Math. Proc. Camb. Phil. Soc. (2011), **151**, 145 © *Cambridge Philosophical Society* 2011
doi:10.1017/S0305004111000247
First published online 27 April 2011

A pointwise ergodic theorem for Fuchsian groups

BY ALEXANDER I. BUFETOV

The Steklov Institute of Mathematics, Russian Academy of Sciences
Gubkina Str. 8, 119991*, Moscow, Russia.*
e-mail: aib1@rice.edu

AND CAROLINE SERIES

Mathematics Institute, University of Warwick
Coventry CV 4 *7AL.*
e-mail: C.M.Series@warwick.ac.uk

(*Received* 26 *October* 2010; *revised* 4 *February* 2011)

Abstract

We use Series' Markovian coding for words in Fuchsian groups and the Bowen-Series coding of limit sets to prove an ergodic theorem for Cesàro averages of spherical averages in a Fuchsian group.

1. *Introduction*

1·1. *An ergodic theorem for surface groups*

Let $g \geqslant 2$ and let Γ be the fundamental group of a surface of genus g endowed with a set of generators

$$\{a_1, \ldots, a_{2g}, b_1, \ldots, b_{2g}\} \tag{1·1}$$

satisfying the standard commutator relation

$$\prod_{i=1}^{2g} [a_i, b_i] = 1.$$

For $g \in \Gamma$, let $|g|$ stand for the length of the shortest word in the generators (1·1) representing g. Let

$$S(n) = \{g \in \Gamma : |g| = n\}$$

be the sphere of radius n in Γ, and let K_n be the cardinality of $S(n)$. A special case of the main result of this note is the following pointwise ergodic theorem for Γ:

THEOREM A. *Let Γ act ergodically on a probability space (X, ν) by measure-preserving transformations, and, for $g \in \Gamma$, let T_g be the corresponding transformation. Then for any $\varphi \in L_1(X, \nu)$ we have*

$$\frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{K_n} \sum_{g \in S(n)} \varphi \circ T_g \longrightarrow \int_X \varphi d\nu \tag{1·2}$$

both ν-almost surely and in $L_1(X, \nu)$ as $N \to \infty$.

Our main result, Theorem B, is that a similar ergodic theorem applies much more generally to any finitely generated Fuchsian group with a suitable choice of generating set, see Section 2·5 for a precise statement. For the finite volume cocompact case and functions in L_2, the result is due to Fujiwara and Nevo [**8**, theorem 4]. For numerous examples of actions to which this theorem applies, see [**9**, **14**]. Theorem B is derived from a pointwise ergodic theorem for free semigroups from [**7**] whose formulation we shortly recall. For previous literature on pointwise ergodic theorems for actions of various classes of non-amenable discrete groups, see for example [**4**, **9**, **11**, **12**, **13**, **15**, **16**] and also the comprehensive recent survey on free group actions in [**1**].

1·2. *Ergodic theorems for free semigroups*

Let (X, ν) be a probability space and let $T_1, \ldots, T_m \colon X \to X$ be ν-preserving transformations.

Denote by W_m the set of all finite words in the symbols $1, \ldots, m$:

$$W_m = \{w = w_1 \cdots w_n,\ w_i \in \{1, \ldots, m\}\}.$$

The length of a word w is denoted by $|w|$. For each $w \in W_m$, $w = w_1 \cdots w_n$, define the transformation T_w by the formula

$$T_w = T_{w_1} \cdots T_{w_n}.$$

Now let A be an $m \times m$-matrix with non-negative entries. For each $w \in W_m$, $w = w_1 \cdots w_n$, set

$$A(w) = A_{w_1 w_2} \cdots A_{w_{n-1} w_n}.$$

Now let φ be a measurable function on X and for each $n = 0, 1, \ldots,$ consider the expression

$$s_n^A \varphi = \frac{\sum_{|w|=n} A(w) \varphi \circ T_w}{\sum_{|w|=n} A(w)}, \tag{1·3}$$

where we assume that the denominator does not vanish.

Furthermore, consider Cesàro averages of the "sphere averages" s_n^A and set:

$$c_N^A(\varphi) = \frac{1}{N} \sum_{n=0}^{N-1} s_n^A(\varphi). \tag{1·4}$$

Definition 1. A matrix A with non-negative entries is called irreducible if for some $n > 0$ all entries of the matrix $A + A^2 + \cdots + A^n$ are positive.

Definition 2. A matrix A with non-negative entries is called strictly irreducible if A is irreducible and AA^T is irreducible (here A^T stands for the transpose of A).

A measurable subset $Y \subset X$ will be called $T_1, \ldots, T_m$- invariant if its characteristic function χ_Y satisfies $T_1 \chi_Y = \cdots = T_m \chi_Y = \chi_Y$. Denote by $\mathcal{B}$ the σ-algebra of all $T_1, \ldots, T_m$- invariant subsets of X. Given $\varphi \in L_1(X, \nu)$, denote by $\mathbb{E}(\varphi|\mathcal{B})$ the conditional expectation of φ with respect to $\mathcal{B}$.

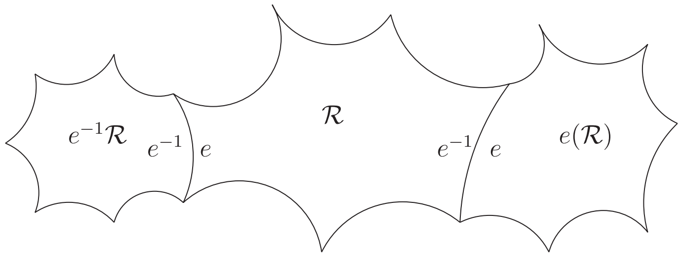

Fig. 1. Labelling the sides of the fundamental domain $\mathcal{R}$. The label e appears interior to $\mathcal{R}$ on the side of adjacent to the region $e^{-1}\mathcal{R}$.

We now recall [**6**, corollary 2]:

PROPOSITION 1. *Let A be a strictly irreducible* $m \times m$ *matrix. Let* $T_1, \ldots, T_m$ *be measure-preserving self-maps of a probability space* (X, ν). *Let* $\mathcal{B}$ *be the* σ*-algebra of* $T_1, \ldots, T_m$*-invariant measurable sets. Then for each* $\varphi \in L_1(X, \nu)$ *we have* $c_N^A(T)\varphi \to \mathbb{E}(\varphi|\mathcal{B})$ *almost everywhere and in* $L_1(X, \nu)$ *as* $N \to \infty$.

Theorems A and B will be derived from Proposition 1 using the Markov coding for Fuchsian groups introduced in [**5**], see also [**17**, **20**].

2. *Markov maps for Fuchsian groups*

Let Γ be a finitely generated non-elementary Fuchsian group acting in the hyperbolic disk $\mathbb{D}$. The Markov coding, originally introduced in [**5**] to encode limit points of Γ as infinite words in a fixed set of generators, also gives a canonical shortest form for words in Γ. The coding is defined relative to a fixed choice of fundamental domain $\mathcal{R}$ for Γ, which we suppose is a finite sided convex polygon with vertices contained in $\mathbb{D} \cup \partial\mathbb{D}$, such that the interior angle at each vertex is strictly less than π. By a *side* of $\mathcal{R}$ we mean the closure in $\mathbb{D}$ of the geodesic arc joining a pair of adjacent vertices. We allow the infinite area case in which some adjacent vertices on $\partial\mathbb{D}$ are joined by an arc contained in $\partial\mathbb{D}$; we do not count these arcs as sides of $\mathcal{R}$. We assume that the sides of $\mathcal{R}$ are paired; that is, for each side s of $\mathcal{R}$ there is a (unique) element $e \in \Gamma$ such that $e(s)$ is also a side of $\mathcal{R}$ and such that $\mathcal{R}$ and $e(\mathcal{R})$ are adjacent along $e(s)$. (Notice that this includes the possibility that $e(s) = s$, in which case e is elliptic of order 2 and the side s contains the fixed point of e in its interior. The condition that the vertex angle is strictly less than π excludes the possibility that the fixed point of e is counted as a vertex of $\mathcal{R}$.)

