pw)^2 Z_{k-1}$ we can write the above as a linear combination of the following nine terms:

$$\begin{aligned}
(14) \quad & \mathbb{E}_x(w^2 Z_{k+d} w^2 Z_k) - \mathbb{E}_x(w^2 Z_{k+d}) \mathbb{E}_x(w^2 Z_k), \\
(15) \quad & \mathbb{E}_x(w^2 Z_{k+d} w Z_k Pw Z_{k-1}) - \mathbb{E}_x(w^2 Z_{k+d}) \mathbb{E}_x(w Z_k Pw Z_{k-1}), \\
(16) \quad & \mathbb{E}_x(w^2 Z_{k+d} (Pw)^2 Z_{k-1}) - \mathbb{E}_x(w^2 Z_{k+d}) \mathbb{E}_x((Pw)^2 Z_{k-1}), \\
(17) \quad & \mathbb{E}_x(w Z_{k+d} Pw Z_{k+d-1} w^2 Z_k) - \mathbb{E}_x(w Z_{k+d} Pw Z_{k+d}) \mathbb{E}_x(w^2 Z_k), \\
(18) \quad & \mathbb{E}_x(w Z_{k+d} Pw Z_{k+d-1} w Z_k Pw Z_{k-1}) \\
& \quad - \mathbb{E}_x(w Z_{k+d} Pw Z_{k+d}) \mathbb{E}_x(w Z_k Pw Z_{k-1}), \\
(19) \quad & \mathbb{E}_x(w Z_{k+d} Pw Z_{k+d-1} (Pw)^2 Z_{k-1}) \\
& \quad - \mathbb{E}_x(w Z_{k+d} Pw Z_{k-1}) \mathbb{E}_x((Pw)^2 Z_{k-1}), \\
(20) \quad & \mathbb{E}_x((Pw)^2 Z_{k+d-1} w^2 Z_k) - \mathbb{E}_x((Pw)^2 Z_{k+d-1}) \mathbb{E}_x(w^2 Z_k), \\
(21) \quad & \mathbb{E}_x((Pw)^2 Z_{k+d-1} w Z_k Pw Z_{k-1}) \\
& \quad - \mathbb{E}_x((Pw)^2 Z_{k+d-1}) \mathbb{E}_x(w Z_k Pw Z_{k-1}), \\
(22) \quad & \mathbb{E}_x((Pw)^2 Z_{k+d-1} (Pw)^2 Z_{k-1}) \\
& \quad - \mathbb{E}_x((Pw)^2 Z_{k+d-1}) \mathbb{E}_x((Pw)^2 Z_{k-1}).
\end{aligned}$$

The following estimates will prove useful.

Lemma 8.2. *Let a, b be as in Proposition 4.3. For $g, h \in C_{a,b}(X, \mathbb{R})$ and $p > 0$ we have*

- (i) *for each $x \in X$, $|Q^p g(x)| \leq \text{Const } \rho^p$;*
- (ii) *$|\nu(Q^p g \cdot h)| < \text{Const } \rho^p$.*

Proof. Statement (i) follows immediately from the definition of $C_{a,b}(X, \mathbb{R})$ and Proposition 4.3.

To prove (ii) note that

$$\begin{aligned}
|\nu(Q^p g \cdot h)| & \leq \int |Q^p g(x)| |h(x)| d\nu(x) \\
& \leq \int \|Q^p g\|_{a,b} a(d(x, x_0)) |h|_a a(d(x, x_0)) d\nu(x) \\
& \leq \text{Const } \rho^p \int a(d(x, x_0))^2 d\nu(x).
\end{aligned}$$

As $a(t) = 1 + t^\alpha$ and $\alpha < 1/2$, we have $a(t) = O(t)$. Hence it is sufficient to prove that $\int d(x, x_0) d\nu(x) < \infty$. In the case of finitely many maps T_j this is immediate from [BDEG, Theorem 2.4]. For the case of infinitely many maps, we argue as follows. From [Pe, Lemma 2.3], we have that

$$\sup_{x \in X, n \geq 0} \mathbb{E}_x(d(Z_n(x, i), x_0)) = M < \infty.$$

For each $m \geq 0$, define $d_m(x) = \min\{d(x, x_0), m\}$. This is a bounded Lipschitz function, hence $P^n d_m(x) \rightarrow \int d_m d\nu$. As for each n ,

$$P^n d_m(x) = \mathbb{E}_x(d_m(Z_n(x, i))) \leq \mathbb{E}_x(d(Z_n(x, i), x_0)) \leq M < \infty,$$

we have $\int d_m d\nu \leq M < \infty$ and the result follows by the Monotone Convergence Theorem. \square

To estimate (14), (16), (20) and (22) we use the following calculation. Recall that $x \in X$ is fixed. For brevity, we shall write Pg for $Pg(x)$, etc., in what follows.

