Ergod. Th. & Dynam. Sys. (2019), **39**, 1971–2005 © *Cambridge University Press, 2017*
doi:10.1017/etds.2017.83

Birkhoff sum fluctuations in substitution dynamical systems

ELLIOT PAQUETTE† and YOUNGHWAN SON‡

† *Department of Mathematics, The Ohio State University, USA*
(e-mail: paquette.30@osu.edu)
‡ *Department of Mathematics, POSTECH, Korea*
(e-mail: yhson@postech.ac.kr)

(*Received* 24 *April* 2016 *and accepted in revised form* 24 *February* 2017)

Abstract. We consider the deviation of Birkhoff sums along fixed orbits of substitution dynamical systems. We show distributional convergence for the Birkhoff sums of eigenfunctions of the substitution matrix. For non-coboundary eigenfunctions with eigenvalue of modulus 1, we obtain a central limit theorem. For other eigenfunctions, we show convergence to distributions supported on Cantor sets. We also give a new criterion for such an eigenfunction to be a coboundary, as well as a new characterization of substitution dynamical systems with bounded discrepancy.

1. *Introduction*

Let $\mathcal{A}$ be a finite set of letters. Let $\mathcal{A}^*$ be the collection of all finite words using letters from $\mathcal{A}$. Let θ be a *substitution* on $\mathcal{A}$, i.e. a map $\mathcal{A} \to \mathcal{A}^*$. This can be extended to a map $\mathcal{A}^* \to \mathcal{A}^*$ by concatenation, i.e. for all $a_1 \cdots a_k \in \mathcal{A}^*$, we define

$$\theta(a_1 a_2 \cdots a_k) = \theta(a_1)\theta(a_2) \cdots \theta(a_k).$$

Define $\mathcal{A}^{\mathbb{N}}$ to be all the sequences using elements of $\mathcal{A}$; we can extend θ further to the map $\mathcal{A}^{\mathbb{N}} \to \mathcal{A}^{\mathbb{N}}$, again by concatenation. Also, for any finite or infinite word $u = u_1 u_2 \ldots$, we let $u_{<k} = u_1 u_2 \cdots u_{k-1}$ with $u_{<1}$ the empty word. We further define $u_{\leq k}$ analogously.

Define a map $(\ell(\cdot)) : \mathcal{A}^* \to \mathbb{R}^{\mathcal{A}}$, which, for any word $w = a_1 a_2 \cdots a_k$ and any $a \in \mathcal{A}$,

$$(\ell(w))_a = |\{1 \leq i \leq k \,:\, a_i = a\}|.$$

Define the *θ-matrix* M associated to θ as the $|\mathcal{A}| \times |\mathcal{A}|$ integer-valued matrix, so that $M_{a,b} = (\ell(\theta(b)))_a$ for all $a, b \in \mathcal{A}$, that is, the number of occurrences of a in $\theta(b)$. A substitution is called *primitive* if there is a number $k > 0$ so that $M^k_{a,b} > 0$ for all $a, b \in \mathcal{A}$. We will assume from here on that θ is a primitive substitution.

An infinite sequence $u = (u_n)_{n=1}^{\infty} \in \mathcal{A}^{\mathbb{N}}$ is called a *fixed point* if there is $k \in \mathbb{N}$ such that $\theta^k(u) = u$. In general, one can find $a \in \mathcal{A}$ and $k \in \mathbb{N}$ such that $\theta^k(a)$ begins with a. It is easy to check that $u = \lim_{m\to\infty} \theta^{km}(a)$ is a fixed point.

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

For any function $f : \mathcal{A} \to \mathbb{C}$, define the map $S_f : \mathcal{A}^* \to \mathbb{C}$ by the rule that for any $w = a_1 a_2 \cdots a_k \in \mathcal{A}^*$,

$$S_f(w) = f(a_1) + f(a_2) + \cdots + f(a_k).$$

Substitution systems (see §4 for more background) are uniquely ergodic and hence, for a fixed point u, we have by the Birkhoff ergodic theorem that

$$\lim_{N\to\infty} \frac{S_f(u_{\le N})}{N} \to \int f \, d\mu$$

for some measure μ.

We will study, in a sense, the first order correction term to this convergence. As a motivating example, consider the case of the irrational circle rotation, let α be any irrational number and consider any interval $I = [0, x)$ with $0 < x < 1$. Write

$$Z_\alpha(N; I) = \sum_{1\le n\le N} \mathbf{1}\{0 \le \{n\alpha\} < x\}.$$

Then by unique ergodicity of the irrational rotation we have that

$$\frac{1}{N} Z_\alpha(N; I) \to x.$$

The fluctuations of this ergodic average from x can be described by the following theorem of Beck.

THEOREM 1.1. (Beck [**Bec10**, **Bec11**]) *Suppose that α is a quadratic irrational and $I = [0, x)$ has a rational end point x. There are constants $C_1 = C_1(\alpha, x)$ and $C_2 = C_2(\alpha, x)$ such that for any real numbers $-\infty < t < \infty$,*

$$\frac{1}{N}\left|\left\{1 \le n \le N : \frac{(Z_\alpha(n; I) - nx) - C_1 \log N}{C_2\sqrt{\log N}} \le t\right\}\right| \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-y^2/2}\, dy.$$

This is to say that the fluctuations of the Birkhoff sum are asymptotically normally distributed. Recently, a new dynamical proof of the above theorem for the case $x = \frac{1}{2}$ was obtained by studying renormalization properties of the linear flow on an infinite staircase [**ADDS15**].

In the case of a fixed point of a substitution, we show a central limit theorem for eigenfunctions f of M with eigenvalues λ_f of modulus 1. For some eigenfunctions f, it is possible that f is a *coboundary*, meaning that $\{S_f(u_{\le n})\}_{n=1}^{\infty}$ is bounded. For these f, no central limit theorem is possible, and we give a characterization of eigenfunctions f that have this property in Proposition 3.1. Otherwise, if f is not a coboundary, appropriately scaling $S_f(u_{\le n})$, the fluctuations of the Birkhoff sum will also be asymptotically normal. We begin by giving the easiest of our theorems to formulate, where $\lambda_f = 1$ (see §3 for the full formulation).

THEOREM 1.2. *Let f be a left eigenfunction of M with eigenvalue $\lambda_f = 1$, so that f is not a coboundary. There are constants c_1, c_2 so that for all real t,*

$$\frac{1}{N}\left|\left\{1 \le n \le N : \frac{S_f(u_{\le n}) - c_1 \log_\lambda(N)}{c_2\sqrt{\log_\lambda(N)}} \le t\right\}\right| \to \int_{-\infty}^{t} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx.$$

(Here λ is the Perron–Frobenius eigenvalue of the θ-matrix M.)

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

The condition that f is an eigenfunction of modulus 1 is essential to this theorem. Indeed, if f is an eigenfunction of modulus not equal to 1, the asymptotic distribution of the Birkhoff sums is non-Gaussian (see Theorems 3.1 and 3.2). Conversely, for any eigenfunction f with eigenvalue of modulus 1 which is not a coboundary, we show a central limit theorem.

The reason why eigenfunctions of M play a special role here is that they satisfy a certain renormalization identity. Specifically, for an eigenfunction f of M with eigenvalue λ_f, we have that for any word w,

$$S_f(\theta^k(w)) = \lambda_f^k S_f(w). \tag{1}$$

Hence, in the case where λ_f has modulus 1, words of all different scales have the same contribution to the Birkhoff sum, due to which we can eventually prove a central limit theorem.

Let (X_θ, T) be the substitution subshift of bi-infinite sequences associated to the primitive substitution θ (see §4 for background). We are also interested in studying the behavior of ergodic sums for any point $v \in X_\theta$, in the same sense of Theorem 1.2. In this case, one sees that the *empirical mean* (meaning the average of $(S_f(v_{\le n}))_{n=1}^N$) is not in general $c_1 \log_\lambda N + o(\sqrt{\log N})$. Hence, it is necessary to subtract a term $a_{v,N}$, which is not simply $c_1 \log_\lambda(N)$, to get a central limit theorem. This is intriguing in that *every* orbit in X_θ satisfies a central limit theorem, with the only difference being a small difference in how different orbits are centered.

See Theorem 3.4 for details.

There exists another type of central limit theorem for Birkhoff sums of eigenfunctions in substitution dynamical systems, the one studied in [**BBH14**]. It is known that if θ is primitive, there exists a unique ergodic T-invariant measure μ on X_θ. Hence, for an eigenfunction f, it is possible to consider $S_f(v_{\le n})$ as a random variable on the probability space $(X_\theta, \mathcal{B}, \mu)$, i.e. to estimate

$$F_{n,f}(t) := \mu\big(v \in X_\theta \ : \ S_f(v_{\le n}) \le t\sqrt{\log n}\big).$$

For similar reasons as to why one expects a central limit theorem for the empirical distribution of $(S_f(v_{\le n}))_{n=1}^N$ (i.e. what is shown in Theorem 1.2), one expects that $S_f(v_{\le n})$ may be normally distributed under μ as $n \to \infty$.

When formulated in this way it is necessary to consider subsequences $S_f(v_{\le [t\lambda^n]})$, as the distribution under μ of these Birkhoff sums depends greatly on t. It is shown [**BBH14**, Theorem 1.2] that for rational t, these sequences converge in distribution to $Z \cdot K$, where Z is a standard normal and K is an independent, finitely supported random variable, which can possibly take the value 0 (nonetheless, it is shown that K is identically 0 only if f is a coboundary). It is conjectured [**BBH14**, Conjecture 1.4] that for Lebesgue-almost every t, K is supported on a single non-zero point.

2. *Preliminaries*

2.1. *Probability background.* We will occasionally use probability formalism where convenient. We may say, for example, that Z is a real-valued random variable with a standard normal distribution, without specifying the probability space or naming the

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

probability measure. In this case, we are only interested in distributional properties of Z. We will use $\mathbb{P}$ as a placeholder for this measure, for example

$$\mathbb{P}[Z \in A] = \int_A \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx.$$

We will also use $\mathbb{E}$ to denote integration against the distribution of a random variable, for example

$$\mathbb{E} f(Z) = \int_{\mathbb{R}} f(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx.$$

Our main theorems are stated in terms of distributional convergence of random variables, also known as *weak* convergence of their distributions. There are many equivalent definitions of weak convergence, which we will freely interchange as convenient. The equivalence of these definitions usually goes by the name of the *portmanteau* lemma.

LEMMA 2.1. (Portmanteau lemma) *Let E be a metric space and let $\mu, \mu_1, \mu_2, \ldots$ be Borel probability measures on E. The following are equivalent. If they occur, we write $\mu_n \Rightarrow \mu$ and say that μ_n converges weakly to μ.*

(1) *For all bounded continuous functions f, $\int f d\mu_n \to \int f d\mu$.*
(2) *For all Lipschitz continuous functions f, $\int f d\mu_n \to \int f d\mu$.*
(3) *For all measurable A with $\mu(\partial A) = 0$, $\mu_n(A) \to \mu(A)$.*
(4) *If $E = \mathbb{R}^d$, then for all A of the form $(-\infty, x_1] \times \cdots \times (-\infty, x_d]$ with $\mu(\partial A) = 0$, $\mu_n(A) \to \mu(A)$.*

See [**Kle14**, Theorem 13.16]. A sequence of random variables $X_1, X_2, \ldots$ is said to *converge in distribution* to X if their distributions, i.e. $\mu_n(\cdot) := \mathbb{P}[X_n \in \cdot]$, converge weakly to $\mu(\cdot) := \mathbb{P}[X \in \cdot]$. We again write $X_n \Rightarrow X$ for distributional convergence.

It is also possible to metrize weak convergence. We will make use of such a metric later on. Define the *bounded-Lipschitz metric* on the space of probability measures on a metric space E as

$$d_{\mathrm{BL}}(\mu, \nu) = \sup_{\substack{f: E \to \mathbb{R} \\ \|f\|_\infty \le 1 \\ \|f\|_{\mathrm{Lip}} \le 1}} \left| \int f\, d\mu - \int f\, d\nu \right|.$$

When E is a separable metric space, it is a theorem of Dudley [**Dud66**] that weak convergence is equivalent to convergence in the bounded-Lipschitz metric.

Theorems 3.3, 3.4 and A.1 show distributional convergence to a complex normal distribution. A complex random variable Z with mean 0 is said to have a complex normal distribution with positive definite covariance matrix $\Gamma = \begin{bmatrix} \mathbb{E}(\Re Z)^2 & \mathbb{E}(\Re Z \Im Z) \\ \mathbb{E}(\Re Z \Im Z) & \mathbb{E}(\Im Z)^2 \end{bmatrix}$ if for any Borel measurable $A \subseteq \mathbb{R}^2$,

$$\mathbb{P}[Z \in A] = \int_A \frac{1}{2\pi\sqrt{\det \Gamma}} e^{-(1/2) x^t \Gamma^{-1} x}\, dx_1\, dx_2.$$

It follows from this definition that both of $\Re Z$ and $\Im Z$ are normally distributed. Note that Z is completely determined by $\mathbb{E}|Z|^2$ and $\mathbb{E}Z^2$. If $\mathbb{E}\Re Z \Im Z = 0$, then the real and imaginary parts of Z are in fact independent normal random variables.

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

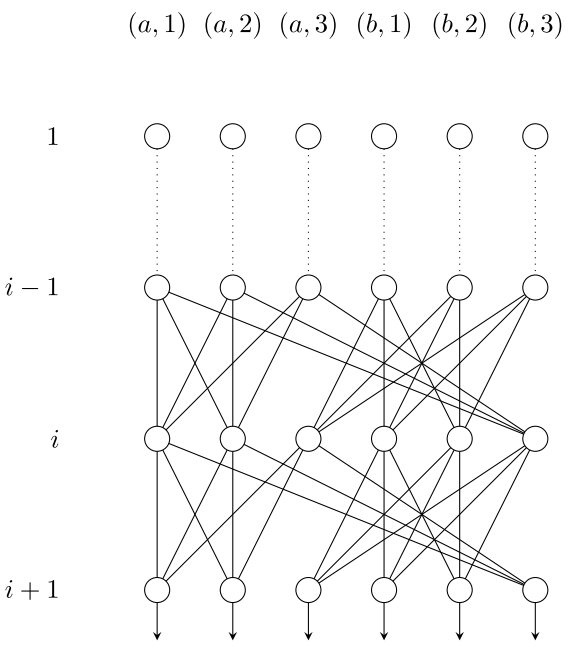


FIGURE 1. Here we give a graphical representation of path space for the substitution $\theta(a) = aab$, $\theta(b) = bba$.

2.2. *Path space.* In this section, following [**Liv88**], we will introduce a coding for substitution sequences. This coding represents a substitution sequence as a path on an ordered Bratteli diagram. (See [**DHS99**] for the general theory of Bratteli diagrams.) Define the *state space* $\mathcal{X}$ by

$$\mathcal{X} = \{(a, j) : a \in \mathcal{A}, 1 \le j \le |\theta(a)|\}. \tag{2}$$

Also, for each $p \in \mathbb{N}$, define the *path space* $\mathcal{X}^{*,p} \subseteq \mathcal{X}^p$ of all sequences $\{(v_i, k_i)\}_{i=1}^p$ so that for all $1 \le i < p$, $v_i = \theta(v_{i+1})_{k_{i+1}}$.

We will now define a coding of the collection of strict prefixes of $w = \theta^p(a)$. Let $n \le |\theta^p(a)|$ be a positive integer. One can find $v_p, v_{p-1}, \ldots, v_1 \in \mathcal{A}$ and positive integers $k_p, k_{p-1}, \ldots, k_1$ such that

$$w_{[1,n)} = \theta^{p-1}(\theta(v_p)_{<k_p}) \cdot \theta^{p-2}(\theta(v_{p-1})_{<k_{p-1}}) \cdots \theta(\theta(v_2)_{<k_2}) \cdot \theta(v_1)_{<k_1},$$

where

$$v_p = a, \quad 1 \le k_p \le |\theta(a)|$$

and, for $i = 1, \ldots, p-1$,

$$v_i = \theta(v_{i+1})_{k_{i+1}}, \quad 1 \le k_i \le |\theta(v_i)|.$$

Since this expression is unique, there is an injection $\Psi_{a,p} : \{1, 2, \ldots, |\theta^p(a)|\} \to \mathcal{X}^{*,p}$ given by

$$\Psi_{a,p}(n) = (v_1, k_1)(v_2, k_2) \cdots (v_p, k_p). \tag{3}$$

We also define the infinite path space $\mathcal{X}^{*,\infty} \subseteq \mathcal{X}^\infty$ as the collection of all sequences $\{(v_i, k_i)\}_{i=1}^\infty$ so that for all $i \ge 1$, $v_i = \theta(v_{i+1})_{k_{i+1}}$. See Figure 1 for the case $\theta(a) = aab$, $\theta(b) = bba$. On this space, there is a natural Markov measure associated to θ, which

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

we denote by SMPM_∞. It is a primitive, stationary Markov chain with some transition matrix $\mathfrak{p}$. It can be defined combinatorially (see (7)) or as the unique invariant measure under the *adic transformation* (see §4).