We denote by $\partial\mathcal{R}$ the union of the sides of $\mathcal{R}$, in other words, $\partial\mathcal{R}$ is the part of the boundary of $\mathcal{R}$ inside the disk $\mathbb{D}$. Each side of $\partial\mathcal{R}$ is assigned two labels, one interior to $\mathcal{R}$ and one exterior, in such a way that the interior and exterior labels are mutually inverse elements of Γ. We label the side $s \subset \partial R$ interior to $\mathcal{R}$ by e if e carries s to another side $e(s)$ of $\mathcal{R}$, while we label the same side exterior to $\mathcal{R}$ by e^{-1}, see Figure 1. With this convention, $\mathcal{R}$ and $e^{-1}(\mathcal{R})$ are adjacent along the side whose *interior* label is e, while the side $e(s)$ has interior label e^{-1}.

Let Γ_0 be the set of labels of sides of $\mathcal{R}$. The labelling extends to a Γ-invariant labelling of all sides of the tessellation $\mathcal{T}$ of $\mathbb{D}$ by images of $\mathcal{R}$. The conventions have been chosen in such a way that if two regions $g\mathcal{R}$, $h\mathcal{R}$ are adjacent along a common side s, then $h^{-1}g \in \Gamma_0$ and the label on s interior to $g\mathcal{R}$ is $h^{-1}g$, while that on the side interior to $h\mathcal{R}$ is $g^{-1}h$. Suppose that O is a fixed basepoint in $\mathcal{R}$ and that γ is an oriented path in $\mathbb{D}$ from O to

gO, $g \in \Gamma$, which avoids all vertices of $\mathcal{T}$, passing through in order adjacent regions $\mathcal{R} = g_0\mathcal{R}, g_1\mathcal{R}, \ldots, g_n\mathcal{R} = g\mathcal{R}$. Then the labels of the sides crossed by γ, read in such a way that if γ crosses from $g_{i-1}\mathcal{R}$ into $g_i\mathcal{R}$ we read off the label $e_i = g_{i-1}^{-1}g_i$ of the common side interior to $g_i\mathcal{R}$, are in order $e_1, e_2, \ldots, e_n$ so that $g = e_1e_2\cdots e_n$. This proves the well known fact that Γ_0 generates Γ, see for example [**2**]. In particular, if we read off labels round a small loop round vertex v of $\mathcal{R}$, we obtain the *vertex cycle* at v with corresponding *vertex relation* $e_1e_2\cdots e_n = \mathrm{id}$. The generating sets implicit in Theorem B are obtained in this way.

Following [**5**], the fundamental domain $\mathcal{R}$ is said to have *even corners* if for each side s of $\mathcal{R}$, the complete geodesic in $\mathbb{D}$ which extends s is contained in the sides of $\mathcal{T}$. This condition is satisfied for example, by the regular $4g$-gon of interior angle $\pi/2g$ whose sides can be paired with the generating set of Theorem A to form a surface of genus g. If the path γ from O to gO described above is a hyperbolic geodesic and $\mathcal{R}$ has even corners, then the corresponding representation of g by the word $e_1e_2\cdots e_n$ is shortest in (Γ, Γ_0), see [**3, 20**].

Let $|\partial\mathcal{R}|$ denote the number of sides of $\mathcal{R}$. In [**5**] we showed that, subject to certain restrictions if $|\partial\mathcal{R}| \leqslant 4$, one can associate to any fundamental domain with even corners an alphabet $\mathcal{A}$ and a transition matrix P, so that $\mathcal{A}$ is mapped by a finite-to-one map π onto Γ_0, in such a way that the obvious extension of π to a map from the set of finite sequences with alphabet $\mathcal{A}$ and allowed transitions defined by P to the group Γ is surjective. We call this map π, the *alphabet map*. A crucial feature of the alphabet map is that every word in its image is shortest in the word metric on (Γ, Γ_0), see [**20**] and Theorem 2 below. In particular π preserves length, that is, the image under π of a sequence of n symbols in $\mathcal{A}$ is an element $g \in \Gamma$ of shortest length n relative to the generators Γ_0. Thus to sum over $g \in \Gamma$ for which $|g| = n$ as required by Theorems A and B, we need only sum over all allowable finite words of length n in the alphabet $\mathcal{A}$.

It follows that in order to apply Proposition 1, we need only check whether the transition matrix P is irreducible and strictly irreducible and that π is, in a precise sense, almost bijective to Γ. Our main work is to show that (subject to some restrictions if $|\partial\mathcal{R}| < 5$) this is indeed the case, see Propositions 9, 13 and 15.

Notice that the statements of Theorems A and B only involve enumerating words in (Γ, Γ_0) and are independent of the precise geometry of $\mathcal{R}$. Thus for example one can replace the regular $4g$-gon with any hyperbolic octagon whose interior angles sum to 2π and with generators given by the same combinatorial pattern of side pairings. We elaborate on this observation in Section 2·1·2, where we explain why requiring a generating set which comes from a fundamental domain with even corners is much more general that it appears, leading to the general statement in Theorem B.

2·1. *The coding*

We briefly review the coding as explained in [**20**], in which it appears in a simpler and more general form than in the original version [**5**].

By abuse of notation from now on we think of the tessellation $\mathcal{T}$ as the union of its sides, precisely $\mathcal{T} = \bigcup\{g(\partial\mathcal{R}) : g \in \Gamma\}$. We assume throughout our discussion that $\mathcal{R}$ has even corners, so that $\mathcal{T}$ is a union of complete geodesics in $\mathbb{D}$. Let $\mathcal{P} \subset \partial\mathbb{D}$ be the collection of endpoints of those geodesics in $\mathcal{T}$ which meet $\partial\mathcal{R}$ (crucially this includes lines which meet $\partial\mathcal{R}$ only in a vertex of $\mathcal{R}$). The points of $\mathcal{P}$ partition $\partial\mathbb{D} - \mathcal{P}$ into connected open intervals I; we denote the collection of all these intervals by $\mathcal{I}$.

Let $s = s(e)$ be the side of $\mathcal{R}$ whose *exterior* label is e. The extension of s into a complete geodesic lies in $\mathcal{T}$, separating $\mathbb{D}$ into two half planes, one of which contains the interior of $\mathcal{R}$ and one of which contains the interior of $e\mathcal{R}$, see Figure 1. Let $L(e)$ denote the open arc on $\partial\mathbb{D}$ bounding the component which contains $e\mathcal{R}$, see Figure 3. Each interval $I \in \mathcal{I}$ is contained in $L(e)$ for at least one and at most two sides of $\partial\mathcal{R}$ (see Lemma 5 below for a full justification of this fact). For each $I \in \mathcal{I}$, choose $e = e(I) \in \Gamma_0$ such that $I \subset L(e)$. We define a map $f\colon \partial\mathbb{D} - \mathcal{P} \to \partial\mathbb{D}$ by $f(x) = e(I)^{-1}(x)$ for $x \in I$. Extend f to a (discontinuous) possibly two valued map on $\mathcal{P}$ in the obvious way. As observed in [**5**, **20**], the map f is *Markov* in the sense that for any $J \in \mathcal{I}$, $f(I) \cap J \neq \varnothing$ implies that $f(I) \supset J$. To see this, it is clearly sufficient to show that $f(\mathcal{P}) \subset \mathcal{P}$, independently of which of the possibly two choices we make for f. So suppose $\xi \in \mathcal{P}$ is an endpoint of an interval $I \subset L(e)$ and that $f(\xi) = e(I)^{-1}(\xi)$. Write $e = e(I)$. From the definitions, ξ is an endpoint of a geodesic t which meets the closure of the side $s = s(e)$ of $\mathcal{R}$. From the definition of the labelling, $e^{-1}(s(e)) = s(e^{-1})$ is also a side of $\mathcal{R}$. Hence $f(\xi) = e^{-1}(\xi)$ is an endpoint of $e^{-1}(t)$ and $e^{-1}(t)$ must meet the closure of $s(e^{-1})$, hence $f(\xi) \in \mathcal{P}$.