Lemma 8.3. *Let $g, h \in C_{a^2, b+b^2}(X, \mathbb{R})$. Then*

$$|\mathbb{E}_x(gZ_{p+q}hZ_p) - \mathbb{E}_x(gZ_{p+q})\mathbb{E}_x(hZ_p)| \leq \text{Const } \rho^q,$$

where the constant is independent of p, q .

Proof. Observe that

$$\begin{aligned} & |\mathbb{E}_x(gZ_{p+q}hZ_p) - \mathbb{E}_x(gZ_{p+q})\mathbb{E}_x(hZ_p)| \\ &= |P^p(hP^qg) - P^{p+q}g \cdot P^ph| \\ &= |P^p(\nu(g)h + Q^qg \cdot h) - (\nu(g) + Q^{p+q}g)(\nu(h) + Q^ph)| \\ &= |\nu(Q^qg \cdot h) + Q^p(Q^qg \cdot h) - \nu(h)Q^{p+q}g - Q^{p+q}g \cdot Q^ph| \\ &\leq \text{Const } \rho^q. \end{aligned} \quad \square$$

To estimate (17) and (19) we use:

Lemma 8.4. *Let $g, h \in C_{a,b}(X, \mathbb{R})$. Then*

$$|\mathbb{E}_x(gZ_{p+q}PgZ_{p+q-1}hZ_p) - \mathbb{E}_x(gZ_{p+q}PgZ_{p+q-1})\mathbb{E}_x(hZ_p)| \leq \text{Const } \rho^q,$$

where the constant is independent of p, q .

Proof. Observe that, using a simple extension of Lemma 6.1,

$$\begin{aligned} & \mathbb{E}_x(gZ_{p+q}PgZ_{p+q-1}hZ_p) \\ &= P^p((P^{q-1}(Pg)^2)h) \\ &= P^p(\nu((Pg)^2)h + Q^{q-1}((Pg)^2)h) \\ &= \nu((Pg)^2)\nu(h) + \nu(Q^{q-1}((Pg)^2)h) + \nu((Pg)^2)Q^ph + Q^p(Q^{q-1}((Pg)^2)h) \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E}_x(gZ_{p+q}PgZ_{p+q-1})\mathbb{E}_x(hZ_p) \\ &= P^{p+q-1}((Pg)^2)P^ph \\ &= (\nu((Pg)^2) + Q^{p+q-1}((Pg)^2))(\nu(h) + Q^ph) \\ &= \nu((Pg)^2)\nu(h) + \nu((Pg)^2)Q^ph + \nu(h)Q^{p+q-1}((Pg)^2) \\ &\quad + Q^{p+q-1}((Pg)^2)Q^ph. \end{aligned}$$

Noting that the constant terms and terms only involving Q^p cancel and applying Lemma 8.2, we see that the difference of the above terms is at most $\text{Const } \rho^q$. \square

To estimate (15) and (21) we use:

Lemma 8.5. *Let $g, h \in C_{a,b}(X, \mathbb{R})$ and suppose that $\nu(h) = 0$. Then*

$$|\mathbb{E}_x(gZ_{p+q}hZ_pPhZ_{p-1}) - \mathbb{E}_x(gZ_{p+q})\mathbb{E}_x(hZ_pPhZ_{p-1})| \leq \text{Const } \rho^q,$$

where the constant is independent of p, q .

Proof. First note that $Ph = Qh$. It is easy to see that

$$\begin{aligned} & \mathbb{E}_x(gZ_{p+q}hZ_pPhZ_{p-1}) \\ &= P^{p-1}((P((P^qg)h))Ph) \\ &= \nu(g)\nu((Qh)^2) + \nu(Q(Q^qg \cdot h)Qh) + \nu(Q^qg \cdot h)Q^ph + \nu(g)Q^{p-1}((Qh)^2) \\ &\quad + Q^{p-1}(Q(Q^qg \cdot h)Qh) \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E}_x(gZ_{p+q})\mathbb{E}_x(hZ_pPhZ_{p-1}) \\ &= \nu(g)\nu((Qh)^2) + \nu(g)Q^{p-1}((Qh)^2) + \nu((Qh)^2)Q^{p+q}g \\ & \quad + Q^{p+q}g \cdot Q^{q-1}((Qh)^2). \end{aligned}$$

The constant terms and terms involving only Q^p cancel, and Lemma 8.2 (with an obvious modification to control the term $\nu(Q(Q^qg \cdot h)Qh)$) implies the result. \square

Finally, we estimate (18). Recall that $\nu(w) = 0$ so that $Pw = Qw$. First note that

$$\begin{aligned} & \mathbb{E}_x(wZ_{k+d}PwZ_{k+d-1}wZ_kPwZ_{k-1}) \\ &= P^{k-1}((P((P^{d-1}((Pw)^2))w))Pw) \\ &= \nu((Qw)^2)\nu((Qw)^2) + \nu(Q(Q^{d-1}((Qw)^2)w)Qw) + \nu(Q^{d-1}((Qw)^2)w)Q^k w \\ & \quad + \nu((Qw)^2)Q^{k-1}((Qw)^2) + Q^{k-1}(Q(Q^{d-1}((Qw)^2)w)Qw) \end{aligned}$$