Remark 2.1. We adopt a convention used by [**Liv88**] in defining a path space. This convention differs from [**BBH14**, **Buf14**, **DHS99**]; however, it is not difficult to see that the induced Bratteli diagrams are isomorphic. The advantage in using the coding in (3) is that for an eigenfunction f of a substitution matrix, we can express

$$S_f(w_{[1,n)}) = \sum_{i=1}^{p} \lambda_f^{i-1} S_f(\theta(v_i)_{<k_i}),$$

which is a statistic of the vertices visited by the path, instead of the edges. In particular, this makes $S_f(w_{[1,n)})$ an *additive functional* of the Markov chain, the theory of which is well developed (at least when $\lambda_f = 1$; see [**MW00**, **KV86**]).

For every $x = (a, j)$, $y = (b, k) \in \mathcal{X}$, let $\mathrm{SSIM}^1_{x,y} = \mathbf{1}\{a = \theta(b)_k\}$. We then define, inductively,

$$\mathrm{SSIM}^p_{x,y} = \sum_{z \in \mathcal{X}} \mathrm{SSIM}^{p-1}_{x,z}\, \mathrm{SSIM}^1_{z,y}\,.$$

Hence, $\mathrm{SSIM}^p_{x,y}$ counts the number of elements in $\mathcal{X}^{*,p+1}$ started from x and ended at y. Let $\mathrm{SSIM}^p_{*,y} = \sum_{x \in \mathcal{X}} \mathrm{SSIM}^p_{x,y}$.

We can see that $\mathrm{SSIM}^1_{\cdot,\cdot}$ is a primitive matrix: since θ is primitive, there exists k such that for any $c, d \in \mathcal{A}$, c appears in $\theta^k(d)$. This means that there exist $m_1, m_2, \ldots, m_k$ so that $\theta(\theta(\cdots(\theta(d)_{m_1})_{m_2}\cdots)_{m_k} = c$. Thus, for any (a, i), $(b, j) \in \mathcal{X}$, $a = \theta(\cdots(\theta(b)_j)_{m_1}\cdots)_{m_k}$, so $\mathrm{SSIM}^{k+1}_{\cdot,\cdot} \geq 1$.

2.3. *Eigenfunctions of* $\mathrm{SSIM}^1_{\cdot,\cdot}$. We begin by observing that for all $(a, j) \in \mathcal{X}$ and all $b \in \mathcal{A}$,

$$M_{a,b} = \sum_{k=1}^{|\theta(b)|} \mathrm{SSIM}^1_{(a,j),(b,k)}.$$

Let $(\rho(b))_b$ be a right Perron–Frobenius eigenfunction of M with Perron–Frobenius eigenvalue λ and define the map $\hat{\rho}((b, k)) = \rho(b)$ for all $(b, k) \in \mathcal{X}$. Then we have that for any $(a, j) \in \mathcal{X}$,

$$\begin{aligned}
\lambda\hat{\rho}(a, j) &= \lambda\rho(a) \\
&= \sum_b M_{a,b}\rho(b) \\
&= \sum_{(b,k)\in\mathcal{X}} \mathrm{SSIM}^1_{(a,j),(b,k)}\, \rho(b) \\
&= \sum_{(b,k)\in\mathcal{X}} \mathrm{SSIM}^1_{(a,j),(b,k)}\, \hat{\rho}(b, k).
\end{aligned}$$

Hence, as it is non-negative, $\hat{\rho}$ is a right Perron–Frobenius eigenfunction of $\mathrm{SSIM}^1_{\cdot,\cdot}$ with eigenvalue λ; the same calculation shows that any right eigenfunction of M can be canonically associated to a right eigenfunction of $\mathrm{SSIM}^1_{\cdot,\cdot}$.

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

Conversely, a similar calculation shows that any left eigenfunction $\hat{g}$ of $\mathrm{SSIM}^1_{\cdot,\cdot}$ gives rise to a left eigenfunction of M with the same eigenvalue by the formula $g(a) = \sum_{j=1}^{|\theta(a)|} \hat{g}(a, j)$, provided that g is non-zero. Furthermore, by the eigenvector equation, we have

$$\begin{aligned}\lambda_g \hat{g}(b, k) &= \sum_{(a,j)\in\mathcal{X}} \hat{g}(a, j)\, \mathrm{SSIM}^1_{(a,j),(b,k)} \\ &= \sum_{(a,j)\in\mathcal{X}} \hat{g}(a, j)\mathbf{1}\{\theta(b)_k = a\} \\ &= \sum_{a\in\mathcal{A}} g(a)\mathbf{1}\{\theta(b)_k = a\} \\ &= g(\theta(b)_k).\end{aligned}$$

Hence, if $\hat{\sigma}$ is the left Perron–Frobenius eigenfunction of $\mathrm{SSIM}^1_{\cdot,\cdot}$, normalized so that $\sum_{x\in\mathcal{X}} \hat{\sigma}(x)\hat{\rho}(x) = 1$, then $\sigma(a) = \sum_{j=1}^{|\theta(a)|} \hat{\sigma}(a, j)$ is the left Perron–Frobenius eigenfunction of M normalized so that $\sum_{a\in\mathcal{A}} \sigma(a)\rho(a) = 1$ and $\hat{\sigma}(b, k) = \lambda^{-1}\sigma(\theta(b)_k)$.

2.4. *Measures on path space.* To prove our theorems, we must consider the measure ν_N that arises on $\mathcal{X}^{*,p}$ by choosing an integer $n \in \{1, 2, \ldots, N\}$ uniformly at random. This means that for every $n \in \{1, 2, \ldots, N\}$,

$$\nu_N(\{\psi_{a,p}(n)\}) = \frac{1}{N},$$

where $|\theta^{p-1}(a)| < N \le |\theta^p(a)|$.

We will decompose ν_N in terms of other measures. For every $p \in \mathbb{N}$ and $y \in \mathcal{X}$, define the probability measure $\mathrm{UPM}_{y,p}$ for each $\mathbf{x} = x_1x_2\cdots x_p \in \mathcal{X}^{*,p}$ by

$$\mathrm{UPM}_{y,p}(\{\mathbf{x}\}) = \frac{\mathbf{1}\{x_p = y\}}{\mathrm{SSIM}^{p-1}_{*,y}}.$$

Then the following is immediate.

PROPOSITION 2.1. *Let $a \in \mathcal{A}$ and $p \in \mathbb{N}$ be such that $|\theta^{p-1}(a)| < N \le |\theta^p(a)|$. Let $\Psi_{a,p}(N) = (v_1, k_1)(v_2, k_2)\cdots(v_p, k_p)$. Then, for every $\mathbf{x} = x_1x_2\cdots x_p \in \mathcal{X}^{*,p}$, we have*

$$\nu_N(\{\mathbf{x}\}) = \sum_{\ell=1}^{p} \sum_{1\le j<k_\ell} \frac{\mathrm{SSIM}^\ell_{*,(v_\ell,j)}}{N}\, \mathrm{UPM}_{(v_\ell,j),\ell}(\{x_1x_2\cdots x_\ell\}).$$

We can also write, for any $\mathbf{x} = x_1x_2\ldots x_p \in \mathcal{X}^p$, that

$$\mathrm{UPM}_{y,p}(\{\mathbf{x}\}) = \frac{\mathbf{1}\{x_p = y\}}{\mathrm{SSIM}^{p-1}_{*,y}} \prod_{i=1}^{p-1} \mathrm{SSIM}^1_{x_i,x_{i+1}}.$$

This family of measures has a (inhomogeneous) Markov chain structure. To express this, define a collection h^y_p of Markov transition matrices given by, for all $x, z \in \mathcal{X}$,

$$h^y_p(x, z) := \frac{\mathrm{SSIM}^1_{x,z}\, \mathrm{SSIM}^{p-2}_{z,y}}{\mathrm{SSIM}^{p-1}_{x,y}},$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

which is a transition matrix as for any $x, y \in \mathcal{X}$ and any $p \in \mathbb{N}$,

$$\sum_{z\in\mathcal{X}} h_p^y(x, z) = \sum_{z\in\mathcal{X}} \frac{\mathrm{SSIM}_{x,z}^1 \,\mathrm{SSIM}_{z,y}^{p-2}}{\mathrm{SSIM}_{x,y}^{p-1}} = 1.$$

This allows us to express $\mathrm{UPM}_{y,p}$ for any $\mathbf{x} = x_1x_2\cdots x_p \in \mathcal{X}^p$ as

$$\begin{aligned}&\mathrm{UPM}_{y,p}(\{\mathbf{x}\})\\ &= \frac{\mathrm{SSIM}_{x_1,y}^{p-1}}{\mathrm{SSIM}_{*,y}^{p-1}} h_p^y(x_1, x_2) h_{p-1}^y(x_2, x_3)\cdots h_3^y(x_{p-2}, x_{p-1}) \frac{\mathrm{SSIM}_{x_{p-1},x_p}^1}{\mathrm{SSIM}_{x_{p-1},y}^{p-1}} \mathbf{1}\{x_p = y\}. \quad (4)\end{aligned}$$

By the Perron–Frobenius theorem, as $\mathrm{SSIM}_{\cdot,\cdot}^1$ is primitive, we have for all $x, y \in \mathcal{X}$,

$$\lim_{p\to\infty} \lambda^{-p}\,\mathrm{SSIM}_{x,y}^p = \hat{\sigma}(y)\hat{\rho}(x). \quad (5)$$

Hence, the limit

$$\mathfrak{p}(x, z) = \lim_{p\to\infty} h_p^y(x, z)$$

exists for all $x, z \in \mathcal{X}$ and is independent of y. In fact, there is a constant $c > 0$ so that for all $p \in \mathbb{N}$,

$$\sup_{x,z\in\mathcal{X}} |\mathfrak{p}(x, z) - h_p^y(x, z)| \le e^{-cp}. \quad (6)$$

Further, we have the following explicit formula for $\mathfrak{p}$:

$$\mathfrak{p}(y, z) = \frac{\mathrm{SSIM}_{y,z}^1\,\hat{\rho}(z)}{\lambda\hat{\rho}(y)}. \quad (7)$$

It also follows that the limit

$$\mathfrak{u}(x_1) = \lim_{p\to\infty} \frac{\mathrm{SSIM}_{x_1,y}^{p-1}}{\mathrm{SSIM}_{*,y}^{p-1}} = \frac{\hat{\rho}(x_1)}{\sum_{x\in\mathcal{X}} \hat{\rho}(x)}$$

exists. This motivates the definition of the following Markov measure on $\mathcal{X}^{*,p}$, where for any $\mathbf{x} = x_1x_2\cdots x_p$,

$$\mathrm{MPM}_p(\{\mathbf{x}\}) = \mathfrak{u}(\{x_1\}) \prod_{i=1}^{p-1} \mathfrak{p}(x_i, x_{i+1}). \quad (8)$$

We will also define a stationary version of MPM_p .

Let $\mathfrak{m}$ be the invariant measure of $\mathfrak{p}$ on $\mathcal{X}$. Define, for $\mathbf{x} = x_1x_2\cdots x_p \in \mathcal{X}^{*,p}$,

$$\mathrm{SMPM}_p(\{\mathbf{x}\}) = \mathfrak{m}(\{x_1\}) \prod_{i=1}^{p-1} \mathfrak{p}(x_i, x_{i+1}). \quad (9)$$

It follows that the invariant measure $\mathfrak{m}$ must be $\mathfrak{m}(y) = \hat{\sigma}(y)\hat{\rho}(y)$, as

$$\sum_{y\in\mathcal{X}} \hat{\sigma}(y)\hat{\rho}(y)\mathfrak{p}(y, z) = \sum_{y\in\mathcal{X}} \frac{\hat{\sigma}(y)\,\mathrm{SSIM}_{y,z}^1\,\hat{\rho}(z)}{\lambda} = \hat{\sigma}(z)\hat{\rho}(z).$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

Given a function $f : \mathcal{A} \to \mathbb{C}$, define a probability measure $\mathfrak{n}$ on $\mathcal{A}$ by the following formula:

$$\int_{\mathcal{A}} f(a)\mathfrak{n}(da) = \int_{\mathcal{X}} S_f(\theta(a)_{<j})\mathfrak{m}(d(a, j)) = \sum_{(a,j)\in\mathcal{X}} \hat{\sigma}(a, j)\hat{\rho}(a, j)S_f(\theta(a)_{<j}). \tag{10}$$

This measure will be used to express the drift in the central limit theorems as well as the condition for being a coboundary.

Example 2.1. Let $\mathcal{A} = \{a, b\}$ and θ be given by

$$\theta : a \to aab, \quad b \to bba.$$

Then the θ-matrix M is

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

- Eigenvalues $\lambda = 3$, $\lambda_f = 1$.
- Left eigenvectors $\sigma = [1, 1]$, $f = [1, -1]$.
- Right Perron–Frobenius eigenvector $\rho = \binom{1}{1}$.
- The state space is denoted by $\mathcal{X} = \{(a, 1), (a, 2), (a, 3), (b, 1), (b, 2), (b, 3)\}$. Then $\mathrm{SSIM}^1_{\cdot,\cdot}$ is given by

$$\mathrm{SSIM}^1_{\cdot,\cdot} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

- Left and right Perron–Frobenius eigenvectors $\hat{\sigma}$ and $\hat{\rho}$ of $\mathrm{SSIM}^1_{\cdot,\cdot}$ are

$$\hat{\sigma} = [1/6, 1/6, 1/6, 1/6, 1/6, 1/6], \quad \hat{\rho}^t = [1, 1, 1, 1, 1, 1].$$

- The invariant measure is $\mathfrak{m}(y) = \hat{\sigma}(y)\hat{\rho}(y) = \frac{1}{6}$ for any $y \in \mathcal{X}$.
- From (7), $\mathfrak{p} = \frac{1}{3}\,\mathrm{SSIM}^1_{\cdot,\cdot}$.

We have that

$$\int_{\mathcal{A}} f(a)\mathfrak{n}(da) = \sum_{(a,j)\in\mathcal{X}} \hat{\sigma}(a, j)\hat{\rho}(a, j)S_f(\theta(a)_{<j}) = (0 + 1 + 2 + 0 - 1 - 2)\frac{1}{6} = 0. \tag{11}$$

2.5. *Reversed path space.* When working with eigenfunctions with eigenvalue $|\lambda_f| > 1$, it is more convenient to work with a *reversed path space*. Define the reversed path space $\mathcal{X}_r^{*,p} \subset \mathcal{X}^p$ as the reversals of all sequences in $\mathcal{X}^{*,p}$. Fix $a \in \mathcal{A}$ and consider $\mathcal{X}_r^{*,p}$ as embedded in $\mathcal{X}^\infty$ by appending to any element the infinite sequence $(a, 1)(a, 1)\cdots$. Make $\mathcal{X}^\infty$ into a topological space by endowing it with the product topology and let $\mathcal{X}_r^{*,\infty}$ be the closure of $\bigcup_p^\infty \mathcal{X}_r^{*,p}$. Now define $\Psi_a^r : \mathbb{N} \to \mathcal{X}^\infty$ by

$$\Psi_a^r(N) = (v_p, k_p)(v_{p-1}, k_{p-1})\cdots(v_1, k_1)(a, 1)(a, 1)\ldots,$$

where (v_i, k_i) are those appearing in $\Psi_{a,p}(N)$ for $|\theta^{p-1}(a)| < N \le |\theta^p(a)|$.