2·1·1. *The alphabet map*

We define our alphabet by setting $\mathcal{A} = \mathcal{I}$, in other words, $\mathcal{A}$ is the collection of all the intervals defined by the subdivision of $\partial\mathbb{D}$ by points in $\mathcal{P}$. We define a transition matrix $P = (p_{I,J})$ by $p_{I,J} = 1$ if $f(I) \supset J$ and $p_{I,J} = 0$ otherwise. Let Σ_F denote the set of finite sequences $I_{i_0} \cdots I_{i_n}$ with $I_{i_r} \in \mathcal{I}$ such that $p_{I_{i_{r-1}},I_{i_r}} > 0$ for all $r = 0, \ldots, n-1$. Thus Σ_F consists of all allowable finite sequences in the subshift on the symbols $I \in \mathcal{I}$ with transition rule specified by P.

The *alphabet map* $\pi\colon \mathcal{I} \to \Gamma_0$ is the map which associates to each $I \in \mathcal{I}$ the element $e \in \Gamma_0$ corresponding to our choice of e for which $I \subset L(e)$, equivalently for which $f_{|I} = e^{-1}$. This extends in an obvious way to a map $\pi : \Sigma_F \to \Gamma$. Recall that a product of n elements of Γ_0 is *shortest* (with respect to the generators Γ_0) if it cannot be expressed as a product of less than n elements of Γ_0. An important feature of the coding is the following result which follows from [**19**, theorem 5·10 and corollary 5·11], see also [**2**, theorem 2·8].

THEOREM 2. *Suppose that $\mathcal{R}$ has even corners and that either*:

(i) $|\partial\mathcal{R}| \geqslant 5$; *or*

(ii) $|\partial\mathcal{R}| = 4$ *and, if in addition all vertices of $\mathcal{R}$ lie in $\mathbb{D}$, then at least three geodesics in $\mathcal{T}$ meet at each vertex*; *or*

(iii) $|\partial\mathcal{R}| = 3$ *and at least one vertex of $\mathcal{R}$ is on $\partial\mathbb{D}$.*

Then the alphabet map π is surjective to Γ. Moreover every word in $\pi(\Sigma_F)$ is shortest with respect to the geometric generators Γ_0 associated to $\mathcal{R}$ as above, and each element $g \in \Gamma$ has a unique *representation in $\pi(\Sigma_F)$.*

In what follows, we always assume that $\mathcal{R}$ satisfies one of the hypotheses of Theorem 2. Uniqueness means that any $g \in \Gamma$ has a unique representation as a word $e_{i_1} \cdots e_{i_n}$ in the image of π; however we may have two distinct sequences $I_{i_1} \cdots I_{i_n}$ and $I_{j_1} \cdots I_{j_n}$ with $\pi(I_{i_r}) = \pi(I_{j_r})$ for all r. A key point in the proof of Theorem B is to show that π is nevertheless almost injective, precisely:

PROPOSITION 3. *Let $g \in \Gamma$. Then $\pi^{-1}(g) \leqslant 2$ and $\#\{g\colon |g| = n, \pi^{-1}(g)| > 1\}/n \to 0$ as $n \to \infty$.*

2·1·2. *Ubiquity of even corners*

The condition that $\mathcal{R}$ have even corners may seem very special. However our result depends only on the combinatorics of the generating set, so that regions which do not have even corners may still have side pairings which satisfy the required conditions. More precisely, we note the following facts:

(i) many standard fundamental domains, for example the symmetrical $4g$-gon for a surface of genus g, and the standard fundamental domain for $SL(2, \mathbb{Z})$, do have even corners. Moreover the condition depends only on the geometry of $\mathcal{R}$ and not on the pattern of side pairings, so that for example the symmetrical $4g$-gon with opposite sides paired would work equally well;

(ii) a simple observation going back to Koebe shows that the group corresponding to any closed hyperbolic surface has a fundamental domain with even corners. To see this we have only to choose smooth closed geodesics for the sides of $\mathcal{R}$. This is always possible; see [**3**] for a picture;

(iii) if Γ has no torsion but contains parabolics then one can always choose a fundamental polygon with all vertices on $\partial\mathbb{D}$. Such a polygon certainly has even corners (and in fact Γ is then a free group);

(iv) we showed in [**5**] that every finitely generated Fuchsian group is quasiconformally equivalent to one which has a fundamental domain with even corners. The deformation can be chosen to preserve the combinatorial pattern of sides and side pairings of $\mathcal{R}$. Since our results only depend on the group and not on the specific hyperbolic structure, it is sufficient to work with the deformed group for which the fundamental domain does have the even corner property;

(v) we show in [**17**] that, subject to hypotheses essentially the same as those of Theorem 2, one can always find a partition of $\partial\mathbb{D}$ and a map f whose combinatorial properties are identical to those which pertain when $\mathcal{R}$ has even corners. One could work directly in this setting, but with the disadvantage that without the geometry of $\mathcal{R}$ at ones disposal it would be much harder to follow what is going on.

(vi) Despite the above comments, one should be clear that our results depend heavily on the choice of $\mathcal{R}$ and the geometrical generating set Γ_0.

Remark 4. Let Σ denote the space of all infinite sequences $I_{i_0}I_{i_1}\ldots$ allowed by the transition matrix P. Let $\Lambda(\Gamma)$ denote the limit set of Γ. We showed in [**17**] that, modulo the exceptional cases excluded by Theorem 2, the obvious map defined by "f-expansions" induces a surjection $\pi(\Sigma) \to \Lambda(\Gamma)$ which is injective except on a countable number of points where it is two-to-one. (The exceptional points are essentially the endpoints of infinite special chains, see [**17**] Proposition 4·6 and Section 2·4 below.)

If Γ contains no parabolics, then we show in [**17**] that Hausdorff measure in dimension δ, where δ is the exponent of convergence of Γ, lifts to a Gibbs measure on Σ. In this case, Hausdorff measure is the so-called Patterson measure on $\Lambda(\Gamma)$. In particular, if $\mathbb{D}/\Gamma$ is compact then Lebesgue measure on $\partial\mathbb{D}$ is Gibbs. In [**18**] we studied random walks on the Cayley graph of Γ. We showed that if the transition probabilities are finitely supported on Γ, then hitting distribution on $\Lambda(\Gamma)$ is Gibbs. Note however that the obvious actions of Γ on these spaces are not measure preserving, so our ergodic theorem does not apply in this context.

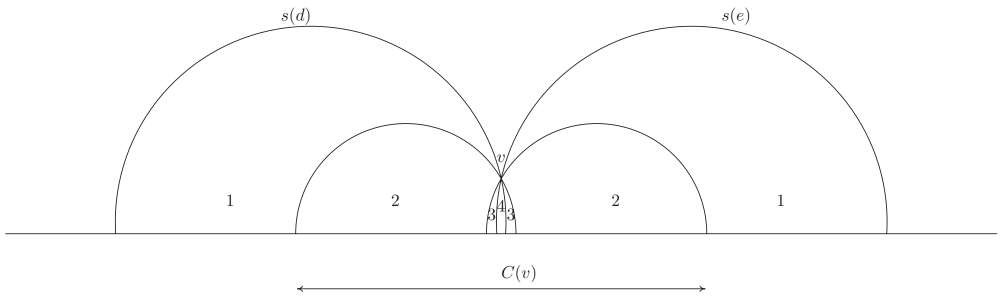

Fig. 2. Intervals round a vertex v of $\mathcal{R}$. In the case shown, $n(v) = 4$, numbers indicating the levels. All intervals except those of level 0 and 1 are contained in $\mathcal{I}$. The crown $C(v)$ is the union of all intervals of levels $2, \ldots, n(v)$.

2·2. *Irreducibility*

In this section we show that the transition matrix P associated to the map f is irreducible, which unfortunately means delving in some detail into its mechanics. The starting point is the following simple but crucial result which is [**4**, lemma 2·2].

