Ergodic Theorems of Free Groups

by

George Wynne

MA469 4th Year Project

Submitted to The University of Warwick

Mathematics Institute

April, 2018

Contents

1 Introduction **1**

2 Markov Operators **1**
2.1 Introduction to Markov Operators 1
2.2 Preliminary convergence results 2
2.3 Relation to spherical averages 4
2.4 Argument of Lemma 2.9 ii). 6
2.5 Proof of Lemma 2.12 9
2.6 Proof of Lemma 2.13 11
2.7 Argument of Lemma 2.9 i). 16
2.8 Summary 17

3 Main paper **18**
3.1 Extending results to more general Markov chains 18
3.2 Set up 18
3.3 Main theorem 19
3.4 Motivation 20
3.5 Proof of Theorem 3.9 23
3.6 Proof of Theorem 3.6 26

4 Fundamental group of the surface of genus two **30**
4.1 Markov Coding 30
4.2 Applying Theorem 3.6 31

1 Introduction

This project will outline the results of [3] and recent paper [1]. In particular it will cover Markov operators and their use in showing convergence of ergodic averages of free group actions. Section 2 will go through the contents of [3] which builds the theory of Markov operators and establishes many techniques that are used in [1] which is covered in Section 3. Section 4 will provide additional explanation of an application discussed in [1].

I wish to emphasize that all results here are entirely taken from these sources as well as most of the proofs. My main contribution is filling in details of proofs found in the cited papers. In particular I have added details to the proofs of Lemma 2.4, Proposition 2.10, Proposition 3.16 and the explanation of the application in Section 4. I found errors in the conditional expectation result of the Markov operator in [1] and in Section 3.4 I have given a more thorough exposition of the equivalence relations, using the relation E_n^X given in a correction provided by one of the co-authors

I have rearranged the order of results to highlight the importance of showing triviality of tail σ-algebras and how convergence results follow from there. For a history of the development of ergodic averages of actions of free group and the use of Markov operators in these problems see Section 1.1 of [1].

I would like to thank Lewis Bowen, Olga Romaskevich, Alexander Bufetov and Richard Sharp for their input in making this project.

2 Markov Operators

2.1 Introduction to Markov Operators

In this chapter I shall give an overview of Markov operators, in particular their use in proofs of ergodic theorems of actions of free groups. Their use for this application can be attributed to R.I Grigorchuk [7], J.-P. Thouvenot (oral communication) and A. Bufetov [2].

Definition 2.1. *Let (X, μ) be a measure space. A linear operator $P : L^1(X, \mu) \to L^1(X, \mu)$ is called a measure-preserving Markov operator if it satisfies the following three conditions:*

1. *P maps constant functions to themselves*
2. *P maps non-negative functions to non-negative functions*
3. *P preserves the L_1, L_∞ norms*

This definition is motivated as a continuous generalization of multiplication of the transpose of a stochastic matrix with a vector, which occurs in the discrete setting. The transpose of a stochastic matrix will have columns summing to one representing the probabilities associated with a set of outcomes. So when multiplied by a positive vector, the resulting vector is positive. Markov operators generalize matrices to operators, vectors to functions and sums to integrals.

2.2 Preliminary convergence results

Convergence results relating to group actions of finitely generated free groups can be written in terms of convergence results of series of Markov operators. Before detailing how this correspondence works I will first give a preliminary convergence result of Markov operators and the set-up it requires such as tail σ-algebras and the induced Markov measure.

Let (X, μ) be a probability space and let $\mathbf{X} = \{(x_i)_{i=0}^{\infty} : x_i \in X\}$ be the corresponding sequence space. Given a measure preserving Markov operator $P : L^1(X, \mu) \to L^1(X, \mu)$ we define the induced Markov measure $\tilde{\mu}$ as follows.

We give $\mathbf{X}$ the product σ-algebra $\mathcal{B}(\mathbf{X})$. This is the σ-algebra generated by sets of the form $\prod_{i=0}^{\infty} A_i$ where $A_i \in \mathcal{B}(X)$, the σ-algebra of μ-measurable sets, for all i and all but finitely many $A_i = X$. Consider a function $F : \mathbf{X} \to \mathbb{C}$ of the form

$$F(x_0, x_1, x_2, \ldots) = f_0(x_0) f_1(x_1) f_2(x_2) \ldots$$

where $f_i \in L^{\infty}(X, \mu)$ for all i and $f_i = 1$ for all but finitely many i. Note that the integral makes sense since $P1 = 1$ and $f_i = 1$ for all but finitely many f_i. Then $\tilde{\mu}$ is the measure such that for any function F of this form we have

$$\int_{\mathbf{X}} F \, \mathrm{d}\tilde{\mu} = \int_X f_0 P(f_1 P(f_2 P(f_3 \cdots)) \cdots) \, \mathrm{d}\mu$$

The measure of sets of the type outlined above are then defined as

$$\tilde{\mu}\left(\prod_{i=0}^{\infty} A_i\right) = \int_{\mathbf{X}} F \, \mathrm{d}\tilde{\mu}$$

where $F = \chi_{A_0} \cdot \chi_{A_1} \cdot \chi_{A_2} \cdots$.

Now given the space $(\mathbf{X}, \tilde{\mu})$ I will define some σ-algebras relating to P to allow us to use the reverse martingale convergence theorem.

Definition 2.2. *Let $f_k : \mathbf{X} \to X$ be defined as $f_k(x) = x_k$. Then define the following σ-algebras as follows:*

1. *$\mathcal{F}_0$ the minimal complete σ-algebra on $(\mathbf{X}, \tilde{\mu})$ such that f_0 is measurable*

2. *$\mathcal{F}_{\geq n}$ the minimal complete σ-algebra on $(\mathbf{X}, \tilde{\mu})$ such that f_k is measurable for $k \geq n$*

3. *$\mathcal{F}_\infty = \bigcap_{n \geq 0} \mathcal{F}_{\geq n}$*

We call $\mathcal{F}_\infty$ the tail σ-algebra. Given a function $f \in L^1(X, \mu)$ we will define $f^0 \in L^1(\mathbf{X}, \tilde{\mu})$ by $f^0(x) = f(x_0)$. We are now ready to state the first convergence results of Markov operators.

Proposition 2.3. *Let P be a measure-preserving Markov operator on the probability space (X, μ). Then the tail σ-algebra $\mathcal{F}_\infty$ is trivial if and only if for any $\varphi \in L^1(X, \mu)$ the sequence $\left(Q^n \varphi\right)_{n \geq 0}$ converges in L^1 to the constant $\int_X \varphi \, \mathrm{d}\mu$, where $Q = P^*$.*

To prove this result we first need the following lemma whose proof gives some instruction on how the induced Markov measure behaves.

Lemma 2.4. *Let (X, μ) be a probability space and P a measure-preserving Markov operator. Let $(\mathbf{X}, \tilde{\mu})$ be the sequence space with the induced Markov measure and $\varphi \in L^1(X, \mu)$. Then $\mathbb{E}(\varphi^0 | \mathcal{F}_{\geq n}) = (Q^n \varphi)^0 \circ \sigma^n$, where $Q = P^*$ and σ is the shift map on $\mathbf{X}$.*

Proof. From the definition of conditional expectation, we wish to show that for every $A \in \mathcal{F}_{\geq n}$ we have $\int_A \varphi^0 \, \mathrm{d}\tilde{\mu} = \int_A (Q^n \varphi) \circ \sigma^n \, \mathrm{d}\tilde{\mu}$. First, by Kolmogorov's extension theorem it suffices to show this only for A in the generating set of $\mathcal{F}_{\geq n}$, so a set of the form $\prod_{i=0}^\infty A_i$ where $A_i = X$ for $0 \leq i \leq n-1$ and $A_i \in \mathcal{B}(X)$ for $i \geq n$ with all but finitely many A_i equal to X. We may write the integral over A as follows, noting that the integrand is now a function of the form used in the definition of $\tilde{\mu}$.

$$\int_A \varphi^0 \, \mathrm{d}\tilde{\mu} = \int_{\mathbf{X}} \varphi(x_0) \cdot \prod_{i=0}^\infty \chi_{A_i}(x_i) \, \mathrm{d}\tilde{\mu}$$

Note that this integral makes sense since all but finitely many of the χ_{A_i} are equal to 1. Now we use the definition of $\tilde{\mu}$.

$$\int_{\mathbf{X}} \varphi(x_0) \cdot \prod_{i=0}^\infty \chi_{A_i}(x_i) \, \mathrm{d}\tilde{\mu} = \int_X \varphi \chi_{A_0} P(\chi_{A_1} P(\chi_{A_2} P(\cdots \chi_{A_{n-1}} P(\chi_{A_n} P(\cdots) \cdots) \, \mathrm{d}\mu$$

Note that the first n indicator functions are just 1 since the corresponding A_i are the whole space X.

$$\begin{aligned}\int_X \varphi \chi_{A_0} P(\chi_{A_1} P(\chi_{A_2} P(\cdots \chi_{A_{n-1}} P(\chi_{A_n} P(\cdots)) \, \mathrm{d}\mu &= \int_X \varphi P^n(\chi_{A_n} P(\chi_{A_{n+1}} P(\cdots) \cdots)) \, \mathrm{d}\mu \\ &= \int_X (Q^n \varphi) \chi_{A_n} P(\chi_{A_{n+1}} P(\cdots)) \, \mathrm{d}\mu\end{aligned}$$

Since P preserves L^1 norm we have $\int f = \int Pf$ for positive functions f. Since we can split any L^1 function into the difference of two positive functions and P is linear the equality holds for all functions in L^1. So we can pad out the front of the integrand with n indicator

functions on the entire space (which are all just 1) and n copies of P to get it in the desired form.

$$\begin{aligned}
&\int_X (Q^n\varphi)\chi_{A_n} P(\chi_{A_{n+1}} P(\cdots))\,\mathrm{d}\mu \\
&\qquad = \int_X P^n((Q^n\varphi)\chi_{A_n} P(\chi_{A_{n+1}} P(\cdots))\,\mathrm{d}\mu \\
&\qquad = \int_X \chi_{A_0} P(\chi_{A_1} P(\cdots P(\chi_{A_{n-1}} P((Q^n\varphi)\chi_{A_n} P(\chi_{A_{n+1}} P(\cdots)\cdots)\,\mathrm{d}\mu \\
&\qquad = \int_{\mathbf{X}} (Q^n\varphi)(x_n)\chi_A(x)\,\mathrm{d}\tilde{\mu} \\
&\qquad = \int_A (Q^n\varphi)^0 \circ \sigma^n\,\mathrm{d}\tilde{\mu}
\end{aligned}$$

□

This proof demonstrates the way the Markov measure works. We see that the conditional expectation has the Q^n term in since we know that the coordinates of a set in $\mathcal{F}_{\geq n}$ that are not equal to X occur only from position n and onwards. One could interpret this with the same heuristics used in martingale theory: our best guess for events past time n is all the information we have up to time n. We now prove Proposition 2.3, following [4].

Proof. Take $\varphi \in L^1(X,\mu)$, by the lemma we have $\mathbb{E}(\varphi^0|\mathcal{F}_{\geq n}) = (Q^n\varphi)^0 \circ \sigma^n$. By the reverse martingale theorem, $\mathbb{E}(\varphi^0|\mathcal{F}_{\geq n}) \xrightarrow{L^1(\mathbf{X},\tilde{\mu})} \mathbb{E}(\varphi^0|\mathcal{F}_\infty)$. Note that $\mathbb{E}(\varphi^0|\mathcal{F}_\infty) = \int_X \varphi\,\mathrm{d}\mu$ if an only if $\mathcal{F}_\infty$ is trivial. Without loss of generality we may assume that $\int_X \varphi\,\mathrm{d}\mu = 0$. Using this assumption and the result of the lemma we have that the tail σ-algebra is trivial if and only if $(Q^n\varphi)^0 \circ \sigma^n \xrightarrow{L^1(\mathbf{X},\tilde{\mu})} 0$. Since the shift is invariant with respect to the induced Markov measure we have

$$\|(Q^n\varphi)^0 \circ \sigma^n\|_{L^1(\mathbf{X},\tilde{\mu})} = \|(Q^n\varphi)^0\|_{L^1(\mathbf{X},\tilde{\mu})} = \|Q^n\varphi\|_{L^1(X,\mu)}$$

Therefore $(Q^n\varphi)^0 \circ \sigma^n \xrightarrow{L^1(\mathbf{X},\tilde{\mu})} 0$ is equivalent to $Q^n\varphi \xrightarrow{L^1(X,\mu)} 0$ as required. □

2.3 Relation to spherical averages

I will now describe the Markov operator which was introduced by Grigorchuk, Bufetov and Thouvenot. Let $\mathcal{A} = \{1,2,\ldots,r,-1,-2,\ldots,-r\}$ be the set of states of a Markov chain (these will come to represent the generators of a free group and their inverses) and $\Pi = (p_{ij})_{i,j\in\mathcal{A}}$ a corresponding stochastic matrix. Assume we have a corresponding stationary distribution $p = (p_i)_{i\in\mathcal{A}}$ with all entries positive.

For each $a \in \mathcal{A}$ associate a function $T_a : X \to X$ and extend their action onto $L^1(X,\mu)$ by

$T_a\psi = \psi \circ T_{-a}$ and for any word $w = w_1 \dots w_n$, $w_i \in \mathcal{A}$ define the following

$$T_w = T_{w_1} \dots T_{w_n} \qquad p(w) = p_{w_n w_{n-1}} \dots p_{w_2 w_1} \qquad \pi(w) = p_{w_n} p(w)$$

We are interested in the following spherical average

$$S_n = \sum_{|w|=n} \pi(w) T_w$$

To write this in terms of a Markov operator we will work with the space $Y = X \times \mathcal{A}$ and the measure $v = \mu \times p$. Define the operator $P : L^1(Y,v) \to L^1(Y,v)$ as

$$P\varphi(x,i) = \sum_{j \in \mathcal{A}} p_{ij} \varphi(T_i x, j)$$

For $n = 2$ we have

$$\begin{aligned} P^2\varphi(x,i) &= \sum_{j \in \mathcal{A}} p_{ij} P\varphi(T_i x, j) \\ &= \sum_{j,k \in \mathcal{A}} p_{ij} p_{jk} \varphi(T_j T_i x, k) \end{aligned}$$

By induction we see that for $n > 1$ this gives

$$P^n\varphi(x,i) = \sum_{|w|=n-1, j \in \mathcal{A}} p_{i w_{n-1}} p(w) p_{w_1 j} \varphi(T_w T_i x, j)$$

Now take any $\psi \in L^1(X,\mu)$ and take a $\varphi \in L^1(Y,v)$ with $\varphi(x,a) = \psi(x)\ \forall a \in \mathcal{A}$. Using the above formula for P^n we get

$$S_n\psi(x) = \sum_{i \in \mathcal{A}} p_i (P^n\varphi)(x,i)$$

This equality holds because we are essentially prefixing words of length $n-1$ with elements $i \in \mathcal{A}$, thus making all words of length n.