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

We will now give another description of $\mathrm{UPM}_{y,p}$, which is useful for reversed path space. Define a collection $q_p(y, z)$ of Markov transition matrices given by

$$q_p(y, z) := \frac{\mathrm{SSIM}_{*,z}^{p-2}}{\mathrm{SSIM}_{*,y}^{p-1}} \mathrm{SSIM}_{z,y}^{1},$$

which is a transition matrix as for any $y \in \mathcal{X}$ and any $p \in \mathbb{N}$,

$$\sum_{z\in\mathcal{X}} q_p(y, z) = \sum_{z\in\mathcal{X}} \frac{\mathrm{SSIM}_{*,z}^{p-2}\, \mathrm{SSIM}_{z,y}^{1}}{\mathrm{SSIM}_{*,y}^{p-1}} = 1.$$

This allows us to express $\mathrm{UPM}_{y,p}$ for any $\mathbf{x} = x_1x_2 \cdots x_p \in \mathcal{X}^{*,p}$ as

$$\mathrm{UPM}_{y,p}(\{\mathbf{x}\}) = \prod_{i=1}^{p-1} q_{p-i+1}(x_{p-i+1}, x_{p-i})\, \mathbf{1}\{y = x_p\}. \tag{12}$$

As $\mathrm{SSIM}_{\cdot,\cdot}^{1}$ is a primitive matrix, by the Perron–Frobenius theorem, the limit

$$\mathfrak{p}^*(y, z) = \lim_{p\to\infty} q_p(y, z) = \frac{\hat{\sigma}(z)\, \mathrm{SSIM}_{z,y}^{1}}{\lambda\hat{\sigma}(y)} \tag{13}$$

exists for all $y, z \in \mathcal{X}$. We also have that there is a constant $c > 0$ so that for all $p \in \mathbb{N}$,

$$\sup_{y,z\in\mathcal{X}} |\mathfrak{p}^*(y, z) - q_p(y, z)| \le e^{-cp}. \tag{14}$$

Let $\mathfrak{a}$ be any probability measure on $\mathcal{X}$ and define a Markov measure $\mathrm{RMPM}_{\mathfrak{a},\infty}$ on $\mathcal{X}_r^{*,\infty}$ by, for any cylinder set $[\mathbf{x}] = [x_1x_2 \ldots x_p]$,

$$\mathrm{RMPM}_{\mathfrak{a},\infty}([\mathbf{x}]) = \mathfrak{a}(x_1) \prod_{i=1}^{p-1} \mathfrak{p}^*(x_i, x_{i+1}). \tag{15}$$

3. *Main results*

Let θ be a primitive substitution and let $u = (u_n)_{n=1}^{\infty}$ be any fixed point of θ. Denote the Perron–Frobenius eigenvalue of the θ-matrix by λ. For eigenfunctions f of M with eigenvalue λ_f having $|\lambda_f| < 1$, it is well known that the Birkhoff sums $S_f(u_{\le N})$ stay bounded. We show that they also converge in distribution to a bounded random variable. This is a relatively simple consequence of Markov chain mixing.

THEOREM 3.1. *Let f be a left eigenfunction of M with eigenvalue λ_f with $|\lambda_f| < 1$. Let K_N be a random variable with uniform distribution on $\{1, 2, \ldots, N\}$. For $\mathbf{x} = (v_1, k_1)(v_2, k_2) \cdots \in \mathcal{X}^{*,\infty}$, define $W_f(\mathbf{x}) = \sum_{i=1}^{\infty} \lambda_f^{i-1} S_f(\theta(v_i)_{<k_i})$. Then*

$$S_f(u_{\le K_N}) \Rightarrow W_f(\mathbf{X}),$$

where $\mathbf{X}$ has the distribution of MPM_{∞}.

In the case that $|\lambda_f| > 1$, on the other hand, the Birkhoff sums will not in general have a distributional limit. In fact, there are many distributional limit points of $S_f(u_{\le K_N}) N^{-\log_\lambda(\lambda_f)}$ as $N \to \infty$. We show that by choosing different subsequences, it is possible to get different distributional limits, although their distributions are closely related.

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

THEOREM 3.2. *Let $a = u_1$, and let f be a left eigenfunction of M with eigenvalue λ_f having $\lambda > |\lambda_f| > 1$. Let N_ℓ be a sequence with $N_\ell \to \infty$ so that*

$$\Psi_a^r(N_\ell) \to \mathbf{z} = (\rho_1, \kappa_1)(\rho_2, \kappa_2) \cdots \in \mathcal{X}_r^{*,\infty}.$$

Suppose that $\hat{\rho}$ is normalized so that $\sum_{x\in\mathcal{X}} \hat{\rho}(x) = 1$. Define a probability measure on $\mathcal{X}$ by

$$\mathfrak{a}((v, k)) = \frac{1}{R}\hat{\sigma}((v, k)) \sum_{q=1}^{\infty} \mathbf{1}\{v = \rho_q \text{ and } k < \kappa_q\}\lambda^{1-q},$$

where $R > 0$ is chosen so that $\mathfrak{a}$ is a probability measure.

Let K_{N_ℓ} be a random variable with uniform distribution on $\{1, 2, \ldots, N_\ell\}$. For $\mathbf{x} = (v_1, k_1)(v_2, k_2) \cdots \in \mathcal{X}^{*,\infty}$, define $U_f(\mathbf{x}) = \sum_{i=1}^{\infty} \lambda_f^{-i} S_f(\theta(v_i)_{<k_i})$. Let $p = p(\ell) \in \mathbb{N}$ be such that $|\theta^{p-1}(a)| < N_\ell \le |\theta^p(a)|$. Then

$$\frac{S_f(u_{\le K_{N_\ell}})}{N_\ell^{\log_\lambda |\lambda_f|} e^{ip(\ell)\arg\lambda_f}} \Rightarrow \frac{U_f(\mathbf{X})}{R^{\log_\lambda |\lambda_f|}},$$

where $\mathbf{X}$ has the distribution of $\mathrm{RMPM}_{\mathfrak{a},\infty}$.

Remark 3.1.

(1) Theorem 3.2 can be alternatively formulated as

$$\frac{S_f(u_{\le K_{N_\ell}})}{\lambda_f^{p(\ell)}} \Rightarrow \frac{U_f(\mathbf{X})}{R^{\log_\lambda |\lambda_f|}}.$$

(2) An equivalent condition to requiring that

$$\Psi_a^r(N_\ell) \to \mathbf{z} = (\rho_1, \kappa_1)(\rho_2, \kappa_2) \cdots \in \mathcal{X}_r^{*,\infty}$$

is to require that $(N_\ell/\lambda^{p(\ell)}) \to t \in [\lambda^{-1}, 1]$.

Remark 3.2. In both of Theorems 3.1 and 3.2, the support of the limiting measure is a Cantor set. This can be seen by noting that W_f (and U_f) are continuous functions from $\mathcal{X}^\infty$ with the product topology, which is a Cantor set.

Remark 3.3. Theorem 3.2 considers a (deterministic) fixed point and a random length K_{N_ℓ}. One could also consider choosing a random sequence w chosen according to the probability measure from the dynamical system $(X_\theta, \mathcal{B}, \mu, T)$ (see §4 for the definition). In this case it is also possible to prove distributional convergence of $S_f(w_{\le N_\ell})/\lambda_f^{p(\ell)}$, as $\ell \to \infty$. This type of convergence has already been investigated and in fact this convergence follows as a special case of [**Buf14**, Theorems 5 and 6]. We will elaborate on this correspondence, but we require some notation to do so. See Remark 4.3 for more details.

We will soon formulate our main theorems when $|\lambda_f| = 1$, but, before doing so, we give a characterization of eigenfunction coboundaries. Recall that a continuous function $f : X \to \mathbb{C}$ is called a coboundary if there exists a continuous function g such that $f = g - g \circ T$. By the Gottschalk–Hedlund theorem [**GH55**], if X is a compact metric space and $T : X \to X$ is a minimal homeomorphism, and if $f : X \to \mathbb{C}$ is continuous,

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

then f is a coboundary if and only if there exists $K < \infty$ such that $|\sum_{n=1}^{N} f(T^n x)| \le K$ for all $N \in \mathbb{N}$ and $x \in X$. Kornfeld and Lin [**KL97**] obtained a more general result: if X is a compact Hausdorff space and T is an irreducible Markov operator on $C(X)$, then $\sup_N \|\sum_{n=1}^{N} f \circ T^n\| < \infty$ if and only if f is a coboundary.

For left eigenfunctions f of M, we show a further characterization of coboundaries.

PROPOSITION 3.1. *Let $(X_\theta, \mathcal{B}, \mu, T)$ be a substitution dynamical system associated to a primitive substitution θ. Suppose that f is a left eigenfunction of M with eigenvalue λ_f having $|\lambda_f| = 1$. Then the following are equivalent.*

(i) *The function $w \mapsto f(w_1)$ from X_θ to $\mathbb{C}$ is a coboundary.*

(ii) *There exists $w \in X_\theta$ such that $\sup_N |S_f(w_{\le N})| < \infty$.*

(iii) *There is a function $h : \mathcal{A} \to \mathbb{C}$ so that for all $(a, j) \in \mathcal{X}$,*

$$S_f(\theta(a)_{<j}) = \int_{\mathcal{A}} f(c)\mathfrak{n}(dc) + h(a) - \lambda_f^{-1} h(\theta(a)_j).$$

(iv) *There is a function $h : \mathcal{A} \to \mathbb{C}$ so that the following hold.*

(a) *For all $(a, j) \in \mathcal{X}$ with $1 \le j < |\theta(a)|$,*

$$f(\theta(a)_j) = \lambda_f^{-1}(h(\theta(a)_j) - h(\theta(a)_{j+1})).$$

(b) *For all $a \in \mathcal{A}$,*

$$\int_{\mathcal{A}} f(c)\mathfrak{n}(dc) + h(a) - \lambda_f^{-1} h(\theta(a)_1) = 0.$$

Remark 3.4. We say that a function $f : \mathcal{A} \to S^1$ is a coboundary in the sense of Host (see [**Hos86**] or [**Fog02**, Definition 7.3.13]) if there is a function $h : \mathcal{A} \to S^1$ so that for all admissible two-letter words ab, $h(b) = h(a) f(a)$. If f satisfies condition (iv) of Proposition 3.1, then $a \mapsto e^{i\Re f(a)}$ and $a \mapsto e^{i\Im f(a)}$ are coboundaries in the sense of Host.

For an eigenfunction f of M which is not a coboundary with eigenvalue λ_f having $|\lambda_f| = 1$, let Z_f be the following normal random variable.

(1) If $\lambda_f \notin \mathbb{R}$, then Z_f is a complex normal variable. Further, $Z_f = X + iY$, where X, Y are independent, centered normal distributions with $\mathbb{E}X^2 = \mathbb{E}Y^2$ and, letting $g = f - \int f \, d\mathfrak{n}$, we have

$$\mathbb{E}|Z_f|^2 = \mathbb{E}|g(X_1)|^2 + \sum_{k=2}^{\infty} 2\mathbb{E}\Re[\lambda_f^{k-1} g(X_1)\overline{g(X_k)}],$$

where $(X_1, X_2, \ldots)$ has the distribution of SMPM_∞. By Proposition 3.1 and Theorem A.1, this variance is 0 if and only if f is a coboundary.

(2) If $\lambda_f \in \mathbb{R}$ and f is real, then we have that Z_f is real and has the same variance as above. Again, the variance is 0 if and only if f is a coboundary.

THEOREM 3.3. *Let f be a left eigenfunction of M with eigenvalue λ_f with $|\lambda_f| = 1$, so that f is not a coboundary. Let K_N be a random variable with uniform distribution on $\{1, 2, \ldots, N\}$. Then, if $\lambda_f \ne 1$, as $N \to \infty$,*

$$\frac{S_f(u_{\le K_N})}{\sqrt{\log_\lambda(N)}} \Rightarrow Z_f.$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

If $\lambda_f = 1$, *then*

$$\frac{S_f(u_{\leq K_N}) - \log_\lambda(N) \int_{\mathcal{A}} f(a)\mathfrak{n}(da)}{\sqrt{\log_\lambda(N)}} \Rightarrow Z_f.$$

The combination of Theorems 3.1, 3.2 and 3.3 allows us to give a new complete description of systems with *bounded discrepancy*. The first such description is due to [**Ada04**, Corollary 2], but it is non-trivial to see a direct derivation of either corollary from the other. We say that a fixed point u has bounded discrepancy if for every $a \in \mathcal{A}$, with $f_a : \mathcal{A} \to \mathbb{R}$ given by $f_a(b) = \mathbf{1}\{a = b\}$,

$$\sup_{N\in\mathbb{N}} |S_{f_a}(u_{\leq N}) - Nq(\{a\})| < \infty,$$

where q is the occurrence frequency of a, i.e. $q(a) = \lim_{N\to\infty} N^{-1} S_{f_a}(u_{\leq N})$.

Remark 3.5. The occurrence frequency q can be computed using the Perron–Frobenius eigenfunctions of M. Recall that $M^k_{a,b} = |\{j \in \mathbb{N} : (\theta^k(b))_j = a\}|$. Using

$$\lim_{k\to\infty} \lambda^{-k} M^k_{a,b} = \sigma(b)\rho(a)$$

(cf. (5)), it is easily seen that $q(a) = \rho(a)/\sum_{b\in\mathcal{A}} \rho(b)$.

COROLLARY 3.1. *Suppose that* u *is a fixed point of a primitive substitution* θ. *Then* u *has bounded discrepancy if and only if:*

(1) *all eigenvalues of* M *except the Perron–Frobenius eigenvalue have modulus less than or equal to* 1;
(2) *the geometric multiplicity of each eigenvalue of modulus* 1 *equals its algebraic multiplicity, i.e. each Jordan block in the Jordan form of* M *having eigenvalue of modulus* 1 *is one dimensional;*
(3) *each eigenfunction* f *with eigenvalue equal to* 1 *is a coboundary in the sense of Proposition 3.1.*

Proof. Let $\mathcal{W} \subset \{f : \int f\, dq = 0\}$ be all those functions so that

$$\sup_{N\in\mathbb{N}} |S_f(u_{\leq N})| < \infty.$$

Note that this is a vector space. Hence, u has bounded discrepancy if and only if $\dim \mathcal{W} = |\mathcal{A}| - 1$, as the functions $\{f_a - \int f_a\, dq : a \in \mathcal{A}\}$ span the space $\{f : \int f\, dq = 0\}$. Given a basis of generalized eigenfunctions $f_0, f_1, f_2, \ldots, f_r$, with f_0 the Perron–Frobenius eigenfunction, we have that $\int f_i\, dq = 0$, $(1 \leq i \leq r)$ and hence $\{f : \int f\, dq = 0\}$ is also spanned by $f_1, f_2, \ldots, f_r$. Hence, the necessity of the first and third conditions follows by Theorems 3.2 and 3.3. For the second condition, suppose that λ_f were an eigenvalue with $|\lambda_f| = 1$, so that, in the Jordan form of M, there is a non-trivial Jordan block. Then, by [**Ada04**, Theorem 1], there are functions with unbounded discrepancy. Hence, all three conditions are necessary.

Conversely, suppose that all three conditions are satisfied. Then we can give a basis of generalized eigenfunctions $f_0, f_1, f_2, \ldots, f_r$, where f_0 is the Perron Frobenius eigenfunction. For those that correspond to eigenvalues of modulus less than 1, it is easily checked using the path space decomposition that these have bounded discrepancy. For those with modulus 1, we have that their Birkhoff sums remain bounded as they are eigenfunction coboundaries. □

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

3.1. *Typical orbits.* So far, we have focused on proving theorems for a fixed point u of θ. We now show how Theorem 3.3 changes when instead of looking at u, we look at other sequences v from the orbit closure of u (see §4 for the relevant background).

THEOREM 3.4. *Let $(X_\theta, \mathcal{B}, \mu, T)$ be the substitution dynamical system arising from the primitive substitution θ. Let f be a left eigenfunction of M with eigenvalue λ_f with $|\lambda_f| = 1$ which is not a coboundary. Let K_N be a random variable with uniform distribution on $\{1, 2, \ldots, N\}$. Then, for every $v \in X_\theta$, there is a sequence $(a_{v,N})_{N=1}^{\infty}$ so that if $\lambda_f \neq 1$*

$$\frac{S_f(v_{\leq K_N}) - a_{v,N}}{\sqrt{\log_\lambda(N)}} \Rightarrow Z_f$$

as $N \to \infty$ and if $\lambda_f = 1$,

$$\frac{S_f(v_{\leq K_N}) - a_{v,N} - \log_\lambda(N) \int_{\mathcal{A}} f(a)\mathfrak{n}(da)}{\sqrt{\log_\lambda(N)}} \Rightarrow Z_f.$$

Further, for μ-almost every v, there is a constant $C > 0$ independent of v so that

$$\begin{aligned} \limsup_{N\to\infty} \frac{|a_{v,N} - \log_\lambda(N) \int_{\mathcal{A}} f(a)\mathfrak{n}(da)|}{\sqrt{\log N \log\log\log N}} &= C \quad \text{for } \lambda_f = 1, \\ \limsup_{N\to\infty} \frac{|a_{v,N}|}{\sqrt{\log N \log\log\log N}} &= C \quad \text{for } \lambda_f \neq 1. \end{aligned} \tag{16}$$

Hence, the fixed point differs from typical orbits in that $a_{v,N} \equiv 0$, while all orbits of eigenfunctions of modulus 1 give central limit theorems.