LEMMA 5. *Suppose that $|\partial\mathcal{R}| > 3$. Then any two distinct complete geodesics in $\mathcal{T}$ which meet $\partial\mathcal{R}$ are either disjoint in $\mathbb{D} \cup \partial\mathbb{D}$, or intersect in a vertex of $\mathcal{R}$.*

The most interesting case is when lines $t, t' \subset \mathcal{T}$ meet $\partial\mathcal{R}$ in the two vertices at the opposite ends of some side s, neither being coincident with s. The lemma asserts that t and t' are disjoint. Note also that there is a choice in the definition of f only if $L(e) \cap L(d) \neq \emptyset$. It follows from Lemma 5 that this occurs only if $s(e)$ and $s(d)$ are adjacent sides of $\mathcal{R}$.

Now we establish some notation. Let v be a vertex of $\mathcal{R}$. Let $n(v)$ be the number of geodesics of $\mathcal{T}$ which meet at v (so that $2n(v)$ copies of $\mathcal{R}$ meet at v). The endpoints on $\partial\mathbb{D}$ of the $n(v)$ complete geodesics in $\mathcal{T}$ which meet at v partition their complement in $\partial\mathbb{D}$ into $2n(v)$ open intervals, each of which is (the interior of the closure of) a union of intervals $J \in \mathcal{I}$. We assign to each of these new intervals a level, denoted lev(.), as follows. The interval $L(d) \cap L(e)$ bounded by the endpoints of the extensions of sides $s(d)$ and $s(e)$ of $\mathcal{R}$ which meet at v has level $n(v)$. The interval opposite $L(d) \cap L(e)$ has level 0. The remaining intervals going round in both directions from 0 to $n(v)$ have in order levels $1, 2, \ldots, n(v)-1$, see Figure 2. It follows from Lemma 5 that the intervals of all levels except 0 and 1 are also intervals in the set $\mathcal{I}$, and moreover that there is a choice in the definition of f only on intervals of level $n(v)$. Finally define the *crown* of v, denoted $C(v)$, to be the interior of the closure of the union of the intervals of levels $n(v), n(v)-1, \ldots, n(v)-2$ at v, see Figure 2. Thus $C(v)$ is an open interval on $\partial\mathbb{D}$. Notice that if $v \in \partial\mathbb{D}$ then $C(v) = \emptyset$.

For each $e \in \Gamma_0$ define ∂e to be the two vertices of $\mathcal{R}$ at the ends of the side $s(e)$, and let $C(e) = \overline{C(v) \cup C(w)}$ where $\partial e = \{v, w\}$. Also let $M(e) = L(e) - \overline{\bigcup_{d \neq e} L(d)}$ and $A(e) = L(e) - \overline{C(\partial e)}$, see Figure 3. If $|\partial\mathcal{R}| > 3$ then Lemma 5 implies that $A(e) \neq \emptyset$. Note that if $x \in M(e)$ then $f(x) = e^{-1}(x)$.

We are finally ready to start our proof that the transition matrix P is irreducible. In what follows we shall say that a constant depends only on $\mathcal{R}$, when we mean that it depends on $\mathcal{R}$ and the combinatorial pattern of side pairings of $\mathcal{R}$. We introduce various such constants and denote all of them by K.

LEMMA 6. *Suppose $|\partial\mathcal{R}| > 3$. Then there exists $K \in \mathbb{N}$, depending only on $\mathcal{R}$, such that for any $I \in \mathcal{I}$, we have $f^r(I) \supset A(e)$ for some r with $0 \leqslant r \leqslant K$ and some $e \in \Gamma_0$.*

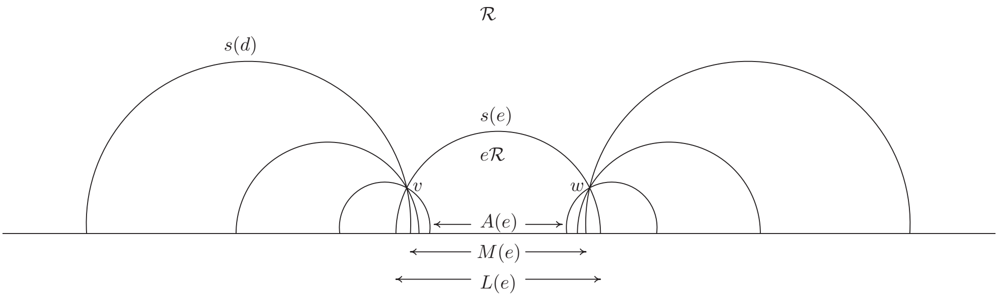

Fig. 3. Intervals subtended by a side of $\mathcal{R}$. If the vertices of $s(e)$ are v, w then $A(e)$ is the subset of $L(e)$ which is contained in neither $C(v)$ nor $C(w)$, while $M(e)$ consists of all of $L(e)$ except the 'ambiguous' top level intervals at either end of $L(e)$. Thus $f_{|M(e)} = e^{-1}$ while $f_{|L(e)\cap L(d)}$ can have value either e^{-1} or d^{-1}.

Proof. If I is not already of the form $A(e)$ for some $e \in \Gamma_0$, then it is in the crown of some vertex v of $\mathcal{R}$, and hence $\mathrm{lev}(I) = r > 1$. Suppose that $I \subset L(e)$ and that $f_{|I} = e^{-1}$. Then f carries $s(e)$ to the side $e^{-1}(s) = s(e^{-1})$, so that $f(v) = e^{-1}v$ is a vertex of $s(e^{-1})$. Following the discussion at the start of Section 2, the cyclic order of labels around the vertices v and $f(v)$ is the same, and by inspection one sees that $f(I)$ is an interval of level $r-1$ at $e^{-1}v$. Since the sets $A(e)$ are contained in the union of level 1 sets associated to all vertices of $\mathcal{R}$, the result follows.

LEMMA 7. *Let $e \in \Gamma_0$. Then*:
(i) $f(M(e)) \supset C(v)$ *for any vertex* $v \notin \partial e^{-1}$;
(ii) $f(M(e)) \supset A(d)$ *for any* $d \neq e^{-1}$.

Proof. By definition, $f_{|M(e)} = e^{-1}$ so that f carries $s(e)$ to $s(e^{-1})$. Moreover e^{-1} carries $L(e)$ to the complement in $\partial\mathbb{D}$ of $L(e^{-1})$. We need to find the image under e^{-1} of $M(e) \subset L(e)$. Let V, W denote the endpoints of $L(e^{-1})$ on $\partial\mathbb{D}$ and let V_1, W_1 denote the points in $\mathcal{P}$ adjacent to V, W and outside $L(e^{-1})$. Then $e^{-1}(M(e))$ is the interval on $\partial\mathbb{D}$ bounded by V_1, W_1 and not containing $L(e^{-1})$. This covers all of $\partial\mathbb{D}$ except for $A(e^{-1})$ and parts of $C(\partial e^{-1})$.

LEMMA 8. *Suppose that $|\partial\mathcal{R}| > 3$. Then there exists $K \in \mathbb{N}$, depending only on $\mathcal{R}$, such that for any $e \in \Gamma_0$, and any $I \in \mathcal{I}$ which is contained in $M(e)$, we have $f^r(A(e)) \supset I$ for some $r \leqslant K$.*

Proof. By definition, $f_{|A(e)} = e^{-1}$, and e^{-1} maps $e\mathcal{R}$ (which is the copy of $\mathcal{R}$ adjacent to $\mathcal{R}$ along $s = s(e)$) to $\mathcal{R}$. The endpoints of $A(e)$ are the endpoints on $\partial\mathbb{D}$ of the extensions of the sides of $e\mathcal{R}$ adjacent to s, see Figure 3. Thus the endpoints of $f(A(e))$ are the endpoints on $\partial\mathbb{D}$ of the extensions of the sides of $\mathcal{R}$ adjacent to $e^{-1}(s) = s(e^{-1})$. If these sides are $s(d)$, $s(d')$, then provided that $e \neq e^{-1}$, we see that $f(A(e))$ is the complement in $\partial\mathbb{D}$ of $L(e^{-1}) \bigcup L(d) \bigcup L(d')$. This gives the result (with $r = 1$) in the case in which $s(e)$ is neither equal to nor adjacent to $s(e^{-1})$. Since in both of the exceptional cases e is necessarily elliptic, this in particular proves the result whenever Γ is torsion free.