I will state the main theorem from [3] and then show how this correspondence between the spherical average and P is used to simplify the proof. Since we will be modelling the generators of the group as a Markov chain there will of course be some assumptions made on the stochastic matrix Π which dictates the dynamics of the chain i.e. the dynamics of the group.

Definition 2.5. *The matrix Π generates the free group if $p_{ij} = 0$ is equivalent to $i + j = 0$.*

Definition 2.6. *The symmetry condition on Π is*

$$p_i = p_{-i} \qquad p_{-i,-j} = \frac{p_j p_{ji}}{p_i}$$

Remark 2.7. *In terms of the Markov chain the first definition states that the chain can't bounce between a generator and it's inverse. The symmetry condition is equivalent to saying S_n is self-adjoint.*

We are now nearly ready to state the main result of [3]. Denote by $\mathbb{F}_r$ the free group on r generators and suppose it acts on (X,μ) by measure-preserving transformations. Let $T_1,\dots,T_r : X \to X$ be transformations corresponding to the r generators. Define $T_{-i} := T_i^{-1}$ for $i \in \{1,\dots,r\}$. For $g \in \mathbb{F}_r$ we define T_g in the natural way where if the reduced form of g is $g = a_1 \dots a_n$ then $T_g = T_{a_1} \circ \cdots T_{a_n}$. We will again use the extension of the action to $L^1(X,\mu)$ given by $T_g\varphi = \varphi \circ T_{g^{-1}}$ for $g \in \mathbb{F}_r$. Let $\mathcal{A}$, Π, p be as before in this section and $\mathbb{F}_r^2$ denote the elements of $\mathbb{F}_r$ that have even word length.

Theorem 2.8. *Let (X,μ) be a Lebesgue probability space. Assume the matrix Π generates the free groups and satisfies the symmetry condition. Then for any $\varphi \in L\log L(X,\mu)$, the sequence $S_n\varphi$ converges as $n \to \infty$ both μ-almost everywhere and in $L^1(X,\mu)$ to a $\mathbb{F}_r^2$-invariant function.*

Using the correspondence between the Markov operator P and spherical averages and the way we can decompose group actions into ergodic components, to prove Theorem 2.12 it is enough to prove the following. We will be using the same space (Y,v) defined previously in this section.

Lemma 2.9. *Suppose Π generates the free group and satisfies the symmetry condition. Suppose the action of $\mathbb{F}_r^2$ on (X,μ) is ergodic. Then for any $\varphi \in L\log L(Y,v)$ we have $P^n\varphi \to \int_Y \varphi \,\mathrm{d}v$ both*

i). v-almost everywhere

ii). in $L^1(Y,v)$

In the subsequent subsections I shall prove Lemma 2.9 following the arguments made in [3].

2.4 Argument of Lemma 2.9 ii).

Proposition 2.10. *The operator adjoint to P is given by*

$$P^*\varphi(x,i) = \sum_{j\in\mathcal{A}} \frac{p_j p_{ji}}{p_i} \varphi(T_{-j}x, j)$$

Proof. Take $\psi, \varphi \in L^1(Y, v)$

$$\begin{aligned}
\int_{X\times\mathcal{A}} \psi(x,k)\cdot P\varphi(x,k)\,\mathrm{d}(\mu\times p) &= \sum_{i\in\mathcal{A}} \int_{X\times\{i\}} \psi(x,k) P\varphi(x,k)\,\mathrm{d}(\mu\times p) \\
&= \sum_{i\in\mathcal{A}} \int_{X\times\{i\}} \psi(x,k) \sum_{j\in\mathcal{A}} p_{kj}\varphi(T_k x, j)\,\mathrm{d}(\mu\times p) \\
&= \sum_{j\in\mathcal{A}} \Big(\sum_{i\in\mathcal{A}} \int_{X\times\{i\}} \psi(x,k) p_{kj}\varphi(T_k x, j)\,\mathrm{d}(\mu\times p)\Big) \\
&= \sum_{j\in\mathcal{A}} \Big(\sum_{i\in\mathcal{A}} \int_{X\times\{i\}} \psi(T_{-k}x, k) p_{kj}\varphi(x, j)\,\mathrm{d}(\mu\times p)\Big) \\
&= \sum_{j\in\mathcal{A}} \Big(\sum_{i\in\mathcal{A}} \int_{X\times\{i\}} \psi(T_{-i}x, i) p_{ij}\varphi(x, j)\,\mathrm{d}(\mu\times p)\Big) \\
&= \sum_{j\in\mathcal{A}} \Big(\sum_{i\in\mathcal{A}} \int_{X\times\{j\}} \psi(T_{-i}x, i) \frac{p_i p_{ij}}{p_j}\varphi(x, j)\,\mathrm{d}(\mu\times p)\Big) \\
&= \sum_{j\in\mathcal{A}} \int_{X\times\{j\}} \sum_{i\in\mathcal{A}} \frac{p_i p_{ij}}{p_j} \psi(t_{-i}x, i)\varphi(x, j)\,\mathrm{d}(\mu\times p) \\
&= \int_{X\times\mathcal{A}} \sum_{i\in\mathcal{A}} \frac{p_i p_{ik}}{p_k} \psi(T_{-i}x, i)\varphi(x, k)\,\mathrm{d}(\mu\times p) \\
&= \int_{X\times\mathcal{A}} P^*\psi(x,k)\cdot\varphi(x,k)\,\mathrm{d}(\mu\times p)
\end{aligned}$$

In the above I used k as a dummy variable to represent the $\mathcal{A}$ input of the function. The variables i, j represent a partition of $\mathcal{A}$. The sixth line can be interpreted as a sort of change of measure procedure, this is because we had partitioned the region being integrated over in a way that the $\mathcal{A}$ input into ψ is constant. So given a fixed $i, j \in \mathcal{A}$ we have

$$\int_{X\times\{i\}} \psi(T_{-i}x, i)\,\mathrm{d}(\mu\times p) = p_i \int_X \psi(T_{-i}x, i)\,\mathrm{d}\mu = \frac{p_i}{p_j}\int_{X\times\{j\}} \psi(T_{-i}x, i)\,\mathrm{d}(\mu\times p)$$

□

If we define $U\varphi(x,i) = \varphi(T_i x, -i)$ then U is a unitary operator, $U^2 = \mathrm{Id}$ and $P = UP^*U$. Note since the coefficients in P and P^* are positive we have $Pf \geq Pg, P^*f \geq P^*g$ when $f \geq g$ for $f, g \in L^1(Y, v)$.

Lemma 2.11. *Suppose Π generates the free group and satisfies the symmetry condition. Then there exists a positive constant c depending only on Π such that for any non-negative $\varphi \in L^1(Y, v)$ and any $n > 0$,*

$$(P^*)^n P^n \varphi \geq cUP^{2n-1}\varphi$$

Proof. We first consider the case $n = 1$. By definition we have

$$P^*P\varphi(x,i) = \sum_{k\in\mathcal{A}}\sum_{j\in\mathcal{A}} \frac{p_j p_{ji}}{p_i} p_{jk}\varphi(x,k)$$

Fix any $i,k \in \mathcal{A}$, since Π generates the free group we have $\sum_{j\in\mathcal{A}} \frac{p_j p_{ji}}{p_i} p_{jk} > 0$ because there are at most two terms in that sum which are zero, the rest are positive. This means we can write

$$P^*P\varphi(x,i) = \sum_{k\in\mathcal{A}} c_{i,k}\varphi(x,k)$$

where $c_{i,k} > 0$ and only depends on the entries of Π. Since

$$UP\varphi(x,i) = \sum_{k\in\mathcal{A}} p_{-i,k}\varphi(x,k)$$

the result holds for $n = 1$. We continue by induction, assuming the result holds for $n = k$. We have $(P^*)^k P^k \varphi \geq cUP^{2k}\varphi$, $\forall \varphi \in L^1(Y,v)$, therefore

$$\begin{aligned}(P^*)^{k+1}P^k\varphi &\geq cP^*UP^{2k}\varphi \\ &= cUP^{2k+1}\varphi\end{aligned}$$

for all $\varphi \in L^1(Y,v)$, in particular we can take any $\psi \in L^1(Y,v)$ and let $\varphi = P\psi$ and the inequality holds. This gives

$$(P^*)^{k+1}P^{k+1}\psi \geq cUP^{2k+1}\psi$$

as required. □

Using Proposition 2.3 and the way that $P = UP^*U$ it remains to show the triviality of the tail σ-algebra of P^*. To do this we will use the following two lemmas whose proof will rely on multiple further results. Lemma 2.12 is a version of the $0-2$ law for Markov operators [9].

Lemma 2.12. *Let P be an arbitrary measure-preserving Markov operator on a probability space (Z,μ).*

i). If the tail σ-algebra of P is trivial then for any $\varphi,\psi \in L^2(Z,\mu)$

$$\int_Z (Q^*)^n\varphi \cdot (Q^*)^n\psi \,\mathrm{d}\mu \to \int_Z \varphi \,\mathrm{d}\mu \int_Z \psi \,\mathrm{d}\mu$$

as $n \to \infty$

ii). If the tail σ-algebra of Q is non-trivial then for any $\varepsilon > 0$ there exist positive functions $\varphi,\psi \in L^\infty(Z,\mu)$ of integral 1 such that

$$\limsup_{n\to\infty} \int_Z (Q^*)^n\varphi \cdot (Q^*)^n\psi \,\mathrm{d}\mu < \varepsilon$$

Lemma 2.13. *Under the assumptions of Lemma 2.9, for any $\varphi, \psi \in L^2(Y, v)$,*

$$\int_Y P^n \varphi \cdot \psi \, \mathrm{d}v \to \int_Y \varphi \, \mathrm{d}v \int_Y \psi \, \mathrm{d}v$$

We now show triviality of the tail σ-algebra of P using the contrapositive of Lemma 2.12 ii). Take any positive $\varphi, \psi \in L^\infty(Y, v)$ of integral 1, then

$$\int_Y P^n \varphi \cdot P^n \psi \, \mathrm{d}v = \int_Y (P^*)^n P^n \varphi \cdot \psi \, \mathrm{d}v \geq c \int_Y U P^{2n-1} \varphi \cdot \psi \, \mathrm{d}v \to c \int_Y \varphi \, \mathrm{d}v \int_Y \psi \, \mathrm{d}v$$

The inequality is from Lemma 2.11 and the convergence from Lemma 2.13. Lemma 2.12 gives triviality of tail σ-algebra of $P^* = UPU$ which gives us triviality of the tail σ-algebra of P. Proposition 2.3 gives $P^n \xrightarrow{L^1} \int_Y \varphi \, \mathrm{d}v, \ \forall \varphi \in L^1(Y, v)$. It remains to prove the results we have used.

2.5 Proof of Lemma 2.12

Theorem 2.14 (Rota's Theorem)**.** *Let (Z, μ) be a probability space. Let Q be a measure-preserving Markov operator on $L^1(Z, \mu)$. Then for any $\varphi \in L \log L(Z, \mu)$ the sequence $(Q^*)^n Q^n \varphi$ converges μ-almost everywhere and in L^1 as $n \to \infty$.*

Proof. We shall be using a technique similar to the proof of Proposition 2.3. Let $\tilde{\mu}$ be the Markov measure on the one-way infinite sequence space of Z induced by Q. Previously we had defined $\tilde{\mu}$ only one-way infinite sequence spaces but the definition for the bi-infinite sequence space which we shall denote $\mathbf{Z}$ is analogous. We use the bi-infinite product σ-algebra generated by sets of the form $\prod_{i=-\infty}^{\infty} A_i$ where $A_i \in \mathcal{B}(Z)$ for all i, and all but finitely many of the A_i are equal to Z. Then let $j = \min\{i \in \mathbb{Z} : A_i \neq Z\}$ and define

$$\tilde{\mu}\left(\prod_{i=-\infty}^{\infty} A_i\right) = \tilde{\mu}\left(\prod_{i=0}^{\infty} A_{i+j}\right)$$

then extend to the whole product σ-algebra.

We also need to define tail σ-algebras for the bi-infinite sequence space. As in Definition 2.2 take any $k \in \mathbb{Z}$ and let $f_k : \mathbf{Z} \to Z$ be defined as $f_k(z) = z_k$. Then for $n, m \in \{-\infty\} \cup \mathbb{Z} \cup \{\infty\}$ with $n \leq m$ let $\mathcal{F}_n^m$ be the minimal σ-algebra on $(\mathbf{Z}, \tilde{\mu})$ such that f_i is measurable for all $n \leq i \leq m$.