3.2. *Examples.*

Example 3.1. ($\lambda_f = 1$) Let $\mathcal{A} = \{a, b\}$ and θ_1, θ_2 and θ_3 be given by

$$\begin{aligned} \theta_1 &: a \to aab, \quad b \to bba, \\ \theta_2 &: a \to aab, \quad b \to bab, \\ \theta_3 &: a \to aba, \quad b \to bab. \end{aligned}$$

The θ_i-matrix M is given by

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Then $f = [1, -1]$ is a (left) eigenvector of M corresponding to the eigenvalue $\lambda_f = 1$. Using the information obtained in Example 2.1, we see that:

(i) θ_1: it has drift 0 from equation (11). Also, f is not a coboundary, otherwise

$$1 = f(\theta(a)_1) = h(\theta(a)_1) - h(\theta(a)_2) = h(a) - h(a) = 0,$$

which is impossible. So, we have a central limit theorem;

(ii) θ_2: similarly, we see that f is not a coboundary. In this case we see a logarithmic drift since

$$\begin{aligned} \int_{\mathcal{A}} f(a)\mathfrak{n}(da) &= \sum_{(a,j)\in\mathcal{X}} \hat{\sigma}(a, j)\hat{\rho}(a, j) S_f(\theta(a)_{<j}) \\ &= (0 + 1 + 2 + 0 - 1 + 1)\frac{1}{6} = \frac{1}{2}. \end{aligned}$$

So, we have a central limit theorem with non-zero drift;

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

(iii) θ_3: a fixed point of this substitution is periodic:

$$ababab\cdots.$$

So, there is no central limit theorem.

Example 3.2. (Rational eigenvalues) Let $\mathcal{A} = \{a, b\}$ and θ be given by

$$\theta : a \to abb, \quad b \to baa.$$

Then the θ-matrix M is

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

and eigenvalues and corresponding (left) eigenvectors of M are

$$\lambda = 3,\ \sigma = [1, 1], \quad \lambda_f = -1,\ f = [1, -1].$$

By a similar computation to part (i) of Example 3.1, we see that f is not a coboundary, so we have a central limit theorem.

Example 3.3. (Irrational eigenvalues) (Adapted from [**BBH14**, Proposition 4.1]). Let $\mathcal{A} = \{a, b, c, d\}$ and, for every $n \geq 1$, θ_n be given by

$$\theta_n = \begin{cases} a \to ad, \\ b \to adbbd, \\ c \to ad(bc)^{n+1}bd, \\ d \to ad(bc)^n bd. \end{cases}$$

Then the characteristic polynomial of the θ_n-matrix M_n has the roots $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ such that:

(i) $\lambda_4 = 1/\lambda_1$, $\lambda_3 = 1/\lambda_2$ and $\lambda_1 > |\lambda_2| = |\lambda_3| > \lambda_4$;

(ii) λ_2 and λ_3 are of modulus 1, so they are Salem numbers;

(iii) if $\lambda_2 = e^{2\pi i\alpha}$, then α is irrational.

In matrix theory, various authors have studied how to determine the d-tuples of complex numbers which can occur as the eigenvalues of a primitive matrix. Especially, Boyle and Handelman [**BH91**] formulated the 'spectral conjecture'. Later Kim, Ormes and Roush [**KOR00**] obtained the following result.

For $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_d)$, denote

$$\mathrm{tr}(\Lambda^n) = \sum_{i=1}^{d} (\lambda_i)^n \quad \text{and} \quad \mathrm{tr}_n(\Lambda) = \sum_{k|n} \mu\left(\frac{n}{k}\right) tr(\Lambda^k),$$

where μ is the Möbius function.

THEOREM 3.5. *Let $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_d)$ be a d-tuple of non-zero complex numbers with $|\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_d|$. There exists a primitive integer matrix A such that $\det(I - At) = \prod_{i=1}^{d}(1 - \lambda_i t)$ if and only if:*

(i) *the polynomial $\prod_{i=1}^{d}(1 - \lambda_i t)$ has integer coefficients;*

(ii) *$\lambda_1 > |\lambda_i|$ for $i = 2, 3, \ldots, d$*

(iii) *$\mathrm{tr}_n(\Lambda) \geq 0$ for all $n \geq 1$.*

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

Note that for a matrix A, there is $m \geq 0$ such that $\det(tI - A) = t^m \prod_{i=1}^{d}(t - \lambda_i)$ if and only if $\det(I - At) = \prod_{i=1}^{d}(1 - \lambda_i t)$.

Example 3.4. (Eigenfunction coboundaries) Let $\mathcal{A} = \{a, b, c, d\}$ and θ be given by

$$\theta : a \to ab, \quad b \to ca, \quad c \to cd, \quad d \to ac.$$

Then the θ-matrix M is

$$\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Eigenvalues of M are

$$\lambda_1 = 2, \quad \lambda_2 = -1, \quad \lambda_3 = 1, \quad \lambda_4 = 0$$

and corresponding (left) eigenvectors of M are

$$\sigma = [1, 1, 1, 1], \quad [-1, 2, -1, 2], \quad [-1, 0, 1, 0], \quad [1, -1, -1, 1].$$

A right Perron–Frobenius eigenvector ρ is given by $\rho^t = [2/6, 1/6, 2/6, 1/6]$, so that $\sum \sigma(a)\rho(a) = 1$. Then we have that $\mathfrak{m}$ has the form

$$(1/6, 1/6, 1/12, 1/12, 1/6, 1/6, 1/12, 1/12)$$

on the state space $\mathcal{X} = \{(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2), (d, 1), (d, 2)\}$.

(i) $\lambda_f = 1$ and $f = [-1, 0, 1, 0]$. One has

$$\int_{\mathcal{A}} f(a)\mathfrak{n}(da) = 0$$

and the function $h : \mathcal{A} \to \mathbb{R}$ such that $h(a) = h(d) = 0$, $h(b) = h(c) = 1$ satisfies the conditions (iii) and (iv) in Proposition 3.1, so f is a coboundary.

(ii) $\lambda_f = -1$, $f = [-1, 2, -1, 2]$. We can obtain

$$\int_{\mathcal{A}} f(a)\mathfrak{n}(da) = -\frac{1}{2}$$

and it is not difficult to show that f is not a coboundary by checking condition (iii) in Proposition 3.1.

Example 3.5. (Fibonacci substitution) The Fibonacci substitution θ is given by

$$\theta : a \to ab, \quad b \to a.$$

The θ-matrix of the substitution is

$$M = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Let $\alpha = (-1 + \sqrt{5})/2$. Then M has eigenvalues $\lambda_1 = 1 + \alpha$ and $\lambda_2 = -\alpha$.

Let $u = (u_n)_{n=1}^{\infty} = \lim_{n\to\infty} \theta^n(a)$. The sequence u also can be obtained by a rotation by α (cf. [**Fog02**, Proposition 5.4.9]):

$$u_n = a \quad \text{if } \{n\alpha\} \in [1 - \alpha, 1), \quad u_n = b \quad \text{if } \{n\alpha\} \in [0, 1 - \alpha).$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

Let f be a function such that $f(a) = 0 - (1-\alpha)$ and $f(b) = 1 - (1-\alpha)$. Then, for $I = [0, 1-\alpha)$,

$$Z_\alpha(n; I) - n(1-\alpha) = \sum_{k=1}^{n} f(u_k) = S_f(u_{\le n}).$$

Moreover, $f = (-1+\alpha, \alpha)$ is an eigenvector corresponding to the eigenvalue $\lambda_2 = -\alpha$, so $|\lambda_2| < 1$. For Beck's case, if $I = [0, x)$ for $x \in \mathbb{Q}$, then we have a central limit theorem for $Z_\alpha(n; I) - n(1-\alpha)$. From Theorem 3.1, when $x = 1-\alpha$, the limiting distribution of $Z_\alpha(n; I) - n(1-\alpha)$ is supported on a Cantor set and has a different normalization than when $x \in \mathbb{Q}$.

4. *Substitution dynamical systems*

Let $\mathcal{A}$ be a finite set of letters, endowed with the discrete topology, and let $\mathcal{A}^{\mathbb{Z}}$ have the product topology, so that $\mathcal{A}^{\mathbb{Z}}$ is a compact metric space and the shift map T given by $(Tu)_n = u_{n+1}$ is a homeomorphism. The pair $(\mathcal{A}^{\mathbb{Z}}, T)$ is called the full shift on the alphabet $\mathcal{A}$. If X is a closed T-invariant subset of $\mathcal{A}^{\mathbb{N}}$, the pair (X, T) is called a *subshift*.

Given $x \in \mathcal{A}^{\mathbb{Z}}$, let $L(x)$ be the set of all finite words appearing in x. The language of θ, denoted by L_θ, is the set of all finite words occurring in $\theta^n(a)$ for some $n \ge 0$ and $a \in \mathcal{A}$. Let $X_\theta = \{x \in \mathcal{A}^{\mathbb{Z}} : L(x) \subset L(\theta)\}$. Then X_θ is closed in $\mathcal{A}^{\mathbb{Z}}$ and invariant under the shift. We denote by T the restriction of the shift to X_θ. The pair (X_θ, T) is called the *(two-sided) substitution subshift* associated to θ. It is known that (X_θ, T) is minimal and uniquely ergodic (see [**Fog02**] or [**Que10**]).

Remark 4.1. For a given substitution θ, there exist two letters $a, b \in \mathcal{A}$ and $k \in \mathbb{N}$ such that:

- a is the last word of $\theta^k(a)$;
- b is the first word of $\theta^k(b)$;
- $ab \in L(\theta)$.

Then there exists $v \in \mathcal{A}^{\mathbb{Z}}$ such that $v_{-1} = a$, $v_0 = b$ and $\theta^k(v) = v$. We say that v is a *(two-sided) fixed point* of θ. In this case we have $X_\theta = \overline{\{T^n v : n \in \mathbb{Z}\}}$.

Remark 4.2. We also can define the *one-sided substitution subshift* associated to θ by the following. By the construction above, there is a fixed point u for θ. Let $\tilde{X}_\theta$ be the orbit closure $\overline{\{T^n u : n \in \mathbb{N}_0\}}$ (where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$). The pair $(\tilde{X}_\theta, T)$ is the substitution subshift. This definition can be checked to be independent of the choice of fixed point u. The projection $\pi : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{N}}$ maps X_θ onto $\tilde{X}_\theta$ and (X, T, π) is the natural extension of $(\tilde{X}_\theta, T)$, that is, for every dynamical system (Y, S) and every factor map $\phi : Y \to \tilde{X}_\theta$, there exists a unique factor map $\psi : Y \to X_\theta$ with $\pi \circ \psi = \phi$. See [**DHS99**] for details.

4.1. *Desubstitution.*

Let θ be a primitive substitution with non-periodic fixed point. The result on recognizability by [**Mos92**, **Mos96**] allows one to desubstitute w in X_θ (cf. [**Fog02**] for details):

$$w = \cdots w_{-m} \cdots w_{-1} w_0 w_1 \cdots w_n \cdots = \cdots \theta(y_{-1})\theta(y_0)\theta(y_1) \cdots,$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

where w_0 lies in $\theta(y_0)$. Furthermore, by [**CS01**], there exists a unique sequence $(p_i, c_i, s_i)_{i\in\mathbb{N}_0} \in (\mathcal{A}^* \times \mathcal{A} \times \mathcal{A}^*)^{\mathbb{N}_0}$ such that $\theta(c_{i+1}) = p_i c_i s_i$ and

$$w = \cdots \theta^2(p_2)\theta(p_1)p_0.c_0s_0\theta(s_1)\theta^2(s_2)\cdots,$$

which is called a prefix–suffix decomposition of w. If only finitely many s_i are non-empty, then there exist $a \in \mathcal{A}$ and $l, s \in \mathbb{N}$ such that

$$x_{[0,\infty)} = c_0s_0\theta(s_1)\theta^2(s_2)\cdots\theta^l(s_l)\lim_{n\to\infty}\theta^{ns}(a).$$

Similarly, if only finitely many p_i are non-empty, then there exist $b \in \mathcal{A}$ and $m, t \in \mathbb{N}$ such that

$$x_{(-\infty,-1]} = \lim_{n\to\infty}\theta^{nt}(b)\theta^m(p_m)\cdots\theta(p_1)p_0.$$

4.2. *Adic transformations.* As $p \to \infty$, MPM_p induces the Markov measure MPM_∞ on the infinite path space $\mathcal{X}^{*,\infty}$. Following Livshits [**Liv88**], we define the adic transformation T_A on $\mathcal{X}^{*,\infty}$ as follows: given $\mathbf{x} = (v_1, k_1)(v_2, k_2)(v_3, k_3)\cdots \in \mathcal{X}^{*,\infty}$:

- if $k_1 < |\theta(v_2)|$, then

$$T_A(\mathbf{x}) = (v_1, k_1 + 1)(v_2, k_2)(v_3, k_3)\cdots;$$

- otherwise, let ℓ be the smallest positive integer such that $k_\ell < |\theta(v_{\ell+1})|$; then

$$T_A(\mathbf{x}) = (b_1, 1)(b_2, 1)\cdots(b_{\ell-1}, 1)(b_\ell, k_\ell + 1)(v_{\ell+1}, k_{\ell+1})(v_{\ell+2}, k_{\ell+2})\cdots,$$

  where $b_\ell = \theta(v_{\ell+1})_{j_\ell+1}$ and $b_i = \theta(b_{i+1})_1$ for $1 \le i \le \ell - 1$.

It is known that the adic transformation is uniquely ergodic and it is measurably isomorphic to the substitution subshift.

Consider the set C of the form

$$C = [(v_1, k_1)\cdots(v_m, k_m)] \setminus [(v_1, k_1)\cdots(v_{m-1}, k_{m-1})(v_m, |\theta(v_m)|)(v_{m+1}, k_{m+1})]. \tag{17}$$

Then, for large N, $\nu_N(C) = \nu_N(T_A(C))$. Since the sets of the form (17) generate the Borel σ-algebra on MPM_∞, by Proposition 5.3 below, T_A is invariant with respect to MPM_∞. Then $(\mathcal{X}^{*,\infty}, \mathrm{MPM}_\infty, T_A)$ is measurably isomorphic to the substitution subshift (X_θ, μ, T).

Also, as $p \to \infty$, SMPM_p induces the stationary Markov measure SMPM_∞ on the infinite path space $\mathcal{X}^{*,\infty}$ with the invariant measure $\mathfrak{m}$ given by $\mathfrak{m}(y) = \hat{\sigma}(y)\hat{\rho}(y)$ and the transition probability given by $p(y, z) = \mathrm{SSIM}^1_{y,z}\,\hat{\rho}(z)/\lambda\hat{\rho}(y)$, where λ is the Perron–Frobenius eigenvalue of $\mathrm{SSIM}^1_{\cdot,\cdot}$ and $\hat{\sigma}$ and $\hat{\rho}$ are corresponding left and right eigenvectors with $\sum_y \hat{\sigma}(y)\hat{\rho}(y) = 1$. It is known [**Par64**] that SMPM_∞ is the maximal measure for the topological Markov shift on $\mathcal{X}^{*,\infty}$.

Remark 4.3. We will now discuss how [**Buf14**, Theorem 5] can be applied to show distributional convergence for $S_f(w_{\le N})$ when w is chosen according to μ. To relate what we do to [**Buf14**], we need to consider a two-sided Markov compactum. Hence, we introduce $\mathcal{X}^{*,\mathbb{Z}} \subseteq \mathcal{X}^{\mathbb{Z}}$ as the collection of all sequences $\{(v_i, k_i)\}_{-\infty}^{\infty}$ so that for all $i \in \mathbb{Z}$, $v_i = \theta(v_{i+1})_{k_{i+1}}$. This space supports a natural Markov measure that extends SMPM_∞.

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

It can be defined as the limit of uniform measures on admissible paths $\{(v_i, k_i)\}_{-K}^{K}$, $K \to \infty$, or as the extension of SMPM_∞ to a two-sided Markov chain using the reversed transition matrix.