(where we have used the facts that $\nu(w) = 0$ and $\nu Q = 0$) and that

$$\begin{aligned} & \mathbb{E}_x(wZ_{k+d}PwZ_{k+d-1})\mathbb{E}_x(wZ_kPwZ_{k-1}) \\ &= (P^{k+d-1}((Qw)^2))(P^{k-1}(Qw)^2) \\ &= (\nu((Qw)^2) + Q^{k+d-1}((Qw)^2))(\nu((Qw)^2) + Q^{k-1}((Qw)^2)). \end{aligned}$$

Multiplying this out, we see that the constant terms and terms involving only Q^k again cancel. Applying Lemma 8.2 (and an obvious modification to control the term $\nu(Q(Q^{d-1}((Qw)^2)w)Qw)$) implies that (18) is bounded above by $\text{Const } \rho^d$.

Combining the above and Lemmas 8.3, 8.4, 8.5 we see that

$$|\mathbb{E}(X_{k+d}X_k)| = |\mathbb{E}_x(u_{k+d}^2u_k^2) - \mathbb{E}_x(u_{k+d}^2)\mathbb{E}_x(u_k^2)| \leq \text{Const } \rho^d,$$

where the constant is independent of k, d . By the discussion above, this implies that $(\text{IV}) = O(n^{\frac{1}{2}+\varepsilon})$ for any $\varepsilon > 0$. This concludes the proof of Theorem 1.4.

REFERENCES

- [BDEG] M.F. Barnsley, S.G. Demko, J.H. Elton and J.S. Geronimo, *Invariant measures for Markov processes arising from iterated function systems with place-dependent probabilities*. Ann. Inst. H. Poincaré Probab. Statist., **24** (1988), 367–394. MR0971099 (89k:60088)
- [DK] 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. MR0779712 (86g:28025)
- [DF] P. Diaconis and D. Freedman, *Iterated random functions*, SIAM Review **41** (1999), 45–76. MR1669737 (2000c:60102)
- [E1] J.H. Elton, *An ergodic theorem for iterated maps*, Ergod. Th. & Dynam. Sys., **7** (1987), 481–488. MR0922361 (89b:60156)
- [E2] J.H. Elton, *A multiplicative ergodic theorem for Lipschitz maps*, Stoch. Proc. Appl., **34** (1990), 39–47. MR1039561 (91a:47010)
- [F] W. Feller, *An Introduction to Probability Theory and its Applications, vol. II*, Wiley, New York, 1966. MR0210154 (35:1048)
- [FMT] M.J. Field, I. Melbourne and A. Török, *Decay of correlations, central limit theorems and approximation by Brownian motion for compact lie group extensions*, Ergod. Th. & Dynam. Sys., **23** (2003), 87–110. MR1971198 (2004m:37046)
- [HH] P. Hall and C. Heyde, *Martingale Limit Theory and its Applications*, Academic Press, New York, 1980. MR0624435 (83a:60001)

- [MT] I. Melbourne and A. Török, *Central limit theorems and invariance principles for time-one maps of hyperbolic flows*, Commun. Math. Phys., **229** (2002), 57–71. MR1917674 (2003k:37012)
- [Pe] M. Peigné, *Iterated function systems and spectral decomposition of the associated Markov operator*, Publ. Inst. Rech. Math. Rennes 1993, Univ. Rennes I, Rennes, 1993. MR1347702
- [PS] W. Philipp and W. Stout, *Almost sure invariance principles for partial sums of weakly dependent random variables*, Mem. Amer. Math. Soc., **161** (1975), Amer. Math. Soc., Prov., Rhode Island. MR0433597 (55:6570)
- [Po] M. Pollicott, *Contraction in mean and transfer operators*, Dynam. & Stab. Syst., **16** (2001), 97–106. MR1835908 (2002c:37080)
- [PoS] M. Pollicott and R. Sharp, *Invariance principles for interval maps with an indifferent fixed point*, Commun. Math. Phys., **229** (2002), 337–346. MR1923178 (2003k:37016)
- [S1] V. Strassen, *An invariance principle for the law of the iterated logarithm*, Z. Wahr., **3** (1964), 211–226. MR0175194 (30:5379)
- [S2] V. Strassen, *Almost-sure behaviour of sums of independent random variables and martingales*, Proc. Fifth Berkeley Symp. Math. Stat. Prob., **2** (1965), 314–343.

SCHOOL OF MATHEMATICS, UNIVERSITY OF MANCHESTER, OXFORD ROAD, MANCHESTER M13 9PL, UNITED KINGDOM

E-mail address: cwalkden@maths.man.ac.uk