Then following a similar derivation as in the proof of Proposition 2.3 and still using the notation that $\varphi^0(z) = \varphi(z_0)$, for any $\varphi \in L \log L(Z, \mu)$ we obtain the following expectation

formulas.

$$\mathbb{E}(\varphi^0|\mathcal{F}_{-\infty}^{-n})(z) = ((Q^n\varphi)\circ\sigma^{-n})(z) = Q^n\varphi(z_{-n})$$
$$\mathbb{E}(\mathbb{E}(\varphi^0|\mathcal{F}_{-\infty}^{-n})|\mathcal{F}_0)(z) = ((Q^*)^nQ^n\varphi)^0(z) = ((Q^*)^nQ^n\varphi)(z_0)$$

Write $X_n = Q^n\varphi$, then we see that X_n is a backward martingale that is adapted to the decreasing sequence of σ-algebras $\mathcal{F}_{-\infty}^0 \supseteq \mathcal{F}_{-\infty}^{-1} \subseteq \mathcal{F}_{-\infty}^{-2} \supseteq \ldots$. Therefore by the backward martingale convergence theorem we get $\mathbb{E}(\varphi^0|\mathcal{F}_{-\infty}^{-n}) \to \mathbb{E}(\varphi^0|\bigcap_{n\geq 0}\mathcal{F}_{-\infty}^{-n})$. Then using that $\varphi \in L^1$ to get the bound in Proposition 1.19 of [16] we get that $\mathbb{E}(\mathbb{E}(\varphi^0|\mathcal{F}_{-\infty}^{-n})|\mathcal{F}_0)$ converges as required. □

Proposition 2.15. *Suppose that the tail σ-algebra of Q^* is trivial. Then for all $\varphi \in L\log L(Z,\mu)$, $\lim_{n\to\infty}(Q^*)^nQ^n\varphi = \int \varphi \,\mathrm{d}\mu$*

Proof. This follows easily from Theorem 2.14 since the expectation with respect to trivial σ-algebra is just the integral. □

Proposition 2.16. *Let Q be a measure-preserving Markov operator on a probability space (Z,μ). If the tail σ-algebra of Q^* is trivial then for any $\varphi \in L^2(Z,\mu)$ we have $Q^n\varphi \xrightarrow{L^2} \int \varphi$.*

Proof. Using Proposition 2.15, given a measure-preserving Markov operator with trivial tail σ-algebra and $\varphi \in L^2(Z,\mu)$ (which due to the linearity of Q we may assume without loss of generality has integral 0) we have

$$\int_Z (Q^n\varphi)^2 \,\mathrm{d}\mu = \int_Z (Q^*)^nQ^n\varphi\cdot\varphi \,\mathrm{d}\mu \to 0$$

as $n \to \infty$ by Proposition 2.15 □

The following is a proof of Lemma 2.12

Proof. Throughout this proof $\mathbf{Z}$ will denotes the bi-infinite sequence space of Z and $\tilde{\mu}$ will be the Markov measure induced by Q as in the proof of Theorem 2.14. The first part of the lemma is immediate from Proposition 2.16. For the second part fix $\varepsilon > 0$ and let $\mathcal{F}_\infty$ be the tail σ-algebra of Q. We are assuming it is non-trivial so let $A \in \mathcal{F}_\infty$ be such that $0 < \tilde{\mu}(A) < 1$. Define Φ, Ψ on $\mathbf{Z}$ as follows

$$\Phi(z) = \frac{\chi_A(z)}{\tilde{\mu}(A)} \qquad \Psi(z) = \frac{\chi_{A^c}(z)}{\tilde{\mu}(A^c)}$$

We see both are positive, bounded, tail measurable, have integral one and are orthogonal in the sense that $\Phi\cdot\Psi = 0$. Let M be a constant which bounds both of them and let $\varphi_k(z_k) = \mathbb{E}(\Phi(z)|\mathcal{F}_{\leq k}), \psi_k(z_k) = \mathbb{E}(\Psi(z)|\mathcal{F}_{\leq k})$. We use z_k as the input into the functions to emphasize that there is only dependence on the kth entry of the sequence input, this

follows from similar expectation formulas done in previous arguments. Then φ_k, ψ_k are both positive and bounded my M. By the martingale convergence theorem we have $\varphi_k(z_k) \to \Phi(z)$ and $\psi_k(z_k) \to \Psi(z)$ both $\tilde{\mu}$ almost everywhere are in $L^1(\mathbf{Z}, \tilde{\mu})$ as $k \to \infty$. Since $A \in \mathcal{F}_\infty$ we have that both Φ, Ψ are $\mathcal{F}_\infty$ measurable. By Proposition 1.19 in [16] we have

$$\mathbb{E}(\varphi_k(z_k)|\mathcal{F}_\infty) \to \Phi \qquad \mathbb{E}(\psi_k(z_k)|\mathcal{F}_\infty) \to \Psi$$

with the convergence being $\tilde{\mu}$ almost everywhere and in $L^1(\mathbf{Z}, \tilde{\mu})$. Since Φ, Ψ are orthogonal we can choose a K such that $\forall\, k > K$

$$\int_{\mathbf{Z}} \mathbb{E}(\varphi_k(z_k)|\mathcal{F}_\infty)\mathbb{E}(\psi_k(z_k)|\mathcal{F}_\infty)\,\mathrm{d}\tilde{\mu} < \varepsilon$$

Following a similar argument as in Subection 2.2 we have,

$$\mathbb{E}(\varphi_k(z_k)|\mathcal{F}_{\geq n+k}) = (Q^*)^n \varphi_k(z_{n+k}) \qquad \mathbb{E}(\psi_k(z_k)|\mathcal{F}_{\geq n+k}) = (Q^*)^n \psi_k(z_{n+k})$$

Deriving the above equality hinges around the fact that $\varphi_k(z_k)$ depends only on the k-th coordinate of an element in $\mathbf{Z}$ and given any set $A \in \mathcal{F}_{\geq n+k}$ the first coordinate which could have a nontrivial indicator associated with it is $n + k$. Therefore the integral of $\varphi_k(z_k)\chi_A$ over $\mathbf{Z}$ with respect to $\tilde{\mu}$ has n compositions of Q between $\varphi_k(z_k)$ and $\chi_{A_{n+k}}$, these can be moved to the front using the definition of Q^*.

Finally, we can say that as $n \to \infty$,

$$\begin{aligned}\int_Z (Q^*)^n \varphi_k(z) \cdot (Q^*)^n \psi_k(z)\,\mathrm{d}\mu &= \int_{\mathbf{Z}} \mathbb{E}(\varphi_k(z_k)|\mathcal{F}_{\geq n+k})\mathbb{E}(\psi_k(z_k)|\mathcal{F}_{\geq n+k})\,\mathrm{d}\tilde{\mu} \\ &\to \int_{\mathbf{Z}} \mathbb{E}(\varphi_k(z_k)|\mathcal{F}_\infty)\mathbb{E}(\psi_k(z_k)|\mathcal{F}_\infty)\,\mathrm{d}\tilde{\mu} < \varepsilon\end{aligned}$$

□

2.6 Proof of Lemma 2.13

Note that in the proof of Lemma 2.12 we didn't use the Markov operator P specifically, instead we always assumed that we were using an arbitrary measure-preserving Markov operator. Recall how we were previously using the Markov operator P, space $Y = X \times \mathcal{A}$ and measure $v = \mu \times p$.

In the following we will be using the space $(\mathbf{Y}, \tilde{v})$ where $\mathbf{Y}$ is the bi-infinite sequence space of Y and $\tilde{v}$ is the Markov measure induced by P. Let σ_P be the shift on $(\mathbf{Y}, \tilde{v})$. The following is a result from [13] (Lemma 9).

Lemma 2.17. *A measure-preserving system $(X, \mathcal{B}, \mu, T)$ is mixing if and only if*

$$\lim_{n \to \infty} \int (f \circ T^n) g \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu \int g \, \mathrm{d}\mu$$

for every $f, g \in L^2(X, \mu)$.

We can now see that if σ_P is mixing then by Lemmma 2.17 we will have Lemma 2.13. This is because for any $\varphi, \psi \in L^2(Y, v)$ we may define $f, g \in L^2(\mathbf{Y}, \tilde{v})$ by $f(y) = \varphi(y_0)$, $g(y) = \psi(y_0)$. Then we have

$$\begin{aligned} \int_{\mathbf{Y}} (f \circ \sigma_P^n) g \, \mathrm{d}\tilde{v} &= \int_{\mathbf{Y}} \varphi(y_n) \psi(y_0) \, \mathrm{d}\tilde{v} \\ &= \int_Y (P^n \varphi) \psi \, \mathrm{d}v \end{aligned}$$

The second equality is due to the definition of $\tilde{v}$. Therefore to show Lemma 2.13 it suffices to show σ_P is mixing.

Definition 2.18. *Let $(X, \mathcal{B}, \mu)$ be a standard probability space and T a measure-preserving transformation. We say T has the K-property if there exists a sub σ-algebra $\mathcal{K} \subset \mathcal{B}$ such that the following three properties holds*

i). $\mathcal{K} \subset T\mathcal{K}$

ii). $\bigvee_{n=0}^{\infty} T^n \mathcal{K} = \mathcal{B}$

iii). $\bigcap_{n=0}^{\infty} T^{-n} \mathcal{K} = \{X, \emptyset\}$

Note that if T has the K-property then it is strong mixing, for a proof of this see [12] (page 52). The K-property is a natural thing to look at since the σ-algebras we have been using so far are generated by the shift map. To simplify the following arguments we will use a system that is isomorphic to $(\mathbf{Y}, \tilde{v}, \sigma_P)$, before the explanation first recall the definition of an isomorphism between two systems.

Definition 2.19. *We say two measure-preserving systems $(X, \mathcal{B}, \mu, T)$ and $(Y, \mathcal{C}, m, S)$ are isomorphic if there exists $M \in \mathcal{B}$ and $N \in \mathcal{C}$ such that*

i). $TM \subset M$, $SN \subset N$,

ii). $\mu(M) = 1$, $m(N) = 1$

and there exists a bijection $F : M \to N$ such that

i). F, F^{-1} *are measurable and measure-preserving*

ii). $F \circ T = S \circ F$

Let $\mathbf{A}$ be the space of bi-infinite sequences of elements of $\mathcal{A}$ and $\sigma_{\mathcal{A}}$ be the shift on $\mathbf{A}$. Let p_Π be the $\sigma_{\mathcal{A}}$-invariant Markov measure on $\mathbf{A}$ corresponding to Π and its stationary

distribution p. Define the skew product $T : X \times \mathbf{A} \to X \times \mathbf{A}$ by

$$T(x,w) = (T_{w_0}x, \sigma_{\mathcal{A}}(w),)$$

T preserves the measure $\mu \times p_\Pi$ since p_Π is $\sigma_{\mathcal{A}}$-invariant and T_{w_0} is a measure-preserving transformation for every $w_0 \in \mathcal{A}$.

Lemma 2.20. *The systems $(X \times \mathbf{A}, p_\Pi \times \mu, T)$ and $(\mathbf{Y}, \tilde{v}, \sigma_P)$ are isomorphic.*

Proof. For motivation why the two might be isomorphic note that the map T retains information about the sequence in the form of the actions that have been applied to x. For each $y \in \mathbf{Y}$ denote the n-th component as (i_n, x_n) with $i_n \in \mathcal{A}$, $x_n \in X$. Let $w(y) = (i_n)_{n\in\mathbb{Z}}$ It is natural to think that the bijection will involve just mapping y to $w(y) \in \mathbf{A}$. As for the other coordinate, note that $\mathbf{Y}_{inv} = \{y \in \mathbf{Y} : x_1 = T_{i_0}x_0, x_2 = T_{i_1}x_1, \ldots$ and $x_{-1} = T_{-i_{-1}}x_0, x_{-2} = T_{-i_{-2}}x_0, \ldots\}$ is such that $\tilde{v}(\mathbf{Y}_{inv}) = 1$. This is because $\sigma_P(\mathbf{Y}_{inv}) = \mathbf{Y}_{inv}$. The following diagram demonstrates the structure of elements in this set, from the diagram it is clear why $\mathbf{Y}_{inv}$ is shift-invariant.

$$\cdots\cdots \quad x_{-2} \xrightarrow{T_{i_{-2}}} x_{-1} \xrightarrow{T_{i_{-1}}} x_0 \xrightarrow{T_{i_0}} x_1 \xrightarrow{T_{i_1}} x_2 \quad \cdots\cdots$$

Therefore almost every $y \in \mathbf{Y}$ is completely determined by its $w(y)$ and x_0. Writing $x(y) = x_0$ we define $F(y) = (x(y), w(y))$ and note that using $M = \mathbf{Y}_{inv}$, $N = X \times \mathbf{A}$ in the definition shows that the two systems are isomorphic as desired. □

Lemma 2.21. *Assume $\mathbb{F}_r^2$ acts ergodically on (X, μ) and assume the matrix Π generates the free group and satisfies the symmetry condition. Then the system $(\mathbf{Y}, \tilde{v}, \sigma_P)$ has the K-property.*

The proof of Lemma 2.21 is based on the Rohlin-Sinai theorem below, which is discussed in more detail in [14]. The theorem involves the Pinsker σ-algebra which is defined as $\pi(T) = \{A \in \mathcal{B}(X) | h(T, \{A, A^c\}) = 0\}$ so it is the σ-algebra of sets whose canonical partitions have zero entropy. We say T has completely positive entropy if $\pi(T)$ is trivial. Note $\pi(T) = \pi(T^{-1})$.

Theorem 2.22 (Rohlin-Sinai)**.** *The following two results are true*

i). T has the K-property if and only if T has completely positive entropy.

ii). If a measurable partition ζ is such that $\zeta \subset T\zeta$ and $\bigvee_{n\in\mathbb{Z}} T^n\zeta = \varepsilon$, where ε is the partition of the space into individual points, then $\pi(T) \subset \bigcap_{n=0}^{\infty} T^n\zeta$.

Before we show that the former system in Lemma 2.20 has the K-property, we will first establish some notation. Write $\mathcal{F}_m^k(\mathbf{Y})$ as the σ-algebra in $\mathbf{Y}$ generated by y_l for $m \le l \le k$, likewise for $\mathcal{F}_m^k(\mathbf{A})$ as a σ-algebra in $\mathbf{A}$. Write $\mathcal{F}_{\ge n}$ for $\mathcal{F}_n^\infty$ and $\mathcal{F}_k$ for $\mathcal{F}_k^k$. Denote by $\mathcal{B}(X)$ the σ-algebra of all μ measurable subsets of X, by $\mathcal{B}(X \times \mathbf{A})$ the σ-algebra of all $\mu \times_\Pi$ measurable subsets of $X \times \mathbf{A}$.

Consider $\mathcal{G}_+ = \mathcal{B}(X) \times \mathcal{F}_{\geq 0}(\mathbf{A})$ which is the σ-algebra representing the future of the Markov chain. Since T shifts the elements of $\mathbf{Y}$ to the left i.e. moving one step forwards in time, we have $\mathcal{G}_+ \subset T\mathcal{G}_+$ and likewise since T^{-1} moves a step backwards in time, we obtain $\bigvee_{k\in\mathbb{Z}} T^k\mathcal{G}_+ = \mathcal{B}(X \times \mathbf{A})$. Therefore by Theorem 2.22 ii). we have $\pi(T) \subset \bigcap_{n=0}^{\infty} T^n\mathcal{G}_+ = \mathcal{G}_+$. Now consider $\mathcal{G}_- = \mathcal{B}(X) \times \mathcal{F}_{\leq 0}(\mathbf{A})$. $\mathcal{G}_- \subset T^{-1}\mathcal{G}_-$ and $\bigvee_{k\in\mathbb{Z}} T^k\mathcal{G}_- = \mathcal{B}(X \times \mathbf{A})$ so again by Theorem 2.22 ii). $\pi(T) = \pi(T^{-1}) \subset \mathcal{G}_-$. So overall we have $\pi(T) \subset \mathcal{G}_+ \wedge \mathcal{G}_-$.