Now suppose that $s(e)$ and $s(e^{-1})$ are adjacent with common vertex $v \in \mathbb{D}$, so that e is elliptic with fixed point v. Reasoning as above, we see that $f(A(e))$ covers $M(x)$ for any side $s(x)$ neither equal nor adjacent to $s(e^{-1})$. Fix one such x, which is possible since $|\partial\mathcal{R}| > 3$.

Since by our assumption $s(e)$ and $s(e^{-1})$ are adjacent, we have $x \neq e, e^{-1}$. By Lemma 7 (i), $f(M(x))$ covers all crowns except $C(w)$ for $w \in \partial x^{-1}$. In particular, $f(M(x)) \supset C(v)$. Let $I \in \mathcal{I}$ be the level 2 interval contained in $C(v) \cap L(e)$. Then $f(I)$ is the level 1 interval in $C(v) \cap L(e)$, which is equal to $L(e) - C(v)$. Thus suitable powers of f applied to $C(v)$ cover all of $L(e) \supset M(e)$ which gives the result.

Now consider the case $e = e^{-1}$. First assume that ∂R has at least 5 sides. (Remember we count the edge containing the fixed point of e as one side.) In this case, there exist x, y, distinct from each other and from e, such that $f(A(e))$ covers $M(x)$ and $M(y)$. Now $f(M(x))$ covers $A(e)$ since $e \neq x^{-1}$. In addition, $f(M(x))$ and $f(M(y))$ together cover all crowns except for those crowns $C(w)$ with $w \in \partial x^{-1} \cap \partial y^{-1}$. This implies that $f(M(x)) \cup f(M(y))$ covers $C(v)$ for $v \in \partial e$, which gives the result.

Finally, suppose that ∂R has 4 sides. In this case, $f(A(e))$ only covers $M(x)$ for x the side opposite e. As usual, $f(M(x))$ covers $A(e)$. If $x = x^{-1}$ then $f(M(x))$ covers $M(e)$. Otherwise, x^{-1} is adjacent to e and $f(M(x))$ covers $M(y)$ where y is the fourth side of ∂R (opposite x^{-1}). In this case we have $y = y^{-1}$. Letting v, w be the vertices of $s(e)$ adjacent to $s(x^{-1})$ and $s(y)$ respectively, we see that $f(M(y))$ covers $C(v)$ and $f(M(x))$ covers $C(w)$. The result follows.

PROPOSITION 9. *The Markov chain P is irreducible.*

Proof. We have to show that there exists K, depending only on $\mathcal{R}$, such that for any $I, J \in \mathcal{I}$, we have $f^r(I) \supset J$ for some $0 \leqslant r \leqslant K$.

Assume first that $|\partial R| > 3$. By Lemma 6, we may as well assume that $I = A(e)$ for some $e \in \Gamma_0$. By Lemma 8, it will be enough to show that images of $M(e)$ cover $\partial\mathbb{D}$. By Lemma 7, $f(M(e))$ covers all crowns except $C(\partial e^{-1})$ and all sets $A(x)$ with $x \neq e^{-1}$. Since $|\partial\mathcal{R}| > 3$ we may choose x, y distinct from each other and from e and e^{-1} such that $f(M(e)) \supset A(x) \cup A(y)$. By Lemmas 7 and 8, there exists $r < K$ such that $f^r(A(x))$ covers $A(e^{-1})$ and all crowns except $C(\partial x^{-1})$. Likewise $f^s(A(y))$ covers all crowns except $C(\partial y^{-1})$ for some $s < K$. Now by choice x, y and e are distinct and so $C(\partial x^{-1}) \cap C(\partial y^{-1}) \cap C(\partial e^{-1}) = \emptyset$. The result follows.

Finally, we have to consider the case in which $|\partial R| = 3$. Notice that this is the only case in which it is possible that $A(e) = \emptyset$. By hypothesis, at least one vertex of $\mathcal{R}$ is on $\partial\mathbb{D}$. There are only three possible cases:

(i) $\mathcal{R}$ has three vertices on $\partial\mathbb{D}$;

(ii) $\mathcal{R}$ has two vertices v, w on $\partial\mathbb{D}$. The side joining v, w is paired to itself by an order two elliptic x; the remaining two sides are paired to each other by an elliptic e with fixed point at the third (finite) vertex u;

(iii) $\mathcal{R}$ has one vertex v on $\partial\mathbb{D}$. The two sides meeting at v are paired to each other by e, the third side is paired to itself by an order two elliptic x.

Case (ii). Set $B(e^{\pm}) = L(e^{\pm}) - C(u)$. Note that $f(B(e^{\pm})) = L(x)$ and $f(L(x)) = \partial\mathbb{D} - L(x)$. Furthermore, for each $I \subset C(u)$ it is clear that $f^r(I) = (B(e^{\pm}))$ for some $r \leqslant n(u)$. This proves the result.

Case (iii). Let u be the finite endpoint of side e and let $J(u) = C(u) - L(e)$. Define $J(w)$ similarly relative to w the finite endpoint of e^{-1}. Note that $f(J(u)) \supset A(e^{-1})$ and $f(J(w)) \supset A(e)$. It follows that the image of every interval in $C(u) \cup C(v)$ eventually covers either $A(e)$ or $A(e^{-1})$. Further, a bounded image of $A(e)$ covers $L(e) \cup J(u)$ and a bounded image of $A(e^{-1})$ covers $L(e^{-1}) \cup J(w)$. The result follows.

Case (i) is easier and is left to the reader.

2·3. *Strict irreducibility*

We now investigate strict irreducibility of the transition matrix P. It is well known and easy to see that P is strictly irreducible if the equivalence relation $\sim$ on $\mathcal{I}$ generated by $I \sim J$ if $f(I) \cap f(J) \neq \emptyset$ has just one equivalence class. We show that if $|\partial\mathcal{R}| > 4$ then f is always strictly irreducible, while if $|\partial\mathcal{R}| \leqslant 4$ the map f may or may not be strictly irreducible depending on the precise arrangement of $\mathcal{R}$ and its side pairings. In particular, the continued fraction map associated to the standard fundamental domain for $SL(2, \mathbb{Z})$ is *not* strictly irreducible.

LEMMA 10. *The Markov chain associated to any choice of Markov map f for the fundamental domain $|z| > 1$, $-1/2 < \Re z < 1/2$ is not strictly irreducible.*

Proof. The continuations of the sides of $\partial\mathcal{R}$ through the two vertices at $(1 \pm \sqrt{3}i)/2$ meet the real axis $\mathbb{R}$ in the 7 points $-2, -1, -1/2, 0, 1/2, 1, 2$ which partition $\mathbb{R}$ into 8 intervals which we number $\mathbf{1}, \mathbf{2}, \ldots, \mathbf{8}$ in order from left to right, thus for example $\mathbf{3}$ denotes the interval $(-1, -1/2)$. The map f is defined as:

$$f(x) = x + 1 \text{ for } x \in \mathbf{1} \cup \mathbf{2},$$
$$f(x) = -1/x \text{ for } x \in \mathbf{3} \cup \mathbf{4} \cup \mathbf{5} \cup \mathbf{6},$$
$$f(x) = x - 1 \text{ for } x \in \mathbf{7} \cup \mathbf{8}.$$

There is a choice for f on the overlap regions $\mathbf{4}$ and $\mathbf{5}$: for definiteness we have taken the usual choice $f(x) = -1/x$ for $x \in \mathbf{4} \cup \mathbf{5}$ which is associated to the continued fraction map.