Inside of $\mathcal{X}^{*,\mathbb{Z}}$, we define a copy of $\mathcal{X}^{*,\infty}$ by the injective map

$$\iota(\{(v_i, k_i)\}_1^\infty) = \ldots (v_{-2}, 1)(v_{-1}, 1)(v_0, 1)(v_1, k_1)(v_2, k_2)\ldots,$$

where $(v_j)_{-\infty}^0$ are chosen to make the sequence admissible. The $(v_0, 1)$ term occupies the 0 coordinate of the sequence. On this two-sided Markov compactum, it is possible to define a continuous flow $h_t^+(\mathbf{x})$, following the notation of [**Buf14**], with respect to an order topology (see [**Ito78**, **Ver95**] or [**SS14**]). For our purposes, it is enough to know that $\iota(T_A\mathbf{x}) > \iota(\mathbf{x})$ and to compute the length of time required in the flow so that $h_t^+(\iota(\mathbf{x})) = \iota(T_A\mathbf{x})$. Specifically, we have that

$$h^+_{\hat{\sigma}(x_0)}(\iota(\mathbf{x})) = \iota(T_A\mathbf{x}). \tag{18}$$

Hence, pulling these dynamics back to $\mathcal{X}^{*,\infty}$, we get the suspension flow over T_A with roof function $\hat{\sigma}$. (To verify (18), observe that the Q matrix in [**Buf14**] is the transpose of $\mathrm{SSIM}^1_{\cdot,\cdot}$. Hence, the right eigenvector of Q is $\hat{\sigma}$. Further, as in Theorem 3.2, we pick $\hat{\rho}$ to have the normalization $\sum \hat{\rho}(a) = 1$. In this case $\Phi_1^+(\iota(\mathbf{x}), \iota(T_A\mathbf{x})) = \hat{\sigma}(x_0)$.)

This identification allows us to express Birkhoff sums in the substitution dynamical system as integrals over the flow h_t^+. Specifically, for any function $f : \mathcal{A} \to \mathbb{C}$ and any $w \in X_\theta$,

$$\sum_{i=1}^{N} f((T^i w)_0) = \sum_{i=1}^{N} \hat{f}((T_A^i \mathbf{x})_0),$$

where for $(v, k) \in \mathcal{X}$, $\hat{f}(v, k) = f(v)$. Moreover,

$$\sum_{i=1}^{N} f((T^i w)_0) = \int_0^{t(N)} \frac{\hat{f}((h_s^+(x))_0)}{\hat{\sigma}((h_s^+(x))_0)}\, ds,$$

where $t(N) = \sum_{i=1}^{N} \hat{\sigma}((T_A^i\mathbf{x})_0)$. We can now apply [**Buf14**, Theorems 5 and 6] to show that for an eigenfunction f with eigenvalue λ_f having $|\lambda_f| > 1$,

$$\frac{\sum_{i=1}^{[\tau\lambda^\ell]} f((T^i w)_0)}{\lambda_f^\ell} \Rightarrow_{\ell\to\infty} \alpha(f)\eta_f(\tau),$$

where η_f and $\alpha(f)$ have an explicit description in terms of the Markov compactum (see [**Buf14**, Theorem 6 and subsequent Remark]).

5. *Comparing measures on path space*

Recall that for $\mathbf{x} = x_1 x_2 \cdots x_p \in \mathcal{X}^{*,p}$,

$$\mathrm{MPM}_p(\{\mathbf{x}\}) = \mathfrak{u}(\{x_1\}) \prod_{i=1}^{p-1} \mathfrak{p}(x_i, x_{i+1}).$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

We will show that both of $\mathrm{UPM}_{y,p}$ and MPM_p are very similar for large p. To compare them, we recall the notion of total variation distance. For two measures μ and ν on a common measure space (X, Ω), the total variation distance d_{TV} is given by

$$d_{TV}(\mu, \nu) = \sup_{A \in \Omega} |\mu(A) - \nu(A)|.$$

The following is now an exercise in *coupling* (see [**LPW09**, Chs. 4 and 5] for an introduction).

PROPOSITION 5.1. *For any $r < p \in \mathbb{N}$, let $S^r : \mathcal{X}^{*,p} \to \mathcal{X}^{*,p-r}$ be given by*

$$S^r(x_1 x_2 \dots x_p) = x_1 x_2 \dots x_{p-r}.$$

For every $c_1 > 0$, there is a constant $c_2 > 0$ such that for all $p \in \mathbb{N}$, all $y \in \mathcal{X}$ and all integers $r > c_2 \log p$,

$$d_{TV}(\mathrm{MPM}_p \circ S^r, \mathrm{UPM}_{y,p} \circ S^r) < p^{-c_1}.$$

Proof. We begin by defining coupling of two probability measures on a common probability space. Suppose that ν_1 and ν_2 are probability measures on a probability space $(X, \mathcal{B})$. A coupling of ν_1 and ν_2 is a probability measure γ on the product space $(X \times X, \mathcal{B} \otimes \mathcal{B})$ such that marginals are ν_1 and ν_2. The total variation norm of ν_1 and ν_2 can be expressed in terms of couplings of ν_1 and ν_2. Specifically, by [**LPW09**, Proposition 4.7],

$$d_{TV}(\nu_1, \nu_2) = \inf_{\gamma:\ \text{couplings}} \gamma\{(x, z) \in X \times X : x \neq z\}.$$

To bound the total variation norm of $\mathrm{MPM}_p \circ S^r$ and $\mathrm{UPM}_{y,p} \circ S^r$, it suffices to construct a coupling γ such that $\gamma\{(x, y) \in \mathcal{X}^{*,p} \times \mathcal{X}^{*,p} : x \neq y\} < p^{-c_1}$. We recall for convenience (4), which states that for $\mathbf{x} = x_1 x_2 \cdots x_p$,

$$\begin{aligned}&\mathrm{UPM}_{y,p}(\{\mathbf{x}\})\\ &= \frac{\mathrm{SSIM}^{p-1}_{x_1,y}}{\mathrm{SSIM}^{p-1}_{*,y}} h^y_p(x_1, x_2) h^y_{p-1}(x_2, x_3) \dots h^y_3(x_{p-2}, x_{p-1}) \frac{\mathrm{SSIM}^1_{x_{p-1},x_p}}{\mathrm{SSIM}^{p-1}_{x_{p-1},y}} \mathbf{1}\{x_p = y\}.\end{aligned}$$

For a fixed $x \in \mathcal{X}$, $h^y_k(x, \cdot)$ and $\mathfrak{p}(x, \cdot)$ are probability measures. For any pair $(x, z) \in \mathcal{X}^2$ and any $2 < k \leq p$, we define a coupling $G_k((x, z), (\cdot, \cdot))$ of $h^y_k(x, \cdot)$ and $\mathfrak{p}(z, \cdot)$ such that $G_k((x, z), (\cdot, \cdot))$ attains $d_{TV}(h^y_k(x, \cdot), \mathfrak{p}(z, \cdot))$. This coupling can in fact be given explicitly; see [**LPW09**, Remark 4.8]. We also let $G_2((x, z), (\cdot, \cdot))$ be any coupling of δ_y and $\mathfrak{p}(z, \cdot)$, and we let H_p be the coupling that attains the total variation distance of the measures

$$\frac{\mathrm{SSIM}^{p-1}_{\cdot,y}}{\mathrm{SSIM}^{p-1}_{*,y}} \quad \text{and} \quad \mathfrak{u}(\{\cdot\}).$$

Hence, by Perron–Frobenius theory, we have

$$H_p(x \neq z) \ll e^{-cp}.$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

Now we define the coupling γ on $\mathcal{X}^{*,p} \times \mathcal{X}^{*,p}$:

$$\gamma(\mathbf{x}, \mathbf{z}) = H_p((x_1, z_1)) \prod_{i=1}^{p-1} G_{p-i+1}((x_i, z_i), (x_{i+1}, z_{i+1})),$$

so that γ is the law of a Markov chain on $\mathcal{X}^{*,p} \times \mathcal{X}^{*,p}$.

For any measures μ_1, μ_2 on a countable space X, we have by [**LPW09**, Proposition 4.2] that $d_{TV}(\mu_1, \mu_2) = \frac{1}{2} \sum_{x \in X} |\mu_1(x) - \mu_2(x)|$. Hence, for all $k \le p$, by (6),

$$d_{TV}(\mathfrak{p}(a, \cdot), h_k^y(a, \cdot)) \le |\mathcal{X}| e^{-ck}.$$

Then

$$\begin{aligned}
\gamma(\exists j \le p - r : x_j \ne z_j) &= \sum_{j=1}^{p-r} \gamma(x_j \ne z_j, x_k = z_k \quad (\text{for all } k < j)) \\
&\le \gamma(x_1 \ne z_1) + \sum_{j=2}^{p-r} \gamma(x_j \ne z_j | x_{j-1} = z_{j-1}) \\
&\le H_p(x_1 \ne z_1) + \sum_{j=2}^{p-r} \sup_{x \in \mathcal{X}} d_{TV}(\mathfrak{p}(x, \cdot), h_{p-j+2}(x, \cdot)) \\
&\ll e^{-cr}.
\end{aligned}$$

Taking $r = C \log p$ for sufficiently large C completes the proof. □

We can also compare SMPM_p and MPM_p in a similar way, which is a standard result on primitive Markov chains.

PROPOSITION 5.2. *For any $r < p \in \mathbb{N}$, let $L^r : \mathcal{X}^{*,p} \to \mathcal{X}^{*,p-r}$ be given by*

$$L^r(x_1 x_2 \dots x_p) = x_{r+1} x_{r+2} \dots x_p.$$

For every $c_1 > 0$, there is a constant $c_2 > 0$ such that for all $p \in \mathbb{N}$ and all integers $r > c_2 \log p$,

$$d_{TV}(\mathrm{MPM}_p \circ L^r, \mathrm{SMPM}_p \circ L^r) < p^{-c_1}.$$

For a proof, see [**LPW09**, (5.2)].

As a consequence of Propositions 5.1 and 2.1, we have the following result.

PROPOSITION 5.3. *For every $c_1 > 0$, there is a constant $c_2 > 0$ so that for all $p \in \mathbb{N}$, $a \in \mathcal{A}$, $N \in \mathbb{N}$ with $|\theta^{p-1}(a)| \le N < |\theta^p(a)|$ and all integers $r > c_2 \log p$,*

$$d_{TV}(\mathrm{MPM}_p \circ S^r, \nu_N \circ S^r) \ll p^{-c_1}.$$

Proof. By Proposition 2.1,

$$\nu_N \circ S^r = \sum_{\ell=1}^{p} \sum_{1 \le j < k_\ell} \frac{\mathrm{SSIM}^{\ell}_{*,(v_\ell, j)}}{N} \mathrm{UPM}_{(v_\ell, j), \ell} \circ S^{(r-(p-l))_+}.$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

We also have that

$$\mathrm{MPM}_p \circ S^r = \sum_{\ell=1}^{p} \sum_{1\le j<k_\ell} \frac{\mathrm{SSIM}^{\ell}_{*,(v_\ell,j)}}{N} \mathrm{MPM}_p \circ S^r.$$

Note that by stationarity, $\mathrm{MPM}_p \circ S^r = \mathrm{MPM}_{p-r}$.

For any event A and large r as in Proposition 5.1, let $r_0 = [r/2]$. Then

$$\begin{aligned}
&|\nu_N \circ S^r(A) - \mathrm{MPM}_p \circ S^r(A)| \\
&\quad \le \sum_{\ell=1}^{p-r_0} \sum_{1\le j<k_\ell} \frac{\mathrm{SSIM}^{\ell}_{*,(v_\ell,j)}}{N} \\
&\qquad + \sum_{\ell=p-r_0}^{p} \sum_{1\le j<k_\ell} \frac{\mathrm{SSIM}^{\ell}_{*,(v_\ell,j)}}{N} |\mathrm{UPM}_{(v_\ell,j),\ell} \circ S^{r-(p-l)}(A) - \mathrm{MPM}_l \circ S^{r-(p-l)}(A)| \\
&\quad \le \frac{|\theta^{p-r_0}(a)|}{N} + (p-r_0)^{-c_1},
\end{aligned}$$

where we have applied Proposition 5.1 to the third line. Then, for some $\alpha > 0$,

$$\frac{|\theta^{p-r_0}(a)|}{N} \ll \frac{1}{|\lambda^{r_0-1}|} \ll p^{-\alpha c_2}.$$

Picking c_2 sufficiently large, the result follows. □

6. *Proofs for* $|\lambda_f| = 1$

Theorem 3.3 will follow immediately from Proposition 6.1 combined with Proposition 3.1. However, Proposition 3.1 relies on Proposition 6.1, so we present Proposition 6.1 first.

PROPOSITION 6.1. *Suppose that f is a left eigenfunction of M with eigenvalue λ_f with $|\lambda_f| = 1$ for which there is no function $h : \mathcal{A} \to \mathbb{C}$ so that for all $c \in \mathcal{A}$,*

$$S_f(\theta(c)_{<j}) = \int_{\mathcal{A}} f(b)\mathfrak{n}(db) + h(c) - \lambda_f^{-1} h(\theta(c)_j).$$

Let K_N be a random variable with uniform distribution on $\{1, 2, \ldots, N\}$ and let $a \in \mathcal{A}$ be fixed. Then if $\lambda_f \ne 1$, as $N \to \infty$,

$$\sup_{\substack{\ell\in\mathbb{N} \\ |\theta^\ell(a)|\ge N}} d_{BL}\left(\frac{S_f(\theta^\ell(a)_{\le K_N})}{\sqrt{\log_\lambda(N)}}, Z_f \right) \to 0.$$

If $\lambda_f = 1$, as $N \to \infty$,

$$\sup_{\substack{\ell\in\mathbb{N} \\ |\theta^\ell(a)|\ge N}} d_{BL}\left(\frac{S_f(\theta^\ell(a)_{\le K_N}) - \log_\lambda(N) \int_{\mathcal{A}} f(b)\mathfrak{n}(db)}{\sqrt{\log_\lambda(N)}}, Z_f \right) \to 0.$$

Proof. Given N, one can find p_N such that $|\theta^{p_N-1}(a)| < N \le |\theta^{p_N}(a)|$. Now we define a function $\check{f}$ on $\mathcal{X}$ by $\check{f}(v_i, k_i) = S_f(\theta(v_i)_{<k_i})$. Then we have for any $1 \le K \le N$ that

$$\lambda_f S_f(\theta^\ell(a)_{<K}) = \sum_{i=1}^{\ell} \lambda_f^i \check{f}((v_i, k_i)),$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

where $\Psi_{a,\ell}(K) = (v_1, k_1)(v_2, k_2) \dots (v_\ell, k_\ell)$. For $m > p_N$, we have that $k_m = 1$. Hence, $\check{f}(v_m, k_m) = 0$ for all these m, and we have

$$\lambda_f S_f(\theta^\ell(a)_{<K}) = \sum_{i=1}^{p_N} \lambda_f^i \check{f}((v_i, k_i)),$$

We will show a central limit theorem for $\lambda_f S_f(\theta^\ell(a)_{<K_N}) - \sum_{i=1}^{p_N} \lambda_f^i \int \check{f}\, d\mathfrak{m}/\sqrt{\log_\lambda N}$. The desired central limit theorems follow immediately from this.

Let $Z = (1/\sqrt{\log_\lambda N}) \sum_{i=1}^{p_N} \lambda_f^i (\check{f}((v_i, k_i)) - \int \check{f}\, d\mathfrak{m})$, which is a function on $\mathcal{X}^{*,p}$. Then we have that

$$\frac{\lambda_f S_f(\theta^\ell(a)_{<K}) - \sum_{i=1}^{p_N} \lambda_f^i \int \check{f}\, d\mathfrak{m}}{\sqrt{\log_\lambda(N)}} = Z(\Psi_{a,p_N}(K)).$$

By definition of ν_N, we therefore have that for any bounded Lipschitz function ϕ,

$$\begin{aligned}
&\frac{1}{N} \sum_{K=1}^{N} \phi\left(\frac{\lambda_f S_f(\theta^\ell(a)_{<K}) - \sum_{i=1}^{p_N} \lambda_f^i \int \check{f}\, d\mathfrak{m}}{\sqrt{\log_\lambda(N)}} \right) \\
&\quad = \mathbb{E}\phi\left(\frac{\lambda_f S_f(\theta^\ell(a)_{<K_N}) - \sum_{i=1}^{p_N} \lambda_f^i \int \check{f}\, d\mathfrak{m}}{\sqrt{\log_\lambda(N)}} \right) \\
&\quad = \mathbb{E}\phi(Z(\Psi_{a,p_N}(K_N))) \\
&\quad = \int_{\mathcal{X}^{*,p}} \phi(Z(\mathbf{x})) \nu_N(d\mathbf{x}).
\end{aligned}$$

We will show that for any bounded Lipschitz function ϕ,

$$\int_{\mathcal{X}^{*,p}} \phi(Z(\mathbf{x})) \nu_N(d\mathbf{x}) \to \int_{\mathbb{C}} \phi(x) \Phi(dx),$$

where Φ is the probability measure given by $\Phi(A) = \mathbb{P}[Z_f \in A]$.