It is clear that

$$(\mathcal{B}(X) \times \mathcal{F}_{\geq 0}(\mathbf{A})) \wedge (\mathcal{B}(X) \times \mathcal{F}_{\leq 0}(\mathbf{A})) = \mathcal{B}(X) \times \mathcal{F}_0(\mathbf{A}).$$

If we let $\mathcal{G}_k = \mathcal{B}(X) \times \mathcal{F}_k(\mathbf{A})$ then $\pi(T) \subset \wedge_{k\in\mathbb{Z}}\mathcal{G}_k$. This is because as shown above $\pi(T) \subset \mathcal{G}_0$, $T\pi(T) = \pi(T)$ and $T^k\mathcal{G}_0 = \mathcal{G}_k$.

Our overall aim is to show that $\pi(T)$ is trivial. We have shown a property of $\pi(T)$ which relates it to $\mathcal{F}_k$, $k \in \mathbb{Z}$. We have also previously shown expectation results using $\mathcal{F}_k$, so it is natural to try and use expectation results in showing the triviality of $\pi(T)$. In particular we will take the expectation of a $\pi(T)$-measurable function and use it to show that the function must be constant.

Let $\varphi : X \times \mathbf{A} \to \mathbb{R}$ be $\pi(T)$-measurable. Therefore it is $\mathcal{G}_k$-measurable for every $k \in \mathbb{Z}$, so for any $k \in \mathbb{Z}$ there exists $\varphi_k : Y \to \mathbb{R}$ such that $\varphi(x,w) = \varphi_k(,w_k)$. Therefore for every $k \in \mathbb{Z}$ we have $\mathbb{E}(\mathbb{E}(\varphi|\mathcal{G}_k)|\mathcal{G}_0) = \varphi = \varphi_0$. But we know that for any $n \in \mathbb{N}$ we have

$$\mathbb{E}(\mathbb{E}(\varphi|\mathcal{G}_n)|\mathcal{G}_0) = P^n(P^*)^n\varphi_0$$
$$\mathbb{E}(\mathbb{E}(\varphi|\mathcal{G}_{-n})|\mathcal{G}_0) = (P^*)^nP^n\varphi_0$$

Therefore for every $n \in \mathbb{N}$ we have

$$(P^*)^nP^n\varphi_0 = P^n(P^*)^n\varphi_0 = \varphi_0$$

Take any $a,b \in \mathbb{R}$ and consider $\varphi_0^{-1}(a,b)$. Fix $(x,i) \in \varphi_0^{-1}(a,b)$ then for any $w \in \mathbf{A}$ with $w_0 = i$ we have $\varphi(x,w) = \varphi_0(x,w_0) \in (a,b)$. Now choose any $j \in \mathcal{A}$ and fix $w \in \mathbf{A}$ with $w_0 = i, w_1 = j$. Then $\varphi(x,w) = \varphi_1(x,j) = \varphi_0(x,i) \in (a,b)$. Now take any $k \in \mathcal{A}$ and fix $w' \in \mathbf{A}$ with $w_0 = k, w_1 = j$. Then

$$\varphi_0(x,i) = \varphi(x,w) = \varphi_1(x,j) = \varphi(x,w') = \varphi_0(x,k) \in (a,b)$$

Therefore we see that for some $(x,i) \in \varphi_0^{-1}(a,b)$ we have $(x,j) \in \varphi_0^{-1}(a,b)$ for every $j \in \mathcal{A}$. This means that φ_0 does not depend on $\mathcal{A}$. The same argument could be applied to any φ_k, therefore φ does not depend on $\mathcal{A}$. This means that to show that φ is constant it suffices to show that φ_0 is constant.

Definition 2.23. *A subset $A \subset Y$ is called P-invariant if $P\chi_A = \chi_A$.*

Proposition 2.24. *Suppose Π satisfies the symmetry condition and generates the free group. Then a set A is P^*P-invariant if and only if χ_A does not depend on $\mathcal{A}$.*

Proof. By definition of P, P^* we have,

$$P^*P\chi_A(x,i) = \sum_{l\in\mathcal{A}}\sum_{k\in\mathcal{A}} \frac{p_k p_{ki}}{p_i} p_{kl}\chi_A(x,l)$$

First assume χ_A does not depend on $\mathcal{A}$. Then $\chi_A(x,i) = \chi_A(x,l)\ \forall\, l \in \mathcal{A}$. Therefore

$$\begin{aligned}
\sum_{k\in\mathcal{A}}\sum_{l\in\mathcal{A}} \frac{p_k p_{ki}}{p_i} p_{kl}\chi_A(x,l) &= \sum_{k\in\mathcal{A}}\sum_{l\in\mathcal{A}} \frac{p_k p_{ki}}{p_i} p_{kl}\chi_A(x,i) \\
&= \sum_{k\in\mathcal{A}} \frac{p_k p_{ki}}{p_i}\chi_A(x,i) \\
&= \sum_{k\in\mathcal{A}} p_{-i.-k}\chi_A(x,i) \\
&= \chi_A(x,i)
\end{aligned}$$

The second and fourth lines are due to Π being stochastic, and the third is due to the symmetry condition.

Now assume χ_A is P^*P-invariant. Note that by a similar derivation as above we have that

$$\sum_{l,k\in\mathcal{A}} \frac{p_k p_{ki}}{p_i} p_{kl} = 1$$

Also note that $a_{i,l} := \sum_{k\in\mathcal{A}} \frac{p_k p_{ki}}{p_i} p_{kl} > 0$ because due to Π generating the free group and p_i being positive for all $i \in \mathcal{A}$ we have that at most only two terms in that sum can be zero. Therefore $\chi_A(x,i) = \sum_{l\in\mathcal{A}} a_{i,l}\chi_A(x,l)$ which gives $\chi_A(x,i) = \chi_A(x,l)$ for all $l \in \mathcal{A}$ or else that equality wouldn't hold. □

Lemma 2.25. *Suppose Π generates the free group and satisfies the symmetry condition. Suppose a set $A \subset Y$ is both P^*P and $(P^*)^2P^2$ invariant. Then χ_A does not depend on $\mathcal{A}$ and is $\mathbb{F}_r^2$-invariant.*

Proof. By Proposition 2.24 we have that χ_A does not depend on $\mathcal{A}$. We may write

$$\begin{aligned}
\chi_A(x,i) = (P^*)^2P^2\chi_A(x,i) &= \sum_{j,k,l,m\in\mathcal{A}} \frac{p_j p_{ji}}{p_i}\frac{p_k p_{kj}}{p_j} p_{kl}p_{lm}\chi_A(T_lT_{-j}x,m) \\
&= \sum_{j,k,l\in\mathcal{A}} \frac{p_j p_{ji}}{p_i}\frac{p_k p_{kj}}{p_j} p_{kl}\chi_A(T_lT_{-j}x,i)
\end{aligned}$$

The last equality comes from uses that Π is stochastic and χ_A doesn't depend on $\mathcal{A}$. As

in the previous proof let $a_{j,l} = \sum_{k\in\mathcal{A}} \frac{p_k p_{kj}}{p_j} p_{kl}$, again this is positive due to Π generating the free group. So we may write the above sum as

$$\sum_{j\in\mathcal{A}} \sum_{l\in\mathcal{A}} \frac{p_j p_{ji}}{p_i} a_{j,l} \chi_A(T_l T_{-j} x, i)$$

this equality holds for all i. Therefore we get that $\chi_A(x,i) = \chi_A(T_l T_{-j} x, i)$ for all $l, j \in \mathcal{A}$, thus χ_A is $\mathbb{F}_r^2$-invariant. □

We have already deduced that $\chi_{\varphi_0^{-1}(a,b)}$ does not depend on $\mathcal{A}$. We know that φ_0 is P^*P and $(P^*)^2P^2$ invariant, therefore the $\varphi_0^{-1}(a,b)$ is P^*P and $(P^*)^2P^2$ invariant. So by Lemma 2.25 we see $\varphi_0^{-1}(a,b)$ is also $\mathbb{F}_r^2$ invariant. In the assumptions of Lemma 2.21 we have that $\mathbb{F}_r^2$ acts on X ergodically, therefore $\varphi_0^{-1}(a,b)$ has measure 0 or 1. Since a, b were arbitrary we can conclude that φ_0, hence φ, is constant which proves Lemma 2.21.

Thus the system $(\mathbf{A} \times X, p_\Pi \times \mu, T)$, and hence the system $(\mathbf{Y}, \tilde{v}, \sigma_P)$, has the K-property. Therefore σ_P is mixing as required to prove Lemma 2.13. So he have shown Lemma 2.9 ii).

2.7 Argument of Lemma 2.9 i).

To prove the result we will be using the Orlicz norm, see [18]. Let (Z, μ) be an arbitrary probability space, the Orlicz norm on the space $L\log L(Z,\mu)$ is defined as

$$\|\varphi\|_{L\log L} = \inf\{c : \int_Z \frac{|\varphi|}{c} \cdot \log\Big(\frac{|\varphi|}{c} + 2\Big)\, \mathrm{d}\mu \leq 1\}$$

We obtain the following maximal inequality from [11] (Pages 68 and 70) by using the expectation formulas of $(Q^*)^n Q^n$ as martingales.

Lemma 2.26. *Let (Z,μ) be a probability space and let Q be a measure-preserving Markov operator on $L^1(Z,\mu)$.*

i). For any $p > 1$ there exists a constant $A_p > 0$ such that for any $\varphi \in L^p(Z,\mu)$ we have

$$\|\sup_n (Q^*)^n Q^n \varphi\|_{L^p} \leq A_p \|\varphi\|_{L^p}$$

ii). There exists a constant $A > 0$ such that for any $\varphi \in L\log L(Z,\mu)$,

$$\|\sup_n (Q^*)^n Q^n \varphi\|_{L^1} \leq A \|\varphi\|_{L\log L}$$

Using Lemma 2.26 and the inequality in Lemma 2.11 gives us the following.

Lemma 2.27. *Let $p > 1$ then we have both*

i). There exists a constant $p > 1$ such that for any $\varphi \in L^p(Y,v)$,

$$\|\sup_n P^{2n}\varphi\|_{L^p} \leq A_p\|\varphi\|_{L^p}$$

ii). There exists a constant $A > 0$ such that for any $\varphi \in L\log L(Y,v)$,

$$\|\sup_n P^{2n}\varphi\|_{L^1} \leq A\|\varphi\|_{L\log L}$$

We now combine the results of Proposition 2.16 and Lemma 2.27 together with a density argument. Take any $\varphi \in L^2(Y,v)$ with $\int_Y \varphi \, \mathrm{d}v = 0$. Then $\|P^n\varphi\|_{L^2} \to 0$ by Proposition 2.16. Using the inequality in Lemma 2.27 for any $k \in \mathbb{N}$ we have

$$\|\sup_n P^{2n+2k}\varphi\|_{L^2} \leq A_2\|P^{2k}\varphi\|_{L^2}$$

with the right hand side tending to 0 as $k \to \infty$. This gives us pointwise convergence of $P^{2n}\varphi$ to $\int \varphi$ for $\varphi \in L^2(Y,v)$. Having already proven Lemma 2.9 ii). we have convergence in L^1 of the entries sequence $P^n\varphi$, which gives us pointwise convergence of the entire sequence $P^n\varphi$ for $\varphi \in L^2(Y,v)$. Since $L^2(Y,v)$ is dense in $L\log L(Y,v)$, we can use a density argument along with Lemma 2.27 ii). to get pointwise convergence of $P^n\varphi$ for any $\varphi \in L\log L(Y,v)$. This completes the proof of Lemma 2.9.

2.8 Summary

We wrote the spherical average S_n, whose sum index depended on n, in terms of a sum over the elements of $\mathcal{A}$ of powers of the Markov operator P. Convergence in L^1 followed from Proposition 2.3 once we had the triviality of the tail σ-algebra of P, this is due to the way which we can frame convergence of P^n in terms of martingale convergence. To get this triviality we used the equality $P^* = UPU$ and the Markov operator version of the $0-2$ law, Proposition 2.12. In particular we use the contrapositive on part two of that result which required us to show an inequality involving $\int P^n\varphi P^n\psi$. To show this we used the inequality of Lemma 2.11, which depends explicitly on the formulation of P and does not hold for any arbitrary Markov operator, and Lemma 2.13. In particular Π generating the free group and satisfying the symmetry condition are critical for this proof.

The pointwise convergence part of Lemma 2.9 hinges on Lemma 2.26, which are maximal inequalities that are obtained by writing $(Q^*)^nQ^n\varphi$ in terms of expectations conditioned on σ-algebras depending on n, hence as martingales, and using martingale maximal inequalities. In view of Lemma 2.11 we get Lemma 2.27, a maximal inequality involving P^{2n}. Proposition 2.16 follows from Theorem 2.14 and then a density argument gives the

desired pointwise convergence.

3 Main paper

3.1 Extending results to more general Markov chains

The previous section outlined convergence results for situations where the entries of Π generate the free group and satisfy the symmetry condition. This section will outline Bowen, Bufetov, Romaskevich [1] which has conditions only concerning zero and non-zero entries of Π rather than relationships between the entries.

Whereas in the previous section our alphabet $\mathcal{A}$ was in direct correspondance with the generators of $\mathbb{F}_r$ and their inverses, we will now consider any finite alphabet V and a labelling map $\mathcal{L} : V \to \mathbb{F}_r$. So in the previous example the $\mathcal{L}$ we were using was bijective, it is important to note that this is not an explicit assumption for the following arguments. Indeed, the reason for this generality on $\mathcal{L}$ is due to the way that there are times when we give a free group a Markov coding we don't have $\mathcal{L}$ always bijective. Instead we require assumptions on Π which relate to the graph that it induces, paths in that graph and what $\mathcal{L}$ maps those paths to.

A lot of arguments are mirrored from the previous section. Most importantly the use of writing a spherical average in terms of powers of a Markov operator, writing powers of that Markov operator in terms of expectations with respect to certain σ-algebras and then using martingale convergence results.The majority of the proof of the main result will be dedicated to a technical intermediate result showing equality of two σ-algebras which will then be used to show convergence.