It is easy to write down the transition matrix P for f. We find $\mathbf{1} \to \mathbf{1} \cup \mathbf{2}$, $\mathbf{2} \to \mathbf{3} \cup \mathbf{4}$, $\mathbf{3} \to \mathbf{7}$, $\mathbf{4} \to \mathbf{8}$, $\mathbf{5} \to \mathbf{1}$, $\mathbf{6} \to \mathbf{2}$, $\mathbf{7} \to \mathbf{5} \cup \mathbf{6}$, and $\mathbf{8} \to \mathbf{7} \cup \mathbf{8}$. From this we easily see that there are four equivalence classes under $\sim$: $\{\mathbf{3}, \mathbf{4}, \mathbf{8}\}$, $\{\mathbf{5}, \mathbf{6}, \mathbf{1}\}$, $\{\mathbf{2}\}$ and $\{\mathbf{7}\}$. This gives the result. We remark that even had we made the other choice for f on either of $\mathbf{3}$ and $\mathbf{6}$, then there are still at least two equivalence classes.

Remark 11. Even though strict irreducibility of P fails in this case, it is still possible to prove the required convergence of Cesàro averages. Combine the above intervals into four larger ones: $J_1 = \mathbf{1} \cup \mathbf{2}$, $J_2 = \mathbf{3} \cup \mathbf{4}$, $J_3 = \mathbf{5} \cup \mathbf{6}$ and $J_4 = \mathbf{7} \cup \mathbf{8}$. The map f is still Markov with respect to the new partition $\mathfrak{J} = \{J_i\}$ and the alphabet map π factors through the obvious map from $\mathfrak{I}$ to $\mathfrak{J}$. The transition matrix for $\mathfrak{J}$ is again irreducible but not strictly irreducible. However it is easily seen to be aperiodic and this is enough for Proposition 1, see [**6**, proposition 1].

The alphabet map $\pi(J_1) = S^{-1}$, $\pi(J_2) = \pi(J_3) = Q$, $\pi(J_4) = S$ gives the well-known representation of words in $PSL(2, \mathbb{Z})$ in the form $\cdots S^{n_i} Q S^{-n_{i+1}} Q S^{n_{i+2}} Q \cdots$ where $n_i > 0$ for all i and the sequence may begin and end in any of the 3 symbols $S^{\pm}$, Q. The coding has to divide Q into two states J_2, J_3 in order to prevent transitions of the form SQS or $S^{-1}QS^{-1}$.

Another interesting example is furnished by the group $\Gamma = \langle a, b, c : a^2 = b^2 = c^2 \rangle$ where $\mathcal{R}$ is the ideal triangle with vertices $0, 1, \infty$ and $a, b, c \in PSL(2, \mathbb{Z})$ are elliptics of order two with fixed points at i, $(1+i)/2$ and $1+i$ respectively. In this case one checks that f *is* strictly irreducible.

We base our general proof of strict irreducibility on the following lemma.

LEMMA 12. *Suppose that there exists a family of open intervals* $J_0, \dots, J_m \in \mathcal{I}$ *such that*:

(i) $\bigcup_{i=0}^{m} f(J_i)$ *covers* $\partial\mathbb{D} - \mathcal{P}$;
(ii) $f(J_i) \cap f(J_{i+1}) \neq \emptyset$ *for* $i = 0, \dots, m-1$.

Then the Markov chain associated to the Markov map f *is strictly irreducible.*

Proof. By assumption (i), every interval $I \in \mathcal{I}$ is equivalent to at least one of the J_i. By assumption (ii), $J_i \sim J_{i+1}$ for all $0 \leqslant i < m$. The result follows.

PROPOSITION 13. *Suppose that* $|\partial\mathcal{R}| \geqslant 5$. *Then the Markov chain associated to the Markov map* f *is strictly irreducible.*

Proof. We show the sets $A(e)$, $e \in \Gamma_0$, satisfy the requirements of Lemma 12. Suppose that the extensions of the sides of $\mathcal{R}$ adjacent to e^{-1} meet $\partial\mathbb{D}$ in points V and W. Then $f(A(e))$ is the interval between V and W and not containing $L(e^{-1})$. Since $|\partial\mathcal{R}| \geqslant 5$, this set of intervals overlaps round $\partial\mathbb{D}$ in the required manner.

If $|\partial\mathcal{R}| = 4$, then f may or may not be strictly irreducible. For example, suppose that $\mathcal{R}$ has two opposite vertices $v, w \in \partial\mathbb{D}$ while the other opposite pair are in $\mathbb{D}$. Suppose the sides adjacent to v are paired, and equally the sides adjacent to w. Then one can verify directly that $\sim$ has two equivalence classes. The idea is that the points v and w divide $\partial\mathbb{D}$ into two halves E and F say. One checks easily that the image of every interval I is contained either completely in E, or completely in F, so the intervals whose images fall in these two halves cannot be equivalent.

On the other hand, if $\mathcal{R}$ has 4 sides all of which lie in $\mathbb{D}$, then by hypothesis we assume that at least three geodesics in $\mathcal{T}$ meet at each vertex. One can check that the images of the level one intervals at each vertex cover $\partial\mathbb{D}$ in the manner required by Lemma 12.

We have already studied similar phenomena when $|\partial\mathcal{R}| = 3$.

2·4. *The alphabet map*

Finally we prove Proposition 3. We begin by recalling some further terminology from [**3**, **5**, **20**].

Let $e_{i_0} \cdots e_{i_n}$ be a word in the generators Γ_0. Since Γ_0 consists of side pairing transformations of $\mathcal{R}$, the regions $\mathcal{R}$ and $e_{i_r}\mathcal{R}$, and more generally $e_{i_0} \cdots e_{i_{r-1}}\mathcal{R}$ and $e_{i_0} \cdots e_{i_r}\mathcal{R}$ for $0 < r < n$, have a common side. The word $e_{i_0} \cdots e_{i_n}$ is called a *cycle* if in the tesselation of $\mathbb{D}$ by images of $\mathcal{R}$, the regions $\mathcal{R}$, $e_{i_0}\mathcal{R}$, $e_{i_0}e_{i_1}\mathcal{R}, \dots, e_{i_0} \cdots e_{i_n}\mathcal{R}$ are arranged in order round a common vertex $v \in \mathbb{D}$, see Section 2. (According to this definition, a single letter e is always a cycle provided that at least one of the vertices ∂e is in $\mathbb{D}$.) Cycles $e_{i_0} \cdots e_{i_s}$, $e_{j_0} \cdots e_{j_t}$ are called *consecutive* if there exists $e \in \Gamma_0$ such that $e_{i_0} \cdots e_{i_s} e$ and $e^{-1} e_{j_0} \cdots e_{j_t}$ are both cycles, see [**3**] for more details. This means that $e_{i_0} \cdots e_{i_s}$ is a cycle at v and that $e_{j_0} \cdots e_{j_t}$ is a cycle with the same orientation at w, where v and w are the endpoints of the side e^{-1} of $\mathcal{R}$, see Figure 4.

The word $e_{i_0} \cdots e_{i_N}$ is called a *special chain* if it consists of a sequence of consecutive cycles $B_1 B_2 \cdots B_n$ at vertices $v_1, \dots, v_n$ such that B_1 has length at most $n(v_1) - 1$, B_n has length at most $n(v_n)$ and B_i has length exactly $n(v_i) - 1$ for $1 < i < n$. The geometrical meaning of this definition is that the sequences of copies of $\mathcal{R}$ corresponding to a special chain all touch a common hyperbolic line $t \subset \mathcal{T}$, all except possibly the first or last one lying on the same side of t.

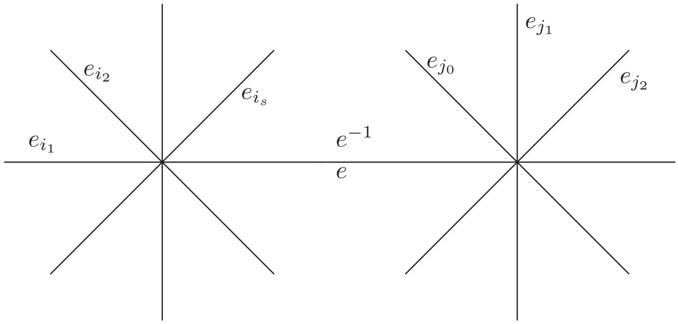

Fig. 4. Consecutive cycles.