Let $r = [(\log p_N)^2]$ and write $Z = X + Y$, where

$$Y = \frac{1}{\sqrt{\log_\lambda N}} \sum_{i=r+1}^{p_N - r} \lambda_f^i \left(\check{f}((v_i, k_i)) - \int \check{f}\, d\mathfrak{m} \right).$$

Hence, $\|X\|_\infty \ll (\log\log N)^2/\sqrt{\log N}$. Then

$$\int \phi(Z)\, d\nu_N = \int \phi(Y)\, d\nu_N + \int \phi(Z) - \phi(Y)\, d\nu_N.$$

As ϕ is Lipschitz, the second integral is at most $\|\phi\|_{\mathrm{Lip}} \|X\|_\infty = o(1)$. Hence

$$\begin{aligned}
\int \phi(Z)\, d\nu_N &= \int \phi(Y \circ S^r \circ L^r)\, d(\nu_N \circ S^r \circ L^r) + o(1) \\
&= \int \phi(Y \circ S^r \circ L^r)\, d\, \mathrm{SMPM}_{p_N} \circ S^r \circ L^r \\
&\quad + O(d_{TV}(\nu_N \circ S^r \circ L^r, \mathrm{SMPM}_{p_N} \circ S^r \circ L^r)) + o(1).
\end{aligned}$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

By Propositions 5.3 and 5.2, $d_{TV}(\nu_N \circ S^r \circ L^r, \mathrm{SMPM}_{p_N} \circ S^r \circ L^r) \to 0$ as $N \to \infty$. Therefore,

$$\int \phi(Z)\, d\nu_N = \int \phi(Y \circ S^r \circ L^r)\, d\,\mathrm{SMPM}_{p_N-2r} + o(1).$$

Again using that X is uniformly small and the Lipschitzness of ϕ, we conclude that

$$\int \phi(Z)\, d\nu_N = \int \phi(Z)\, d\,\mathrm{SMPM}_{p_N} + o(1).$$

All said, we have shown that

$$d_{\mathrm{BL}}(\nu_N \circ Z^{-1}, \mathrm{SMPM}_{p_N} \circ Z^{-1}) \to 0$$

as $N \to \infty$ uniformly in ℓ.

Now the theorem follows from Theorem A.1 and the observation that $p_N/\log_\lambda N \to 1$, provided that we show that there is no $\check{h}$ satisfying $P^*P\check{h} = \check{h}$ so that $\check{f} = \int \check{f}(x)\mathfrak{m}(dx) + \check{h} - \lambda_f P\check{h}$, where $\mathfrak{p}^*$ and $\mathfrak{p}$ are defined in (13) and (7) and

$$(Ph)(x) = \sum_{x\in\mathcal{X}} \mathfrak{p}(x, y)h(y) \quad \text{and} \quad (P^*h)(x) = \sum_{x\in\mathcal{X}} \mathfrak{p}^*(x, y)h(y).$$

Suppose that there were such an $\check{h}$. Then, by Theorem A.1, we have that

$$W = \sum_{i=1}^{p_N} \lambda_f^i \left(\check{f}((v_i, k_i)) - \int \check{f}\, d\mathfrak{m} \right)$$

is uniformly bounded in N for SMPM_{p_N}-almost every path. As SMPM_p has full support on $\mathcal{X}^{*,p}$, for every p, we have that there is a C so that

$$\sup_{N>0} \sup_{(v_i,k_i)_{i=1}^{p_N}\in\mathcal{X}^{*,p_N}} |W((v_i, k_i)_{i=1}^{p_N})| < C.$$

Hence, we also have that

$$\sup_{N>0} \sup_{(v_i,k_i)_{i=1}^{p_N}\in\mathcal{X}^{*,p_N}} |\lambda_f^{-p_N-1} W((v_i, k_i)_{i=1}^{p_N})| < C.$$

Observe that

$$\lambda_f^{-p_N-1} W = \sum_{i=1}^{p_N} \lambda_f^{-(p_N-i+1)} \left(\check{f}((v_i, k_i)) - \int \check{f}\, d\mathfrak{m} \right).$$

Hence, by Theorem A.1 applied to the reversed chain with transition matrix $\mathfrak{p}^*$, there must be an $\hat{h}$ so that $PP^*\hat{h} = \hat{h}$ and $\check{f} = \int \check{f}\, d\mathfrak{m} + \hat{h} - \lambda_f^{-1} P^*\hat{h}$.

We now turn to characterizing those $\hat{h}$ for which $PP^*\hat{h} = \hat{h}$. First, we evaluate PP^*: by equations (13) and (7),

$$\begin{aligned} \mathfrak{p}\mathfrak{p}^*((a, j), (b, k)) &= \sum_{(c,\ell)\in\mathcal{X}} \mathfrak{p}((a, j), (c, \ell))\mathfrak{p}^*((c, \ell), (b, k)) \\ &= \sum_{(c,\ell)\in\mathcal{X}} \frac{\mathbf{1}\{\theta(c)_\ell = a\}\hat{\rho}(c, \ell)}{\lambda\hat{\rho}(a, j)} \frac{\mathbf{1}\{\theta(c)_\ell = b\}\hat{\sigma}(b, k)}{\lambda\hat{\sigma}(c, \ell)}. \end{aligned}$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

Observe that this is non-zero if and only if $a = b$. Hence, $\mathfrak{pp}^*$ is a block matrix, with $|\mathcal{A}|$ many blocks, each of which is positive. Moreover, we have that $\mathfrak{pp}^*$ is right stochastic, as it is the product of two right stochastic matrices, and hence

$$1 = \sum_{(b,k)\in\mathcal{X}} \mathfrak{pp}^*((a,\, j),\, (b,\, k)) = \sum_{k=1}^{|\theta(a)|} \mathfrak{pp}^*((a,\, j),\, (a,\, k)).$$

Thus, each of these blocks is itself right stochastic. As a consequence, the eigenspace of $\mathfrak{pp}^*$ with eigenvalue 1 has dimension $|\mathcal{A}|$ and is spanned by those functions $(a,\, j) \mapsto h(a)$, where $h : \mathcal{A} \to \mathbb{C}$ is any function.

Hence, we have that $\hat{h}(a,\, j) = h(a)$ for some function h. Evaluating $P^*\hat{h}$, we have

$$\begin{aligned} P^*\hat{h}(a,\, j) &= \sum_{(b,k)\in\mathcal{X}} \mathfrak{p}^*((a,\, j),\, (b,\, k))h(b) \\ &= \sum_{(b,k)\in\mathcal{X}} \frac{\mathbf{1}\{\theta(a)_j = b\}\hat{\sigma}(b,\, k)}{\lambda\hat{\sigma}(a,\, j)} h(b) \\ &= h(\theta(a)_j) \sum_{(b,k)\in\mathcal{X}} \frac{\mathbf{1}\{\theta(a)_j = b\}\hat{\sigma}(b,\, k)}{\lambda\hat{\sigma}(a,\, j)} \\ &= h(\theta(a)_j) \sum_{(b,k)\in\mathcal{X}} \mathfrak{p}^*((a,\, j),\, (b,\, k)) \\ &= h(\theta(a)_j). \end{aligned}$$

Therefore, we have that

$$\check{f}(a,\, j) = \int \check{f}(x)\mathfrak{m}(dx) + h(a) - \lambda_f^{-1} h(\theta(a)_j),$$

which contradicts the hypothesis on f that we assumed in the statement of the proposition. □

7. *Proof of coboundary proposition*

Proof of Proposition 3.1. (i) ⇔ (ii): This follows from that (X_θ, T) is minimal.
(iii) ⇔ (iv): This follows from the identities $f(\theta(a)_j) = S_f(\theta(a)_{<j+1}) - S_f(\theta(a)_{<j})$ and $S_f(\theta(a)_{<1}) = 0$.
(i) ⇒ (iii): By Proposition 6.1, if (iii) does not hold, then for a fixed point $u = u_1u_2, \ldots,$ $\limsup_{N\to\infty} |S_f(u_{\le N})| = \infty$.
(iii) ⇒ (ii): Let u be a fixed point. Then, by (3), for each N there exist $(v_1, k_1), \ldots,$ $(v_{p_N}, k_{p_N}) \in \mathcal{X}$ so that

$$S_f(u_{<N}) = \sum_{i=1}^{p_N} \lambda_f^{i-1} S_f(\theta(v_i)_{<k_i}).$$

Then, since $v_i = \theta(v_{i+1})_{k_i+1}$,

$$\begin{aligned} S_f(u_{<N}) &= \sum_{i=1}^{p_N} \lambda_f^{i-1} \left[\int_{\mathcal{A}} f(c)\mathfrak{n}(dc) + h(v_i) - \lambda_f^{-1} h(\theta(v_i)_{k_i}) \right] \\ &= \sum_{i=1}^{p_N} \left[\lambda_f^{i-1} \int_{\mathcal{A}} f(c)\mathfrak{n}(dc) \right] - \lambda_f^{-1} h(\theta(v_1)_{k_1}) + \lambda_f^{p_N-1} h(v_{p_N}). \end{aligned}$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

If $\lambda_f \neq 1$, then we are done. If $\lambda_f = 1$, one can find $a \in \mathcal{A}$ and $k \in \mathbb{N}$ so that $\theta^k(a)_1 = a$. Then, for $0 \leq i \leq k-1$,

$$\int_{\mathcal{A}} f(c)\mathfrak{n}(dc) + h(\theta^i(a)_1) - h(\theta^{i+1}(a)_1) = 0.$$

Adding these sums, we have $\int f(x)\mathfrak{n}(dx) = 0$. Thus, $S_f(u_{\leq N})$ is bounded.

8. *Proof of Theorem 3.4*

Let $u(a, k) = \theta^k(a)$ for $a \in \mathcal{A}$ and $k, N \in \mathbb{N}$. For $n \leq |u(a, k)|$, define

$$Z_n^N(a, k) = \begin{cases} S_f(u(a, k)_{\leq n}) - \log_\lambda(N) \displaystyle\int_{\mathcal{A}} f(b)\mathfrak{n}(db) & \text{if } \lambda_f = 1, \\ S_f(u(a, k)_{\leq n}) & \text{otherwise.} \end{cases}$$

LEMMA 8.1. *Let f be a left eigenfunction of M with eigenvalue λ_f having $|\lambda_f| = 1$ which is not a coboundary. For any Borel measurable set $A \subset \mathbb{C}$ (or $\mathbb{R}$ in the case λ_f and f are real), let $\Phi(A) = \mathbb{P}(Z_f \in A)$. For any Borel measurable set $A \subset \mathbb{C}$ with $\Phi(\partial A) = 0$,*

$$\max_{\substack{a \in \mathcal{A} \\ k \in \mathbb{N}: |u(a,k)| \geq N}} \left| \frac{1}{N} \left| \left\{ 1 \leq k \leq N : \frac{Z_n^N(a, k)}{\sqrt{\log_\lambda N}} \in A \right\} \right| - \Phi(A) \right| \to 0 \quad \text{as } N \to \infty.$$

Proof. As $\mathcal{A}$ is finite, this follows directly from Proposition 6.1. □

Let $M = |u(a, k)|$. For $n \leq M$, define

$$Y_n^N(a, k) = \begin{cases} -S_f(u(a, k)_{[M-n+1,M]}) - \log_\lambda(N) \displaystyle\int_{\mathcal{A}} f(b)\mathfrak{n}(db) & \text{if } \lambda_f = 1, \\ -S_f(u(a, k)_{[M-n+1,M]}) & \text{otherwise.} \end{cases}$$

LEMMA 8.2. *For any Borel-measurable set $A \subset \mathbb{C}$ with $\Phi(\partial A) = 0$, as $N \to \infty$,*

$$\max_{\substack{a \in \mathcal{A} \\ k \in \mathbb{N}: |u(a,k)| \geq N}} \left| \frac{1}{N} \left| \left\{ 1 \leq n \leq N : \frac{Z_n^N(a, k)}{\sqrt{\log_\lambda N}} \in A \right\} \right| - \frac{1}{N} \left| \left\{ 1 \leq n \leq N : \frac{Y_n^N(a, k)}{\sqrt{\log_\lambda N}} \in A \right\} \right| \right| \to 0. \tag{19}$$

Proof. Let $M = |\theta^k(a)|$. Note that

$$\begin{aligned} -S_f(u(a, k)_{[M-n+1,M]}) &= -S_f(u(a, k)) + S_f(u(a, k)_{[1,M-n]}) \\ &= -\lambda_f^k f(a) + S_f(u(a, k)_{[1,M-n]}). \end{aligned}$$

Then

$$\left\{ 1 \leq k \leq M : \frac{Y_n^M(a, k)}{\sqrt{\log_\lambda M}} \in A \right\} = \left\{ 1 \leq k \leq M : \frac{Z_n^M(a, k)}{\sqrt{\log_\lambda M}} \in A + \frac{\lambda_f^k f(a)}{\sqrt{\log_\lambda M}} \right\}.$$

Thus, (19) holds along the subsequence $N_k = |\theta^k(a)| \uparrow \infty$. Given a substitution θ, we can define a reverse substitution $\tilde{\theta}$ by $\tilde{\theta}(a) = a_n a_{n-1} \cdots a_1$ for $\theta(a) = a_1 a_2 \cdots a_n$. Since (19) holds along the subsequence N_k, if there exists a central limit theorem with a drift for θ with eigenfunction f and there exists a central limit theorem for $\tilde{\theta}$ and $-f$, then one has the same drift for $\tilde{\theta}$ and $-f$. Now we can see that $-S_f(u(a, k)_{[M-n+1,M]})$ is a Birkhoff

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

sum of $-f$ on the substitution system associated to $\tilde{\theta}$. By Lemma 8.1, for some Φ_1 and Φ_2,

$$\max_{\substack{a\in\mathcal{A}\\ k\in\mathbb{N}:|u(a,k)|\geq N}} \left|\frac{1}{N}\left|\left\{1\leq k\leq N : \frac{Z_n^N(a,k)}{\sqrt{\log_\lambda N}}\in A\right\}\right| - \Phi_1(A)\right| \to 0,$$

$$\max_{\substack{a\in\mathcal{A}\\ k\in\mathbb{N}:|u(a,k)|\geq N}} \left|\frac{1}{N}\left|\left\{1\leq k\leq N : \frac{Y_n^N(a,k)}{\sqrt{\log_\lambda N}}\in A\right\}\right| - \Phi_2(A)\right| \to 0.$$

Since (19) holds along the subsequence $N_k = |\theta^k(a)| \uparrow \infty$, $\Phi_1 = \Phi_2$, so the proof is completed. □

Proof of Theorem 3.4. We will prove it for the case $\lambda_f = 1$. The proof for $\lambda_f \neq 1$ is analogous. Let us consider the prefix–suffix decomposition of v:

$$v = \cdots \theta^2(p_2)\theta(p_1)p_0 \cdots c_0 s_0 \theta(s_1)\theta^2(s_2)\cdots . \tag{20}$$

Case I. If only finitely many s_i are non-empty, then for some $a\in\mathcal{A}$ and l, $k\in\mathbb{N}$, $v_{[0,\infty)} = c_0 s_0 \theta(s_1)\cdots\theta^l(s_l)u$ and $u = \lim_{n\to\infty}\theta^{kn}(a)$. For sufficiently large n, $S_f(v_{\leq n}) = S_f(u_{\leq n}) + O(1)$. Thus, the result follows from Theorem 3.3.