3.2 Set up

(X, μ) is a probability space, $\mathbb{F}_r$ the free group on r generators. Let $\mathbb{F}_r$ act on X by measure-preserving transformations, V a finite alphabet and $\mathcal{L} : V \to \mathbb{F}_r$ a labelling map. Let Π be a stochastic matrix with rows and columns indexed by V. Assume Π has a strictly positive stationary distribution and call it p. Let $G = (V, E)$ denote the directed graph on V with edge set

$$E := \{(w, v) : \Pi_{vw} > 0\} = \{(w, v) : \Pi^T_{wv} > 0\}$$

A directed path in G of length n is a sequence $s = (s_1, \ldots, s_n) \in V^n$ such that $(s_i, s_{i+1}) \in E$ for all i. Concatenation of a path q of length m and a path p of length n is done in the

natural way of $pq = (p_1, \ldots, p_n, q_1, \ldots, q_m) \in V^{n+m}$. We denote

$$\mathcal{L}(s) := \mathcal{L}(s_1) \cdots \mathcal{L}(s_n) \in \mathbb{F}_r, \;\; T_s = T_{\mathcal{L}(s)}$$

We define spherical averages $S_n : L^1(X, \mu) \to L^1(X, \mu)$

$$S_n \varphi(x) = \sum_{s=(s_1,\ldots,s_n)} \pi(s) \varphi(T_s x)$$

where we have used the notation of $\pi(s)$ from the previous section.

3.3 Main theorem

Definition 3.1. *A subgraph $H \subset G$ is called good of order k if it consists of vertices u, w and directed paths p, q, p^*, q^* of length k such that*

i). upw, uqw, pq^*p, qp^*q are directed paths in G

ii). $\mathcal{L}(p^*) = \mathcal{L}(p)^{-1}$ and $\mathcal{L}(q^*) = \mathcal{L}(q)^{-1}$

Remark 3.2. *A good subgraph does not have to be an induced subgraph.*

Definition 3.3. *For each $v \in V$ let $\Gamma_v \leq \mathbb{F}_r$ be the subgroup generated by all elements of the form $\mathcal{L}(p)$ where pv is a directed path from v to itself in G i.e. p is a path starting at v and ending at a node adjacent to v.*

Definition 3.4. *Π is admissable of order k if the following three conditions all holds*

i). G contains a good subgraph of order k

ii). G is strongly connected

iii). there is some $v \in V$ such that $\Gamma_v = \mathbb{F}_r$

Remark 3.5. *For there to exist a $v \in V$ such that $\Gamma_v = \mathbb{F}_r$ the map $\mathcal{L}$ has to be surjective onto the set of generators of $\mathbb{F}_r$ in the sense that for every generator α of $\mathbb{F}_r$ there has to exist some $w \in V$ such that α appears as a letter of the word $\mathcal{L}(w)$ or $\mathcal{L}(w)^{-1}$.*

Theorem 3.6. *Let (X, μ) be a probability space V a finite alphabet and Π a stochastic matrix with entries indexed by V. Suppose Π has a positive stationary distribution p. Let $\mathbb{F}_r$ be the fre group on r generators and suppose $\mathbb{F}_r$ has an action on (X, μ) that preserves μ. Let $\mathcal{L} : V \to \mathbb{F}_r$ be a labelling map. Let $E = \{(w, v) : \Pi^T_{w,v} > 0\}$ be the edges of $G = (V, E)$ and assume Π is admissable of order k. Then for every $f \in L^1(X, \mu)$*

$$\frac{1}{2k} \sum_{i=0}^{2k-1} S_{n+i} f \to \mathbb{E}(f|\mathcal{I})$$

converges in L^1 as $n \to \infty$ where $\mathcal{I}$ is the σ-algebra of $\mathbb{F}_r$-invariant measurable subsets.

3.4 Motivation

As in the previous section we may first write this spherical average in terms of a Markov operator. Define $P : L^1(V \times X) \to L^1(V \times X)$ as follows

$$P\varphi(u,x) = \sum_{w \in V} p_{uw} \varphi(w, T_u x)$$

Therefore

$$P^n \varphi(u,x) = \sum_{s_1,\ldots,s_n \in V} \Pi_{(s_1,\ldots,s_n,u)} \varphi(s_1, T_{(s_2,\ldots,s_n,u)} x)$$

Remark 3.7. *Previously we summed p_{ij} over j, now we are summing p_{wv} over w, this is deliberate.*

Using this definition of P^n and supposing that $\varphi(u,x) = \psi(x)\ \forall u \in V$, for some $\psi \in L^1(X)$ (in other words, φ doesn't depend on its V input) then we have

$$\begin{aligned}
P^n \varphi(u,x) &= \sum_{s_1,\ldots,s_n \in V} p_{u,s_n} \ldots p_{s_2,s_1} \varphi(s_1, T_{(s_2,\ldots,s_n,u)} x) \\
&= \sum_{s_1,\ldots,s_n \in V} p_{u,s_n} \ldots p_{s_2,s_1} \varphi(u, T_{(s_2,\ldots,s_n,u)} x) \\
&= \sum_{s_2,\ldots,s_n \in V} p_{u,s_n} \ldots p_{s_3,s_2} \sum_{s_1 \in V} p_{s_2,s_1} \varphi(u, T_{(s_2,\ldots,s_n,u)} x) \\
&= \sum_{s_2,\ldots,s_n \in V} p_{u,s_n} \ldots p_{s_3,s_2} \varphi(u, T_{(s_2,\ldots,s_n,u)} x)
\end{aligned}$$

We get from the penultimate line to the last line using the way that Π is stochastic and $\varphi(u, T_{(s_2,\ldots,s_n,u)} x)$ doesn't depend on s_1. Now using this we may deduce the following.

$$\begin{aligned}
S_n \psi(x) &= \sum_{s=(s_1,\ldots,s_n)} \pi(s) \psi(T_s x) \\
&= \sum_{w \in V} p_w \sum_{s=(s_1,\ldots,s_{n-1})} \Pi_{(s_1,\ldots,s_{n-1},w)} \varphi(w, T_{(s_2,\ldots,s_{n-1},w)} x) \\
&= \sum_{w \in V} p_w P^n(\varphi)(w,x)
\end{aligned}$$

So once again the convergence of S_n becomes a question of convergence of P^n. But since we are dealing with a more general matrix Π we require that averaging factor k as in the statement of Theorem 3.6.

In particular the lack of direct correspondence between our finite alphabet and the generators means that we can't use something similar to the operator U in the previous section, because the group actions won't cancel out in the sum due to the lack of regularity of $\mathcal{L}$.

Hence we can't use convergence results of P^* to help get the desired result. We used

conditional expectations with respect to σ-algebras of sequence spaces to get convergence results of P^*, for example in Proposition 2.15. Therefore it would not be helpful to look at expectations with respect to such σ-algebras in this more general case.

If one looks at the powers of P we see that the lack of conditions on $\mathcal{L}$ translates into a lack of regularity on the group action part of P. If you had an $x = T^{-1}_{(t_1,\ldots,t_n)}y$ for some $t_1,\ldots,t_n \in V, y \in X$ then some cancelling would occur in the sum. Given the lack of regularity on $\mathcal{L}$ there is potentially more than one set of n elements that combine to give the same group action. Define $\sigma_X : V^{\mathbb{N}} \times X \to V^{\mathbb{N}} \times X$ as $\sigma_X(s,x) = (\sigma s, T_{\mathcal{L}(s_1)^{-1}}x)$. This is similar to formula 2.8 in [10] and allows us to retain information about the first terms of a sequence in $V^{\mathbb{N}}$ via the foot print of the group action on x. So if there is $(s,x),(t,y) \in V^{\mathbb{N}} \times X$ and $n_0 \in \mathbb{N}$ such that $\sigma_X^{n_0}(s,x) = \sigma_X^{n_0}(t,y)$ then we have that $T^{-1}_{\mathcal{L}(s_1,\ldots,s_{n_0})}x = T^{-1}_{\mathcal{L}(t_1,\ldots,t_{n_0})}y$ and that the entries of s and t agree from n_0 onward, in particular we have $\sigma_X^m(s,x) = \sigma_X^m(t,y)\ \forall m \geq n_0$.

So letting $\pi : V^{\mathbb{N}} \times X \to V \times X$ be the projection map defined by $\pi(s,x) = (s_1,x)$ we see that for $(s,x),(t,y)$ satisfying the equality above we have $(P\varphi)(\pi \circ \sigma_X^m)(s,x) = (P\varphi)(\pi \circ \sigma_X^m)(t,y)$ for all $m \geq n_0$. This motivates the use of the equivalence relation E_n^X where $(s,x)E_n^X(t,y)$ if $\sigma_X^n(s,x) = \sigma_X^n(t,y)$. Note that $(s,x)E_n^X(t,y)$ implies $(s,x)E_m^X(t,y)$ for $m \geq n$ so if we let $\mathcal{F}_n^E$ denote the σ-algebra generated by unions of equivalence classes of E_n^X then $\mathcal{F}_n^E$ forms a decreasing sequence of σ-algebras. Write $\mathcal{F}_{sync} = \bigcap_{n=0}^{\infty} \mathcal{F}_n^E$, then we see that $\mathcal{F}_{sync}$ is the σ-algebra generated by unions of equivalence classes of the equivalence relation E_{sync}^X defined as $(s,x)E_{sync}^X(t,y)$ if there exists $n \in \mathbb{N}$ such that $(s,x)E_n^X(t,y)$. So if we can relate P to an expectation with respect to $\mathcal{F}_n^E$ then we can use reverse martingale convergence theorem.

Indeed if we use the Markov measure induced by Π on $V^{\mathbb{N}}$ defined as follows

$$\tilde{v}\left\{\left(s \in V^{\mathbb{N}} : s_i = t_i\ \forall 1 \leq i \leq n\right)\right\} = p_{t_n}\Pi_t = p_{t_n}\Pi_{t_n,t_{n-1}}\cdots\Pi_{t_2,t_1}$$

Then using the product measure $\tilde{v} \times \mu$ on $V^{\mathbb{N}} \times X$ and taking any $\varphi \in L^1(V \times X)$ we have

$$\begin{aligned}(P^n\varphi)(\pi \circ \sigma_X^n)(s,x) &= (P^n\varphi)(s_{n+1}, T^{-1}_{(s_1,\ldots,s_n)}x) \\ &= \sum_{t_1,\ldots,t_n \in V} \Pi_{(t_1,\ldots,t_n,s_{n+1})}\varphi(t_1, T_{(t_2,\ldots,t_n,s_{n+1})}T^{-1}_{(s_1,\ldots,s_n)}x) \\ &= \mathbb{E}[\varphi \circ \pi | \mathcal{F}_n^E](s,x)\end{aligned}$$

We can generalise the equivalence relation E_n^X by noting that we could be in a situation where for some $(s,x),(t,y) \in V^{\mathbb{N}} \times X$ and some $n,m \in \mathbb{N}, n \neq m$ we have $\sigma_X^n(s,x) = \sigma_X^m(t,y)$. So first define the equivalence relation R_n on $V^{\mathbb{N}}$, $n \in \mathbb{N}$, so that sR_nt if there exists an $l \in \mathbb{Z}$ and $N \in \mathbb{N}$ such that $s_{ln+i} = t_i\ \forall\, i > N$, write R_{sync} for R_0 (the equivalence

relation stating that the tails are the same without needing shifting) and note that if $d|n$ then sR_nt implies sR_dt. Define $\alpha : R_1 \to \mathbb{F}_r$ as

$$\alpha(s,t) = \mathcal{L}(s_1)\cdots\mathcal{L}(s_{N+l})\cdot(\mathcal{L}(t_1)\cdots\mathcal{L}(t_N))^{-1}$$

where $N \in \mathbb{N}$, $l \in \mathbb{Z}$ are such that $s_{l+i} = t_i \,\forall i > N$. Define R_n^X as $(s,x)R_n^X(t,y)$ if sR_nt and $\alpha(s,t)y = x$. This is equivalent to saying $(s,x)R_n^X(t,y)$ if there exists $l \in \mathbb{Z}, N \in \mathbb{N}$ such that $\sigma_X^{ln+N}(s,x) = \sigma_X^N(t,y)$. Therefore $(s,x)E_n^X(t,y)$ implies $(s,x)R_{sync}^X(t,y)$ and $(s,x)R_{sync}^X(t,y)$ is equivalent to there existing an $n \in \mathbb{N}$ such that $(s,x)E_n^X(t,y)$, that is to say that R_{sync}^X is the same as E_{sync}^X. Let $\mathcal{F}_n^R$ be the σ-algebra generated by unions of equivalence classes of R_n^X and note that $\mathcal{F}_{sync} = \mathcal{F}_0^R$.

Lemma 3.8. *R_n^X is generated by σ_X^n and R_{sync}^X in the sense that if $(s,x)R_n^X(t,y)$ then there exist (t',y') and $l \in \mathbb{Z}$ such that $(s,x)R_{sync}^X(t',y')$ and $\sigma_X^{ln}(t',y') = (t,y)$.*

Proof. Suppose $(s,x)R_n^X(t,y)$, then there exists some $l \in \mathbb{Z}$ and $N \in \mathbb{N}$ such that $s_{ln+i} = t_i \,\forall\, i \geq N$ and

$$\alpha(s,t) = \mathcal{L}(s_1)\cdots\mathcal{L}(s_{ln+N})\cdot(\mathcal{L}(t_1)\cdots\mathcal{L}(t_N))^{-1}y = x$$

Assume without loss of generality that $l \in \mathbb{N}$. The case for l negative is analogous. Take any $a_1,\ldots,a_{ln} \in V$ and let $t' = (a_1,\ldots,a_{ln},t_1,t_2,\ldots) \in V^{\mathbb{N}}$ then $t'_i = s_i \,\forall\, i \geq ln+N$. Let $y' = \mathcal{L}(a_1)\cdots\mathcal{L}(a_{ln})y$. Then we have

$$\begin{aligned}
\alpha(s,t')y' &= \mathcal{L}(s_1)\cdots\mathcal{L}(s_{ln+N})\cdot(\mathcal{L}(a_1)\cdots\mathcal{L}(a_{ln})\mathcal{L}(t_1)\cdots\mathcal{L}(t_N))^{-1}y' \\
&= \mathcal{L}(s_1)\cdots\mathcal{L}(s_{ln+N})\cdot(\mathcal{L}(t_1)\cdots\mathcal{L}(t_N))^{-1}\cdot(\mathcal{L}(a_1)\cdots\mathcal{L}(a_{ln}))^{-1}y' \\
&= \mathcal{L}(s_1)\cdots\mathcal{L}(s_{ln+N})\cdot(\mathcal{L}(t_1)\cdots\mathcal{L}(t_N))^{-1}y \\
&= \alpha(s,t)y \\
&= x
\end{aligned}$$

This shows that $(s,x)E_{ln+N}^X(t',y')$, therefore $(s,x)R_{sync}^X(t',y')$. Finally by working through the similar calculations we see that $\sigma_X^{ln}(t',y') = (t,y)$ as required. □

Note also that $\mathcal{F}_{sync} \subset \mathcal{F}_n^R$ for all $n \in \mathbb{N}$. A reasonable question is whether there is a $n^* \in \mathbb{N}$ such that $\mathcal{F}_{sync} = \mathcal{F}_{n^*}^R$ and this is where the assumption that Π induces a graph G that contains a good subgraph H of order k is used. This condition allows us to know equalities involving $\mathcal{L}(p)^{-1}$ for a certain path in p in G. This is a crucial assumption since we do not have any other assumptions on $\mathcal{L}$ which reveal information about the inversion of the elements in the group that it maps to.