Remark 14. Special chains are intimately connected to the solution of the word problem for Fuchsian groups given in [**3**, **17**, **19**]. Special chains are shortest words and two shortest words $V = e_{i_0} \cdots e_{i_N}$ and $W = e_{j_0} \cdots e_{j_N}$ with $e_{i_r} \neq e_{j_r}$ for all r represent the same element of Γ only if either both V and W are single cycles of length $n(v)$, or are both sequences of consecutive cycles of lengths $n(v_1)-1, n(v_2)-1, \ldots, n(v_{n-1})-1, n(v_n)$ and $n(v_1), n(v_2)-1, \ldots, n(v_{n-1})-1, n(v_n)-1$ respectively. In the latter case, the sequences of copies of $\mathcal{R}$ corresponding to the words $e_{i_0} \cdots e_{i_N}$ and $e_{j_0} \cdots e_{j_N}$ meet along a common line in $\mathcal{T}$. A *special chain* is any sequence of the above form.

PROPOSITION 15. *Let* $\pi : \Sigma \to \Gamma$ *be the alphabet map. Then* $\pi^{-1}(e_{i_0} \ldots e_{i_N}) \leqslant 2$ *with equality if and only if* $e_{i_0} \cdots e_{i_N}$ *ends in a special chain.*

Proof of Proposition 3. This is an immediate corollary of Proposition 15. Notice that a special chain is completely specified by its initial two letters (which determine the direction of the cycle at the first vertex) and the length of the initial cycle. It follows that there exists a constant K depending only on $\mathcal{R}$ such that the total number of special chains of length exactly n is bounded by K, independent of n. Since the total number of words in Γ of length n grows exponentially with n, and since a special chain can be continued to arbitrary length, the result follows.

We establish several lemmas before proving Proposition 15.

LEMMA 16. *Suppose that* $\pi(I_{i_0} \cdots I_{i_n}) = \pi(I_{j_0} \cdots I_{j_m})$. *Then* $n = m$ *and* $\pi(I_{i_r}) = \pi(I_{j_r})$ *for* $r = 0, \ldots, n$. *Moreover if* $I_{i_k} \neq I_{j_k}$ *for some* k *then* $I_{i_r} \neq I_{j_r}$ *for any* $r > k$.

Proof. Suppose that $\pi(I_{i_0} \cdots I_{i_n}) = \pi(I_{j_0} \cdots I_{j_m})$. Since the images of both sequences are shortest, $n = m$. Moreover because of unique representation in Γ by sequences in the image of π, see Theorem 2, we have $\pi(I_{i_r}) = \pi(I_{j_r})$ for $r = 0, \ldots, n$.

Now suppose that $I_{i_k} \neq I_{j_k}$. By definition, $f|_I = \pi(I)^{-1}$. Since $\pi(I_{i_k}) = \pi(I_{j_k})$, we see that f is injective on $I_{i_k} \cup I_{j_k}$ and the result follows.

LEMMA 17. *Suppose that* $\pi(I_{i_0} \cdots I_{i_n}) = \pi(I_{j_0} \cdots I_{j_n})$ *with* $I_{i_0} \neq I_{j_0}$, *and suppose that* $0 \leqslant r < n$. *Then*:

(i) *if* $I_{i_r} \subset C(v) \cap L(e)$ *for* $e \in \Gamma_0$ *and* $v \in \partial e$, *then*; *either* $I_{j_r} \subset C(v)$ *or* $I_{j_r} = A(e)$;

(ii) *if* $I_{i_r} = A(e)$ *for* $e \in \Gamma_0$ *then* I_{j_r} *is adjacent to* $A(e)$;

(iii) *if* $I_{i_r}, I_{j_r} \subset C(v) \cap L(e)$ *for* $e \in \Gamma_0$ *and some vertex* $v \in \partial e$, *and if* $\mathrm{lev}(I_{j_r}) <$ $lev(I_{i_r}) = k$ *and* $r + k \leqslant n$, *then* $f^{k-1}(I_{i_r}) = I_{i_{r+k-1}} = A(d)$ *for some* $d \in \Gamma_0$.

Proof.
Assertion (i). Let w be the other vertex in ∂e. If the result is false, then $I_{j_r} \subset C(w) \cap L(e)$. Let $s(x)$ and $s(y)$ be the sides of $\mathcal{R}$ adjacent to $s(e^{-1})$. Note that $f|_{I_{j_r}} = e^{-1}$ so that $f(C(v) \cap L(e)) \subset L(x)$, say, and $f(C(w) \cap L(e)) \subset L(y)$.

By the hypotheses of Theorem 2 either $|\partial\mathcal{R}| > 3$ or $\mathcal{R}$ has a vertex at ∞. In both cases $L(x)$ and $L(y)$ are disjoint. In the first case this is clear by Lemma 5. For the second, observe that we may assume that both vertices in ∂e^{-1} are in $\mathbb{D}$, since otherwise at least one of $C(v)$ and $C(w)$ is empty and there is nothing to prove. This forces $\pi(I_{i_{r+1}}) \nsim \pi(I_{j_{r+1}})$ contrary to hypothesis, which gives the result.
Assertion (ii). Suppose first that $e \nsim e^{-1}$. Observe that the image under e^{-1} of any interval in $I \subset L(e)$ but not adjacent to $A(e)$ is contained in $C(\partial e^{-1})$ and is thus contained in $L(e^{-1}) \cup L(x) \cup L(y)$ where as above $s(x)$, $s(y)$ are the two sides of $\mathcal{R}$ adjacent to $s(e^{-1})$. On the other hand, the image under e^{-1} of $A(e)$ is outside $L(e^{-1}) \cup L(x) \cup L(y)$. Thus provided $L(x)$ and $L(y)$ are disjoint, $\pi(J) \notin \{e, x, y\}$ for any $J \in \mathcal{I}$ contained in $f(A(e))$. The result follows as above if $|\partial\mathcal{R}| > 3$. The special case $|\partial\mathcal{R}| = 3$ is easily treated separately.

Finally suppose that $e = e^{-1}$. If $I \subset C(\partial e)$ then $f(I) \subset C(\partial e)$ but $f(A(e))$ is *outside* $L(e)$ which is impossible.
Assertion (iii). The map f decreases level and at each stage with $t < k$, $I_{i_{r+t}}$ and $I_{j_{r+t}}$ are in the crown of a common vertex. At step $k-1$, $I_{i_{r+k-1}}$ has level 1 so that by (i), $I_{i_{r+k-1}} = A(d)$ for some $d \in \Gamma_0$.

Proof of Proposition 15. Suppose that

$$\pi(I_{i_0} \cdots I_{i_n}) = \pi(I_{j_0} \cdots I_{j_n}).$$

Without loss of generality we may as well assume that $I_{i_0} \nsim I_{j_0}$. By Lemma 16, $I_{i_r} \nsim I_{j_r}$ for any $r > 0$.

Suppose first that $r < n$ and that $I_{i_r} \subset C(v) \cap L(e)$ has level $k > 1$ in the clockwise direction starting at the highest level interval at vertex v. (The proof if the interval is in the anticlockwise direction is similar; obviously the direction of all cycles are then reversed.) Then $I_{i_{r+1}}$ has level $k-1$ at vertex $e^{-1}v$ and is outside $L(e^{-1})$. Thus $f(I_{i_r}) \subset L(e')$ where $s(e')$ is the side of $\mathcal{R}$ adjacent to $s(e^{-1})$ in clockwise order round $\partial\mathcal{R}$. Therefore $e^{-1}e'^{-1} = \pi(I_{i_r})\pi(I_{i_{r+1}})$ is an anticlockwise cycle at $e^{-1}v$.

Inductively, it follows that $\pi(I_{i_r})\pi(I_{i_{r+1}}) \cdots \pi(I_{i_{r+k-1}})$ is an anticlockwise cycle. Further, by Lemma 17 (iii) we see that $I_{i_{r+k-1}} = A(d)$ for some $d \in \Gamma_0$ and that $I_{j_{r+k-1}}$ is adjacent to $I_{i_{r+k-1}}$, where $d^{-1} = \pi(I_{i_{r+k-1}})$.