Case II. Otherwise, define $N_l = |c_0 s_0 \theta(s_1)\cdots\theta^l(s_l)|$. Note that $N_l \uparrow \infty$. For any positive integer N, we choose N_ℓ such that $N_l \leq N < N_{\ell+1}$. Then set $a_{v,N} = S_f(v_{[1,N_\ell]})$. Note that

$$S_f(v_{[1,n]}) = \begin{cases} S_f(v_{[1,N_\ell]}) - S_f(v_{(n,N_\ell]}) & \text{if } n\leq N_\ell,\\ S_f(v_{[1,N_\ell]}) + S_f(v_{(N_\ell,n]}) & \text{if } n > N_l.\end{cases}$$

Then

$$\begin{aligned}
&\frac{1}{N}\left|\left\{1\leq n\leq N : \frac{S_f(v_{[1,n]}) - a_{v,N} - \log_\lambda(N)\int_{\mathcal{A}} f(b)\mathfrak{n}(db)}{\sqrt{\log_\lambda N}}\in A\right\}\right|\\
&\quad = \frac{N_\ell}{N}\frac{1}{N_\ell}\left|\left\{1\leq n\leq N_\ell : \frac{-S_f(v_{(n,N_\ell]}) - \log_\lambda(N)\int_{\mathcal{A}} f(b)\mathfrak{n}(db)}{\sqrt{\log_\lambda N}}\in A\right\}\right|\\
&\qquad + \frac{N-N_\ell}{N}\frac{1}{N-N_\ell}\left|\left\{N_\ell < n\leq N : \frac{S_f(v_{(N_\ell,n]}) - \log_\lambda(N)\int_{\mathcal{A}} f(b)\mathfrak{n}(db)}{\sqrt{\log_\lambda N}}\in A\right\}\right|.
\end{aligned} \tag{21}$$

Fix $\epsilon > 0$ small so that if M satisfies that $\epsilon N \leq M \leq N$, then $\log_\lambda N/\log_\lambda M = 1 + O(\epsilon)$ and $\log_\lambda (M/N) = 1 + O(\epsilon)$.

For any N, we have the following three cases.

Case (i): $\epsilon N \leq N_\ell \leq N$ and $\epsilon N \leq N - N_\ell \leq N$. If N is large enough so that N_ℓ is large,

$$\begin{aligned}
&\frac{-S_f(v_{(n,N_\ell]}) - \log_\lambda(N)\int_{\mathcal{A}} f(b)\mathfrak{n}(db)}{\sqrt{\log_\lambda N}}\\
&\quad = \frac{-S_f(v_{(n,N_\ell]}) - \log_\lambda(N_\ell)\int_{\mathcal{A}} f(b)\mathfrak{n}(db)}{\sqrt{\log_\lambda N_\ell}}(1+O(\epsilon)) + o_N(1).
\end{aligned}$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

So, from Lemma 8.2,

$$\frac{1}{N_l}\left|\left\{1 \le n \le N_\ell : \frac{-S_f(v_{(n,N_\ell]}) - \log_\lambda(N)\int_{\mathcal{A}} f(b)\mathfrak{n}(db)}{\sqrt{\log_\lambda N}} \in A\right\}\right|$$

is close to $\Phi(A)$ for small ϵ and large N. Similarly,

$$\frac{1}{N - N_\ell}\left|\left\{N_\ell < n \le N : \frac{S_f(v_{(N_\ell,n]}) - \log_\lambda(N)\int_{\mathcal{A}} f(b)\mathfrak{n}(db)}{\sqrt{\log_\lambda N}} \in A\right\}\right|$$

is also close to $\Phi(A)$. Thus, (21) is close to $\Phi(A)$.

Case (ii) $N_\ell < \epsilon N$. Then

$$\frac{N_\ell}{N}\frac{1}{N_\ell}\left|\left\{1 \le n \le N_\ell : \frac{-S_f(v_{(n,N_\ell]}) - \log_\lambda(N)\int_{\mathcal{A}} f(b)\mathfrak{n}(db)}{\sqrt{\log_\lambda N}} \in A\right\}\right| < \epsilon.$$

Also, $N - N_\ell > (1-\epsilon)N \ge \epsilon N$. Using the same argument as in Case (i), one can see that (21) is close to $\Phi(A)$.

Case (iii) $N - N_\ell < \epsilon N$. It is similar to Case (ii).

Now it remains to show equation (16). For almost every v, the prefix–suffix decomposition (20) satisfies Case II above; thus, $a_{v,N}$ is given by $a_{v,N} = S_f(v_{[1,N_\ell]})$ as above. Moreover, we claim that there exists C_0 such that for almost every v,

$$\limsup_{\ell\to\infty} \frac{\log_\lambda N_{\ell+1}/N_\ell}{\log_\lambda \ell} < C_0. \tag{22}$$

Note that for some $t > 0$, there exists $c > 0$, denoting $m = t/c \in \mathbb{N}$, such that if $\log_\lambda (N_{\ell+1}/N_\ell) > t$, then $s_{\ell-m}, s_{\ell-m+1}, \ldots, s_l$ are empty words. Since the sequences $(p_i, c_i, s_i)_{i\in\mathbb{N}_0}$ are primitive homogeneous Markov chains, for some $\alpha > 0$,

$$\mathbb{P}\left[\log_\lambda \frac{N_{\ell+1}}{N_\ell} > t\right] < e^{-\alpha t}.$$

Choosing $t = (2/\alpha)\log \ell$ and applying the Borel–Cantelli lemma, the claim follows.

Now we see that

$$S_f(v_{[1,N_\ell]}) = S_f(c_0) + S_f(s_0) + \lambda_f S_f(s_1) + \cdots + \lambda_f^\ell S_f(s_\ell).$$

As $S_f(v_{[1,N_\ell]})$ is an additive functional of the finite state Markov chain $(p_i, c_i, s_i)_{i\in\mathbb{N}_0}$, the law of the iterated logarithm holds (see Theorem A.1). If $\lambda_f = 1$, this implies that there is some $C > 0$ such that

$$\limsup_{\ell\to\infty} \frac{|S_f(v_{[1,N_\ell)} - \ell\int_{\mathcal{A}} f(a)\mathfrak{n}(da)|}{\sqrt{\ell \log\log \ell}} = C.$$

From (22), the desired conclusion holds. For general λ_f with $|\lambda_f| = 1$, a similar argument completes the proof. □

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

9. *Proofs for* $|\lambda_f| \neq 1$

We begin with the proof of Theorem 3.1.

Proof of Theorem 3.1. Given N, one can find p_N such that $|\theta^{p_N-1}(a)| < N \leq |\theta^{p_N}(a)|$. We again define a function $\check{f}$ on $\mathcal{X}$ by $\check{f}(v_i, k_i) = S_f(\theta(v_i)_{<k_i})$. Then we again have that for any $1 \leq K \leq N$ that

$$\lambda_f S_f(u_{<K}) = \sum_{i=1}^{p_N} \lambda_f^i \check{f}((v_i, k_i)),$$

where $\Psi_{a,p_N}(K) = (v_1, k_1)(v_2, k_2) \dots (v_{p_N}, k_{p_N})$.

For any natural number p, let $Z_p = \sum_{i=1}^{p} \lambda_f^i \check{f}((v_i, k_i))$, which is a function on $\mathcal{X}^{*,p}$. We naturally embed $\mathcal{X}^{*,p}$ as the initial coordinates of $\mathcal{X}^{*,\infty}$, and thus also consider Z_p as a function on $\mathcal{X}^{*,\infty}$. Also, let $Z_\infty = W_f = \sum_{i=1}^{\infty} \lambda_f^i \check{f}((v_i, k_i))$, a function on $\mathcal{X}^{*,\infty}$. Observe that

$$\sup_{p \geq 1} \sup_{\mathbf{x} \in \mathcal{X}^{*,\infty}} |Z_p(\mathbf{x})| < \infty.$$

Moreover, we have that uniformly in p, for all $n > p$,

$$\sup_{\mathbf{x} \in \mathcal{X}^{*,\infty}} |Z_p(\mathbf{x}) - Z_n(\mathbf{x})| \ll |\lambda_f|^p. \tag{23}$$

Hence, the same estimate holds for the difference of Z_p and Z_∞.

We will show that for any bounded uniformly continuous function ϕ,

$$\int_{\mathcal{X}^{*,p}} \phi(Z_{p_N}(\mathbf{x})) \nu_N(d\mathbf{x}) \to \int_{\mathbb{C}} \phi(Z_\infty(\mathbf{x}))\, \mathrm{MPM}_\infty(d\mathbf{x}),$$

which will complete the proof. Exactly as in the proof of Proposition 6.1, using (23) and Proposition 5.3, we have that

$$\int \phi(Z_{p_N})\, d\nu_N = \int \phi(Z_{p_N})\, d\, \mathrm{MPM}_{p_N} + o(1) = \int \phi(Z_{p_N})\, d\, \mathrm{MPM}_\infty + o(1).$$

But, by uniform continuity of ϕ and (23), we have that

$$\int \phi(Z_{p_N})\, d\, \mathrm{MPM}_\infty = \int \phi(Z_\infty)\, d\, \mathrm{MPM}_\infty + o(1),$$

so the proof is complete. □

Proof of Theorem 3.2. Let $p = p(\ell) \in \mathbb{N}$ be such that $|\theta^{p-1}(a)| < N_\ell \leq |\theta^p(a)|$. For every $\mathbf{x} = x_1 x_2 \cdots x_p \in \mathcal{X}^{*,p}$, we have by Proposition 2.1 that

$$\nu_{N_\ell}(\{\mathbf{x}\}) = \sum_{q=1}^{p} \sum_{1 \leq j < k_q} \frac{\mathrm{SSIM}_{*,(v_q,j)}^{p-q+1}}{N_\ell}\, \mathrm{UPM}_{(v_q,j),p-q+1}(\{x_1 x_2 \cdots x_{p-q}\}) \tag{24}$$

(caution: we have used the reversed path here), where $\Psi_a^r(N_\ell) = (v_1(\ell), k_1(\ell))$ $(v_2(\ell), k_2(\ell)) \cdots$.

We will begin by showing that as a measure on $\mathcal{X}^\infty$, ν_{N_ℓ} converges. Let $\tilde{\nu}_{N_\ell}$ be a measure on $\mathcal{X}^\infty$ given by the property that for any $1 \leq K < N_\ell$, $\tilde{\nu}_{N_\ell}(\Psi_a^r(K)) = \nu_{N_\ell}(\Psi_{a,p}(K))$. It follows that for any cylinder $[\mathbf{x}] = [x_1 x_2 \dots x_p]$,

$$\tilde{\nu}_{N_\ell}([\mathbf{x}]) = \nu_{N_\ell}(x_p x_{p-1} \dots x_1). \tag{25}$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

We will show that $\tilde{\nu}_{N_\ell} \Rightarrow \mathrm{RMPM}_{\mathfrak{a},\infty}$, which is equivalent to showing that for any fixed cylinder $[\mathbf{x}] = [x_1 x_2 \dots x_k]$,

$$\tilde{\nu}_{N_\ell}([\mathbf{x}]) \to \mathrm{RMPM}_{\mathfrak{a},\infty}([\mathbf{x}]). \tag{26}$$

Recall by (5) that

$$\lim_{p\to\infty} \lambda^{-p}\, \mathrm{SSIM}^p_{x,y} = \hat{\sigma}(y)\hat{\rho}(x).$$

Moreover, we have that

$$\lambda^{-p}\, \mathrm{SSIM}^p_{x,y} = \hat{\sigma}(y)\hat{\rho}(x) + O(e^{-cp})$$

uniformly in x and y by Perron–Frobenius theory and (5).

By the convergence of $\Psi^r_a(N_\ell)$, we therefore have that for every $r \in \mathbb{N}$, there is an $\ell_0(r)$ sufficiently large so that for all $\ell > \ell_0$,

$$\sum_{q=1}^{r} \sum_{j=1}^{k_q(\ell)-1} \frac{\mathrm{SSIM}^{p-q+1}_{*,(v_q(\ell),j)}}{\lambda^p} = \sum_{q=1}^{r} \sum_{1\le j<\kappa_q} \hat{\sigma}((\rho_q,\, j))\lambda^{1-q} + O(e^{-cp} + \lambda^{-r}),$$

where we have used that $1 = \sum_{x\in\mathcal{X}} \hat{\rho}(x)$. Furthermore, we have that

$$\sum_{q=r+1}^{p(\ell)} \sum_{j=1}^{k_q(\ell)-1} \frac{\mathrm{SSIM}^{p-q+1}_{*,(v_q(\ell),j)}}{\lambda^p} = O(\lambda^{-r})$$

uniformly in N_ℓ. By (24), we have that

$$N_\ell = \sum_{q=1}^{p} \sum_{1\le j<k_q} \mathrm{SSIM}^{p-q+1}_{*,(v_q,j)}$$

and hence

$$\frac{N_\ell}{\lambda^{p(\ell)}} = \sum_{q=1}^{r} \sum_{1\le j<\kappa_q} \hat{\sigma}((\rho_q,\, j))\lambda^{1-q} + O(e^{-cp(\ell)} + \lambda^{-r})$$

uniformly in r for all $\ell > \ell_0(r)$. Define

$$R = \lim_{\ell\to\infty} \frac{N_\ell}{\lambda^{p(\ell)}} = \sum_{q=1}^{\infty} \sum_{1\le j<\kappa_q} \hat{\sigma}((\rho_q,\, j))\lambda^{1-q}. \tag{27}$$

Then $\mathfrak{a}((v,\, k))$ is given by

$$\mathfrak{a}((v,\, k)) = \frac{1}{R}\hat{\sigma}((v,\, k)) \sum_{q=1}^{\infty} \mathbf{1}\{v = \rho_q \text{ and } k < \kappa_q\}\lambda^{1-q}$$

and, from (27) and (24), (26) follows.

Recall that $U_f(\mathbf{x}) = \sum_{i=1}^{\infty} \lambda_f^{-i} S_f(\theta(v_i)_{<k_i})$. This is a bounded $\mathbb{C}$-valued continuous function from $\mathcal{X}_r^{*,\infty}$ under the product topology. Hence, for any bounded uniformly continuous function $\phi : \mathbb{C} \to \mathbb{R}$, by the definition of $\tilde{\nu}_{N_\ell} \Rightarrow \mathrm{RMPM}_{\mathfrak{a},\infty}$, we have that

$$\lim_{\ell\to\infty} \int \phi(U_f(\mathbf{x}))\tilde{\nu}_{N_\ell}(d\mathbf{x}) = \int \phi(U_f(\mathbf{x}))\mathrm{RMPM}_{\mathfrak{a},\infty}(d\mathbf{x}).$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

The remainder of the proof now proceeds in the same manner as the proof of Theorem 3.1, with some minor changes. We define a function $\check{f}$ on $\mathcal{X}$ by $\check{f}(v_i, k_i) = S_f(\theta(v_i)_{<k_i})$. Then we have for any $1 \le K \le N_\ell$,

$$\lambda_f^{-p(\ell)} S_f(u_{<K}) = \sum_{i=1}^{p(\ell)} \lambda_f^{-i} \check{f}((v_i, k_i)),$$

where $\Psi_a^r(K) = (v_1, k_1)(v_2, k_2)\cdots$.

For any natural number p, let $U_p = \sum_{i=1}^{p} \lambda_f^{-i} \check{f}((v_i, k_i))$, which is a function on $\mathcal{X}_r^{*,\infty}$. Then we have that

$$\sup_{\mathbf{x}\in\mathcal{X}^{*,\infty}} |U_p(\mathbf{x}) - U_f(\mathbf{x})| \ll \lambda_f^{-p}. \tag{28}$$

Then we have that for any bounded uniformly continuous ϕ,

$$\begin{aligned}\mathbb{E}(\phi(\lambda_f^{-p(\ell)} S_f(u_{<K_{N_\ell}}))) &= \int \phi(U_{p(\ell)}(\mathbf{x}))\tilde{\nu}_{N_\ell}(d\mathbf{x}) \\ &= \int \phi(U_f(\mathbf{x}))\tilde{\nu}_{N_\ell}(d\mathbf{x}) + o(1).\end{aligned}$$

Hence, we have shown that $\lambda_f^{-p(\ell)} S_f(u_{<K_{N_\ell}}) \Rightarrow U_f(\mathbf{x})$ with $\mathbf{x}$ distributed according to $\mathrm{RMPM}_{\mathfrak{a},\infty}$. As

$$\frac{N_\ell^{\log_\lambda |\lambda_f|}}{|\lambda_f|^{p(\ell)}} = \frac{\lambda^{p(\ell)\log_\lambda |\lambda_f|}}{|\lambda_f|^{p(\ell)}} \frac{N_\ell^{\log_\lambda |\lambda_f|}}{\lambda^{p(\ell)\log_\lambda |\lambda_f|}} \to R^{\log_\lambda |\lambda_f|},$$

it follows that

$$\frac{S_f(u_{<K_{N_\ell}})}{N_\ell^{\log_\lambda |\lambda_f|} e^{ip(\ell)\arg\lambda_f}} = \frac{|\lambda_f|^{p(\ell)}}{N_\ell^{\log_\lambda |\lambda_f|}} \frac{S_f(u_{<K_{N_\ell}})}{\lambda_f^{p(\ell)}} \Rightarrow \frac{1}{R^{\log_\lambda |\lambda_f|}} U_f(\mathbf{x})$$

and the proof is complete. □

Acknowledgements. The authors would like to thank Omri Sarig for suggesting this problem, as well as useful discussion and comments. EP is partially supported by NSF Postdoctoral Fellowship DMS-1304057.