Theorem 3.9. *If the directed graph G, induced by Π in the way outlined above, contains a good subgraph of order k then $\mathcal{F}_{2k}^R = \mathcal{F}_{sync}$.*

3.5 Proof of Theorem 3.9

As previously noted it is clear that $\mathcal{F}_{sync} \subset \mathcal{F}_n^R$ for all $n \in \mathbb{N}$. For a given $n \in \mathbb{N}$ if we wish to show the reverse inclusion then it suffices to show that every $f \in L^1(V^{\mathbb{N}} \times X)$ which is invariant on R_{sync}^X equivalence classes is also invariant on R_n^X equivalence classes. Lemma 3.8 shows that σ_X^n together with R_{sync}^X generates the equivalence classes of R_n^X. So it suffices to show that if $f \in L^1(V^{\mathbb{N}} \times X)$ is R_{sync}^X invariant then $f \circ \sigma_X^n = f$. This is because for any $(s,x)R_n^X(t,y)$ we have $(t,y) = \sigma_X^{nl}(t',y')$ for some $l \in \mathbb{Z}$ with $(t',y')R_{sync}^X(t,y)$.

We are going to show this is the case for $n = 2k$. The main thrust of the argument is using the assumption that G contains a good subgraph of order H combined with the fact that the standard shift map σ is ergodic, we know that almost every $s \in V^{\mathbb{N}}$ passes through the path upw or uqw infinitely often. This uses the way that the measure of the cylinder set of sequences starting with upw or uqw is positive and Proposition 2.14 in [6].

For $s \in V^{\mathbb{N}}$ let $s_{[n,m]} = (s_n, s_{n+1}, \ldots, s_m)$ and let $s_{[n,\infty)} = (s_n, s_{n+1}, \ldots)$. Define the function $\tau_n : V^{\mathbb{N}} \to \mathbb{N}$ so that $\tau(s)$ is the n-th occurence of either upw or uqw, therefore

$$s_{[\tau_n(s),\tau_n(s)+k+1]} \in \{upw, uqw\}$$

Define $w_n : V^{\mathbb{N}} \to V^{\mathbb{N}}$ by

$$w_n(s) = \begin{cases} s_{[1,\tau_n(s)]} q s_{[\tau_n(s)+k+1,\infty)} & \text{if } s_{[\tau_n(s),\tau_n(s)+k+1]} = upw \\ s_{[1,\tau_n(s)]} p s_{[\tau_n(s)+k+1,\infty)} & \text{if } s_{[\tau_n(s),\tau_n(s)+k+1]} = uqw \end{cases}$$

So w_n replaces looks for the n-th occurence of upw or uqw and swaps p for q or vice versa, therefore is self-invertible.

Define $\psi_n : V^{\mathbb{N}} \to V^{\mathbb{N}}$ in terms of its composition with w_n as follows

$$\begin{aligned}(\psi_n \circ w_n)(s) &= \begin{cases} w_n(s)_{[2k+1,\tau_n(s)+k]} p^* w_n(s)_{[\tau_n(s)+1,\infty)} & \text{if } s_{[\tau_n(s),\tau_n(s)+k+1]} = upw \\ w_n(s)_{[2k+1,\tau_n(s)+k]} q^* w_n(s)_{[\tau_n(s)+1,\infty)} & \text{if } s_{[\tau_n(s),\tau_n(s)+k+1]} = uqw \end{cases} \\ &= \begin{cases} s_{[2k+1,\tau_n(s)]} q p^* q s_{[\tau_n(s)+k+1,\infty)} & \text{if } s_{[\tau_n(s),\tau_n(s)+k+1]} = upw \\ s_{[2k+1,\tau_n(s)]} p q^* p s_{[\tau_n(s)+k+1,\infty)} & \text{if } s_{[\tau_n(s),\tau_n(s)+k+1]} = uqw \end{cases}\end{aligned}$$

So $(\psi_n \circ w_n)$ looks for the n-th occurence of upw or uqw, shifts by $2k$ and swaps p for qp^*q or q for pq^*p.

Since V is finite we know there is a uniform bound on the ratio of non-zero entries of Π so let $C > 0$ be a constant such that

$$C^{-1} \leq \frac{d(w_n^{-1})_* \tilde{v}}{d\tilde{v}} \leq C, \; C^{-1} \leq \frac{d((\psi_n w_n)^{-1})_* \tilde{v}}{d\tilde{v}} \leq C$$

Finally let $d_{V^{\mathbb{N}}}$ be the metric on $V^{\mathbb{N}}$ defined as $d_{V^{\mathbb{N}}}(s,t) = \frac{1}{n}$ where $s_i = t_i \, \forall \, i < n$.

Proposition 3.10. *For every $n > 2k+1$ and $s \in V^{\mathbb{N}}$ we have*

i). $d_{V^{\mathbb{N}}}\Big((\psi_n \circ w_n)(s), (\sigma^{2k} \circ w_n)(s)\Big) \leq \frac{1}{\tau_n(s)-k}$

ii). $d_{V^{\mathbb{N}}}(s, w_n(s)) \leq \frac{1}{\tau_n(s)}$

iii). $w_n(s) R_{sync} s$ *and* $\psi_n(s) R_{sync} s$

iv). $\alpha((\psi_n \circ w_n)(s), w_n(s)) = \alpha((\sigma^{2k} \circ w_n)(s), s)$

v). $\forall \, f \in L^1(V^{\mathbb{N}})$ *we have* $\|f \circ w_n\|_1 \leq C\|f\|_1$ *and* $\|f \circ \psi_n\|_1 \leq C^2\|f\|_1$

Proof. For ease of notation I will not use the composition symbols in this proof. The first two items are immediate from the definitions of w_n and ψ_n. For item iii). note that the tails of $w_n(s)$ and s are the same therefore $w_n(s) R_{sync} s$ and since $\psi_n w_n(s)$ has the same tails as $w_n(s)$, without needing to shift, we have $\psi_n w_n(s) R_{sync} s$. Therefore $\psi_n w_n(w_n(s)) R_{sync} w_n(s)$, and as w_n has itself as an inverse we conclude that $\psi_n(s) R_{sync} s$.

For part iv). fix $n > 2k+1$ and without loss of generality suppose $s_{[\tau_n(s), \tau_n(s)+k+1]} = upw$. Let $N = \tau_n(s)$. We have $\psi_n w_n(s)_i = w_n(s)_i \, \forall \, i > N$, therefore

$$\begin{aligned}
&\alpha(\psi_n w_n(s), w_n(s)) \\
&= \mathcal{L}(\psi_n w_n(s)_1) \cdots \mathcal{L}(\psi_n w_n(s)_N) \cdot \Big(\mathcal{L}(w_n(s)_1) \cdots \mathcal{L}(w_n(s)_N)\Big)^{-1} \\
&= \mathcal{L}(s_{2k+1}) \cdots \mathcal{L}(s_N) \mathcal{L}(q_1) \cdots \mathcal{L}(q_k) \mathcal{L}(p_k)^{-1} \cdots \mathcal{L}(p_1)^{-1} \Big(\mathcal{L}(s_1) \cdots \mathcal{L}(s_N)\Big)^{-1}
\end{aligned}$$

Since $(\sigma^{2k} w_n(s), s)_{i-2k} = w_n(s)_i = s_i \, \forall \, i > N+k$ we have

$$\begin{aligned}
\alpha(\sigma^{2k} w_n(s), s) &= \mathcal{L}(\sigma^{2k} w_n(s)_1) \cdots \mathcal{L}(\sigma^{2k} w_n(s)_{N-k}) \Big(\mathcal{L}(s_1) \cdots \mathcal{L}(s_{N+k})\Big)^{-1} \\
&= \mathcal{L}(w_n(s)_{2k+1}) \cdots \mathcal{L}(w_n(s)_{N+k}) \Big(\mathcal{L}(s_1)) \cdots \mathcal{L}(s_{N+k})\Big)^{-1} \\
&= \mathcal{L}(s_{2k+1}) \cdots \mathcal{L}(s_N) \mathcal{L}(q_1) \cdots \mathcal{L}(q_k) \Big(\mathcal{L}(s_1) \cdots \mathcal{L}(s_N) \mathcal{L}(p_1) \cdots \mathcal{L}(p_k)\Big)^{-1} \\
&= \alpha(\psi_n w_n(s), w_n(s))
\end{aligned}$$

For part v). we use the C chosen in the passage preceding the statement of the proposition to get that for every $f \in L^1(V^{\mathbb{N}})$ we have

$$\|f \circ w_n\|_1 \leq C\|f\|_1, \qquad \|f \circ \psi_n w_n\| \leq C\|f\|_1$$

and since w_n is self-inverse we have

$$\|f \circ \psi_n\|_1 = \Big\|f \circ \psi_n w_n \circ w_n^{-1}\Big\|_1 \leq C\|f \circ \psi_n w_n\|_1 \leq C^2\|f\|_1$$

□

Proposition 3.11. *For $n > 2k+1$ there exist measurable maps $\Phi_n, \Psi_n, \Omega_n : V^{\mathbb{N}} \times X \to V^{\mathbb{N}} \times X$ such that*

i). $\forall f \in L^1(V^{\mathbb{N}} \times X)$, $\lim_{n\to\infty} \left\| f \circ \Psi_n \circ \Omega_n - f \circ \sigma_X^{2k} \circ \Phi_n \right\|_1 = 0$;

ii). $\forall f \in L^1(V^{\mathbb{N}} \times X)$, $\lim_{n\to\infty} \left\| f \circ \Omega_n - f \right\|_1 = 0$;

iii). $\forall s \in V^{\mathbb{N}}$ *we have* $\Phi_n(s) R^X_{sync} s$ *and* $\Psi_n(s) R^X_{sync} s$.

Proof. For $n > 2k+1$, $n \in \mathbb{N}$ let ψ_n and w_n be as in Proposition 3.10. Define the following functions

$$\begin{aligned}
\Omega_n(s,x) &:= (w_n s, x) \\
\Phi_n(s,x) &:= (w_n s, \alpha(w_n s, s)x) \\
\Psi_n(s,x) &:= (\psi_n s, \alpha(\psi_n s, s)x)
\end{aligned}$$

From Proposition 3.10 iv). we know that the graphs of ψ_n and w_n satisfy R_{sync}. Therefore the definition of Φ_n and Ψ_n means their graphs satisfy R^X_{sync}. Let d_X be a metric on X the induces its Borel structure and makes X into a compact space. Define $d_*((s,x),(s',x')) = d_X(x,x') + d_{V^{\mathbb{N}}}(s,s')$. By Proposition 3.10 ii). and iii). we have

$$d_*(\Omega(s,x),(s,x)) = d_{V^{\mathbb{N}}}(w_n s, s) \leq \frac{1}{\tau_n(s)} \leq \frac{1}{n}$$

and

$$\begin{aligned}
\Psi_n \Omega_n(s,x) &= (\psi_n w_n s, \alpha(\psi_n w_n s, w_n s)x) \\
\sigma_X^{2k} \Phi_n(s,x) &= \sigma_X^{2k}(w_n s, \alpha(w_n s, s)x) = (\sigma^{2k} w_n s, \alpha(\sigma^{2k} w_n s, w_n s)\alpha(w_n s, s)x) \\
&= (\sigma^{2k} w_n s, \alpha(\sigma^{2k} w_n s, s)x) = (\sigma^{2k} w_n s, \alpha(\psi_n w_n s, w_n s)x)
\end{aligned}$$

Therefore Proposition 3.10 i). gives us

$$d_*((\Psi_n \circ \Omega_n)(s,x), (\sigma_X^{2k} \circ \Phi_n)(s,x)) \leq \frac{1}{n-k}$$

Now take any continuous function f on $V^{\mathbb{N}} \times X$. Then as our metric makes X compact we have that f is uniformly continuous. This means that for any given $\varepsilon > 0$ there exists a $N \in \mathbb{N}$ such that $n > N$ implies $|(f \circ \Psi_n \circ \Omega_n)(s,x) - (f \circ \sigma_X^{2k} \circ \Phi_n)(s,x)| < \varepsilon$ for every $(s,x) \in V^{\mathbb{N}} \times X$, therefore as we are in a probability space we have

$$\lim_{n\to\infty} \left\| f \circ \Psi_n \circ \Omega_n - f \circ \sigma_X^{2k} \circ \Phi_n \right\|_1 = 0$$

Similarly we have

$$\lim_{n\to\infty} \left\| f \circ \Omega_n - f \right\|_1 = 0$$

By item v). of Proposition 3.10 we have that the operators $f \mapsto f \circ \Omega_n$, $f \mapsto f \circ \Phi_n$ and $f \mapsto f \circ \Psi_n$ are all bounded for $f \in L^1(V^{\mathbb{N}} \times X)$, with the bound being independant of n. Since V is finite and $\tilde{v}$ is the Markov measure we also have that $f \mapsto f \circ \sigma_X^{2k}$ is bounded. As the continuous functions are dense in $L^1(V^{\mathbb{N}} \times X)$ by a simple triangle inequality argument we are done. □

We now prove Theorem 3.9. As previously noted if suffices to show that if $f \in L^1(V^{\mathbb{N}} \times X)$ is R^X_{sync} then $f \circ \sigma_X^{2k} = f$. Let Φ_n, Ψ_n, Ω_n be as in Proposition 3.11. We know that the graph of Ψ_n is contained in R^X_{sync}, therefore $f \circ \Psi_n = f$. If $(s,x) R^X_{sync} (t,y)$ then $\sigma_X(s,x) R^X_{sync} \sigma_X(t,y)$, therefore we have that $f \circ \sigma_X^{2k}$ is R^X_{sync} invariant. Additionally since the graph of Φ_n is contained in R^X_{sync} we have $f \circ \sigma_X^{2k} \circ \Phi_n = f \circ \sigma_X^{2k}$.