Let the sides of $\partial\mathcal{R}$ in clockwise order from $s(d)$ be $s(d), s(c), s(b)$ (so that the interior labels of sides in clockwise order are d^{-1}, c^{-1}, b^{-1}). Let v be the common vertex of $s(d)$ and $s(c)$ and let w be the common vertex of $s(c)$ and $s(b)$. From the proof of Lemma 17 (ii), it follows that $I_{j_{r+k}}$ is necessarily the highest level interval at v, and that $\pi(I_{j_{r+k}}) = c$.

Now $\pi(I_{i_{r+k-1}})\pi(I_{j_{r+k}}) = d^{-1}c$ is an anticlockwise cycle at v, hence so also is $\pi(I_{i_r})\pi(I_{i_{r+1}}) \ldots \pi(I_{i_{r+k-1}})c$. Furthermore $c^{-1}\pi(I_{j_{r+k}}) = c^{-1}b$ is an anticlockwise cycle at w. This means the cycles

$$\pi(I_{i_r})\pi(I_{i_{r+1}}) \cdots \pi(I_{i_{r+k-1}}) \text{ and } \pi(I_{j_{r+k}}) = \pi(J_{j_{r+k}})$$

are consecutive. Moreover $I_{i_{r+k}}$ and $I_{j_{r+k}}$ are both contained in the crown at w, in fact $I_{i_{r+k}}$ is the level $n(w)-1$ interval in the crown at w, adjacent to $I_{j_{r+k}}$ going around clockwise.

Now the first part of the argument repeats so that $\pi(I_{j_{r+k}})$ is (assuming n is sufficiently large) the first term in an anticlockwise cycle of length of length $n(w) - 1$. A similar argument in the case $I_{i_0} = A(e)$ completes the proof.

2·5. *An ergodic theorem for Fuchsian groups*

As before, let Γ be a finitely generated non-elementary Fuchsian group acting in the hyperbolic disk $\mathbb{D}$, and assume that a fundamental domain $\mathcal{R}$ for Γ has even corners and satisfies $|\partial\mathcal{R}| \geqslant 5$. As before, let Γ_0 be the generating set corresponding to $\mathcal{R}$. For $g \in \Gamma$, let $|g|$ be the length of the shortest word in Γ_0 representing Γ, and for $n \in \mathbb{N}$, let

$$S(n) = \{g \in \Gamma : |g| = n\}$$

be the sphere of radius n in Γ. Finally, let K_n be the cardinality of $S(n)$. Observe that K_n grows exponentially and so, by the Borel–Cantelli Lemma, the contribution of "non-injective" elements of Proposition 3 to the spherical averages is negligible. Propositions 1, 3 and 13 now imply

THEOREM B. *Let Γ act ergodically on a probability space (X, ν) by measure-preserving transformations, and, for $g \in \Gamma$, let T_g be the corresponding transformation. Suppose Γ_0 is a geometric set of generators associated to a Markov partition as in Section 2, and suppose either that $|\partial\mathcal{R}| \geqslant 5$, or, if $|\partial\mathcal{R}| < 5$, that the associated transition matrix is strictly irreducible. Then, measuring word length with respect to the generators Γ_0, for any $\varphi \in L_1(X, \nu)$ we have*

$$\frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{K_n}\sum_{g\in S(n)}\varphi\circ T_g \longrightarrow \int_X \varphi d\nu \tag{2·1}$$

both ν-almost surely and in $L_1(X, \nu)$ as $N \to \infty$.

Theorem B implies as a special case Theorem A.

In special cases, the averages for groups for which the transition matrix is not strictly irreducible may also converge, see Remark 11.

Acknowledgments. A.I.B. is an Alfred P. Sloan Research Fellow. During work on this project, he was supported in part by Grant MK-4893.2010.1 of the President of the Russian Federation, by the Programme on Mathematical Control Theory of the Presidium of the Russian Academy of Sciences, by the Programme 2.1.1/5328 of the Russian Ministry of Education and Research, by the Edgar Odell Lovett Fund at Rice University, by the National Science Foundation under grant DMS 0604386, and by the RFBR-CNRS grant 10-01-93115.

REFERENCES

[**1**] C. ANANTHARAMAN ET AL. *Théorèmes ergodiques pour les actions de groupes.* Monographie **41** de L' Enseignment Mathématique (Genève, 2010).

[**2**] A. BEARDON. An introduction to hyperbolic geometry. *Ergodic Theory and Symbolic Dynamics in Hyperbolic Spaces, T. Bedford, M. Keane and C. Series eds.* (Oxford University Press, 1991).

[**3**] J. BIRMAN and C. SERIES. Dehn's algorithm revisited, with application to simple curves on surfaces. *Combinatorial Group Theory and Topology, S. Gersten and J. Stallings eds. Ann. of Math. Studies* III (*Princeton* University Press, 1987), 451–478.

[**4**] L. BOWEN. Invariant measures on the space of horofunctions of a word hyperbolic group. *Ergodic Theory Dynam. Syst.* **30** (2010), 97–129.

[**5**] R. BOWEN and C. SERIES. Markov maps associated with Fuchsian groups. *Inst. Hautes Études Sci. Pubi. Math.* **50** (1979), 153–170.

[**6**] A. I. BUFETOV. Convergence of spherical averages for actions of free groups. *Ann. of Math.* (2), **155** (2002), 929–944.

[**7**] A. I. BUFETOV. Markov averaging and ergodic theorems for several operators. *Topology, ergodic theory, real algebraic geometry,* 39–50, *Amer. Math. Soc. Transl. Ser.* 2 **202** (Amer. Math. Soc., 2001).

[**8**] K. FUJIWARA and A. NEVO. Maximal and pointwise ergodic theorems for word-hyperbolic groups. *Ergodic Theory Dynam. Syst.* **18** (1998), 843–858.

[**9**] A. GORODNIK and A. NEVO. The ergodic theory of lattice subgroups. *Ann. of Math. Stud.* **172** (Princeton University Press, 2010).

[**10**] R. I. GRIGORCHUK. Ergodic theorems for the actions of a free group and a free semigroup. (Russian) *Mat. Zametki* **65** (1999), 779–783 (English trans: *Math. Notes* **65** (1999), 654–657).

[**11**] R. I. GRIGORCHUK. An ergodic theorem for actions of a free semigroup. (Russian) *Tr. Mat. Inst. Steklova* **231** (2000), Din. Sist., Avtom. i Beskon. Gruppy, 119–133 (English trans: *Proc. Steklov Inst. Math.* 2000 **231**, 113–127).

[**12**] G. A. MARGULIS, A. NEVO and E. M. STEIN. Analogs of Wiener's ergodic theorems for semisimple Lie groups. II. *Duke Math. J.* **103** (2000), 233–259.

[**13**] A. NEVO. Harmonic analysis and pointwise ergodic theorems for noncommuting transformations. *J. Amer. Math. Soc.* **7** (1994), 875–902.

[**14**] A. NEVO. Pointwise ergodic theorems for actions of groups. *Handbook of dynamical systems. Vol.* 1*B*, 871–982 (Elsevier, Amsterdam, 2006).

[**15**] A. NEVO and E. M. STEIN. A generalization of Birkhoff's pointwise ergodic theorem. *Acta Math.* **173** (1994), 135–154.

[**16**] A. NEVO and E. M. STEIN. Analogs of Wiener's ergodic theorems for semisimple groups. I. *Ann. of Math.* 2 **145** (1997), 565–595.

[**17**] C. SERIES. The infinite word problem and limit sets in Fuchsian groups. *Ergodic Theory Dynam. Syst.* **1** (1981), 337–360.

[**18**] C. SERIES. Martin boundaries of random walks on Fuchsian groups. *Israel J. Math.* **44** (1983), 221–240.

[**19**] C. SERIES. Geometrical Markov coding of geodesics on surfaces of constant negative curvature. *Ergodic Theory Dynam. Syst.* **6** (1986), 601–625.

[**20**] C. SERIES. Geometrical methods of symbolic coding. *Ergodic Theory and Symbolic Dynamics in Hyperbolic Spaces, T. Bedford, M. Keane and C. Series eds.* (Oxford University Press, 1991).