A. *Appendix. Central limit theorem*

In this section, we give a proof of the exact version of the Markov chain central limit theorem that we will need. Let $X_1, X_2, \ldots$ be a primitive Markov chain on a finite state space $\mathcal{X}$ with invariant measure π. Let $\mathbb{P}$ denote the probability measure of this Markov chain on $\mathcal{X}^{\mathbb{N}}$. Let $p(x, y)$ be the transition matrix of the Markov chain, i.e. $p(x, y) = \mathbb{P}(X_2 = y \mid X_1 = x)$. Hence, $p(x, y)$ is a right stochastic matrix and π, its invariant measure, is the positive left Perron–Frobenius eigenvector of p whose $\|\cdot\|_1$ norm is 1.

We also define the reversed transition matrix p^* given by $p^*(x, y) = p(y, x)\pi(y)/\pi(x)$, which is also a right stochastic matrix. It is easily checked that if $Y_j = X_{n-j+1}$ for $j = 1, 2, \ldots, n$, then $(Y_j)_{j=1}^n$ are n steps of a stationary Markov

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

chain with transition matrix p^*. Further, p^* is the Hilbert space adjoint of p with respect to the inner product on $\mathbb{C}^{\mathcal{X}}$ given by $(f, g)_\pi = \sum_{x\in\mathcal{X}} f(x)\overline{g(x)}\pi(x)$.

We also define the operators P and P^* on $\mathbb{C}^{\mathcal{X}}$ by $(Ph)(x) = \sum_{x\in\mathcal{X}} p(x, y)h(y)$ and $(P^*h)(x) = \sum_{x\in\mathcal{X}} p^*(x, y)h(y)$. Let $\mathscr{F}_N = \sigma(X_1, X_2, \ldots, X_N)$, the σ-algebra generated by the first N states of the Markov chain. Then we have that for any $N \in \mathbb{N}$, $(Ph)(X_N) = \mathbb{E}(h(X_{N+1})|\mathscr{F}_N)$.

THEOREM A.1. *Let $f : \mathcal{X} \to \mathbb{C}$ and $\lambda \in \mathbb{C}$ with $|\lambda| = 1$. Then either:*

(1) *if there is a function $h : \mathcal{X} \to \mathbb{C}$ satisfying $P^*Ph = h$ and $f = \int f\, d\pi + h - \lambda Ph$, then we have*

$$\sup_{N\in\mathbb{N}} \left| \sum_{i=1}^{N} \lambda^i (f(X_i) - \int f d\pi) \right| \le 2\|h\|_\infty$$

almost surely; or

(2) *if there is no such function, we have that*

$$\frac{1}{\sqrt{N}} \left(\sum_{i=1}^{N} \lambda^i (f(X_i) - \int f d\pi) \right) \Rightarrow Z,$$

where Z has a complex normal distribution with $\mathbb{E}|Z|^2 > 0$. If $\lambda \in \mathbb{R}$ and $f : \mathcal{X} \to \mathbb{R}$, then Z is a real normal distribution. If $\lambda \notin \mathbb{R}$, then real and imaginary parts of Z are independent and have identical variance. Further, we always have that

$$\mathbb{E}|Z|^2 = \mathbb{E}|g(X_1)|^2 + \sum_{k=2}^{\infty} 2\mathbb{E}\Re[\lambda^{k-1} g(X_1)\overline{g(X_k)}],$$

where $g(x) = f(x) - \int f\, d\pi$. Finally, we have that there is a constant $C > 0$ so that with probability 1,

$$\limsup_{N\to\infty} \frac{|\sum_{i=1}^{N} \lambda^i (f(X_i) - \int f d\pi)|}{\sqrt{N \log\log N}} = C.$$

Proof. We will use the martingale central limit theorem to prove the convergence (see [**HH80**, Theorem 3.2]). To do so, we will show that $Y_N = \sum_{i=1}^{N} \lambda^i g(X_i)$ is nearly a martingale. The first step towards doing so is to show that there is a function $h(x)$ so that $g(x) = h(x) - \lambda(Ph)(x)$. Because p is a primitive stochastic matrix, it has Perron–Frobenius eigenvalue 1 and its other eigenvalues have modulus strictly less than 1. Hence, if $\lambda \neq 1$, then $I - \lambda P$ is invertible, so one can find a unique $h(x)$. If $\lambda = 1$, then the space W of $r(x)$ with $\int r(x)\pi(dx) = 0$ is $\Im(I - P)$ since:

(i) W has dimension $|\mathcal{X}| - 1$;

(ii) $\Im(I - P)$ has dimension $|\mathcal{X}| - 1$ by considering eigenvalues of p;

(iii) $\Im(I - P) \subset W$:

$$\int h(x)\pi(dx) - \int (Ph)(x)\pi(dx) = \mathbb{E}h(X_1) - \mathbb{E}(\mathbb{E}(h(X_2)|\mathscr{F}_1)) = 0.$$

Note that the kernel of $I - P$ is just the constant functions and hence we may choose $h(x)$ to have $\int h\, d\pi = 0$. Having made this choice, h is uniquely determined. Notice that in the case that $\lambda \neq 1$, we have that h satisfies this condition as well.

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

Let $Z_N = \sum_{i=1}^{N-1} \lambda^{i+1}[h(X_{i+1}) - (Ph)(X_i)]$. Note that

$$Y_N = Z_N + \lambda h(X_1) - \lambda^{N+1}(Ph)(X_N).$$

Also, we can see that Z_N is a martingale:

$$\mathbb{E}(Z_{N+1}|\mathscr{F}_N) = Z_N + \lambda^{N+1}\mathbb{E}(h(X_{N+1}) - (Ph)(X_N)|\mathscr{F}_N) = Z_N.$$

Furthermore,

$$\mathbb{E}|Z_N|^2 = \sum_{i=1}^{N-1} \mathbb{E}|h(X_{i+1}) - (Ph)(X_i)|^2 = (N-1)\mathbb{E}|h(X_2) - (Ph)(X_1)|^2$$

by the orthogonality of martingale increments. We now show that this variance is 0 if and only if the function h satisfies $P^*Ph = h$. If the variance is 0, then we have that Z_N almost surely vanishes and so the first conclusion of the theorem follows from the definition of Y_N. Conversely, if $\mathbb{E}|h(X_2) - (Ph)(X_1)|^2 \neq 0$, the second conclusion of the theorem follows immediately from the martingale central limit theorem and the law of the iterated logarithm for martingales [**Sto70**].

We now expand the squares to get

$$\begin{aligned}\mathbb{E}|h(X_2) - (Ph)(X_1)|^2 &= \mathbb{E}(h(X_2) - (Ph)(X_1))(\overline{h}(X_2) - \overline{(Ph)}(X_1)) \\ &= \mathbb{E}|h(X_1)|^2 - \mathbb{E}|(Ph)(X_1)|^2 \\ &= (h, h)_\pi - (Ph, Ph)_\pi \\ &= ((I - P^*P)h, h)_\pi. \end{aligned} \tag{A.1}$$

In the second equality, we have used that

$$\begin{aligned}\mathbb{E}(\overline{h}(X_2)(Ph)(X_1)) &= \mathbb{E}(\mathbb{E}(\overline{h}(X_2)|\mathscr{F}_1)(Ph)(X_1)) \\ &= \mathbb{E}((\overline{Ph})(X_1)(Ph)(X_1)) \\ &= \mathbb{E}|(Ph)(X_1)|^2.\end{aligned}$$

The operator $I - P^*P$ is Hermitian positive semidefinite and hence the variance is 0 if and only if $(I - P^*P)h = 0$. In the case that $\lambda \neq 1$, the solution to $(I - \lambda P)h = g$ is unique and hence we are done as this is the only possible h that could satisfy the criterion in (1). If $\lambda = 1$, the collection of h' for which $(I - \lambda P)h' = g$ just differ from h by constant functions. Hence, $P^*Ph' = h'$ if and only if $P^*Ph = h$. It remains to show the formula for the limiting variance of Z. We have that

$$\mathbb{E}|Z|^2 = \lim_{N\to\infty} \frac{\mathbb{E}|Z_N|^2}{N} = \mathbb{E}|h(X_2) - (Ph)(X_1)|^2.$$

By (A.1), we therefore have that

$$\begin{aligned}\mathbb{E}|Z|^2 &= (h, h)_\pi - (Ph, Ph)_\pi \\ &= (h - \lambda Ph, h)_\pi + (\lambda Ph, h - \lambda Ph)_\pi \\ &= (g, h)_\pi + (\lambda Ph, g)_\pi \\ &= (((I - \overline{\lambda}P^*)^{-1} + \lambda P(I - \lambda P)^{-1})g, g)_\pi.\end{aligned}$$

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

These inverses always exist on the space of functions W. Expanding the inverses as Neumann series, we arrive at

$$\mathbb{E}|Z|^2 = (g, g)_\pi + \sum_{k=1}^{\infty}[(\overline{\lambda}^k P^{*k} g, g)_\pi + (\lambda^k P^k g, g)_\pi].$$

The desired formula for the variance now follows using the identities $P^k g(x) = \mathbb{E}[g(X_{k+1})|X_1 = x]$ and $P^{*k} g(x) = \mathbb{E}[g(X_1)|X_{k+1} = x]$.

To show that the real and imaginary parts of Z are independent and have the same variance in the case $\lambda \notin \mathbb{R}$, observe that, again by the orthogonality of martingale increments,

$$\mathbb{E}Z_N^2 = \sum_{i=1}^{N-1} \mathbb{E}\lambda^{2(i+1)}(h(X_{i+1}) - (Ph)(X_i))^2 = \mathbb{E}(h(X_2) - (Ph)(X_1))^2 \sum_{i=1}^{N-1} \lambda^{2(i+1)}.$$

In particular, if $\lambda \notin \mathbb{R}$, we have that

$$\mathbb{E}Z^2 = \lim_{N\to\infty} \frac{\mathbb{E}Z_N^2}{N} = 0.$$

This means that

$$0 = \mathbb{E}Z^2 = \mathbb{E}(\Re Z)^2 - \mathbb{E}(\Im Z)^2 + 2i\mathbb{E}(\Re Z \Im z).$$

As $(\Re Z, \Im Z)$ are jointly Gaussian and their covariance is 0, they are independent. Further, the variances of the real and imaginary parts match. □

References

[**Ada04**] B. Adamczewski. Symbolic discrepancy and self-similar dynamics. *Ann. Inst. Fourier (Grenoble)* **54**(7) (2004), 2201–2234.

[**ADDS15**] A. Avila, D. Dolgopyat, E. Duryev and O. Sarig. The visits to zero of a random walk driven by an irrational rotation. *Israel J. Math.* **207**(2) (2015), 653–717.

[**BBH14**] X. Bressaud, A. I. Bufetov and P. Hubert. Deviation of ergodic averages for substitution dynamical systems with eigenvalues of modulus 1. *Proc. Lond. Math. Soc. (3)* **109**(2) (2014), 483–522.

[**Bec10**] J. Beck. Randomness of the square root of 2 and the giant leap. Part 1. *Period. Math. Hungar.* **60**(2) (2010), 137–242.

[**Bec11**] J. Beck. Randomness of the square root of 2 and the giant leap. Part 2. *Period. Math. Hungar.* **62**(2) (2011), 127–246.

[**BH91**] M. Boyle and D. Handelman. The spectra of nonnegative matrices via symbolic dynamics. *Ann. of Math. (2)* **133**(2) (1991), 249–316.

[**Buf14**] A. I. Bufetov. Finitely-additive measures on the asymptotic foliations of a Markov compactum. *Mosc. Math. J.* **14**(2) (2014), 205–224, 426.

[**CS01**] V. Canterini and A. Siegel. Automate des préfixes–suffixes associé à une substitution primitive. *J. Théor. Nombres Bordeaux* **13**(2) (2001), 353–369.

[**DHS99**] F. Durand, B. Host and C. Skau. Substitutional dynamical systems, Bratteli diagrams and dimension groups. *Ergod. Th. & Dynam. Sys.* **19**(4) (1999), 953–993.

[**Dud66**] R. M. Dudley. Convergence of Baire measures. *Studia Math.* **27** (1966), 251–268.

[**Fog02**] N. Pytheas Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics (Lecture Notes in Mathematics, 1794)*. Eds. V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel. Springer, Berlin, 2002.

[**GH55**] W. H. Gottschalk and G. A. Hedlund. Topological dynamics. *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 1955, p. 36.

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press

[**HH80**] P. Hall and C. C. Heyde. *Martingale Limit Theory and its Application (Probability and Mathematical Statistics)*. Academic Press [Harcourt Brace Jovanovich], New York, London, 1980.

[**Hos86**] B. Host. Valeurs propres des systèmes dynamiques définis par des substitutions de longueur variable. *Ergod. Th. & Dynam. Sys.* **6**(4) (1986), 529–540.

[**Ito78**] S. Ito. A construction of transversal flows for maximal Markov automorphisms. *Tokyo J. Math.* **1**(2) (1978), 305–324.

[**KL97**] I. Kornfeld and M. Lin. Coboundaries of irreducible Markov operators on $C(K)$. *Israel J. Math.* **97** (1997), 189–202.

[**Kle14**] A. Klenke. *Probability Theory: A Comprehensive Course*, 2nd edn. Universitext. Springer, London, 2014.

[**KOR00**] K. H. Kim, N. S. Ormes and F. W. Roush. The spectra of nonnegative integer matrices via formal power series. *J. Amer. Math. Soc.* **13**(4) (2000), 773–806 (electronic).

[**KV86**] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104**(1) (1986), 1–19.

[**Liv88**] A. N. Livshits. Sufficient conditions for weak mixing of substitutions and of stationary adic transformations. *Mat. Zametki* **44**(6) (1988), 785–793, 862.

[**LPW09**] D. A. Levin, Y. Peres and E. L. Wilmer. *Markov Chains and Mixing Times*. Eds. J. G. Propp and D. B. Wilson. American Mathematical Society, Providence, RI, 2009.

[**Mos92**] B. Mossé. Puissances de mots et reconnaissabilité des points fixes d'une substitution. *Theoret. Comput. Sci.* **99**(2) (1992), 327–334.

[**Mos96**] B. Mossé. Reconnaissabilité des substitutions et complexité des suites automatiques. *Bull. Soc. Math. France* **124**(2) (1996), 329–346.

[**MW00**] M. Maxwell and M. Woodroofe. Central limit theorems for additive functionals of Markov chains. *Ann. Probab.* **28**(2) (2000), 713–724.

[**Par64**] W. Parry. Intrinsic Markov chains. *Trans. Amer. Math. Soc.* **112** (1964), 55–66.

[**Que10**] M. Queffélec. *Substitution Dynamical Systems—Spectral Analysis (Lecture Notes in Mathematics, 1294)*, 2nd edn. Springer, Berlin, 2010.

[**SS14**] O. Sarig and M. Schmoll. Adic flows, transversal flows, and horocycle flows. *Ergodic Theory and Dynamical Systems (De Gruyter Procedings in Mathematics)*. De Gruyter, Berlin, 2014, pp. 241–259.

[**Sto70**] W. F. Stout. A martingale analogue of Kolmogorov's law of the iterated logarithm. *Z. Wahrscheinlichkeitsth. Verw. Geb.* **15** (1970), 279–290.

[**Ver95**] A. M. Vershik. The adic realizations of the ergodic actions with the homeomorphisms of the Markov compact and the ordered Bratteli diagrams. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) (Teor. Predstav. Din. Sistemy Kombin. i Algoritm. Metody. I)* **223** (1995), 120–126, 338.

https://doi.org/10.1017/etds.2017.83 Published online by Cambridge University Press