Putting this all together we may make the bound:

$$\begin{aligned}
\left\| f - f \circ \sigma_X^{2k} \right\|_1 &= \left\| f - f \circ \sigma_X^{2k} \circ \Phi_n \right\|_1 \\
&\leq \left\| f - f \circ \Psi_n \circ \Omega_n \right\|_1 + \left\| f \circ \Psi_n \circ \Omega_n - f \circ \sigma_X^{2k} \circ \Phi_n \right\|_1 \\
&= \left\| f - f \circ \Omega_n \right\|_1 + \left\| f \circ \Psi_n \circ \Omega_n - f \circ \sigma_X^{2k} \circ \Phi_n \right\|_1
\end{aligned}$$

Taking $n \to \infty$ gives $f = f \circ \sigma_X^{2k}$.

3.6 Proof of Theorem 3.6

Recall that $(s,x) R^X_n (t,y)$ implies $(s,x) R^X_1 (t,y)$ for all $n \geq 1$. We have defined $\mathcal{F}^R_n$ to be the σ-algebra formed by unions of equivalence classes of R^X_n. Take any equivalence class of R^X_n, then it is also an equivalence class of R^X_1 so $\mathcal{F}^R_1 \subset \mathcal{F}^R_n$. To see this take any equivalence class of R^X_1, A say, and $(s,x) \in A$. Let $A^n_{s,x}$ be the equivalence class of (s,x) with respect to R^X_n. Then $A \subset \bigcup_{(s,x)\in A} A^n_{s,x}$ is clear and the reverse inclusion follows since for any $(t,y) \in A^n_{s,x}$ we have $(s,x) R^X_n (t,y)$ hence $(s,x) R^X_1 (t,y)$ so $(t,y) \in A$.

Lemma 3.12. *$\mathcal{F}^R_n$ is the σ-algebra of σ^n_X-invariant measurable subsets of $V^{\mathbb{N}} \times X$*

Proof. I will show the case for $n = 1$, all other cases are anaolgous. Take any equivalence class of R^X_1 and call it A. To show $A \subset \sigma_X(A)$ take any $(s,x) \in A$, $u \in V$ and defined $s' = (u, s_1, s_2, \ldots)$ and $x' = \mathcal{L}(u)x$. I claim $(s,x) = \sigma_X(s',x')$ and $(s',x') \in A$. The first of these is immediate. For the second take any $(t,y) \in A$, then there exists $l \in \mathbb{Z}, N \in \mathbb{N}$ (depending on (t,y)) such that $\sigma_X^{l+N}(s,x) = \sigma_X^N(t,y)$ and since $\sigma_X(s',x') = (s,x)$ we get $\sigma_X^{l+1+N}(s',x') = \sigma_X^{l+N}(s,x) = \sigma_X^N(t,y)$ as required.

For $\sigma_X(A) \subset A$ take $(s',x') = \sigma_X(s,x) \in \sigma_X(A)$. Repeating the above argument we see that $\sigma_X^{l-1+N}(s',x') = \sigma_X^{l+N}(s,x) = (t,y)$. $\square$

Therefore we can use von Neumann's mean ergodic theorem to get that for every $f \in L^1(V^{\mathbb{N}} \times X)$

$$\frac{1}{nk}\sum_{i=0}^{nk-1} f \circ \sigma_X^i \to \mathbb{E}[f|\mathcal{F}_1^R]$$

in L^1 as $n \to \infty$. We can now use the tower rule with $\mathcal{F}_1^R \subset \mathcal{F}_k^R$ to get

$$\frac{1}{nk}\sum_{i=0}^{nk-1} \mathbb{E}[f \circ \sigma_X^i|\mathcal{F}_k^R] \to \mathbb{E}[f|\mathcal{F}_1^R]$$

Using Lemma 3.12, $\mathbb{E}[f \circ \sigma_X^{k+i}|\mathcal{F}_k^R] = \mathbb{E}[f \circ \sigma_X^i|\mathcal{F}_k^R]$ for any i. So for any n we get

$$\frac{1}{nk}\sum_{i=0}^{nk-1} \mathbb{E}[f \circ \sigma_X^i|\mathcal{F}_k^R] = \frac{1}{k}\sum_{i=0}^{k-1} \mathbb{E}[f \circ \sigma_X^i|\mathcal{F}_k^R]$$

This has proven the following Lemma.

Lemma 3.13. *For any $f \in L^1(V^{\mathbb{N}} \times X)$ and any $k \in \mathbb{N}$,*

$$\frac{1}{k}\sum_{i=0}^{k-1} \mathbb{E}[f \circ \sigma_X^i|\mathcal{F}_k^R] = \mathbb{E}[f|\mathcal{F}_1^R]$$

Given that $\mathcal{F}_1^R$ is the σ-algebra of σ_X-invariant subsets, if we can show that σ_X is ergodic then we will have that $\mathcal{F}_1^R$ is trivial. The function σ_X is very similar to the function in Kakutani's random ergodic theorem [10], so we can try to use that to show σ_X is ergodic.

Definition 3.14. *Let $(S,\mathcal{B},m)$ and $(Y,\mathcal{F},\mu)$ be probability spaces.*

i). A one to one mapping φ of S onto itself is called a $\mathcal{B}$-m-measure preserving transformation if $B \in \mathcal{B}$ implies $\varphi(B) \in \mathcal{B}$, $\varphi^{-1}(B) \in \mathcal{B}$ and $m(\varphi(B)) = m(\varphi^{-1}(B) = m(B)$.

ii). Let $\Phi = \{\varphi_y \mid y \in Y\}$ be a family of $\mathcal{B}$-m-measure preserving transformations. Φ is called $(\mathcal{B},\mathcal{F})$-measurable if $B \in \mathcal{B}$ implies $\{(s,y) \mid \varphi_y(s) \in B\} \in \mathcal{B} \times \mathcal{F}$.

iii). $B \in \mathcal{B}$ is called Φ-invariant if $m(\varphi_y(B)\Delta B) = 0$ for μ-almost every $y \in Y$. Φ is called ergodic if every Φ-invariant set $B \in \mathcal{B}$ has measure 0 or 1.

Theorem 3.15 (Kakutani's random erogdic theorem)**.** *Let $(S,\mathcal{B},m)$ and $(Y,\mathcal{F},\mu)$ be two probability spaces. Let $\Phi = \{\varphi_y \mid y \in Y\}$ be a $(\mathcal{B},\mathcal{F})$-measurable family of $\mathcal{B}$-m-measure preserving transformations φ_y defined on S. Then the following are equivalent:*

i). $\Phi = \{\varphi_y \mid y \in Y\}$ is ergodic

ii). $\varphi^* : S \times Y^{\mathbb{N}} \to S \times Y^{\mathbb{Z}}$ defined by $\varphi^*(s,w) = (\sigma(w), \varphi_{w_0}(s))$ is ergodic.

Proposition 3.16. *Let $\Pi, V, \mathcal{L}$ be as in the statement of Theorem 3.6 with $v \in V$ such that $\Gamma_v = \mathbb{F}_r$. Then $\mathcal{F}_1^R$ is the σ-algebra generated by all sets of the form $V^{\mathbb{N}} \times A$ where $A \subset X$ is a measurable $\mathbb{F}_r$-invariant set. In particular if the action of $\mathbb{F}_r$ on X is ergodic then $\mathcal{F}_1^R$ is trivial.*

Proof. By decomposing into erogdic components we may assume that $\mathbb{F}_r$ acts ergodically. As previous discussed it suffices to show that σ_X is ergodic since $\mathcal{F}_1^R$ is generated by σ_X.

The main idea is that we need to use Kakutani's random ergodic theorem, but in the statement of that theorem the family of transformations is indexed by a measure space whereas group actions are indexed by elements in the group. $\mathcal{L}$ is used to map from elements of V into the group but we don't want to put assumptions on the way $\mathcal{L}$ does this mapping.

The idea is to use the way that almost every element of $V^{\mathbb{N}}$ returns to v infinitely often, so we may view an element of $V^{\mathbb{N}}$ as a sequence of paths from v to a node adjacent to v, rather than a sequence of elements of V. This is where the assumptions of $\Gamma_V = \mathbb{F}_r$ is used.

In the statement of Kakutani's random ergodic theorem let $S = X$, $Y = V_{pv}$ be the set of paths p in G such that pv is a loop and $\Phi = \{\mathcal{L}(p)^{-1} \mid p \in V_{pv}\}$. Then any set that is Φ invariant is going to be invariant under any finite combination of $\mathcal{L}(p)$ and $\mathcal{L}(p)^{-1}$ for all $p \in V_{pv}$. But we have assumed that the $\mathcal{L}(p)$ generate the free group, meaning that a Φ-invariant set would be $\mathbb{F}_r$-invariant. We have assumed the action of $\mathbb{F}_r$ is ergodic, therefore the set would have measure 0 or 1 meaning that Φ is ergodic.

Now we need to phrase the function in the second part of Kakutani's theorem in terms of $V^{\mathbb{N}}$ instead of $V_{pv}^{\mathbb{N}}$. This is done by defining $Z \subset V^{\mathbb{N}} \times X$ to be the set of (s,x) such that $s_1 = v$ and letting $T : Z \to Z$ be defined as $T(s,x) = \sigma_X^n(s,x)$ where $n \geq 1$ is the smallest natural such that $\sigma_X^n(s,x) \in Z$.

To see how this is equivalent to the function in the theorem, take any $(s,x) \in Z$ with $s_i = v$ for infinitely many i and suppose $n_1 \geq 1$ is the smallest natural such that $s_{n_1} = v$, then $T(s,x) = (\sigma^{n_1}(s), (\mathcal{L}(s_1) \cdots \mathcal{L}(s_{n_1}))^{-1}x)$. (s,x) can be identified with an element $(\tilde{s},x) \in V_{pv}^{\mathbb{N}} \times X$ where $\tilde{s}_1 = (s_1, \ldots, s_{n_1})$ and $\tilde{s}_2 = (s_{n_1+1}, \ldots, s_{n_2})$ with n_2 the second smallest natural such that $s_{n_2} = v$. Then

$$\varphi^*(\tilde{s},x) = (\sigma(\tilde{s}), \mathcal{L}(\tilde{s}_1)) = ((p^2, p^3, \ldots), (\mathcal{L}(s_1) \cdots \mathcal{L}(s_{n_1}))^{-1}x)$$

If we unpack $(p^2, p^3, \ldots)$ into a sequence of elements we see that it is equal to $\sigma^{n_1}(s)$ as required. Therefore we have that T is ergodic.

To show that σ_X is ergodic take any $W \subset V^{\mathbb{N}} \times X$ that is measurable, σ_X-invariant and has positive measure. Then $Z \cap W$ is T-invariant (W is σ_X-invariant and Z ensures after shifting the first element of the sequence term will be v). Also we have that $(\tilde{v} \times \mu)(Z \cap$

$W) > 0$. To see this suppose otherwise and let $W_v^n \subset W$ be the set $\{(s,x) \in W \mid s_n = v\}$. Then $\sigma_X^{n-1}(W_v^n) \subset W_v^1$ and because $W_v^1 \subset W$, with σ_X preserving the measure, we get that $(\tilde{v} \times \mu)(W_v^n) = 0$ for all $n \geq 1$. We can decompose $W = I \dot{\cup} N$ where I is the subset of W of elements (s,x) where v occurs in s infinitely often, and N the subset where v occurs finitely often. As mentioned before we have $(\tilde{v} \times \mu)(N) = 0$, therefore $(\tilde{v} \times \mu)(I) > 0$. Let $I' = \cup_{n=1}^{\infty} W_v^n$, the subset of W of elements where v occurs at least once in the sequence term. Then $I \subset I'$ but $(\tilde{v} \times \mu)(I') \leq \sum_{n=1}^{\infty} (\tilde{v} \times \mu)(W_v^n) = 0$ which gives $(\tilde{v} \times \mu)(I) = 0$, a contradiction.

So as T is ergodic we have that $W \cap Z = Z$ up to measure 0. However $\cup_{i=0}^{\infty} \sigma_X^i(Z) = V^{\mathbb{N}} \times X$ due to strong connectivity of G. Using $W \cap Z = Z \subset W$ we have the following

$$V^{\mathbb{N}} \times X = \cup_{i=0}^{\infty} \sigma_X^i(Z) \subset \cup_{i=0}^{\infty} \sigma_X^i(W) = \cup_{i=0}^{\infty} W = W$$

therefore W is conull and σ_X is ergodic. □

We are now ready to derive the final results. First recall that by the reverse martingale convergence theorem

$$(P^n \varphi)(\pi \circ \sigma_X^n) = \mathbb{E}[\varphi \circ \pi \mid \mathcal{F}_n^E] \to \mathbb{E}[\varphi \circ \pi \mid \mathcal{F}_{sync}]$$

in $L^1(V^{\mathbb{N}} \times X)$. Using Theorem 3.9 we get

$$(P^n \varphi)(\pi \circ \sigma_X^n) \to \mathbb{E}[\varphi \circ \pi \mid \mathcal{F}_{2k}^R]$$

in $L^1(V^{\mathbb{N}} \times X)$. Conditioning on $\mathcal{F}_{2k}^R$ commutes with σ_X^i for any $i \geq 0$ because σ_X does not change equivalence classes of R_{2k}^X. Therefore for $i \geq 0$

$$(P^n \varphi)(\pi \circ \sigma_X^{n+i}) \to \mathbb{E}[\varphi \circ \pi \circ \sigma_X^i \mid \mathcal{F}_{2k}^R]$$

Since $\mathbb{E}[\varphi \circ \pi \circ \sigma_X^i \mid \mathcal{F}_{2k}^R] = \mathbb{E}[\varphi \circ \pi \circ \sigma_X^{2k+1} \mid \mathcal{F}_{2k}^R]$ we can write for any $0 \leq i < 2k$

$$(P^n \varphi)(\pi \circ \sigma_X^{n-i}) \to \mathbb{E}[\varphi \circ \pi \circ \sigma_X^{2k-i} \mid \mathcal{F}_{2k}^R]$$

in $L^1(V^{\mathbb{N}} \times X)$. Lemma 3.13 gives

$$\frac{1}{2k} \sum_{i=0}^{2k-1} (P^n \varphi)(\pi \circ \sigma_X^{n-i}) \to \frac{1}{2k} \sum_{i=0}^{2k-1} \mathbb{E}[\varphi \circ \sigma_X^i \mid \mathcal{F}_{2k}^R] = \mathbb{E}[\varphi \circ \pi \mid \mathcal{F}_1^R]$$

and under the assumption that the action is ergodic, using Proposition 3.16 gives

$$\frac{1}{2k} \sum_{i=0}^{2k-1} (P^n \varphi)(\pi \circ \sigma_X^{n-i}) \to \mathbb{E}[\varphi \circ \pi \mid \mathcal{F}_1^R] = \int \varphi \, \mathrm{d}(v \times \mu)$$

in $L^1(V^{\mathbb{N}} \times X)$. Note that we can split up P^n into $P^{n-i}P^i$ to get

$$(P^n\varphi)(\pi \circ \sigma_X^{n-i}) = (P^{n-i}(P^i(\varphi))(\pi \circ \sigma_X^{n-i}) \to \mathbb{E}[(P^i\varphi) \circ \pi \mid \mathcal{F}_{2k}^R]$$

in $L^1(V^{\mathbb{N}} \times X)$. Similarly,

$$(P^{n+i}\varphi)(\pi \circ \sigma_X^n) = (P^n(P^i(\varphi))(\pi \circ \sigma_X^n) \to \mathbb{E}[(P^i\varphi) \circ \pi \mid \mathcal{F}_{2k}^R]$$

in $L^1(V^{\mathbb{N}} \times X)$. Using this we see

$$\frac{1}{2k}\sum_{i=0}^{2k-1}(P^{n+i}\varphi)(\pi \circ \sigma_X^n) \to \int \varphi \, \mathrm{d}(v \times \mu)$$

in $L^1(V^{\mathbb{N}} \times X)$. Due to linearity of P and the integral we may assume without loss of generality that $\int \varphi \, \mathrm{d}(v \times \mu) = 0$, so we have

$$\left\| \frac{1}{2k}\sum_{i=0}^{2k-1}(P^{n+i}\varphi)(\pi \circ \sigma_X^n) \right\|_{L^1(V^{\mathbb{N}} \times X)} \to 0$$

which means

$$\left\| \frac{1}{2k}\sum_{i=0}^{2k-1} P^{n+i}\varphi \right\|_{L^1(V \times X)} \to 0$$

Therefore if $\varphi(v,x) = \phi(x)$ for some $\phi \in L^1(X)$ then by the previously outlined change of variable argument we have $(S_n\phi)(x) = \sum_{u \in V} p_u (P^n\varphi)(u,x)$. This gives us the following.

$$\left\| \frac{1}{2k}\sum_{i=0}^{2k-1} S_{n+i}\phi \right\|_{L^1(X)} = \left\| \frac{1}{2k}\sum_{i=0}^{2k-1}\sum_{u \in V} p_u P^{n+i}\varphi \right\|_{L^1(V \times X)} \leq \left\| \frac{1}{2k}\sum_{i=0}^{2k-1} P^{n+i}\varphi \right\|_{L^1(V \times X)} \to 0$$

which completes the proof of Theorem 3.6.

4 Fundamental group of the surface of genus two

4.1 Markov Coding

I shall be following the example in [1] (Section 1.2) of the regular hyperbolic octagon with internal angles $\frac{\pi}{4}$ in the Poincaré disc, $\mathbb{D}$. As explained in [17] (Chapter 19) this hyperbolic octagon, call it A, is a fundamental domain for a group action of $\Lambda = \langle a,b,c,d | [a,b][c,d] = e \rangle$ acting on $\mathbb{D}$ by side-pairing isometries. In particular if we let $\Lambda_g = \{a,b,c,d,a^{-1},b^{-1},c^{-1},d^{-1}\}$ then we can arrange the group action such that for every $s \in \Lambda_g$, $sA \cap A$ is a side of A.

This group action will make A tessellate the plane, a good way to visualise this is using

the animation that can be found at [8]. Let $\mathcal{T} = \cup_{g \in \Lambda} g\partial A$ be the union of the boundaries of the actions of Λ on A. Let $\mathcal{P} \subset \mathbb{D}$ be the end points of the geodesics in $\mathcal{T}$ that meet A, indeed we will only obtain end points from geodesics which only meet A at a vertex. These points partition $\partial\mathbb{D}$ into a collection of intervals which I shall denote $\mathcal{I}$. As can be seen from the animation in [8] each vertex has six corresponding points and an image of A and these intervals is given in [1] where the hyperbolic octagon has been drawn as a circle.

I will now outline the Markov coding in [15] and [5]. For any $s \in \Lambda_g$ we know that sA forms a sort of clam shell shape with an edge in common with A. This edge can be extended as a geodesic which cuts $\mathbb{D}$ into two half-spaces. Let $L(s)$ be the open arc on $\partial\mathbb{D}$ which bounds the half-space that contains sA.

For each $I \in \mathcal{I}$ choose $s_I \in \Lambda_g$ such that $I \subset L(s_I)$. For each I there are either one of two choices for s_I. In particular it is only the eight intervals which occur in the overlaps of the $L(s)$ for $s \in \Lambda_g$. These are the intervals which are the middle intervals of the five intervals formed from the six rays coming from each vertex of A. Now define $f : \partial\mathbb{D} \to \mathbb{D}$ by $f(x) = s_I x$ for $x \in I$.

Proposition 4.1. *Take $I, J \in \mathcal{I}$ and suppose that $f(I) \cap J \neq \emptyset$ then $f(I) \supset J$.*

Proof. Due to the properties of the group action we know that an interval $I \in \mathcal{I}$ will be mapped to an interval under f and the end points of I will be mapped to the end points of $f(I)$. Therefore if we can show that $f(\mathcal{P}) \subset \mathcal{P}$ then we can conclude that $f(I)$ is a union of intervals in $\mathcal{I}$, from which the result follows. A proof of $f(\mathcal{P}) \subset \mathcal{P}$ can be found in [5]. □

4.2 Applying Theorem 3.6

Let $V = \mathcal{I}$, $E = \{(I, J) \in V \times V : f(I) \subset J\}$ and $G = (V, E)$ be the graph we shall be applying the theorem to. Let $\mathbb{F}_4$ be the free group on four generators and our labelling map $\mathcal{L} : V \to \mathbb{F}_4$ be the map $\mathcal{L}(I) = s_I$.

The following is Theorem 1.5 in [1], most steps are obtained from other references but I shall mainly explain involvement of f in the proof.

Theorem 4.2. *Given the current set up, if $(\Pi)_{v,w \in V}$ is any stochastic matrix with $\Pi_{v,w} > 0$ for all $(w, v) \in E$ then it is admissable of order one.*

Proof. First recall the definition of admissable, Definition 3.4. The fact that G is strongly connected comes from the way that Π is irreducible which is proved in [5].

To see that there is a $v \in V$ such that $\Gamma_v = \mathbb{F}_4$ we can analyse the behaviour of the map f on a subset of $\mathcal{I}$. For every $s \in \Lambda_g$ let I_s be the central interval associated with the side

transform s. See the diagram in [1] for a visualisation. Additionally these I_s intervals are completely analogous to the intervals $A(e)$ as described in [5].

For these intervals we have that for every $s,t \in \mathcal{I}$, $(I_s, I_t) \in E$ if an only if $t \neq s^{-1}$ and I_t is not adjacent to $I_{s^{-1}}$. To see this we will use a limit set argument. Take any point $x \in I_s$. Then x can be expressed as an infinite reduced word made of letters from Λ_g and it must start with s, indeed the only points in $\mathbb{D}$ whose infinite word representations start with s are in $L(s)$. Then as $x \in I_s \subset L(s)$ we have that $f(x) = s^{-1}x$. All points in $I_{s^{-1}}$ have infinite words starting with s^{-1} so $f(x) \notin I_{s^{-1}}$. To see why this is the case first write $x = sx'$ for some infinite word x' which cannot start with s^{-1} since x was a reduced word. Since $f(x) = s^{-1}x = s^{-1}sx' = x'$ and we can therefore conclude that $f(x) \notin I_{s^{-1}}$.

As for the adjacency statement, this follows from the way that the I_s are the central intervals associated with each $s \in \Lambda_g$. More precisely we know that for a given $s \in \Lambda_g$ the hyperbolic octagon is mapped via s in a way that pairs a side. This transformation preserves the ordering of the sides too. For the rest of the argument to make visualization easier I will, without loss of generality, set $s = b$ and again use the diagram in [1]. After the action of b, if the action of either adjacent side i.e. c or a^{-1} is done then the clam shell will be mapped to an octagon which shares a side with the clam shell, whose contents will either be to the left or to the right of the the clam shell. The animation can again help with visualisation. Both of these new octagons each have an edge which is not shared by the clam shell but is incident to a vertex of A. When the geodesics that these edge lie on are extended to the boundary they will meet at the points on the boundary which will give the end points of the intervals that are adjacent to I_b.

So the mapping of bc or ba^{-1} will result in any point on the boundary being mapped into one of the intervals immediately adjacent to I_b, vice versa any point on the boundary whose infinite reduced word starts with bc or ba^{-1} must be in one of the two intervals immediately adjacent to I_b. So given an $x = bx' \in I_b$ we know that $f(x) = b^{-1}x = b^{-1}bx' = x'$ does not start with a or c^{-1} and so cannot be in I_a or $I_{c^{-1}}$. The reason that for every $s \in \Lambda_g \setminus \{c, a^{-1}\}$ are points in I_b whose words start with bs is that the clam shell will get mapped to another octagon sharing an egde with the clam shell but within the same boundary formed by extending the geodesics of the two edges of the clam shell that are paired under the action of a and c^{-1} on the clam shell. Indeed, under the actions of a^{-1} or c the clam shell is mapped to another octagon outside of the inner most rays.

So given this characterisation of edges in G we may identify some loops in G. Taking $v = a$ we can see there is a loop from I_a to itself and a path to I_c, $I_{c^{-1}}$, I_d and $I_{d^{-1}}$. which means Γ_v contains $\mathcal{L}(I_a) = a$, $\mathcal{L}(I_a, I_c) = ac$, $\mathcal{L}(I_a, I_d, I_c) = adc$, $\mathcal{L}(I_a, I_d, I_b) = adb$ which together generate $\mathbb{F}_4$ i.e. $\Gamma_v = \mathbb{F}_4$.

To complete the proof it only remains to show that G contains a good subgraph of order one. To see this let $u = w = I_a$, $p = (I_a)$, $q = (I_c)$, $p^* = (I_{a^{-1}})$, $q^* = (I_{c^{-1}})$. Then it is

easy to see that upw, uqw, pq^*p, qp^*q are all directed paths in G and $\mathcal{L}(p^*) = \mathcal{L}(p)^{-1}$ and $\mathcal{L}(q^*) = \mathcal{L}(q)^{-1}$. $\square$

This now gives us the following result.

Theorem 4.3. *Suppose Λ acts on a probability space (X, μ) with measure-preserving actions then with $G = (V, E)$ as outlined above and $(\Pi)_{v,w \in V}$ a stochastic matrix such that$\Pi_{v,w} > 0$ for all $(w, v) \in E$ with positive stationary distribution p. Given $f \in L^1(X, \mu)$ we have*

$$S_n f(x) = \sum_{g \in V^n} p_{g_n} \Pi_g (T_g f)(x) \xrightarrow{L^1} \mathbb{E}(f|\Lambda_I)$$

Where Λ_I is the σ-algebra of Λ-invariant subsets.

References

[1] L. Bowen, A. Bufetov, O. Romaskevich. *Mean convergence of Markovian spherical averages for measure-preserving actions of the free group*, Geom Dedicata, Volume 181, pages 293-306, 2016

[2] A. I. Bufetov *Operator ergodic theorems for actions of free semigroups and groups*, Functional Analysis and its Applications, Volume 34, pages 239-251, 2000

[3] A. I. Bufetov. *Convergence of spherical averages for actions of free groups*, Annals of Mathematics, Volume 155, pages 929-944, 2002

[4] A. I. Bufetov, A. Klimenko. *On Markov operators and ergodic theorems for group actions*, European Journal of Combinatorics, Volume 33, pages 1427-1443, 2012

[5] A. Bufetov, C. Series. *A pointwise ergodic theorem for Fuchsian groups*, Mathematical Proceedings of the Cambridge Philosophical Society, Volume 151, Issue 1, pages 145-159, 2011

[6] M. Einsiedler, T. Ward. *Ergodic theory with a view towards number theory*, Springer, 2013

[7] R. I. Grigorchuk. *Ergodic theorems for actions of free semigroups and groups*, Mathematical Notes, Volume 65, pages 654-657, 1999

[8] Ryan Hoban, Hyperbolic octagon tiling plane
`http://egl.math.umd.edu/projects/archive/Oct_Pics/fuschian.gif`

[9] V. A. Kaimanovich. *Measure-Theoretic boundaries of Markov chains, 0-2 laws and entropy*, Harmonic Analysis and Discrete Potential Theory, pages 145-181, Springer, 1992

[10] S. Kakutani. *Random ergodic theorems and Markoff processes with a stable distribution*, Proceedings of the Second Berkley Symposium on Mathematical Statistics and Probability, 1950

[11] J. Neveu. *Discrete Parameter Martingales* Oxford University Press, 1975

[12] W. Parry. *Topics in ergodic theory*, Cambridge University Press, 1981

[13] A. Quas. *Ergodicity and mixing properties*, Encyclopedia of Complex Systems, Springer 2009

[14] V. A. Rohlin, Y. G. Sinai. *The structre and properties of invariant measurable partitions*, Doklady Akademii Nauk, Volume 141, pages 1038-1041, 1961

[15] C. Series. *Geometrical methods of symbolic coding*, Ergodic Theory and Symbolic Dynamics in Hyperbolic Space, Oxford University Press 1991

[16] P. Sousi. *Advanced Probability Course Notes*, University of Cambridge

[17] C. Walkden. *Hyperbolic Geometry Lecture Notes*, University of Manchester

[18] A. Zygmund. *Trigonometric Series*, Cambridge University Press, 1968